



Table des matières

Chap 1 : Séries numériques	1
I Définition et somme totale	1
I.1 Généralités	1
I.2 Changement de borne / de termes	2
I.3 Calculs	3
II Critères de convergence / divergence	4
II.1 Divergence grossière	4
II.2 Séries à termes de signe constant	4
II.2-a Propriétés élémentaires des séries à termes positifs	4
II.2-b Critère de comp. de deux séries à termes positifs	5
II.2-c Equivalence	5
II.2-d Séries à termes négatifs	7
II.3 Convergence absolue	7
III Exemples fondamentaux	8
III.1 Séries de Riemann $\sum \frac{1}{n}$ et $\sum \frac{1}{n^2}$	8
III.2 Séries géométriques	8
III.3 Série "dérivée" $(\sum nq^{n-1})_{n \in \mathbb{N}}$	9
III.4 Série "dérivée seconde" $(\sum n(n-1)q^{n-2})_{n \in \mathbb{N}}$	10
III.5 Série exponentielle	11
III.6 Application	12
IV Et si on change l'ordre des termes ?	12
Chap 2 : Ensembles et Dénombrement (rappels)	13
I Vocabulaire des ensembles et des applications	13
I.1 Ensembles	13
I.2 Applications	13
II Dénombrement	14

II.1 Notion de cardinal	14
II.2 Formules	14
Chap 3 : Probabilités (Généralités)	15
I Le triplet fondamental	15
I.1 Cas des probabilités finies	15
I.1-a Rappel de définition et propriétés élémentaires	15
I.1-b Problématiques, limites et insuffisances de la théorie de l'univers fini	16
I.2 Cas général : mesure de probabilité	16
I.2-a Réunions et intersections "dénombrables"	16
I.2-b Notion de tribu	17
I.2-c La tribu des Boréliens	17
I.2-d Espace probabilisé	18
I.3 Cohérence avec les propriétés connues	19
II Probabilités conditionnelles	19
II.1 Définition et exemples de calculs	19
II.2 Formules de Bayes	20
II.3 Formule des probabilités composées	21
III Indépendance	21
IV systèmes (quasi)-complets	22
IV.1 Définition	22
IV.2 Formule(s) des probabilités totales	24
V Variables aléatoires	25
V.1 Variables aléatoires quelconques	25
V.2 Loi d'une v.a. par fonction de répartition	26
V.2-a Valeurs de la variable : notion de support	26
V.2-b Fonction de répartition d'une variable aléatoire	27
V.2-c Fonction dite "de répartition"	31
V.3 Indépendance de variables aléatoires	31
V.3-a Cas de deux variables	31
V.3-b De manière générale :	32
Chap 4 : Polynômes	33
I Généralités sur les polynômes	33
I.1 Définitions et vocabulaire de base	33
I.2 Polynôme nul et unicité de l'écriture d'un polynôme	34

I.3 Degré d'un polynôme	34
II Opérations sur les polynômes	34
II.1 Définitions et propriétés élémentaires	34
II.2 Résultat des opérations sur les coefficients	35
II.3 Résultat des opérations sur les degrés	36
III Décomposition d'un polynôme en sous-facteurs	37
III.1 Divisibilité	37
III.2 Racine et divisibilité	37
III.3 Racines dans \mathbb{C} et décomposition dans $\mathbb{C}[X]$	42
Chap 5 : Rappels équations diff. linéaires	45
I Ordre 1 $y' + ay = c$	45
II Ordre 2 à coefficients constants $ay'' + by' + cy = d(x)$	47
Chap 6 : Espaces vectoriels	51
I Généralités	51
I.1 Structure d'un espace vectoriel	51
I.2 Sous-espace vectoriel	52
II Famille génératrice, famille libre et base	54
II.1 S.e.v. engendré (Vect) et famille génératrice	54
II.2 Famille libre	55
II.3 Base	57
II.3-a Définition	57
II.3-b Existence d'une base : extraction d'une famille génératrice	58
III Dimension, rang et recherche de base	58
III.1 Dimension	58
III.2 Rang d'une famille	60
III.3 Théorème de la base incomplète	61
IV Coordonnées	61
IV.1 Unicité de la décomposition dans une base	61
IV.2 Notion de coordonnées	62
IV.3 Changement de base	63
IV.3-a Si on a les coordonnées de u dans \mathcal{B}' et on veut $M_{\mathcal{B}}(u)$	63
IV.3-b Si on a les coordonnées de u dans \mathcal{B} et on veut $M_{\mathcal{B}'}(u)$	64
IV.3-c Propriétés des matrices de passage	65

Chap 7 : Intégrale généralisée	67
I Rappels de notations, vocabulaire et d'intégration sur un intervalle $[a, b]$	67
I.1 Intégration par partie sur un segment	67
I.2 Changement de variable sur une intégrale $\int_a^b f$	67
I.2-a Intégration par substitution et changement de variable $x = \varphi(t)$	68
I.2-b Changement de variable (le vrai!) avec changement de variable $t = \psi(x)$	69
I.3 Primitivation	70
II Convergence d'une intégrale (impropre ou) généralisée	71
II.1 Déf. d'une intégrale généralisée sur un intervalle semi-ouvert	71
II.2 Intégrales faussement impropres	72
II.3 Définition d'une intégrale généralisée sur un intervalle ouvert	73
II.4 Définition d'une intégrale généralisée sur une réunion d'intervalles ouverts	74
III Manipulation des intégrales impropres	75
III.1 Linéarité	75
III.2 IPP	75
III.3 Changement de variable	77
III.3-a Le théorème et son application	77
III.3-b Cas des fonctions paires et impaires	78
III.4 Bilan des principaux résultats	79
IV Fonctions positives et convergence absolue	79
V Exemples fondamentaux	83
Chap 8 : Équations différentielles autonomes	85
I Définition et quelques propriétés	85
II Résolutions graphiques : méthode d'Euler	87
III Exemples de résolutions en dynamique des populations	88
III.1 Modèle de Malthus	88
III.2 Modèle Logistique	88
III.3 Le modèle de Gompertz	89
Chap 9 : Variables aléatoires discrètes	91
I Généralités	91
I.1 Définition	91

I.2 Loi d'une v.a.d.	91
I.3 Fonction de répartition	92
II Exemples fondamentaux	93
II.1 Rappels de Probabilités finies	93
II.1-a Loi uniforme	93
II.1-b Loi binomiale $\mathcal{B}(n; p)$	95
II.2 Loi géométrique	95
II.3 Loi de Poisson	97
III Espérance et variance	99
III.1 Espérance	99
III.2 Moments	101
III.3 Variance	102
III.4 Exemples fondamentaux	103
III.4-a Loi uniforme discrète	103
III.4-b Loi de Bernoulli, $\mathcal{B}(p)$	103
III.4-c Loi Binomiale, $\mathcal{B}(n; p)$	103
III.4-d Loi géométrique sur \mathbb{N}^* ou \mathbb{N}	103
III.4-e Loi de Poisson, $\mathcal{P}(\lambda)$	104
Chap 10 : Couples de va. aléatoires discrètes	107
I Loi	107
I.1 Loi conjointe	107
I.2 Loi marginale	108
I.3 Loi conditionnelle	109
II Corrélation	109
II.1 Espérances	109
II.2 Covariance	111
II.3 Coefficient de corrélation linéaire	114
II.4 Indépendance	114
III Convolution	117
III.1 Généralités	117
III.2 Exemples	119
III.2-a Convolution de Lois binomiales	119
III.2-b Convolution de lois de Poisson	119

Chap 11 : Applications linéaires	123
I Généralités	123
I.1 Notion d'application linéaire	123
I.2 Opérations sur les applications linéaires	125
I.3 Lien entre matrices et applications linéaires en dimension finie	126
I.3-a Matrice d'une application linéaire en dimension finie	126
I.3-b Changement de base dans un eV en tant qu'application linéaire	130
I.3-c Changement de base dans "E" et/ou "F" pour des AL	131
II Image, noyau, injectivité, surjectivité	134
II.1 Image (d'un sev ou de E)	134
II.2 Noyau	135
II.3 Injection, surjection	137
III Applications linéaires en dimension finie	139
III.1 Rang et matrices	139
III.2 Image d'une base et injectivité	139
III.3 Théorème du rang et conséquences	143
Chap 12 : Variables à densité	145
I Densité	145
I.1 Définitions et généralités	145
I.2 Si je sais qu'une variable admet une densité	148
I.3 Si je veux montrer qu'une variable admet une densité	151
II Moments, Espérance et variance	153
III Exemples fondamentaux	157
III.1 Loi uniforme	157
III.2 Loi exponentielle	160
III.3 Loi normale	162
III.3-a Loi normale centrée réduite	162
III.3-b Loi normale	163
IV Opérations sur les variables aléatoires	164
IV.1 Multiplication et addition par un scalaire	164
IV.2 Somme de va indépendantes	165
V Annexe : Tables et calculatrice	167

Chap 13 : Produit scalaire dans \mathbb{R}^n	171	II.2 Convergence en loi : cas des supports dans \mathbb{N}	211
I Généralités	171	II.2-a En général	211
I.1 Définitions et propriétés élémentaires	171	II.2-b Approximation d'une loi binomiale par une loi de Poisson	211
I.2 Changement de base	174	III Théorème central limite : deux formes	213
II Orthogonalité	177	III.1 Si on connaît σ^2	213
II.1 Définitions et propriétés élémentaires	177	III.1-a Le théorème ; première forme	213
II.2 Traduction des formules de changement de base dans une base ortho- normée	181	III.1-b Appl.1 : approximation d'une loi binomiale par une loi nor- male	214
II.3 Diagonalisation dans une base orthonormée	183	III.1-c Appl 2 : approximation d'une loi de Poisson par une loi normale	215
III Distance et projections orthogonales	184	III.1-d Les approximations en bref	216
Chap 14 : Diagonalisation	191	III.2 Si on ne connaît pas σ^2 : Deuxième forme du TCL	216
I Valeur propre et vecteur propre	191	IV Introduction aux tests de conformité	218
I.1 Valeur propre d'un endomorphisme	191	IV.1 Principe d'un test de conformité sur un exemple	219
I.2 Vecteurs propres et espaces propres d'un endomorphisme	191	IV.2 Résumé de vocabulaire et principe général	220
I.3 Stratégies de calcul des valeurs propres d'un endomorphisme	192	IV.3 Tests bilatéraux	221
I.4 Stratégies de calcul des espaces propres d'un endomorphisme	194	IV.3-a Tests de conformité d'une moyenne connaissant σ^2	221
I.5 Valeurs propres d'une matrice	195	IV.3-b Test de conformité ne connaissant pas σ^2	223
II Diagonalisation	197	IV.4 Tests unilatéraux	223
II.1 Définitions	197	IV.4-a On sait déjà que $\mu_A \geq \mu$; donc cas de H_1 : " $\mu_A > \mu$ "	224
II.2 Nombre de valeurs propres / dimension des espaces propres	197	IV.4-b Cas de H_1 : " $\mu_A < \mu$ "	225
II.3 Diagonalisabilité sur \mathbb{R} ou \mathbb{C} ?	200		
II.4 Formule de changement de base	201		
III Applications : puissances $n^{\text{èmes}}$	201		
III.1 Suites récurrentes linéaires	201		
III.2 Sauts aléatoires par un exemple	202		
Chap 15 : Théorèmes limites	205		
I Loi faible des grands nombres et approximation de la moyenne, d'une proba- bilité et de la variance	205		
I.1 Problématique et vocabulaire sur un exemple :	205		
I.2 Inégalités	206		
I.3 Loi faible : approximation de la moyenne	207		
I.4 Application : approximation des probabilités	208		
I.5 Application : approximation de la variance	208		
II Convergence en loi et variables à supports entiers	209		
II.1 Convergence en loi : cas général	210		

SÉRIES NUMÉRIQUES

CHAPITRE 1

Introduction :

Rappelons qu'une suite numérique est une suite à valeurs dans \mathbb{R} ou dans \mathbb{C} . Pour toute suite numérique $(u_n)_{n \geq N}$, on peut donc construire une nouvelle suite $(S_n)_{n \geq N}$ définie par

$$S_n = \sum_{k=N}^n u_k$$

Ce chapitre consiste à étudier ce type de suites, appelées "séries", qui seront ensuite utilisées dans le but de définir la théorie sur les "probabilités discrètes".

I Définition et somme totale

I-1 Généralités

Définition

Une suite $(S_n)_{n \geq N}$ est appelée *série* s'il existe une suite $(u_n)_{n \geq N}$ telle que

$$\forall n \in \mathbb{N}, \quad n \geq N \quad S_n = \sum_{k=N}^n u_k$$

Dans ce cas, on appelle

- u_n le *terme général* de la série $(S_n)_{n \geq N}$;
- S_n la *somme partielle d'ordre n* de la série $(S_n)_{n \geq N}$.

et on note de manière plus explicite

$$\sum_{n \geq N} u_n$$

pour désigner la série $(S_n)_{n \geq N}$.

Exemple 1 :

$$S_n = \sum_{k=1}^n \underbrace{k^2}_{u_k}. \text{ On a}$$

$$S_1 = \underbrace{1}_{u_1}, \quad S_2 = \underbrace{1}_{u_1} + \underbrace{2^2}_{u_2}, \quad \dots, \quad S_n = \underbrace{1}_{u_1} + \underbrace{2^2}_{u_2} + \dots + \underbrace{n^2}_{u_n}$$

Contre-Exemple(s) :

- 1 ■ $A_n = \sum_{k=n+1}^{2n} u_k$ n'est pas une somme partielle, puisque les deux bornes dépendent de n . En revanche, on peut exprimer A_n grâce à des sommes partielles :

$$A_n = S_{2n} - S_n, \text{ où } S_n = \sum_{k=N}^n u_k \text{ où } N \text{ est fixé.}$$

- 2 ■ $B_n = \sum_{k=0}^n \frac{1}{n+k}$ n'est a priori pas une somme partielle, puisque le terme général dépend aussi de n . Néanmoins, dans cet exemple, on peut également exprimer B_n grâce à des sommes partielles :

$$B_n = \sum_{k=0}^n \frac{1}{n+k} = \sum_{k=n}^{2n} \frac{1}{k} = S_{2n} - S_{n-1}$$

où $S_n = \sum_{k=1}^n \frac{1}{k}$.

Définition

On dit (trivialement) que la série $\sum_{n \geq N} u_n$ *converge* (resp. *diverge*) si la suite $(S_n)_{n \geq N}$ *converge* (resp. *diverge*), où $S_n = \sum_{k=N}^n u_k$ pour tout $n \geq N$.

En cas de convergence, la limite sera notée $\lim_{n \rightarrow +\infty} \sum_{k=N}^n u_k = \sum_{k=N}^{+\infty} u_k$ et sera appelée *somme totale* de la série $\sum_{n \geq N} u_n$.

Exemples :

- 2 ■ La série $\sum_{n \geq 0} 1$ est divergente. En effet, $\sum_{k=0}^n 1 = n+1 \xrightarrow{n \rightarrow +\infty} +\infty$
- 3 ■ La série $\sum_{n \geq 0} n$ est divergente. En effet, $\sum_{k=0}^n k = \frac{n(n+1)}{2} \xrightarrow{n \rightarrow +\infty} +\infty$.
- 4 ■ Si $|q| < 1$, la série $\sum q^n$ est convergente. En effet, $\sum_{k=0}^n q^k = \frac{1 - q^{n+1}}{1 - q} \xrightarrow{n \rightarrow +\infty} \frac{1}{1 - q}$, d'où $\sum_{k=0}^{+\infty} q^k = \frac{1}{1 - q}$

OBJETS

Ne pas mélanger les types d'objets en présence :

objet :	u_n	$S_n = \sum_{k=N}^n u_k$	$\sum_{k=N}^{+\infty} u_k = \sum_{n=N}^{+\infty} u_n$	$\sum_{n \geq N} u_n = \sum_{k \geq N} u_k$
type :	nombre réel qui dépend de n		nombre réel : limite indépendante de n	suite (liste de nombres)

■ Exemple 5 :

On peut dire que $\sum_{n=0}^{+\infty} \frac{1}{2^n} = \frac{1}{1-\frac{1}{2}} = 2$, mais $\sum_{n \geq 0} \frac{1}{2^n}$ ne vaut **pas** $\frac{1}{1-\frac{1}{2}}$. C'est une suite.

On a
$$\sum_{n \geq 0} \frac{1}{2^n} = (S_n) = \left(\underbrace{1}_{S_0}, \underbrace{1 + \frac{1}{2}}_c, \underbrace{1 + \frac{1}{2} + \frac{1}{2^2}}_c, \dots \right) \quad (\text{infinité de termes})$$

⚠ Remarque :

Si la limite de la série est $\pm\infty$, on s'autorise toutefois souvent **par abus de notation** à écrire $\sum_{k=N}^{+\infty} u_k = \pm\infty$, mais dans ce cas, on ne parle pas de somme totale, puisque ce nombre n'existe pas. (Oui, on rappelle en effet que ∞ n'est pas un nombre...)

? Exercice 1

Montrer que $\sum_{n \geq 0} 2^n$ est divergente et que $\sum_{n=0}^{+\infty} 2^n = +\infty$.

Solution

Pour tout $n \in \mathbb{N}$, on a $\sum_{k=0}^n 2^k = \frac{1-2^{n+1}}{1-2} = 2^{n+1} - 1 \xrightarrow{n \rightarrow +\infty} +\infty$

🌿 Définition

De même que pour une suite, on appelle *nature* de la série le caractère convergent ou divergent de la série.

⚠ Remarque :

Tout comme les suites, il existe des séries qui divergent autrement que vers $\pm\infty$!

■ Exemple 6 :

Si $u_n = (-1)^n$ pour tout $n \in \mathbb{N}$. Alors $S_n = \sum_{k=0}^n u_k = \begin{cases} 0 & \text{si } n \text{ est impair} \\ 1 & \text{sinon} \end{cases}$

Les sous-suites de (S_n) d'indices pairs et impairs convergent donc vers des limites différentes. Autrement dit, la série $\sum u_n$ diverge.

⚠ DIFFÉRENCE SUITE (u_n) / SÉRIE $\sum_{n \geq N} u_n$

➤ Ne pas confondre la convergence de la suite (u_n) et celle de la série $\sum u_n$!

■ Exemple 7 :

Si $u_n = 1$, on a $\sum_{k=1}^n 1 = n \xrightarrow{n \rightarrow +\infty} +\infty$, alors que $\lim_{n \rightarrow +\infty} u_n = 1$. La suite (u_n) converge, alors que $\sum_{n \in \mathbb{N}} u_n$ diverge.

I-2 Changement de borne / de termes

⚠ VARIABLES

L'objet sur lequel on fait la limite est le "n" du $S_n = \sum_{k=N}^n$ et non le "k" qui est juste une variable muette. Ainsi, la borne de départ N est importante dans le calcul de la limite :

■ Exemple 8 :

Si $|q| < 1$, la série $\sum_{n \geq 2} q^n$ est convergente. En effet,

$$\sum_{k=2}^n q^k = \sum_{k=0}^n q^k - (q^0 + q^1) = \frac{1-q^{n+1}}{1-q} - (1+q) \xrightarrow{n \rightarrow +\infty} \frac{1}{1-q} - (1+q) = \frac{q^2}{1-q}$$

D'où $\sum_{n=0}^{+\infty} q^n = \frac{1}{1-q}$ et $\sum_{n=2}^{+\infty} q^n = \frac{q^2}{1-q} \neq \sum_{n=0}^{+\infty} q^n$

Propriété 1

Pour une suite $(u_n)_{n \geq N}$ et pour tout $M \in \mathbb{N}$, $M \geq N$, les séries $\sum_{n \geq N} u_n$ et $\sum_{n \geq M} u_n$ sont de même nature.

Démonstration :

Pour tout $n \geq M$, on a $\sum_{k=N}^n u_k = \underbrace{\sum_{k=N}^{M-1} u_k}_{\text{indépendant de } n} + \sum_{k=M}^n u_k$

Le terme $\sum_{k=N}^{M-1} u_k$ étant constant par rapport à n , la convergence de $\left(\sum_{k=N}^n u_k\right)$ est donc équivalente à la convergence de $\left(\sum_{k=M}^n u_k\right)$. □

⚠ CONFUSION

➤ Même nature ne veut pas dire même somme totale! (cf ex précédent)

⚠ Remarque :

S'il n'y a aucune ambiguïté sur la borne de départ N de la série (ou si elle n'a pas d'importance), on notera le plus souvent $\sum u_n$ pour désigner la série $\sum_{n \geq N} u_n$.

Propriété 2

Si (u_n) et (v_n) ne diffèrent que d'un nombre fini de termes, alors $\sum u_n$ et $\sum v_n$ sont de même nature.

Méthode 2 - Décomposition

Propriété 3 (Combinaisons linéaires de séries)

Soient $\sum_{n \geq N} u_n$ et $\sum_{n \geq N} v_n$ deux séries et $\lambda \in \mathbb{C}$.

- i - Si $\sum_{n \geq N} u_n$ converge, alors la suite $\sum_{n \geq N} \lambda u_n$ converge vers $\lambda \sum_{n=N}^{+\infty} (u_n)$.
- ii - Si $\sum_{n \geq N} u_n$ et $\sum_{n \geq N} v_n$ convergent, alors la série $\sum_{n \geq N} (u_n + v_n)$ converge vers $\sum_{n=N}^{+\infty} u_n + \sum_{n=N}^{+\infty} v_n$.
- iii - Si $\sum_{n \geq N} u_n$ converge et $\sum_{n \geq N} v_n$ diverge, alors $\sum_{n \geq N} (u_n + v_n)$ diverge.
- iv - Si $\sum_{n \geq N} u_n$ diverge et $\sum_{n \geq N} v_n$ diverge, alors on ne peut rien dire sur $\sum_{n \geq N} (u_n + v_n)$.

Démonstration :

C'est tout simplement une conséquence directe des propriétés sur les suites. \square

Exemple 10 :

La série $\sum_{n \geq 0} \left(\frac{1}{2}\right)^{n+1} = \frac{1}{2} \sum_{n \geq 0} \frac{1}{2}^n$ converge. (application de i)

En effet, comme $|\frac{1}{2}| < 1$, on sait que $\sum_{n \geq 0} \frac{1}{2}^n$ converge et la série initiale converge par multiplication.

Exemple 11 :

La série $\sum_{n \geq 0} \left(\frac{1}{6} + \left(\frac{1}{2}\right)^{n+1}\right)$ diverge. (application de iii)

En effet, on voit pour commencer que : $\sum_{n \geq 0} \frac{1}{6} = \frac{1}{6} \sum_{n \geq 0} 1$.

On a déjà vu que la série $\sum_{n \geq 0} 1$ diverge et que $\sum_{n \geq 0} \left(\frac{1}{2}\right)^{n+1}$ converge. Par combinaison linéaire de séries, on sait donc que la somme va diverger.



Remarque :

En cas de divergence vers l'infini, le principe des limites usuelles s'applique toutefois. Par exemple, si $\sum_{n=N}^{+\infty} u_n = +\infty$ et $\sum_{n=N}^{+\infty} v_n = +\infty$, alors $\sum_{n=N}^{+\infty} (u_n + v_n) = +\infty$. En revanche, si les deux limites infinies sont de signe opposé, alors c'est une forme indéterminée.

Démonstration :

Sachant que la convergence ne dépend pas de la borne de départ, pour des raisons de facilité d'écriture, on pourra supposer que la suite est indexée sur \mathbb{N} .

On pose $\Omega = \{n \in \mathbb{N} \mid u_n \neq v_n\}$. Ω est un sous ensemble majoré de \mathbb{N} . Il possède donc une borne supérieure, notée M .

Par définition de M , on a

$$\left(\sum u_n\right)_{n \geq M+1} = \left(\sum v_n\right)_{n \geq M+1}.$$

Par le lemme précédent, on sait que $\sum_{n \in \mathbb{N}} u_n$ est de même nature que $\left(\sum u_n\right)_{n \geq M+1}$.

De même, $\sum_{n \in \mathbb{N}} v_n$ est de même nature que $\left(\sum v_n\right)_{n \geq M+1}$. Par transitivité, on en déduit que $\sum_{n \in \mathbb{N}} u_n$ et $\sum_{n \in \mathbb{N}} v_n$ sont de même nature. \square

I-3 Calculs

Méthode 1 - Calcul de la somme totale par télescopage

Le télescopage est une méthode utilisée (dans des cas limités) pour aboutir au calcul explicite de la somme partielle puis de la somme totale. Il consiste à écrire le terme général u_n sous la forme d'une somme d'éléments de même type :

$$u_n = v_{n+1} - v_n,$$

(exemple $\frac{1}{n}$ et $\frac{1}{n+1}$), qui, par changement d'indices, produiront l'élimination progressive de la plupart des termes.

Exemple 9 :

Calcul de la somme de la série $\sum_{n \geq 1} \frac{1}{n(n+1)}$:

pour $u_n = \frac{1}{n(n+1)}$, on écrit $u_n = \underbrace{\frac{1}{n}}_{v_n} - \underbrace{\frac{1}{n+1}}_{v_{n+1}}$. Ainsi, par télescopage :

$$\sum_{k=1}^n \frac{1}{k(k+1)} = \sum_{k=1}^n v_k - \sum_{k=1}^n v_{k+1} = \sum_{k=1}^n v_k - \sum_{k=2}^{n+1} v_k = v_1 - v_{n+1} = 1 - \frac{1}{n+1}$$

Autrement dit, la série est convergente et $\sum_{k=1}^{+\infty} \frac{1}{k(k+1)} = 1$.

II Critères de convergence / divergence

II-1 Divergence grossière

Proposition 4

Si $\sum u_n$ converge, alors $\lim_{n \rightarrow +\infty} u_n = 0$

Démonstration :

Si $\sum u_n$ converge, on note S la somme totale de la série et S_n la somme partielle d'ordre n . On a alors

$$\square \quad u_n = S_n - S_{n-1} \xrightarrow{n \rightarrow +\infty} S - S = 0.$$

Conséquence :

Si la suite (u_n) ne tend pas vers 0, alors $\sum u_n$ ne peut être convergente.

■ Exemples :

12 ■ $\sum_{n \in \mathbb{N}^*} \ln(2 + \frac{1}{n})$ ne peut être convergente, car $\lim_{n \rightarrow +\infty} \ln(2 + \frac{1}{n}) = \ln 2$.

13 ■ $\sum_{n \in \mathbb{N}} \sin n \frac{\pi}{2}$ ne peut être convergente, car $(\sin n \frac{\pi}{2})$ est une suite divergente.

 **RÉCIPROQUE FAUSSE !**
 Il existe des séries divergentes, dont le terme général tend vers 0 :

Proposition 5

La série $\sum_{n \geq 1} \frac{1}{n}$ est divergente.

Démonstration :

$$\text{On pose } S_n = \sum_{k=1}^n \frac{1}{k}. \text{ Alors } S_{2n} - S_n = \sum_{k=n+1}^{2n} \frac{1}{k} \geq n \frac{1}{2n} = \frac{1}{2}$$

Si (S_n) était convergente, en notant S la somme totale, par passage à la limite des deux cotés de l'inégalité, on obtiendrait $\underbrace{S - S}_0 \geq \frac{1}{2}$, ce qui est absurde. \square

 **Définition**
 Si une série $\sum u_n$ est telle que (u_n) ne tend pas vers 0, on dit que la série *diverge grossièrement*.

II-2 Séries à termes de signe constant

II.2-a) Propriétés élémentaires des séries à termes positifs

Commentaires :

Comme on ne sait pas toujours calculer facilement les sommes partielles, pour déterminer la convergence d'une série dans le cas où elle ne diverge pas grossièrement, on peut être en proie à certaines difficultés. Néanmoins (et heureusement), il existe un certain nombre de critères nous permettant de faciliter cette étude. Les critères de base consistent à se ramener à des séries dont la convergence est déjà connue.

Lemme 6

On se donne $\sum u_n$ une série à termes positifs au moins à partir d'un certain rang. On a alors :

i → La série $\sum u_n$ est croissante.

ii → La série $\sum u_n$ converge si et seulement si elle est majorée. (et s'il y a convergence, sa somme totale est inférieure à son majorant.)

iii → La série $\sum u_n$ diverge vers $+\infty$ ssi elle n'est pas majorée.

Démonstration :

Soit $\sum u_n$ une série à terme positifs. On peut supposer que $\sum u_n = \sum_{n \in \mathbb{N}} u_n$.

On pose $S_n = \sum_{k=0}^n u_k$. Alors $S_{n+1} - S_n = u_n \geq 0$. La suite (S_n) est donc croissante.

Comme (S_n) est croissante, les critères de convergence sur les suites croissantes nous indiquent que $\sum_{n \in \mathbb{N}} u_n$ est convergente si et seulement si elle est majorée. Les

critère de majoration de la limite s'appliquent également.

De même que pour la partie précédente, les critères de convergence sur les suites croissantes nous indiquent que si $\sum u_n$ n'est pas majorée, elle diverge vers $+\infty$. \square

■ Exemple 14 :

Montrons que la série $\sum_{n \geq 1} \frac{1}{n3^n}$ converge :

Notons

$$S_n = \sum_{k=1}^n \frac{1}{k3^k}$$

Remarquons que la série est à termes positifs. Elle est donc croissante. De plus, pour tout $k \geq 1$, on a

$$\frac{1}{k3^k} \leq \frac{1}{3^k}$$

Ainsi, on a

$$S_n \leq \sum_{k=1}^n \frac{1}{3^k} = \frac{1}{3} \sum_{i=0}^{n-1} \frac{1}{3^i} = \frac{1}{3} \frac{1 - (\frac{1}{3})^n}{1 - \frac{1}{3}} = \frac{1}{2} \left(1 - \left(\frac{1}{3} \right)^n \right) \leq \frac{1}{2}$$

Ainsi, au total, on a pour tout $n \in \mathbb{N}^*$:

$$S_n \leq \frac{1}{2}$$

C'est donc une suite croissante et majorée, donc convergente, avec même

$$0 \leq \sum_{n=1}^{+\infty} \frac{1}{n3^n} \leq \frac{1}{2}$$

Commentaires :

En réalité, ce théorème n'est pas du tout optimal dans l'exemple précédent. On se servira plutôt du théorème qui suit :

II.2-b) Critère de comp. de deux séries à termes positifs

Théorème 7 de comparaison des séries à termes positifs (TCSTP)

On se donne deux suites (u_n) et (v_n) telles que, à partir d'un certain rang N ,

- (u_n) et (v_n) sont **positives**
- $u_n \leq v_n$.

Alors

- i* → Si $\sum v_n$ converge, alors $\sum u_n$ converge et $\sum_{n=N}^{+\infty} u_n \leq \sum_{n=N}^{+\infty} v_n$.
- ii* → Si $\sum u_n$ diverge, alors $\sum v_n$ diverge.

Démonstration :

Les séries $\sum_{n \geq N} u_n$ et $\sum_{n \geq N} v_n$ (resp. $\sum_{n \geq N} v_n$ et $\sum_{n \geq N} u_n$) sont de même nature. Il suffit donc d'étudier la convergence de $\sum_{n \geq N} u_n$ et $\sum_{n \geq N} v_n$.

Clairement, on obtient $0 \leq \sum_{k=N}^n u_k \leq \sum_{k=N}^n v_k$.

De plus, d'après le lemme précédent, comme tous les termes de rang plus grand que N sont positifs, les séries $\left(\sum_{n \geq N} u_n\right)$ et $\left(\sum_{n \geq N} v_n\right)$ sont toutes deux croissantes.

- Supposons que la série $\sum_{n \in \mathbb{N}} v_n$ converge.

$$\begin{aligned} \text{On a } \sum v_n \text{ converge} &\Leftrightarrow \sum_{n \geq N} v_n \text{ converge} \\ &\Leftrightarrow \sum_{n \geq N} v_n \text{ est majorée} \\ &\Rightarrow \sum_{n \geq N} u_n \text{ est majorée} \\ &\Leftrightarrow \sum_{n \geq N} u_n \text{ est convergente} \\ &\Leftrightarrow \sum u_n \text{ est convergente} \end{aligned}$$

En cas de convergence, le passage à la limite donne le reste.

- Supposons que la série $\sum_{n \in \mathbb{N}} u_n$ diverge.

Cette partie est simplement la contraposée de la première. \square

■ Exemples :

- 15 ■ La série $\sum \left(\frac{1}{n}\right)^n$ converge.

En effet, on observe que pour tout $n \geq 2$, $0 \leq \left(\frac{1}{n}\right)^n \leq \left(\frac{1}{2}\right)^n$. Or la série $\sum \left(\frac{1}{2}\right)^n$ est convergente (comme vu sur un exemple précédent.) Ainsi, par théorème de comparaison des séries à termes positifs, $\left(\sum \left(\frac{1}{n}\right)^n\right)_{n \geq 2}$ converge.

- 16 ■ La série $\sum_{n \geq 2} \frac{n+1}{n(n-1)}$ diverge.

On constate en effet aisément que

$$\frac{n+1}{n(n-1)} \geq \frac{1}{n} \geq 0 \quad \forall n \geq 2.$$

La divergence de la série $\left(\sum \frac{1}{n}\right)_{n \geq 2}$, permet de conclure.

II.2-c) Equivalence

Théorème 8

On se donne deux suites (u_n) et (v_n) telles que,

- i* → à partir d'un certain rang N , (u_n) et/ou (v_n) sont **positives**
 - ii* → $u_n \sim v_n$.
- Alors $\sum u_n$ converge ssi $\sum v_n$ cv (elles sont de même nature)

Démonstration :

Les deux séries étant équivalentes, à partir d'un certain rang, il existe une suite h_n telle que

$$u_n = h_n v_n$$

avec $h_n \xrightarrow[n \rightarrow +\infty]{} 1$. Ainsi, il existe $M \geq N$ tel que pour tout $n \geq M$,

$$\frac{1}{2} \leq h_n \leq \frac{3}{2}.$$

Les suites (u_n) et (v_n) étant de signe constant à partir du rang N (supposons ici positives, le cas négatif est laissé en exercice), pour tout $n \geq M$,

$$0 < \frac{1}{2} v_n \leq u_n \leq \frac{3}{2} v_n$$

Par TCSTP sur les deux premiers membres, on déduit que si $\sum u_n$ converge, alors $\sum v_n$ converge également.

Les membres de droite donnent, toujours par TCSTP, que si $\sum v_n$ converge, alors $\sum u_n$ converge également.

(Ou alors, l'énoncé étant symétrique en u_n et v_n , il suffit d'échanger les deux suites u_n et v_n .) CQFD \square

 **Remarque :**

De manière équivalente, la conclusion du théorème précédent peut également être " $\sum u_n$ diverge ssi $\sum v_n$ diverge".

Commentaires :

Il est écrit " (u_n) et/ou (v_n) sont positives". En effet, si l'une des deux est positive et qu'elles sont équivalentes, cela signifie qu'automatiquement, l'autre sera également positive à partir d'un certain rang. Il n'est donc pas nécessaire de vérifier que les deux sont positives. Une seule suffit.

■ **Exemple 17 :**

Montrons que la série $\sum \ln\left(1 + \frac{1}{n(n+1)}\right)$ converge :

On observe trivialement que les termes de la série sont tous positifs. De plus, comme

$$\frac{1}{n(n+1)} \xrightarrow{n \rightarrow +\infty} 0$$

On sait que $\ln\left(1 + \frac{1}{n(n+1)}\right) \sim \frac{1}{n(n+1)}$. Or, nous avons déjà montré par télescopage que la série $\sum \frac{1}{n(n+1)}$ était convergente. En conclusion, $\sum \ln\left(1 + \frac{1}{n(n+1)}\right)$ converge.

 **Exercice 2**

Montrer que la série $\sum \ln\left(1 - \frac{1}{n(n+1)}\right)$ converge. (*Attention, ici les séries sont négatives !!*)

Solution

Comme dans l'exemple précédent, on voit que pour tout $n \in \mathbb{N}^*$:

$$0 \leq -\ln\left(1 - \frac{1}{n(n+1)}\right) \sim \frac{1}{n(n+1)}$$

Ce qui, par convergence de la série positive $\sum \frac{1}{n(n+1)}$ et par équivalence des termes généraux nous dit que $-\sum \ln\left(1 - \frac{1}{n(n+1)}\right)$ converge et donc également $\sum \ln\left(1 + \frac{1}{n(n+1)}\right)$

 **ERREUR DE RAISONNEMENT**

Équivalence des termes généraux ne signifie EN AUCUN CAS que les séries le sont équivalentes !!

■ **Exemple 18 :**

Remarquons que

$$\frac{1}{n(n+1)} \sim \frac{1}{n(n-1)}$$

(on peut passer par $\frac{1}{n(n+1)} \sim \frac{1}{n^2} \sim \frac{1}{n(n-1)}$). Mais

$$\sum_{n \geq 2} \frac{1}{n(n+1)} \not\sim \sum_{n \geq 2} \frac{1}{n(n-1)}$$

En effet, on sait que la série $\sum_{n \geq 2} \frac{1}{n(n+1)}$ est convergente par télescopage (p.3).

Nous avons d'ailleurs également calculé (p.3) :

$$\sum_{k=1}^n \frac{1}{k(k+1)} = 1 - \frac{1}{n+1} \sim 1$$

D'où

$$\sum_{k=2}^n \frac{1}{n(n+1)} = 1 - \frac{1}{n+1} - \frac{1}{1(1+1)} = \frac{1}{2} - \frac{1}{n+1} \sim \frac{1}{2}$$

Comme $\frac{1}{2} \not\sim 1$, on a donc

$$\sum_{k=1}^n \frac{1}{k(k+1)} = 1 - \frac{1}{n+1} \not\sim \sum_{k=2}^n \frac{1}{k(k+1)} = 1 - \frac{1}{n+1}$$

 **VÉRIFICATION HYPOTHÈSES**

Le théorème précédent s'applique également à des séries dont les termes sont "négatifs à partir d'un certain rang". En revanche, attention à **ne pas l'appliquer si les termes ne sont pas de signe constant !**

■ **Exemple 19 :**

Avec $u_n = \frac{(-1)^n}{\sqrt{n}} + \frac{1}{n}$, on observe que :

$$u_n \sim \frac{(-1)^n}{\sqrt{n}}$$

Or, $\sum \frac{(-1)^n}{\sqrt{n}}$ est convergente (admis ici) mais $\sum u_n = \sum \frac{(-1)^n}{\sqrt{n}} + \sum \frac{1}{n}$ est ainsi la combinaison linéaire d'une série convergente et d'une série divergente. Elle est donc bien **divergente**, contrairement à $\sum \frac{(-1)^n}{\sqrt{n}}$ qui elle, est convergente !



Remarque :

Tous les résultats de la section II-2 peuvent être transposés plus ou moins directement au cas où on a u_n et v_n **tous deux négatifs à partir d'un certain rang** N . (On pourra éventuellement s'interroger sur ce point.)

Néanmoins, afin de ne pas multiplier (et surtout mélanger) les résultats à retenir, il est plus aisé pour vous de poser simplement deux autres suites a_n, b_n telles que $a_n = -u_n, b_n = -v_n$. Ainsi, on a

$$a_n, b_n \geq 0 \quad \forall n \geq N$$

et on peut alors tenter de conclure dans un premier temps sur les séries $\sum a_n = -\sum u_n$ et $\sum b_n = -\sum v_n$. L'important est que le signe soit **constant et identique pour les deux séries** à partir d'un certain rang.

? Exercice 3

Montrer que la série $\sum_{n \in \mathbb{N}} \ln \frac{1}{n+1} e^{-n+1}$ converge.

(On pourra utiliser que $\ln(1+x) \leq x \quad \forall x \geq 0$)

Solution

Si on pose $u_n = \ln \frac{1}{n+1} e^{-n+1}$, on constate, comme $0 < \frac{1}{n+1} \leq 1$, que

$$\ln \frac{1}{n} = -\ln n < 0 \text{ et donc } u_n \leq 0 \quad \forall n \in \mathbb{N}.$$

On pose alors $a_n = -\ln \frac{1}{n+1} e^{-n+1} = \ln(n+1) e^{-n+1} \geq 0 \quad \forall n \in \mathbb{N}$. D'après les inégalités de l'énoncé, on sait alors que

$$0 \leq a_n \leq n \left(\frac{1}{e}\right)^{n-1} \quad \forall n \in \mathbb{N}$$

Or, comme $\frac{1}{e} \in [0, 1[$, la série $\sum n \left(\frac{1}{e}\right)^{n-1}$ est donc une série géométrique dérivée convergente. Par TCSTP, la série $\sum a_n$ est donc convergente. Par multiplication par -1 , il en va donc de même pour $\sum u_n$.

II-3 Convergence absolue

Commentaires :

S'il est souvent utile de pouvoir se ramener à l'étude d'une série déjà connue, le paragraphe précédent ne permet de le faire que pour des séries à termes tous positifs (ou tous négatifs) à partir d'un certain rang. Or, en général, les séries sont beaucoup plus diversifiées. Malgré tout, un des critères de convergence, appelé convergence absolue, va permettre de se ramener, dans beaucoup de cas, à l'étude d'une série à termes positifs.

Proposition 9

On se donne une série quelconque $\sum u_n$.

Si la série $\sum |u_n|$ converge, alors la série $\sum u_n$ converge.

On appelle ceci la *convergence absolue*.

Démonstration :

On ne perd rien à supposer que $\sum u_n = \sum_{n \in \mathbb{N}} u_n$ (c'est-à-dire que la série est indexée sur \mathbb{N} tout entier.)

On pose
$$u_n^+ = \max(u_n, 0) = \begin{cases} u_n & \text{si } u_n \geq 0 \\ 0 & \text{sinon} \end{cases}$$

et
$$u_n^- = -\min(u_n, 0) = \begin{cases} -u_n & \text{si } u_n \leq 0 \\ 0 & \text{sinon} \end{cases}$$

de sorte que $u_n = u_n^+ - u_n^-$ avec

$$0 \leq u_n^+ \leq |u_n| \quad \text{et} \quad 0 \leq u_n^- \leq |u_n|$$

pour tout $n \in \mathbb{N}$.

Ainsi,

$$\sum_{k=0}^n (u_k^+ - u_k^-) = \sum_{k=0}^n u_k^+ - \sum_{k=0}^n u_k^- \quad (*)$$

La série $\sum_{n \in \mathbb{N}} u_n^+$ est à termes positifs et u_n^+ est majoré par $|u_n|$, qui est, par hypothèse, le terme général d'une série convergente. La série $\sum_{n \in \mathbb{N}} u_n^+$ est donc convergente. De même, la série $\sum_{n \in \mathbb{N}} u_n^-$ est convergente.

Comme $\lim_{n \rightarrow +\infty} \sum_{k=0}^n u_k^+$ et $\lim_{n \rightarrow +\infty} \sum_{k=0}^n u_k^-$ existent, on en déduit l'existence de leur différence grâce à l'égalité (*), autrement dit, $\sum_{n \in \mathbb{N}} u_n$ converge. \square



RÉCIPROQUE FAUSSE

➤ voyons ceci dans l'exemple suivant :

■ Exemple 20 :

Soit la série harmonique alternée : $\sum_{n \geq 1} \frac{(-1)^{n+1}}{n}$.

Elle n'est pas absolument convergente, mais est convergente. (non démontré ici, mais cf. fin de chapitre (p.12) pour une observation graphique.)

III Exemples fondamentaux

Méthode 3 - Détermination d'une convergence

Si le terme général de la série est de signe quelconque éventuellement fluctuant : on pourra chercher à prouver sa convergence absolue ou à décomposer en somme des séries dont on connaît la convergence.

Si le terme général de la série est positif : on pourra tenter de majorer / minorer par une suite plus simple ou alors en chercher un équivalent plus simple.

Si le terme général de la série est négatif : on peut transposer les théorèmes sur les séries à termes positifs, ou alors poser la série $\sum(-u_n)$

Dans tous les cas, les séries auxquelles on peut se référer sont celles qui suivent dans cette partie. Toutes les autres sont à redémontrer.

III-1 Séries de Riemann $\sum \frac{1}{n}$ et $\sum \frac{1}{n^2}$

Proposition 10

Rappel : La série $\sum \frac{1}{n}$ est divergente.

Démonstration :

Cette démonstration a déjà été effectuée dans la partie "divergence grossière". \square

Proposition 11

La série $\sum \frac{1}{n^2}$ est convergente.

Démonstration :

Soit $n \in \mathbb{N}^*$, $n > 1$. On a $n \geq n - 1$, d'où

$$n^2 \geq n(n - 1)$$

Ainsi, par décroissance de la fonction inverse sur $]0; +\infty[$, on a

$$0 < \frac{1}{n^2} \leq \frac{1}{n(n - 1)}$$

Or, la série $\sum_{n \geq 2} \frac{1}{n(n-1)}$ est convergente (déjà démontré par télescopage.) Ainsi, par critère de comparaison des séries à termes positifs, on obtient la convergence de la série $\sum_{n \geq 2} \frac{1}{n^2}$ et donc celle de $\sum \frac{1}{n^2}$. \square

? Exercice 4

Montrer que toute série de type $\sum \frac{1}{n^\alpha}$ est :

- convergente si $\alpha \geq 2$.
- divergente si $\alpha \leq 1$.

Solution

- Si $\alpha \geq 2$:

On a pour tout $n \in \mathbb{N}^*$:

$$0 \leq \frac{1}{n^\alpha} \leq \frac{1}{n^2}.$$

Or $\sum \frac{1}{n^2}$ converge. Ainsi, par TCSTP, la série $\sum \frac{1}{n^\alpha}$ converge.

- Si $\alpha \leq 1$:

On a pour tout $n \in \mathbb{N}^*$:

$$\frac{1}{n^\alpha} \geq \frac{1}{n} \geq 0.$$

Or $\sum \frac{1}{n}$ diverge. Ainsi, par TCSTP, la série $\sum \frac{1}{n^\alpha}$ diverge.

⚠ Remarque :

Les résultats précédents ne permettent pas de conclure sur la convergence ou la divergence des séries de type $\sum \frac{1}{n^\alpha}$ si $1 < \alpha < 2$. Ces résultats existent mais ne sont pas à notre programme.

? Exercice 5

Montrer que $\sum e^{-\frac{2}{n^2}}$ converge.

Solution

Comme $\frac{2}{n^2} \xrightarrow[n \rightarrow +\infty]{} 0$, on a

$$e^{-\frac{2}{n^2}} \sim \frac{2}{n^2}$$

Or, la série $\sum \frac{2}{n^2} = 2 \sum \frac{1}{n^2}$ est une série positive convergente et donc $\sum e^{-\frac{2}{n^2}}$ converge.

III-2 Séries géométriques

🌿 Définition

On appelle *série géométrique* toute série $\sum_{n \geq N} q^n$, où $q \in \mathbb{C}$.

Théorème 12

- $\sum_{n \in \mathbb{N}} q^n$ est convergente si et seulement si $|q| < 1$.
- Si $|q| < 1$ la somme totale existe et vaut $\sum_{k=0}^{+\infty} q^k = \frac{1}{1-q}$.

Démonstration :

• Tout d'abord, si $|q| \geq 1$, la série est grossièrement divergente puisque $q^n \not\rightarrow 0$.

• Si $|q| < 1$:

On sait que $\sum_{k=0}^n q^k = \frac{1-q^{n+1}}{1-q}$. Or, comme $|q| < 1$, on a

$$q^{n+1} \xrightarrow{n \rightarrow +\infty} 0$$

On en déduit donc la proposition directement par considération de limite sur la formule. \square

III-3 Série "dérivée" $(\sum_{n \in \mathbb{N}} nq^{n-1})_{n \in \mathbb{N}}$

Théorème 13

- $\sum_{n \in \mathbb{N}^*} nq^{n-1}$ est convergente si et seulement si $|q| < 1$.
- Si $|q| < 1$ la somme totale existe et vaut $\sum_{k=0}^{+\infty} kq^{k-1} = \sum_{k=1}^{+\infty} kq^{k-1} = \frac{1}{(1-q)^2}$.

Moyen mnémotechnique : Pour retenir ce résultat, on "dérive" terme à terme par rapport à q l'expression $\frac{1}{1-q} = \sum_{k=0}^{+\infty} q^k$, pour obtenir $\frac{1}{(1-q)^2} = \sum_{k=1}^{+\infty} kq^{k-1}$.

INTERPRÉTATION FAUSSE
 **Ne pas voir dans ce moyen mnémotechnique un quelconque résultat sur les dérivées de séries !**

Démonstration :

• Supposons que $|q| \geq 1$.

Alors $n|q|^n \geq n$ et donc

$$n|q|^n \xrightarrow{n \rightarrow +\infty} +\infty$$

ce qui signifie que

$$nq^n \not\xrightarrow{n \rightarrow +\infty} 0$$

Il y a divergence grossière de la série.

• Supposons maintenant que $|q| < 1$. Montrons que la série converge.

Méthode 1 : (par dérivée)

On pose $f_n(q) = \sum_{k=0}^n q^k$, alors $f'_n(q) = \sum_{k=1}^n kq^{k-1}$

Comme on sait que $f_n(q) = \frac{1-q^{n+1}}{1-q}$, on obtient

$$\begin{aligned} f'_n(q) &= \left(\frac{1-q^{n+1}}{1-q} \right)' = \frac{-(n+1)q^n(1-q) - (1-q^{n+1})(-1)}{(1-q)^2} \\ &= \frac{-(n+1)q^n + (n+1)q^{n+1} + 1 - q^{n+1}}{(1-q)^2} \\ &= \frac{nq^{n+1} - (n+1)q^n + 1}{(1-q)^2} \end{aligned}$$

On en déduit que $\sum_{k=1}^n kq^{k-1} = \frac{nq^{n+1} - (n+1)q^n + 1}{(1-q)^2} \xrightarrow{n \rightarrow +\infty} \frac{1}{(1-q)^2}$, ce qui montre la proposition.

Méthode 2 : par considération combinatoire

$$\begin{array}{rcl} 1 + & q + & q^2 + & q^3 + & \dots & + q^{n-1} & = & \sum_{k=0}^{n-1} q^k = \frac{1-q^n}{1-q} \\ & q + & q^2 + & q^3 + & \dots & + q^{n-1} & = & \sum_{k=1}^{n-1} q^k = q \frac{1-q^{n-1}}{1-q} \\ & & & & & & & \dots \\ & & & & & q^N + & \dots & + q^{n-1} & = & \sum_{k=N}^{n-1} q^k = q^N \frac{1-q^{n-N}}{1-q} \\ & & & & & & & \dots \\ & & & & & & & q^{n-1} & = & \sum_{k=n-1}^{n-1} q^k = q^{n-1} \frac{1-q}{1-q} \end{array}$$

Par sommation sur les colonnes, on obtient la somme partielle

$$\sum_{k=1}^n kq^{k-1} = \sum_{N=0}^n q^N \frac{1-q^{n-N}}{1-q} = \frac{1}{1-q} \sum_{N=0}^n (q^N - q^n) = \frac{1}{1-q} \left(\sum_{N=0}^n q^N - nq^n \right)$$

Par passage à la limite, on obtient la convergence de la série et la somme totale.

Méthode 3 : par introduction du "1"

On pose $S_n = \sum_{k=0}^n q^k$ et $S'_n = \sum_{k=0}^n kq^{k-1}$. Alors

$$\begin{aligned} S'_n &= \sum_{k=1}^n kq^{k-1} = \sum_{k=1}^n (k-1+1)q^{k-1} \\ &= \sum_{k=1}^n (k-1)q^{k-1} + \sum_{k=1}^n q^{k-1} \\ &= qS'_{n-1} + S_{n-1} = q(S'_n - nq^{n-1}) + S_{n-1} \end{aligned}$$

D'où

$$S'_n = \frac{S_{n-1} - nq^n}{1 - q}$$

Par passage à la limite, on trouve encore une fois le résultat annoncé.

□

Exercice 6

Montrer que $\sum_{n \in \mathbb{N}} nq^n$ est **convergente** si et seulement si $|q| < 1$ et que dans ce cas, la

somme totale vaut $\sum_{n=0}^{+\infty} nq^n = \frac{q}{(1-q)^2}$.

Solution

On a $\sum_{n \geq 0} nq^n = q \sum_{n \geq 0} nq^{n-1}$ pour tout réel q .

Le reste n'est qu'un simple passage à la limite : $\sum_{n \geq 0} nq^{n-1}$ étant convergente, on a que

$\sum_{n \geq 0} nq^n$ est convergente et

$$\sum_{n \geq 0} nq^n = q \sum_{n \geq 0} nq^{n-1} = \frac{q}{(1-q)^2}$$

III-4 Série "dérivée seconde" $(\sum_{n \in \mathbb{N}} n(n-1)q^{n-2})_{n \in \mathbb{N}}$

Théorème 14

■ $\sum_{n \geq 2} n(n-1)q^{n-2}$ est convergente si et seulement si $|q| < 1$.

■ Si $|q| < 1$ la somme totale existe et vaut $\sum_{k=2}^{+\infty} k(k-1)q^{k-2} = \frac{2}{(1-q)^3}$

Moyen mnémotechnique : Pour retenir ce résultat, comme pour $\sum nq^{n-1}$, on "dérive" terme à terme par rapport à q^n l'expression $\frac{1}{(1-q)^2} = \sum_{k=0}^{+\infty} kq^{k-1}$, pour obtenir

$$\frac{2}{(1-q)^3} = \sum_{k=0}^{+\infty} k(k-1)q^{k-2}.$$

⚠ INTERPRÉTATION FAUSSE

Encore une fois, ne pas voir dans ce moyen mnémotechnique un quelconque résultat sur les dérivées de séries!

Démonstration :

• Supposons maintenant que $|q| \geq 1$.

Alors, pour tout $n \in \mathbb{N}, n \geq 2, n(n-1)|q|^{n-2} \geq 1$ ce qui entraîne la divergence grossière.

• Supposons maintenant que $|q| < 1$. Montrons que la série converge.

Méthode 1 : par dérivée

On pose $g_n(q) = \sum_{k=1}^n kq^{k-1}$, alors

$$g'_n(q) = \sum_{k=0}^n k(k-1)q^{k-2}$$

Dans la démonstration de la partie précédente, on a obtenu

$$g_n(q) = \sum_{k=1}^n kq^{k-1} = \frac{nq^{n+1} - (n+1)q^n + 1}{(1-q)^2}.$$

Autrement dit, on obtient

$$\begin{aligned} g'_n(q) &= \left(\frac{nq^{n+1} - (n+1)q^n + 1}{(1-q)^2} \right)' \\ &= \frac{(n(n+1)q^n - n(n+1)q^{n-1})(1-q)^2 + 2(1-q)(nq^{n+1} - (n+1)q^n + 1)}{(1-q)^4} \\ &= \frac{(n(n+1)q^n - n(n+1)q^{n-1})(1-q) + 2(nq^{n+1} - (n+1)q^n + 1)}{(1-q)^3} \end{aligned}$$

Par passage à la limite, on obtient $\lim_{n \rightarrow +\infty} g'_n(q) = \frac{2}{(1-q)^3}$. Ainsi, on en déduit la convergence de la série, ainsi que sa somme totale

$$\sum_{k=0}^{+\infty} k(k-1)q^{k-2} = \frac{2}{(1-q)^3}$$

Méthode 2 : par introduction du "1"... ou d'un "2" ...

On pose $S_n = \sum_{k=0}^n q^k$, $S'_n = \sum_{k=0}^n kq^{k-1}$ et $S''_n = \sum_{k=0}^n k(k-1)q^{k-2}$.

On a

$$\begin{aligned} S''_n &= \sum_{k=0}^n k(k-1)q^{k-2} = \sum_{k=2}^n k(k-1)q^{k-2} \\ &= \sum_{k=2}^n (k-2+2)(k-1)q^{k-2} = \sum_{k=2}^n (k-1)(k-2)q^{k-2} + 2 \sum_{k=2}^n (k-1)q^{k-2} \\ &= \sum_{k=1}^{n-1} k(k-1)q^{k-1} + 2 \sum_{k=1}^{n-1} kq^{k-1} \\ &= qS''_{n-1} + 2S'_{n-1} = q(S''_n - n(n-1)q^{n-2}) + 2S'_{n-1} \end{aligned}$$

On en déduit que

$$(1-q)S''_n = 2S'_{n-1} - n(n-1)q^{n-1} \xrightarrow{n \rightarrow +\infty} \frac{2}{(1-q)^2}$$

D'où le résultat annoncé. □

? Exercice 7

Montrer que $\sum_{n \in \mathbb{N}} n^2 q^n$ est convergente si et seulement si $|q| < 1$ et que, dans ce cas, la somme totale vaut $\sum_{n=0}^{+\infty} n^2 q^n = \frac{q(1+q)}{(1-q)^3}$.

Solution

On pose $S_n(q) = \sum_{k=0}^n k^2 q^k$.

- Si $|q| \geq 1$, alors il y a divergence grossière.
- Si $|q| < 1$, on a

$$S_n(q) = \sum_{k=0}^n k(k-1+1)q^k = \sum_{k=0}^n k(k-1)q^k + \sum_{k=0}^n kq^k = q^2 \sum_{k=0}^n k(k-1)q^{k-2} + \sum_{k=0}^n kq^k$$

Or, on sait que $\sum_{k=0}^n k(k-1)q^k$ et $\sum_{k=0}^n kq^k$ sont toutes deux convergentes, d'où

$$\begin{aligned} \sum_{k=0}^{+\infty} k^2 q^k &= q^2 \frac{2}{(1-q)^3} + \frac{q}{(1-q)^2} \\ &= q \left(\frac{2q}{(1-q)^3} + \frac{1-q}{(1-q)^3} \right) \\ &= \frac{q(1+q)}{(1-q)^3} \end{aligned}$$

? Exercice 8

De manière générale, montrer que, pour $a \in \mathbb{N}$, si $|q| < 1$, la série $\sum_{n \in \mathbb{N}} n(n-1)\dots(n-a+1)q^{n-a}$ est convergente et

$$\sum_{n=0}^{+\infty} \frac{n!}{(n-a)!} q^{n-a} = \sum_{n=0}^{+\infty} n(n-1)\dots(n-a+1)q^{n-a} = \frac{a!}{(1-q)^{a+1}}$$

Solution

Cette démonstration peut se faire par récurrence sur \mathbb{N} en utilisant les méthodes d'introduction déjà vues. Notamment, par

$$\sum_{n=0}^{+\infty} n(n-1)\dots(n-a+1)q^{n-a} = \sum_{n=0}^{+\infty} (n-a+a)(n-1)\dots(n-a+1)q^{n-a}$$

III-5 Série exponentielle

Définition

On appelle *série exponentielle* toute série $\sum_{n \in \mathbb{N}} \frac{x^n}{n!}$, où $x \in \mathbb{R}$.

Proposition 15

- Pour tout $x \in \mathbb{R}$, la série $\sum_{n \in \mathbb{N}} \frac{x^n}{n!}$ est convergente.
- $\sum_{k=0}^{+\infty} \frac{x^k}{k!} = e^x$

Démonstration :

Cette démonstration est hors programme et ne figure ici qu'à titre "culturel!"

On pose $S_n = \sum_{k=0}^n \frac{x^k}{k!}$. On montre facilement par récurrence sur $n \in \mathbb{N}$ que

$$e^x = \sum_{k=0}^n \frac{x^k}{k!} + \int_0^x \frac{(x-t)^n}{n!} e^t dt$$

ce qui donne

$$S_n = e^x - \int_0^x \frac{(x-t)^n}{n!} e^t dt.$$

On pose $I(x) = \int_0^x \frac{(x-t)^n}{n!} e^t dt$. Par changement de variable $t = \frac{t}{u}$, on obtient

$$I(x) = x^{n+1} \int_0^1 \frac{(1-t)^n}{n!} e^{xu} du$$

Ainsi,

$$\begin{aligned} |I(x)| &\leq |x^{n+1}| \int_0^1 \left| \frac{(1-t)^n}{n!} e^{xu} \right| du \\ &\leq \frac{|x^{n+1}|}{n!} \underbrace{\int_0^1 e^{xu} du}_{\text{constant par-rapport à } n} \xrightarrow{n \rightarrow +\infty} 0 \end{aligned}$$

□

Méthode 4 - Calcul explicite de la somme totale

Pour déterminer la convergence puis la somme totale d'une série $\sum_{n \in \mathbb{N}} u_n$, on peut chercher à décomposer en somme de séries connues.

? Exercice 9

Déterminer la convergence et la somme totale de $\sum_{n \in \mathbb{N}} \frac{n+4}{3^n}$

IV Et si on change l'ordre des termes ?

Dans une somme finie de nombres réels, on sait (commutativité dans \mathbb{R}) que l'on peut additionner les nombres dans n'importe quel ordre. Plus précisément, pour toute permutation $\varphi : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ d'indices, on sait que

$$a_1 + \dots + a_n = a_{\varphi(1)} + \dots + a_{\varphi(n)}$$

La question ici est de savoir si oui ou non le même raisonnement est vrai dans le cas d'une somme infinie de termes. Nous allons tenter de nous convaincre graphiquement que la réponse est NON !

Considérons la "série harmonique alternée" $S = \sum_{n \geq 1} \frac{(-1)^{k+1}}{k}$ convergente (vers $\ln 2$). C'est la somme, "de gauche à droite", des éléments

$$1, \quad -\frac{1}{2}, \quad \frac{1}{3}, \quad -\frac{1}{4}, \quad \frac{1}{5}, \quad -\frac{1}{6}, \quad \frac{1}{7}, \dots$$

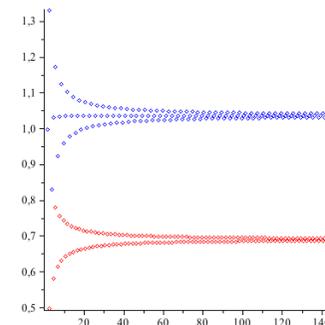
Tentons maintenant d'additionner les termes d'une autre manière. Nous prendrons 2 positifs pour un négatif. "De gauche à droite" :

$$1, \quad \frac{1}{3}, \quad -\frac{1}{2}, \quad \frac{1}{5}, \quad \frac{1}{7}, \quad -\frac{1}{4}, \quad \dots$$

Nous admettrons que cette série converge également.

Ces deux séries (série harmonique en bas ... en rouge, l'autre en haut ... en bleu) sont

représentées sur le graphique suivant :



Si on veut bien admettre que nous pouvons faire confiance au graphique en ce qui concerne le comportement asymptotique, nous constaterons que les sommes totales sont nettement différentes, ce qui devrait nous convaincre qu'en cas de somme infinie, en général, il faut faire attention à l'ordre dans lequel on somme les éléments !

Toutefois, la majorité des cas raisonnables que nous aurons à étudier ne poseront pas ce type de problèmes. En effet :

Théorème 16

Si une série $(\sum_{n \in \mathbb{N}} u_n)$ est absolument convergente, alors la somme des u_n peut se faire dans n'importe quel ordre.

Démonstration : admise. \square

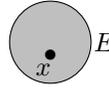
Corollaire

Si une suite $(u_n)_{n \in \mathbb{N}}$ positive est telle que $(\sum_{n \in \mathbb{N}} u_n)$ converge, alors, alors la somme des u_n peut se faire dans n'importe quel ordre.

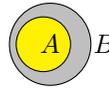
I Vocabulaire des ensembles et des applications

I-1 Ensembles

■ **Élément / appartenance** : Quand x est dans l'ensemble E , on note $x \in E$. On dit que x est un *élément* de E ou que x *appartient* à E .

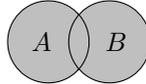


■ **Inclusion** : On dit que l'ensemble A est *inclus* dans l'ensemble B et on note $A \subset B$ si tous les éléments de A sont aussi dans B . On dit également que A est un *sous ensemble* de B .



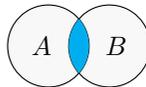
■ **Réunion** : Si A et B sont deux sous ensembles d'un ensemble E , on appelle *réunion* des ensembles A et B et on note $A \cup B$ l'ensemble contenant tous les éléments de A et B . Ainsi

$$x \in A \cup B \iff x \in A \text{ ou } x \in B$$

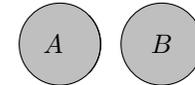


■ **Intersection** : Si A et B sont deux sous ensembles d'un ensemble E , on appelle *intersection* des ensembles A et B et on note $A \cap B$ l'ensemble ne contenant que les éléments qui sont à la fois dans A et dans B . Ainsi

$$x \in A \cap B \iff x \in A \text{ et } x \in B$$



■ **Ensembles disjoints / incompatibles** : Si A et B sont deux sous ensembles d'un ensemble E , on dit qu'ils sont *disjoints* ou *incompatibles* si A et B n'ont aucun élément en commun. On note $A \cap B = \emptyset$.



VOCABULAIRE

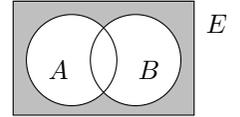
Il ne faut pas confondre "disjoints" et "distincts". (Distincts signifiant simplement $A \neq B$.)

■ **Complémentaire** : Si A est un sous ensemble d'un ensemble E , on appelle *complémentaire de A (dans E)* l'ensemble noté A^c ou \bar{A} , constitué de tous les éléments de E qui ne sont pas dans A . Ainsi, si $x \in E$,

$$x \in \bar{A} \iff x \notin A$$

S'il peut y avoir confusion d'ensembles " E ", on note plutôt $E \setminus A$ (pour E privé de A). Ainsi,

$$x \in E \setminus A \iff x \in E \text{ et } x \notin A$$



■ **Partition** : Soit $\mathcal{F} = \{A_i\}_{i \in I}$ une famille de sous ensembles d'un ensemble E . On dit que \mathcal{F} est une *partition* de E si et seulement si

$$\begin{cases} A_i \cap A_j = \emptyset & \text{si } i \neq j, \text{ où } i, j \in I \\ \bigcup_{i \in I} A_i = E \end{cases}$$

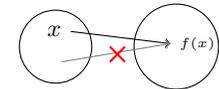


■ **Produit cartésien de n ensembles**. Soit $\mathcal{F} = \{A_i\}_{i=1, \dots, n}$ une famille de sous ensembles d'un ensemble E . On note $A_1 \times \dots \times A_n$ l'ensemble formé de tous les n -uplets (x_1, \dots, x_n) où $x_1 \in A_1, \dots, x_n \in A_n$.

I-2 Applications

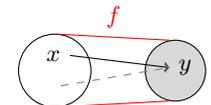
■ **Injection** : On dit qu'une application $f : E \rightarrow F$ est *injective* s'il n'existe au maximum qu'un antécédent par f pour chaque élément de F , autrement dit,

$$f(x) = f(y) \implies x = y$$



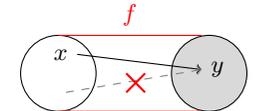
■ **Surjection** : On dit qu'une application $f : E \rightarrow F$ est *surjective* s'il existe au minimum un antécédent par f pour chaque élément de F , autrement dit,

$$y \in Y \implies \exists x \in E \text{ tel que } y = f(x)$$



■ **Bijection** : On dit qu'une application $f : E \rightarrow F$ est *bijective* s'il existe exactement un antécédent par f pour chaque élément de F , c'est-à-dire que f est injective et surjective, autrement dit,

$$y \in Y \implies \exists! x \in E \text{ tel que } y = f(x)$$



■ **Composée de deux bijections** : Si $f : E \rightarrow F$ et $g : F \rightarrow G$ sont bijectives, alors $g \circ f$ est bijective.

■ **Image d'une partie** : Si $f : E \rightarrow F$ et $A \subset E$, on note $f(A) = \{f(x) \mid x \in A\} \subset F$.

■ **Application réciproque** : Si $f : E \rightarrow F$ est une application bijective, on note $f^{-1} : F \rightarrow E$ l'unique application telle que

$$f^{-1} \circ f(x) = x \quad \forall x \in E$$

II Dénombrement

II-1 Notion de cardinal

■ **Cardinal** : On appelle *cardinal* d'un ensemble fini E le nombre d'éléments contenus dans E . On note ce nombre $\text{card } E$ ou quelquefois $|E|$.

Dénombrer des cas revient à déterminer le cardinal d'un ensemble.

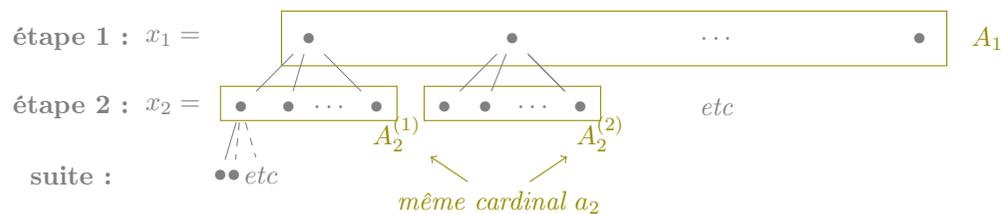
■ **Cardinal et bijection** : Si E est un ensemble fini et $f : E \rightarrow F$ une bijection, alors F est fini et $\text{card } E = \text{card } F$.

"Modéliser" une situation signifie "réaliser une bijection avec un ensemble plus simple à dénombrer".

■ **Principe multiplicatif** : Si $A = A_1 \times \dots \times A_n$, alors

$$\text{card } A = \text{card } A_1 \times \text{card } A_2 \times \dots \times \text{card } A_n$$

De manière générale, si les éléments de A peuvent se décomposer comme des n -listes (x_1, \dots, x_n) liées par exemple à des étapes successives dont les possibilités $A_i^{(j)}$ pour l'étape i fixée sont en nombre identique ($\text{card}(A_i^{(1)}) = \text{card}(A_i^{(2)}) = \dots$)



$$\text{Alors } \text{card } A = a_1 \times a_2 \times \dots \times a_n$$

$$\text{avec } a_1 = \text{card } A_1, a_2 = \text{card } A_2^{(1)} = \text{card } A_2^{(2)} = \dots, a_3 = \dots$$

■ **Cardinal de l'ensemble des parties** : Si E est un ensemble fini. On note

$$\mathcal{P}(E) = \{A \subset E\}$$

c'est-à-dire l'ensemble constitué de tous les sous ensembles de E . On a

$$\text{card } \mathcal{P}(E) = 2^n, \quad \text{où } n = \text{card } E$$

Par exemple, si $E = \{1, 2, 3\}$, on a

$$\mathcal{P}(E) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \underbrace{\{1, 2, 3\}}_E\}$$

On a $n = 3$ éléments dans E et retrouve bien $2^3 = 8$ éléments dans $\mathcal{P}(E)$.

II-2 Formules

■ **Nombre d'arrangements** : Pour $n, p \in \mathbb{N}^*$, $p \leq n$, on appelle *nombre d'arrangements de p nombres parmi n* le nombre d'éléments de l'ensemble

$$\{(i_1, \dots, i_p) \mid i_k \in \llbracket 1; n \rrbracket \forall k = 1, \dots, p \text{ et où les } i_k \text{ sont deux à deux distincts}\}$$

c'est

$$\frac{n!}{(n-p)!}$$

■ **Nombre de permutations** : Pour $n \in \mathbb{N}^*$, on appelle *nombre de permutations de n éléments* le nombre d'éléments de l'ensemble

$$\{(i_1, \dots, i_n) \mid i_k \in \llbracket 1; n \rrbracket \forall k = 1, \dots, n \text{ et où les } i_k \text{ sont deux à deux distincts}\}$$

c'est

$$n!$$

■ **Nombre de combinaisons** : Pour $n, p \in \mathbb{N}^*$, $p \leq n$, on appelle *nombre de combinaisons de p nombres parmi n* le nombre d'éléments de l'ensemble

$$\{\{i_1, \dots, i_p\} \mid i_k \in \llbracket 1; n \rrbracket \forall k = 1, \dots, p \text{ et où les } i_k \text{ sont deux à deux distincts}\}$$

c'est

$$\binom{n}{p} = \frac{n!}{p!(n-p)!}$$

■ **Formule du binôme par dénombrement** : Soient $x, y \in \mathbb{R}$ et $n \in \mathbb{N}$. On rappelle que

$$(x+y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}$$

Démonstration :

$$(x+y)^n = \underbrace{(x+y)(x+y)\dots(x+y)}_{n \text{ fois}}$$

Le développement de l'expression de droite par distributivité donne toutes les sommes des monômes du type $x^k y^{n-k}$: lorsqu'on a choisi dans quelle parenthèse prendre les " k x ", on prend obligatoirement les " y " dans les $n - k$ parenthèses restantes. Il s'en suit donc que l'on a autant de monômes $x^k y^{n-k}$ que de choix des k " x " possibles dans les n parenthèses, c'est-à-dire $\binom{n}{k}$. \square

VERSION ALLEGÉE
 Pour éviter tout excès de technicité, certaines définitions utilisées dans ce chapitre sont un peu édulcorées par rapport à celles véritablement utilisées par les probabilistes. Elles sont cependant à notre programme telles que données ici.

I Le triplet fondamental

I-1 Cas des probabilités finies

I.1-a) Rappel de définition et propriétés élémentaires

Notation : Pour un ensemble Ω , on note $\mathcal{P}(\Omega)$ l'ensemble de toutes les parties de Ω :

$$\mathcal{P}(\Omega) = \{A \mid A \subset \Omega\}$$

Remarque :
 Si Ω est de cardinal fini n , on a $\text{card}(\mathcal{P}(\Omega)) = 2^n$.

Définition (rappel)
 Pour tout ensemble **fini** non vide Ω , on appelle *mesure (ou loi) de probabilité* p sur $(\Omega, \mathcal{P}(\Omega))$ toute application $p : \mathcal{P}(\Omega) \rightarrow \mathbb{R}$ telle que

- i ■ $\forall A \in \mathcal{P}(\Omega)$, on a $0 \leq p(A) \leq 1$
- ii ■ $p(\emptyset) = 0$; $p(\Omega) = 1$
- iii ■ Pour tout couple (A, B) d'éléments incompatibles de $\mathcal{P}(\Omega)$, on a

$$p(A \sqcup B) = p(A) + p(B)$$

Le triplet $(\Omega, \mathcal{P}(\Omega), p)$ est alors appelé *espace probabilisé*.

Étant donné un espace probabilisé fini $(\Omega, \mathcal{P}(\Omega), p)$, on rappelle également le vocabulaire suivant :

- on appelle Ω l'*ensemble fondamental* (ou l'*univers*).
- Un élément de Ω est appelé *épreuve* (ou événement fondamental).
- Un élément de $\mathcal{P}(\Omega)$ est appelé *évènement*.

? Exercice 1

Montrer que la condition $p(\emptyset) = 0$ de la définition est redondante.

Solution

On a $\Omega = \emptyset \cup \Omega$ une réunion d'ensembles incompatibles. Ainsi, on a

$$p(\Omega) = p(\emptyset) + p(\Omega)$$

Il suffit d'isoler $p(\emptyset)$ d'un côté de l'égalité pour obtenir directement

$$p(\emptyset) = p(\Omega) - p(\Omega) = 0$$

Propriété 17

On rappelle également que la propriété iii de la définition nous amène à un résultat plus général qui est le suivant :

Si $A_1, \dots, A_n \in \mathcal{P}(\Omega)$ sont des ensembles deux à deux incompatibles, alors

$$p(A_1 \sqcup \dots \sqcup A_n) = p(A_1) + \dots + p(A_n)$$

■ Exemple 1 :

Une compagnie pétrolière dispose de 4 bateaux numérotés de 1 à 4. Les bateaux 3 et 4 sont plus récents. L'expérience est la suivante : pour faire un voyage, le compagnie choisit un des bateaux, avec une probabilité deux fois plus importante de choisir un bateau récent qu'un bateau ancien.

On modélise cette expérience par le triplet fondamental $(\Omega, \mathcal{P}(\Omega), p)$ suivant :
 $\Omega = \{1, 2, 3, 4\}$ p décrit par

i	1	2	3	4
$p(\{i\})$	a	a	$2a$	$2a$

où $a \in [0, 1]$. Pour trouver a , il nous suffit d'utiliser la propriété ci-dessus avec les 4 ensembles 2 à 2 incompatibles $A_i = \{i\}$ pour $i \in \Omega$. Ainsi :

$$a + a + 2a + 2a = p(A_1) + \dots + p(A_4) = p(A_1 \cup \dots \cup A_4) = p(\Omega) = 1$$

D'où

$$a = \frac{1}{6}$$

On peut également trouver de la même manière par exemple

$$p(\{1, 3\}) = p(\{1\}) + p(\{3\}) = \frac{1}{6} + \frac{1}{3} = \frac{1}{2}$$

! Remarque :

Généralement, à des fins de simplicité d'écriture, on préfère utiliser l'abus de notation $p(i)$ à la place de $p(\{i\})$ si $i \in \Omega$.

Première problématique : connaît-on Ω ?

Pour une expérience donnée, on a deux possibilités pour l'univers :
 - **on connaît à l'avance** exactement l'ensemble des résultats possibles (*par exemple, un lancé de dé à 6 faces*)
 - **on ne connaît pas à l'avance** l'ensemble des résultats (*par exemple, si on étudie l'ensemble les effets secondaires d'un médicament*)
 Dans le premier cas, on peut énoncer clairement un univers modélisant l'expérience (*par exemple, $\Omega = \{1, \dots, 6\}$*).
 Dans le deuxième cas, on sait simplement qu'il existe. On ne connaît pas ses éléments et on ne sait pas du tout combien il en contient. On doit donc se contenter d'exploiter ce qu'on sait.

Deuxième problématique : Peut-on se contenter d' Ω fini ?

Faisons une expérience simple :

On lance une pièce de monnaie et on s'arrête dès que l'on obtient "pile".
 On s'intéresse dans un premier temps aux résultats exacts de l'expérience.
 On peut par exemple représenter chaque épreuve par une série de lettres P et F qui représentent respectivement l'obtention d'un pile ou d'un face. (FFFP représenterait donc l'obtention de trois "faces" avant l'obtention du premier pile.) On a

$$\Omega = \{P, FP, F \dots FP \mid \text{la longueur de } F \dots F \text{ est quelconque.}\}$$

L'univers ainsi constitué peut-il être fini? Non. (*à méditer*). **Peut-on construire un univers fini pour décrire l'ensemble des résultats de cette expérience? Non.** (*à méditer*).

On s'intéresse dans un second temps plus simplement au rang d'apparition du premier pile. Par exemple, ce nombre serait "4" dans le cas de "FFFP"). En modélisant par 0 si Pile n'apparaît jamais, L'univers obtenu est

infini là aussi... $\Omega = \mathbb{N}$

Troisième problématique : peut-on appliquer la définition d'une mesure de probabilité finie a à Ω infini ?

La réponse à cette question est là aussi : **Non**. Pour plusieurs raisons :

- 1 ■ Avec la définition "basique" donnée jusque là, on ne peut pas calculer la probabilité d'une réunion infinie d'événements.
- 2 ■ Se contenter d'utiliser comme ensemble d'événements $\mathcal{P}(\Omega)$ se révèle en réalité trop complexe pour l'étude des probabilités en général, par exemple dans le cas où $\Omega = \mathbb{R}$ qui contient trop de sous-ensembles "compliqués" (et HP) qu'il est inutile de prendre en compte lors de calculs de probabilités.

Cet aspect technique de construction d'univers sera très brièvement évoqué cette année. (*cf ci-après*.)

I.2-a) Réunions et intersections "dénombrables"

Commentaires :

*Un univers constitué d'une infinité d'éléments doit certainement pouvoir se "découper" en une infinité d'ensembles !
 Par exemple, $\Omega = \mathbb{N}$ est constitué d'une infinité de petits morceaux qui sont les $\{i\}$ pour tout entier naturel i . Par le principe de réunion, vient alors naturellement l'idée d'écrire*

$$\mathbb{N} = \bigcup_{i=0}^{+\infty} \{i\}$$

C'est une réunion "infinie" qu'il nous faut donc définir mathématiquement (on en profite pour décrire également les intersections !) :



Définition

Les opérateurs de réunion et d'intersection infinie ont le même sens que dans le cadre fini :

- 1 ■ L'événement $\bigcup_{n=1}^{+\infty} A_n$ correspond à *"au moins un des événements A_n se produit."*
- 2 ■ L'événement $\bigcap_{n=1}^{+\infty} A_n$ correspond à *tous les événements A_n se produisent.*

Une réunion (resp. intersection) avec des indices dans \mathbb{N} s'appelle une *réunion (resp. intersection) dénombrable*.



Remarque :

Pour ce qui est des probabilités, si on a un univers $\Omega = \mathbb{N} = \bigcup_{i=0}^{+\infty} \{i\}$, on aura alors très rapidement envie (*oui, oui !*) d'écrire par exemple

$$p(\mathbb{N}) = \sum_{i=0}^{+\infty} p(i)$$

mais pour l'instant, la définition de probabilités finies ne nous le permet pas. (On peut faire seulement des sommes finies.)

I.2-b) Notion de tribu

■ Exemple 2 :

Imaginons par exemple l'expérience suivante :

On lance deux fléchettes sur une cible carré lointaine de côté 1 mètre. La distance entre les deux impacts nous amène alors à considérer l'univers $\Omega = [0, \sqrt{2}]$ (unité en mètre). On a donc un univers qui n'est ni fini, ni \mathbb{N} , mais un intervalle de \mathbb{R} .

Commentaires :

Dans le cadre de probabilités en général, on souhaite pouvoir :

- calculer des probabilités de réunions (finies ou dénombrables)
- calculer des probabilités d'intersections (finies ou dénombrables)
- calculer des probabilités de complémentaires
- dire que $p(\Omega) = 1$
- et si $\Omega = \mathbb{R}$ (ou tout intervalle de \mathbb{R}), on veut pouvoir calculer des probabilités d'intervalles (ouverts, fermés, semi-ouverts)

Si on conserve la définition

$$p : \mathcal{P}(\Omega) \rightarrow [0; 1]$$

(Non, le domaine de définition d'une probabilité n'est pas Ω !)

La généralisation du reste de la définition devrait prendre en compte le fait que \mathbb{R} contient des sous-ensembles "compliqués", qui ne peuvent par exemple pas s'écrire sous une forme de réunion ou d'intersection dénombrable. (exemple HP)

Plutôt que de compliquer inutilement cette définition en tenant compte d'ensembles auxquels on ne s'intéresse pas en calcul de probabilités, on va plutôt restreindre le domaine de définition de p à des cas "plus simples à étudier" :

Imaginons donné un triplet (Ω, \mathcal{T}, p) , avec $\mathcal{T} \subset \mathcal{P}(\Omega)$. Si $p : \mathcal{T} \rightarrow [0; 1]$ doit être une probabilité, il est clair qu'elle doit au moins conserver les propriétés de base d'une probabilité finie ainsi que celles évoquées ci-dessus. On doit donc au moins avoir

- $\emptyset \in \mathcal{T}$ et $\Omega \in \mathcal{T}$;
- $A \in \mathcal{T} \Rightarrow \bar{A} \in \mathcal{T}$;
- $A, B \in \mathcal{T} \Rightarrow A \cup B \in \mathcal{T}$. Plus généralement $A_1, \dots, A_n, \dots \in \mathcal{T} \Rightarrow \bigcup_{n=1}^{+\infty} A_n \in \mathcal{T}$.
- $A, B \in \mathcal{T} \Rightarrow A \cap B \in \mathcal{T}$. Plus généralement $A_1, \dots, A_n, \dots \in \mathcal{T} \Rightarrow \bigcap_{n=1}^{+\infty} A_n \in \mathcal{T}$.

En retirant les conditions redondantes, on obtient la définition suivante :

Définition

Soit Ω un ensemble fondamental et $\mathcal{T} \subset \mathcal{P}(\Omega)$. On dit que \mathcal{T} est une *tribu* sur Ω si

- $\Omega \in \mathcal{T}$,
- \mathcal{T} est stable par passage au complémentaire, (i.e. si $A \in \mathcal{T}$, alors $\bar{A} \in \mathcal{T}$.)
- \mathcal{T} est stable par réunion dénombrable : (i.e. pour toute suite $(A_n)_{n \in \mathbb{N}}$ de \mathcal{T} , on a $\bigcup_{n=0}^{+\infty} A_n \in \mathcal{T}$.)

Propriété 18

Si \mathcal{T} est une tribu, alors elle est également stable par intersection.

Démonstration : admise. \square

■ Contre-Exemple(s) :

- 3 ■ Si $\Omega = \{1, 2, 3, 4\}$, l'ensemble $\{\emptyset, \Omega, \{1\}, \{1, 2\}\}$ n'est pas une tribu.
- 4 ■ Si $\Omega = \mathbb{Z}$, l'ensemble $\{\emptyset, \Omega, \{1\}, \{2\}, \{3\}, \dots\}$ n'est pas une tribu.

■ Exemples :

Soit Ω un ensemble fondamental.

- 5 ■ $\{\emptyset, \Omega\}$ est une tribu.
- 6 ■ $\mathcal{P}(\Omega)$ est une tribu quelquesoit Ω .
- 7 ■ Si $\Omega = \{1, 2, 3, 4\}$, une tribu sur Ω peut être $\mathcal{T} = \{\emptyset, \Omega, \{1, 2\}, \{3, 4\}\}$
- 8 ■ Si $\Omega = \mathbb{N}$, une tribu sur Ω peut être $\mathcal{T} = \{\emptyset, \Omega, 2\mathbb{N}, 2\mathbb{N} + 1\}$

Définition

Pour tout ensemble Ω et toute tribu \mathcal{T} sur Ω , on appelle le couple (Ω, \mathcal{T}) un *espace probabilisable* (ou *espace mesurable*).

I.2-c) La tribu des Boréliens

Commentaires :

Si l'ensemble fondamental est \mathbb{R} ou un intervalle de \mathbb{R} , au vue des raisons évoquées précédemment, on choisit donc généralement une tribu plus petite que $\mathcal{P}(\Omega)$:

Définition et Proposition

Si l'univers $\Omega \subset \mathbb{R}$ est un intervalle, il existe une tribu ayant les propriétés suivantes :

- elle contient l'ensemble $\mathcal{C} = \{[a, b] \mid a, b, \in \Omega\}$
- il n'existe pas de tribu plus petite contenant \mathcal{C}

On la note $\mathcal{B}(\Omega)$. On l'appelle *tribu des boréliens* sur Ω .

Concrètement :

On peut montrer (**et on va simplement retenir**) que cette tribu fait tout ce dont on a besoin dans \mathbb{R} : elle contient les intervalles de tout type (fermés, bornés, semi-ouverts) leurs réunion, intersections, complémentaires, etc. . .

De plus, conformément à ce qui a évoqué en début de paragraphe, $\mathcal{B}(\mathbb{R}) \neq \mathcal{P}(\mathbb{R})!$ (*démo toujours HP*)

I.2-d) Espace probabilisé

Commentaires :

| Tout d'abord, on récupère le vocabulaire des espaces finis :

Définition

Étant donné un espace probabilisé (Ω, \mathcal{T}, p) ,

- on appelle Ω l'*ensemble fondamental* (ou l'*univers*).
- Un élément de Ω est appelé *épreuve*.
- Un élément de \mathcal{T} est appelé *événement*.
- Pour un événement $A \in \mathcal{T}$, l'événement $\bar{A} = A^c$ est appelé *événement contraire de A*.
- Si deux événements A et B sont disjoints, alors ils sont dits *incompatibles*.

Commentaires :

| Ensuite on généralise la notion de mesure de probabilités :

Définition

Étant donné un ensemble Ω et une tribu \mathcal{T} sur Ω , on appelle *mesure de probabilité* p sur (Ω, \mathcal{T}) toute application $p : \mathcal{T} \rightarrow \mathbb{R}$ telle que

- 1 ■ $\forall A \in \mathcal{T}$, on a $0 \leq p(A) \leq 1$
- 2 ■ $p(\emptyset) = 0$; $p(\Omega) = 1$
- 3 ■ *axiome de Σ -additivité* :

Pour toute suite $(A_n)_{n \in \mathbb{N}}$ d'événements deux à deux incompatibles de \mathcal{T} , la série $\sum_{n \in \mathbb{N}} p(A_n)$ est convergente et $p\left(\bigsqcup_{n=0}^{+\infty} A_n\right) = \sum_{n=0}^{+\infty} p(A_n)$.

Le triplet (Ω, \mathcal{T}, p) est alors appelé *espace probabilisé*.

Remarque :

- Si on a $A_n = \emptyset$ (et donc $p(A_n) = 0$) pour tout $n > N$, on obtient alors, **sous réserve d'incompatibilité des A_n** : $p\left(\bigsqcup_{n=0}^N A_n\right) = \sum_{n=0}^N p(A_n)$.
- Si on suppose Ω fini, $\mathcal{T} = \mathcal{P}(\Omega)$ et que l'on se donne $A_n = \emptyset \forall n \geq 2$ (et donc $p(A_n) = 0$) dans la troisième hypothèse, on **retrouve la définition d'une mesure de probabilité finie** donnée en première année.

Remarque :

- **L'ordre dans lequel on fait la somme des $p(A_n)$ n'importe pas.** On rappelle en effet que nous avons vu dans le chapitre sur les séries que si la série est convergente et de termes positifs (ce qui est le cas ici), on pouvait sommer dans n'importe quel ordre.
- **Si les A_n sont bien incompatibles**, la convergence de la série $\sum_{n \in \mathbb{N}} p(A_n)$ est en réalité une **hypothèse redondante**. En effet, on sait qu'elle est forcément convergente car c'est une série à termes positifs et majorée (par 1) :

$$\forall n \in \mathbb{N}, \quad \sum_{k=0}^n p(A_k) = p\left(\bigsqcup_{k=0}^n A_k\right) \leq 1$$
- Cette année, les tribus utilisées seront, soit $\mathcal{P}(\Omega)$ (probabilités finies ou discrètes), soit la tribu des boréliens (probabilités à densité.)

Commentaires :

| Observons maintenant l'émergence de nouveaux type d'événements :

Exemple 9 :

Choisissons au hasard un nombre dans l'intervalle $[0, 1]$. Intuitivement, réfléchir à la probabilité possible de l'événement A : "Obtenir 0" ainsi qu'à l'événement B : "Obtenir un nombre contenu dans l'intervalle $]0, 1[$ ".

On ne démontrera pas encore les choses ici mais on peut imaginer que si on généralise (**de manière totalement scandaleuse !**) le principe d'équiprobabilité vu dans le cadre des probas finies, on devrait avoir

$$p(A) = \frac{\text{nombre de cas favorables}}{\text{nombre total de cas}} = \ll \frac{1}{\infty} \gg = 0$$

Même si on n'est pas encore en capacité de démontrer les choses proprement, on voit mal quel pourrait être une autre probabilité pour cet événement. **On a donc un événement de probabilité nulle, mais qui n'est pas impossible !**

Pour ce qui est de $p(]0, 1[)$, on écrira plutôt :

$$p(]0, 1[) = 1 - (p(0) + p(1)) = 1 - 0 = 1$$

On a donc un événement de probabilité 1, mais qui n'est pas l'univers !

Définition

Étant donné un espace probabilisé (Ω, \mathcal{T}, p) ,

- Si $A \in \mathcal{T}$ est tel que $A \neq \emptyset$ et $p(A) = 0$, on dit que A est *négligeable*.
- Si $A \in \mathcal{T}$ est tel que $A \neq \Omega$ et $p(A) = 1$, on dit que A est *presque sûr*.

I-3 Cohérence avec les propriétés connues

Remarque :
Les propriétés ci-dessous, déjà connues dans le cadre des probabilités finies se généralisent aisément à toutes les mesures de probabilités. (à traiter en exercice)

Propriété 19

On se donne un espace probabilisé (Ω, \mathcal{T}, p) et des événements $A, B \in \mathcal{T}$.

- Si A et B sont incompatibles, alors $p(A \sqcup B) = p(A) + p(B)$
- $p(\bar{A}) = 1 - p(A)$
- Si $A \subset B$, on a $p(A) \leq p(B)$ et $p(B \setminus A) = p(B) - p(A)$.
- Pour A et B quelconques, $p(A \cup B) = p(A) + p(B) - p(A \cap B)$.

? Exercice 2

Montrer que

$$p(A \cup B \cup C) = (p(A) + p(B) + p(C)) - (p(A \cap B) + p(A \cap C) + p(B \cap C)) + p(A \cap B \cap C)$$

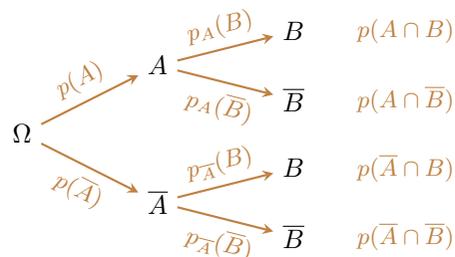
II Probabilités conditionnelles

II-1 Définition et exemples de calculs

Commentaires :

On rappelle que, dans le cas des probabilités finies, on sait que la probabilité d'un événement A connaissant déjà la réalisation d'un autre événement B tel que $p(B) \neq 0$ peut s'écrire (sous réserve de définition) $p_B(A) = \frac{p(A \cap B)}{p(B)}$

Cette formule peut s'interpréter grace de l'arbre de probabilité suivant :



Heureusement, on peut étendre cette notion à l'ensemble de toutes les mesures de probabilités.

Définition

Pour tout événement B de l'espace probabilisé (Ω, \mathcal{T}, p) , vérifiant $p(B) \neq 0$ et $A \in \mathcal{T}$, on appelle *probabilité de A sachant B* et on note $p_B(A)$ la probabilité de l'événement A "sachant que B est déjà vérifié".

Propriété 20

Avec les notations de la définition précédente, l'application

$$p_B : \mathcal{T} \rightarrow \mathbb{R}^+; A \mapsto p_B(A)$$

est bien une mesure de probabilités et on a

$$p(B)p_B(A) = p(A \cap B)$$

Démonstration : admise. \square

? CONFUSION

On voit bien sur la formule qu'il ne faut pas confondre la probabilité "conditionnelle" avec la probabilité de l'intersection !

Commentaires :

L'erreur fréquemment commise est de penser qu'il faut systématiquement utiliser la formule $p_B(A) = \frac{p(A \cap B)}{p(B)}$ pour calculer la probabilité conditionnelle, ce qui est faux. Voyons ci-dessous deux exemples de situation. Dans la première, on utilise la formule, la deuxième non.

? Exercice 3

Un couple a deux enfants. Calculer la probabilité qu'ils aient 1 fille sachant qu'ils ont au moins un garçon.

Solution

En travaillant sur l'univers $\Omega = \{(\underbrace{i}_{1^{\text{er}} \text{ enfant}}, \underbrace{j}_{2^{\text{ème}} \text{ enfant}}) \in \{G, F\}^2\}$ avec la probabilité uniforme et en posant

A : « ils ont une fille » B : « ils ont au moins un garçon »

On a

$$p_B(A) = \frac{p(A \cap B)}{p(B)} = \frac{p(\{(F, G), (G, F)\})}{p(\{(F, G), (G, F), (G, G)\})} = \frac{\frac{1}{2}}{\frac{3}{4}} = \frac{2}{3}$$

? Exercice 4

On dispose de 2 urnes appelées U_1 et U_2 . L'urne U_2 contient autant de boules rouges que de boules bleues.

On lance maintenant un dé équilibré à 6 faces. On note RP : « le résultat est pair ».

Si le résultat est pair, on tire une boule dans l'urne U_2 . On regarde la couleur de la boule obtenue. Décrire p_{RP} c'est-à-dire : donner l'univers, la tribu associée ainsi que la loi de probabilité p_{RP} .

Solution

On sait qu'on a tiré un nombre pair et qu'on tire alors une boule dans l'urne U_2 qui contient autant de boules rouges que bleues. Ainsi, on peut poser

$$\Omega = \{R, B\} \quad \text{où } R \text{ désigne le fait de tirer une boule rouge et } B \text{ une boule bleue.}$$

Alors Ω est fini. On peut donc poser comme tribu $\mathcal{T} = \mathcal{P}(\Omega)$. Étant donné qu'il y a le même nombre de boules rouges et bleues, la probabilité p_{RP} est alors l'équiprobabilité.

Commentaires :

En bref, lors d'expériences menées par étapes, il est souvent utile (et nécessaire) de déterminer la probabilité conditionnelle par d'autres moyens que la formule, par exemple par une analyse directe de la problématique en se plaçant directement dans le cadre du "sachant que". De plus, dans ce cadre, la probabilité conditionnelle sert justement à calculer les $P(A \cap B)$:

? Exercice 5

Dans le cadre de l'exercice précédent, calculer la probabilité d'avoir un résultat pair pour le dé et d'obtenir une boule rouge.

Solution

On cherche $p(RP \cap \{R\})$:

Par probabilités conditionnelles, comme $p(RP) \neq 0$, on a

$$p(RP \cap \{R\}) = p(RP)p_{RP}(\{R\})$$

Or, on peut calculer par équiprobabilité sur les résultats du dé que

$$p(RP) = \frac{1}{2}$$

$$p_{RP}(\{R\}) = \frac{1}{2}$$

D'où

$$p(\text{« avoir un résultat pair pour le dé et obtenir une boule rouge »}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

Propriété 21

Pour tout événement $A, B \in \mathcal{T}$ d'un espace probabilisé (Ω, \mathcal{T}, p) , on a

$$p(A \cap B) = p_B(A)p(B) \quad \text{si } p(B) \neq 0 \quad \text{ou} \quad p(A \cap B) = p_A(B)p(A) \quad \text{si } p(A) \neq 0$$

Notation

Si $p(B) = 0$, par convention, on va poser $p_B(A)p(B) = p(A \cap B) = 0$.

en effet, si $p(B) = 0$, comme $A \cap B \subset B$, on a également $p(A \cap B) = 0$. On généralise donc la formule $p_B(A)p(B) = p(A \cap B)$ dans ce cadre.

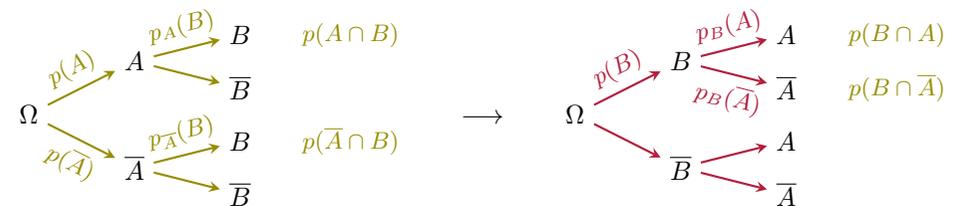
Les formules établies par la suite seront toutes écrites suivant cette convention.

II-2 Formules de Bayes

Après la généralisation de la formule des probabilités conditionnelles la formule de Bayes se généralise donc également sans modification :

Commentaires :

Rappelons que cette formule vise en quelque sorte à "inverser" les étapes, c'est-à-dire par exemple que si l'arbre de gauche ci-dessous est connu, on peut déterminer l'arbre de droite. (Cette formule n'a pas nécessité d'être apprise par coeur. Elle découle immédiatement des autres propriétés connues.)



Théorème 22 Formule de Bayes

Soit (Ω, \mathcal{T}, p) un espace probabilisé. Pour tout $A, B \in \mathcal{T}$ tels que $p(A), p(B) \neq 0$, on a que

$$p_B(A) = \frac{p(A \cap B)}{p(B)} = \frac{p_A(B)p(A)}{p(B)} = \frac{p_A(B)p(A)}{p_A(B)p(A) + p_{\bar{A}}(B)p(\bar{A})}$$

II-3 Formule des probabilités composées

Théorème 23 Formule des probabilités composées

Soient (Ω, \mathcal{T}, p) un espace probabilisé, n un entier supérieur ou égal à 2 et $A_1, \dots, A_n \in \mathcal{T}$. On a alors

$$p(A_1 \cap \dots \cap A_n) = p(A_1)p_{A_1}(A_2)p_{A_1 \cap A_2}(A_3) \dots p_{A_1 \cap \dots \cap A_{n-1}}(A_n)$$

Démonstration : exercice. \square

? Exercice 6

On dispose d'une urne contenant n boules dont 5 boules blanches. On tire sans remise 3 boules. Quelle est la probabilité d'obtenir 3 boules blanches ?

Solution

On note B_i l'événement "obtenir une boule blanche au $i^{\text{ème}}$ tirage". On cherche donc $p(B_1 \cap B_2 \cap B_3)$: D'après la formule des probabilités composées,

$$p(B_1 \cap B_2 \cap B_3) = p(B_1)p_{B_1}(B_2)p_{B_1 \cap B_2}(B_3)$$

Or, par équiabilité des tirages, on a

$$p(B_1) = \frac{5}{n} \quad ; \quad p_{B_1}(B_2) = \frac{4}{n-1} \quad ; \quad p_{B_1 \cap B_2}(B_3) = \frac{3}{n-2}$$

D'où

$$p(B_1 \cap B_2 \cap B_3) = \frac{5}{n} \cdot \frac{4}{n-1} \cdot \frac{3}{n-2}$$

Commentaires :

Cette formule permet de démontrer des résultats que l'on considère souvent comme intuitifs, alors qu'ils ont en réalité (cf formule) une justification bien précise.



Remarque :

Cette formule est souvent utilisée quand les événements A_1, \dots, A_n sont donnés dans un ordre chronologique décroissant ; $A_n \subset A_{n-1} \subset \dots \subset A_1$. La formule s'écrit alors

$$p(A_n) = p(A_1)p_{A_1}(A_2) \dots p_{A_{n-1}}(A_n)$$

? Exercice 7

On reprend l'exemple précédent. On cherche toujours la probabilité d'obtenir 3 boules blanches, mais avec la technique précédente.

Solution

On cherche donc $p(A_3)$ et on sait que $A_3 \subset A_2 \subset A_1$

D'après la formule des probabilités composées pour des événements chronologiques,

$$p(A_3) = p(A_1)p_{A_1}(A_2)p_{A_2}(A_3)$$

Or, par équiabilité des tirages, on a

$$p(A_1) = \frac{5}{n} \quad ; \quad p_{A_1}(A_2) = \frac{4}{n-1} \quad ; \quad p_{A_2}(A_3) = \frac{3}{n-2}$$

D'où

$$p(A_3) = \frac{5}{n} \cdot \frac{4}{n-1} \cdot \frac{3}{n-2}$$

III Indépendance

On se donne un espace probabilisé (Ω, \mathcal{T}, p) .

Commentaires :

Instinctivement un événement A est indépendant de B , si la réalisation de A ne dépend pas de celle de B . EN CONSEQUENCE : une restitution sous forme de formule serait donc la suivante :

$$p_B(A) = p(A)$$

D'un autre côté, cette formule a l'air non symétrique entre A et B (ce qui est contraire à l'instinct. Heureusement, ce n'est qu'une impression.

Propriété 24

On se donne $A, B \in \mathcal{T}$ tels que $p(A), p(B) \neq 0$. Alors on a

$$p_B(A) = p(A) \Leftrightarrow p(A \cap B) = p(A)p(B) \Leftrightarrow p_A(B) = p(B)$$

Démonstration :

Soient $A, B \in \mathcal{T}$ tels que $p(A), p(B) \neq 0$. Alors on a

$$p_B(A) = p(A) \Leftrightarrow p(A \cap B)/p(B) = p(A) \Leftrightarrow p(A \cap B) = p(A)p(B)$$

De même, on pourra écrire

$$\square \quad p(A \cap B) = p(A)p(B) \Leftrightarrow p(B \cap A)/p(A) = p(B) \Leftrightarrow p_A(B) = p(B)$$

Définition

Étant donné $A, B \in \mathcal{T}$, on dit que A et B sont indépendants si $p(A \cap B) = p(A)p(B)$.

Remarque :
Toutes les propriétés vue en première année sur l'indépendance restent vraies. (Les relire et les redémontrer.)

Remarque :
La formule d'indépendance est un équivalent de la question instinctive :

"les deux événements s'influencent-ils l'un l'autre ?"

Ainsi, il n'est pas forcément utile de se jeter immédiatement dans un calcul pour justifier ceci ! Un peu de bon sens suffit souvent à affirmer les choses ...

Par exemple : prenons A : "probabilité d'obtenir Face sur la pièce équilibrée que je lance." et B : "probabilité qu'il fasse beau ce matin" n'ont clairement rien à voir l'un avec l'autre. On peut affirmer sans crainte qu'il y a indépendance...

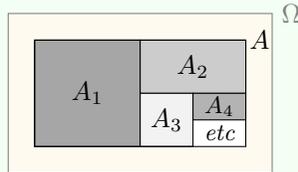
FAUX SEMBLANTS
Attention toutefois aux affirmations trop hâtives ! Les apparences sont parfois trompeuses.

IV systèmes (quasi)-complets

IV-1 Définition

Définition
Étant donné (Ω, \mathcal{T}, p) un espace probabilisé, $A \in \mathcal{T}$ un événement et $I \subset \mathbb{N}$. On dit que la famille $(A_k)_{k \in I}$ est un *système complet d'événements* (ou une partition) de A si :

- $A_k \in \mathcal{T}$ pour tout $k \in I$;
- les événements A_k sont 2 à 2 incompatibles;
- $\bigsqcup_{k \in I} A_k = A$



Commentaires :

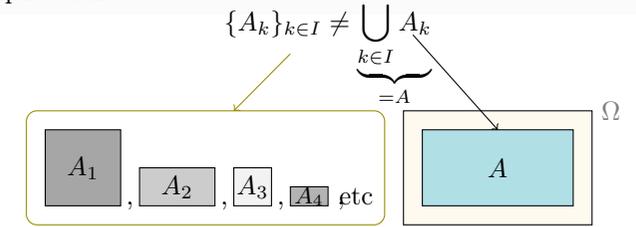
Pour un ensemble A , la famille $(A_k)_{k \in I}$ dans \mathcal{T} est donc un système complet d'événements de A si et seulement si :

- deux A_k distincts ne peuvent se produire en même temps
- l'ensemble des A_k forme toutes les possibilités pour obtenir A .

■ Exemples :

- 10 ■ Si on cherche le rang d'apparition du premier pile lors d'une série de lancers de pile ou face, on pose
 A_n : "Obtenir Pile au rang n " $\forall n \in \mathbb{N}^*$ et A_0 : "Ne jamais obtenir Pile"
 Alors $(A_n)_{n \in \mathbb{N}}$ est un système complet d'événements de l'univers.
- 11 ■ Si $\Omega = \mathbb{N}$ et A l'ensemble des nombres pairs. Alors, en posant $A_n = \{2n\}$ pour tout entier n , on trouve que $(A_n)_{n \in \mathbb{N}}$ est un système complet d'événements de A .

NOTATIONS
Faire attention aux mélanges de notations. Le système complet est une famille d'événements, pas une réunion ! Explication :

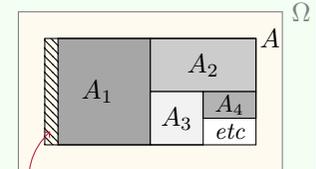


Remarque :
On peut bien entendu avoir $A = \Omega$. On parle alors d'un système complet (ou partition) (de l'univers).

Commentaires :
Dans la suite (cas des variables discrètes ou à densité), nous aurons besoin d'une notion plus souple, qui tient compte des événements négligeables, d'où la définition ci-dessous :

Définition
Étant donné (Ω, \mathcal{T}, p) un espace probabilisé, on dit que la famille $(A_k)_{k \in \mathbb{N}}$ est un *système quasi-complet d'événements* de $A \in \mathcal{T}$ si :

- $A_k \in \mathcal{T}$, $A_k \subset A$ pour tout $k \in \mathbb{N}$;
- les événements A_k sont 2 à 2 incompatibles;
- $p\left(\bigsqcup_{k \in \mathbb{N}} A_k\right) = \sum_{k=0}^{+\infty} p(A_k) = p(A)$.



négligeable

Convention d'écriture pour la suite du cours :

Dans la définition et dans la suite du cours, pour des facilités d'écriture, les indices de la réunion/somme dans les différents théorèmes et résultats sont pris dans \mathbb{N} tout entier, mais rien n'empêche un changement d'indice, (par exemple une indexation sur \mathbb{N}^* comme ci-dessous.)

■ Exemple 12 :

On imagine à nouveau qu'on lance une pièce de monnaie indéfiniment avec

$$\forall n \geq 1, \quad A_n = \text{« Obtenir Pile au } n^{\text{ième}} \text{ lancer »}$$

Montrons que $\{A_n\}_{n \in \mathbb{N}^*}$ est **quasi complet**. Les événements étant 2 à 2 incompatibles, il reste à montrer que $p\left(\bigcup_{n=1}^{+\infty} A_n\right) = 1$:

$$p\left(\bigcup_{n=1}^{+\infty} A_n\right) = 1$$

On sait par σ -additivité que

$$p\left(\bigcup_{n=1}^{+\infty} A_n\right) = \sum_{n=1}^{+\infty} p(A_n)$$

Notons maintenant F_k : « obtenir Face au $k^{\text{ième}}$ lancer ». Alors pour $n \geq 1$:

$$A_n = F_1 \cap \dots \cap F_{n-1} \cap \overline{F_n}$$

Les lancers étant infinis quoi qu'il arrive, les résultats à chaque lancer sont indépendants.

Ainsi,

$$p(A_n) = p(F_1) \dots p(F_{n-1}) p(\overline{F_n})$$

Or, par équiprobabilité, au $i^{\text{ème}}$ lancer, on a une probabilité de $\frac{1}{2}$ d'obtenir Pile ou Face.

Ainsi,

$$p(A_n) = \underbrace{\frac{1}{2} \dots \frac{1}{2}}_{n-1 \text{ fois}} \cdot \frac{1}{2} = \left(\frac{1}{2}\right)^n$$

Maintenant :

$$p\left(\bigcup_{n=1}^{+\infty} A_n\right) = \sum_{n=1}^{+\infty} \left(\frac{1}{2}\right)^n = \frac{1}{1 - \frac{1}{2}} - 1 = 1$$

Le système est donc bien quasi-complet...!



Remarque :

Dans l'exemple précédent, avec la notation A_0 : "Ne jamais obtenir Pile", comme $\{A_n\}_{n \in \mathbb{N}}$ est complet on peut en déduire que A_0 est négligeable :

$$p(A_0) = 1 - p\left(\bigcup_{n=1}^{+\infty} A_n\right) = 1 - 1 = 0$$

? Exercice 8

Réfléchir à ce qui change dans la démonstration de cet exemple si on suppose que l'on s'arrête dès l'obtention du premier Pile.

Solution

Les événements ne sont maintenant plus indépendants. En effet, si par exemple pile au premier lancer ($\overline{F_1}$), alors on ne peut pas obtenir face (F_2) au deuxième, puisque le jeu s'est arrêté... Il faut donc utiliser la formule des probabilités composées :

$$p(A_n) = p(F_1)p(F_1)_{F_1} \dots p(F_1 \cap \dots \cap F_{n-2})(F_{n-1})p(\overline{F_n})$$

Mais fort heureusement, les probabilités sont les mêmes!!



Remarque :

Si dans la définition précédente, on prend $A_k = \emptyset$ pour tout $k > N$, on a en fait un

système quasi-complet fini : $(A_k)_{0 \leq k < N}$ où la somme $\sum_{k=0}^{+\infty} p(A_k)$ devient une somme

finie $\sum_{k=0}^N p(A_k)$.

■ Exemple 13 :

On choisit un nombre au hasard entre 0 et 1 (0 et 1 inclu.).

On associe à cette expérience l'ensemble fondamental $\Omega = [0; 1]$ et la tribu \mathcal{B} des boréliens.

Comparons les systèmes suivants :

$$A_1 = [0; 1/2[\text{ et } A_2 = [1/2; 1] \quad (1)$$

et

$$B_1 =]0; 1/2[\text{ et } B_2 = [1/2; 1] \quad (2).$$

Le système (1) est clairement un système complet pour $\Omega = [0; 1]$, car

$$A_1 \cap A_2 = \emptyset, \quad A_1 \cup A_2 = \Omega$$

Mais même si $B_1 \cap B_2 = \emptyset$,

$$B_1 \cup B_2 =]0; 1] \neq \Omega$$

Donc le système (2) n'est **pas complet**. Néanmoins, la probabilité de tomber **exactement** sur 0 est nulle. (Pour l'instant, il faut se contenter de l'intuition, nous prouverons ceci un peu plus tard.) Ainsi, on a en fait

$$p(B_1 \cup B_2) = 1 - p(\overline{B_1 \cup B_2}) = 1 - p(\{0\}) = 1$$

Le système (2) est donc quasi-complet.

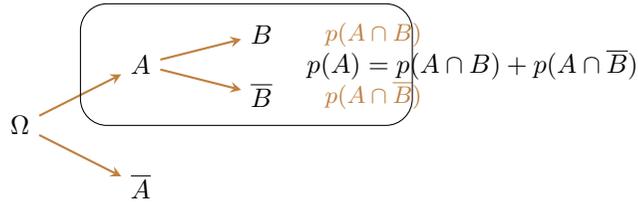
Commentaires :

En observant les définitions, on voit qu'un système complet est un cas particulier de système quasi-complet, toutes les propriétés suivantes sont donc valables dans les deux cadres...

IV-2 Formule(s) des probabilités totales

Commentaires :

Si la formule des probabilités composées consiste à suivre **une branche** d'un arbre de probabilités, la formule des probabilités totales ci-dessous consistera plutôt à analyser la décomposition suivant **ses extrémités**.



Voyons par étape comment ceci se généralise à des systèmes infinis :

Propriété 25 (Découpage par système quasi-complet)

On se donne un espace probabilisé (Ω, \mathcal{T}, p) . Soit B un événement quelconque. Si $(A_n)_{n \in \mathbb{N}}$ est un **système (quasi)-complet d'événements** de Ω (ou de A tel que $B \subset A$), alors

$$p(B) = p\left(B \cap \bigsqcup_{n \in \mathbb{N}} A_n\right) = \sum_{n=0}^{+\infty} p(B \cap A_n).$$

Démonstration :

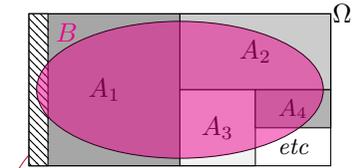
Posons C le complémentaire dans Ω de $\bigsqcup_{n=0}^{+\infty} A_n$. Alors $\Omega = C \sqcup \bigsqcup_{n=0}^{+\infty} A_n$ puis

$$\begin{aligned} p(B) &= p(B \cap \Omega) = p\left(B \cap C \sqcup \bigsqcup_{n=0}^{+\infty} A_n\right) \\ &= p\left((B \cap C) \sqcup (B \cap \bigsqcup_{n=0}^{+\infty} A_n)\right) = p(B \cap C) + p\left(B \cap \bigsqcup_{n=0}^{+\infty} A_n\right) \end{aligned}$$

Or, $B \cap C \subset C$, d'où $0 \leq p(B \cap C) \leq p(C) = 0$ et donc $p(B \cap C) = 0$. On en déduit que

$$P(B) = P\left(B \cap \bigsqcup_{n=0}^{+\infty} A_n\right) = \sum_{n=0}^{+\infty} P(B \cap A_n)$$

□



négligeable

? Exercice 9

On lance une pièce de monnaie indéfiniment. Calculer la probabilité de l'événement B : « le deuxième Pile arrive juste après le premier. »

Solution

Nous avons démontré tout à l'heure qu'en posant $A_n =$ « Obtenir Pile au $n^{\text{ième}}$ lancer » pour tout $n \in \mathbb{N}^*$, le système $\{A_n\}_{n \in \mathbb{N}^*}$ est quasi complet. Ainsi, par découpage en système quasi-complet, on a

$$P(B) = \sum_{n=1}^{+\infty} P(B \cap A_n)$$

Or, en reprenant les notations utilisées dans la démonstration de " $\{A_n\}_{n \in \mathbb{N}^*}$ est quasi complet" à la page 23, on trouve que

$$B \cap A_n = F_1 \cap \dots \cap F_{n-1} \cap \bar{F}_n \cap \bar{F}_{n+1}$$

D'où la encore par indépendance,

$$p(B \cap A_n) = p(F_1) \dots p(F_{n-1}) p(\bar{F}_n) p(\bar{F}_{n+1}) = \left(\frac{1}{2}\right)^{n+1}$$

Et donc

$$p(B) = \sum_{n=1}^{+\infty} \left(\frac{1}{2}\right)^{n+1} = \left(\frac{1}{2}\right)^2 \sum_{n=1}^{+\infty} \left(\frac{1}{2}\right)^{n-1} = \left(\frac{1}{2}\right)^2 \frac{1}{1 - \frac{1}{2}} = \frac{1}{2}$$

(intuitif et logique. Après un pile arrive un autre pile avec « 1 chance sur 2 » !)

Théorème 26 (Formule des probabilités totales)

Étant donné (Ω, \mathcal{T}, p) un espace probabilisé, et (A_0, \dots, A_n, \dots) un système quasi-complet d'événements de (Ω, \mathcal{T}, p) . Alors, pour tout événement $B \in \mathcal{T}$, on a

$$p(B) = \sum_{k=0}^{+\infty} p_{A_k}(B) p(A_k) \quad (\text{avec } p_{A_k}(B) p(A_k) = 0 \text{ si } p(A_k) = 0)$$

Démonstration :

On sait déjà, d'après le découpage en système quasi-complet démontré précédemment, que

$$p(B) = \sum_{k=0}^{+\infty} p(B \cap A_k) \quad (*)$$

Or, pour tout $k = 0, \dots, n$, on a $p(B \cap A_k) = p_{A_k}(B)p(A_k)$. On réinjecte dans (*) et c'est terminé. \square

■ **Exemple 14 :**

Suivant une étude (non contractuelle), on sait que la probabilité p_k pour une famille d'avoir k enfants est telle que

$$p_0 = p_1 = a, \quad p_k = (1 - 2a)2^{-(k-1)} \quad \forall k \geq 2.$$

On part du principe que la probabilité d'avoir une fille ou un garçon est la même. Quelle est dans ce cas la probabilité d'avoir exactement 2 garçons sur l'ensemble des enfants eu par le couple ?

Notons G l'événement "avoir deux garçons" et E_n : "La famille a n enfants." Alors le système $\{E_n\}_{n \geq 2}$ est un système complet de l'événement G et on a alors

$$P(G) = \sum_{n=2}^{+\infty} P(G \cap E_n) = \sum_{n=2}^{+\infty} P_{E_n}(G)P(E_n)$$

Or $P_{E_n}(G) = P(\text{"Il y a deux garçons et } n - 2 \text{ filles"})$ L'événement ci-dessus est l'ensemble des permutations de la liste $GG \underbrace{F \dots F}_{n-2 \text{ fois}}$.

De plus, chacune de ces listes est de probabilité égale à $\frac{1}{2^n}$ par indépendance. Ainsi, comme on a $\binom{n}{2}$ permutations possibles,

$$P_{E_n}(G) = \binom{n}{2} \frac{1}{2^n} = \frac{n(n-1)}{2} \frac{1}{2^n}$$

D'où

$$P(G) = \sum_{n=2}^{+\infty} \frac{n(n-1)}{2} \frac{1}{2^n} (1-2a)2^{-(n-1)} = (1-2a) \sum_{n=2}^{+\infty} n(n-1) \frac{1}{4^n} = (1-2a) \frac{1}{4^2} \frac{2}{(1-\frac{1}{4})^3}$$

i.e.

$$P(G) = \frac{8}{27}(1-2a)$$

v Variables aléatoires

Rappel : Dans le cadre des variables aléatoires **finies**, en première année, avec un univers Ω fini, nous avons défini une variable aléatoire simplement comme une application

$$X : \Omega \rightarrow \mathbb{R}.$$

Nous avons alors régulièrement utilisé des notations de type « $X = i$ » pour désigner l'événements "la variable aléatoire X prend la valeur i " (pour $i \in \mathbb{R}$), ou « $X \in B$ »

pour désigner l'événement "la variable aléatoire X prend ses valeurs dans B " (pour B un ensemble de réels).

Rappelons que cela signifie la chose suivante :

$$(X = i) = \{w \in \Omega \mid X(w) = i\} \quad (\text{antécédents de } i \text{ par } X)$$

$$(X \in B) = \{w \in \Omega \mid X(w) \in B\} \quad (\text{antécédents de tous les éléments de } B \text{ par } X)$$

■ **Exemple 15 :**

On lance deux dés et on appelle S la somme obtenue. Pour l'univers

$$\Omega = \{(i, j) \mid i, j \in \llbracket 1; 6 \rrbracket\}$$

et

$$S : \Omega \rightarrow \mathbb{R}; (i, j) \mapsto i + j$$

l'événement « $S = 4$ » correspond à

$$(S = 4) = \{w \in \Omega \mid S(w) = 4\} = \{(1, 3); (2, 2); (3, 1)\}$$

De même « $S \leq 3$ » correspond à

$$(S \leq 3) = \{w \in \Omega \mid S(w) \leq 3\} = \{(1, 1); (1, 2); (2, 1)\}$$

et à y regarder de plus près, « $S \in]2, 3[$ » correspond à

$$(S \in]2, 3[) = \emptyset$$

V-1 Variables aléatoires quelconques

Commentaires :

Dans le cadre futur des variables construites sur un univers étant un intervalle I de \mathbb{R} , il faut adapter la théorie à la présence d'une tribu potentiellement différente de $\mathcal{P}(\Omega)$:

Définition

On appelle variable aléatoire réelle sur (Ω, \mathcal{T}, P) toute fonction $X : \Omega \rightarrow \mathbb{R}$ telle que :

1. (Ω, \mathcal{T}, P) soit un espace probabilisé
2. pour tout $a \in \mathbb{R}$, $(X \leq a) \in \mathcal{T}$.



Remarque :

La deuxième condition peut également se traduire ainsi :

pour tout $a \in \mathbb{R}$, $(X \leq a)$ est un événement.

Il se trouve que grâce aux propriétés des tribus, tous les ensembles de type « $X \in B$ » avec B raisonnable auront également un sens et seront également des événements bien définis :

Propriété 27

Si X est une variable aléatoire et I un intervalle ou une réunion finie d'intervalles, alors $(X \in I)$ est un événement.

Commentaires :

Le but n'étant pas de rentrer dans des excès de technicité, on ne fera pas d'exemple chiffré ici. On se contentera des faits suivants, qui nous assurent que grand nombres de fonctions sont des variables aléatoires. (Vous n'êtes pas censés savoir démontrer qu'une fonction est une variable aléatoire.)

■ Exemple 16 :

Pour Ω fini, toute fonction $X : \Omega \rightarrow \mathbb{R}$ est une variable aléatoire sur $(\Omega, \mathcal{P}(\Omega), P)$ quelquesoit la probabilité p .

Proposition 28

Si Ω est un intervalle ou une réunion finie d'intervalle de \mathbb{R} , toute fonction continue ou continue par morceaux est une variable aléatoire sur $(\Omega, \mathcal{B}(\Omega), P)$.

Proposition 29

Soient X, Y deux variables aléatoires réelles définies sur le même espace probabilisé (Ω, \mathcal{T}, P) , alors

- $X + Y$ est une variable aléatoire.
- XY est une va.
- Si Y ne s'annule pas, $\frac{X}{Y}$ est une variable aléatoire.

Proposition 30

Soit f une fonction continue sur $\mathcal{D}_f \subset \mathbb{R}$ et X une variable aléatoire sur (Ω, \mathcal{T}, P) telle que $X(\Omega) \subset \mathcal{D}_f$, alors $f \circ X$ est une variable aléatoire.

■ Exemple 17 :

Si X, Y, Z sont des va sur (Ω, \mathcal{T}, P) , alors $\ln \frac{|XY-Z|+1}{1+Z^2}$ est une va.

V-2 Loi d'une v.a. par fonction de répartition

V.2-a) Valeurs de la variable : notion de support

Commentaires :

L'intérêt d'une variable aléatoire sur (Ω, \mathcal{T}, P) est notamment de pouvoir "se débarrasser" de l'univers d'origine, sur lequel il était peut être difficile d'effectuer des calculs.

On constate qu'en pratique, il n'y a nul besoin de connaître Ω . En revanche, la probabilité P est bien celle de l'espace (Ω, \mathcal{T}, P) . En général, on ne sait pas la décrire intégralement, mais on souhaite pouvoir calculer les probabilités du type $P(X \in B)$ ou $P(X = b)$ ou encore $P(X \leq b)$.

■ Exemple 18 :

Effectuons une infinité de lancers d'une pièce équilibrée. L'univers que l'on peut associer à cette expérience est le suivant :

$$\Omega = \{(u_n)_{n \in \mathbb{N}^*} \mid u_n \in \{P, F\}\}$$

Un élément est donc une suite (infinie) de F et de P qui donne l'ensemble des faces obtenues chaque lancer, par exemple

$$(F, F, P, F, P, \dots)$$

La probabilité d'un tel événement n'est pas calculable à l'aide de l'indépendance ou des probabilités composées (car on a une infinité de valeurs...)...*(à méditer)*

En revanche, si on note maintenant

$$X : \Omega \rightarrow \mathbb{N}; w \mapsto \begin{cases} \text{le rang d'apparition du premier Pile s'il apparait au moins une fois} \\ 0 \text{ sinon} \end{cases}$$

X sera bien une variable aléatoire (admis). Si on ne s'intéresse maintenant qu'à la seule valeur de X , on peut facilement calculer $P(X = n)$ pour tout $n \in \mathbb{N}$. On l'a déjà fait dans l'exemple sur le système complet p.23 :

$$P(X = n) = \begin{cases} \frac{1}{2^n} & \forall n \in \mathbb{N}^* \\ 0 & \text{si } n = 0 \end{cases}$$

Commentaires :

Dans l'exemple ci-dessus, X a d'un côté un **univers de définition** constitué des suites (P, F, \dots) et de l'autre un ensemble de **valeurs** $0, 1, \dots$

De manière générale, pour toute fonction, l'ensemble de ses valeurs est défini par :

$$X(\Omega) = \{X(\omega) \mid \omega \in \Omega\}$$

Notons alors que, comme il s'agit de l'ensemble de toutes les valeurs possibles, on a

$$P(X \in X(\Omega)) = P(\Omega) = 1.$$

■ Exemples :

- 19 ■ Sur l'exemple précédent (lancer infini d'une pièce et X rang d'apparition du premier Pile), on a

$$X(\Omega) = \mathbb{N}$$

- 20 ■ Si X est le résultat du choix d'un nombre dans l'intervalle $[0, 1]$ alors trivialement

$$X(\Omega) = [0, 1]$$

- 21 ■ Si on regarde le nombre de boules rouges que l'on peut obtenir en sélectionnant de manière simultanée 10 boules dans une urne contenant 6 boules rouges et 6 boules blanches, on a

$$X(\Omega) = \llbracket 4, 6 \rrbracket$$

Commentaires :

Étant donné la possibilité d'avoir des événements négligeables, on aimerait se laisser un peu plus de souplesse. On aimerait par exemple, dans l'exemple de la première apparition de Pile, pouvoir se débarrasser du cas $X = 0$ qui est négligeable :



Définition

Si X est une v.a. définie sur l'espace probabilisé (Ω, \mathcal{T}, P) , on appelle *support de X* et on note $Supp(X)$ tout sous-ensemble de $X(\Omega)$ tel que $P(X \in Supp(X)) = 1$.

Commentaires :

Dit "simplement" cela correspond à l'ensemble des valeurs que peut prendre la variable X où on se préoccupe pas de ce qui est négligeable. On va donc voir qu'un support n'est pas forcément défini de manière unique (cf ci-dessous)

■ Exemples :

- 22 ■ Un support trivial (et **maximum** !) pour toute variable aléatoire est $X(\Omega)$:

En effet, on vient de revoir plus haut que

$$P(X \in X(\Omega)) = 1.$$

- 23 ■ Sur la série infinie de lancers d'une pièce, si X est le rang d'apparition du premier Pile, comme $P(X \in \mathbb{N}^*) = 1$, on peut poser

$$Supp(X) = \mathbb{N}^*$$

- 24 ■ Si X est le résultat du choix d'un nombre dans l'intervalle $[0, 1]$ alors (sous-réserve d'admettre pour l'instant que 0 est de probabilité nulle), on a par exemple

$$Supp(X) =]0, 1]$$

mais il y a d'autres possibilités ! En effet, si 0 est de probabilité nulle, il en va certainement de même pour 1, ainsi

$$Supp(X) =]0, 1[$$

Et on pourrait peut être en retirer encore d'autres (mais tout ceci arrivera un peu plus tard...)

- 25 ■ Si on regarde le nombre de boules rouges que l'on peut obtenir en sélectionnant de manière simultanée 10 boules dans une urne contenant 6 boules rouges et 6 boules blanches, aucun élément de $\llbracket 4, 6 \rrbracket$ n'est de probabilité nulle. Ainsi, on ne peut rien retirer et on ne peut trouver de support plus petit que

$$Supp(X) = \llbracket 4, 6 \rrbracket$$



NOTATION TOLÉRÉE

Souvent, dans les ouvrages (et certaines fois au concours), les notions $X(\Omega)$ et $Supp(X)$ sont confondues. C'est un abus de notation qui est toléré. Néanmoins, dans ce cours, la différence sera faite systématiquement afin de ne pas commettre d'erreur de réflexion.

V.2-b) Fonction de répartition d'une variable aléatoire

Commentaires :

On a vu qu'une variable aléatoire ne dépendait pas de l'univers sur lequel elle est construite mais seulement de son support. Pour déterminer sa loi, dans tous les cas, on peut utiliser la "fonction de répartition" ; La notion de "loi" d'une variable aléatoire commune à toutes les variables est celle de la fonction de répartition, qui, on va le voir, permet de décrire complètement l'ensemble des probabilités d'une variable quelle qu'elle soit.

Définition
 Étant donnée une variable aléatoire X , on appelle *fonction de répartition* de X la fonction

$$F_X : \mathbb{R} \rightarrow [0, 1]$$

$$t \mapsto P(X \leq t)$$

■ Exemple 26 :

On choisit simultanément 10 boules dans une urne contenant 6 boules rouges et 6 boules blanches. On note X le nombre de boules rouges obtenues et on note F la fonction de répartition de X . On cherche à déterminer $F(2), F(4), F(4.5), F(5), F(6)$

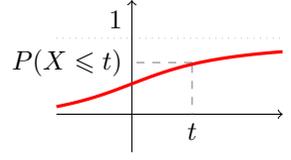
- ★ On a facilement que $F(2) = P(X \leq 2) = 0$ car l'événement est impossible.
 - ★ A contrario, $F(6) = P(X \leq 6) = 1$ car c'est l'événement certain.
 - ★ On a $F(4) = P(X \leq 4) = P(X = 4)$ car il n'y a pas d'autre possibilité.
 - ★ Comme les valeurs de X sont nécessairement entières, on a $F(4,5) = P(X \leq 4,5) = P(X \leq 4) = F(4)$.
 - ★ Pour la même raison, on a $F(5) = P(X \leq 5) = P(X = 4) + P(X = 5)$.
- Pour résumer après avoir fini les calculs (exercice) :

k	2	4	4.5	5	6
$P(X = k)$	0	$\frac{5}{22}$	$\frac{5}{22}$	$\frac{6}{11}$	$\frac{5}{22}$

Remarque :
 Dans la situation de l'exemple précédent, on a trouvé les valeurs de F seulement pour $t \in \{2, 4, 4.5, 5, 6\}$. Elle n'est donc pas entièrement déterminée. Il faudrait préciser l'ensemble de tous les $F(t)$ pour tous les t réels. On en est loin!

Commentaires :

La fonction de répartition est une fonction réelle tout à fait classique, qui n'est pas une variable aléatoire. Elle peut se dessiner à l'aide d'un graphique tout à fait habituel. Elle peut par exemple avoir cette allure là, (avec toujours des ordonnées clairement comprises entre 0 et 1) :



En abscisse : les valeurs de t pour tout t réel et en ordonnées, les probabilités associées. Nous allons commencer par voir quelques valeurs particulières (que nous aurons commencer à aborder dans l'exemple précédent.)

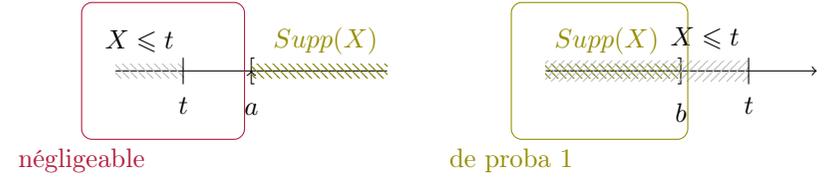
Propriété 31

Si $Supp(X) \subset [a, +\infty[$, alors $F_X(t) = 0 \quad \forall t < a$.
 De même, si $Supp(X) \subset]-\infty, b]$, alors $F_X(t) = 1 \quad \forall t \geq b$

Et en français ?? :

- ★ Si les valeurs de X sont toutes supérieures à a , alors, si $t < a$, l'événement $X \leq t$ est impossible.
- ★ Si les valeurs de X sont toutes inférieures à b , alors, si $t \geq b$, l'événement $X \leq b$ est certain.

Illustrations :



Démonstration :
 • Si $Supp(X) \subset [a, +\infty[$:
 Soit $t < a$. Alors $(X \leq t) \subset (X < a) \subset (X \in \overline{Supp(X)})$. D'où

$$0 \leq P(X \leq t) \leq P(X < a) \leq P(X \in \overline{Supp(X)}) = 1 - P(X \in Supp(X)) = 1 - 1 = 0$$
 et donc $P(X \leq t) = 0$.

• Si $Supp(X) \subset]-\infty, b]$:
 Soit $t \geq b$. Alors $(X \in Supp(X)) \subset (X \leq b) \subset (X \leq t)$. D'où

$$1 = P(X \in Supp(X)) \leq P(X \leq b) \leq P(X \leq t) \leq 1$$
 et donc $P(X \leq t) = 1$.
 □

De plus, il vient rapidement une propriété élémentaire :

Propriété 32

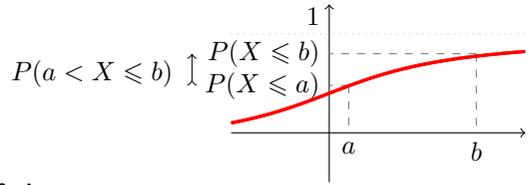
Soit F_X la fonction de répartition d'une variable aléatoire X . Alors F_X est une fonction croissante.

Démonstration :
 Soient $u, v \in \mathbb{R}$ tels que $u \leq v$. Comme $(X \leq u) \subset (X \leq v)$, on obtient immédiatement $F_X(u) = P(X \leq u) \leq P(X \leq v) = F_X(v)$. □

Continuons l'interprétation :

Propriété 33

Si $a < b$, $P(a < X \leq b)$ désigne la "différence de hauteur" entre $P(X \leq a)$ et $P(X \leq b)$.



Démonstration :

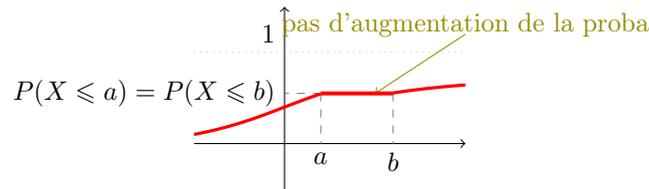
$$P(a < X \leq b) = P(X \leq b \setminus X \leq a) = P(X \leq b) - P(X \leq a) \quad \square$$

De la croissance de F_X et du résultat précédent, on tire également que :

Propriété 34

Soient $a, b \in \mathbb{R}$ tels que $a < b$. Alors

$$P(a < X \leq b) = 0 \text{ ssi } F_X \text{ est constante sur } [a, b].$$



Voyons maintenant les inégalités strictes :

Propriété 35

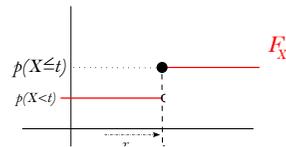
Soit F_X la fonction de répartition d'une v.a. X et $t \in \mathbb{R}$. Alors

$$P(X < t) = \lim_{\substack{x \rightarrow t \\ x < t}} P(X \leq x) = \lim_{\substack{x \rightarrow t \\ x < t}} F_X(x)$$

Démonstration : admise.

□

Explication sur un graphique :



Remarque :

Soit F_X la fonction de répartition d'une v.a. X . Alors,

$$P(X = t) = P(X \leq t) - P(X < t)$$

Ce qui veut dire que :

$$P(X = t) \neq 0 \text{ si et seulement si } t \text{ est un point de discontinuité de } F_X$$

et

$P(X = t)$ est la "hauteur" de la discontinuité de la fonction de répartition en t .

Commentaires :

Nous allons voir maintenant 2 exemples : un dans le cadre des probabilités finies, et l'autre dans le cadre d'un support $[0, 1]$

■ Exemple 27 :

Revenons à : "On choisit simultanément 10 boules dans une urne contenant 6 boules rouges et 6 boules blanches. On note X le nombre de boules rouges obtenues." On cherche à établir la fonction de répartition F_X de X .

• Tout d'abord, le support :

C'est une variable finie avec $Supp(X) = \{4, 5, 6\} \subset [4, 6]$. Ainsi,

$$F_X(x) = \begin{cases} 0 & \text{si } x < 4 \\ 1 & \text{si } x \geq 6 \end{cases}$$

• Les intervalles où F_X est constante :

On sait que $P(4 < X < 5) = P(5 < X < 6) = 0$, ainsi, F_X est constante entre 4, 5 puis 5, 6.

• Les valeurs restantes :

Nous avons déjà déterminé les valeurs en 4, 5, 6 dans un exemple précédent. On avait

k	4	5	6
$P(X = k)$	$\frac{5}{22}$	$\frac{6}{11}$	$\frac{5}{22}$

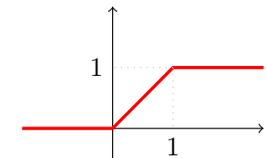
D'où la fonction $F_X(x) = \begin{cases} 0 & \text{si } x < 4 \\ \frac{5}{22} & \text{si } 4 \leq x < 5 \\ \frac{5}{22} + \frac{6}{11} & \text{si } 5 \leq x < 6 \\ 1 & \text{si } x \geq 6 \end{cases}$

■ Exemple 28 :

Loi uniforme sur $]0; 1]$

On choisit un nombre X au hasard dans l'intervalle $]0; 1]$. Montrons que la fonction de répartition de X est F_X définie comme suit :

$$F_X(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ x & \text{si } x \in]0; 1[\\ 1 & \text{si } x \geq 1 \end{cases}$$



• Tout d'abord, $Supp(X) =]0; 1]$, d'où :

$$F_X(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ 1 & \text{si } x \geq 1 \end{cases}$$

• Soit $x \in]0; 1[$. Montrons que $F_X(x) = x$:

De manière intuitive, on peut dire que la probabilité d'obtenir un nombre dans l'intervalle $]0; x]$ est x fois la probabilité d'obtenir un nombre dans $]0; 1]$. i.e.

$$P(0 < X \leq x) = x.P(0 < X \leq 1) = x$$

Nous allons le montrer en plusieurs étapes : d'abord $x = \frac{1}{n}$ pour $n \in \mathbb{N}^*$, puis $x = \frac{k}{n} \in]0, 1]$ pour $n, k \in \mathbb{N}^*$, et enfin x réel quelconque dans $]0, 1]$. Pour commencer, remarquons qu'on a

$$F_X(x) = P(X \leq x) = P(0 < X \leq x)$$

★ Pour $x = \frac{1}{n}$ avec $n \in \mathbb{N}^*$:

Par choix équiprobable, pour tout $n \in \mathbb{N}^*$, on a

$$P\left(0 < X \leq \frac{1}{n}\right) = P\left(\frac{1}{n} < X \leq \frac{2}{n}\right) = \dots = P\left(\frac{n-1}{n} < X \leq \frac{1}{n}\right)$$

Sommons ces valeurs. On obtient d'une part

$$\sum_{k=0}^{n-1} P\left(\frac{k}{n} < X \leq \frac{k+1}{n}\right) = P([0; 1]) = 1$$

et d'autre part

$$\sum_{k=0}^{n-1} P\left(\frac{k}{n} < X \leq \frac{k+1}{n}\right) = n.P\left(0 < X \leq \frac{1}{n}\right)$$

d'où

$$p\left(0 < X \leq \frac{1}{n}\right) = \frac{1}{n}$$

★ Pour $x = \frac{k}{n}$ avec $n, k \in \mathbb{N}^*$ où $k \leq n$:

Comme $\left\{\left(\frac{j}{n} < X \leq \frac{j+1}{n}\right)\right\}_{j=0, \dots, k-1}$ et un système complet de $(0 < X \leq \frac{k}{n})$, on a

$$P\left(0 < X \leq \frac{k}{n}\right) = \sum_{j=0}^{k-1} P\left(\frac{j}{n} < X \leq \frac{j+1}{n}\right) = \frac{k}{n}$$

★ Pour x réel quelconque dans $]0, 1]$:

On cherche à encadrer x entre deux nombres rationnels du type $\frac{k}{n}$ et $\frac{k+1}{n}$. Remarquons alors que pour tout $x \in]0; 1[$ et tout $n \in \mathbb{N}^*$, on a

$$\frac{k}{n} \leq x < \frac{k+1}{n} \Leftrightarrow k \leq nx < k+1 \Leftrightarrow k = [nx]$$

D'où, si $x \in]0, 1[$, en posant $k = [nx]$, par croissance de p , pour tout $n \in \mathbb{N}^*$, on a

$$\underbrace{P\left(0 < X \leq \frac{k}{n}\right)}_{=\frac{k}{n}} \leq P(0 < X \leq x) \leq P\left(0 < X < \frac{k+1}{n}\right) \leq \underbrace{P\left(0 < X \leq \frac{k+1}{n}\right)}_{=\frac{k+1}{n}}$$

D'où

$$\frac{[nx]}{n} \leq P(0 < X \leq x) \leq \frac{[nx] + 1}{n}$$

et, par passage à la limite quand n tend vers l'infini, et théorème des gendarmes $\lim_{n \rightarrow +\infty} \frac{[nx]}{n} = \lim_{n \rightarrow +\infty} \frac{[nx] + 1}{n} = x$ (exercice en considérant que $[nx] \in [nx - 1, nx]$)

$$P(0 < X \leq x) = x$$

ce qui achève la démonstration.

Au final, on constate sur une fonction de répartition que :

Propriété 36

Soit F_X la fonction de répartition d'une variable aléatoire X . Alors, les conditions suivantes sont vérifiées :

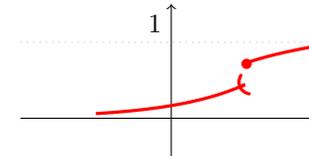
- i) $0 \leq F_X(t) \leq 1 \quad \forall t \in \mathbb{R}$.
- ii) F_X est une fonction croissante.
- iii) F_X est une fonction continue à droite en tout point $t \in \mathbb{R}$.
- iv) $\lim_{t \rightarrow -\infty} F_X(t) = 0$ et $\lim_{t \rightarrow +\infty} F_X(t) = 1$

Démonstration :

- i) Soit $t \in \mathbb{R}$. $F_X(t) = P(X \leq t) \in [0; 1]$. ii) déjà vu
le reste est hors programme. \square

Commentaires :

La condition iii) prendra tout son sens dans le futur chapitre sur les variables aléatoires discrètes. Elle consiste à dire qu'en cas de discontinuité, les points inclus se présentent toujours de la manière suivante :



V.2-c) Fonction dite "de répartition"

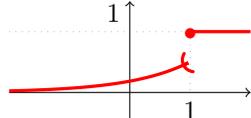
Définition

De manière générale, toute fonction $F : \mathbb{R} \rightarrow \mathbb{R}$ vérifiant les conditions

- i) $0 \leq F(t) \leq 1 \quad \forall t \in \mathbb{R}$.
 - ii) F est une fonction croissante.
 - iii) F est une fonction continue à droite en tout point $t \in \mathbb{R}$.
 - iv) $\lim_{t \rightarrow -\infty} F(t) = 0$ et $\lim_{t \rightarrow +\infty} F(t) = 1$
- est appelée *fonction de répartition*.

Commentaires :

Cette fonction peut être définie de manière indépendante d'une quelconque variable aléatoire. Si on définit par exemple la fonction suivante :

$$F(x) = \begin{cases} \frac{e^x}{2e} & \text{si } x < 1 \\ 1 & \text{si } x \geq 1 \end{cases}$$


Alors F vérifie bien les conditions ci-dessus. C'est une fonction de répartition, et elle n'a pas été construite à partir d'une expérience. Néanmoins, les variables aléatoires ne sont jamais bien loin :

Propriété 37

Pour toute fonction de répartition F , il existe une variable aléatoire X de fonction de répartition F_X telle que

$$F = F_X$$

Autrement dit, toute à toute fonction de répartition, on peut associer une variable aléatoire.

Démonstration : admise. \square

Commentaires :

Avec un peu plus d'habitude, nous verrons que nous pouvons déduire des fonctions de répartition toutes les informations utiles : support, type de variable, différentes probabilités, etc... Nous verrons ceci plus en détail dans les chapitres spécifiques.

V-3 Indépendance de variables aléatoires

V.3-a) Cas de deux variables

Définition

Soient X et Y deux variables aléatoires définies sur un même espace probabilisé (Ω, \mathcal{T}, P) . On dit que X et Y sont *indépendantes* si, pour tout $A, B \in \mathcal{T}$, on a

$$P(X \in A \cap Y \in B) = P(X \in A) P(Y \in B)$$

Remarque :

Tout comme dans le cas des événements, la définition est une traduction mathématique de l'idée instinctive :

Le résultat d'une des variables n'influence pas l'autre.

Il est donc inutile de se jeter dans des calculs (qui peuvent être fastidieux!) quand la situation est claire.

Proposition 38

Deux variables aléatoires X, Y définies sur un même espace probabilisé sont indépendantes si et seulement si, pour tout $x, y \in \mathbb{R}$,

$$P(X \leq x \cap Y \leq y) = F(x)F(y)$$

Démonstration : admise. \square

? Exercice 10

On choisit deux nombres au hasard de manière indépendante dans l'intervalle $]0; 1]$. Quelle est la probabilité d'obtenir deux nombres inférieurs ou égaux à $\frac{1}{2}$?

Solution

On pose X et Y les variables donnant les deux résultats. Ce sont deux variables aléatoires indépendantes suivant des lois uniformes sur $]0; 1]$. Alors

$$P(X \leq \frac{1}{2}, Y \leq \frac{1}{2}) = F_X\left(\frac{1}{2}\right) F_Y\left(\frac{1}{2}\right) = \left(\frac{1}{2}\right)^2 = \frac{1}{4}$$

Définition
 Soient $(X_k)_{k \in K}$ une famille (finie ou infinie) de variables aléatoires définies sur un même espace probabilisé (Ω, \mathcal{T}, P) . On dit que les variables X_k sont *mutuellement indépendantes* si, pour toute famille $(A_k)_{k \in \mathbb{N}}$ d'événements, on a

$$P\left(\bigcap_{k \in K} (X_k \in A_k)\right) = \prod_{k \in K} P(X_k \in A_k)$$

et voici quelques propriétés de bon sens :

Propriété 39

Si $(X_k)_{k \in K}$ une famille de variables aléatoires mutuellement indépendantes, alors toute sous-famille de variables est mutuellement indépendante.

Démonstration : exercice. \square

Lemme 40 des coalitions

Si $(X_1, \dots, X_n, X_{n+1}, \dots, X_{n+p})$ est une famille de variables mutuellement indépendantes, alors, pour tout fonction, $u : \mathbb{R}^n \rightarrow \mathbb{R}$, $v : \mathbb{R}^p \rightarrow \mathbb{R}$, les variables $u(X_1, \dots, X_n)$ et $v(X_{n+1}, \dots, X_{n+p})$ sont indépendantes.

Commentaires :

Le lemme ci-dessus signifie simplement qu'on peut associer les variables indépendantes de toute manière souhaitée. Pour peu qu'on les divise au moins en 2 paquets disjoints, ça reste bien indépendant.

I Généralités sur les polynômes

Dans tout ce chapitre, \mathbb{K} désignera \mathbb{R} ou \mathbb{C} et n un entier naturel positif ($n \in \mathbb{N}$). Ce chapitre a pour but de généraliser les notions sur les polynômes de $\mathbb{R}[X]$ vues en première année aux polynômes à coefficients complexes.

I-1 Définitions et vocabulaire de base

Définition
On note X la **fonction identité** de \mathbb{K} dans \mathbb{K} , c'est-à-dire :

$$\begin{aligned} X : \mathbb{K} &\rightarrow \mathbb{K} \\ x &\mapsto x \end{aligned}$$

CONFUSION DE NOTATION
Ne pas confondre X qui est une application et le réel x ci-dessus. Ainsi, si par exemple on écrit $P = X$, P est bien une application qui peut s'évaluer en n'importe quel nombre réel. Par exemple ici, $P(2) = 2$.

Définition
Soit a un élément de \mathbb{K} . On appelle *monôme de degré n et de coefficient a* l'application

$$\begin{aligned} aX^n : \mathbb{K} &\rightarrow \mathbb{K} \\ x &\mapsto ax^n \end{aligned}$$

De manière générale :

Définition
Soit P une fonction de \mathbb{K} dans \mathbb{K} . On dit que P est un polynôme à coefficients dans \mathbb{K} s'il existe un entier naturel m ainsi que $m + 1$ éléments de $\mathbb{K} : a_0, a_1, a_2, \dots, a_m$, tels que :

$$P = a_0 + a_1X + a_2X^2 + \dots + a_mX^m$$

c'est-à-dire :

$$\begin{aligned} P : \mathbb{K} &\rightarrow \mathbb{K} \\ x &\mapsto a_0 + a_1x + a_2x^2 + \dots + a_mx^m \end{aligned}$$

■ **Exemple 1** :
Les termes $-\pi X^2, (1 + 2i)X$ et $\sqrt{2}$ sont les trois monômes constituant le polynôme $P = -\pi X^2 + (1 + 2i)X + \sqrt{2}$.

Notation :
Il arrive fréquemment d'observer la notation $P(X)$ au lieu de P . (par exemple $P(X) = \frac{1}{2}X^2 - i\sqrt{3}X + 1$)
Celle-ci est à comprendre comme une composition d'applications, c'est-à-dire :

$$P(X) = P \circ X = P \circ id = P$$

Moralité, écrire P ou écrire $P(X)$ est en réalité la même chose. (contrairement à $P(2)$ et P qui sont deux objets différents ! L'un est un nombre et l'autre une fonction...)

CONFUSION
Considérant la notation précédente, si $x \in \mathbb{K}$, remarquons alors que $P(X)$ et $P(x)$ sont deux objets de **types différents** ! $P(x) \in \mathbb{K}$ est un nombre. C'est l'image de $x \in K$ par l'application $P(X)$.

■ **Exemple 2** :
Si $P(X) = X^2 - i\sqrt{3}X + 1$ est un polynôme, alors pour tout $x \in \mathbb{C}$, $P(x) = x^2 - i\sqrt{3}x + 1$ est un nombre complexe. Par exemple, pour $x = i$, on a $P(i) = i^2 - i^2\sqrt{3} + 1 = \sqrt{3} \in \mathbb{C}$

Définition
On note $\mathbb{K}[X]$ l'ensemble des polynômes à coefficients dans \mathbb{K} .

Définition
Les éléments de $\mathbb{R}[X]$ sont appelés *polynômes à coefficients réels*.
Les éléments de $\mathbb{C}[X]$ sont appelés *polynômes à coefficients complexes*.

Remarque :
Dans l'écriture $P = a_0 + a_1X + \dots + a_nX^n$, on ne demande pas nécessairement que $a_n \neq 0$. Ainsi, quitte à rajouter (ou suivant les cas, retirer) des a_i nuls (par exemple $a_{n+1}, \dots, a_m = 0$), on peut également écrire que

$$P = a_0 + a_1X + \dots + a_nX^n + a_{n+1}X^{n+1} + \dots + a_mX^m \quad \text{avec } n \neq m$$

Cette écriture sera très pratique pour la suite des définitions et propriétés.

■ **Exemple 3** :
 $1 + 2X + X^3 = 1 + 2X + X^3 + 0 \cdot X^4 + 0 \cdot X^5 \quad (n = 3, m = 5)$

I-2 Polynôme nul et unicité de l'écriture d'un polynôme

Notation :

On note 0 le polynôme nul. C'est le polynôme suivant :

$$\begin{aligned} 0 : \mathbb{K} &\rightarrow \mathbb{K} \\ x &\mapsto 0 \end{aligned}$$

CONFUSION

C'est bien une application dont l'image est nulle pour **tout** $x \in \mathbb{R}$, à ne pas confondre avec une application qui s'annule en l'un ou l'autre point. Par exemple,

$$\begin{aligned} P = 1 + X : \mathbb{K} &\rightarrow \mathbb{K} \\ x &\mapsto 1 + x \end{aligned}$$

s'annule en $x = -1$, mais ce n'est en aucun cas le polynôme nul, puisque si on prend par exemple $x = 1$, on obtient $P(1) = 1 + 1 \neq 0$. En réalité, il n'y a qu'une seule possibilité pour qu'un polynôme soit nul :

Proposition 41

Soient $a_0, a_1, \dots, a_n \in \mathbb{K}$. On pose $P = a_0 + a_1X + \dots + a_nX^n$. On a :

$$P = 0 \iff a_0 = a_1 = \dots = a_n = 0.$$

Démonstration :

On peut démontrer ceci par exemple par récurrence sur n , on se référera à la démonstration de première année sur les polynômes dans $\mathbb{R}[X]$ (et au cas échéant, on admettra qu'elle se généralise à $\mathbb{C}[X]$.) \square

Commentaires :

De cette proposition découle le corollaire suivant qui permet donc l'identification des coefficients de deux polynômes égaux :

Corollaire *Unicité des coefficients*

Soient m un entier naturel, ainsi que $a_0, a_1, \dots, a_n \in \mathbb{K}$ et $b_0, b_1, \dots, b_m \in \mathbb{K}$. On pose $P = a_0 + a_1X + \dots + a_nX^n$ et $Q = b_0 + b_1X + \dots + b_mX^m$. Quitte à rajouter des $a_i = 0$ ou $b_j = 0$, on peut supposer que $n = m$. Alors

$$P = Q \iff \begin{cases} a_0 = b_0 \\ \vdots \\ a_n = b_n \end{cases}$$

Démonstration :

La démonstration vue en première année se généralise aisément ici au cas de $\mathbb{C}[X]$. Elle repose sur la proposition précédente. \square

I-3 Degré d'un polynôme

Définition

Soit P un polynôme non nul défini par $P = a_0 + a_1X + \dots + a_nX^n$ où $a_0, \dots, a_n \in \mathbb{K}$.

- On appelle *degré de P* le plus grand entier N tel que $a_N \neq 0$. On le note $\deg(P)$.
- On appelle *terme dominant de P* (ou monôme dominant) le monôme a_NX^N .
- On appelle *coefficient dominant de P* le nombre a_N précédemment défini.

Exemple 4 :

| Si $P = 0 \times X^3 + \pi X^2 + 0 \times X + i$, on a $\deg P = 2$ avec P de coeff. dominant π .

Convention

La définition précédente ne fonctionne pas si P est le polynôme nul. Pour des futurs raisons de calcul, on pose par convention

$$\deg(0) = -\infty.$$

Notation :

On note $\mathbb{K}_n[X]$ l'ensemble des polynômes à coefficients dans \mathbb{K} dont le degré est inférieur ou égal à n .

CONFUSION

Contrairement à une erreur fréquemment commise, on ne demande pas à ce que ce le degré d'un polynôme de $\mathbb{K}_n[X]$ soient "égal" à n , mais bien "inférieur ou égal". Par exemple :

$$X \in \mathbb{C}_2[X].$$

Exemple 5 :

| $X^2 - iX + e^{i\frac{\pi}{4}}$ appartient à $\mathbb{C}_2[X]$ mais $X^3 + iX$ n'appartient pas à $\mathbb{C}_2[X]$.

II Opérations sur les polynômes

II-1 Définitions et propriétés élémentaires

Les polynômes sont des applications. Toutes les opérations dans $\mathbb{R}[X]$ découlent donc de celles des fonctions réelles. Pour des raisons de cohérence de définition, on généralise donc ceci aux polynômes de $\mathbb{C}[X]$.

Définition
Soient P et Q deux polynômes à coefficients dans \mathbb{K} et λ un élément de \mathbb{K} .

- On définit la fonction $P + Q$ sur \mathbb{K} par :
$$\forall x \in \mathbb{K}, (P + Q)(x) = P(x) + Q(x).$$
- On définit la fonction λP sur \mathbb{K} par :
$$\forall x \in \mathbb{K}, (\lambda P)(x) = \lambda P(x)$$
- On définit la fonction $P \times Q$ sur \mathbb{K} par :
$$\forall x \in \mathbb{K}, (P \times Q)(x) = P(x) \times Q(x)$$
- On définit la fonction $P \circ Q$ sur \mathbb{K} par :
$$\forall x \in \mathbb{K}, (P \circ Q)(x) = P(Q(x)).$$

Exercice 1
Posons $P = 4X^2 - 5X + \pi$ et $Q = 3X + iX^3$, montrer que $P \circ Q = -4X^6 + 24iX^4 - 5iX^3 + 36X^2 - 15X + \pi$.

Solution

$$\begin{aligned} P \circ Q &= 4(3X + iX^3)^2 - 5(3X + iX^3) + \pi \\ &= 4(9X^2 - X6 + 6iX^4) - 15X - 5iX^3 + \pi \\ &= -4X^6 + 24iX^4 - 5iX^3 + 36X^2 - 15X + \pi \end{aligned}$$

Cette définition donne lieu aux propriétés suivantes :

Propriété 42

Soient P, Q et R trois polynômes à coefficients dans $\mathbb{K}[X]$ et $\lambda \in \mathbb{K}$. On a :

1. $P + Q = Q + P$	7. $P \times (Q \times R) = (P \times Q) \times R$
2. $P + (Q + R) = (P + Q) + R$	8. $P \times (Q + R) = P \times Q + P \times R$
3. $P + 0 = P$	9. $P \times 1 = P$
4. $P + (-P) = 0$	10. $(P + Q) \circ R = P \circ R + Q \circ R$
5. $P \times Q = Q \times P$	11. $(\lambda P) \circ Q = \lambda(P \circ Q)$
6. $\lambda(Q \times R) = (\lambda Q) \times R = Q \times (\lambda R)$	12. $(P \times Q) \circ R = (P \circ R) \times (Q \circ R)$
	13. $P \circ (Q \circ R) = (P \circ Q) \circ R$

Remarque :
On peut calculer $P \circ Q$ et $Q \circ P$ avec $P(X) = 1 + X + X^2$ et $Q(X) = X^2$. On constate qu'en général

$$P \circ Q \neq Q \circ P$$

La commutativité du produit ($PQ = QP$) quant à elle entraîne :

Proposition 43 Formule du binôme de Newton

Soient P et Q deux polynômes à coefficients dans \mathbb{K} , on a :

$$(P + Q)^n = \sum_{k=0}^n \binom{n}{k} P^k Q^{n-k}$$

NOTION DE DÉRIVÉE
La notion de **dérivée ne se généralise pas** dans votre programme aux polynômes quelconques dans $\mathbb{C}[X]$. Pour revoir la question de la dérivée dans $\mathbb{R}[X]$, on se référera donc au cours de première année.

II-2 Résultat des opérations sur les coefficients

De même que tout à l'heure, les résultats des opérations sur les polynômes s'étendent à $\mathbb{C}[X]$. On va les expliciter ci-dessous. Pour les besoins des énoncés, on rappelle que si $P = a_0 + a_1X + \dots + a_NX^N$ est un polynôme de degré N , alors, pour tout $n \in \mathbb{N}$ tel que $n > N$, on a également

$$P = a_0 + a_1X + \dots + a_NX^N + a_{N+1}X^{N+1} + \dots + a_nX^n$$

avec $a_{N+1}, \dots, a_n = 0$.

Proposition 44 (opérations élémentaires)

Soient P et Q deux polynômes à coefficients dans \mathbb{K} et λ un élément de \mathbb{K} . On note :

$$P = a_0 + a_1X + \dots + a_nX^n \quad \text{et} \quad Q = b_0 + b_1X + \dots + b_nX^n$$

(n n'étant pas forcément le degré de P et Q !)

- $P + Q$ est un polynôme et on a :
$$P + Q = (a_0 + b_0) + (a_1 + b_1)X + \dots + (a_n + b_n)X^n.$$
- λP est un polynôme et on a : $\lambda P = \lambda a_0 + \lambda a_1X + \dots + \lambda a_nX^n$.
- $P \times Q$ est un polynôme et on a (en posant $a_i = 0$ et $b_i = 0$ si i n'est pas dans $\llbracket 0, n \rrbracket$)
$$P \times Q = \sum_{k=0}^{2n} c_k X^k \quad \text{avec, pour tout } m \in \llbracket 0, 2n \rrbracket, c_m = \sum_{k=0}^m a_k b_{m-k}$$



Remarque :

En ce qui concerne la multiplication, la formule avec les "c_k", n'est en général utilisée que dans un cadre théorique. Plus concrètement, si on veut faire le produit de deux polynôme, on le fait généralement en respectant les règles de développement classique des parenthèses!

■ Exemple 6 :

Si $P = 1 + iX$ et $Q = X^2 + 3$, alors

$$P \times Q = (1 + iX)(X^2 + 3) = (X^2 + 3) + iX(X^2 + 3) = 3 + 3iX + X^2 + iX^3$$



Remarque :

La formule énoncée dans la proposition se rapproche un peu plus de la technique du produit "à vue", où on cherche à déterminer au vol les coefficients devant les X^k .

■ Exemple 7 :

Si $P = 1 + iX$ et $Q = X^2 + 3$, alors on sait qu'en faisant le produit, on obtiendra une constante et X, X^2, X^3 .

$$P \times Q = ? + ?X + ?X^2 + ?X^3$$

On observe ici que par développement :

- la constante s'obtient en faisant 1×3 .
- le coefficient devant X s'obtient seulement en faisant $iX \times 3$
- le coefficient devant X^2 s'obtient seulement en faisant $1 \times X^2$
- le coefficient devant X^3 s'obtient seulement en faisant $iX \times X^2$

D'où

$$P \times Q = 3 + 3iX + X^2 + iX^3$$

II-3 Résultat des opérations sur les degrés

Propriété 45

Soient P et Q deux polynômes à coefficients dans \mathbb{K} et $\lambda \in \mathbb{K}$.

- On a : $\deg(P + Q) \leq \max(\deg(P), \deg(Q))$.
- Si on sait que $\deg(P) \neq \deg(Q)$ alors $\deg(P + Q) = \max(\deg(P), \deg(Q))$.
- Si $\lambda \neq 0$ alors : $\deg(\lambda P) = \deg(P)$.
- On a : $\deg(P \times Q) = \deg(P) + \deg(Q)$.

Démonstration :

La démonstration de première année se généralise sans problème au cas $\mathbb{C}[X]$. \square

■ Exemples :

Le cas de $P + Q$ dépend du fait que les coefficients dominants se "compensent" ou non :

8 ■ Si $P = 1 + iX - \sqrt{2}X^3$ et $Q = \pi - X^3$, alors $P + Q = (1 + \pi) + iX + (\sqrt{2} - 1)X^3$.
Ici, $\deg(P + Q) = \deg P = \deg Q$

9 ■ Si $P = 1 + iX + X^3$ et $Q = \pi - X^3$, alors $P + Q = (1 + \pi) + iX$.
Ici, $\deg(P + Q) < \deg P, \deg Q$



Remarque :

La propriété $\deg(P \times Q) = \deg(P) + \deg(Q)$ est bien valable également si P ou Q (ou les deux) est nul. Elle respecte bien la convention " $\deg 0 = -\infty$ ". En effet, si par exemple P est nul, on a également $P \times Q = 0$ et alors

$$\deg(P \times Q) = -\infty$$

et

$$\deg(P) + \deg(Q) = -\infty + \deg(Q) = -\infty$$

car il ne peut y avoir de forme indéterminée ($\deg(Q)$ est soit un nombre fini, soit $-\infty$.)

Corollaire

Avec les notations précédentes, $\deg P^n = n \deg P$

Démonstration :

Cette démonstration se fait par récurrence sur n , avec le passage de l'hérédité qui se fait grâce à $P^{n+1} = P^n P$ et à la propriété du degré de $P \times Q$. \square

Corollaire Intégrité de $\mathbb{K}[X]$

Soient P et Q deux polynômes à coefficients dans \mathbb{K} , on a :

$$P \times Q = 0 \iff P = 0 \text{ ou } Q = 0.$$

Démonstration :

D'après l'hypothèse, on a $\deg(P \times Q) = \deg(0) = -\infty$.
D'après les propriétés, on a

$$\deg(P \times Q) = \deg(P) + \deg(Q)$$

D'où

$$\deg(P) + \deg(Q) = -\infty$$

ce qui n'est possible que si l'un au moins parmi $\deg(P)$ ou $\deg(Q)$ est $-\infty$, c'est-à-dire que si l'un au moins ds polynômes P ou Q est nul. \square

Corollaire simplification dans $\mathbb{K}[X]$

Soient $A, B, C \in \mathbb{K}[X]$, avec A non nuls

$$AB = AC \iff B = C$$

Démonstration :

$$AB = AC \iff A(B - C) = 0 \underset{A \neq 0}{\iff} B - C = 0$$

□

III Décomposition d'un polynôme en sous-facteurs

III-1 Divisibilité

Définition

Soient A et P dans $\mathbb{K}[X]$. On dit que A *divise* P et on note $A|P$ s'il existe un polynôme Q à coefficients dans \mathbb{K} tel que : $P = A \times Q$. On dit alors que :

- P est un *multiple* de A ou P est *divisible* par A .
- A est un *diviseur* de P .
- Q est le *quotient* de P par A .

Exemples :

- 10 ■ $X - i$ et $X + i$ divisent tous deux $X^2 + 1$ car $X^2 + 1 = (X + i) \times (X - i)$.
- 11 ■ Tout polynôme divise 0.

CONFUSION

Ne pas confondre la divisibilité des polynômes dans $\mathbb{K}[X]$ et la divisibilité d'éventuels coefficients entiers !

Exemple 12 :

Le polynôme constant 2 divise $X + i$. On a en effet

$$X + i = 2 \left(\frac{1}{2}X + \frac{i}{2} \right)$$

N'oublions pas que les coefficients peuvent être quelconques dans \mathbb{K} . Rien ne les oblige à être entiers ...

Remarque :

Partant de ce principe, on peut constater que les polynômes constants non nuls divisent n'importe quel autre polynôme.

Proposition 46

Soient A et P deux polynômes à coefficients dans \mathbb{K} . Alors

$$A \text{ divise } P \implies \deg(A) \leq \deg(P).$$

SENS DE L'IMPLICATION

Le résultat précédent n'est en aucun cas une équivalence !

Exemple 13 :

Si $Q = 1 + X$ et $P = 1 - X + X^2$, alors on a bien $\deg Q \leq \deg P$, mais Q ne divise pas P . (La vérification pourra se faire simplement à l'aide des racines. Argument que nous reverrons ici dans la suite du chapitre ci-dessous.)

III-2 Racine et divisibilité

Définition

Soient $P \in \mathbb{K}[X]$ et $\alpha \in \mathbb{K}$.

On dit que α est une *racine* de P (ou un *zéro* de P) si $P(\alpha) = 0$.

Exemple 14 :

i est racine de $P = X^2 + 1$ car $P(i) = i^2 + 1 = 0$.

ou de manière plus générale sur les polynômes réels de degré 2 :

Exemple 15 :

Soit $P = aX^2 + bX + c \in \mathbb{R}[X]$ avec $a \neq 0$. On pose $\Delta = b^2 - 4ac$. On pose δ tel que $\delta^2 = \Delta$. Ainsi, les racines de P sont

$$r_1 = \frac{-b - \delta}{2a}, \quad \frac{-b + \delta}{2a}$$



Remarque :

Dit autrement, au cas par cas, ceci signifie que :

$$\begin{cases} \text{si } \Delta > 0 : & \delta = \sqrt{\Delta} \text{ et } r_1 = \frac{-b-\sqrt{\Delta}}{2a}, \quad \frac{-b+\sqrt{\Delta}}{2a} \\ \text{si } \Delta = 0 : & \delta = 0 \text{ et il y a une unique racine } r = \frac{-b}{2a} \\ \text{si } \Delta < 0 : & \delta = i\sqrt{|\Delta|} \text{ et } r_1 = \frac{-b-i\sqrt{|\Delta|}}{2a}, \quad \frac{-b+i\sqrt{|\Delta|}}{2a} \end{cases}$$

? Exercice 2

Vérifier que les racines précédentes fonctionnent également pour les polynômes à coefficients complexes.

Solution

Dans tous les cas, on a

$$\begin{aligned} a(X - r_1)(X - r_2) &= aX^2 - a(r_1 + r_2) + ar_1r_2 = aX^2 - a\frac{-2b}{2a} + a\frac{(-b)^2 - \delta^2}{4a^2} \\ &= aX^2 + bX + \frac{(-b)^2 - \Delta}{4a^2} = aX^2 + bX + a\frac{4ac}{4a^2} \\ &= aX^2 + bX + c \end{aligned}$$

Ainsi, r_1 et r_2 sont toutes deux des racines de P .

On rappelle ce résultat obtenu en première année, que l'on peut démontrer facilement à l'aide du TVI :

Propriété 47

Tout polynôme réel de degré impair admet au moins une racine réelle.

■ Exemple 16 :

Le polynôme $P = X^5 - 3X^2 + 1$ admet au moins une racine réelle (même si on ne sait pas la déterminer a priori...)

Proposition 48

Soient P un polynôme non constant à coefficients dans \mathbb{K} et $\alpha \in \mathbb{K}$. On a :

$$\alpha \text{ est racine de } P \iff (X - \alpha) \text{ divise } P.$$

Démonstration :

De même que pour les autres résultats, on peut généraliser aisément la démonstration de première année. \square

? Exercice 3

Montrer que $X + 1$ divise $X^5 + 3X^4 + X - 1$.

Solution

On note que $X + 1 = X - (-1)$. Or, si on note $P = X^5 + 3X^4 + X - 1$, on a

$$P(-1) = (-1)^5 + 3(-1)^4 + (-1) - 1 = -1 + 3 - 1 - 1 = 0$$

D'où $(X + 1)$ divise P .

Si on prend plusieurs racines, les facteurs s'enchaînent :

Propriété 49

Soient $\alpha_1, \dots, \alpha_m \in \mathbb{K}$ deux à deux distincts, et $P \in \mathbb{K}[X]$. Alors

$$\text{Il existe } Q \in \mathbb{K}[X] \text{ tel que } \left| \begin{array}{l} P = (X - \alpha_1) \dots (X - \alpha_m)Q \\ \Leftrightarrow \end{array} \right| \alpha_1, \dots, \alpha_m \in \mathbb{K} \text{ sont racines de } P$$

De plus, le polynôme Q ainsi défini est unique et $\deg Q = n - m$ avec $n = \deg P$.

Démonstration :

• Unicité :

Si $P = (X - \alpha_1) \dots (X - \alpha_m)Q_1$ et $P = (X - \alpha_1) \dots (X - \alpha_m)Q_2$, on a

$$(X - \alpha_1) \dots (X - \alpha_m)Q_1 = (X - \alpha_1)^{k_1} \dots (X - \alpha_m)Q_2$$

Comme $(X - \alpha_1) \dots (X - \alpha_m) \neq 0$, on a par simplification polynomiale :

$$Q_1 = Q_2$$

• \Rightarrow :

Les α_i sont clairement racines de P .

• \Leftarrow :

On le démontre ici par récurrence sur m .

★ Initialisation : Pour $m = 1$, c'est la proposition précédente.

★ Hérédité : On suppose que c'est vrai pour m et on le montre pour $m + 1$.

$\alpha_1, \dots, \alpha_m \in \mathbb{K}$ étant en particulier par hypothèse racines de P , l'hypothèse de récurrence nous dit qu'il existe Q_1 tel que

$$P = (X - \alpha_1) \dots (X - \alpha_m)Q_1$$

Mais α_{m+1} est également racine de P . Ainsi

$$0 = P(\alpha_{m+1}) = \underbrace{(\alpha_{m+1} - \alpha_1) \dots (\alpha_{m+1} - \alpha_m)}_{\neq 0} Q_1(\alpha_{m+1})$$

D'où $Q_1(\alpha_{m+1}) = 0$; i.e. α_{m+1} est racine de Q_1 . Alors, par la proposition précédente, il existe Q tel que

$$Q_1 = (X - \alpha_{m+1})Q \quad \text{avec } Q(\alpha_{m+1}) \neq 0$$

D'où

$$P = (X - \alpha_1) \dots (X - \alpha_m)(X - \alpha_{m+1})Q$$

CQFD \square

? Exercice 4

Montrer que $(X - 1)(X - 2)$ divise le polynôme $P = X^3 - 2X^2 - X + 2$

Solution

On a que $P(1) = P(2) = 0$, d'où le résultat.

Corollaire

Un polynôme non nul de degré n admet au plus n racines deux à deux distinctes.

Démonstration :

Si on note m le nombre de racines 2 à 2 distinctes du polynôme et si on note $\alpha_1, \dots, \alpha_m$ ces racines, alors, d'après la propriété précédente, on sait que $(X - \alpha_1) \dots (X - \alpha_m)$ divise P .

Par les considération de degré, on sait alors que

$$\deg((X - \alpha_1) \dots (X - \alpha_m)) \leq \deg P$$

mais $\deg((X - \alpha_1) \dots (X - \alpha_m))$ correspond également au nombre m ici de racines. D'où

$$m \leq \deg P$$

\square

■ Exemple 17 :

Le polynôme $X^3 + 2X$ n'a au maximum que 3 racines distinctes. (Si on fait le calcul ici on s'aperçoit que c'est 0, $i\sqrt{2}$ et $-i\sqrt{2}$.)

Le polynôme $X^3 + 2X^2 + X$ n'a au maximum que 3 racines distinctes. Si on les calcule, on s'aperçoit en effet qu'il n'y a que 0 et -1 .

⚠ ÉCRITURE

La recherche des racines non connues d'un polynôme doit se faire avec une certaine réserve au niveau des notations. Regardons ceci de plus près :

- soit on veut vérifier si un nombre α est racine, auquel cas, il n'y a pas de problème, on peut écrire sans soucis

$$P(\alpha) = \dots (\text{on remplace } X \text{ par } \alpha \text{ dans l'expression de } P(X)).$$

- soit on veut chercher les racines sous forme d'équation. Par exemple, on cherche les racines réelles de $X^4 - 2X$. Or

$$\text{l'écriture } X^4 - X = 0 \Leftrightarrow \dots \quad \text{est FAUSSE!!}$$

Rappelons en effet que $X^4 - X$ est une application et que $X^4 - X = 0$ signifie : "l'application $X^4 - X$ est nulle", ce qui est faux, car la seule possibilité pour un polynôme nul est que tous ses coefficients soient nuls. Or, ici ses coefficients ne le sont manifestement pas...

Pour solutionner ce problème, on cherchera les racines en posant plutôt l'équation $x^4 - x = 0$ où $x \in \mathbb{R}$.

(Pour information les solutions sont 0 et 1. Ainsi, $X^4 - X = X(X - 1)Q$ avec Q un polynôme de degré 2 sans racine dans \mathbb{R} .)

⚠ Remarque :

On notera que dans la propriété précédente sur la décomposition (p.38), on ne suppose pas que $\{\alpha_1, \dots, \alpha_m\}$ forme nécessairement l'ensemble de toutes les racines de P . Il peut y en avoir d'autres. cf. *exercices suivants*

? Exercice 5

On a vu précédemment que $(X - 1)(X - 2)$ divise le polynôme $P = X^3 - 2X^2 - X + 2$. Déterminer "à vue" l'ensemble de toutes les racines de P par une recherche de racine évidente.

Solution

De part son degré, ce polynôme admet au maximum trois racines. On en a déjà 2 avec 1 et 2. Il en reste une seule éventuelle à trouver. On peut tenter des racines "évidentes" et on constate qu'en effet, -1 est elle aussi racine. Autrement dit, on est au complet et l'ensemble des racines de P est $\{-1, 1, 2\}$

Corollaire

Un polynôme de degré n qui a au moins $n + 1$ racines est nul.

Démonstration :

Notons P le polynôme en question et supposons par l'absurde qu'il soit non nul. Son degré n devrait majorer son nombre de racines, qui est par hypothèse $\geq n + 1$. ca donnerait $n \geq n + 1$. Nous voici avec une belle contradiction! \square

 **Remarque :**

Le corollaire précédent s'applique en particulier sur une infinité de racines : un polynôme ayant une infinité de racines est nécessairement le polynôme nul.

 **Exercice 6**

Montrer que la fonction sin ne peut pas être un polynôme.

Solution

Supposons que $P = \sin$ soit un polynôme. Comme on a $\sin(n\pi) = 0$ pour tout $n \in \mathbb{N}$, on a une infinité de racines distinctes pour P , ce qui n'est possible que si $P = 0$. Ainsi, la fonction sin serait nulle, ce qui est faux!

 **Exercice 7**

Soient $P, Q \in \mathbb{K}[X]$ deux polynômes. Montrer que si on a une infinité de valeurs $\alpha_0, \alpha_1, \dots \in \mathbb{K}$, deux à deux distinctes, telles que

$$P(\alpha_i) = Q(\alpha_i) \quad \forall i \in \mathbb{N}$$

Alors on a tout simplement $P = Q$ (c'est-à-dire $P(x) = Q(x)$ pour tout $x \in \mathbb{R}$.)

Solution

Il suffit de poser le polynôme $H = P - Q$ qui a alors une infinité de racines $\alpha_i \forall i \in \mathbb{N}$ et qui est donc nul.

 **Définition**

Soient $P \in \mathbb{K}[X]$, $\alpha \in \mathbb{K}$ et $n \in \mathbb{N}^*$. On dit que :

- α est racine de P d'ordre (ou de multiplicité) au moins k si $(X - \alpha)^k$ divise P .
- α est racine de P d'ordre exactement k si $(X - \alpha)^k$ divise P et $(X - \alpha)^{k+1}$ ne divise pas P .
- Si α est racine de P d'ordre exactement 1, on dit que α est une *racine simple* de P .
- α est une *racine double* de P si α est racine de P d'ordre exactement 2.
- α est une *racine multiple* de P si α est racine de P d'ordre au moins 2.

Tout d'abord un petit rappel de première année sur les polynômes réels :

Proposition 50

Soit P un polynôme à coefficients réels. $\alpha \in \mathbb{R}$ est une racine multiple de P ssi

$$P(\alpha) = P'(\alpha) = 0$$

Démonstration :

la démonstration a déjà été vue en première année, nous ne revenons pas sur la question. \square

 **Exercice 8**

Montrer que 2 est racine multiple de $P = X^4 - 2X^3 - 3X^2 + 4X + 4$.

Solution

On vérifie que $P(2) = P'(2) = 0$

 **ATTENTION AUX GÉNÉRALISATIONS ABUSIVES**

Ceci ne peut être généralisé aux polynômes à coefficients complexes ou aux racines complexes, n'ayant pas de définition de la dérivée dans $\mathbb{C}[X]$.

La propriété précédente (p.38) sur la décomposition en facteurs de degrés 1 peut être généralisée à la problématique "avec multiplicités" :

Théorème 51 décomposition suivant les racines

Soient $\alpha_1, \dots, \alpha_m \in \mathbb{K}$ deux à deux distincts, $k_1, \dots, k_m \in \mathbb{N}^*$, $P \in \mathbb{K}[X]$. Alors

$$\begin{array}{l}
 \text{Il existe } Q \in \mathbb{K}[X] \text{ tel que} \\
 P = (X - \alpha_1)^{k_1} \dots (X - \alpha_m)^{k_m} Q \\
 \text{avec } Q(\alpha_i) \neq 0 \text{ pour tout } i = 1, \dots, m
 \end{array}
 \Leftrightarrow
 \begin{array}{l}
 \alpha_1, \dots, \alpha_m \in \mathbb{K} \text{ sont racines de } P \\
 \text{d'ordres de multiplicité respectifs} \\
 k_1, \dots, k_m
 \end{array}$$

De plus, le polynôme Q ainsi défini est unique et $\deg Q = n - (k_1 + \dots + k_m)$ avec $n = \deg P$.

Démonstration :

- Unicité :

Si $P = (X - \alpha_1)^{k_1} \dots (X - \alpha_m)^{k_m} Q_1$ et $P = (X - \alpha_1)^{k_1} \dots (X - \alpha_m)^{k_m} Q_2$, on a

$$(X - \alpha_1)^{k_1} \dots (X - \alpha_m)^{k_m} Q_1 = (X - \alpha_1)^{k_1} \dots (X - \alpha_m)^{k_m} Q_2$$

Comme $(X - \alpha_1)^{k_1} \dots (X - \alpha_m)^{k_m} \neq 0$, on a par simplification polynomiale :

$$Q_1 = Q_2$$

• $\Rightarrow :$

Les α_i sont clairement racines de P . Voyons maintenant leur ordre de multiplicité :

Soit $i \in \{1, \dots, m\}$. Si on pose $T_i = \prod_{\substack{j=1 \\ j \neq i}}^n (X - \alpha_j)^{k_j} Q$, alors

$$P = (X - \alpha_i)^{k_i} T_i$$

et $T_i(\alpha_i) = \prod_{\substack{j=1 \\ j \neq i}}^n (\alpha_i - \alpha_j)^{k_j} Q(\alpha_i) \neq 0$. Donc k_i est bien l'ordre de multiplicité de α_i .

• $\Leftarrow :$

On le démontre ici par récurrence sur m .

* Initialisation : Pour $m = 1$, c'est la définition de la multiplicité de α_1 .

* Hérité : On suppose que c'est vrai pour m et on le montre pour $m + 1$.

$\alpha_1, \dots, \alpha_m \in \mathbb{K}$ étant en particulier par hypothèse racines de P d'ordres de multiplicité respectifs k_1, \dots, k_m , l'hypothèse de récurrence nous dit qu'il existe Q_1 tel que

$$P = (X - \alpha_1)^{k_1} \dots (X - \alpha_m)^{k_m} Q_1$$

Mais α_{m+1} est également racine de P . Ainsi

$$0 = P(\alpha_{m+1}) = \underbrace{(\alpha_{m+1} - \alpha_1)^{k_1} \dots (\alpha_{m+1} - \alpha_m)^{k_m}}_{\neq 0} Q_1(\alpha_{m+1})$$

D'où $Q_1(\alpha_{m+1}) = 0$; i.e. α_{m+1} est racine de Q_1 . On note n son ordre de multiplicité dans Q_1 . Alors il existe Q tel que

$$Q_1 = (X - \alpha_{m+1})^n Q \quad \text{avec } Q(\alpha_{m+1}) \neq 0$$

D'où

$$P = (X - \alpha_1)^{k_1} \dots (X - \alpha_m)^{k_m} (X - \alpha_{m+1})^n Q$$

On note

$$T = (X - \alpha_1)^{k_1} \dots (X - \alpha_m)^{k_m} Q$$

alors

$$P = (X - \alpha_{m+1})^n T$$

et on peut observer là encore, en réinjectant, que

$$T(\alpha_{m+1}) \neq 0$$

Ainsi, par définition, le n ci-dessus est également l'ordre de multiplicité de α_{m+1} dans P , c'est-à-dire que $n = m + 1$. CQFD \square

? Exercice 9

On sait que $(X - 1)(X - 2)$ divise le polynôme $P = X^3 - 2X^2 - X + 2$. Déterminer par calcul toutes les racines de P .

Solution

1 et 2 sont racines. Comme $\deg P = 3$, d'après le théorème de décomposition, il existe alors $a, b \in \mathbb{R}$ tels que $P = (X - 1)(X - 2)(aX + b)$. On a donc

$$X^3 - 2X^2 - X + 2 = aX^3 - (a + 2a + b)X^2 + \dots$$

L'identification des coefficients nous dit alors que

$$a = 1, \quad a + 2a + b = 2, \quad \text{i.e. } a = 1, \quad b = -1$$

et donc $P = (X - 1)(X - 2)(X + 1)$. Les racines de P sont donc $-1, 1, 2$.

Corollaire somme des multiplicités

Soit $P \in \mathbb{K}[X]$. Si $\alpha_1, \dots, \alpha_m \in \mathbb{K}$ sont des racines deux à deux distinctes de P de multiplicités respectives $k_1, \dots, k_m \in \mathbb{N}^*$. Alors

$$k_1 + \dots + k_m \leq \deg(P)$$

Démonstration :

C'est une simple conséquence du théorème précédent : comme $(X - \alpha_1)^{k_1} \dots (X - \alpha_m)^{k_m}$ divise P , alors

$$\deg((X - \alpha_1)^{k_1} \dots (X - \alpha_m)^{k_m}) \leq \deg P$$

et

$$\deg((X - \alpha_1)^{k_1} \dots (X - \alpha_m)^{k_m}) = k_1 + \dots + k_m$$

D'où

$$k_1 + \dots + k_m \leq \deg P$$

\square

Intéprétation et corollaire du corollaire précédent :

Corollaire Nombre de racines

Un polynôme non nul de degré n admet au plus n racines comptées avec leur multiplicité.

■ Exemple 18 :

Le polynôme $X^3 + 2X^2 + X$ n'a au maximum que 3 racines distinctes. Si on les calcule, on s'aperçoit qu'on n'a que 0 et -1 .

Néanmoins, on sait que la somme des multiplicités de chacune d'entre elles ne peut dépasser 3. Ainsi, on sait qu'au maximum, l'une sera de multiplicité maximum 2 et l'autre de multiplicité 1.

On s'aperçoit qu'on a en effet ici $X^3 + 2X^2 + X = X(X^2 + 2X + 1) = X(X + 1)^2$. Ainsi, -1 est racine double et 0 racine simple.

? Exercice 10

Déterminer l'ensemble des racines de $P = X^4 - 2X^3 - 3X^2 + 4X + 4$ sachant qu'il a deux racines réelles doubles, dont 2 et que $P'(X) = (X - 2)(X - 1)(X + 1)$

Solution

Le polynôme P est réel. Une racine α est double ssi $P(\alpha) = P'(\alpha) = 0$. Les racines étant ici doubles, on peut donc les chercher parmi celles de P' . On sait déjà que 2 est racine double. Vérifions maintenant si 1 est racine double :

$$P(1) = 1 - 2 - 3 + 4 + 4 = 4 \neq 0$$

1 n'est pas racine. L'autre racine double ne peut donc être que -1 . (On peut éventuellement le vérifier en constatant que $P(-1) = 0$)

Les deux racines doubles sont donc 2 et -1 . Il ne peut y avoir d'autres racines, puisque la somme des multiplicités vaut au moins $2 + 2 = 4 = \deg P$. On a donc ici l'ensemble de toutes les racines.

III-3 Racines dans \mathbb{C} et décomposition dans $\mathbb{C}[X]$

Certains polynômes, construits au départ dans \mathbb{R} , peuvent néanmoins avoir des racines complexes. On prend pour exemple $P = X^2 + 1 \in \mathbb{R}[X]$, qui n'a aucune racine réelle mais admet exactement deux racines complexes i et $-i$. On peut alors l'écrire en deux facteurs

$$\underbrace{X^2 + 1}_{\in \mathbb{R}[X]} = \underbrace{(X - i)}_{\in \mathbb{C}[X]} \underbrace{(X + i)}_{\in \mathbb{C}[X]}$$

mais il n'existe en revanche aucune décomposition dans $\mathbb{R}[X]$. Sinon, on aurait $X^2 + 1 = (X - a)(X - b)$ dans $\mathbb{R}[X]$, c'est-à-dire deux racines a, b réelles.

Pour certaines nécessités de factorisation, on est donc amené à considérer les expressions dans $\mathbb{C}[X]$. Nous allons donc voir dans cette partie quelques propriétés spécifiques amenés par les complexes.

⚠ Remarque :

Notons que dans l'exemple précédent, les deux racines trouvées i et $-i$ sont conjuguées. D'ailleurs, on peut au passage rappeler que dans le cas plus général des racines d'un polynôme P de degré 2, avec $P = a + bX + cX^2$ avec $a \neq 0$, en cas de discriminant négatif ($\Delta = b^2 - 4ac < 0$), on obtient également deux racines conjuguées :

$$r_1 = \frac{-b - i\sqrt{|\Delta|}}{2a} \quad r_2 = \frac{-b + i\sqrt{|\Delta|}}{2a}$$

et bien en réalité, ceci est vrai quel que soit le degré du polynôme mais avec néanmoins une "petite" (!) restriction aux polynômes réels :

Propriété 52

Soit P un polynôme à coefficients réels. Si $\alpha \in \mathbb{C}$ est racine de P , alors $P(\bar{\alpha}) = 0$, c'est-à-dire que $\bar{\alpha}$ est également racine de P .

Démonstration :

On note $P = a_0 + a_1X + \dots + a_nX^n$ avec par hypothèse $a_0, \dots, a_n \in \mathbb{R}$. On sait que $P(\alpha) = 0$. Comme $P(\alpha)$ est un nombre, on peut prendre son complémentaire "de chaque côté de l'égalité" :

$$\overline{P(\alpha)} = \bar{0}$$

Or, d'une part,

$$\overline{P(\alpha)} = \overline{a_0 + a_1\alpha + \dots + a_n\alpha^n} = \bar{a}_0 + \bar{a}_1 \bar{\alpha} + \dots + \bar{a}_n \bar{\alpha}^n$$

Mais comme les a_i sont tous réels, on a

$$\bar{a}_i = a_i \quad \forall i = 0, \dots, n$$

et donc

$$\overline{P(\alpha)} = a_0 + a_1\bar{\alpha} + \dots + a_n\bar{\alpha}^n = P(\bar{\alpha})$$

D'autre part, on a également

$$\bar{0} = 0$$

d'où, en combinant les deux résultats et en réinjectant dans la première ligne, on obtient

$$P(\bar{\alpha}) = 0$$

c'est-à-dire que le conjugué de α est racine également \square

Commentaires :

Ce résultat est valable que α soit complexe non réel, ou réel. En effet, si jamais $\alpha \in \mathbb{R}$, la propriété dit simplement que si α est racine, alors $\bar{\alpha} = \alpha$ est racine. Bref, il n'y a pas spécialement de contradiction, mais juste un "non résultat" !



IL FAUT DES COEFFICIENTS RÉELS

Il faut bien faire attention ici à avoir un polynôme dont **tous** les coefficients sont **réels**! En effet : Le polynôme $X - i \notin \mathbb{R}[X]$ a i comme racine, mais son conjugué $-i$ n'est pas racine. (à vérifier...)



Remarque :

Ce résultat nous dit en particulier que si on a trouvé une racine complexe α d'un polynôme réel P , alors on en a automatiquement une autre $\bar{\alpha}$. Ainsi, $(X - \alpha)(X - \bar{\alpha})$ divise P . Or,

$$Q = (X - \alpha)(X - \bar{\alpha}) = X^2 - (\alpha + \bar{\alpha})X + \alpha\bar{\alpha} = X^2 - 2\text{Re}(\alpha)X + |\alpha|^2$$

qui est du coup un polynôme réel. Ainsi, on a trouvé un polynôme réel Q de degré 2 tel que Q divise P .

? Exercice 11

Montrer que i est racine de $P = -X^3 + X^2 - X + 1$ et en déduire que $X^2 + 1$ divise P .

Solution

On a que $P(i) = -i^3 + i^2 - i + 1 = i - 1 - i + 1 = 0$.

Comme P est un polynôme réel, on sait alors que $-i$ est également racine de P et donc que $Q = (X - i)(X + i)$ divise P . Or, $Q = X^2 + 1$, d'où le résultat.

Voyons maintenant un résultat très important sur les polynômes, qui est la base de la décomposition en sous-facteurs :

Théorème 53 De D'Alembert

Soit P un polynôme à coefficients réels ou complexes **non constant**. P admet au moins une racine dans \mathbb{C} .

Démonstration : admise. \square

Corollaire

Tout polynôme P de $\mathbb{K}[X]$ s'écrit de façon unique dans $\mathbb{C}[X]$ (à l'ordre près) sous la forme :

$$P = \lambda \times (X - \alpha_1)^{k_1} \times \dots \times (X - \alpha_m)^{k_m}$$

avec :

- λ le coefficient dominant de P .
- m un entier naturel.
- $\alpha_1, \dots, \alpha_m \in \mathbb{C}$ les racines de P de multiplicités respectives k_1, \dots, k_m .
- $k_1 + k_2 + \dots + k_m$ le degré de P .

■ Exemple 19 :

Si $\sqrt{2}$ est le coefficient dominant d'un polynôme P dont les racines sont i et 2 avec des multiplicités respectives de 3 et 1 alors P est nécessairement le polynôme suivant :

$$P = \sqrt{2}(X - i)^3 \times (X - 2).$$



Définition

Factoriser dans $\mathbb{C}[X]$ (en produit de facteurs irréductibles) signifie qu'il faut trouver cette décomposition en facteurs de degré 1.

? Exercice 12

Montrer que i et 1 sont racines de $P = -X^3 + X^2 - X + 1$. En déduire une factorisation dans \mathbb{C} de P .

Solution

On a déjà vu dans l'exercice précédent que $P(i) = 0$. De même,

$$P(1) = 0$$

Ainsi, i et 1 sont bien des racines.

Le polynôme P étant réel, on sait que $-i$ est racine également.

De plus, P étant de degré 3, il ne peut pas y avoir plus de 3 racines. On les a donc toutes et on obtient dans la foulée l'existence de Q de degré $3 - 3 = 0$ tel que

$$P = (X - i)(X + i)(X + 1)Q$$

Q étant constant, c'est le coefficient dominant, c'est-à-dire -1 . Ainsi,

$$P = -(X - i)(X + i)(X + 1)$$

Dans ce chapitre, on utilisera régulièrement l'abus de notation $y = y(x)$ pour une fonction y , qui confond la fonction y avec son image $y(x)$. On veillera d'autant plus à bien comprendre de quel objet on parle lors des calculs !

I Ordre 1

$$y' + ay = c$$

Soient $a, c : I \rightarrow \mathbb{R}$ deux fonctions continues sur un intervalle ouvert I . On considère l'équation différentielle

$$y' + ay = c \quad \text{sur } I$$

où on rappelle que ceci un abus de notation pour désigner l'équation :

$$y'(x) + a(x)y(x) = c(x) \quad \forall x \in I$$

Si $c \neq 0$, (i.e. si l'équation n'est pas homogène), cette résolution se fait en plusieurs étapes

- recherche des solutions de l'équation homogène $y' + a(x)y = 0$
- recherche d'une solution particulière.
- Constitution des solutions générales avec le principe de superposition.
- Éventuellement recherche de solutions particulières avec "condition initiale".

sachant que l'ordre dans lequel sont traitées ces étapes, ainsi que la méthode utilisée, peut dépendre des données du problème ! Il faut donc bien lire l'énoncé avant de se lancer dans une résolution générale... Revoyons tout ceci ci-dessous.

Résolution d'une équation homogène $(E_0) : y' + a(x)y = 0$

Théorème 54

L'ensemble des solutions de l'équation $y' + a(x)y = 0$ est

$$\mathcal{S}_0 = \{y : I \rightarrow \mathbb{R} \mid y(x) = \lambda \underbrace{e^{-A(x)}}_{y_0(x)}, \lambda \in \mathbb{R}\}$$

où A est une primitive quelconque de a sur I .

■ Exemple 1 :

On cherche les fonctions réelles y, C^1 sur $]0; +\infty[$ vérifiant

$$y'(x) = \frac{1}{2x}y(x) \quad \forall x \in]0; +\infty[$$

Une primitive sur $]0; +\infty[$ de $a(x) = -\frac{1}{2x}$ est $A(x) = -\frac{1}{2} \ln x$, ainsi, les solutions sont toutes les fonctions

$$x \mapsto \lambda e^{\frac{1}{2} \ln x} = \lambda \sqrt{x}, \quad \text{où } \forall \lambda \in \mathbb{R}$$

Résolution d'une équation générale $(E) : y' + a(x)y = c(x)$, où $c \neq 0$

Dans tous les cas, il nous faut résoudre l'équation homogène.

Nous avons ensuite besoin d'au moins une solution de l'équation générale, que l'on appellera "solution particulière". Notons la ici y_p . Si l'énoncé ne nous fait pas calculer de solution particulière, il faut aller en chercher une :

- Si a et c sont constants, il existe également une solution particulière constante (cf ci-dessous)
- Si a et c sont des polynômes avec $\deg a \leq \deg c$, on pourra tenter de chercher une solution y_0 polynomiale, mais on n'est pas certain d'en trouver une. (cf ci-dessous)
- Dans les autres cas, on pourra tenter la méthode de variation de la constante. Il existe toujours une solution, mais on n'est pas forcément certain de pouvoir l'expliquer. (cf ci-dessous)

Une fois une solution trouvée, on appliquera le principe de superposition :

Théorème 55

L'ensemble des solutions de l'équation (E) est

$$\mathcal{S}_0 = \{y : I \rightarrow \mathbb{R} \mid y(x) = \lambda e^{-A(x)} + y_p(x), \lambda \in \mathbb{R}\}$$

- Si a et c sont constants :

Il existe une solution particulière sous la forme $y_p(x) = \frac{c}{a}$.

Analyse : Supposons que y_p soit une fonction constante. Alors, $y'_p = 0$. Ainsi,

$$y_p \text{ est solution de l'équation } (E) \Leftrightarrow ay_p = c \Leftrightarrow y_p = \frac{c}{a}$$

Synthèse : Par les équivalents précédents, il existe bien sur une fonction constante y_p qui répond à l'équation.

- Si a et c sont des polynômes non nuls :

Supposons $y_p = P$ un polynôme donné. On note $n = \deg P$.

Analyse :

$$y_p \text{ est solution de ce l'équation } (E) \Leftrightarrow P' + aP = c$$

Equation aux degrés :

De l'équation précédente, on tire

$$\deg(P' + aP) = \deg c$$

Or, comme $\deg P' < \deg P$, on a

$$\deg(P' + aP) = \deg P + \deg a$$

D'où :

$$\begin{cases} \deg P = \deg c - \deg a & \text{si } \deg a \leq \deg c \\ \text{une contradiction} & \text{sinon} \end{cases}$$

On réinjecte :

Si on arrive à une contradiction, c'est qu'une telle solution est impossible. Sinon, on récupère le degré obtenu et on passe à un calcul plus précis. (*cf. ci-dessous*)

■ Exemple 2 :

Déterminer une solution polynomiale de l'équation (E) : $y' + xy = x^2 + 1$:

Soit P un polynôme. On note n son degré.

$$P \text{ est solution de } (E) \Leftrightarrow \underbrace{P' + xP}_{\text{de degré } n+1} = \underbrace{x^2 + 1}_{\text{de degré } 2}$$

Ainsi, $n + 1 = 2$ et donc $n = 1$. On peut donc poser $P = \alpha x + \beta$. Alors

$$\begin{cases} P \text{ est solution} \Leftrightarrow \alpha + x(\alpha x + \beta) = x^2 + 1 & \forall x \in \mathbb{R} \\ \Leftrightarrow \alpha x^2 + \beta x + \alpha = x^2 + 1 & \forall x \in \mathbb{R} \end{cases}$$

ce qui est possible dès que $\alpha = 1$, $\beta = 0$. Ainsi, la fonction $y_p = x$ est solution de l'équation.

■ Exemple 3 :

Existe-t-il une solution polynomiale à l'équation (E) : $y' + xy = x^2$:

Ici, " $\deg a < \deg c$ ", mais néanmoins, il n'existe pas de solution polynomiale : Avec les mêmes arguments et notations que pour l'exemple ci-dessus, on arrive à

$$\begin{cases} P \text{ est solution} \Leftrightarrow \alpha x^2 + \beta x + \alpha = x^2 & \forall x \in \mathbb{R} \end{cases}$$

ce qui est impossible car on aurait $1 \stackrel{\text{coeff dominant}}{=} \alpha \stackrel{\text{coeff constant}}{=} 0$, ce qui est contradictoire.

Il n'existe donc en effet pas de solution polynomiale à cette équation.

• Recherche d'une solution particulière de (E) par Méthode de Variation de la Constante (MVC)

On recherche une solution particulière sous la forme $y_p(x) = \lambda(x) e^{-A(x)}$ pour tout $x \in I$, avec λ une fonction dérivable sur I .

Analyse : Si on réinjecte y_p dans l'équation (E), on trouve un candidat pour λ tel que

$$\text{Démonstration : } \lambda'(x) = c(x) e^{A(x)}.$$

on a, pour $x \in I$, $y_p'(x) = \lambda'(x) e^{-A(x)} - \lambda(x) a(x) e^{-A(x)} = \lambda'(x) e^{-A(x)} - a(x) y_p(x)$.
D'où

$$y_p' + a y_p = c \Leftrightarrow \lambda'(x) e^{-A(x)} = c(x) \Leftrightarrow \lambda'(x) = c(x) e^{A(x)}$$

□

Synthèse : On choisit λ une primitive quelconque de $x \mapsto c(x) e^{A(x)} = \frac{c(x)}{y_0(x)}$.

On vérifie que $y_p(x) = \lambda(x) e^{-A(x)}$ est une solution particulière de (E) sur I .

■ Exemple 4 :

Résolvons l'équation $y'(x) = \frac{1}{2x} y(x) + 2$ sur $x \in]0; +\infty[$.

D'après la résolution précédente, on sait déjà que les solutions de l'équation homogène sont les fonctions

$$x \mapsto \lambda \sqrt{x}, \quad \text{où } \forall \lambda \in \mathbb{R}$$

Utilisons maintenant la MVC pour déterminer une solution particulière. Posons $y_p(x) = \lambda(x) \sqrt{x}$ où λ est une fonction $C^1(]0; +\infty[)$.

Alors,

$$\begin{aligned} y_p \text{ est solution particulière} &\Leftrightarrow y_p' = \frac{1}{2x} y_p + 2 \\ &\Leftrightarrow \dots \text{ (calcul, cf ci-dessus)} \\ &\Leftrightarrow \lambda'(x) = \frac{2}{\sqrt{x}} \end{aligned}$$

Par exemple,

$$\lambda(x) = 4\sqrt{x} \text{ convient}$$

Autrement dit, une solution particulière est :

$$y_p : x \in]0; +\infty[\mapsto \lambda(x) \sqrt{x} = 4x$$

D'où, par principe de superposition, l'ensemble de toutes les solutions de l'équation :

$$x \in]0; +\infty[\mapsto \lambda \sqrt{x} + 4x \quad \forall \lambda \in \mathbb{R}$$

Équation avec condition initiale

La recherche de ces solutions se fait toujours **après** avoir résolu l'équation dans son intégralité...

■ Exemple 5 :

Résoudre l'équation $y'(x) = \frac{1}{2x}y(x) + 2$ sur $x \in]0; +\infty[$ avec la condition initiale $y(1) = 5$.

L'ensemble des solutions de (E) sont les fonctions

$$x \in]0; +\infty[\mapsto \lambda\sqrt{x} + 4x \quad \forall \lambda \in \mathbb{R}$$

Soit y une de ces fonctions. Alors :

$$y(1) = 1 \Leftrightarrow \lambda\sqrt{1} + 4 = 5 \Leftrightarrow \lambda = 1$$

D'où la (seule) solution avec condition initiale $y(1) = 5$:

$$x > 0 \mapsto \sqrt{x} + 4x$$



Remarque :

Dans le cadre des équations différentielles d'ordre 1, une condition initiale amènera toujours à une seule solution (au maximum). *à méditer...*

II Ordre 2 à coefficients constants $ay'' + by' + cy = d(x)$

Soient $a, b, c \in \mathbb{R}$, $a \neq 0$ et $d : I \rightarrow \mathbb{K}$ une fonction continue sur un **intervalle** ouvert I . On considère l'équation différentielle

$$(E) \quad ay'' + by' + cy = d$$

C'est-à-dire que l'on cherche l'ensemble des fonctions $\mathcal{C}^2(I)$ vérifiant l'équation. Les étapes sont les mêmes que pour l'ordre 1.

Résolution de l'équation homogène $(E_0) : ay'' + by' + cy = 0$

On pose l'équation caractéristique

$$(C) \quad ar^2 + br + c = 0$$

On note Δ le discriminant de C .

Théorème 56

- Si $\Delta > 0$, on note r_1, r_2 les deux solutions réelles distinctes de (C),

$$\mathcal{S}_0 = \left\{ \begin{array}{l} y : I \rightarrow \mathbb{R} \\ x \mapsto \lambda_1 e^{r_1 x} + \lambda_2 e^{r_2 x} \end{array}, \lambda_1, \lambda_2 \in \mathbb{R} \right\}$$

- Si $\Delta = 0$, on note r_1 la seule solution (réelle) de (C),

$$\mathcal{S}_0 = \left\{ \begin{array}{l} y : I \rightarrow \mathbb{R} \\ x \mapsto (\lambda x + \mu) e^{r_1 x} \end{array}, \lambda, \mu \in \mathbb{R} \right\}$$

- Si $\Delta < 0$, on note $r = \alpha + i\beta$ une solution de (C), ($\alpha, \beta \in \mathbb{R}$)

$$\mathcal{S}_0 = \left\{ \begin{array}{l} y : I \rightarrow \mathbb{R} \\ x \mapsto e^{\alpha x} (\lambda_1 \cos(\beta x) + \lambda_2 \sin(\beta x)) \end{array}, \lambda_1, \lambda_2 \in \mathbb{R} \right\}$$

Commentaires : On remarque qu'une conséquence de ce théorème est que l'ensemble des solutions de l'équation (E_0) est un \mathbb{R} -e.v. de dimension 2. On le note \mathcal{S}_0 .

■ Exemple 6 :

Résolvons l'équation

$$(E_0) : y'' + 2y' + y = 0$$

L'équation caractéristique de (E_0) est

$$r^2 + 2r + r = 0$$

de racine double

$$r = -1$$

L'ensemble des solutions à valeurs dans \mathbb{R} de (E_0) est

$$\mathcal{S} = \{y : x \in \mathbb{R} \mapsto (\lambda + \mu x)e^{-x} \mid \lambda, \mu \in \mathbb{R}\}$$

Solution particulière y_p de (E) si d est une constante

Pour une **constante** d et une équation sous la forme

$$ay'' + by' + cy = d$$

Une solution particulière est

$$y_p : x \mapsto \frac{d}{a}$$

(facile à vérifier !)

Solution particulière y_p de (E) si $d(x) = P(x)e^{\beta x}$

Supposons que $d(x) = P(x)e^{\beta x}$, où β est réel et P est un polynôme réel. Alors on cherche une solution particulière $y_p(x) = Q(x)e^{\beta x}$ de la manière suivante :

★ Analyse : Recherche du degré de Q :

$$\begin{aligned} y_P \text{ est sol. part.} &\Leftrightarrow ay_P'' + by_P' + cy_P = Pe^{\beta x} \quad \forall x \\ &\Leftrightarrow \dots \text{ calcul} \\ &\Leftrightarrow e^{\beta x} (aQ'' + (2a\beta + b)Q' + (\beta^2 + b\beta + c)Q) = Pe^{\beta x} \quad \forall x \\ &\Leftrightarrow aQ'' + \underbrace{(2a\beta + b)}_B Q' + \underbrace{(\beta^2 + b\beta + c)}_C Q = P \end{aligned}$$

Le degré de P est donc également celui de l'expression $aQ'' + BQ' + CQ$. Comme on a

$$\deg Q \geq \deg Q' \geq \deg Q'',$$

on a

$$\begin{cases} \text{si } C \neq 0, \deg Q = \deg P \\ \text{si } C = 0 \text{ mais } B = 0, \deg Q' = \deg P \\ \text{si } C = B = 0, \text{ comme } a \neq 0, \deg Q'' = \deg P \end{cases}$$

En bref (HP) : ceci pourrait se traduire de la manière suivante en terme de polynôme caractéristique :

$$\deg Q(x) = \begin{cases} \deg P & \text{si } \beta \text{ n'est pas racine du polynôme caractéristique.} \\ \deg P + 1 & \text{si } \beta \text{ est racine simple du polynôme caractéristique.} \\ \deg P + 2 & \text{si } \beta \text{ est racine double du polynôme caractéristique.} \end{cases}$$

★ Synthèse : Recherche d'une solution s'écrivant sous la forme proposée :

On pose $Q = a_0 + a_1X + \dots + a_nX^n$ avec n le degré trouvé précédemment et on procède par équivalence pour déterminer y_p comme solution de l'équation :

y_P solution **ssi** ...

on pourra ainsi reprendre et poursuivre les calculs effectués dans la partie "analyse" en l'appliquant au degré trouvé.

■ Exemple 7 :

On considère l'équation

$$(E) \quad y'' + 2y' + y = xe^{-x}$$

On propose de trouver une solution particulière sous la forme $y_P = Qe^{-x}$ où Q est un polynôme réel.

On a

$$y_P = Q'e^{-x} - Qe^{-x} = (Q' - Q)e^{-x}, \quad y_P'' = (Q'' - 2Q' + Q)e^{-x}$$

Ainsi,

$$\begin{aligned} y_P \text{ est sol. part.} &\Leftrightarrow y_P'' + 2y_P' + y_P = xe^{-x} \quad \forall x \\ &\Leftrightarrow \dots \text{ calculs} \\ &\Leftrightarrow Q'' + 0 \cdot Q' + 0 \cdot Q = x \quad (*) \end{aligned}$$

On obtient alors $\deg Q'' = \deg x = 1$, d'où

$$\deg Q = 3.$$

On pose alors pour tout $x \in \mathbb{R}$:

$$y_P(x) = (a + bx + cx^2 + dx^3)e^{-x}$$

Ainsi, d'après (*) et avec $Q'' = 2c + 6x$, y_P est solution de (E) ssi pour tout x :

$$2c + 6dx = x$$

i.e. ssi

$$\begin{cases} c = 0 \\ d = \frac{1}{6}, \end{cases}$$

les a, b pouvant être choisis de manière quelconque. On sait donc qu'on a comme solution particulière la fonction y_P telle que

$$y_P : x \mapsto \frac{1}{6}x^3 e^{-x}$$

⚠ Remarque :

Dans l'exemple précédent, l'ensemble des **fonctions** $y = (a + bx)e^{-x}$ **étant déjà solutions de l'équation homogène**, en réalité, on pouvait poser, dès lors qu'on a su que $\deg Q = 3$ la proposition

$$y_P(x) = (cx^2 + dx^3)e^{-x} \quad \text{au lieu de} \quad y_P(x) = (a + bx + cx^2 + dx^3)e^{-x}$$

(la partie $(a + bx)e^{-x}$ n'apportant aucune contribution au second membre.)

Solution particulière y_p de (E) si $d(x) = P(x)$

C'est tout simplement un cas particulier du cas précédent, avec $\beta = 0$. En effet,

$$d(x) = P(x)e^{0x}$$

On peut donc proposer donc une recherche de solution sous forme polynomiale par exemple.

Principe de superposition : solution particulière y_p de (E) si $d(x) = d_1(x) + d_2(x) + \dots$

Si y_1 est solution de $ay'' + by' + cy = d_1$ et y_2 est solution de $ay'' + by' + cy = d_2$, alors

$$y_P = y_1 + y_2$$

est solution de $ay'' + by' + cy = d_1 + d_2$. (Facile à montrer !)

■ Exemple 8 :

On souhaite trouver une solution particulière de l'équation

$$(E) : y'' + 2y' + y = x(e^{-x} + e^x) + 1$$

On connaît déjà l'ensemble des solutions de l'équation homogène par un exemple précédent :

$$S_0 = \{y : x \in \mathbb{R} \mapsto (\lambda + \mu x)e^{-x} \mid \lambda, \mu \in \mathbb{R}\}$$

★ Recherche d'une solution particulière de $(E_1) : y'' + 2y' + y = xe^{-x}$:

Dans un exemple précédent, on a trouvé $y_1(x) = \frac{1}{6}x^3e^{-x}$.

★ Recherche d'une solution particulière de $(E_2) : y'' + 2y' + y = xe^x$:

Le même type de recherche que précédemment donne une solution particulière : $y_2 : x \mapsto \frac{1}{4}(x-1)e^x$.

★ Recherche d'une solution particulière de $(E_3) : y'' + 2y' + y = 1$:

Le second membre est constant, une solution particulière est donc y_3 avec

$$y_3(x) = 1$$

★ En conclusion, une solution particulière de (E) est

$$y_P(x) = \frac{1}{6}x^3e^{-x} + \frac{1}{4}(x-1)e^x + 1$$

Solution générale de (E)

Théorème 57

L'ensemble des solutions de l'équation $(E) : ay'' + by' + cy = d$ est

$$S = S_0 + y_p(x)$$

■ Exemple 9 :

On souhaite résoudre l'équation

$$(E) : y'' + 2y' + y = x(e^x + e^{-x}) + 1$$

Les exemples précédents nous permettent d'affirmer que l'ensemble des solutions à valeurs dans \mathbb{R} sont toutes les fonctions

$$y : x \in \mathbb{R} \mapsto (\lambda + \mu x)e^{-x} + \underbrace{\frac{1}{6}x^3e^{-x} + \frac{1}{4}(x-1)e^x + 1}_{y_P(x)}, \quad \text{où } \lambda, \mu \in \mathbb{R}$$

Recherche de la (ou des) solution(s) avec conditions initiales :

On rappelle que la recherche de ces solutions se fait toujours **après** avoir résolu l'équation dans son intégralité... Ainsi, si y est solution de (E) avec une ou plusieurs conditions initiales, il faudra ensuite résoudre le système associé.

■ Exemple 10 :

Résoudre l'équation

$$(E) : y'' + 2y' + y = x(e^x + e^{-x}) + 1, \quad \text{sachant que } y(0) = 1$$

L'ensemble des solutions de (E) sont les fonctions obtenues dans l'exemple précédent :

$$y : x \in \mathbb{R} \mapsto (\lambda + \mu x)e^{-x} + \underbrace{\frac{1}{6}x^3e^{-x} + \frac{1}{4}(x-1)e^x + 1}_{y_P(x)}, \quad \text{où } \lambda, \mu \in \mathbb{R}$$

Or, pour y de la forme ci-dessus :

$$\begin{aligned} y(0) = 1 &\Leftrightarrow (\lambda + \mu 0)e^{-0} + 0e^{-0} + \frac{1}{4}(0-1)e^0 + 1 = 1 \\ &\Leftrightarrow \lambda - \frac{1}{4} + 1 = 1 \\ &\Leftrightarrow \lambda = -\frac{1}{4} \end{aligned}$$

En conclusion, les solutions de (E) avec la condition initiale demandée sont :

$$y : x \in \mathbb{R} \mapsto \left(\mu x - \frac{1}{4}\right)e^{-x} + \frac{1}{6}x^3 e^{-x} + \frac{1}{4}(x-1)e^x + 1, \quad \text{où } \mu \in \mathbb{R}$$



Remarque :

Si on veut fixer λ et μ , il faut au moins 2 conditions initiales dans le cadre des équations d'ordre 2... (à méditer)...

Avec une seule condition initiale, il se peut que l'on n'obtienne pas une valeur directe de λ ou μ , mais plutôt une condition reliant λ et μ .

■ Exemple 11 :

Résoudre l'équation précédente avec cette fois-ci la condition initiale $y(1) = 1$ et non plus $y(0) = 1$.

On reprend les notations précédentes. Alors

$$\begin{aligned} y(1) = 0 &\Leftrightarrow (\lambda + \mu)e^{-1} + \frac{1}{6}e^{-1} + 0e^1 + 1 = 1 \\ &\Leftrightarrow \left(\lambda + \mu + \frac{1}{6}\right)e^{-1} = 0 \\ &\Leftrightarrow \mu = -\lambda - \frac{1}{6} \end{aligned}$$

En conclusion, les solutions de (E) avec la condition initiale demandée sont :

$$y : x \in \mathbb{R} \mapsto \left(\lambda - \left(\lambda + \frac{1}{6}\right)x\right)e^{-x} + \frac{1}{6}x^3 e^{-x} + \frac{1}{4}(x-1)e^x + 1, \quad \text{où } \lambda \in \mathbb{R}$$

I Généralités

I-1 Structure d'un espace vectoriel

Dans tout ce paragraphe, on se donne $\mathbb{K} = \mathbb{R}$ ou \mathbb{C} .

Définition
 Étant donné un ensemble E , on appelle

- Loi interne, toute application $E \times E \rightarrow E$.
 On notera $u \boxplus v$ l'image de (u, v) par cette application et l'espace ainsi défini sera noté (E, \boxplus) .
- Loi externe, toute application de $\mathbb{K} \times E$ dans E .
 On note $\lambda \cdot u$ l'image de (λ, u) par cette application. L'espace ainsi défini sera noté (E, \cdot) .

Exemples :

1 ■ Pour $E = \mathbb{K}^n$, les loi classiques sont

Loi interne : $E \times E \rightarrow E$ et loi externe : $\mathbb{K} \times E \rightarrow E$
 $(u, v) \mapsto u + v$ $(\lambda, u) \mapsto \lambda u$

2 ■ Avec $E =]0; +\infty[$:

	Oui	Non
Loi interne	$E \times E \rightarrow \mathbb{R}$ $(x, y) \mapsto \sqrt{xy}$ $(x \boxplus y = \sqrt{xy})$	$E \times E \rightarrow \mathbb{R}$ $(x, y) \mapsto \ln(\frac{1}{2} + xy)$
Loi externe	$\mathbb{R} \times E \rightarrow \mathbb{R}$ $(\lambda, x) \mapsto \lambda^2 x$ $(\lambda \cdot x = \lambda^2 x)$	$\mathbb{R} \times E \rightarrow \mathbb{R}$ $(\lambda, x) \mapsto \lambda x$

Définition
 On appelle \mathbb{K} -*espace vectoriel* (ou espace vectoriel sur \mathbb{K}) tout ensemble E muni d'une loi \boxplus , et d'une loi \cdot telles que :

- (E, \boxplus) est un groupe commutatif :
 - loi interne : $u \boxplus v \in E \quad \forall u, v \in E$.
 - Commutativité : $u \boxplus v = v \boxplus u \quad \forall u, v \in E$.
 - élément neutre : Il existe $0_E \in E$ tel que $0_E \boxplus u = u \boxplus 0_E = u \quad \forall u \in E$ (donc E est **non vide** !)
 - symétrique : Pour tout $u \in E$, il existe $v \in E$ tel que $u \boxplus v = 0_E$
 - associativité : Pour tous $u, v, w \in E$, $u \boxplus (v \boxplus w) = (u \boxplus v) \boxplus w$
- Distributivité / associativité de \cdot :
 - loi externe : $\forall u \in E, \alpha \in \mathbb{K}$ on a $\alpha \cdot u \in E$.
 - distributivité de \cdot par rapport à \boxplus : $\forall \alpha \in \mathbb{K}, u, v \in E$, on a $\alpha \cdot (u \boxplus v) = \alpha \cdot u \boxplus \alpha \cdot v$.
 - distributivité de \boxplus par rapport à \cdot : $\forall \alpha, \beta \in \mathbb{K}, u \in E$, on a $(\alpha + \beta) \cdot u = \alpha \cdot u \boxplus \beta \cdot u$.
 - Associativité : pour tous $\alpha, \beta \in \mathbb{K}, u \in E$, on a $(\alpha\beta) \cdot u = \alpha \cdot (\beta \cdot u)$.
 - Élément neutre : Pour tout $u \in E$, on a $1 \cdot u = u$.

On note (E, \boxplus, \cdot) cet espace.

Contre-Exemple(s) :

- \mathbb{N}, \mathbb{Q} munis des additions et loi externe habituelle (multiplication), ne sont pas des espaces vectoriels sur \mathbb{R} , (ni sur \mathbb{C} !) (La loi n'est pas externe.)
- Considérons $E =]0, +\infty[$ muni de la loi \boxplus définie par $x \boxplus y = \sqrt{xy} \quad \forall x, y \in E$.
 - Loi interne : On a bien $x, y \in E \Rightarrow x \boxplus y > 0$, d'où $x \boxplus y \in E$ (déjà vu)
 - Élément neutre : $x \boxplus y = x \Leftrightarrow \sqrt{xy} = x \Leftrightarrow \sqrt{x}(y - \sqrt{x}) = 0 \Leftrightarrow \sqrt{x} = 0$ ou $y = \frac{1}{\sqrt{x}}$: pas d'élément y commun à tous les x . Il n'y a donc pas d'élément neutre. (E, \boxplus) ne peut donc pas donner lieu à un espace vectoriel.

Exemple 3 :

À voir sur la feuille d'exercices

Définition
 Les éléments d'un espace vectoriel E sont appelées des *vecteurs*.
 Les éléments de \mathbb{K} sur lequel est construit E sont appelés des *scalaires*

Théorème 58

Les ensembles suivants sont des espaces vectoriels sur \mathbb{K} :

- $(\mathbb{K}^n, +, \cdot)$
- $(\mathbb{K}[X], +, \cdot)$
- $(\mathcal{M}_{n,p}(\mathbb{K}), +, \cdot)$
- $(\mathcal{F}(I, E), +, \cdot)$ avec (E, \boxplus, \bullet) un \mathbb{K} -e.v., I un intervalle et

$$f + g : I \rightarrow \mathbb{K} \quad \text{et} \quad \lambda \cdot f : I \rightarrow \mathbb{K}$$

$$x \mapsto f(x) \boxplus g(x) \quad \quad \quad x \mapsto \lambda \bullet (f(x))$$



Notation :

On écrira souvent "ev" ou "e.v." pour "espace vectoriel" et \mathbb{K} -ev" pour "espace vectoriel sur \mathbb{K} ".
Plus généralement, quand il n'y aura pas de doute possible, on pourra dire "espace vectoriel" au lieu de " \mathbb{K} -espace vectoriel".

Propriété 59

Si (E, \boxplus, \cdot) est un espace vectoriel, alors :

- i ■ l'élément neutre est unique.
- ii ■ l'élément symétrique est unique.
- iii ■ pour tout $u \in E$, on a $0_E = 0 \cdot u$.
- iv ■ pour tout $u \in E$, le vecteur $(-1) \cdot u$ est le symétrique de u .

Démonstration :

- Montrons que l'élément neutre est unique :

Soient 0_E et $0'_E$ deux éléments neutres dans E . Alors

$$0_E \underbrace{=}_{0'_E \text{ est neutre}} 0_E + 0'_E \underbrace{=}_{0_E \text{ est neutre}} 0'_E$$

- Montrons que l'élément symétrique est unique :

Soit $u \in E$. Supposons qu'il ait deux symétriques v et w , alors

$$v = v + 0_E = v + (u + w) = (v + u) + w = 0_E + w = w$$

- Montrons que pour tout $u \in E$, on a $0_E = 0 \cdot u$:

Soit $u \in E$. On a

$$0 \cdot u = (0 + 0) \cdot u = 0 \cdot u + 0 \cdot u$$

En ajoutant de chaque côté le symétrique de $0 \cdot u$, on obtient $0 \cdot u = 0_E$.

- Montrons que pour tout $u \in E$, le vecteur $(-1) \cdot u$ est le symétrique de u :

Soit $u \in E$. On a

$$\square \quad 1 \cdot u + (-1) \cdot u = 0 \cdot u = 0_E$$



Notation :

À partir de maintenant, on adoptera la notation $+$ en remplacement de \boxplus , sauf s'il peut y avoir confusion.



Notation :

La propriété préc. (iv) légitime la notation $-u$ pour désigner le symétrique de u . On écrira donc " $u - v$ " à la place de " $u + \text{symétrique de } v$ ".

I-2

Sous-espace vectoriel



Définition

Soit $(E, +, \cdot)$ un \mathbb{K} -espace vectoriel. On appelle *sous espace vectoriel* ("sev") de E tout sous ensemble F non vide de E tel que $(F, +, \cdot)$ soit un \mathbb{K} -espace vectoriel.

- **Exemple 4 :**

| Dans \mathbb{R}^3 , on a vu en première année que $\text{Vect}((1, 0, 0)) \subset \mathbb{R}^3$ était un sev de \mathbb{R}^3 .

Théorème 60 (Caractérisation des sev 1)

Soient $(E, +, \cdot)$ un \mathbb{K} -espace vectoriel. F est un sous espace vectoriel de E si et seulement si les conditions suivantes sont vérifiées :

- i ■ F est non vide
- ii ■ $F \subset E$.
- iii ■ $\forall u, v \in F$, on a $u + v \in F$ (stabilité de la loi interne)
- iv ■ $\forall \lambda \in \mathbb{K}, u \in F$, on a $\lambda u \in F$. (stabilité de la loi externe).

Démonstration :

- Si F est un sev de E :

- $F \subset E$ par définition de F .
- $(F, +, \cdot)$ est en particulier un espace vectoriel, donc iii et iv sont vérifiées.
- F est non vide par définition.

- Supposons que les conditions soient vérifiées, montrons que $(F, +, \cdot)$ est un \mathbb{K} -ev.

Loi $+$:

- La loi est bien interne par hypothèse.
- associativité : héritée de E .
- commutativité : héritée de E .
- Existence de l'élément neutre dans F : 0_E est neutre pour F , car $0_E \in F$ et si $u \in F$, alors $u \in E$ et $0_E + u = u$
- Symétrique : Soit $u \in F$. Comme E est un ev, un symétrique de u est $(-1) \cdot u \in F$ par stabilité de la loi externe.

Loi \cdot :

- Distributivité /associativité à gauche : héritée de E .
- Element neutre 1 valable sur E donc valable sur F .

□

Les conditions *ii* et *iv* avec $\lambda = 0$ dans la propriété précédente donnent en particulier :

Corollaire

Tout sev F d'un espace vectoriel E contient 0_E . De plus, $0_E = 0_F$.

Démonstration :

F étant non vide, Il existe $u \in F$. On sait que $0_E \underset{\text{ppé précédente}}{=} 0 \cdot u \underset{\text{par iv}}{\in F}$. \square



Remarque :

La condition *i*) peut être simplement remplacée par " $0_E \in F$." En effet, dans ce cas, F est bien non vide! D'où la méthode suivante :

Méthode (Pour déterminer si F est un sev :)

En pratique, pour vérifier si un sous ensemble F d'un espace vectoriel E est un sev, on pourra donc commencer par vérifier que $F \subset E$ puis $0_E \in F$.

* Si $0_E \notin F$, alors F n'est pas un sev.

* Si $0_E \in F$, alors on vérifie les conditions de stabilité.

On obtient donc un premier exemple de sous-espace vectoriel, à connaître :

Proposition 61

Pour tout $n \in \mathbb{N}$, $\mathbb{K}_n[X]$ est un espace vectoriel sur \mathbb{K} (et \mathbb{R}).

Démonstration : exercice. \square

Théorème 62 (Caractérisation des sev 2)

Soient $(E, +, \cdot)$ un espace vectoriel. F est un sous espace vectoriel de E si et seulement si les conditions suivantes sont vérifiées :

i ■ $F \subset E$

ii ■ $0_E \in F$ (ou alors " F non vide")

iii ■ $\forall u, v \in F, \lambda \in \mathbb{K}, \lambda u + v \in F$

Démonstration :

Par le théorème de caractérisation 1, il suffit de vérifier que la condition *iii*) est équivalente au svstème des deux conditions $\begin{cases} 1- \forall u, v \in F, \text{ on a } u + v \in F \\ 2- \forall \lambda \in \mathbb{K}, u \in F, \text{ on a } \lambda u \in F. \end{cases}$

• Supposons que les conditions du TCSEV1 sont vérifiées, alors $\forall u, v \in E, \lambda \in \mathbb{K}$:

$$\underbrace{\underbrace{\lambda u}_{\in F \text{ par iv}} + v}_{\in F \text{ par iii}}$$

- Supposons la condition vérifiée : $\forall u, v \in E, \lambda \in \mathbb{K}, \lambda u + v \in E$. Montrons que la condition 1- est vérifiée : Il suffit de prendre $\lambda = 1$
- Montrons que la condition 2- est vérifiée : Il suffit de prendre $v = 0_E$. \square

? Exercice 1

1. Dans \mathbb{R}^2 , toute droite passant par l'origine est un \mathbb{R} -ev.
2. Dans un \mathbb{K} -espace vectoriel E , tout ensemble $\{\alpha u \mid \alpha \in \mathbb{K}\}$ où $u \in E$ fixé est un sev de E . (ex dans \mathbb{R}^2 .)

Solution

1. D est une droite de \mathbb{R}^2 passant par l'origine ssi il existe $a, b \in \mathbb{R}$ tels que $D = \{(x, y) \mid ax + by = 0\}$. Ainsi, :
 - $F \subset \mathbb{R}^2$ qui est un \mathbb{R} -ev.
 - $\vec{0} \in \mathbb{R}^2$ car $a \times 0 + b \times 0 = 0$.
 - Pour tout $u = (x, y), v = (x', y') \in D, \lambda \in \mathbb{R}$, on a $\lambda u + v = (\lambda x + x', \lambda y + y')$, avec

$$a(\lambda x + x') + b(\lambda y + y') = \lambda(ax + by) + (ax' + by') = \lambda \times 0 + 0 = 0$$

D'où, par TCSEV, D est un sev de \mathbb{R}^2 .

2. On note D un tel ensemble. On a
 - $D \subset \mathbb{R}^2$ qui est un \mathbb{R} -ev.
 - $\vec{0} = 0u \in D$.
 - Soient $x = \alpha u, y = \beta v \in D$ et $\lambda \in \mathbb{R}$. Alors

$$x + \lambda y = (\alpha + \lambda\beta)u \in D.$$

D'où, par TCSEV, D est un sev de \mathbb{R}^2 .

Corollaire de la caractérisation 2 :

pour tout intervalle I , l'ensemble $\mathcal{C}^n(I, \mathbb{R})$ est un \mathbb{R} -ev, où $n \in \mathbb{N} \cup \{\infty\}$.

Démonstration :

On a $\mathcal{C}^n(I, \mathbb{R}) \subset \mathcal{F}(I, \mathbb{R})$ qui est un \mathbb{R} -eV d'élément neutre f_0 la fonction nulle.

* $f_0 \in \mathcal{C}^n(I, \mathbb{R})$

* Soient $f, g \in \mathcal{C}^n(I, \mathbb{R}), \lambda \in \mathbb{R}$. Alors, par résultat de cours, on sait que $f + \lambda g$ est également dans $\mathcal{C}^n(I, \mathbb{R})$. Ainsi, $\mathcal{C}^n(I, \mathbb{R})$ est bien stable par C-L.

* En conclusion, par TCSEV, $\mathcal{C}^n(I, \mathbb{R})$ est bien un sev de $\mathcal{F}(I, \mathbb{R})$, et c'est donc bien un \mathbb{R} -eV.

\square

Proposition 63

Toute intersection de sous espaces vectoriels d'un même \mathbb{K} -ev est un \mathbb{K} -ev.

Démonstration :

Soit $(F_i)_{i \in I}$ une famille quelconque de \mathbb{K} -sous-espaces vectoriels d'un \mathbb{K} -espace vectoriel E . Montrons que $F = \bigcap_{i \in I} F_i$ est un \mathbb{K} -sev de E . Utilisons la caractérisation des sev.

- F est non vide : $0_E \in F_i \quad \forall i \in I$, d'où $0_E \in F$.
- Soient $\lambda \in \mathbb{K}$ et $x, y \in F$. Ainsi, comme tous les F_i sont des sev, par caractérisation des sev, on a

$$\lambda u + y \in F_i \quad \forall i \in I$$

autrement dit

$$\lambda u + y \in F$$

et donc, toujours par caractérisation des sev, on en déduit que F est bien un sev de E . \square



Remarque :

Une intersection de sous espaces vectoriels d'un \mathbb{K} -ev E n'est jamais vide, elle contient au moins toujours $\{0_E\}$.

II Famille génératrice, famille libre et base

(rappel : on considère toujours $\mathbb{K} = \mathbb{R}$ ou \mathbb{C} .) On dira également "espace vectoriel" pour " \mathbb{K} -espace vectoriel" tout le long du reste du chapitre.

II-1 S.e.v. engendré (Vect) et famille génératrice

Définition

Soit E un \mathbb{K} -ev et $a_1, \dots, a_r \in E$. On dit que $v \in E$ est *combinaison linéaire de* a_1, \dots, a_r s'il existe $\lambda_1, \dots, \lambda_r \in \mathbb{K}$ tels que

$$v = \lambda_1 a_1 + \dots + \lambda_r a_r$$

■ **Exemple 5 :**

Pour $E = \mathbb{C}^3$ et $a_1 = \begin{pmatrix} 1 \\ i \\ -1 \end{pmatrix}$, $a_2 = \begin{pmatrix} -i \\ 1 \\ -2 \end{pmatrix}$ alors $v = \begin{pmatrix} 2+i \\ 2i-1 \\ 0 \end{pmatrix}$ est combinaison linéaire de a_1, a_2 car $v = 2a_1 - a_2$.



Définition

Soit E un \mathbb{K} -ev et $A = \{a_1, \dots, a_r\} \subset E$ une famille finie de E . On note $\text{Vect}(A)$ l'ensemble de toutes les combinaisons linéaires possibles entre les éléments de A :

$$\text{Vect}(A) = \{\alpha_1 a_1 + \dots + \alpha_r a_r \mid \alpha_i \in \mathbb{K} \forall i = 1, \dots, r\}.$$

■ **Exemples :**

- 6 ■ Dans \mathbb{C}^3 , on pose $A = \{(1, 0, 0), (0, 1, 0)\}$. $\text{Vect}(A) = \{(x, y, 0) \mid x, y \in \mathbb{C}\}$.
- 7 ■ Dans $\mathbb{R}[X]$, on pose $A = \{1, X\}$. $\text{Vect}(A) = \mathbb{R}_1[X]$.

Ci-dessous une propriété immédiate de contenu très utile !

Propriété 64

Dans les conditions de la définition précédente, on a :

- i ■ $A \subset \text{Vect}(A)$ (i.e. $\text{Vect}(A)$ contient tous les éléments de A) ;
- ii ■ $\text{Vect}(A)$ est un \mathbb{K} sous-espace vectoriel de E ;
- iii ■ $\text{Vect}(A)$ est "le plus petit \mathbb{K} -sous espace vectoriel de E contenant A ", au sens où :

si $A \subset F$ où F est un \mathbb{K} -sev de E , alors $\text{Vect}(A) \subset F$

Démonstration :

- i) $a_i = \alpha_1 a_1 + \dots + \alpha_r a_r \in \text{Vect}(A)$ avec $\alpha_j = 0$ pour tout j sauf $j = i$ où $\alpha_i = 1$.
- ii) exercice
- iii) Soit $u \in \text{Vect}(A)$. Alors il existe $\alpha_1, \alpha_r \in \mathbb{K}$ tels que $u = \alpha_1 a_1 + \dots + \alpha_r a_r$. Or, F est un \mathbb{K} -ev. Ainsi, comme $a_1, \dots, a_r \in F$, et que $\alpha_1, \alpha_r \in \mathbb{K}$, on a de fait $\alpha_1 a_1 + \dots + \alpha_r a_r \in F$. D'où $u \in F$ et donc au final

$$\square \quad \text{Vect}(A) \subset F$$

■ **Exemple 8 :**

Dans \mathbb{R}^3 , on pose $A = \{(1, 1, 0), (2, 0, 0)\}$.

$$\text{Vect}(A) = \{(x, y, 0) \mid x, y \in \mathbb{R}\} :$$

en effet, si on pose $F = \{(x, y, 0) \mid x, y \in \mathbb{R}\}$, on a clairement $A \subset F$, donc $\text{Vect}(A) \subset F$. On connaît déjà la notion de dimension dans \mathbb{R}^3 , on peut alors justifier ensuite que $\dim A = \dim F$, donc on a bien $\text{Vect}(A) = F$.

■ Exemple 9 :

Dans $\mathbb{R}[X]$, on pose $A = \{X, X - 1\}$.

$$\text{Vect}(A) = \mathbb{R}_1[X] :$$

De même que tout à l'heure, on a clairement $A \subset F$ et donc $\text{Vect}(A) \subset F$. Des arguments de dimension pourront également plus tard justifier l'égalité. En revanche, si on veut prouver ceci maintenant, il faut justifier différemment, en montrant que $F \subset A$: Ainsi, si $P \in \mathbb{R}_1[X]$, il existe $a, b \in \mathbb{R}$ tels que

$$P = a + bX = -a(X - 1) + aX + bX = -a(X - 1) + (a + b)X \in \text{Vect}(A)$$

c'est fini!

De la propriété de tout à l'heure, on tire maintenant le vocabulaire suivant :

🌿 Définition

Soit E un \mathbb{K} -ev et $A = \{a_1, \dots, a_r\} \subset E$ une famille finie de E . On appelle $\text{Vect}(A)$ le \mathbb{K} sous-espace vectoriel de E engendré par A et on dit que A est **une famille génératrice** de F .

■ Exemples :

10 ■ Dans \mathbb{R}^3 , une famille génératrice de $F = \{(x, y, z) | x - y + 2z = 0\}$ est

$$(1, 1, 0), (-2, 0, 1) :$$

$$(x, y, z) \in F \Leftrightarrow x - y + 2z = 0 \Leftrightarrow x = y - 2z \Leftrightarrow (x, y, z) = y(1, 1, 0) + z(-2, 0, 1)$$

11 ■ La famille $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ est génératrice de \mathbb{R}^3 , car

$$\begin{aligned} \text{Vect}((1, 0, 0), (0, 1, 0), (0, 0, 1)) &= \{x(1, 0, 0) + y(0, 1, 0) + z(0, 0, 1) | x, y, z \in \mathbb{R}\} \\ &= \{(x, y, z) | x, y, z \in \mathbb{R}\} = \mathbb{R}^3 \end{aligned}$$

12 ■ La famille $\{1, X, X^2\}$ est de même génératrice de $\mathbb{R}_2[X]$.

13 ■ Une famille génératrice de $\{P \in \mathbb{C}_2[X] | P(0) = 0\}$ est X, X^2 car

$$\{P \in \mathbb{C}_2[X] | P(0) = 0\} = \{X(aX + b) | a, b \in \mathbb{C}\} = \{aX^2 + bX | a, b \in \mathbb{C}\} = \text{Vect}(X, X^2)$$

■ Contre-Exemple(s) :

Il n'y a pas de famille génératrice finie pour $\mathbb{K}[X]$.

En effet, (par l'absurde) : supposons qu'il y en ait une, notée ici \mathcal{F} .

Soit $N = \max\{\deg P | P \in \mathcal{F}\}$. Alors $\mathcal{F} \subset \mathbb{K}_N[X]$ et donc $\mathbb{K}[X] = \text{Vect}(\mathcal{F}) \subset \mathbb{K}_N[X]$, ce qui est absurde.

II-2 Famille libre

🌿 Définition

• Soit $(x_i)_{1 \leq i \leq n}$ une famille finie de vecteurs d'un espace vectoriel E . On dit que la famille est **libre** (ou que les vecteurs sont **linéairements indépendants**) si

$$\forall \alpha_1, \dots, \alpha_n \in \mathbb{R}, \quad \alpha_1 x_1 + \dots + \alpha_n x_n = 0 \iff \alpha_1 = \dots = \alpha_n = 0$$

• Une famille $(x_i)_{1 \leq i \leq n}$ non libre est dite **liée**.

Autrement dit : Une famille est libre, si on ne peut pas exprimer les vecteurs les uns en fonction des autres. S'il existe un (ou plusieurs) vecteur(s) qui peut s'écrire en fonction des autres, alors elle est trivialement liée.

■ Exemple 14 :

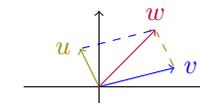
Si $A = \{a_1, \dots, a_r\}$ est une famille finie de E et que $v \in \text{Vect}(A)$, alors la famille $\{a_1, \dots, a_r, v\}$ est liée.

■ Exemple 15 :

Dans \mathbb{R}^2 , le schéma ci-contre représente trois vecteurs liés car

$$w = u + v, \text{ i.e. } u + v - w = 0.$$

En revanche, les vecteurs u, v sont libres (ainsi que les vecteurs u, w ou les vecteurs v, w . à méditer...)



⚠️ CONFUSION

Il existe toujours $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ tels que $\alpha_1 x_1 + \dots + \alpha_n x_n = 0$. (il suffit en effet de prendre tous les $\alpha_i = 0$.)

Ce n'est pas ça qu'il faut vérifier et ceci est une erreur fréquemment commise !

Donc en pratique, le sens \Leftarrow est évident. Le seul sens à vérifier est \Rightarrow , c'est-à-dire : "Soient $\alpha_1, \dots, \alpha_n \in \mathbb{R}$, tels que $\alpha_1 x_1 + \dots + \alpha_n x_n = 0$. Montrons que $\alpha_1 = \dots = \alpha_n = 0$ "

? Exercice 2

(révisions) Dans \mathbb{C}^3 , montrer que $\{(1, 3, -1), (2, 1, i), (1, 0, 0)\}$ est une famille libre.

Solution

On note $\mathcal{F} = \{u, v, w\}$ la famille en question (dans l'ordre énoncé). Soient $a, c, b \in \mathbb{C}$ tels que $au + bv + cw = 0$. On a :

$$au + bv + cw = 0 \Leftrightarrow \begin{cases} a + 2b + c = 0 \\ 3a + b = 0 \\ -a + ib = 0 \end{cases} \Leftrightarrow \dots \Leftrightarrow a = b = c = 0$$

Ainsi, la famille est libre.

On note par ailleurs que d'après le cours de première année, on peut également démontrer la liberté par l'argument suivant : " (u, v, w) est libre équivaut à montrer que le rang de la

matrice $\begin{pmatrix} 1 & 2 & 1 \\ 3 & 1 & 0 \\ -1 & i & 0 \end{pmatrix}$ vaut 3".

■ Exemple 16 :

Montrer que la famille $\{f : x \mapsto x, g : x \mapsto e^{-x}, h : x \mapsto 1 + \cos x\} \subset \mathcal{C}^\infty(]-1, +\infty[)$ est une famille libre.

Soient $\alpha, \beta, \gamma \in \mathbb{R}$ tels que

$$\alpha f + \beta g + \gamma h = 0 \quad (*)$$

On a

$$\begin{aligned} (*) &\Leftrightarrow \forall x > -1, \quad \alpha f(x) + \beta g(x) + \gamma h(x) = 0 \\ &\Leftrightarrow \forall x > -1, \quad \alpha x + \beta e^{-x} + \gamma(1 + \cos x) = 0 \end{aligned}$$

Avec $x = 0$, on obtient $\beta + \gamma = 0$. On propose ensuite de dériver l'expression :

$$\alpha - \beta e^{-x} - \gamma \sin x = 0$$

et donc, en $x = 0$ là aussi, on a $\alpha - \beta = 0$. Finalement, on propose de passer à la limite en $+\infty$ sur la dérivée et on trouve que, si $\gamma \neq 0$, on arrive à

$$\alpha - \gamma \sin x \xrightarrow[n \rightarrow +\infty]{} 0$$

ce qui est impossible car \sin diverge, sauf dans le cas où $\gamma = 0$.

Au final, on obtient le système

$$\begin{cases} \beta + \gamma = 0 \\ \alpha - \beta = 0 \end{cases}$$

Ce qui après une très rapide résolution donne la seule possibilité $\alpha = \beta = \gamma = 0$.

? Exercice 3

(révisions)

1. Montrer que dans un espace vectoriel, **deux** vecteurs sont linéairement indépendants ssi ils sont non colinéaires.
2. Montrer que dans \mathbb{K}^3 , les vecteurs $(1, 1, 1)$, $(0, 6, 5)$ et $(-2, 4, 3)$ ne sont pas linéairement indépendants.

Solution

1. Notons u, v deux vecteurs.

- Supposons qu'ils soient colinéaires. Dans ce cas, il existe $\lambda \in \mathbb{K}$ tel que $u = \lambda v$ (ou $v = \lambda u$), auquel cas on a $1 \times u + (-\lambda)v = 0$, (ou $1 \times v + (-\lambda)u = 0$). Dans les deux cas, $a = b = 0$ n'est pas la seule solution à l'équation $au + bv = 0$. Ils ne sont donc pas linéairement indépendants.
- Supposons maintenant qu'ils ne soient pas linéairement indépendants. Il existe donc a, b non tous deux nuls tels que $au + bv = 0$. Si $a \neq 0$, on a alors $u = -\frac{b}{a}v$. Si $b \neq 0$, on a $v = -\frac{a}{b}u$. Dans les deux cas, u et v sont bien colinéaires!

2. On peut tout simplement vérifier que $\text{rg} \begin{pmatrix} 1 & 0 & -2 \\ 1 & 6 & 4 \\ 1 & 5 & 3 \end{pmatrix} < 3$ ou alors montrer que le système $au + bv + cw = 0$ (où u, v, w sont les trois vecteurs considérés) admet au moins une liste de solution (a, b, c) non nulle.

? Exercice 4

Vérifier que dans $\mathcal{F}(\mathbb{R}, \mathbb{R})$, les fonctions $f : x \mapsto \cos^2(x)$, $g : x \mapsto \cos(2x)$, $h : x \mapsto 1$ ne sont pas linéairement indépendantes.

Solution

On utilise directement les formules de trigonométrie. En effet, on a

$$\forall x \in \mathbb{R}, \quad \cos(2x) = \cos^2(x) - \sin^2(x) = \cos^2(x) - (1 - \cos^2(x)) = 2 \cos^2(x) - 1$$

i.e.

$$g = 2f - h$$

La famille n'est donc pas libre.

? Exercice 5

On note $I =]1, +\infty[$. Montrer que les fonctions $f : x \mapsto \arctan \frac{1-x}{1+x}$, $g : x \mapsto \arctan x$, $h : x \mapsto 1$ dans $\mathcal{F}(I, \mathbb{R})$ ne sont pas linéairement indépendantes.

Solution

La relation ne saute pas forcément aux yeux ! Il faut aller chercher les coefficients ! Pour ce faire, on pose $\alpha, \beta, \gamma \in \mathbb{R}$ tels que

$$\alpha f + \beta g + \gamma h = 0$$

Évalué en $x = 0$ et $x = 1$, on obtient le système

$$\begin{cases} \frac{\pi}{4}\alpha + \gamma = 0 \\ \frac{\pi}{4}\beta + \gamma = 0 \end{cases}$$

qui admet une infinité de solutions $(\alpha, \beta, \gamma) \in \text{Vect}(1, 1, -\frac{\pi}{4})$. Une relation possible est donc peut-être $f + g - \frac{\pi}{4}h = 0$, mais ceci n'est pas démontré pour tout x . Il faut encore le vérifier !

L'étape suivante consiste alors à valider l'égalité pour tout $x \in \mathbb{R}$. Pour ce faire, on peut par exemple procéder de la manière suivante ici :

- On montre que la fonction est constante en montrant que la fonction $f + g - \frac{\pi}{4}h$ est de dérivée nulle (faire le calcul.)
- Il existe alors une constante c telle que $f + g - \frac{\pi}{4}h = c$. En évaluant en $x = 0$, on a $0 = c$. D'où la relation démontrée

$$f + g - \frac{\pi}{4}h = 0$$

Propriété 65

Toute famille de polynômes non nuls de degrés deux à deux distincts est libre.

Démonstration :

Montrons par récurrence sur le nombre n de polynômes de $\mathbb{K}[X]$ que la famille est libre.

- Initialisation : Soit $n = 1$. P_1 est un polynôme non nul. La famille $\{P_1\}$ est donc libre.
- Hérite : Supposons la propriété vraie pour $n \in \mathbb{N}^*$ et montrons qu'elle est vraie pour $n + 1$. Soient P_1, \dots, P_{n+1} des polynômes non nuls de degrés deux à deux distincts. Quitte à renommer, on peut supposer que

$$\deg P_1 < \dots < \deg P_{n+1}$$

Soient $\lambda_1, \dots, \lambda_{n+1} \in \mathbb{K}$ tels que

$$\lambda_1 P_1 + \dots + \lambda_{n+1} P_{n+1} = 0$$

On note N le degré de P_{n+1} . On a

$$\alpha_{n+1} P_{n+1}(X) = \underbrace{-(\lambda_1 P_1 + \dots + \lambda_n P_n)}_{\text{de degré} < N}$$

Comme $\deg P_{n+1} = N$, il est nécessaire que $\lambda_{n+1} = 0$, sinon on aboutit à une contradiction sur les degrés. D'où de plus

$$\lambda_1 P_1 + \dots + \lambda_n P_n = 0$$

Or la famille P_1, \dots, P_n est libre par hypothèse de récurrence, d'où

$$\lambda_1 = \dots = \lambda_n = \lambda_{n+1} = 0.$$

On en conclut donc la liberté de la famille P_1, \dots, P_{n+1} .

□

II-3 Base

II.3-a) Définition



Définition

Soit E un \mathbb{K} -espace et $(x_i)_{1 \leq i \leq n}$ une famille de vecteurs de E . Si elle est libre et génératrice de E , on dit que c'est une *base* de E .

Propriété 66

- Base canonique de \mathbb{K}^n : $\mathcal{B} = \{(1, 0, 0, \dots), (0, 1, 0, \dots), \dots, (\dots, 0, 0, 1)\}$ est une base du \mathbb{K} -ev \mathbb{K}^n .
- Base canonique de $\mathbb{K}_n[X]$: $\mathcal{B} = \{1, X, \dots, X^n\}$ est une base du \mathbb{K} -ev $\mathbb{K}_n[X]$.

Démonstration :

Dans chaque cas, on a trivialement que les \mathcal{B} sont des familles génératrices respectives de \mathbb{K}^n et $\mathbb{K}_n[X]$.

- Dans le cas de \mathbb{K}^n , on montre que la famille est libre avec les techniques habituelles.
- Dans le cas de $\mathbb{K}_n[X]$, on sait d'après le cours sur les polynômes que la famille est libre car il s'agit d'une famille de polynômes non nuls de degrés deux à deux distincts. □

? Exercice 6

(révisions) Montrer que $\mathcal{B} = \{(1, 3, -1), (2, 1, i), (1, 0, 0)\}$ est une base de \mathbb{C}^3 .

Solution

On a déjà montré lors d'un exercice précédent que la famille était libre. Le cours de première année nous dit alors que si une famille libre de \mathbb{C}^3 est de cardinal $\dim \mathbb{C}^3 = 3$, alors c'est une base. C'est bien le cas ici. La famille est donc bien une base de \mathbb{C}^3 . (On peut également démontrer le caractère générateur de la famille, mais c'est beaucoup plus long...)

■ Exemple 17 :

Montrons que la famille $\mathcal{F} = (1 + X, 2)$ est une base de $\mathbb{R}_1[X]$.

On a tout d'abord $\mathcal{F} \subset \mathbb{R}_1[X]$ une famille de polynômes de degrés 2 à deux distincts, donc libre.

Montrons ensuite que \mathcal{F} est une famille génératrice de $\mathbb{R}_1[X]$, i.e. $\mathbb{R}_1[X] = \text{Vect}(\mathcal{F})$. On pose $P = aX + b \in \mathbb{R}_1[X]$ un polynôme quelconque. \mathcal{F} est génératrice ss'il existe $\alpha, \beta \in \mathbb{R}$ tels que $P = \alpha(1 + X) + \beta \times 2$.
Or,

$$\begin{aligned} P = \alpha(1 + X) + \beta \times 2 &\Leftrightarrow aX + b = \alpha + 2\beta + \alpha X \\ &\Leftrightarrow \begin{cases} \alpha + 2\beta = b \\ \alpha = a \end{cases} \Leftrightarrow \begin{cases} \beta = \frac{1}{2}(b - a) \\ \alpha = a \end{cases} \end{aligned}$$

Étant donné qu'il existe bien une solution à l'équation, la famille est bien génératrice. (Plus tard, on verra de plus que le fait que la solution est unique assure également la liberté de la famille en même temps!)

En conclusion, la famille étant libre et génératrice de $\mathbb{R}_1[X]$, c'est une base de $\mathbb{R}_1[X]$.

⚠ Remarque :

Dans certains espaces, on ne peut pas trouver de famille finie vérifiant les propriétés libres et génératrices. (Par exemple, pour $\mathbb{K}[X]$ et $\mathcal{F}(\mathbb{R}; \mathbb{R})$.) (démontré pour $\mathbb{K}[X]$ et en conséquence pour $\mathcal{F}(\mathbb{R}; \mathbb{R})$.)

II.3-b) Existence d'une base : extraction d'une famille génératrice

Théorème 67

Dans un espace vectoriel E , de toute éventuelle famille génératrice A finie non vide, on peut extraire une base de $F = \text{Vect}(A)$.

Démonstration : admise. \square

Commentaires :

Ce théorème dit essentiellement que si un espace vectoriel admet au moins une famille génératrice finie, alors il admet également une base, c'est-à-dire ici une famille libre et génératrice que l'on peut extraire de la famille génératrice "de départ".

■ Exemple 18 :

Soit $A = \{u_1, u_2, u_3\} \subset \mathbb{R}^3$ tels que $u_1 = (1, 0, 2)$, $u_2 = (1, 1, 1)$ et $u_3 = (0, -1, 1)$. On considère $F = \text{Vect}(A)$. Alors A est génératrice finie de F . On peut en extraire une base de F .

On sait que la famille A n'est pas libre et on a

$$u_3 = u_1 - u_2$$

Ainsi, (u_1, u_2) est encore génératrice de F . Or, ce sont deux vecteurs non colinéaires, donc ils sont libres. On a donc la famille

$$B = \{u_1, u_2\} \subset A$$

qui est génératrice de F et libre. C'est bien une base de F extraite de A .

Commentaires :

La technique est donc de retirer des éléments un par un en faisant attention à ce que la famille soit encore génératrice à chaque fois jusqu'à l'obtention d'une famille libre. Dans certains cas, un raisonnement par dimension pourra également être utile et accélérer un peu les choses. (cf p. 60)

III Dimension, rang et recherche de base

III-1 Dimension

Théorème 68

Si un espace vectoriel admet une base de cardinal n , alors toutes les bases de E sont de cardinal n .

Démonstration :

Soient \mathcal{B} et \mathcal{B}' deux bases de E . Supposons qu'elles soient de cardinal différent. Quitte à les échanger, on peut supposer que $\text{card } \mathcal{B}' < \text{card } \mathcal{B}$. La base \mathcal{B}' est donc en particulier une famille libre dans l'espace $\text{Vect}(\mathcal{B})$. On va montrer que ceci n'est pas possible, ce qui imposera $\text{card } \mathcal{B}' = \text{card } \mathcal{B}$.

• Montrons par récurrence sur p que "toute famille de $p + 1$ éléments de $F_p = \text{Vect}(u_1, \dots, u_p)$ est forcément liée dans E ".

* Initialisation : Si $p = 1$:

Soit $G = \{g_1, g_2\}$ une famille de $F_p = \text{Vect}(u)$. Alors il existe $\alpha, \beta \in \mathbb{K}$ tels que $g_1 = \alpha u$ et $g_2 = \beta u$.

Si $\alpha = \beta = 0$, alors $g_1 = g_2 = 0$, la famille G est donc liée.

Sinon, l'un des deux au moins est non nul et de plus on a clairement

$$\alpha g_2 - \beta g_1 = \alpha\beta(u - u) = 0$$

La famille G est donc bien liée.

* Hérité : Supposons que la propriété est vraie pour $p \in \mathbb{N}$:

Soit $G = \{g_1, \dots, g_{p+2}\}$ une famille de $F_{p+1} = \text{Vect}(u_1, \dots, u_{p+1})$. Il existe alors $\lambda_1, \dots, \lambda_{p+2} \in \mathbb{R}$, $v_1, \dots, v_{p+1} \in F_p = \text{Vect}(u_1, \dots, u_p)$ tels que

$$\begin{aligned} g_1 &= \lambda_1 u_{p+1} + v_1 \\ &\dots \\ g_{p+1} &= \lambda_{p+1} u_{p+1} + v_{p+1} \\ g_{p+2} &= \lambda_{p+2} u_{p+1} + v_{p+2} \end{aligned}$$

Si $\lambda_1 = \dots = \lambda_{p+1} = \lambda_{p+2} = 0$, alors G est une famille de F_p . En conséquence, par hypothèse de récurrence, $\{g_1, \dots, g_{p+1}\}$ est liée, et donc, la famille $G = \{g_1, \dots, g_{p+2}\}$ est liée.

Sinon, il existe un λ non nul. Quitte à renommer la famille, on peut supposer que c'est λ_{p+2} . En réécrivant la dernière ligne du système ci-dessus, on peut donc exprimer u_{p+1} en fonction de g_{p+2} et de F_p de la manière suivante :

$$u_{p+1} = \frac{1}{\lambda_{p+2}} g_{p+2} - v_{p+2}$$

On réinjecte dans le système :

$$\begin{aligned} g_1 &= \lambda_1 \frac{1}{\lambda_{p+2}} g_{p+2} - v_{p+2} + v_1 \\ &\dots \\ g_{p+1} &= \lambda_{p+1} \frac{1}{\lambda_{p+2}} g_{p+2} - v_{p+2} + v_{p+1} \\ u_{p+1} &= \frac{1}{\lambda_{p+2}} g_{p+2} - v_{p+2} \end{aligned}$$

On en déduit que $g_1 - \frac{\lambda_1}{\lambda_{p+2}} g_{p+2}, \dots, g_{p+1} - \frac{\lambda_{p+1}}{\lambda_{p+2}} g_{p+2} \in F_p$. Par hypothèse de récurrence, ces éléments sont liés, c'est-à-dire qu'il existe $\alpha_1, \dots, \alpha_{p+1}$ non tous nuls tels que

$$\alpha_1 \left(g_1 - \frac{\lambda_1}{\lambda_{p+2}} g_{p+2} \right) + \dots + \alpha_{p+1} \left(g_{p+1} - \frac{\lambda_{p+1}}{\lambda_{p+2}} g_{p+2} \right) = 0$$

Autrement dit,

$$\alpha_1 g_1 + \dots + \alpha_{p+1} g_{p+1} + \left(-\alpha_1 \frac{\lambda_1}{\lambda_{p+2}} - \alpha_{p+1} \frac{\lambda_{p+1}}{\lambda_{p+2}} \right) g_{p+2} = 0$$

La famille G est donc bien liée.

• On note $n = \text{card } \mathcal{B}$. On vient donc de montrer que toute famille de cardinal $n + 1$ est liée. La famille \mathcal{B}' est de cardinal supérieur ou égal à $n + 1$. Elle contient donc une famille de cardinal $n + 1$ qui sera liée. En conclusion, \mathcal{B}' est liée, ce qui est contradictoire. \square



Définition

Un espace vectoriel E est dit de *dimension* n s'il admet une base de cardinal n . On note $\dim E = n$.

On dit qu'un espace vectoriel est de *dimension finie* s'il existe un entier $n \in \mathbb{N}$ tel que $\dim E = n$.

Exemples :

19 ■ \mathbb{K}^n est de dimension n , une base est la base canonique $(1, 0, \dots, 0), (0, 1, 0, \dots, 0), \dots, (0, \dots, 0, 1)$

20 ■ $\mathbb{K}_n[X]$ est de dimension $n + 1$. Une base est la base canonique $1, X, \dots, X^n$.

Corollaire

Soit E un espace vectoriel de dimension n . Alors :

- i ■ Toute famille libre admet au plus n éléments.
- ii ■ Toute famille génératrice admet au moins n éléments.
- iii ■ Une famille libre de n éléments est une base.
- iv ■ Une famille génératrice de n éléments est une base.

Démonstration :

i) : C'est ce qui a été fait dans la démonstration du théorème précédent.

ii) : S'il existe une famille génératrice avec $r < n$ éléments, alors on peut en extraire une base de cardinal $s \leq r$ et donc $s < n$, ce qui est impossible.

iii) : On note \mathcal{F} une famille libre à n éléments non génératrice. Alors il existe $v \in E$ tel que $v \notin \text{Vect}(\mathcal{F})$. Ainsi, la famille $\mathcal{F} \cup \{v\}$ est une famille libre contenant $n + 1$ éléments, ce qui contredit i.)

iv) On note \mathcal{F} une famille génératrice à n éléments. Si elle n'est pas libre, alors on peut retirer un élément v de manière à ce qu'elle reste encore génératrice. On peut ensuite extraire une base de la famille $\mathcal{F} \setminus \{v\}$ qui sera donc de cardinal $\leq n - 1 < n$, ce qui est contradictoire. \square

Dit autrement :

- i ■ Si E admet une famille libre de cardinal p , alors $\dim E \geq p$.
- ii ■ Si E admet une famille génératrice de cardinal p , alors $\dim E \leq p$.
- iii ■ Si E admet une famille libre de cardinal p qui n'est pas génératrice, alors $\dim E > p$.
- iv ■ Si E admet une famille génératrice de cardinal p qui n'est pas libre, alors $\dim E < p$.

Voici un exemple d'utilisation de ces propriétés dans le cadre de la recherche d'une base par extraction d'une famille libre déjà vu autrement :

■ Exemple 21 :

Soit $E = \mathbb{R}_3[X]$. Alors la famille $\mathcal{F} = \{2, 3 + X, X^2 - 1, X^3 + 2X - 4\}$ est une base de E :

En effet, la famille étant une famille de polynômes non nuls de degrés 2 à 2 distincts, elle est donc libre.

De plus, $\text{card } \mathcal{F} = 4 = \dim \mathbb{R}_3[X]$.

Ainsi, les deux arguments réunis font que c'est bien une base de $\mathbb{R}_3[X]$.

Ces résultats peuvent nous servir également à une justification lors de l'extraction d'une base. Par exemple :

■ Exemple 22 :

Soit $A = \{u_1, u_2, u_3\} \subset \mathbb{R}^3$ tels que $u_1 = (1, 0, 2)$, $u_2 = (1, 1, 1)$ et $u_3 = (0, -1, 1)$. On considère $F = \text{Vect}(A)$. Alors A est génératrice finie de F par définition. On peut en extraire une base de F .
 A contient deux vecteurs non colinéaires u_1 et u_2 . Ainsi, (d'après *i*)

$$\dim F \geq 2$$

De plus, on peut montrer que A n'est pas libre (mais génératrice), donc (Ppé *iv*)

$$\dim \text{Vect}(A) < \text{card}(A) = 3$$

Ainsi :

$$\dim \text{Vect}(A) = 2$$

Or, $B = \{u_1, u_2\}$ est une famille libre de cardinal $2 = \dim \text{Vect}(A)$, c'est donc une base de F extraite de A . (*iii*)

Proposition 69

Soit F un sous espace vectoriel d'un ev E de dimension finie n . Alors, F est de dimension finie et $\dim F \leq \dim E$.

Démonstration :

Soit \mathcal{B} une base de F . Alors $\mathcal{B} \subset E$ et c'est donc une famille libre de E . Ainsi, $\dim E \geq \text{card}(\mathcal{B}) = \dim F$. □

Corollaire

Si F et G sont deux sous-espaces vectoriels d'un espace vectoriel E , alors $F = G \iff (F \subset G \text{ et } \dim F \geq \dim G)$

■ Exemple 23 :

Dans $\mathbb{R}[X]$, on pose $A = \{X, X - 1\}$. $\text{Vect}(A) = \mathbb{R}_1[X]$:

On a trivialement $A \subset F \Rightarrow \text{Vect}(A) \subset F$. Maintenant, comme $X, X - 1$ sont non colinéaires, ils sont libres et donc

$$\dim A \geq 2 = \dim F, \quad \text{Ainsi, } F = A.$$

III-2 Rang d'une famille

Définition

On appelle *rang* d'une famille A la dimension de l'espace vectoriel qu'elle engendre. On note $rg(A)$.

■ Exemples :

- 24 ■ $rg\{1, 2X + 1\} = 2$
- 25 ■ $rg\{(1, 1, 0), (3, 3, 0)\} = 1$

 Remarque :

Les propriétés de la partie précédente nous permettent de dire que :

Propriété 70

- i ■ $rg(a_1, \dots, a_p) \leq p$.
- ii ■ $rg(a_1, \dots, a_p) < p$ ssi la famille $\{a_1, \dots, a_p\}$ n'est pas libre.
- iii ■ $rg(a_1, \dots, a_p) = p$ ssi la famille $\{a_1, \dots, a_p\}$ est libre.
- iv ■ $rg(a_1, \dots, a_p) = p$ ssi la famille $\{a_1, \dots, a_p\}$ est une base.

Démonstration :

i) : issue du *ii*) du Corollaire p.59

ii) : Si la famille $A = \{a_1, \dots, a_p\}$ est libre, alors elle est libre et génératrice de $\text{Vect}(A)$, donc

$$rg(A) = \dim \text{Vect}(A) = p$$

Ainsi, on obtient

$$rg(a_1, \dots, a_p) < p \implies \text{la famille } \{a_1, \dots, a_p\} \text{ n'est pas libre.}$$

Si la famille $A = \{a_1, \dots, a_p\}$ n'est pas libre, alors c'est la propriété *iv*) du Corollaire p.59 qui nous dit que

$$rg(a_1, \dots, a_p) = \dim \text{Vect}(A) < p$$

iii) : conséquence directe de *i*) et *ii*)

iv) : La famille (a_1, \dots, a_p) est nécessairement génératrice de $\text{Vect}(A)$. C'est donc une conséquence immédiate de *iii*). □

III-3 Théorème de la base incomplète

Théorème 71 de la base incomplète

Supposons que E est un \mathbb{K} -ev dans il existe une famille génératrice finie. Alors, toute famille libre de E peut être complétée en une base de E .

Démonstration : admise. \square

■ Exemple 26 :

Dans \mathbb{R}^3 , on veut compléter $\mathcal{L} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \right\}$ en une base \mathbb{R}^3 .

On connaît une famille génératrice de \mathbb{R}^3 qui est la base canonique :

$$\mathcal{G} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$$

ainsi que la dimension de \mathbb{R}^3 : $\dim \mathbb{R}^3 = 3$, ce qui va nous simplifier la tâche !

On va donc compléter la famille \mathcal{L} par des éléments de \mathcal{G} en veillant à garder une famille libre jusqu'à obtenir une famille génératrice. Ici, connaissant la dimension, on va donc compléter jusqu'à obtention d'une famille libre de cardinal 3.

Par exemple, on constate que

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$$

est une famille libre de cardinal 3 dans \mathbb{R}^3 , c'est donc une base de \mathbb{R}^3 .

■ Exemple 27 :

On souhaite compléter la famille $\mathcal{L} = (3, 1 + X, X + X^2)$ en une base de $\mathbb{R}_4[X]$.

Pour commencer, \mathcal{L} est libre par famille de polynômes non nuls de degrés 2 à 2 distincts. L'espace vectoriel $E = \mathbb{R}_4[X]$ de dimension 5 admet comme famille génératrice $\mathcal{G} = (1, X, X^2, X^3, X^4, X^5)$.

On peut donc avec les mêmes arguments que tout à l'heure la compléter en une base de $\mathbb{R}_4[X]$:

$$= (3, 1 + X, X + X^2, X^3, X^4, X^5)$$

IV Coordonnées

IV-1 Unicité de la décomposition dans une base

Théorème 72 de décomposition dans une base

Soit E un espace vectoriel.

La famille $\mathcal{B} = (e_1, \dots, e_n)$ d'éléments de E est une base de E si et seulement si, pour tout vecteur $u \in E$, il existe une unique décomposition $u = \alpha_1 e_1 + \dots + \alpha_n e_n$.

Démonstration :

- Supposons que $\mathcal{B} = (e_1, \dots, e_n)$ est une base de E . Supposons qu'il existe un élément $u \in E$ ayant deux décompositions $u = \alpha_1 e_1 + \dots + \alpha_n e_n$ et $u = \alpha'_1 e_1 + \dots + \alpha'_n e_n$.

Ainsi, en soustrayant les deux égalités, on obtient

$$(\alpha_1 - \alpha'_1)e_1 + \dots + (\alpha_n - \alpha'_n)e_n = 0_E$$

La famille (e_1, \dots, e_n) étant libre, on obtient par définition

$$\alpha_1 - \alpha'_1 = \dots = \alpha_n - \alpha'_n = 0$$

- Supposons que toute décomposition est unique. Montrons que la famille est libre :

Supposons qu'il existe $\alpha_1, \dots, \alpha_n \in \mathbb{K}$ tel que

$$\alpha_1 e_1 + \dots + \alpha_n e_n = 0_E$$

Par unicité de la décomposition, comme $0_E = 0e_1 + \dots + 0e_n$, on obtient

$$\alpha_1 = \dots = \alpha_n = 0$$

\square

? Exercice 7

D'après l'exercice précédent, on sait que $\mathcal{B} = \{(1, 3, -1), (2, 1, i), (1, 0, 0)\}$ est une base de \mathbb{C}^3 . Déterminer l'unique décomposition de $(0, 4, -1 + i)$ dans cette base \mathcal{B} .

Solution

A vue, on constate que $\underbrace{(1, 3, -1)}_{u_1} + \underbrace{(2, 1, i)}_{u_2} - 3 \underbrace{(1, 0, 0)}_{u_3} = (0, 4, -1 + i)$

D'où

$$(0, 4, -1 + i) = u_1 + u_2 - 3u_3$$

On sait que c'est la seule décomposition possible car \mathcal{B} est une base.

? Exercice 8

Montrer que la famille de polynômes $\{3 + X, X, X^2 + 1\}$ est une base de $\mathbb{K}_2[X]$ et déterminer l'unique décomposition de 1 dans $1 + X - 2X^2$ dans cette base.

Soient $P = a + bX + cX^2$ un polynôme quelconque de $\mathbb{R}_2[X]$ et $\alpha, \beta, \gamma \in \mathbb{K}$. Alors

$$P = \alpha(3 + X) + \beta X + \gamma(1 + X^2) \quad \begin{array}{c} \Leftrightarrow \\ \text{par identification} \end{array} \quad P = (3\alpha + \gamma) + (\alpha + \beta)X + \gamma X^2$$

$$\Leftrightarrow \begin{cases} a = 3\alpha + \gamma \\ b = \alpha + \beta \\ c = \gamma \end{cases} \Leftrightarrow \dots$$

$$\Leftrightarrow \begin{cases} \alpha = \frac{1}{3}(a - c) \\ \beta = \frac{1}{3}(a - c) - b \\ \gamma = c \end{cases}$$

Il existe donc une seule et unique solution à la décomposition de P pour tout P . La famille $\{3 + X, X, X^2 + 1\}$ est donc une base de $\mathbb{K}_2[X]$.

De plus, la décomposition de $1 + X - 2X^2$ est obtenue directement avec les coefficients $a = b = 1, c = -2$ ci-dessus :

$$1 + X - 2X^2 = 1 \cdot (3 + X) - 0 \cdot X - 2 \cdot (X^2 + 1)$$

⚠ Remarque :

Si la famille $\{e_1, \dots, e_n\}$ n'est pas libre, la décomposition n'est pas unique. (cf. exemple ci-dessous.)

■ Exemple 28 :

Supposons que $e_1 = (1, 0, 2), e_2 = (1, 1, 1)$ et $e_3 = (0, -1, 1)$. **La famille n'est pas libre** car

$$e_3 = e_1 - e_2$$

On a alors par exemple

$$e_1 + e_2 + e_3 = (2, 0, 4) = 2e_1 + 0e_2 + 0e_3$$

La décomposition n'est donc pas unique.

⚠ Remarque :

Si la famille $\{e_1, \dots, e_n\}$ n'est pas génératrice de l'espace E , la décomposition de $x \in E$ n'est pas toujours possible. (cf. exemple ci-dessous.)

■ Exemple 29 :

Supposons que $e_1 = (1, 0, 2), e_2 = (1, 1, 1)$. **La famille n'est pas génératrice** de \mathbb{R}^3 . (Elle est de cardinal $2 < 3$; cf. fin du cours.)

Le vecteur $x = (1, 0, 0)$ ne se décompose pas grâce à e_1 et e_2 .

En effet, supposons qu'une décomposition existe :

$$(1, 0, 0) = \alpha(1, 0, 2) + \beta(1, 1, 1)$$

Alors, par unicité des coefficients d'un vecteur de \mathbb{R}^3 ,

$$\begin{cases} 1 = \alpha + \beta \\ 0 = \beta \\ 0 = 2\alpha + \beta \end{cases}$$

Ce qui donne $1 = \alpha = 0$. C'est absurde. **La décomposition n'existe donc pas.**

IV-2 Notion de coordonnées

📖 Notation :

Si E est un espace vectoriel admettant une base $\mathcal{B} = (e_1, \dots, e_n)$, tout vecteur $u \in E$ admet une unique décomposition dans la base \mathcal{B} : $u = \alpha_1 e_1 + \dots + \alpha_n e_n$.

Pour des raisons de lourdeur d'écriture, on préfère généralement se débarrasser des notations e_1, \dots, e_n . On écrit alors à la place :

$$u = (\alpha_1, \dots, \alpha_n)_{\mathcal{B}}$$

■ Exemples :

30 ■ Dans \mathbb{R}^3 , on note $\mathcal{B} = ((1, 1, 0), (0, 1, 1), (1, 0, 1))$ (base de \mathbb{R}^3), alors

$$(2, 1, -1)_{\mathcal{B}} = 2 \cdot (1, 1, 0) + 1 \cdot (0, 1, 1) - 1 \cdot (1, 0, 1) = (1, 3, 0)$$

31 ■ Dans $\mathbb{R}_2[X]$, on note $\mathcal{B} = (1, X, X^2)$ et $\mathcal{B}' = (1, X - 1, X^2)$ deux bases de $\mathbb{R}_2[X]$.

$$(2, -3, 2)_{\mathcal{B}} = 2 - 3X + 2X^2 = -1 \cdot 1 - 3 \cdot (X - 1) + 2 \cdot X^2 = (-1, -3, 2)_{\mathcal{B}'}$$

Commentaires :

Ceci amène également à la définition suivante, qui nous permet de nous ramener à \mathbb{K}^n et qui nous sera bien utile dans les futures formules matricielles :

📖 Définition

On appelle $(\alpha_1, \dots, \alpha_n)$ les **coordonnées** de $u = \alpha_1 e_1 + \dots + \alpha_n e_n$ dans la base $\mathcal{B} = (e_1, \dots, e_n)$ et on note

$$M_{\mathcal{B}}(u) = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} \in \mathbb{K}^n$$

le vecteur associé, nommé **vecteur coordonnées** de u dans \mathcal{B}

 **Notation :**
 Dans le même cadre, il arrive souvent de voir également le vecteur en ligne :

$$M_{\mathcal{B}}(u) = (\alpha_1, \dots, \alpha_n) \quad \text{au lieu de} \quad M_{\mathcal{B}}(u) = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$$

 **NÉANMOINS**
 > seule l'écriture en colonne sera utilisée dans les calculs matriciels à venir.

 **Remarque :**
 $M_{\mathcal{B}}(u) \in \mathbb{K}^n$ et $u \in E$. Généralement, (sauf si $E = \mathbb{K}^n$ et que \mathcal{B} est la base canonique), on a donc

$$M_{\mathcal{B}}(u) \neq u$$

Ce sont donc **deux objets différents** à différencier avec précaution dans les calculs.

■ **Exemples :**

32 ■ Dans \mathbb{K}^3 , on note $\mathcal{B} = ((1, 1, 0), (0, 1, 1), (1, 0, 1))$ (base de \mathbb{K}^3), alors si

$$M_{\mathcal{B}}(u) = (2, 1, -1)$$

on a $u = 2 \cdot (1, 1, 0) + 1 \cdot (0, 1, 1) - 1 \cdot (1, 0, 1) = (1, 3, 0)$

On voit bien que

$$u \neq M_{\mathcal{B}}(u)$$

33 ■ Dans $\mathbb{R}_2[X]$, on note $\mathcal{B} = (1, X, X^2)$ et $\mathcal{B}' = (1, X-1, X^2)$ deux bases de $\mathbb{R}_3[X]$.
 Si

$$M_{\mathcal{B}}(u) = (2, -3, 2)$$

alors $u = 2 - 3X + 2X^2 = -1 \cdot 1 - 3 \cdot (X - 1) + 2 \cdot X^2$

d'où

$$M_{\mathcal{B}'}(u) = (-1, -3, 2)$$

Là encore, $u \neq M_{\mathcal{B}}(u)$, $u \neq M_{\mathcal{B}'}(u)$ et en toute logique $M_{\mathcal{B}}(u) \neq M_{\mathcal{B}'}(u)$.

IV-3 Changement de base

On se donne deux bases $\mathcal{B} = (e_1, \dots, e_n)$ et $\mathcal{B}' = (e'_1, \dots, e'_n)$ et on suppose ici que l'on dispose des coordonnées des vecteurs de \mathcal{B}' dans \mathcal{B} , c'est-à-dire

$$M_{\mathcal{B}}(e'_k) = (\alpha_{1,k}, \dots, \alpha_{n,k}) \quad \forall k = 1, \dots, n$$

■ **Exemple 34 :**

Dans $\mathbb{R}_2[X]$, on considère les bases

$$\mathcal{B} = (1, X, X^2) \quad \text{et} \quad \mathcal{B}' = (\underbrace{1}_{e'_1}, \underbrace{X-1}_{e'_2}, \underbrace{(X-1)(X-2)}_{e'_3})$$

où on a alors

$$M_{\mathcal{B}}(e'_1) = (1, 0, 0), \quad M_{\mathcal{B}}(e'_2) = (-1, 1, 0), \quad M_{\mathcal{B}}(e'_3) = (2, -3, 1)$$

On se pose la question suivante :
Ayant les coordonnées d'un vecteur u dans l'une des deux bases, comment trouver les coordonnées de u dans l'autre ?

IV.3-a) Si on a les coordonnées de u dans \mathcal{B}' et on veut $M_{\mathcal{B}}(u)$

i.e. on a $u = \mu_1 e'_1 + \dots + \mu_n e'_n$ et on cherche $u = \lambda_1 e_1 + \dots + \lambda_n e_n$:

Méthode n° 1 : On développe :

On connaît $M_{\mathcal{B}'}(u) = (\mu_1, \dots, \mu_n)$. Alors

$$u = \mu_1 e'_1 + \dots + \mu_n e'_n = \mu_1 (\text{expression de } e'_1 \text{ en fonction des } e_i) + \dots$$

■ **Exemple 35 :**

Dans les mêmes bases que dans l'exemple précédent, on cherche les coordonnées du polynôme T tel que $M_{\mathcal{B}'}(T) = (2, 0, 1)$ dans la base canonique \mathcal{B} , c'est-à-dire les coefficients α, β, γ tels que

$$T = \alpha e_1 + \beta e_2 + \gamma e_3 = \alpha + \beta X + \gamma X^2$$

On a

$$T = 2 + 0 \cdot (X - 1) + 1 \cdot (X - 1)(X - 2) = 2 + (X^2 - 3X + 2) = X^2 - 3X + 4$$

D'où

$$M_{\mathcal{B}}(T) = (4, -3, 1)$$

Méthode n° 2 : Calcul matriciel

On écrit la matrice composée de l'ensemble des coordonnées de \mathcal{B}' dans \mathcal{B} :

$$M_{\mathcal{B}}(\mathcal{B}') = \begin{matrix} & e'_1 & \dots & e'_n \\ e_1 & \alpha_{1,1} & \dots & \alpha_{1,n} \\ \vdots & \vdots & & \\ e_n & \alpha_{n,1} & \dots & \alpha_{n,n} \end{matrix}$$

■ Exemple 36 :

Dans $\mathbb{R}_2[X]$, on note encore \mathcal{B} la base canonique $\mathcal{B} = (1, X, X^2)$ et :

$$\mathcal{B}' = (\underbrace{1}_{e'_1}, \underbrace{X-1}_{e'_2}, \underbrace{(X-1)(X-2)}_{e'_3}). \text{ alors } M_{\mathcal{B}}(\mathcal{B}') = \begin{matrix} & e'_1 & e'_2 & e'_3 \\ \begin{matrix} 1 \\ X \\ X^2 \end{matrix} & \begin{pmatrix} 1 & -1 & 2 \\ 0 & 1 & -3 \\ 0 & 0 & 1 \end{pmatrix} \end{matrix}$$

Toujours avec $M_{\mathcal{B}'}(T) = (2, 0, 1)$, on a $T = 2e'_1 + 0e'_2 + 1e'_3$. Matriciellement, ceci donne

$$\begin{aligned} M_{\mathcal{B}}(T) &= 2M_{\mathcal{B}}(e'_1) + 0M_{\mathcal{B}}(e'_2) + 1M_{\mathcal{B}}(e'_3) \\ &= 2 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + 0 \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + 1 \begin{pmatrix} 2 \\ -3 \\ 1 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & -1 & 2 \\ 0 & 1 & -3 \\ 0 & 0 & 1 \end{pmatrix}}_{M_{\mathcal{B}}(\mathcal{B}')} \underbrace{\begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}}_{M_{\mathcal{B}'}(T)} = \begin{pmatrix} 4 \\ -3 \\ 1 \end{pmatrix} \end{aligned}$$

On observe alors ci-dessus ici pour résumer que :

$$M_{\mathcal{B}}(T) = M_{\mathcal{B}}(\mathcal{B}')M_{\mathcal{B}'}(T)$$

Ce principe se généralise en fait à tous les cas possibles :

 Définition

Étant données deux bases $\mathcal{B} = (e_1, \dots, e_n)$ et $\mathcal{B}' = (e'_1, \dots, e'_n)$ telles que

$$e'_k = (\alpha_{1,k}, \dots, \alpha_{n,k})_{\mathcal{B}} \quad \forall k = 1, \dots, n$$

on note

$$P_{\mathcal{B}, \mathcal{B}'} = M_{\mathcal{B}}(\mathcal{B}') = \begin{matrix} & e'_1 & \dots & e'_n \\ \begin{matrix} e_1 \\ \vdots \\ e_n \end{matrix} & \begin{pmatrix} \alpha_{1,1} & \dots & \alpha_{1,n} \\ \vdots & & \\ \alpha_{n,1} & \dots & \alpha_{n,n} \end{pmatrix} \end{matrix}$$

qu'on appelle *matrice de passage de la base \mathcal{B} à la base \mathcal{B}'* .

Théorème 73 (Formule de changement de base)

Étant donné un espace vectoriel E et deux bases \mathcal{B} et \mathcal{B}' , pour tout $u \in E$, on a

$$M_{\mathcal{B}}(u) = M_{\mathcal{B}}(\mathcal{B}')M_{\mathcal{B}'}(u)$$

avec $M_{\mathcal{B}}(\mathcal{B}')$ inversible.

Démonstration :

Réécrire l'exemple ci-dessus avec des valeurs générales ! L'inversibilité sera justifiée à la fin du chapitre. □

■ Exemple 37 :

Si $\mathcal{B}' = \{(1, 0, 2), (0, 1, 1), (1, -1, 0)\}$ et $u = (-1, 0, 1)_{\mathcal{B}'}$, alors les coordonnées de u dans la base canonique sont :

$$M_{\mathcal{B}}(u) = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 2 & 1 & 0 \end{pmatrix} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \\ -2 \end{pmatrix}$$

i.e. $u = (0, -1, -2)$.

IV.3-b) Si on a les coordonnées de u dans \mathcal{B} et on veut $M_{\mathcal{B}'}(u)$

i.e. on a $u = \lambda_1 e_1 + \dots + \lambda_n e_n$ et on cherche $u = \mu_1 e'_1 + \dots + \mu_n e'_n$:

■ Exemple 38 :

Dans $\mathbb{R}_3[X]$, avec toujours $\mathcal{B} = (1, X, X^2)$ et : $\mathcal{B}' = (\underbrace{1}_{e'_1}, \underbrace{X-1}_{e'_2}, \underbrace{(X-1)(X-2)}_{e'_3})$,

on cherche les coordonnées du polynôme $T = X^2 - 3X + 4$ dans la base \mathcal{B}' , c'est-à-dire les coefficients α, β, γ tels que

$$T = \alpha e'_1 + \beta e'_2 + \gamma e'_3 = \alpha + \beta(X-1) + \gamma(X-1)(X-2)$$

Méthode n° 1 : Par résolution de système

On développe l'expression de droite, et on résout les équations obtenues pour trouver les coefficients μ_i .

■ Exemple 39 :

Dans l'exemple précédent, on a

$$X^2 - 3X + 4 = (\alpha - \beta + 2\gamma) + X(\beta - 3\gamma) + \gamma X^2$$

La famille $(1, X, X^2)$ étant libre, il y a unicité des coefficients et on peut identifier :

$$\begin{cases} 1 & = \gamma \\ -3 & = \beta - 3\gamma \\ 4 & = \alpha - \beta + 2\gamma \end{cases} \Leftrightarrow \text{résolution...} \Leftrightarrow \begin{cases} \alpha & = 2 \\ \beta & = 0 \\ \gamma & = 1 \end{cases}$$

Méthode n° 1 : Par inversion de matrice

On a vu tout à l'heure que $M_{\mathcal{B}}(u) = M_{\mathcal{B}}(\mathcal{B}')M_{\mathcal{B}'}(u)$. Alors

$$M_{\mathcal{B}'}(u) = M_{\mathcal{B}}(\mathcal{B}')^{-1}M_{\mathcal{B}}(u)$$

En effet, si on note $P = M_{\mathcal{B}}(\mathcal{B}')$, comme elle est inversible, en multipliant à gauche par P^{-1} l'égalité $M_{\mathcal{B}}(u) = PM_{\mathcal{B}'}(u)$, on obtient ce qui est dit dans la méthode.

■ **Exemple 40** :

On considère la base de $\mathbb{R}_2[X]$: $\mathcal{B}' = (1, X - 1, (X - 1)(X - 2))$. On cherche les coordonnées du polynôme $X^2 - 3X + 4$ dans cette base.

On pose $P = \begin{matrix} & e'_1 & \dots & e'_n \\ 1 & \begin{pmatrix} 1 & -1 & 2 \\ 0 & 1 & -3 \\ 0 & 0 & 1 \end{pmatrix} \end{matrix}$ P est inversible avec $P^{-1} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix}$,

alors les coordonnées de $X^2 - 3X + 4$ dans la base \mathcal{B}' sont :

$$M_{\mathcal{B}'}(T) = P^{-1} \begin{pmatrix} 4 \\ -3 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 4 \\ -3 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}$$

IV.3-c) Propriétés des matrices de passage

Pour commencer, on peut remarquer que les passages de bases s'enchaînent telle une relation de Chasles (ceci sera ssimplement dans le cadre des applications linéaires) :

Proposition 74

Soit E un espace vectoriel de dimension finie, avec $\mathcal{B}, \mathcal{C}, \mathcal{D}$ trois bases de E . Alors

$$M_{\mathcal{B}}(\mathcal{C})M_{\mathcal{C}}(\mathcal{D}) = M_{\mathcal{B}}(\mathcal{D})$$

Démonstration : admise. □

Commentaires :

| En bref : passer de \mathcal{B} à \mathcal{C} puis de \mathcal{C} à \mathcal{D} revient à passer de \mathcal{B} à \mathcal{D} .

Corollaire

Étant donné un espace vectoriel E et deux bases \mathcal{B} et \mathcal{B}' , la matrice de passage $M_{\mathcal{B}}(\mathcal{B}')$ est inversible et on a

$$M_{\mathcal{B}}(\mathcal{B}')^{-1} = M_{\mathcal{B}'}(\mathcal{B})$$

Démonstration :

D'après la proposition, on a $M_{\mathcal{B}}(\mathcal{B}')M_{\mathcal{B}'}(\mathcal{B}) = M_{\mathcal{B}}(\mathcal{B}) = Id$ (à méditer !)

De même, $M_{\mathcal{B}'}(\mathcal{B})M_{\mathcal{B}}(\mathcal{B}') = M_{\mathcal{B}'}(\mathcal{B}') = Id$. CQFD □

 **Notation :**

Pour $a, b \in \overline{\mathbb{R}}$, dans ce cours, la notation $|a, b|$ désignera un intervalle d'extrémités a, b qui peut être soit ouvert, soit fermé, soit semi ouvert selon les besoins. (On pourra remarquer que dans cette notation, b peut être strictement plus petit que a . Par exemple, l'intervalle $|3, -\infty|$ a un sens.)

■ **Exemple 1 :**

La notation $|3, 2|$ pourra désigner $[2, 3]$, $[2, 3[$, $]2, 3]$ ou encore $]2, 3[$ suivant les besoins.

La notation $|2, 3|$ désignera $[2, 3[$ ou $]2, 3[$ suivant les besoins.

I Rappels de notations, vocabulaire et d'intégration sur un intervalle $[a, b]$

Rappel de notation : pour désigner l'intégrale d'une fonction continue f sur un intervalle I contenant a et b , (avec non nécessairement $a < b$) on peut écrire :

$$\int_a^b f(t) dt, \quad \int_a^b f$$

I-1 Intégration par partie sur un segment

Théorème 75 Intégration par partie

Soit u, v deux fonctions \mathcal{C}^1 sur un intervalle fermé $[a, b]$, alors

$$\int_a^b u'v = [uv]_a^b - \int_a^b uv'$$

■ **Exemple 2 :**

Calculer $\int_0^{\frac{\pi}{2}} x \cos x dx$

On pose les deux fonctions $u, v \in \mathcal{C}^1([0, \frac{\pi}{2}])$ telles que

$$\begin{aligned} u(x) &= x & u'(x) &= 1 \\ v'(x) &= \cos x & v(x) &= \sin x \end{aligned}$$

Alors,

$$\int_0^{\frac{\pi}{2}} x \cos x dx = \int_0^{\frac{\pi}{2}} u'v \underbrace{=}_{IPP} = \underbrace{[uv]_0^{\frac{\pi}{2}}}_{=\frac{\pi}{2}} - \int_0^{\frac{\pi}{2}} u'v = \frac{\pi}{2} - \int_0^{\frac{\pi}{2}} \sin x dx$$

Or

$$\int_0^{\frac{\pi}{2}} \sin x dx = [-\cos x]_0^{\frac{\pi}{2}} = 1$$

Ainsi

$$\int_0^{\frac{\pi}{2}} x \cos x dx = \frac{\pi}{2} - 1$$

I-2

Changement de variable sur une intégrale $\int_a^b f$

Point de départ dans cette section : une intégrale $\int_a^b f(t)dt$ que l'on souhaite simplifier.

Commentaires :

Le changement de variable dans une intégrale $\int_a^b f$ est une modification raisonnable du "f", destinée à simplifier le calcul :

$$\int_a^b f(t) dt = \int_\alpha^\beta g(x) dx$$

(L'aire sous la courbe de f est la même que la même que sous celle de g , quitte à changer des bornes d'intégration.) Il y a une infinité de telles fonctions mais la formule du "changement de variable" est une technique particulière permettant d'y arriver.

Nous allons voir plusieurs manière de l'exprimer.

I.2-a) Intégration par substitution et changement de variable $x = \varphi(t)$

On rappelle le théorème de cours ci-dessous, dont la démonstration est un principe simple de manipulation de fonctions, et qui est rappelée également.

Théorème 76 dit d'intégration par substitution / changement de variable

Soient $a, b \in \mathbb{R}$, I un intervalle f et φ deux fonctions telles que :

- $g \circ \varphi : [a, b] \xrightarrow{\varphi} I \xrightarrow{g} \mathbb{R}$ soit bien définie
- $g : I \rightarrow \mathbb{R}$ soit continue
- $\varphi : [a, b] \rightarrow I$ soit \mathcal{C}^1

alors on a

$$\int_a^b \underbrace{g(\varphi(t))\varphi'(t)}_{\text{"f(t)"}} dt = \int_{\varphi(a)}^{\varphi(b)} g(x) dx$$

Démonstration :

On part d'un principe simple : la formule de dérivation de la composition sur des fonctions \mathcal{C}^1 :

$$(G \circ \varphi)'(t) = \varphi'(t) \cdot G'(\varphi(t))$$

La formule précédente permet ainsi de réécrire l'intégrale d'une autre manière :

$$\int_a^b \underbrace{\varphi'(t) \cdot G'(\varphi(t))}_{f(t)} dt = \int_a^b (G \circ \varphi)'(t) dt = G(\varphi(b)) - G(\varphi(a)) = \int_{\varphi(a)}^{\varphi(b)} \underbrace{G'(t)}_{g(t)} dt$$

□

Notation :

Comme on peut le voir dans la démonstration, l'égalité du théorème pourrait s'écrire avec la même variable t des deux côtés. Néanmoins, on s'aperçoit que pour des raisons de praticité et de notation dans les calculs, il est plus pertinent (sous peine de confusions) d'utiliser plutôt deux noms de variables différents :

$$\int_a^b \underbrace{\varphi'(t) \cdot G'(\varphi(t))}_{f(t)} dt = \int_{\varphi(a)}^{\varphi(b)} \underbrace{G'(x)}_{g(x)} dx$$

Remarque :

La plupart du temps, ce théorème est fait pour être utilisé dans ce sens là :

$$\int_a^b \underbrace{g(\varphi(t))\varphi'(t)}_{\text{"f(t)"}} dt = \int_{\varphi(a)}^{\varphi(b)} \underbrace{g(x) dx}_{\text{arrivée}}$$

départ

Méthode 1 : Calcul par identification directe des intervenants dans la formule

C'est la méthode la moins courante, mais l'application la plus directe de la formule de substitution :

Étant donné φ (qui est généralement donnée dans l'énoncé), on repère dans l'intégrale $\int_a^b f$ les fonctions φ et g en les mettant en valeur chacune de leur côté pour

$$\text{trouver } g \text{ telle que } \int_a^b f(t) dt = \int_a^b g(\varphi(t))\varphi'(t) dt.$$

Exemple 3 :

Calculons $\int_1^2 \frac{e^{-\frac{1}{t}}}{t^2} dt$ avec le chang. de variable $\varphi(t) = -\frac{1}{t}$

Soit $\varphi(t) = -\frac{1}{t}$ et $g : x \in \mathbb{R} \mapsto e^x > 0$.

de cette manière, $f = g \circ \varphi$ est bien définie.

Le changement de variable φ est \mathcal{C}^1 ($[1, 2]$), tandis que g est continue.

On constate également que

t	$1 (= a)$	$2 (= b)$
$\varphi(t)$	$-1 (= \alpha)$	$-\frac{1}{2} (= \beta)$

La formule de substitution donne alors

$$\int_1^2 \frac{e^{-\frac{1}{t}}}{t^2} dt = \int_1^{\frac{1}{2}} \underbrace{e^{-\frac{1}{t}}}_{g(\varphi(t))} \underbrace{\frac{1}{t^2}}_{\varphi'(t)} dt = \int_{-1}^{-\frac{1}{2}} e^x dx = [e^x]_{-1}^{-\frac{1}{2}} = e^{-\frac{1}{2}} - e^{-1} = \frac{1}{e} (e^{\frac{1}{2}} - 1)$$

Explications générales : Observons le principe de substitution sur la formule :

$$\left(\int_a^b f(t) dt = \right) \int_a^b g(\varphi(t))\varphi'(t) dt = \int_{\alpha}^{\beta} g(u) du$$

De gauche à droite :

- $g(\varphi(t))$ se transforme en $g(u)$:

$$\int_a^b \underbrace{g(\varphi(t))\varphi'(t)}_{g(x)} dt = \int_{\alpha}^{\beta} g(x) dx$$

D'où le fait qu'on rencontre souvent la notation

$$x = \varphi(t)$$

- " $\varphi'(t) dt$ " se transforme en " dx " :

$$\int_a^b g(\varphi(t))\varphi'(t) dt = \int_{\alpha}^{\beta} g(x) dx$$

qui s'écrirait comme une "forme différentielle" (notion détaillée HP)

$$dx = \varphi'(t) dt$$

Commentaires :

L'explication précédente prend son sens de la manière suivante :

- Pour le premier item précédente, : on intègre $g(\varphi(t))$ entre a et b , ce qui signifie qu'on intègre g entre " $\varphi(a)$ " = α et " $\varphi(b)$ " = β
- Le deuxième item $dx = \varphi'(t) dt$ se traduirait comme "la dérivée de x par-rapport à t est $\varphi'(t)$ ", ce qui, en PC, s'écrit également

$$\frac{dx}{dt} = \varphi'(t)$$

En pratique, "en maths", la théorie (qui n'est pas à votre programme) sur ces notations permet de considérer tout ceci comme des **calculs avec 3 quantités "différentes"** qui sont " $dx, \varphi'(t)$ et dt ". que l'on peut multiplier ou diviser à notre guise, avec les significations suivantes :

dx " = infime variation de x ", dt " = infime variation de t "

ainsi que par conséquent :

$$\frac{dx}{dt} = \frac{\text{infime variation de } \varphi(t)}{\text{infime variation de } t} = \varphi'(t) \quad (\text{cf. définition de la dérivée})$$

Ceci peut donc donner lieu à une rédaction du type suivant :

■ Exemple 4 :

Re-calculons $\int_1^2 \frac{e^{-\frac{1}{t}}}{t^2} dt$ avec le chang. de variable $x = -\frac{1}{t}$

Soit $x = -\frac{1}{t}$ qui est $\mathcal{C}^1([1, 2])$, avec $dx = \frac{1}{t^2} dt$

On constate également que

t	1	2
x	-1	$-\frac{1}{2}$

On reconnaît alors

$$\int_1^2 \frac{e^{-\frac{1}{t}}}{t^2} dt = \int_{-1}^{-\frac{1}{2}} \underbrace{e^x}_{g(x)} \underbrace{\frac{1}{t^2} dt}_{dx}$$

où $g = \exp$ continue sur \mathbb{R} . Ainsi, par formule de substitution,

$$\int_1^2 \frac{e^{-\frac{1}{t}}}{t^2} dt = \int_{-1}^{-\frac{1}{2}} e^x dx = \dots = \frac{1}{e} (e^{\frac{1}{2}} - 1)$$

I.2-b) Changement de variable (le vrai!) avec changement de variable $t = \psi(x)$

Dans cette partie, on suppose que φ' ne s'annule pas, ce qui signifie en réalité que φ est **strictement monotone**. En notant $\psi = \varphi^{-1}$, on obtient alors

$$x = \varphi(t) \Leftrightarrow t = \psi(x)$$

⚠ Remarque :

Si φ' ne s'annule pas sur l'intervalle $[a, b]$, on peut alors écrire :

$$dt = \frac{1}{\varphi'(t)} dx \quad \text{ou} \quad \frac{dt}{dx} = \frac{1}{\varphi'(t)}$$

Ce qui est cohérent avec la notation "dérivée de t par rapport à x " qui serait :

$$\frac{dt}{dx} = \psi'(x) \quad \text{ou} \quad dt = \psi'(x) dx$$

car par formule de dérivabilité de l'inverse :

$$\frac{1}{\varphi'(t)} = (\varphi^{-1})'(x) = \psi'(x)$$

où on rappelle au passage que φ est \mathcal{C}^1 ssi ψ est \mathcal{C}^1 .

⚠ Remarque :

Dans le cas où φ (ou ψ) est bijective, le théorème de substitution peut s'écrire d'une manière qui semble moins abordable, mais qui pourtant est nettement plus pratique d'utilisation car on n'a pas à "deviner" le g et que les vérifications se font directement sur f et le changement de variable

Théorème 77 Changement de variable version " $t = \psi(x)$ "

Soient $\alpha, \beta \in \mathbb{R}$, I un intervalle, ainsi que f et ψ deux fonctions telles que :

- $f : [a, b] \rightarrow \mathbb{R}$ soit continue
- $\psi : [a, b] \rightarrow \mathbb{R}$ soit \mathcal{C}^1 et strictement monotone

alors on a

$$\int_a^b f(t) dt = \int_{\psi^{-1}(a)}^{\psi^{-1}(b)} f(\psi(x)) \psi'(x) dx$$

Démonstration :

On pose $\alpha = \psi^{-1}(a)$ et $\beta = \psi^{-1}(b)$. Alors ψ est bien \mathcal{C}^1 sur l'intervalle fermé d'extrémités α, β . le théorème de substitution permet de dire que

$$\int_a^b f(t)dt = \int_{\psi(\alpha)}^{\psi(\beta)} f(t) dt = \int_{\alpha}^{\beta} f(\psi(x))\psi'(x) dx$$

Ce qui s'écrit donc bien

$$\int_a^b f(t) dt = \int_{\psi^{-1}(a)}^{\psi^{-1}(b)} f(\psi(x))\psi'(x) dx$$

□

Méthode 2 : On exprime t en fonction de x

Concrètement, dans les calculs, on peut poser

$$x = \varphi(t) \text{ (ou directement } t = \psi(x))$$

et, si φ' ne s'annule pas sur l'intervalle d'intégration :

$$dt = \psi'(x) dx$$

Le remplacement peut donc être fait directement dans la formule de l'intégrale.

■ **Exemple 5 :**

Re-calculons $\int_1^2 \frac{e^{-\frac{1}{t}}}{t^2} dt$ avec le changement de variable $x = -\frac{1}{t}$.

Posons $x = -\frac{1}{t}$ qui est donc un changement de variable \mathcal{C}^1 et bijectif avec

$$t = -\frac{1}{x} \text{ et } \frac{dt}{dx} = \frac{1}{x^2} > 0$$

sur les intervalles

t	1	2
x	-1	$-\frac{1}{2}$

La formule de changement de variable donne alors (on remplace t par $-\frac{1}{x}$ et dt par $\frac{1}{x^2} dx$) :

$$\int_1^2 \frac{e^{-\frac{1}{t}}}{t^2} dt = \int_{-1}^{-\frac{1}{2}} \underbrace{e^x}_{\frac{e^{-\frac{1}{t}}}{t^2}} \underbrace{x^2 \frac{1}{x^2}}_{\frac{1}{dx}} dx = \int_{-1}^{-\frac{1}{2}} e^x dx = \frac{1}{e} (e^{\frac{1}{2}} - 1)$$

I-3 Primitivation

Pour désigner **une** primitive de f de variable $x \in I$, on peut écrire

$$\int^x f, \text{ ou } \int^x f, \text{ ou souvent simplement } \int f,$$

et **la** primitive F de f qui s'annule en $c \in I$ est définie pour tout $x \in I$ par

$$F(x) = \int_c^x f = \int^x f.$$

Ainsi, pour calculer des primitives, toutes les formules des intégrales peuvent s'adapter, notamment l'IPP et le changement de variable!

■ **Exemple 6 :**

Calculer une primitive de $x \sin x$ sur \mathbb{R}

On pose les deux fonctions $u, v \in \mathcal{C}^1(\mathbb{R})$ telles que

$$\begin{aligned} u(x) &= x & u'(x) &= 1 \\ v(x) &= \sin x & v'(x) &= \cos x \end{aligned}$$

Alors,

$$\int x \sin x dx = \int \underbrace{u'v}_{IPP} = uv - \int u'v = -x \cos x + \int \cos x dx$$

Ainsi

$$\int x \sin x dx = \sin x - x \cos x +$$

■ **Exemple 7 :**

Calculons une primitive de $\frac{e^{-\frac{1}{x}}}{x^2}$ sur $]0, +\infty[$ avec le changement de variable $u = -\frac{1}{x}$.

La fonction demandée étant continue, une primitive existe.

(Ceci est une fonction sous la forme $u'e^u$. Elle est donc théoriquement facile à intégrer mais on souhaite ici illustrer le théorème de changement de variable dans une primitive.)

Posons $u = -\frac{1}{x}$ qui est donc un changement de variable \mathcal{C}^1 et bijectif avec

$$x = -\frac{1}{u} \text{ et } \frac{dx}{du} = \frac{1}{u^2} > 0$$

La formule de changement de variable donne alors (on remplace x par $-\frac{1}{u}$ et dx par $\frac{1}{u^2} du$) :

$$\int \frac{e^{-\frac{1}{x}}}{x^2} dx = \int \underbrace{e^u}_{\frac{e^{-\frac{1}{x}}}{x^2}} \underbrace{u^2 \frac{1}{u^2}}_{\frac{1}{dx}} du = \int e^u du = e^u = e^{-\frac{1}{x}}$$

II Convergence d'une intégrale (impropre ou) généralisée

Rappel de vocabulaire avant de commencer :

Définition

On rappelle qu'une fonction f est dite continue sur un intervalle $]a, b[$ si elle est continue sur $]a, b[$ et si $\lim_{b^-} f = f(b)$.

Pour la deuxième condition, on dit souvent que f est continue en b^- .

Le même type de vocabulaire est évidemment valable sur des intervalles de type $[a, b[$ où f serait alors continue en a^+ .

Exemple 8 :

Soit $f : t \mapsto \begin{cases} 0 & \text{si } t < 0 \\ e^{-t} & \text{si } t \geq 0 \end{cases}$. Alors f est continue sur $] -\infty, 0[$ et $[0, +\infty[$

CONFUSION

Dans l'exemple précédent, f n'est pas continue sur $] -\infty, 0[\cup [0, +\infty[$. En effet, $] -\infty, 0[\cup [0, +\infty[= \mathbb{R}$ et f n'est pas continue sur \mathbb{R} car non continue en 0^- .

II-1 Déf. d'une intégrale généralisée sur un intervalle semi-ouvert

Commentaires :

On observe $\int_a^b f$, où f est une fonction continue au moins sur $]a; b[$. Quand parle-t-on d'intégrale impropre (ou généralisée) ?

- Dans le cas où a ou b (voir les 2) sont infinis.
- Dans le cas où f n'est pas définie ou non continue en a ou b (voir les deux.)

Définition

Soit $f :]a; b[\rightarrow \mathbb{R}$ continue sur $]a, b[$, où $a \in \mathbb{R}$ et $b \in \mathbb{R} \cup \{+\infty\}$.

• Si $\lim_{x \rightarrow b} \int_a^x f$ existe et est finie, alors on dit que $\int_a^b f$ converge (en b), (ou qu'elle est convergente) et dans ce cas, on note

$$\int_a^b f = \lim_{x \rightarrow b} \int_a^x f$$

• Sinon, on dit que l'intégrale $\int_a^b f$ diverge (en b), (ou qu'elle est divergente)

Remarque :

Dans le cas où $\lim_{x \rightarrow b} \int_a^x f = \infty$, il peut arriver que l'on note $\int_a^b f = \infty$.

Exemples :

9 ■ Montrons que $\int_0^{+\infty} te^{-t^2} dt$ est convergente et $\int_0^{+\infty} te^{-t^2} dt = \frac{1}{2}$:

La fonction $f : t \in [0; +\infty[\mapsto te^{-t^2}$ est continue, donc le seul problème est en $+\infty$.

Soit $x \in \mathbb{R}$, $x \geq 0$. On a $\int_0^x te^{-t^2} dt = \left[-\frac{e^{-t^2}}{2} \right]_0^x = \frac{1}{2} (1 - e^{-x^2})$.

Ainsi, $\lim_{x \rightarrow +\infty} \int_0^x te^{-t^2} dt = \frac{1}{2}$

On en déduit que $\int_0^{+\infty} te^{-t^2} dt$ est convergente et $\int_0^{+\infty} te^{-t^2} dt = \frac{1}{2}$.

10 ■ Montrons que $\int_1^{+\infty} \frac{1}{t} dt$ est divergente :

La fonction $f : t \in [1; +\infty[\mapsto \frac{1}{t}$ est continue. Il y a donc un problème en $+\infty$. Soit $x \in [1; +\infty[$. On a

$$\int_1^x \frac{1}{t} dt = [\ln t]_1^x = \ln x \xrightarrow{x \rightarrow 0} +\infty$$

On en déduit que $\int_1^{+\infty} \frac{1}{t} dt$ est divergente. On peut noter de plus que $\int_1^{+\infty} \frac{1}{t} dt = +\infty$

Commentaires :

Par le changement de variables $t = -u$, une intégrale \int_a^b se "transforme" en une intégrale \int_{-b}^{-a} (cf changement de variable un peu plus tard). Par exemple, $\int_0^{+\infty}$ se transforme en $\int_{-\infty}^0$. Il est donc logique que la définition reste similaire si on souhaite intégrer f sur $] -\infty; b[$ ou tout autre intervalle ouvert à gauche :

Définition

Soit $f :]a; b[\rightarrow \mathbb{R}$ continue, où $a \in \mathbb{R} \cup \{-\infty\}$ et $b \in \mathbb{R}$.

• Si $\lim_{x \rightarrow a, x > a} \int_x^b f$ existe et est finie, alors on dit que $\int_a^b f$ converge (en a) et dans ce cas, on a

$$\int_a^b f = \lim_{x \rightarrow a, x > a} \int_x^b f$$

• Sinon, on dit que l'intégrale $\int_a^b f$ est divergente.

■ Exemple 11 :

Montrons que $\int_0^1 \frac{1}{t} dt$ est divergente.

La fonction $f : t \in]0; 1] \mapsto \frac{1}{t}$ est continue. Il y a donc un **problème en 0**. Soit $x \in]0; 1]$.

On a $\int_x^1 \frac{1}{t} dt = [\ln t]_x^1 = -\ln x \xrightarrow{x \rightarrow 0} +\infty$

On en déduit que $\int_0^1 \frac{1}{t} dt$ est divergente. On peut noter de plus que $\int_0^1 \frac{1}{t} dt = +\infty$

⚠ Remarque :

Dans toute la suite, les propriétés ne concernant qu'un seul point critique seront énoncées dans le cas des intégrales du type $\int_a^b f$, où $a \in \mathbb{R}$, $b \in \bar{\mathbb{R}}$ et $f : [a, b[\rightarrow \mathbb{R}$ est continue. Les cas où $a \in \bar{\mathbb{R}}$ et $b \in \mathbb{R}$ se démontreront de la même manière par symétrie du raisonnement ou s'en déduisent par changement de variables.

 **Définition**
On appelle *nature* d'une intégrale $\int_a^b f$ son caractère convergent ou divergent.

⚠ Remarque :

"Donner la nature d'une intégrale" signifie « dire si elle est convergente ou divergente ». De la même façon, dire que deux intégrales sont de "même nature" signifie qu'elles sont soit toutes deux convergentes, soit toutes deux divergentes.

Théorème 78 (Relation de Chasles)

Soient $I \subset \mathbb{R}$ un intervalle de borne supérieure $b \in \bar{\mathbb{R}}$. Si $f : I \rightarrow \mathbb{R}$ est continue, alors, pour tout $a, c \in I$,

$$\int_a^b f \quad \text{et} \quad \int_c^b f \quad \text{ont la même nature.}$$

et de plus, en cas de convergence, la relation de Chasles généralisée est vérifiée :

$$\int_a^b f = \int_a^c f + \int_c^b f$$

Démonstration :

Soient $x \in I$, alors f est continue sur $I_x = [\min(a, c, x), \max(a, c, x)]$ car $I_x \subset I$. La relation de Chasles classique sur une fonction f continue s'applique alors :

$$\int_a^x f = \int_a^c f + \int_c^x f$$

la quantité $\int_a^c f$ est une constante indépendante de x . Ainsi, le passage à la limite quand $x \rightarrow b$ donne le résultat escompté. \square

⚠ Remarque :

Cette propriété permettra de décomposer l'étude d'une intégrale sur des morceaux plus pratiques ou éventuellement déjà étudiés.

■ Exemple 12 :

On a vu que $\int_0^{+\infty} te^{-t^2} dt$ converge. Comme la fonction $t \mapsto te^{-t^2}$ est continue sur \mathbb{R} , on en déduit que $\int_a^{+\infty} te^{-t^2} dt$ converge pour tout $a \in \mathbb{R}$.

⚠ **CONFUSION**

Les intégrales ont peut être même nature mais attention, elles n'ont pas nécessairement même valeur. Par exemple :

$$\int_0^{+\infty} te^{-t^2} dt = \frac{1}{2} \neq \int_{-1}^{+\infty} te^{-t^2} dt (= \frac{1}{2e})$$

II-2 Intégrales faussement impropres

Propriété 79

Si

- f est continue sur $[a, b[$ où $a, b \in \mathbb{R}$ (c'est-à-dire que l'intervalle est **borné**)
- et que f est prolongeable par continuité en b ,

alors

- $\int_a^b f$ est convergente
- et $\int_a^b f = \int_a^b \tilde{f}$ où \tilde{f} est la prolongée de f en b

Démonstration :

Supposons que f soit continue sur $[a, b[$ et prolongeable par continuité en b . On rappelle que, dans ce cas, la prolongée \tilde{f} de f en b est définie par

$$\tilde{f} : t \in [a, b] \mapsto \begin{cases} f(t) & \text{si } t \in [a, b[\\ \lim_{x \rightarrow b, x < b} f(x) & \text{sinon} \end{cases}$$

La fonction \tilde{f} étant continue sur $[a, b]$, la fonction primitive

$$\tilde{F} : x \in [a, b] \mapsto \int_a^x \tilde{f}(t) dt$$

est bien définie et continue sur $[a, b]$.

Or, comme $\tilde{f}(t) = f(t)$ pour tout $t \in [a, b[$, on a également que

$$\tilde{F}(x) = \int_a^x f \quad \forall x \in [a, b[$$

La continuité de \tilde{F} en b signifie donc exactement que

$$\lim_{x \rightarrow b, x < b} \int_a^x f = \lim_{x \rightarrow b, x < b} \tilde{F}(x) = \tilde{F}(b) = \int_a^b \tilde{f}$$

□

Commentaires :

De cette propriété, on déduit que la notation est cohérente avec celle utilisée dans les cas traités pour les intégrales définies sur un segment.

Définition
Si f est continue sur $[a, b[$ et prolongeable en b par continuité, on dit que $\int_a^b f$ est *faussement impropre en b* (ou faussement impropre s'il n'y a pas de doute possible.)

■ **Exemple 13 :**

$\int_0^1 \frac{\sin t}{t} dt$ de même que $\int_0^1 \frac{\cos t - 1}{t^2} dt$ sont toutes deux faussement impropres :

$t \mapsto \frac{\sin t}{t}$ est prolongeable par continuité en 0 par 1 et $t \mapsto \frac{\cos t - 1}{t^2}$ est prolongeable par continuité en 0 par $-\frac{1}{2}$

BIEN VÉRIFIER QUE LA BORNE EST UN NOMBRE FINI
➤ Une intégrale ne peut être faussement impropre en l'infini.

■ **Exemple 14 :**

$\int_1^{+\infty} \frac{1}{t} dt$ diverge, alors que $\lim_{t \rightarrow +\infty} \frac{1}{t} = 0$.

En l'infini, il ne suffit donc pas de vérifier la convergence !

II-3 Définition d'une intégrale généralisée sur un intervalle ouvert

Définition
Soit $f :]a; b[\rightarrow \mathbb{R}$ continue, où $a \in \mathbb{R} \cup \{-\infty\}$ et $b \in \mathbb{R} \cup \{+\infty\}$. On dit que $\int_a^b f$ est *convergente*, s'il existe $c \in]a; b[$ tel que $\int_a^c f$ converge en a et $\int_c^b f$ converge en b . Dans ce cas, $\int_a^b f = \int_a^c f + \int_c^b f$. Sinon, on dit qu'elle *diverge*.

Remarque :

Si on veut démontrer la divergence, il faut ici montrer que **pour tout** $c \in]a; b[$, **l'une au moins** des intégrales $\int_a^c f$ ou $\int_c^b f$ diverge, ce qui n'est pas très pratique. Nous verrons donc ici un exemple du cas de convergence et verrons ensuite comment simplifier la considération de la divergence.

■ **Exemple 15 :**

$\int_{-\infty}^{+\infty} te^{-t^2} dt$ converge :

La fonction $t \mapsto te^{-t^2}$ est continue sur \mathbb{R} . (On a donc un problème en $+\infty$ et $-\infty$: cas de l'intervalle ouvert)

* On a déjà vu que $\int_0^{+\infty} te^{-t^2} dt$ converge (avec $\int_0^{+\infty} te^{-t^2} dt = \frac{1}{2}$).

* Étudions $\int_{-\infty}^0 te^{-t^2} dt$: De la même façon que pour la première, $\forall x \leq 0$, on a

$$\int_x^0 te^{-t^2} dt = - \int_0^x te^{-t^2} dt = - \left[-\frac{e^{-t^2}}{2} \right]_0^x = -\frac{1}{2} (1 - e^{-x^2}).$$

Ainsi, $\lim_{x \rightarrow -\infty} \int_x^0 te^{-t^2} dt = -\frac{1}{2}$ et donc $\int_{-\infty}^0 te^{-t^2} dt$ converge avec

$$\int_{-\infty}^0 te^{-t^2} dt = -\frac{1}{2}$$

En conclusion, $\int_{-\infty}^{+\infty} te^{-t^2} dt$ converge avec de plus

$$\int_{-\infty}^{+\infty} te^{-t^2} dt = \int_0^{+\infty} te^{-t^2} dt + \int_{-\infty}^0 te^{-t^2} dt = \frac{1}{2} - \frac{1}{2} = 0$$

Commentaires :

La propriété suivante permet de généraliser la propriété concernant la convergence, mais également de nous simplifier la technique lorsque l'on souhaite prouver la divergence.

Propriété 80

Soit $f :]a; b[\rightarrow \mathbb{R}$ continue, où $a \in \mathbb{R} \cup \{-\infty\}$ et $b \in \mathbb{R} \cup \{+\infty\}$

- Si $\int_a^b f$ est convergente, alors, **pour tout** $c \in]a; b[$, on a $\int_a^c f$ converge en a et $\int_c^b f$ converge en b . De plus, $\int_a^b f = \int_a^c f + \int_c^b f$
- S'il existe $c \in]a; b[$ tel que $\int_a^c f$ diverge en a ou $\int_c^b f$ diverge en b , alors $\int_a^b f$ diverge.

Démonstration :

i) Supposons que $\int_a^b f$ est convergente. Alors, par définition, il existe $c_0 \in]a; b[$ tel que $\int_a^{c_0} f$ et $\int_{c_0}^b f$ convergent. Soit maintenant $c \in]a; b[$. Par relation de Chasles généralisée, comme f est continue sur I qui contient a, c_0, c , on sait que

$$\int_a^c f \text{ est de même nature que } \int_a^{c_0} f.$$

et que

$$\int_a^c f = \int_a^{c_0} f + \int_{c_0}^c f$$

De même, $\int_c^b f$ est de même nature que $\int_{c_0}^b f$ et

$$\int_c^b f = \int_{c_0}^b f + \int_c^{c_0} f.$$

On en déduit que $\int_a^c f$ et $\int_c^b f$ convergent et que

$$\int_a^c f + \int_c^b f = \int_a^{c_0} f + \underbrace{\int_{c_0}^c f + \int_c^{c_0} f}_{=0} + \int_{c_0}^b f = \int_a^b f$$

ii) c'est exactement la négation de la propriété du i).

□

■ Exemple 16 :

Montrons que $\int_0^{+\infty} \frac{1}{t^2} dt$ diverge :

La fonction $f : t \mapsto \frac{1}{t^2}$ est continue sur $]0; +\infty[$. Elle admet donc un problème en 0 et en $+\infty$. (cas de l'intervalle ouvert)

★ Étude du problème en 0 : étudions par exemple $\int_0^1 \frac{1}{t^2} dt$:

On pose $x \in]0; +\infty[$. $\int_x^1 \frac{1}{t^2} dt = \left[-\frac{1}{t} \right]_x^1 = -1 + \frac{1}{x} \xrightarrow{x \rightarrow 0^+} +\infty$

Ainsi $\int_0^1 \frac{1}{t^2} dt$ diverge et donc $\int_0^{+\infty} \frac{1}{t^2} dt$ diverge.

⚠ IL N'Y A PAS DE FORME INDÉTERMINÉE DANS LA RELATION DE CHASLES

La propriété dit bien qu'il suffit que l'une des deux parties diverge pour avoir la divergence. Cela signifie en particulier que si les deux parties divergent, ce n'est **pas** une forme indéterminée...

II-4 Définition d'une intégrale généralisée sur une réunion d'intervalles ouverts

⚠ ERREUR CLASSIQUE :

Considérons $\int_{-1}^1 \frac{1}{x^2} dx$. Il y a une erreur dans le calcul ci-dessous :

$$\int_{-1}^1 \frac{1}{x^2} dx = \left[-\frac{1}{x} \right]_{-1}^1 = -2 < 0 \quad \text{incohérent avec le fait que } \frac{1}{x^2} \geq 0$$

L'erreur vient du fait que l'on a oublié de vérifier la continuité de la fonction sur $[-1, 1]$. On observe en effet que $x \mapsto \frac{1}{x^2}$ n'est pas continue en 0. Dans ce cas, il faut découper l'intégrale suivant le principe de la définition ci-dessous.

🌿 Définition

Soit $f :]a; c[\cup]c; b[\rightarrow \mathbb{R}$ continue, où $a \in \mathbb{R} \cup \{-\infty\}$, $c \in \mathbb{R}$ et $b \in \mathbb{R} \cup \{+\infty\}$, $a < c < b$.

On dit que $\int_a^b f$ est *convergente*, si $\int_a^c f$ converge en a et c et si $\int_c^b f$ converge en b et c . Sinon, on dit qu'elle *diverge*.

■ Exemple 17 :

$$\int_{-1}^1 \frac{1}{t^2} dt \text{ diverge :}$$

La fonction $t \mapsto \frac{1}{t^2}$ est continue sur $[-1, 0[$ et $]0, 1]$. Or, par calcul direct de primitive, on trouve que $\int_0^1 \frac{1}{t^2} dt$ diverge. Cela suffit pour affirmer que $\int_{-1}^1 \frac{1}{t^2} dt$ diverge.

 Notation :

Dans la suite du cours, I désignera un intervalle ou une réunion d'intervalles ouverts, fermés, ou semi-ouverts, qu'ils soient bornés ou non. Dans ce cas, $\int_I f$ désignera $\int_a^b f$ où $a = \min I$ et $b = \max I$.

 Remarque :

Cette notation ne pourra être utilisée que dans le cas où l'ordre des bornes n'a pas d'importance. Par exemple, pour $I = [0, 3]$, $\int_I f = \int_0^3 f$ n'est a priori pas égale à $\int_3^0 f$.

III Manipulation des intégrales impropres

III-1 Linéarité

Théorème 81 de linéarité de l'intégrale

Soient $f, g :]a, b[\rightarrow \mathbb{R}$ deux fonctions continues et $\lambda \in \mathbb{R}$.

Si $\int_a^b f$ et $\int_a^b g$ sont convergentes, alors $\int_a^b (f + \lambda g)$ est convergente et dans ce cas

$$\int_a^b (f + \lambda g) = \int_a^b f + \lambda \int_a^b g$$

 Si...ALORS...

La réciproque est fautive : il suffit de prendre n'importe quelle fonction f telle que

$$\int_I f \text{ est divergente. Par ex. } \int_0^{+\infty} 1 + (-1) dt = \int_0^{+\infty} 0 dt = 0$$

mais $\int_0^{+\infty} 1 dt + \int_0^{+\infty} (-1) dt$ est une forme indéterminée

III-2 IPP

Théorème 82 (Ipp d'une intégrale impropre)

Soient u, v deux fonctions de classe \mathcal{C}^1 sur l'intervalle $]a, b[$ ($a < b$) telles que

$$\lim_{x \rightarrow a^+} u(x)v(x) \text{ et } \lim_{x \rightarrow b^-} u(x)v(x) \text{ existent et sont finies}$$

On a alors : $\int_a^b u'v$ est de même nature que $\int_a^b uv'$

s'il y a convergence, on a

$$\int_a^b u'v = [uv]_a^b - \int_a^b uv'$$

où $[uv]_a^b$ est défini par

$$[u(t)v(t)]_a^b := \lim_{x \rightarrow b^-} u(x)v(x) - \lim_{x \rightarrow a^+} u(x)v(x).$$

Démonstration :

Soient $c, x \in]a, b[$. Comme u, v sont de classe \mathcal{C}^1 sur $[c, x]$ (ou $[x, c]$ suivant le sens de c, x , on peut appliquer l'Ipp classique :

$$\int_c^x u'(t)v(t) dt = [uv]_c^x - \int_c^x u(t)v'(t) dt$$

Le reste n'est que simple passage à la limite, tout d'abord vers b qui donne une première formule :

$$\int_c^b u'(t)v(t) dt = \lim_{x \rightarrow b^-} u(x)v(x) - u(c)v(c) - \int_c^b u(t)v'(t) dt$$

puis un passage à la limite vers a :

$$\int_c^a u'(t)v(t) dt = \lim_{x \rightarrow a^+} u(x)v(x) - u(c)v(c) - \int_c^a u(t)v'(t) dt$$

on multiplie par -1 , ce qui inverse les bornes :

$$\int_a^c u'(t)v(t) dt = u(c)v(c) - \lim_{x \rightarrow a^+} u(x)v(x) - \int_a^c u(t)v'(t) dt$$

puis on additionne les deux égalités en b, c et a, c qui, grâce à la relation de Chasles, donne le résultat annoncé. \square

 Remarque :

Si u et v sont continues en a ou b (par exemple en a) on sait que $\lim_{x \rightarrow a^+}$ existe. Il est donc inutile de vérifier les deux limites et on a

$$[u(t)v(t)]_a^b = \lim_{x \rightarrow b^-} u(x)v(x) - u(a)v(a).$$

■ Exemple 18 :

Convergence et valeur de $\int_0^{+\infty} te^{-t} dt$ par IPP :

Posons les fonctions $u, v \in \mathcal{C}^1([0; +\infty[)$ telles que :

$$\begin{aligned} u(t) &= t & v'(t) &= e^{-t} \\ u'(t) &= 1 & v(t) &= -e^{-t} \end{aligned}$$

• Étape 1 : On a

$$\lim_{x \rightarrow +\infty} u(x)v(x) = \lim_{x \rightarrow +\infty} -xe^{-x} = 0$$

Les intégrales $\int_0^{+\infty} te^{-t} dt$ et $-\int_0^{+\infty} e^{-t} dt$ ont donc même nature. • Étape 2 : L'intégrale $\int_0^{+\infty} e^{-t} dt$ est convergente vers 1 (exercice), ainsi, on a la

$$\text{convergence de } \int_0^{+\infty} te^{-t} dt$$

• Étape 3 : La valeur de $\int_0^{+\infty} te^{-t} dt$ est finalement donnée par la formule

$$\int_0^{+\infty} te^{-t} dt = [-te^{-t}]_0^{+\infty} + \int_0^{+\infty} e^{-t} dt = 0 + 1 = \boxed{1}$$



OUBLI RÉCURRENT

Énoncé à appliquer avec attention et en plusieurs étapes. Il est **nécessaire de** vérifier la convergence vers un réel de $\lim_{x \rightarrow b, x < b} u(x)v(x)$. En l'absence de cette vérification, deux problèmes peuvent se présenter :

1. La formule peut devenir une forme indéterminée. (cf ex11)
2. Les intégrales peuvent ne pas être de même nature. (cf ex12)

■ Exemple 19 :

On souhaite déterminer la nature de $\int_1^{+\infty} t \ln t dt$ en passant par l'IPP suivante :

$$\begin{aligned} u'(t) &= t & v(t) &= \ln t \\ u(t) &= \frac{t^2}{2} & v'(t) &= \frac{1}{t} \end{aligned}$$

où u, v sont deux fonctions \mathcal{C}^1 sur $[1; +\infty[$.

On a

$$\lim_{t \rightarrow +\infty} u(t)v(t) = \lim_{t \rightarrow +\infty} \frac{t^2}{2} \ln t = +\infty$$

et

$$\int_1^{+\infty} u(t)v'(t) dt = \int_1^{+\infty} \frac{t^2}{2} \frac{1}{t} dt = \int_1^{+\infty} \frac{t}{2} dt = +\infty$$

Cette dernière intégrale étant divergente.

L'expression "de droite" dans l'intégration par partie est donc une forme indéterminée...
On ne peut rien dire sur la convergence avec cette méthode!

■ Exemple 20 :

Considérons $\int_0^{+\infty} t dt$ et effectuons l'IPP suivante :

$$\begin{aligned} u(t) &= 1 & u'(t) &= 0 \\ v'(t) &= t & v(t) &= \frac{t^2}{2} \end{aligned}$$

$u, v \in \mathcal{C}^1([0; +\infty[)$. De plus :

$$\begin{aligned} \lim_{x \rightarrow +\infty} u(x)v(x) &= \lim_{x \rightarrow +\infty} \frac{x^2}{2} = +\infty \\ \int_0^{+\infty} uv' &= \int_0^{+\infty} t dt \quad \text{et} \quad \int_0^{+\infty} u'v = \int_0^{+\infty} 0 dt \end{aligned}$$

Si on ne tient pas compte de la divergence de $\lim_{x \rightarrow +\infty} u(x)v(x)$, on annonce à tort que $\int_0^{+\infty} t dt$ a la même nature que $\int_0^{+\infty} 0 dt$, qui est trivialement convergente. Cherchez l'erreur...

Commentaires :

Mais quelles solutions ?

Une possibilité est généralement de repasser aux IPP "classiques", c'est-à-dire sur des intervalles fermés où la fonction est continue. Pour l'exemple précédent :

■ Exemple 21 :

Soit $x > 0$. en passant par l'IPP suivante :

$$\begin{aligned} u'(t) &= t & v(t) &= \ln t \\ u(t) &= \frac{t^2}{2} & v'(t) &= \frac{1}{t} \end{aligned}$$

u, v sont deux fonctions \mathcal{C}^1 sur $[1; x]$. Alors

$$\begin{aligned} \int_1^x t \ln t dt &= \int_1^x u'(t)v(t) dt = \left[\frac{t^2}{2} \ln t \right]_1^x - \int_1^x \frac{t^2}{2} \frac{1}{t} dt \\ &= \frac{x^2}{2} \ln x - \int_1^x \frac{t}{2} dt \\ &= \frac{x^2}{2} \ln x - \frac{1}{2} \left[\frac{t^2}{2} \right]_1^x = \frac{x^2}{2} \ln x - \frac{x^2}{4} + \frac{1}{4} \end{aligned}$$

Il faut maintenant passer à la limite :

$$\frac{x^2}{2} \ln x - \frac{x^2}{4} + \frac{1}{4} = \underbrace{\frac{x^2}{4}}_{x \rightarrow +\infty \rightarrow +\infty} + \underbrace{\frac{(2 \ln x - 1)}{4}}_{x \rightarrow +\infty \rightarrow +\infty} + \frac{1}{4} \xrightarrow{x \rightarrow +\infty} +\infty$$

L'intégrale est divergente.

III-3 Changement de variable

III.3-a) Le théorème et son application

Théorème 83 : Changement de variable dans une intégrale impropre

Soient $\alpha, \beta \in \mathbb{R}$, I un intervalle, ainsi que f et φ deux fonctions telles que :

- $f : [a, b] \rightarrow \mathbb{R}$ soit continue
- $\varphi : [a, b] \rightarrow \mathbb{R}$ soit \mathcal{C}^1 et strictement monotone d'inverse notée ψ

Alors $\int_a^b f(t) dt$ et $\int_\alpha^\beta f(\psi(x))\psi'(x) dx$ sont de même nature, avec, en cas de convergence,

$$\int_a^b f(t) dt = \int_\alpha^\beta f(\psi(x))\psi'(x) dx$$

où α et β sont définis par

$$\alpha := \lim_{x \rightarrow a^+} \varphi(x) \quad ; \quad \beta := \lim_{x \rightarrow b^-} \varphi(x)$$

Démonstration :

Pour commencer, par monotonie de ψ et donc de ψ^{-1} , on sait que $\alpha = \lim_{x \rightarrow a^+} \psi^{-1}(x)$ ainsi que $\beta = \lim_{x \rightarrow b^-} \psi^{-1}(x)$ existe dans $\mathbb{R} \cup \{\pm\infty\}$.

Réciproquement, $\psi(x) \xrightarrow{x \rightarrow \alpha} a^+$ et $\psi(x) \xrightarrow{x \rightarrow \beta} b^+$.

★ Supposons $\int_a^b f$ converge. Montrons que l'intégrale de droite converge.

Soit $c, y \in]\alpha, \beta[$ (ou $c, t \in]\beta, \alpha[$). Par changement de variable sur un segment fermé :

$$\int_c^y f(\psi(x))\psi'(x) dx = \int_{\psi(c)}^{\psi(y)} f(t) dt$$

Or, par convergence de \int_a^b , et comme $\psi(y) \xrightarrow{y \rightarrow \beta} b$ on a alors

$$\lim_{y \rightarrow \beta} \int_c^y f(\psi(x))\psi'(x) dx = \int_{\psi(c)}^b f(t) dt$$

De même pour l'autre côté :

$$\lim_{y \rightarrow \alpha} \int_c^y f(\psi(x))\psi'(x) dx = \int_{\psi(c)}^a f(t) dt$$

D'où, par relation de Chasles, la convergence de l'intégrale de droite avec effectivement l'égalité demandée.

★ Supposons que l'intégrale de droite converge.

Le même principe s'applique dans l'autre sens, mais avec $c, y \in]a, b[$ et la formule de changement de variable :

$$\int_c^y f(t) dt = \int_{\psi^{-1}(c)}^{\psi^{-1}(y)} f(\psi(x))\psi'(x) dx$$

□

⚠ Remarque :

Avant de passer aux exemples, notons que l'énoncé peut être donné sous deux formes différentes :

"Montrer que l'intégrale [...] converge."

ou un peu plus poussée :

"Montrer que l'intégrale [...] converge et calculer sa valeur."

Le théorème de changement de variable peut s'utiliser pour l'une ou pour l'autre de ces questions, alors profitons-en pour voir sur les exemples suivants comment peut se rédiger la réponse dans ces deux types d'énoncés. (1^{er} exemple avec la seule convergence et 2^{ème} exemple avec également la valeur de l'intégrale.)

■ Exemple 22 :

Montrons que $\int_1^{+\infty} \frac{e^{-\frac{1}{t}}}{t^2} dt$ est une intégrale convergente grâce au changement de variable $x = -\frac{1}{t}$.

On pose le changement de variable $x = -\frac{1}{t}$. Le changement de variable est \mathcal{C}^1 tel que

$$\frac{dx}{dt} = \frac{1}{t^2} > 0$$

Il est donc strictement monotone, donc bijectif sur l'intervalle. On a de plus

$$t = -\frac{1}{x} \quad \text{et} \quad \frac{dt}{dx} = \frac{1}{x^2}$$

x s'applique aux bornes par limites :

t	1	$+\infty$
u	-1	0

L'intégrale est donc de même nature que

$$\int_{-1}^0 \frac{e^x}{(\frac{1}{x})^2} \left(\frac{1}{x^2}\right) dx = \int_{-1}^0 \underbrace{e^x}_{g(x)} dx$$

on obtient une fonction g qui est continue sur \mathbb{R} (et donc sur $[0, 1]$). L'intégrale est donc bien convergente, ainsi que par équivalence, l'intégrale de départ.

■ Exemple 23 :

Montrons que $\int_0^{+\infty} \frac{e^{-\frac{1}{t}}}{t^2} dt$ converge et vaut 1 par changement de variable $x = \frac{1}{t}$.

On pose le changement de variable $x = \frac{1}{t}$. Le changement de variable est \mathcal{C}^1 tel que

$$\frac{dx}{dt} = -\frac{1}{t^2} < 0$$

Il est donc strictement monotone, donc bijectif sur l'intervalle. On a de plus

$$t = -\frac{1}{x} \quad \text{et} \quad \frac{dt}{dx} = \frac{1}{x^2}$$

x s'applique aux bornes par limites :

t	0	$+\infty$
u	$+\infty$	0

Sous réserve de convergence de l'une des deux intégrales, par formule de changement de variable, on a

$$\int_0^{+\infty} \frac{e^{-\frac{1}{t}}}{t^2} dt = -\int_{+\infty}^0 \frac{e^{-x}}{\left(\frac{1}{x}\right)^2} \left(\frac{1}{x^2}\right) dx = \int_0^{+\infty} e^{-x} dx$$

on obtient une fonction g **qui est continue** sur \mathbb{R} (et donc sur $[0, 1]$). L'intégrale est donc bien convergente, ainsi que par équivalence, l'intégrale de départ.

De plus, on a déjà vu que l'intégrale obtenue est convergente, ainsi donc, par équivalence, l'intégrale de départ. De plus, par calcul, on a également vu que $\int_0^{+\infty} e^{-x} dx = 1$, ce qui donne bien le résultat souhaité.

Démonstration :

On pose le changement de variable $u = -t$ dans l'intégrale $\int_a^b f(t) dt$ et on obtient directement, par changement de variable, que

$$-\int_0^{-a} f(-u) du \text{ est de même nature que } \int_0^a f \text{ avec égalité en cas de convergence.}$$

- Si f est paire :

$$\int_{-a}^0 f(u) du \text{ est de même nature que } \int_0^a f$$

$$\text{et en cas de convergence } \int_{-a}^0 f(u) du = \int_0^a f$$

- Si f est impaire :

$$\int_{-a}^0 f(u) du \text{ est de même nature que } \int_0^a f$$

$$\text{et en cas de convergence } \int_{-a}^0 f(u) du = -\int_0^a f. \quad \square$$

Corollaire

Soit $a \in \bar{\mathbb{R}}$ et $f \in \mathcal{C}([-a; a])$ **paire ou impaire**.

Alors $\int_{-a}^a f$ converge ssi $\int_0^a f$ converge. De plus, en cas de convergence

- Si f est impaire, $\int_{-a}^a f(t) dt = 0$

- Si f est paire, $\int_{-a}^a f(t) dt = 2 \int_0^a f(t) dt$

Démonstration :

Si $\int_{-a}^a f$ converge alors il en va de même pour $\int_{-a}^0 f$ ainsi que $\int_0^a f$. De plus,

d'après le corollaire précédent, si $\int_0^a f$ converge, il en va de même de $\int_{-a}^0 f$ et donc, par suite, de $\int_{-a}^a f$. L'équivalence de convergence est donc vérifiée.

D'où, en cas de convergence, par relation de Chasles,

$$\int_{-a}^a f = \int_{-a}^0 f(t) dt + \int_0^a f$$

Le reste est conséquence immédiate des égalités du corollaire précédent. \square

III.3-b) Cas des fonctions paires et impaires

Propriété 84

Soit $a \in \bar{\mathbb{R}}$ et $f \in \mathcal{C}([-a; a])$. Si f est paire ou impaire, la nature de $\int_0^a f$ est la même que celle de $\int_{-a}^0 f$ avec, en cas de convergence :

- $\int_{-a}^0 f = \int_0^a f$ si f est paire

- $\int_{-a}^0 f = -\int_0^a f$ si f est impaire.

III-4 Bilan des principaux résultats

Soit f continue sur une réunion d'intervalles I ou sur un intervalle $[a, b[$, $b \in]a, +\infty[$. $\lambda \in \mathbb{R}$.

ACTION	HYPOTHÈSES	CONCLUSION
Chasles 1	$\int_a^c f$ est convergente $\int_c^b f$ est convergente	$\int_a^b f$ converge et $\int_a^b f = \int_a^c f + \int_c^b f$
Chasles 2	$\int_a^b f$ converge $c \in [a, b]$	$\int_a^c f$ est convergente $\int_c^b f$ est convergente et $\int_a^b f = \int_a^c f + \int_c^b f$
Faussement impropre	b est fini $\lim_b f$ existe	$\int_a^b f$ converge.
Linéarité	$\int_I f$ converge $\int_I g$ converge	$\int_I (f + \lambda g) = \int_I f + \lambda \int_I g$
IPP 1	u, v de classe \mathcal{C}^1 $\lim_{x \rightarrow b^-} u(x)v(x)$ est finie	$\int_a^b u'v$ est de même nature que $\int_a^b uv'$
IPP2	u, v de classe \mathcal{C}^1 $\lim_{x \rightarrow b^-} u(x)v(x)$ est finie $\int_a^b u'v$ converge	$\int_a^b uv'$ converge $\int_a^b u'v = [uv]_a^b - \int_a^b uv'$

Changement de variable	g continue sur $]a, b[$ $u \in \mathcal{C}^1(]a, b[)$ et u' strictement monotone	$\int_a^b g(u(t))u'(t) dt$ et $\int_{u(a)}^{u(b)} g(u) du$ sont de même nature
Changement de variable (suite)	g continue sur $]a, b[$ $u \in \mathcal{C}^1(]a, b[)$ et u' strictement monotone En cas de convergence	$\int_a^b g(u(t))u'(t) dt = \int_{u(a)}^{u(b)} g(u) du$

IV Fonctions positives et convergence absolue

Théorème 85 de positivité

Soient $a, b \in \mathbb{R}$ et $f : [a, b[\rightarrow \mathbb{R}$ tels que

- $a < b$ (**bornes croissantes**)
- f est continue sur $[a, b[$,
- $\int_a^b f$ converge
- $f \geq 0$ sur $[a, b[$,

alors

$$\int_a^b f \geq 0$$

Démonstration :

Le passage à la limite conserve les signes. \square

ORDRE DES BORNES

L'hypothèse de bornes croissantes est importante. En effet, si $a > b$ et que f est continue positive sur $[b, a[$ avec convergence de $\int_a^b f$, on a

$$\int_a^b f = - \int_b^a f \leq 0$$

Corollaire "Comparaison d'intégrales"

Soient $a, b \in \mathbb{R}$ et $f, g : [a, b[\rightarrow \mathbb{R}$ tels que

- $a < b$ (**bornes croissantes**)
- f, g sont continues et d'intégrales convergentes sur $[a, b[$,
- $f \leq g$ sur $[a, b[$,

alors

$$\int_a^b f \leq \int_a^b g$$

Démonstration :

On applique le théorème précédent à $\int_a^b (g - f)(t) dt$, puis la linéarité des intégrales convergentes. \square

Théorème 86 de positivité ... la suite

Pour $a < b$, soit $f : [a, b[\rightarrow \mathbb{R}$ une fonction telle que

- f est continue sur $[a, b[$,
- f est de signe constant sur $[a, b[$,
- $\int_a^b f = 0$,

alors f est nulle sur $[a; b[$.

Démonstration :

Supposons que f soit non nulle sur $[a, b[$. Alors il existe $x_0 \in [a, b[$ tel que $f(x_0) > 0$. Par continuité de f , il existe un voisinage fermé $I \subset [a; b[$ autour de x_0 sur lequel

$$f(x) \geq \frac{f(x_0)}{2} \quad \forall x \in I$$

On note $I = [\alpha, \beta]$. Par relation de Chasles, on a

$$0 = \int_a^b f = \int_a^\alpha \underbrace{f(t)}_{\geq 0} dt + \int_\alpha^\beta \underbrace{f(t)}_{\geq \frac{f(x_0)}{2}} dt + \int_\beta^b \underbrace{f(t)}_{\geq 0} dt \geq (\beta - \alpha) \frac{f(x_0)}{2}$$

ce qui est contradictoire avec le fait que $f(x_0)$ est strictement positif. \square



Remarque :

Ici aussi, la deuxième hypothèse est capitale. Il existe des fonctions non nulles d'intégrale nulle. (les aires au dessous et au dessus de l'axe des abscisses se compensent).

Dans la suite, on se donne $a \in \mathbb{R}, b \in \mathbb{R} \cup \{+\infty\}, a < b$

Proposition 87

On se donne $f : [a; b[\rightarrow \mathbb{R}$ continue **positive** sur $[a, b[$. Alors

$$\int_a^b f \text{ converge} \Leftrightarrow F : x \in [a; b[\mapsto \int_a^x f \text{ est majorée.}$$

Démonstration :

Avant tout F est une primitive de f , qui est positive par hypothèse. Ainsi,

$$F \text{ est croissante } (F' = f \geq 0).$$

Or, on sait qu'une fonction croissante sur $[a, b[$ cv en b ssi elle est majorée. C'est exactement la traduction de la proposition. \square

Théorème 88

On se donne $f, g : [a; b[\rightarrow \mathbb{R}$ continues **positives** sur $[a, b[$ telles que

$$0 \leq f(x) \leq g(x) \quad \forall x \in [a; b[$$

$$\text{Alors } \int_a^b g(t) dt \text{ cv} \Rightarrow \int_a^b f \text{ cv et } 0 \leq \int_a^b f \leq \int_a^b g(t) dt$$

$$\int_a^b f \text{ divergente} \Rightarrow \int_a^b g(t) dt \text{ divergente}$$

Démonstration :

Les fonctions $G : x \in [a; b[\mapsto \int_a^x g(t) dt$, $F : x \in [a; b[\mapsto \int_a^x f$ sont toutes deux continues et croissantes sur $[a; b[$ car de dérivée positive sur $[a; b[$. De plus, Comme G admet une limite en b , on a

$$G(x) \leq \lim_{x \rightarrow b} \int_a^x g(t) dt = \int_a^b g(t) dt \quad \forall x \in [a; b[$$

Par comparaison des intégrales, on a alors

$$0 \leq F(x) = \int_a^x f \leq \int_a^x g(t) dt \leq \int_a^b g(t) dt \quad \forall x \in [a; b[$$

F est donc une fonction croissante et majorée par $\int_a^b g(t) dt$. On en déduit que

$$\lim_{x \rightarrow b} F(x) \text{ existe} \quad \text{et} \quad 0 \leq \lim_{x \rightarrow b} F(x) \leq \int_a^b g(t) dt$$

c'est-à-dire $0 \leq \int_a^x f \leq \int_a^b g(t) dt$ □

? Exercice 1

Montrer que $\int_1^{+\infty} \frac{\cos t - 1}{t^2} dt$ est convergente.

Solution

La fonction $f \mapsto \frac{\cos t - 1}{t^2}$ est continue sur $[1, +\infty[$. Il y a donc potentiellement un problème en $+\infty$.

On observe qu'elle est négative sur tout l'intervalle considéré. On peut donc par exemple poser $g = -f \geq 0$ pour répondre à cette question. Or, dans ce cas, on a

$$0 \leq g(t) = \frac{1 - \cos t}{t^2} \leq \frac{2}{t^2}$$

Or l'intégrale $\int_1^{+\infty} \frac{1}{t^2} dt$ converge (déjà vu) et par théorème de comparaison des intégrales de fonctions positives, $\int_1^{+\infty} g$ converge, ce qui entraîne par la même celle de $\int_1^{+\infty} f$.

★ Conclusion : l'intégrale $\int_1^{+\infty} f$ est convergente car convergente (en $+\infty$).

Remarque : Si c'était $\int_0^{+\infty} \frac{\cos t - 1}{t^2} dt$, on pourrait montrer que l'intégrale est faussement impropre en 0 grâce à un équivalent usuel.

Théorème 89 de convergence absolue

On se donne $f : [a; b[\rightarrow \mathbb{R}$ continue sur $[a, b[$ (où $a < b$).

$$\int_a^b |f| \text{ converge} \quad \Rightarrow \quad \begin{cases} \int_a^b f \text{ converge} \\ \left| \int_a^b f \right| \leq \int_a^b |f(t)| dt \end{cases}$$

Démonstration :

(convergence admise) L'inégalité s'en déduit ensuite simplement par

$$-|f(t)| \leq f(t) \leq |f(t)|$$

puis

$$-\int_a^b |f(t)| dt \leq \int_a^b f \leq \int_a^b |f(t)| dt$$

D'où

$$\left| \int_a^b f \right| \leq \left| \int_a^b |f| \right| = \int_a^b |f(t)| dt$$

□

? Exercice 2

Pour $a \in \mathbb{R}^*$, $b > 0$, montrer que $\int_0^{+\infty} \cos(at)e^{-bt} dt = \frac{b}{a^2 + b^2}$. (avec IPP)

Solution

On note I l'intégrale en question.

★ Tout d'abord, montrons que l'intégrale converge. On remarque pour ce faire tout d'abord que $t \mapsto \cos(at)e^{-bt}$ est continue sur $[0, +\infty[$. L'unique problème se situe donc en $+\infty$.

Ensuite, pour tout $t \geq 0$, on a

$$0 \leq \left| \cos(at)e^{-bt} \right| \leq e^{-bt}$$

Or on peut montrer (exercice ou plus tard référence à la fin du chapitre) que $\int_0^{+\infty} e^{-bt} dt$ est convergente car $b > 0$. En conclusion, par théorème de comparaison des intégrales de fonctions positives et par convergence absolue, l'intégrale de départ converge.

★ Effectuons maintenant une première intégration par partie avec deux fonctions $u, v \in \mathcal{C}^1(\mathbb{R})$:

$$\begin{aligned} u(t) &= \cos(at) & u'(t) &= -a \sin(at) \\ v'(t) &= e^{-bt} & v(t) &= -\frac{1}{b} e^{-bt} \end{aligned}$$

On a

$$u(t)v(t) = -\frac{1}{b} \underbrace{\cos(at)}_{\text{borné}} \underbrace{e^{-bt}}_{\xrightarrow[t \rightarrow +\infty]{} 0 \text{ car } b > 0} \xrightarrow[t \rightarrow +\infty]{} 0$$

Ainsi, comme cette limite est finie et que l'intégrale de départ est convergente, on peut appliquer l'IPP et on a

$$I = \underbrace{[u(t)v(t)]_0^{+\infty}}_{0 - u(0)v(0) = \frac{1}{b}} - \frac{a}{b} \int_0^{+\infty} \sin(at)e^{-bt} dt$$

où on sait que l'intégrale $\int_0^{+\infty} \sin(at)e^{-bt} dt$ est convergente.

De la même manière avec un second changement de variable \mathcal{C}^1

$$u(t) = \sin(at) \quad u'(t) = a \cos(at)$$

$$v'(t) = e^{-bt} \quad v(t) = -\frac{1}{b}e^{-bt}$$

les mêmes arguments donnent lieu à l'obtention de la formule avec intégrales convergentes :

$$\int_0^{+\infty} \sin(at)e^{-bt} dt = \underbrace{[u(t)v(t)]_0^{+\infty}}_{\substack{0-u(0)v(0)=0 \\ a^2 - u(0)v(0)=0}} + \frac{a}{b} I$$

D'où, en réinjectant :

$$I = \frac{1}{b} - \frac{a^2}{b^2} I$$

ce qui, en isolant I donne au final :

$$I = \frac{b}{a^2 + b^2}$$

Théorème 90 sur les fonctions équivalentes

On se donne $f, g : [a, b[\rightarrow \mathbb{R}$ continues sur $[a, b[$ telles que

$$f \sim_b g.$$

avec l'une des deux fonctions de signe constant au voisinage de b . Alors

$$\int_a^b f \quad \text{est de même nature que} \quad \int_a^b g$$

Démonstration :

Quitte à échanger les fonctions, on peut supposer que $\int_a^b g$ converge.

Quitte maintenant à prendre $-g$ ou $-f$, on peut supposer que $g > 0$ (ou $f > 0$) au voisinage de b . Comme $f \sim_b g$, il en va de même pour l'autre.

Par équivalence, on a

$$f = hg \quad \text{avec} \quad \lim_b h = 1$$

Par limite, il existe ensuite $B \in [a, b[$ tel que

$$\forall x \in [B, b[, \quad \text{on a} \quad \frac{1}{2} \leq h \leq \frac{3}{2}$$

D'où, en multipliant par $g \geq 0$:

$$0 \leq \frac{1}{2}g \leq f \leq \frac{3}{2}g$$

Le théorème de comparaison des intégrales permet ensuite de conclure. \square

■ Exemple 24 :

Montrons que $\int_0^{+\infty} e^{-t-e^{-t}} dt$ converge :

$t \mapsto e^{-t-e^{-t}}$ est continue sur $[0, +\infty[$. Par continuité de la fonction exponentielle,

$$\frac{e^{-t-e^{-t}}}{e^{-t}} = e^{-e^{-t}} \xrightarrow{t \rightarrow +\infty} e^0 = 1$$

on a donc

$$e^{-t-e^{-t}} \underset{+\infty}{\sim} e^{-t}$$

Or, l'intégrale $\int_0^{+\infty} e^{-t} dt$ converge (déjà fait précédemment !), donc, par équivalence,

$\int_0^{+\infty} e^{-t-e^{-t}} dt$ converge.



CONFUSION

Ceci ne signifie en aucun cas que les intégrales sont ensuite égales (ou "équivalentes" !)

■ Exemple 25 :

On pose $f(t) = \frac{1+t}{t^2}$. On a

$$f(t) \underset{+\infty}{\sim} \frac{1}{t^2}$$

Or, pour tout $x \geq 1$, on a *calcul laissé au lecteur*

$$\int_1^x f = \frac{3}{2} - \frac{1}{x} - \frac{1}{2x^2}, \quad \text{et} \quad \int_1^x \frac{1}{t^2} dt = 1 - \frac{1}{x}$$

On voit bien que si $x \rightarrow \infty$, on a

$$\int_1^x f(t) dt \sim \frac{3}{2} \quad \text{et} \quad \int_1^x \frac{1}{t^2} dt \sim 1$$

D'où

$$\int_1^x f(t) dt \not\sim \int_1^x \frac{1}{t^2} dt$$

et de plus

$$\int_1^{+\infty} f \neq \int_1^{+\infty} \frac{1}{t^2} dt$$

V Exemples fondamentaux

Théorème 91

$$\int_0^{+\infty} e^{-\lambda t} dt \text{ cv en } +\infty \text{ ssi } \lambda > 0. \text{ Dans ce cas, on a } \int_0^{+\infty} e^{-\lambda t} dt = \frac{1}{\lambda}$$

Démonstration :

$t \mapsto e^{-\lambda t}$ est continue sur $[0; +\infty[$. Le problème est donc seulement en $+\infty$.

★ Cas $\lambda \in \mathbb{R}^*$: Soit $x > 0$. Alors

$$\int_0^x e^{-\lambda t} dt = \left[\frac{e^{-\lambda t}}{-\lambda} \right]_0^x = \frac{e^{-\lambda x} - 1}{-\lambda} \xrightarrow{x \rightarrow +\infty} \begin{cases} \frac{-1}{-\lambda} & \text{si } \lambda > 0 \\ +\infty & \text{si } \lambda < 0 \end{cases}$$

★ Cas $\lambda = 0$: Soit $x > 0$, alors

$$\int_0^x e^{-\lambda t} dt = \int_0^x e^0 dt = \int_0^x 1 dt = x \xrightarrow{x \rightarrow +\infty} +\infty$$

En conclusion, l'intégrale est convergente en $+\infty$ si et seulement si $\lambda > 0$ et dans

ce cas, on a $\int_0^{+\infty} e^{-\lambda t} dt = \frac{1}{\lambda}$. \square

Théorème 92

$$\text{Soit } n \in \mathbb{N}. \int_0^{+\infty} t^n e^{-\lambda t} dt \text{ cv ssi } \lambda > 0. \text{ Dans ce cas, } \int_0^{+\infty} t^n e^{-\lambda t} dt = \frac{n!}{\lambda^{n+1}}$$

Démonstration :

• On pose $\lambda \leq 0$.

On peut considérer que $n \geq 1$ (sinon on se ramène au théorème précédent). Ainsi, pour tout $t \geq 1$, on a

$$t^n e^{-\lambda t} \geq e^{-\lambda t} \geq 0 \quad \forall t \geq 1$$

Or, $\int_1^{+\infty} e^{-\lambda t} dt$ diverge (thm précédent). Par théorème de comparaison des intégrales de fonctions positives, $\int_1^{+\infty} t^n e^{-\lambda t} dt$ (et donc $\int_0^{+\infty} t^n e^{-\lambda t} dt$) diverge.

• On pose $\lambda > 0$.

On pose $f(t) = t^n e^{-\lambda t}$ pour tout $t \geq 0$. On a f continue sur $[0, +\infty[$. Le seul problème est donc en $+\infty$.

Montrons par récurrence sur $n \in \mathbb{N}$ que I_n converge et que $I_n = \frac{n}{\lambda} I_{n-1}$ pour tout $n \in \mathbb{N}$.

★ Initialisation : Le théorème précédent nous dit que $I_0 = \int_0^{+\infty} e^{-\lambda t} dt$ converge et on a bien

$$I_0 = \frac{1}{\lambda} = \frac{0!}{\lambda^{0+1}}.$$

★ Hérité : De $n-1$ à n : On pose l'IPP suivante, avec $u, v \in \mathcal{C}^1([0, +\infty[)$:

$$u = t^n \quad u' = nt^{n-1}$$

$$v' = e^{-\lambda t} \quad v = \frac{e^{-\lambda t}}{-\lambda}$$

Soit $x \in \mathbb{R}^+$. Comme

$$[uv]_0^x = \left[t^n \frac{e^{-\lambda t}}{-\lambda} \right]_0^x = x^n \frac{e^{-\lambda x}}{-\lambda} \xrightarrow{x \rightarrow +\infty} 0$$

, l'intégrale $I_n = \int_0^{+\infty} uv'$ est de même nature que

$$\int_0^{+\infty} u'v = -\frac{n}{\lambda} \int_0^{+\infty} t^{n-1} e^{-\lambda t} dt = -\frac{n}{\lambda} I_{n-1}$$

qui est justement convergente par HR. Ainsi, I_n converge d'une part, et d'autre part, on peut calculer sa valeur par IPP :

$$I_n = \int_0^{+\infty} u'v = [uv]_0^{+\infty} + \frac{n}{\lambda} I_{n-1} = 0 + \frac{n}{\lambda} \cdot \frac{(n-1)!}{\lambda^{n-1+1}} = \frac{n!}{\lambda^{n+1}}$$

\square

CQFD

Théorème 93

$$\int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt = \sqrt{2\pi}$$

Démonstration :

admise \square

Par parité de la fonction on obtient également :

Corollaire

$$\int_0^{+\infty} e^{-\frac{t^2}{2}} dt = \sqrt{\frac{\pi}{2}}$$

? Exercice 3

Montrer que $\int_{-\infty}^{+\infty} e^{-t^2+2t-2} dt$ converge et déterminer sa valeur.

(On pensera à se rapprocher, par un changement de variable opportun, d'une intégrale de type $\alpha \int_{-\infty}^{+\infty} e^{-\frac{u^2}{2}} du$. On trouve $\frac{\sqrt{\pi}}{e}$.)

Solution

On note $f(t) = e^{-t^2+2t-2}$. f est continue sur \mathbb{R} . Il y a donc potentiellement un problème en $\pm\infty$.

On remarque que

$$f(t) = e^{-(t^2-2t+2)} = e^{-((t-1)^2+1)} = e^{-1} e^{-(t-1)^2} = e^{-1} e^{-\frac{1}{2}(\sqrt{2}(t-1))^2}$$

On pose donc le changement de variable

$$u = \sqrt{2}(t-1)$$

On a alors, sur l'intervalle considéré : $\frac{du}{dt} = \sqrt{2} > 0$.

Le changement de variable est donc strictement monotone. De surcroît, on a

$$dt = \frac{1}{\sqrt{2}} du$$

avec le changement de bornes par limites :

t	$-\infty$	$+\infty$
u	$-\infty$	$+\infty$

Ainsi, par changement de variable, on obtient, sous réserve de convergence de l'une des deux intégrales :

$$\int_{-\infty}^{+\infty} e^{-t^2+2t-2} dt = e^{-1} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}u^2} \frac{1}{\sqrt{2}} du$$

Or on sait d'après le cours que $\int_{-\infty}^{+\infty} e^{-\frac{1}{2}u^2} du$ est convergente et que

$$\int_{-\infty}^{+\infty} e^{-\frac{1}{2}u^2} du = \sqrt{2\pi}.$$

Ainsi, l'intégrale de départ est convergente et on obtient en réinjectant :

$$\int_{-\infty}^{+\infty} e^{-t^2+2t-2} dt = e^{-1} \sqrt{\pi}$$

I Définition et quelques propriétés

Définition
 Une *équation différentielle scalaire autonome d'ordre 1* est une équation différentielle du type :

$$y' = F(y)$$
 avec y l'inconnue qui est une fonction dérivable et F une fonction continue. On appelle (E) cette équation dans la suite du cours. Une solution de cette équation est une fonction $y \in C^1$ sur un **intervalle** $I \subset \mathbb{R}$ telle que :

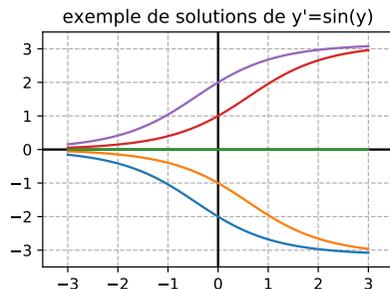
$$\forall t \in I, \quad y'(t) = F(y(t)).$$

Remarque :
 L'ensemble \mathbb{R}^* n'est par exemple pas un intervalle, mais $]0, +\infty[$ et $]-\infty; 0[$ sont deux intervalles distincts.

Commentaires :
 Dans le programme est stipulé que : "Aucune théorie générale ne doit être faite. Toute étude devra être entièrement guidée".

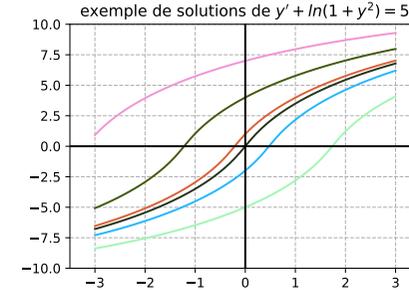
■ **Exemple 1 :**
 L'équation

$$y' = \sin(y)$$
 d'inconnue $y \in C^1(\mathbb{R})$ est une équation autonome d'ordre 1. Voici quelques représentations graphiques de solutions (obtenues avec la fonction `solve_ivp` de Python – pour l'utilisation : cf TP d'informatique)



■ **Exemple 2 :**
 L'équation

$$y' + \ln(1 + y^2) = 5$$
 d'inconnue $y \in C^1(\mathbb{R})$ est une équation autonome d'ordre 1. Voici quelques représentations graphiques de solutions :



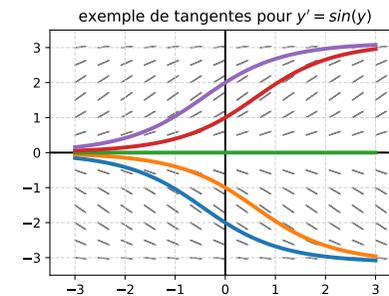
Remarque :
Deux courbes de solutions qui passent par une même ordonnée ont des tangentes parallèles.

En effet, notons k cette ordonnée commune. Dès qu'une courbe passe par cette ordonnée (en une abscisse t), on a $y(t) = k \in \mathbb{R}$. La pente de la tangente à la courbe est

$$y'(t) = F(y(t)) = F(k).$$

qui est donc indépendant de y et de l'abscisse t (ne dépend que de l'ordonnée k .)

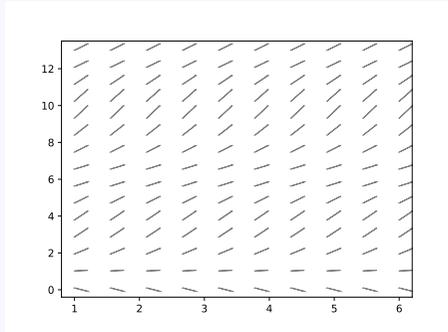
Par exemple, représentons les pentes des tangentes pour pour l'équation $y' = \sin(y)$ d'inconnue y :



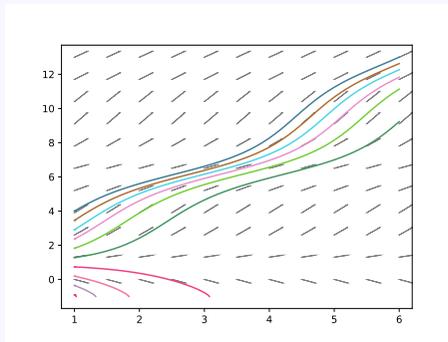
Ceci permet de conjecturer instinctivement que deux courbes de solutions ne se recoupent jamais. (à méditer.)

? Exercice 1

Voici ci-dessous un exemple de tangentes obtenues pour une équation autonome. Esquisser plusieurs courbes de solutions possibles correspondant à ce graphique. (On prendra notamment comme point de départ des courbes des ordonnées inférieures à 1 ainsi que supérieures à 2.)



Solution



Définition et Proposition

Si a est un réel tel que $F(a) = 0$ alors $y : t \mapsto a$ est une solution de (E) . On appelle *équilibre* ou *état stationnaire* les solutions constantes de (E) .

Démonstration :

En réinjectant dans l'expression $y' - F(y)$, on constate que pour $y : t \mapsto a$, on obtient

$$y' - F(y) = 0 - F(a)$$

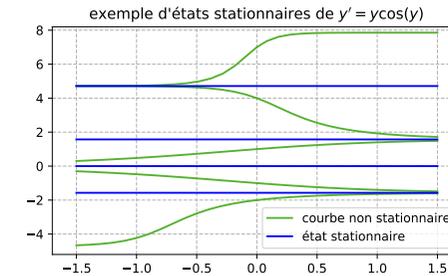
Ainsi, si $F(a) = 0$, on a bien

$$y' - F(y) = 0$$

et $y : t \mapsto a$ est bien une solution constante de l'équation différentielle. \square

■ Exemple 3 :

Les fonctions constantes $y_0 : t \mapsto 0$ et $y_1 : t \mapsto \frac{\pi}{2}$ sont des états stationnaires de l'équation $y' = y \cos y$. (exemples ci-dessous)



Proposition 94

Soit $k \in \mathbb{R}$. On note $J = \{u - k \mid u \in I\}$. Alors :

- $\forall t \in J$, on a $t + k \in I$
- et si y est une solution de l'équation (E) sur I , alors $\varphi : t \in J \mapsto \varphi(t + k)$ avec k réel est solution de (E) sur J .

Commentaires :

Graphiquement, ceci signifie que toute translation horizontale (à gauche ou à droite) d'une courbe est encore la courbe d'une solution

Démonstration :

Avec les notations de la proposition, pour tout $t \in J$, il existe $u \in I$ tel que

$$t = u - k$$

d'où

$$t + k = u \in I$$

Ainsi, $\varphi : t \in J \mapsto y(t + k)$ est bien définie et dérivable et de plus

$$\varphi'(t) = y'(t + k)$$

Or y est solution, donc, pour tout $u \in I$, on a

$$y'(u) = F(y(u))$$

d'où ici

$$\varphi'(t) = y'(t + k) = F(y(t + k)) = F(\varphi(t))$$

i.e. φ est solution de l'équation \square

? Exercice 2

On considère l'équation autonome $y' = \frac{1}{y}$.

1. Montrer que y vérifie cette équation ssi

$$\exists c \in \mathbb{R} \mid y^2(x) = 2x + 2c$$

2. En déduire que l'ensemble des solutions de l'équation est

$$\{y :]-c, +\infty[\mapsto \sqrt{2}\sqrt{c+x}\}$$

3. Vérifier que si y est solution sur I , pour tout $k \in \mathbb{R}$, on a, $\varphi : t \mapsto y(t+k)$ est encore une solution sur $J = I - k$.

Solution

1. On observe que $y' = \frac{1}{y} \Leftrightarrow y'y = 1 \Leftrightarrow \frac{1}{2}(2y^2)' = 1$.
Ceci équivaut donc à

$$\exists c \in \mathbb{R} \mid y^2(x) = 2x + c$$

2. Parmi les possibilités définies précédemment, il nous faut $2x + 2c \geq 0$ sur tout l'ensemble de définition afin que la racine soit bien définie et $2x + 2c \neq 0$ afin que y ne s'annule pas et soit bien dérivable, d'où les solutions

$$\{y :]-c, +\infty[\mapsto \sqrt{2}\sqrt{c+x}\}$$

3. Si $k \in \mathbb{R}$ et $y :]-c, +\infty[\mapsto \sqrt{2}\sqrt{c+x}$ une solution, on a $\varphi : t \in J \mapsto y(t+k) = \sqrt{2}\sqrt{c+k+x} = \sqrt{2}\sqrt{C+x}$ qui est bien définie et dérivable et est bien solution.

II Résolutions graphiques : méthode d'Euler

On rappelle (cf TP d'informatique) que l'on peut obtenir graphiquement des approximations de courbes de solutions grâce à la méthode d'Euler, que vous devez connaître (*mais dont l'énoncé est censé être rappelé en cas d'utilisation.*)

Théorème 95 Méthode d'Euler

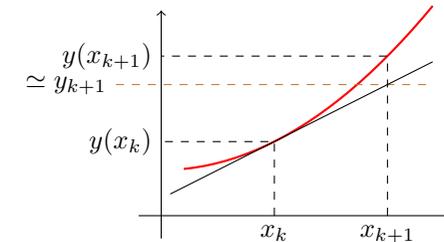
Soient y une solution de (E) sur un intervalle $[\alpha, \beta]$ et n un entier naturel non nul. Soit ensuite $(y_k)_{k \in \llbracket 0, n \rrbracket}$ la suite finie définie par :

$$\begin{cases} y_0 = y(\alpha) \\ h = \frac{\beta - \alpha}{n} \\ y_{k+1} = y_k + hF(y_k) \quad \forall k \in \llbracket 0, n-1 \rrbracket \end{cases}$$

Alors, pour tout $k \in \llbracket 0, n \rrbracket$, si n est assez grand, on a :

$$y_k \simeq y(x_k)$$

qui repose sur l'approximation faite grâce à la définition de la dérivée et qui se traduit graphiquement par le schéma suivant : Ce qui se traduit graphiquement de la manière suivante :



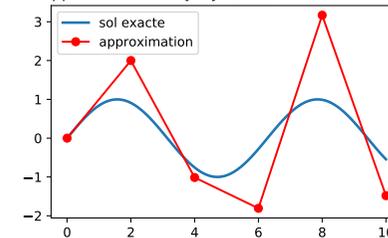
■ Exemple 4 :

Avec l'équation $y' + y = \cos x + \sin x$, sur un intervalle $[\alpha, \beta]$, on a, pour $h = \frac{\beta - \alpha}{n}$,

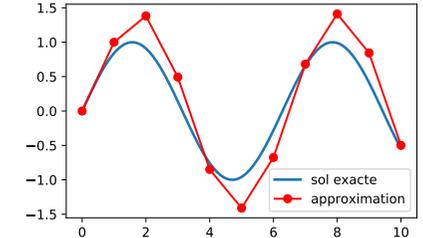
$$y_{k+1} = y_k + h(\cos x_k + \sin x_k - y_k) \quad \forall k \in \llbracket 0, n-1 \rrbracket$$

et où on constate que les approximations deviennent en effet logiquement meilleures quand le nombre n de points augmente :

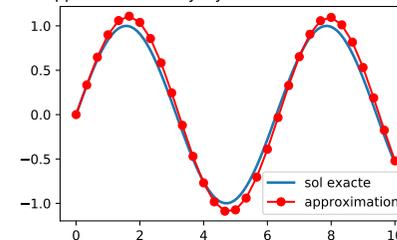
Approximation de $y'+y=\sin x + \cos x$ avec $n=5$



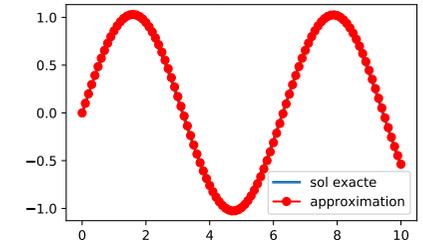
Approximation de $y'+y=\sin x + \cos x$ avec $n=10$



Approximation de $y'+y=\sin x + \cos x$ avec $n=30$



Approximation de $y'+y=\sin x + \cos x$ avec $n=100$



III Exemples de résolutions en dynamique des populations

▲ NOMBRE D'INDIVIDUS

⚡ Dans tous les exemples de cette section, on note $N(t)$ une fonction positive que l'on suppose dérivable et qui désigne le nombre d'individus dans une population donnée, à l'instant t (Le nombre d'individus peut être non entier.)

III-1 Modèle de Malthus

Explications : Dans ce modèle attribué à Thomas Malthus, on considère une population isolée et sans contrainte, au sens suivant :

- espace et nourriture illimités,
- absence de prédateur,
- résistance aux maladies, etc. ...

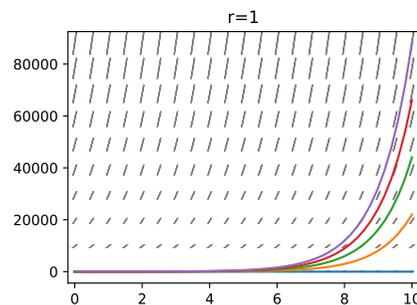
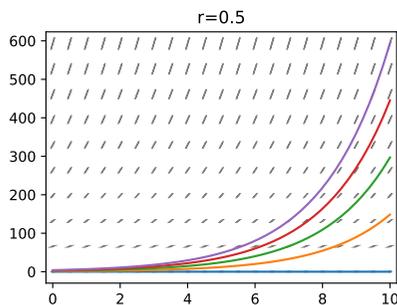
On note r le "taux de croissance" de l'espèce. On veut traduire le fait que plus il y a d'individus, plus ils se reproduisent vite.

🌿 Définition

Dans le modèle élémentaire de Malthus, on estime que

$$N'(t) = rN(t) \quad \forall t \geq 0$$

On obtiendra un graphique du type suivant avec $r = 0.5$ et $r = 1$:



? Exercice 3

Montrer que les solutions de cette équation différentielle sont les fonctions

$$N : t \geq 0 \mapsto N(0)e^{rt}$$

Solution

Il s'agit d'une équation différentielle homogène de type $y' = \text{"constante"} \cdot y$. D'après le cours, on connaît ses solutions. Elles s'écrivent

$$N(t) = \lambda e^{rt}$$

avec $\lambda \in \mathbb{R}$. En prenant $t = 0$, on trouve

$$N(0) = \lambda e^0 = \lambda$$

D'où la solution.

III-2 Modèle Logistique

Explications : Dans ce modèle attribué à Verhulst en 1836, on considère une population isolée du même type que le modèle de Malthus, mais avec une contrainte environnementale, au sens suivant :

- nourriture illimités, absence de prédateur, résistance aux maladies, etc. ...
- mais espace limité.

On note r le "taux de croissance" de l'espèce et K la capacité d'accueil du milieu.

🌿 Définition

Dans ce modèle, on estime que $N'(t) = r \left(1 - \frac{N(t)}{K}\right) N(t) \quad \forall t \geq 0$

Commentaires :

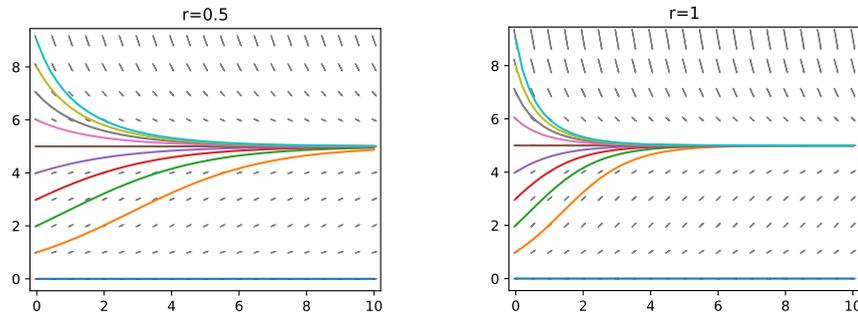
- Quand N est inférieur à la capacité du milieu, on a $1 - \frac{N(t)}{K} > 0$. Ainsi $N' > 0$ et la population augmente.
- Quand N est supérieur à la capacité du milieu, on a $1 - \frac{N(t)}{K} < 0$. Ainsi $N' < 0$ et la population diminue.
- En revanche, quand N est proche de la capacité d'accueil du milieu (par valeurs inférieures), la croissance ralentit de plus en plus. On a

$$\text{Si } N(t) \simeq K \text{ alors } 1 - \frac{N(t)}{K} \simeq 0 \text{ et donc } N' \simeq 0$$

La population stagne

- Tant que N est loin de K par valeurs inférieures, $1 - \frac{N(t)}{K} \simeq 1$, on a $N' \simeq rN$. Le modèle est proche du modèle de Malthus et tout se passe à peu près comme si l'espace était illimité.

On obtiendra un graphique du type suivant avec $r = 0.5$ et $r = 1$:



? Exercice 4

1. Montrer que $N : t \mapsto 0$ et $N : t \mapsto K$ sont des états stationnaires du modèle.
2. Supposons que N ne s'annule pas. Poser pour tout $t \geq 0$, $z(t) = \frac{1}{N(t)}$. Montrer que

$$N'(t) = r \left(1 - \frac{N(t)}{K} \right) N(t) \Leftrightarrow z' = -rz + \frac{r}{K}$$

3. En déduire que solutions du modèle logistique sont :

$$N : t \geq 0 \mapsto \frac{K}{1 + \left(\frac{K}{N(0)} - 1 \right) e^{-rt}}$$

4. Justifier qu'il n'y a pas d'autres états stationnaires.

Solution

1. Trivial car on obtient bien $0 = 0$ en réinjectant.
2. On a $z' = -\frac{N'}{N^2}$. Ainsi,

$$\begin{aligned} N'(t) = r \left(1 - \frac{N(t)}{K} \right) N(t) &\Leftrightarrow \frac{N'(t)}{N^2(t)} = \frac{r}{N(t)} \left(1 - \frac{N(t)}{K} \right) \\ &\Leftrightarrow -z'(t) = rz(t) - \frac{r}{K} \end{aligned}$$

D'où l'équation équivalente $z' = -rz + \frac{r}{K}$.

3. Les techniques de cours habituelles donnent comme solutions à l'équation homogène les fonctions

$$z_0 : t \geq 0 \mapsto \lambda e^{-rt}$$

On cherche comme solution particulière une constante qui est

$$y_p : t \geq 0 \mapsto \frac{1}{K}$$

Ainsi, les solutions générales sont de la forme

$$z(t) : t \geq 0 \mapsto \lambda e^{-rt} + \frac{1}{K}$$

avec ainsi

$$N(t) : t \geq 0 \mapsto \frac{1}{\lambda e^{-rt} + \frac{1}{K}}$$

la condition initiale en $t = 0$ donne

$$\begin{aligned} N(0) = \frac{1}{\lambda e^{-r \cdot 0} + \frac{1}{K}} &\Leftrightarrow \lambda + \frac{1}{K} = \frac{1}{N(0)} \\ &\Leftrightarrow \lambda = \frac{1}{N(0)} - \frac{1}{K} \end{aligned}$$

D'où le résultat :

$$N(t) : t \geq 0 \mapsto \frac{1}{\left(\frac{1}{N(0)} - \frac{1}{K} \right) e^{-rt} + \frac{1}{K}} = \frac{K}{\left(\frac{K}{N(0)} - 1 \right) e^{-rt} + 1}$$

4. Les états stationnaires sont ceux où la dérivée est constante égale à 0. Ce sont donc ceux qui répondent à l'équation

$$r \left(1 - \frac{N(t)}{K} \right) N(t) = 0$$

Ce qui se passe ssi

$$N : t \mapsto 0 \quad \forall t \geq 0 \quad \text{ou} \quad N : t \mapsto K \quad \forall t \geq 0$$

III-3 Le modèle de Gompertz

Explications : Gompertz introduit en 1825 la notion d'évolution du taux de mortalité avec l'âge avançant, avec capacité du milieu limitée.

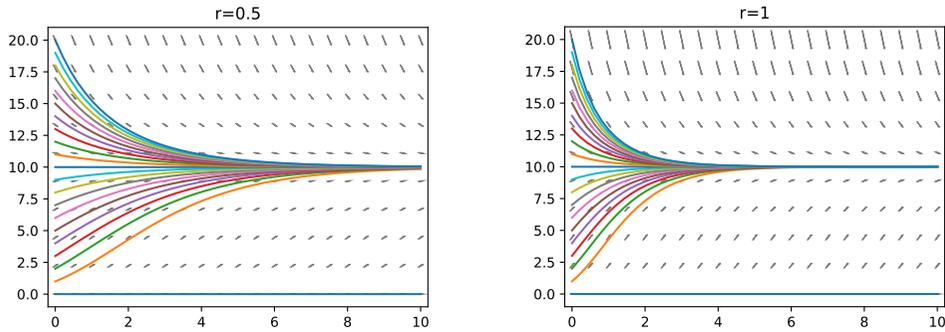
🌿 Définition

Avec les mêmes notations que le modèle logistique, il introduit l'équation :

$$N'(t) = r \ln \left(\frac{K}{N(t)} \right) N(t) \quad \forall t \geq 0$$

Pour tout N application qui ne s'annule pas.

On obtiendra un graphique du type suivant avec $r = 0.5$ et $r = 1$:



Commentaires :

Dans l'hypothèse où $r > 0$:

- Quand N est inférieur à la capacité K du milieu, on a $\ln\left(\frac{K}{N(t)}\right) > 0$. Ainsi $N' > 0$ et la population augmente.
- Quand N est supérieur à la capacité K du milieu, on a $\ln\left(\frac{K}{N(t)}\right) < 0$. Ainsi $N' < 0$ et la population diminue.
- En revanche, quand N est proche de la capacité d'accueil du milieu (par valeurs inférieures), la croissance ralentit de plus en plus. On a

$$\text{Si } N(t) \simeq K \text{ alors } \ln\left(\frac{K}{N(t)}\right) \simeq 0 \text{ et donc } N' \simeq 0$$

La population stagne.

? Exercice 5

On pose ici $r > 0$. On rappelle que dans notre modèle, N ne s'annule pas et est donc > 0 .

1. Déterminer le/les éventuels états stationnaires.
2. Poser pour tout $t \geq 0$, $z(t) = \ln(N(t))$. Montrer que l'équation de Gompertz équivaut à

$$z' + rz = c$$

avec c une constante à déterminer.

3. Résoudre l'équation en z .
4. En déduire que pour toute condition initiale $N(0) > 0$, la solution du modèle logistique est :

$$N : t \geq 0 \mapsto Ke^{-\ln\left(\frac{K}{N(0)}\right)e^{-rt}}$$

Solution

1. Les états stationnaires sont ceux où la dérivée est constante égale à 0. Ce sont donc ceux qui répondent à l'équation

$$r \ln\left(\frac{K}{N(t)}\right) N(t) = 0$$

Or, comme on a supposé $N(t)$ non nulle, il ne reste plus que $\ln\left(\frac{K}{N(t)}\right) = 0$, c'est-à-dire

$$N : t \mapsto K \quad \forall t \geq 0$$

2. On a $z' = \frac{N'}{N}$. Ainsi,

$$\begin{aligned} N'(t) = r \ln\left(\frac{K}{N(t)}\right) N(t) &\Leftrightarrow \frac{N'(t)}{N(t)} = r \ln\left(\frac{K}{N(t)}\right) \\ &\Leftrightarrow z'(t) = r \ln K - rz(t) \end{aligned}$$

D'où l'équation équivalente

$$z' + rz = r \ln K.$$

3. Les techniques de cours habituelles donnent comme solutions à l'équation homogène les fonctions

$$z_0 : t \geq 0 \mapsto \lambda e^{-rt}$$

On cherche comme solution particulière une constante qui est

$$y_p : t \geq 0 \mapsto \ln K$$

Ainsi, les solutions générales sont de la forme

$$z(t) : t \geq 0 \mapsto \lambda e^{-rt} + \ln K$$

4. Comme $N = e^z$, on déduit de la question précédente :

$$N(t) : t \geq 0 \mapsto e^{\lambda e^{-rt} + \ln K}$$

La condition initiale en $t = 0$ donne

$$\begin{aligned} N(0) = e^{\lambda e^{-r \cdot 0} + \ln K} &\Leftrightarrow N(0) = e^{\lambda + \ln K} \\ &\Leftrightarrow \lambda + \ln K = \ln(N(0)) \\ &\Leftrightarrow \lambda = \ln(N(0)) - \ln K = -\ln\left(\frac{K}{N(0)}\right) \end{aligned}$$

D'où le résultat :

$$N(t) : t \geq 0 \mapsto e^{\ln\left(\frac{K}{N(0)}\right)e^{-rt} + \ln K} = Ke^{-\ln\left(\frac{K}{N(0)}\right)e^{-rt}}$$

I Généralités

I-1 Définition

 Définition

On appelle *variable aléatoire réelle discrète* toute variable aléatoire $X : \Omega \rightarrow \mathbb{R}$ telle que $Supp(X)$ soit indexé par un sous-ensemble de \mathbb{N} .

 Remarque :

être indexé par un sous-ensemble de \mathbb{N} signifie simplement que l'on peut "étiqueter" les valeurs de $Supp(X)$ avec des valeurs entières.

■ Exemple 1 :

On lance un dé une infinité de fois. L'ensemble Ω des résultats possibles est

$$\Omega = \llbracket 1; 6 \rrbracket^{\mathbb{N}}$$

On s'intéresse maintenant au rang d'apparition du premier "6". On pose alors

$$X : \Omega \rightarrow \mathbb{R}$$

$$\omega \mapsto \begin{cases} \text{rang d'apparition de 6} & \text{si 6 apparait au moins une fois} \\ 0 & \text{sinon} \end{cases}$$

X est une variable aléatoire discrète car $Supp(X) = \mathbb{N}^* \subset \mathbb{N}$, et même $X(\Omega) = \mathbb{N}$.

■ Exemple 2 :

On note Ω l'ensemble des tailles possibles d'une plante donnée. On pose

$$X : \Omega \rightarrow \mathbb{R}$$

$$\omega \mapsto \frac{(-1)^{\lfloor w \rfloor}}{\lfloor w \rfloor + 1}$$

Alors X est une variable aléatoire discrète, car $Supp(X) = \left\{ \frac{(-1)^n}{n+1} \mid n \in \mathbb{N} \right\}$.

Cet ensemble est indexé par \mathbb{N} .

On lance deux dés à 6 faces de manière indépendante et on note S la somme des deux dés. On a alors

$$Supp(S) = \llbracket 2, 12 \rrbracket \subset \mathbb{N}$$

 Remarque :

Toute variable finie est discrète.

I-2 Loi d'une v.a.d.

 Remarque :

De manière générale, la loi d'une variable aléatoire est caractérisée par sa fonction de répartition. Néanmoins, dans certains cas particuliers, la loi peut être exprimée de manière plus concise. C'est le cas des variables aléatoires discrètes.

Commentaires :

Comme $Supp(X)$ est indexé par un sous-ensemble de \mathbb{N} , il en va de même pour tout $B \subset Supp(X)$. Dans ce cas, $\{(X = k)\}_{k \in B}$ est un système complet d'événement de B , d'où

$$P(X \in B) = \sum_{k \in B} P(X = k)$$

Pour obtenir une probabilité quelconque, il suffit donc de connaître uniquement l'ensemble des $P(X = k)$.

■ Exemple 3 :

Si X est le rang de première apparition du 6 dans une série de lancers indépendants

d'un dé, alors la probabilité que le 6 arrive à un rang pair est $\sum_{k=1}^{+\infty} P(X = 2k)$.

 Définition

On appelle *loi de la v.a.d. X* l'ensemble noté $\mathcal{L}(X)$ constitué de

- $Supp(X)$;
- l'ensemble des $P(X = k)$ avec $k \in Supp(X)$.

■ Exemple 4 :

On s'intéresse encore une fois au rang de première apparition de 6. Avec $\Omega = \{1, 2, \dots, 6\}^{\mathbb{N}}$ l'univers de départ, on avait posé

$$X : \Omega \rightarrow \mathbb{R}$$

$$w \mapsto \begin{cases} \text{rang de première apparition de 6 dans } w & \text{si 6 apparait} \\ 0 & \text{sinon} \end{cases}$$

Avec les différents calculs qui ont été faits précédemment, la loi de X peut s'écrire

$$X(\Omega) = \mathbb{N},$$

$$p(X = n) = \begin{cases} \left(\frac{5}{6}\right)^{n-1} \frac{1}{6} & \forall n \in \mathbb{N}^* \\ 0 & \text{si } n = 0 \end{cases}$$

ou

$$Supp(X) = \mathbb{N}^*,$$

$$p(X = n) = \left(\frac{5}{6}\right)^{n-1} \frac{1}{6} \quad \forall n \in \mathbb{N}^*$$

Commentaires :

Si on prend maintenant $B = \text{Supp}(X)$, $\{(X = k)\}_{k \in \text{Supp}(X)}$ est un système quasi-complet d'événement, on a :

$$1 = P(X \in \text{Supp}(X)) = \sum_{k \in \text{Supp}(X)} P(X = k)$$

On va maintenant voir comment ça se passe "dans l'autre sens" :

Proposition 96

Étant donné un sous-ensemble I de \mathbb{N} , pour toute suite $(p_i)_{i \in I}$ et toute suite $(x_i)_{i \in I}$ de réels, telles que

$$\begin{cases} p_i \geq 0 & \forall i \in I \\ \sum_{i \in I} p_i & \text{converge et de somme totale 1} \end{cases}$$

Alors il existe une variable aléatoire X telle que

$$\begin{cases} \text{Supp}(X) = \{x_i \mid i \in I\} \\ p(X = x_i) = p_i & \forall i \in I \end{cases}$$

Démonstration : admise. \square

■ Exemple 5 :

On pose $p_i = \frac{1}{2^i}$ pour tout $i \in \mathbb{N}^*$. Alors il existe une variable aléatoire X telle que

$$P\left(X = \frac{(-1)^n}{e^n}\right) = \frac{1}{2^n} \quad \forall n \in \mathbb{N}^*$$

En effet, on a :

- $\frac{1}{2^i} \geq 0$ pour tout $i \in \mathbb{N}^*$.
 - $\sum_{n \in \mathbb{N}^*} \frac{1}{2^i}$ qui est une série géométrique convergente, de somme totale
- $$\sum_{i=1}^{+\infty} \frac{1}{2^i} = \frac{1}{2} \sum_{n=0}^{+\infty} \frac{1}{2^n} = \frac{1}{2} \frac{1}{1 - \frac{1}{2}} = 1$$

⚠ CONFUSION

Si X et Y ont même loi, elles ne prennent pas forcément la même valeur en même temps.

■ Exemple 6 :

On note X et Y les résultats des lancers de deux dés équilibrés distincts. X et Y ont même loi, mais n'ont pas toujours même valeur. (On sait bien par expérience que les doubles ne sont pas systématiques!)

I-3 Fonction de répartition

Voyons maintenant que la donnée de la fonction de répartition et de ladite "loi" précédente d'une variable aléatoire sont équivalentes. Rappelons pour commencer la définition d'une fonction de répartition d'une variable aléatoire :

🌿 Définition

Étant donnée une variable aléatoire X , on appelle *fonction de répartition* de X la fonction

$$\begin{aligned} F_X : \mathbb{R} &\rightarrow [0, 1] \\ t &\mapsto P(X \leq t) \end{aligned}$$

■ Exemple 7 :

Reprenons l'exemple de première apparition de 6 dans un jeté infini de dé équilibré. La loi de X est résumée par :

$$\text{Supp}(X) = \mathbb{N}^* \quad \text{et} \quad P(X = n) = \left(\frac{5}{6}\right)^{n-1} \frac{1}{6} \quad \forall n \in \mathbb{N}^*$$

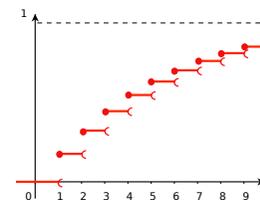
- si $t < 1$, comme $\text{Supp}(X) \subset [1 + \infty[$, on a $F_X(t) = P(X \leq t) = 0$

- Si $t \geq 1$,
- $$F_X(t) = P(X \leq t) = P\left(\bigcup_{k=1}^{\lfloor t \rfloor} (X = k)\right) = \sum_{k=1}^{\lfloor t \rfloor} P(X = k)$$

$$\begin{aligned} \text{D'où} \quad F(t) &= \sum_{k=1}^{\lfloor t \rfloor} P(X = k) = \sum_{k=1}^{\lfloor t \rfloor} \left(\frac{5}{6}\right)^{k-1} \frac{1}{6} = \frac{1}{6} \sum_{k=1}^{\lfloor t \rfloor} \left(\frac{5}{6}\right)^{k-1} \\ &= \underbrace{\frac{1}{6}}_{j=k-1} \sum_{j=0}^{\lfloor t \rfloor - 1} \left(\frac{5}{6}\right)^j = \frac{1}{6} \frac{1 - \left(\frac{5}{6}\right)^{\lfloor t \rfloor}}{1 - \frac{5}{6}} = 1 - \left(\frac{5}{6}\right)^{\lfloor t \rfloor} \end{aligned}$$

Conclusion :

$$F_X(t) = \begin{cases} 0 & \text{si } t < 0 \\ 1 - \left(\frac{5}{6}\right)^{\lfloor t \rfloor} & \text{si } t \geq 0 \end{cases}$$



On observe que la fonction est "en escalier". C'est le cas de manière générale dans le cadre des v.a.d. :

Propriété 97

Soit X une variable aléatoire discrète. Alors sa fonction de répartition F_X est constante sur chaque intervalle $[x_i, x_{i+1}[$ tels que $Supp(X) = \{x_i \mid i \in I\}$ avec $x_i < x_{i+1}$ pour tout i .

Démonstration :

Soient $x, y \in [x_i, x_{i+1}[$. Quitte à échanger x et y , on peut supposer $x < y$. On a

$$\text{Or, } F(x) = P(X \leq x) = P(X \leq y) + P(x < X \leq y)$$

$$0 \leq P(x < X \leq y) \leq P(x_i < X < x_{i+1}) = 0$$

D'où $F(x) = P(X \leq y) = F(y)$. \square

Voyons maintenant ce que signifient les points de discontinuité :

 **Remarque :**

Soit F_X la fonction de répartition d'une v.a.d. X . Alors,

$$p(X = t) = p(X \leq t) - p(X < t)$$

Ce qui veut dire que :

$$p(X = t) \neq 0 \text{ si et seulement si } t \text{ est un point de discontinuité de } F_X$$

et

$$p(X = t) \text{ est la "hauteur" de la discontinuité de la fonction de répartition en } t.$$

De la remarque précédente, on tire :

Propriété 98

Soit F une fonction de répartition d'une variable X , constante entre ses points de discontinuité notés D indexés par un sous-ensemble de \mathbb{N} . Alors

$$Supp(X) = D$$

et X est une variable discrète.

Démonstration :

Le résultat $D \subset Supp(X)$ est immédiat d'après la remarque précédente.

Afin de ne pas faire appel à de la technicité inutile, on se contera de constater ici, que comme la fonction est constante sur chaque intervalle entre les valeurs de D , on a

$$\sum_{x \in D} P(X = x) = \text{"hauteur totale entre 0 et 1"} = 1$$

et donc $Supp(X) = D$ \square

Corollaire

Deux v.a.d. ont même loi si et seulement si elles ont même fonction de répartition.

Démonstration : admise. \square

Vous devez savoir passer d'une fonction de répartition à la variable discrète et réciproquement.

II Exemples fondamentaux

Rappel :

Soit $\Omega_0 = \{x_k \mid k \in I \subset \mathbb{N}\}$. On rappelle que la loi de X est entièrement définie si

$$\sum_{k \in I} \underbrace{p(X = x_k)}_{\geq 0} = 1$$

Dans ce cas, $Supp(X) = \Omega_0$.

II-1 Rappels de Probabilités finies

II.1-a) Loi uniforme

Définition et Proposition

Soit $N \in \mathbb{N}$, $\Omega_0 = \{x_k \mid k = 1, \dots, N\}$ un ensemble fini de cardinal N . On dit que X suit une *loi uniforme* sur Ω_0 si

$$Supp(X) = \Omega_0 \text{ et } p(X = x_k) = \frac{1}{N} \quad \forall k = 1, \dots, N$$

On écrit

$$X \hookrightarrow \mathcal{U}_{\Omega_0}$$

Démonstration :

La suite $(p_k)_{k \in \{1, \dots, N\}}$ définie par $p_k = p(X = x_k)$ est positive et

$$\sum_{k=1}^N p_k = \sum_{k=1}^N \frac{1}{N} = N \frac{1}{N} = 1.$$

\square

Modélisation : Cette loi modélise la survenue équiprobable de N événements ($X = x_1, \dots, X = x_N$).

■ Exemple 8 :

On lance un dé une fois et on note X le nombre obtenu. Alors X suit une loi uniforme sur l'ensemble $\Omega = \{1, \dots, 6\}$.

FORMULE

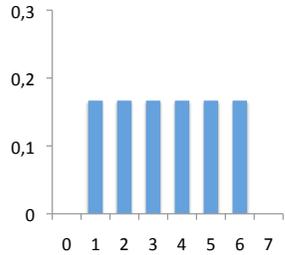
> La loi uniforme est la seule loi pour laquelle on peut affirmer que pour tout A dans

 $Supp(X),$

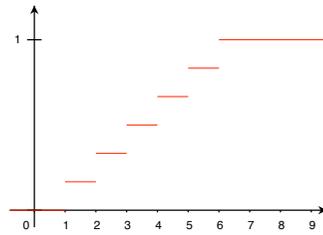
$$p(X \in A) = \frac{\text{card}(A)}{\text{card}(\Omega)}.$$

★ pour $\Omega = \{1, \dots, 6\}$:

Diagramme de probabilités :



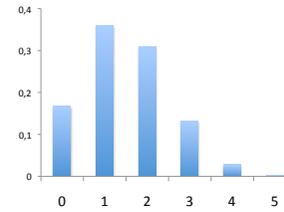
Fonction de répartition :



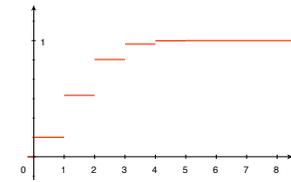
On donne ci-dessous quelques exemples de répartition de lois binomiales :

★ $n = 5$ $p = 0,3$:

Diagramme de probabilités :



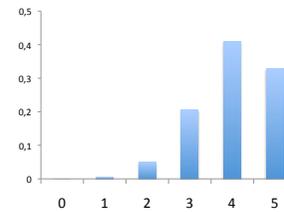
Fonction de répartition :



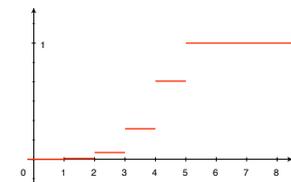
4.

★ $n = 5$ $p = 0,8$:

Diagramme de probabilités :

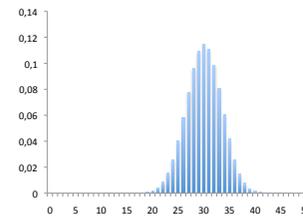


Fonction de répartition :

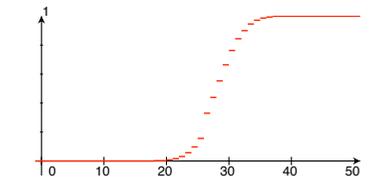


★ $n = 50$ $p = 0,6$:

Diagramme de probabilités :



Fonction de répartition :



II-2 Loi géométrique

Définition et Proposition

Soit $p \in]0; 1[$. On dit que X suit une *loi géométrique* sur \mathbb{N}^* de paramètre p si

$$\text{Supp}(X) = \mathbb{N}^* \quad \text{et} \quad p(X = k) = (1 - p)^{k-1} p \quad \forall k \in \mathbb{N}^*.$$

On note

$$X \hookrightarrow \mathcal{G}_{\mathbb{N}^*}(p) \quad \text{ou} \quad X \hookrightarrow \mathcal{G}(p).$$

II.1-b) Loi binomiale $\mathcal{B}(n; p)$

Dans cette section, on se donne $p \in [0; 1]$ et $n \in \mathbb{N}^*$.

Définition et Proposition

Soient $k \in \mathbb{N}$ et $p_k = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{si } k \in \{0, 1, \dots, n\} \\ 0 & \text{sinon} \end{cases}$

On dit que X suit une loi de Bernoulli de paramètre n et p si

$$\underbrace{\text{Supp}(X) = \{0, \dots, n\}}_{\text{redondant}} \quad \text{et} \quad p(X = k) = p_k \quad \forall k \in \mathbb{N}.$$

On note

$$X \hookrightarrow \mathcal{B}(n; p)$$

Démonstration :

$(p_k)_{k \in \mathbb{N}}$ est une suite positive telle que $\sum_{n \in \mathbb{N}} p_k = 1$. \square

Modélisation : Cette loi modélise le nombre de succès dans une succession d'épreuves de Bernoulli **indépendantes**, où la probabilité de succès à chaque étape est p .

■ Exemple 9 :

On lance n fois un dé et on compte le nombre d'apparition du chiffre 1. Alors X suit une loi binomiale $\mathcal{B}(n; \frac{1}{6})$.

Démonstration :

★ Tout d'abord, $(1 - p)^{k-1}p \geq 0$ pour tout $k \in \mathbb{N}^*$.
 ★ Pour vérifier que c'est une mesure de probabilité discrète, il suffit encore de vérifier que $\sum_{k=1}^{+\infty} (1 - p)^{k-1}p = 1$, ce qui est vrai par somme totale de série géométrique convergente. \square

Modélisation : Cette loi modélise le rang de **première apparition** d'un événement de probabilité p dans une suite infinie de tirages **indépendants**.

■ **Exemple 10 :**

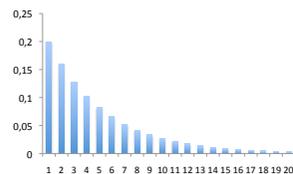
On jette indéfiniment un dé équilibré et on s'intéresse à la première apparition du chiffre 1. À chaque tirage, la probabilité d'apparition de 1 est $p = \frac{1}{6}$. Ainsi, on introduit la variable aléatoire X définie de la manière suivante :

$$X = \begin{cases} 0 & \text{si 1 n'apparaît jamais.} \\ k & \text{si 1 apparaît au rang } k \text{ pour la première fois.} \end{cases}$$

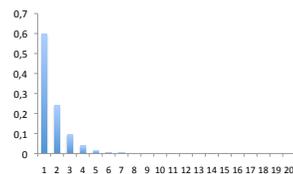
La loi de X est donc $\mathcal{G}\left(\frac{1}{6}\right)$.

On donne ci-dessous quelques exemples de diagramme de probabilités de lois géométriques :

$p = 0,2 :$



$p = 0,6 :$



Propriété 99

Soit $p \in]0; 1[$. La fonction de répartition d'une variable aléatoire de loi $\mathcal{G}(p)$ est définie par

$$F(t) = \begin{cases} 1 - (1 - p)^{\lfloor t \rfloor} & \text{si } t \geq 1 \\ 0 & \text{sinon} \end{cases}$$

Démonstration :

• Si $t < 1$, comme $Supp(X) = \mathbb{N}^*$, il est clair que

$$0 \leq F(t) = P(X \leq t) \leq P(X < 1) = 0,$$

d'où $F(t) = 0$.

$$\begin{aligned} \bullet \text{ Si } t \geq 1, \text{ on a } F(t) = P(X \leq t) &= P(X \leq \lfloor t \rfloor) \\ &= P\left(\bigcup_{k=1}^{\lfloor t \rfloor} (X = k)\right) \\ &= \sum_{k=1}^{\lfloor t \rfloor} P(X = k) \quad (\sigma\text{-add.}) \\ &= \sum_{k=1}^{\lfloor t \rfloor} p(1 - p)^{k-1} = p \frac{1 - (1 - p)^{\lfloor t \rfloor}}{1 - (1 - p)} \\ &= 1 - (1 - p)^{\lfloor t \rfloor} \end{aligned}$$

\square

Définition et Proposition

Soit $p \in]0; 1[$. On pose X et Y deux variables aléatoires telles que

$$X \hookrightarrow \mathcal{G}(p) \quad \text{et} \quad Y = X - 1$$

alors

$$Supp(Y) = \mathbb{N} \quad \text{et} \quad P(Y = k) = (1 - p)^k p \quad \forall k \in \mathbb{N}$$

On note $Y \hookrightarrow \mathcal{G}_{\mathbb{N}}(p)$ et on dit que Y *suit une loi géométrique* sur \mathbb{N} de paramètre p .

Démonstration : exercice. \square

Modélisation : Y modélise ici le nombre d'échecs avant l'apparition du premier succès.

Propriété 100 d'absence de mémoire

Soit X suivant une loi $\mathcal{G}(p)$. Pour tout $n, m > 0$, on a

$$P_{X > n}(X > n + m) = P(X > m)$$

Commentaires :

On peut interpréter ceci de la manière suivante :

“X ne s'épuise pas.”

Si X correspond par exemple au nombre de minutes d'attente avant de réussir à attraper un vendeur dans une boutique, quelqu'un qui déjà attendu n minutes a exactement autant de chance d'attendre encore pendant un temps de m minutes supplémentaires qu'une autre personne qui vient d'arriver. (Il n'y aurait donc pas ici de file d'attente et chaque personne serait servie de manière aléatoire. Oui, c'est particulièrement injuste mais c'est la vie quelquefois !)

Définition et Proposition

Soit $\lambda \in]0; +\infty[$. On dit que X suit une *loi de Poisson* de paramètres λ si

$$\underbrace{\text{Supp}(X) = \mathbb{N}}_{\text{redundant}} \quad \text{et} \quad p(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad \forall k \in \mathbb{N}$$

On note

$X \hookrightarrow \mathcal{P}(\lambda)$

La fonction de répartition d'une variable aléatoire X suivant une loi de Poisson $\mathcal{P}(\lambda)$ est définie par

$$F(t) = 1 - \int_0^\lambda e^{-u} \frac{u^{\lfloor t \rfloor}}{\lfloor t \rfloor!} du$$

Démonstration :

Les $p(X = k)$ sont clairement positifs et d'après le cours sur les séries, on sait que

$\sum \frac{\lambda^k}{k!}$ est convergente, avec

$$\sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} = e^\lambda,$$

d'où

$$\sum_{k=0}^{+\infty} P(X = k) = 1$$

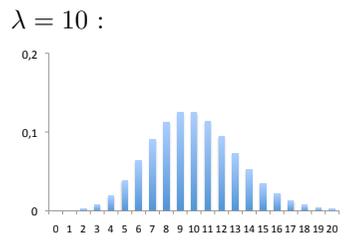
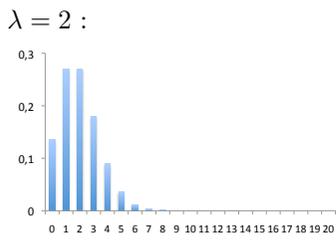
□

Modélisation : Cette loi modélise généralement une fréquentation.

■ **Exemple 11 :**

On compte le nombre X de clients à un guichet pendant un laps de temps T . En général, on suppose que $X \hookrightarrow \mathcal{P}(\lambda)$, où λ est le nombre moyen de client pendant un temps T .

On donne ci-dessous quelques exemples de diagramme de probabilités de lois exponentielles :



Démonstration :

• Montrons par récurrence la formule $F(n) = 1 - \int_0^\lambda e^{-u} \frac{u^n}{n!} du$ pour tout entier $n \in \mathbb{N}$:

Initialisation : Soit $n = 0$.

D'une part, $F(0) = P(X = 0) = e^{-\lambda} \frac{\lambda^0}{0!} = e^{-\lambda}$.

D'autre part, on a $\int_0^\lambda e^{-u} \frac{u^0}{0!} du = \int_0^\lambda e^{-u} du = [-e^{-u}]_0^\lambda = -e^{-\lambda} + 1 = 1 - F(0)$.

Hérédité : Supposons la formule vraie pour un $n \in \mathbb{N}$ et montrons qu'elle reste vraie pour $n + 1$.

On a donc $1 - F(n) = \int_0^\lambda e^{-x} \frac{x^n}{n!} dx$. L'intégration par partie

$$\begin{aligned} u &= e^{-x} & u' &= -e^{-x} \\ v' &= \frac{x^n}{n!} & v &= \frac{x^{n+1}}{(n+1)!} \end{aligned}$$

donne alors

$$1 - F(n) = \left[e^{-x} \frac{x^{n+1}}{(n+1)!} \right]_0^\lambda - \int_0^\lambda -e^{-x} \frac{x^{n+1}}{(n+1)!} dx$$

Or,

$$\left[e^{-x} \frac{x^{n+1}}{(n+1)!} \right]_0^\lambda = e^{-\lambda} \frac{\lambda^{n+1}}{(n+1)!} = P(X = n+1)$$

D'où

$$1 = \underbrace{F(n) + P(X = n+1)}_{F(n+1)} + \int_0^\lambda -e^{-x} \frac{x^{n+1}}{(n+1)!} dx$$

Ce qui donne bien l'égalité attendue

$$F(n+1) = 1 - \int_0^\lambda -e^{-x} \frac{x^{n+1}}{(n+1)!} dx$$

□



Remarque :

Cette fonction de répartition n'est en générale pas utilisée car pour calculer $F(t)$, il

faut faire des IPP, ce qui revient à calculer $F(t) = \sum_{k=0}^{[t]} p(X = k) = e^{-\lambda} \sum_{k=0}^{[t]} \frac{\lambda^k}{k!}$

III Espérance et variance

Notation :

Dans toute cette section, pour des raisons d'écriture, on supposera que X est une variable aléatoire telle que

$$\text{Supp}(X) \subset \{x_k \in \mathbb{R} \mid k \in \mathbb{N}, \text{ avec les } x_k \text{ 2 à 2 distincts.}\}$$

quitte à avoir en réalité un sous-ensemble A d'indices tels que que $p(X = x_k) = 0$ pour tout $k \in A$.

■ Exemple 12 :

Si X est une variable qui suit une loi uniforme sur $\{1, \dots, 8\}$, on aura $\text{Supp}(X) \subset \mathbb{N}$, i.e. $x_i = i \quad \forall i \in \mathbb{N}$, mais avec

$$P(X = x) = 0 \quad \text{si } x = 0 \text{ ou } x \geq 9$$

III-1 Espérance

Commentaires :

En mécanique, le barycentre de points pondérés est un point d'équilibre de la structure. De même, pour une v.a.d., on peut définir la notion de barycentre à travers l'espérance.

Définition et Proposition

Si la série $\sum_{k \in \mathbb{N}} |x_k| p(X = x_k)$ converge, on appelle *espérance de X* et on note $\mathbb{E}[X]$ la valeur (bien définie)

$$\mathbb{E}[X] = \sum_{k \in \mathbb{N}} x_k p(X = x_k)$$

et cette valeur ne dépend pas de l'ordre des x_k .

Démonstration : admise. \square

■ Exemple 13 :

On admet qu'il existe $\alpha \in \mathbb{R}$ tel qu'on ait une variable aléatoire X vérifiant

$$P(X = (-1)^{n+1}n) = \frac{\alpha}{n^3} \quad \forall n \in \mathbb{N}^*$$

Montrons que $\mathbb{E}[X]$ existe.

- Etude de la série $S = \sum |(-1)^{n+1}n|P(X = (-1)^{n+1}n)$:

On a $S = \alpha \sum \frac{1}{n^2}$, qui est une série convergente (*exo*), donc $\mathbb{E}[X]$ existe.

- Valeur de $\mathbb{E}[X]$:

On sait alors que

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

pour information (HP), on a alors

$$\mathbb{E}[X] = \alpha \frac{\pi^2}{12}$$

⚠ Remarque :

Il est **nécessaire de vérifier la convergence absolue**, et ceci pour plusieurs raisons (dont le bon fonctionnement de certaines propriétés) mais surtout, dans un premier temps, **afin que la notion d'espérance soit bien définie de façon unique**. En effet, observons l'exemple suivant :

Soit X une variable aléatoire telle que $\text{Supp}(X) = \left\{ \underbrace{(-1)^n n}_{x_n} \mid n \in \mathbb{N}^* \right\}$ ainsi que,

(pour $\alpha = \frac{6}{\pi^2}$),

$$P(X = (-1)^{n+1}n) = \frac{\alpha}{n^2}$$

(on admet que ceci définit bien la loi d'une variable aléatoire.)

Alors d'une part la série des valeurs absolues :

$$\sum_{k \in \mathbb{N}^*} |x_k| p(X = x_k) = \sum_{k \in \mathbb{N}^*} |(-1)^{n+1}n| \frac{1}{n} = \alpha \sum_{k \in \mathbb{N}^*} \frac{1}{n}$$

est divergente et d'autre part, la série "sans" valeur absolue

$$\sum_{k \in \mathbb{N}^*} x_k p(X = x_k) = \sum_{k \in \mathbb{N}^*} (-1)^{n+1} n \frac{1}{n} = \alpha \sum_{k \in \mathbb{N}^*} \frac{(-1)^{n+1}}{n}$$

est convergente.

On pourrait donc imaginer qu'il existe quand même une "valeur moyenne", sauf qu'il se pose néanmoins un gros problème : la somme totale de cette série dépend de l'ordre dans lequel on additionne les $x_k p(X = x_k)$!

En effet, souvenons-nous que nous avons étudié cette série graphiquement dans le premier chapitre et nous avons constaté qu'en réorganisant l'ordre des termes de la série, nous n'obtenions pas du tout les mêmes sommes totales. On obtiendrait donc **plusieurs valeurs d'espérance possible** ! Ce qui est tout à fait impensable.

Commentaires :

Un des résultats les plus importants concernant le traitement des espérances est le théorème dit "de transfert" énoncé ci-dessous. Il permet, pour toute v.a.d. $f(X)$, de se ramener à la loi de X au lieu d'avoir à calculer la loi de $f(X)$ (pas toujours simple.)

Théorème 102 de transfert

Soient

- I un intervalle (ou réunion finie d'intervalles) contenant $\text{Supp}(X)$
- $g : I \rightarrow \mathbb{R}$ une fonction définie sur $\text{Supp}(X)$

Alors, $\mathbb{E}[g \circ X]$ existe ssi si la série $\sum g(x_k) p(X = x_k)$ est absolument convergente. Alors on a

$$\mathbb{E}[g \circ X] = \sum_{k=0}^{+\infty} g(x_k) p(X = x_k)$$

Démonstration :

(idée de dém.) On note $Y = g \circ X$. Alors, par composition d'application, $Y : \Omega \rightarrow \mathbb{R}$ est une v.a.d. sur Ω . Par définition, l'espérance de Y existe si la série $\sum_{y \in \text{Supp}(Y)} y \cdot p(Y = y)$ cv absolument.

$$\sum_{y \in \text{Supp}(Y)} |y| \cdot p(Y = y) = \sum_{y \in \text{Supp}(Y)} |y| \cdot \sum_{w \in \Omega, Y(w)=y} p(w) = \sum_{y \in \text{Supp}(Y)} \sum_{w \in \Omega, g \circ X(w)=y} |g \circ X(w)| p(w)$$

mais les $X(w)$ sont les x_k . En réorganisant différemment la double somme (dont la convergence est à prouver, ce qui est HP), cette valeur est égale à

$$\sum_{w \in \Omega} |g \circ X(w)| p(w) = \sum_{k \in \mathbb{N}} \left(\sum_{w \in X^{-1}(x_k)} \underbrace{|g \circ X(w)|}_{g(x_k)} p(\{w\}) \right) = \sum_{k \in \mathbb{N}} |g(x_k)| p(X = x_k)$$

Cette somme est convergente, donc celle de départ l'est aussi et donc $\mathbb{E}[Y] = \mathbb{E}[g \circ X]$ existe. En retirant les valeurs absolues et en réitérant le raisonnement, on prouve l'égalité des espérances. \square

? Exercice 1

Soit $X \hookrightarrow \mathcal{G}(p)$ où $p \in]0, 1[$ et $Y = e^X$. Montrer que $\mathbb{E}[Y]$ existe ssi $p > 1 - e^{-1}$.

Solution

D'après le théorème de transfert, $\mathbb{E}[Y]$ existe ssi la série $\sum_{n \in \mathbb{N}^*} e^n P(X = n)$ converge. Or,

$$\sum_{n \in \mathbb{N}^*} e^n P(X = n) = \sum_{n \in \mathbb{N}^*} e^n p(1-p)^{n-1} = ep \sum_{n \in \mathbb{N}^*} (e(1-p))^{n-1} = ep \sum_{i \in \mathbb{N}} (e(1-p))^i$$

Ainsi, la série converge ssi $|e(1-p)| < 1$, i.e. $e(1-p) < 1$, ce qui revient ensuite à $p > 1 - e^{-1}$.

Propriété 103

Soit X une v.a.d.

- $\mathbb{E}[X]$ existe si et seulement si $\mathbb{E}[|X|]$ existe.
- Si $\mathbb{E}[X]$ est finie, alors $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$

Démonstration :

- La fonction $x \mapsto |x|$ est continue. Ainsi, d'après le théorème de transfert, $\mathbb{E}[|X|]$ existe

ssi la série $\sum P(X = x_k) |x_k|$ converge absolument,

autrement dit,

ssi la série $\sum P(X = x_k) |x_k|$ converge.

On constate que c'est exactement la condition d'existence de $\mathbb{E}[X]$.

- Supposons que $\mathbb{E}[X]$ existe. Alors $\mathbb{E}[|X|]$ aussi. De plus, par théorème sur les séries absolument convergentes,

$$|\mathbb{E}[X]| = \left| \sum_{k \in \mathbb{N}} x_k p(X = x_k) \right| \leq \sum_{k \in \mathbb{N}} |x_k| p(X = x_k) \stackrel{\text{transfert}}{=} \mathbb{E}[|X|]$$

\square

Propriété 104 Cas particuliers

Soit X une v.a.d. et $a, b \in \mathbb{R}$ (finis!)

- Si $\text{Supp}(X) = \{a\}$ (i.e. $X = a$ p.s.), alors $\mathbb{E}[X] = a$.
- Si $\text{Supp}(X) \subset [a, b]$, (i.e. $a \leq X \leq b$ p.s.), alors $\mathbb{E}[X]$ existe et $a \leq \mathbb{E}[X] \leq b$

Démonstration :

- trivial, exercice.
- Ceci est conséquence de la proposition suivante. En effet, les variables aléatoires constantes a et b ont des espérances qui existent. Ainsi, d'après la proposition suivante, on obtiendra

$$\mathbb{E}[a] \leq \mathbb{E}[X] \leq \mathbb{E}[b].$$

le *iii*) de cette proposition permet de conclure.

□

■ Exemple 14 :

Soit X une variable de support $\left\{ \frac{(-1)^n}{n}, n \in \mathbb{N}^* \right\}$. Montrer que $\mathbb{E}[X]$ existe.

On a $\text{Supp}(X) \subset [-1, 1]$. Elle est bornée, donc son espérance existe (et d'ailleurs) $-1 \leq \mathbb{E}[X] \leq 1$

Propriété 105 de comparaison de variables positives

Soient X, Y deux v.a.d. positives définies sur un même espace probabilisé. Alors :
Si $X \leq Y$ et $\mathbb{E}[Y]$ existe, alors $\mathbb{E}[X]$ existe et $0 \leq \mathbb{E}[X] \leq \mathbb{E}[Y]$.

Démonstration : admise. □

Théorème 106

Soient X, Y deux v.a.d. définies sur un même espace probabilisé et **admettant une espérance**. On a :

A. Linéarité :

(A1) $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$

(A2) $\mathbb{E}[\lambda X] = \lambda \mathbb{E}[X] \quad \forall \lambda \in \mathbb{R}$.

B. Monotonie :

(B1) $X \geq 0 \implies \mathbb{E}[X] \geq 0$

(B2) $X \geq Y \implies \mathbb{E}[X] \geq \mathbb{E}[Y]$

(B3) X et Y ont même loi $\implies \mathbb{E}[X] = \mathbb{E}[Y]$

Démonstration :

(A1) est admis. (A2) : appliquer le théorème de transfert.

Pour la monotonie, utiliser les propriétés déjà énoncées. □

🌿 Définition

On dit qu'une v.a.d. X est *centrée* si son espérance existe et $\mathbb{E}(X) = 0$.

III-2 Moments

Commentaires :

L'espérance donne une caractéristique moyenne pour une v.a.d.. Avec les moments, on pourra déterminer la dispersion autour de cette moyenne.

🌿 Définition

Soit X une v.a.d. et $r \in \mathbb{N}^*$. Si $\mathbb{E}[X^r]$ existe, on appelle *moment d'ordre r* la valeur :
$$\mathbb{E}[X^r].$$

■ Exemple 15 :

Le moment centré d'ordre 1, s'il existe, est $\mathbb{E}[X]$.

Propriété 107

Si la série $\sum_{n \in \mathbb{N}} x_k^n p(X = x_k)$ converge absolument, le moment d'ordre r existe et

$$\mathbb{E}[X^r] = \sum_{n=0}^{+\infty} x_k^n p(X = x_k)$$

Démonstration :

C'est le théorème de transfert appliqué à $g(x) = x^r$. □

Théorème 108

Soient X une v.a.d. et $r \in \mathbb{N}^*$. Si le moment d'ordre r de X existe, alors le moment d'ordre r' existe pour tout $r' \in \{1, \dots, r\}$.

Démonstration :

Soient $r \in \mathbb{N}^*$ et $r' \in \{1, \dots, r\}$.

Supposons que $\mathbb{E}[X^r]$ existe, c'est-à-dire que la série $\sum_{k \in \mathbb{N}} |x_k|^r p(X = x_k)$ converge.

★ Si $|x_k| \geq 1$, alors, comme $r' \leq r$, on a $|x_k|^{r'} \leq |x_k|^r$

★ Sinon, si $|x_k| \leq 1$, $|x_k|^{r'} \leq 1$.

Autrement dit, dans tous les cas, on a que

$$|x_k|^{r'} \leq \max(1, |x_k|^r) \leq 1 + |x_k|^r$$

$$|x_k|^{r'} p(X = x_k) \leq p(X = x_k) + |x_k|^r p(X = x_k)$$

qui sont tous deux des termes généraux de séries convergentes. Par somme de séries convergentes, la série de terme général $p(X = x_k) + |x_k|^r p(X = x_k)$ est convergente, puis, par comparaison de séries à termes positifs, la série $\sum_{k \in \mathbb{N}} |x_k|^{r'} p(X = x_k)$ est convergente. □



Remarque :

Dans la propriété, on peut remplacer la v.a.d. X par la v.a.d. $X - a$ pour un réel a fixé, on obtient sous réserve de convergence absolue,

$$\mathbb{E}[(X - a)^r] = \sum_{n=0}^{+\infty} p(X = x_k) (x_k - a)^r.$$



Définition

Soit X une v.a.d. et $r \in \mathbb{R}^*$. Si $\mathbb{E}[X^r]$ existe, on appelle *moment centré d'ordre r* la valeur (bien définie : admis).

$$\mathbb{E}[(X - \mathbb{E}[X])^r]$$

III-3 Variance



Définition

Soit X une v.a.d.. Si $\mathbb{E}[X^2]$ existe, le moment centré d'ordre 2 de X existe. On l'appelle *variance* de X . Elle est notée

$$V(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

La racine carrée de cette valeur, notée

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

est appelée *écart-type* de la variable X .



Définition

Une v.a.d. X est dite *réduite* si sa variance existe et $V(X) = 1$.

Vocabulaire : Les v.a.d.

$$X - \mathbb{E}[X] \quad \text{et} \quad \frac{X - \mathbb{E}[X]}{\sigma(X)}$$

sont appelées respectivement variable *centrée* et *centrée réduite* de X .

Théorème 109 de König-Huyghens

Soit X une v.a.d.. Si $\mathbb{E}[X^2]$ existe, on a

$$V(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

Démonstration :

Si le moment d'ordre 2 existe, alors $\mathbb{E}[X]$ existe également. Le reste n'est qu'un petit calcul simple.

$$V(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2 - 2\mathbb{E}[X]X + \mathbb{E}[X]^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

□

Propriété 110

Soient X une v.a.d. admettant un moment d'ordre 2, et $a, b \in \mathbb{R}$. Alors, on a

- X est constante (en loi) si et seulement si $V(X) = 0$.
- $V(aX) = a^2 V(X)$.
- $V(X + b) = V(X)$.

Démonstration :

- ★ Supposons que X est constante, par exemple $X = \alpha$. Alors $\mathbb{E}[X] = \alpha$. D'où

$$V(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[0] = 0$$

- ★ Supposons que $V(X) = 0$. Notons $Y = X - \mathbb{E}[X]$. Montrons que Y est nulle, ce qui montrera que $X = \mathbb{E}[X]$, c'est-à-dire que X est constante.

Notons $P_Y = \sum p_k \delta_{x_k}$ la loi de Y . Par hypothèse, $\mathbb{E}[Y^2]$ existe et vaut 0, c'est-à-dire que la série $\sum_{k \in \mathbb{N}} x_k^2$ converge de somme 0.

C'est une série de terme positif, tous les termes sont donc nuls, i.e. la seule valeur possible de Y est 0.

- On sait que $X - \mathbb{E}[X]$ admet un moment d'ordre 2. En multipliant par une constante, on ne change pas cet état de fait. Or, d'une part,

$$\mathbb{E}[a^2(X - \mathbb{E}[X])^2] = a^2 \mathbb{E}[(X - \mathbb{E}[X])^2] = a^2 V(X)$$

et d'autre part,

$$\mathbb{E}[a^2(X - \mathbb{E}[X])^2] = \mathbb{E}[(aX - a\mathbb{E}[X])^2] = \mathbb{E}[(aX - \mathbb{E}[aX])^2] = V(aX)$$

- *exercice*

□

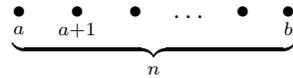
III-4 Exemples fondamentaux

Dans cette section, on récapitule les données obtenues sur les exemples classiques de loi finies et discrètes.

III.4-a) Loi uniforme discrète

On se donne X une variable aléatoire suivant une loi uniforme sur $\{a, a + 1, \dots, b\}$, où $a, b \in \mathbb{Z}, a \leq b$.

$$\begin{aligned} E(X) &= \frac{b+a}{2} \\ V(X) &= \frac{n^2-1}{12} \quad \text{où } n = b - a + 1 \quad (\text{HP}) \end{aligned}$$



Cas particulier,

★ si $a = 1, b = n \in \mathbb{N}^*$, on a

$$E(X) = \frac{n+1}{2}$$

★ si $a = 0, b = n \in \mathbb{N}^*$, on a

$$E(X) = \frac{n}{2}$$

III.4-b) Loi de Bernoulli, $\mathcal{B}(p)$

On se donne X une variable aléatoire suivant une loi de Bernoulli de paramètre p . On a

$$P(X = 0) = q \quad \text{et} \quad P(X = 1) = p, \quad \text{où } q = 1 - p$$

D'où

$$\begin{aligned} E(X) &= p \\ V(X) &= pq \end{aligned}$$

III.4-c) Loi Binomiale, $\mathcal{B}(n; p)$

On se donne X une variable aléatoire suivant une loi Binomiale de paramètre n, p (à valeurs dans $\{0, \dots, n\}$). On a

$$P(X = k) = \binom{n}{k} p^k q^{n-k} \quad \forall k \in \{0, \dots, n\} \quad \text{où } q = 1 - p$$

D'où

$$\begin{aligned} E(X) &= np \\ V(X) &= npq \end{aligned}$$

III.4-d) Loi géométrique sur \mathbb{N}^* ou \mathbb{N}

On se donne X une variable aléatoire suivant une **loi géométrique (sur \mathbb{N}^*)** de paramètre $p \in]0; 1[$, c'est-à-dire que

$$p(X = k) = (1 - p)^{k-1} p \quad \forall k \in \mathbb{N}^*$$

On a

$$\begin{aligned} E(X) &= \frac{1}{p} \\ V(X) &= \frac{q}{p^2} \quad \text{où } q = 1 - p \end{aligned}$$

Démonstration :

On note ici $q = 1 - p$.

• Espérance :

★ Convergence absolue de $\sum_{k \in \mathbb{N}^*} k P(X = k) = p \sum_{k \in \mathbb{N}^*} k q^{k-1}$:

Tous les termes de la série sont positifs, la convergence absolue revient donc à la convergence.

★ Convergence et somme totale : On reconnaît, à la constante p près le terme général de la "série géométrique dérivée" de raison q , avec

$$\sum_{k=1}^{+\infty} k q^{k-1} = \frac{1}{(1-q)^2} = \frac{1}{p^2}.$$

On obtient donc l'existence de l'espérance, avec

$$\mathbb{E}[X] = \sum_{k=1}^{+\infty} k (1-p)^{k-1} p = \frac{1}{p}$$

• Moment d'ordre 2 :

Méthode 1 :

★ Convergence absolue de $\sum_{k \in \mathbb{N}^*} k^2 P(X = k) = p \sum_{k \in \mathbb{N}^*} k^2 q^{k-1}$:

Tous les termes de la série sont positifs, la convergence absolue revient donc à la convergence.

★ Convergence et somme totale : Observons que, comme $q \neq 0$

$$\begin{aligned} \sum_{k \geq 1} k^2 q^{k-1} &= \sum_{k \geq 1} k(k-1+1) q^{k-1} \\ &= q \sum_{k \geq 1} k(k-1) q^{k-2} + \sum_{k \geq 1} k q^{k-1} \end{aligned}$$

Où on reconnaît deux séries géométrique "dérivée seconde" et "géométrique dérivée" convergentes et dont on connaît les sommes totales. Alors la série converge (et

donc l'espérance existe) et on a :

$$\begin{aligned} \mathbb{E}[X^2] &= p \sum_{k=1}^{+\infty} k^2 q^{k-1} = p \sum_{k=1}^{+\infty} k^2 q^{k-1} \\ &= pq \underbrace{\sum_{k=1}^{+\infty} k(k-1)q^{k-2}}_{=\frac{2}{(1-q)^3}} + p \underbrace{\sum_{k=1}^{+\infty} kq^{k-1}}_{=\frac{1}{1-q}^2} \\ &= p \frac{2q}{(1-q)^3} + p \frac{1}{(1-q)^2} = \frac{2q+1-q}{(1-q)^3} = \frac{q+1}{p^2} = \frac{2-p}{p^2} \end{aligned}$$

Méthode 2 :

Vérifions l'existence de $\mathbb{E}[X(X-1)]$:

$\mathbb{E}[X(X-1)]$ existe ssi la série $\sum_{k \in \mathbb{N}^*} k(k-1)p(X=k)$ converge absolument. Tous les termes de la série sont positifs, la convergence absolue revient donc à la convergence.

Soit $n \in \mathbb{N}$. On a

$$\begin{aligned} S_n &= \sum_{k=1}^n k(k-1)p(X=k) = \sum_{k=1}^n k(k-1)p(1-p)^{k-1} \\ &= p(1-p) \sum_{k=2}^n k(k-1)(1-p)^{k-2} \end{aligned}$$

Or, comme $|1-p| < 1$, cette série converge. Ainsi, $\mathbb{E}[X(X-1)]$ existe et vaut

$$\mathbb{E}[X(X-1)] = \sum_{k=1}^{+\infty} k(k-1)p(X=k) = p(1-p) \frac{2}{(1-(1-p))^2} = \frac{2(1-p)}{p^2}$$

De plus, on a

$$X^2 = X(X-1) + X$$

L'espérance de $X(X-1)$ et celle de X existent, ainsi, celle de X^2 existe également et de plus, par linéarité de l'espérance,

$$\mathbb{E}[X^2] = \mathbb{E}[X(X-1)] + \mathbb{E}[X] = \frac{2(1-p)}{p^2} + \frac{1}{p} = \frac{2-p}{p^2}$$

• Variance : X admet un moment d'ordre 2 et d'après la formule de Koenig-Huyghens, on a

$$V(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{2-p}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{2-p-1}{p^2} = \frac{q}{p^2}$$

□

Si maintenant X est une variable aléatoire suivant une loi géométrique sur \mathbb{N} de paramètre $p \in]0; 1[$, i.e.

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

On a

$$\begin{aligned} E(X) &= \frac{1}{p} - 1 = \frac{q}{p} \\ V(X) &= \frac{q}{p^2} \quad \text{où } q = 1-p \end{aligned}$$

Démonstration :

On pose $Y = X + 1$. Alors $Y \hookrightarrow \mathcal{G}(p)$, d'où

$$\mathbb{E}[X] = \mathbb{E}[Y] - 1 = \frac{1}{p} - 1 = \frac{1-p}{p} = \frac{q}{p}$$

et

$$V(X) = V(Y)$$

□

III.4-e) Loi de Poisson, $\mathcal{P}(\lambda)$

On se donne X une variable aléatoire suivant une loi de Poisson de paramètre $\lambda \in]0; +\infty[$.

On a

$$p(X=k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad \forall k \in \mathbb{N}$$

D'où

$$\begin{aligned} E(X) &= \lambda \\ V(X) &= \lambda \end{aligned}$$

Démonstration :

★ Convergence absolue de $\sum_{k \in \mathbb{N}} k e^{-\lambda} \frac{\lambda^k}{k!}$:

Tous les termes de la série sont positifs, la convergence absolue revient donc à la convergence.

★ Convergence et somme totale : $\sum_{k \in \mathbb{N}} k \frac{\lambda^k}{k!} = \sum_{k \in \mathbb{N}^*} k \frac{\lambda^k}{k!} = \sum_{k \in \mathbb{N}^*} \frac{\lambda^k}{(k-1)!} = \lambda \sum_{k \in \mathbb{N}} \frac{\lambda^k}{k!}$

On reconnaît une série exponentielle de paramètre λ , avec $\sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} = e^\lambda$

On obtient donc l'existence de l'espérance, avec

$$\mathbb{E}[X] = \sum_{k=1}^{+\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} e^\lambda = \lambda$$

• Moment d'ordre 2 :

★ Convergence absolue de $\sum_{k \in \mathbb{N}^*} k^2 e^{-\lambda} \frac{\lambda^k}{k!}$:

Tous les termes de la série sont positifs, la convergence absolue revient donc à la convergence.

★ Convergence et somme totale :

$$\begin{aligned}\sum_{k \in \mathbb{N}} k^2 \frac{\lambda^k}{k!} &= \sum_{k \in \mathbb{N}^*} k^2 \frac{\lambda^k}{k!} \\ &= \sum_{k \in \mathbb{N}^*} k(k-1+1) \frac{\lambda^k}{k!} \\ &= \sum_{k \in \mathbb{N}^*} k(k-1) \frac{\lambda^k}{k!} + \sum_{k \in \mathbb{N}^*} k \frac{\lambda^k}{k!}\end{aligned}$$

On reconnaît, à l'extrême droite, la série déjà étudiée ci-dessus, on sait donc que

$$\sum_{k=0}^{+\infty} k \frac{\lambda^k}{k!} = \lambda e^\lambda$$

De plus,

$$\sum_{k \in \mathbb{N}^*} k(k-1) \frac{\lambda^k}{k!} = \lambda^2 \sum_{k \in \mathbb{N}^*} \frac{\lambda^k}{k!}$$

Là encore, on reconnaît une série exponentielle, avec

$$\sum_{k=0}^{+\infty} k(k-1) \frac{\lambda^k}{k!} = \lambda^2 e^\lambda$$

On obtient donc l'existence de l'espérance de X^2 , avec

$$\mathbb{E}[X^2] = \lambda + \lambda^2$$

• Variance :

X admet un moment d'ordre 2 et d'après la formule de Koenig-Huyghens, on a

$$V(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \lambda + \lambda^2 - \lambda^2 = \lambda$$

□

Dans toute cette partie, on supposera données deux variables aléatoires discrètes X et Y , où on notera $Supp(X) \subset \{x_i, i \in \mathbb{N}\}$ et $Supp(Y) \subset \{y_j, j \in \mathbb{N}\}$, avec les x_i deux à deux distincts ou de probabilité nulle, de même que les y_j .

I Loi

I-1 Loi conjointe

Définition

On appelle *vecteur aléatoire discret* toute n -liste (X_1, \dots, X_n) de v.a.d. X_1, \dots, X_n définies sur un même espace probabilisé. En particulier, si $n = 2$, on appelle (X_1, X_2) un *couple* de v.a.d..

Exemple 1 :

On lance un dé une infinité de fois et on note respectivement X et Y les première et deuxième apparitions de 6, avec $X = 0$ si 6 n'apparaît jamais et $Y = 0$ si 6 n'apparaît qu'une fois maximum. Alors (X, Y) est un couple de var. al. discrètes.

Définition

Soit (X, Y) un couple de v.a.d. sur $(\Omega, \mathcal{P}(\Omega), p)$. On appelle *support* du couple (X, Y) et on note $Supp((X, Y))$ tout ensemble $S \subset (X, Y)(\Omega)$ tel que $P((X, Y) \in S) = 1$.

Remarque :

Trivialement, on a

$$Supp((X, Y)) = \{(X(w), Y(w)) \mid w \in \Omega\} \subset Supp(X) \times Supp(Y)$$

ou, de manière équivalente (en loi) **pour les variables discrètes**

$$Supp((X, Y)) = \{(x, y) \in Supp(X) \times Supp(Y) \mid P(X = x, Y = y) \neq 0\}$$

Exemple 2 :

On reprend l'exemple de X et Y resp. les première et deuxième apparitions de 6. Alors

$$Supp((X, Y)) \subset \{(x, y) \in \mathbb{N}^* \mid y > x\} \cup \underbrace{\{(x, 0) \mid x \in \mathbb{N}\}}_{\text{le 6 apparaît au plus une fois}}$$

le 6 apparaît au plus une fois

Définition

Soit (X, Y) un couple de v.a.d.. On appelle *loi du couple* (X, Y) , ou *loi conjointe des variables X et Y* l'application

$$P_{(X, Y)} : Supp((X, Y)) \rightarrow [0; 1] \\ (x, y) \mapsto p((X = x) \cap (Y = y))$$

Exemple 3 :

Commençons à chercher la loi conjointe dans le cas de l'exemple précédent. (1^{ère} et 2^{ème} apparition de 6.)

- $Supp((X, Y)) = \{(n, k) \in \mathbb{N} \mid 1 \leq n < k\} \cup \mathbb{N} \times \{0\}$.
- Soient $n, k \in Supp((X, Y))$. On cherche $p(X = n, Y = k)$:
 * Si $n = 0$, alors pour tout $k \in \mathbb{N}$, $p(X = n, Y = k) = 0$:

En effet, on sait que $X \hookrightarrow \mathcal{G}(\frac{1}{6})$ et donc $p(X = 0) = 0$, d'où, comme $(X = n, Y = k) \subset (X = n)$, on a, pour tout $k \in \mathbb{N}$

$$0 \leq p(X = 0, Y = k) \leq p(X = 0) = 0, \text{ d'où } p(X = n, Y = k) = 0$$

- * Si $1 \leq n < k$, alors $p(X = n, Y = k) = q^{k-2}p^2$, où $p = 1/6$ et $q = 1 - p$:

En effet, posons $A_i =$ "le 6 apparaît au rang i ." Alors

$$p(X = n, Y = k) = p(\overline{A_1} \cap \dots \cap \overline{A_{n-1}} \cap A_n \cap \overline{A_{n+1}} \cap \dots \cap \overline{A_{k-1}} \cap A_k)$$

puis par indépendance des A_i et parce que $p(A_i) = \frac{1}{6}$, on a bien

$$p(X = n, Y = k) = \underbrace{q \dots q}_{n-1 \text{ fois}} p \underbrace{q \dots q}_{\text{de } n+1 \text{ à } k-1} p = q^{n-1} p q^{k-1-n} p = q^{k-2} p^2$$

- * Si $n \in \mathbb{N}^*$, $P(X = n, Y = 0) = 0$:

C'est l'événement "6 n'apparaît qu'une seule fois et c'est au rang n ". Avec le système complet $\{(Y = 0)\} \cup \{(Y = n)\}_{k \geq 2}$, on peut écrire

$$P(X = n) = P(X = n, Y = 0) + \sum_{k=2}^{+\infty} P(X = n, Y = k) \\ = P(X = n, Y = 0) + \sum_{k=n+1}^{+\infty} q^{k-2} p^2 \\ = P(X = n, Y = 0) + pq^{n-1} \text{ (après calcul, laissé en exercice)}$$

D'où $P(X = n) = P(X = n, Y = 0) + P(X = n)$ et donc $P(X = n, Y = 0) = 0$ (Ce résultat peut se redémontrer un peu plus rapidement à l'aide des loi conditionnelles.)

- Conclusion :

$$\text{Supp}((X, Y)) = \{(n, k) \in \mathbb{N}^2 \mid 1 \leq n < k\}$$

$$P(X = n, Y = k) = p^2 q^{k-2} \quad \text{si } 1 \leq n < k$$

■ Exemple 4 :

Avec les variables de l'exemple précédent, calculons $P("X + Y \text{ impair}")$:

Par la décomposition en parité différente, on a

$$P("X + Y \text{ impair}") = P(\underbrace{"X \text{ pair et } Y \text{ impair}"}_A) + P("X \text{ impair et } Y \text{ pair}")$$

Or, comme $P(Y = 1) = 0$, on a $\{(Y = 2j + 1)\}_{j \in \mathbb{N}^*}$ sys. compl. de "Y imp.", d'où

$$\begin{aligned} P(A) &= P\left(X \text{ pair} \cap \left(\bigcup_{j=1}^{+\infty} Y = 2j + 1\right)\right) = \sum_{j=1}^{+\infty} P\left(\underbrace{X \text{ pair, } Y = 2j + 1}_{A_j}\right) \\ &= \sum_{j=1}^{+\infty} \left(\sum_{i=1}^j P(X = 2i, Y = 2j + 1)\right) \quad \text{car } \{(X = 2i)\}_{i=1 \dots j} \text{ sys. q-c. de } A_j. \\ &= \sum_{j=1}^{+\infty} \left(\sum_{i=1}^j p^2 q^{2j+1-2}\right) = p^2 q \sum_{j=1}^{+\infty} j(q^2)^{j-1} \\ &= p^2 q \frac{1}{(1-q^2)^2} = p^2 q \frac{1}{(1-q)^2(1+q)^2} \\ &= \frac{q}{(1+q)^2} \quad \text{car } 1-q = p \end{aligned}$$

et de même (laissé en exercice), on a

$$P("X \text{ impair, } Y \text{ pair}") = \frac{1}{(1+q)^2}$$

D'où la probabilité finale :

$$P("X + Y \text{ impair}") = \frac{q+1}{(1+q)^2} = \frac{1}{1+q}$$

I-2 Loi marginale

🌿 Définition

Soit (X, Y) un couple de v.a.d.. La loi de X est appelée *première loi marginale* de (X, Y) et la loi de Y est appelée *deuxième loi marginale* de (X, Y) .

Commentaires :

| On peut retrouver les lois marginales à partir de la loi conjointe :

Théorème 111

Soit (X, Y) un couple de v.a.d. de loi $p_{(X, Y)}$. Alors

$$\forall x \in \text{Supp}(X), \quad p(X = x) = \sum_{j=0}^{+\infty} p((X = x) \cap (Y = y_j))$$

$$\forall y \in \text{Supp}(Y), \quad p(Y = y) = \sum_{i=0}^{+\infty} p((X = x_i) \cap (Y = y))$$

Démonstration :

On démontre ceci pour $p(X = x)$ (le cas de $p(Y = y)$ étant similaire.)

Par définition de X , la famille $\{(Y = y_j)\}_{j \in \mathbb{N}}$ est un système complet de l'univers Ω . Alors

$$\begin{aligned} p(X = x) &= p((X = x) \cap \Omega) \\ &= p\left((X = x) \cap \bigsqcup_{j=0}^{+\infty} (Y = y_j)\right) \\ &= p\left(\bigsqcup_{j=0}^{+\infty} ((X = x) \cap (Y = y_j))\right) \\ &= \sum_{j=0}^{+\infty} p((X = x) \cap (Y = y_j)) \end{aligned}$$

□

? Exercice 1

On pose X et Y les premières et deuxième apparition de 6 dans une infinité de lancers. Déterminer la loi marginale de Y .

Solution

Tout d'abord, on note que $\text{Supp}(Y) = \mathbb{N} - \{0, 1\}$ car $\text{Supp}((X, Y)) \subset \mathbb{N}^* \times \llbracket 2, +\infty \llbracket$.

Ainsi, si $k \geq 2$,

$$\begin{aligned} p(Y = k) &= \sum_{n=0}^{+\infty} p((Y = k) \cap (X = n)) \\ &= \sum_{n=1}^{k-1} q^{k-2} p^2 = (k-1)q^{k-2} p^2 \end{aligned}$$

Définition

Soit (X, Y) un couple de v.a.d.

- Pour tout $y \in \text{Supp}(Y)$ tel que $p(Y = y) \neq 0$, on appelle *loi conditionnelle de X par-rapport à $[Y = y]$* , l'application

$$\begin{aligned} \text{Supp}(X) &\rightarrow [0; 1] \\ x &\mapsto p_{[Y=y]}(X = x) = \frac{p((X=x) \cap (Y=y))}{p(Y=y)} \end{aligned}$$

- Pour tout $x \in \text{Supp}(X)$ tel que $p(X = x) \neq 0$, on appelle *loi conditionnelle de Y par-rapport à $[X = x]$* , l'application

$$\begin{aligned} \text{Supp}(Y) &\rightarrow [0; 1] \\ y &\mapsto p_{[X=x]}(Y = y) = \frac{p((X=x) \cap (Y=y))}{p(X=x)} \end{aligned}$$

■ Exemple 5 :

Reprenons l'exemple traité jusqu'ici.

- Soit $k \geq 2$. *Loi conditionnelle de X par-rapport à $[Y = k]$* :
Si $k \geq 2$, on a $p(X = n) \neq 0$. La loi conditionnelle sachant $[Y = k]$ est bien définie. Tout d'abord, si k fixé, alors $\text{Supp}(X_{/Y=k}) = \llbracket 1; k-1 \rrbracket$. Ensuite, si $n \in \llbracket 1; k-1 \rrbracket$, on a

$$p_{[Y=k]}(X = n) = \frac{p(Y = k, X = n)}{p(Y = k)} = \frac{q^{k-2}p^2}{(k-1)q^{k-2}p^2} = \frac{1}{k-1}$$

On a

$$X_{/Y=k} \hookrightarrow \mathcal{U}_{\llbracket 1; k-1 \rrbracket} \quad (\text{intuitif?})$$

- Soit $n \geq 1$ *loi conditionnelle de Y par-rapport à $[X = n]$* :
Si $n \geq 1$, on a $p(X = n) \neq 0$. La loi conditionnelle sachant $[X = n]$ est bien définie. De plus,

$$\text{Supp}(Y_{/X=n}) = \llbracket n+1; +\infty \llbracket$$

et, pour tout $k \in \llbracket n+1; +\infty \llbracket$, on a

$$p_{[X=n]}(Y = k) = \frac{p(Y = k, X = n)}{p(X = n)} = \frac{q^{k-2}p^2}{q^{n-1}p} = pq^{k-n-1}$$

On se rend ainsi compte que

$$(Y - n)_{/X=n} \hookrightarrow \mathcal{G}(p)$$

Remarque :

Dans l'exemple précédent, tout se passe comme si on recommençait à zéro à partir de $X = n$ (*phénomène "sans mémoire" !*)

Commentaires :

On note que dans beaucoup de cas, on n'a pas à calculer la loi conditionnelle, mais l'expérience amène à la trouver directement, comme dans l'exemple suivant :

■ Exemple 6 :

On tire une boule au hasard dans une urne contenant n boules identiques numérotées de 1 à n . On note X le résultat. On tire ensuite X fois de manière indépendante sur une cible avec une probabilité p de succès : "atteindre la cible". On note Y le nombre de succès.

Déterminer la loi conditionnelle de Y sachant $(X = i)$ pour $i \in \llbracket 1, n \rrbracket$:

Pour i fixé, sachant que l'événement $(X = i)$ est réalisé, on observe que $Y_{/X=i}$ compte le nombre de succès dans une série d'épreuves de Bernoulli indépendantes de succès "atteindre son objectif" de probabilité p . Ainsi :

$$Y_{/X=i} \hookrightarrow \mathcal{B}(i, p)$$

II Corrélation

II-1 Espérances

Dans cette section, on se donnera un couple de v.a. discrètes (X, Y) . Dans le cadre de l'étude de la corrélation, on aura à étudier l'existence et la valeur de $\mathbb{E}[XY]$. Il existe pour nous deux façons d'aborder cette question :

- ★ La première consiste à considérer la variable $Z = XY$ comme une variable habituelle, en cherchant

$$\text{Supp}(Z) = \{z_k \mid k \in \mathbb{N}\} \quad \text{et} \quad P(Z = z) \text{ pour tout } z \in \text{Supp}(Z)$$

puis valider l'existence et calculer $\mathbb{E}[Z]$ de la manière habituelle.

- ★ La deuxième consiste à utiliser un résultat calculable directement à partir de la loi conjointe :

Théorème 112 de transfert

Soient

- X, Y deux variables finies de support indexé resp. par $\llbracket 0, N \rrbracket$ et $\llbracket 0, M \rrbracket$
- D un sous ensemble de \mathbb{R}^2 contenant $\text{Supp}(X, Y)$,
- $g : D \rightarrow \mathbb{R}$ une fonction

alors

$$E(g(X, Y)) = \sum_{i=0}^N \sum_{j=0}^M g(x_i, y_j) P(X = x_i, Y = y_j) = \sum_{j=0}^M \sum_{i=0}^N g(x_i, y_j) P(X = x_i, Y = y_j)$$

■ Exemple 7 :

On tire une boule au hasard dans une urne contenant n boules identiques numérotées de 1 à n . On note X le résultat. On tire ensuite X fois de manière indépendante sur une cible avec une probabilité p d'atteindre la cible. On note Y le nombre de succès. Calculons $\mathbb{E}[XY^2]$ (On admettra que $\sum_{k=1}^n k^3 = \frac{(n(n+1))^2}{4}$) :

$$\begin{aligned} \mathbb{E}[XY^2] &= \sum_{i=1}^n \sum_{j=0}^n ij^2 P(X=i, Y=j) = \sum_{i=1}^n \sum_{j=0}^i ij^2 P(X=i, Y=j) \\ &= \sum_{i=1}^n \sum_{j=0}^n ij^2 P(X=i) P_{X=i}(Y=j) = \frac{1}{n} \sum_{i=1}^n i \underbrace{\left(\sum_{j=0}^i j^2 P_{X=i}(Y=j) \right)}_{\mathbb{E}[(Y_{/X=i})^2]} \end{aligned}$$

Or, on a $T = Y_{/X=i} \hookrightarrow \mathcal{B}(i, p)$ et donc $\mathbb{E}[T^2] = V_n(T) + \mathbb{E}[T]^2 = ip(1-p) + i^2 p^2$, d'où

$$\begin{aligned} \mathbb{E}[XY^2] &= \frac{p(1-p)}{n} \sum_{i=1}^n i^2 + \frac{p^2}{n} \sum_{i=0}^n i^3 \\ &= p(1-p) \frac{(n+1)(2n+1)}{6} + p^2 \frac{n(n+1)^2}{4} \end{aligned}$$

et donc

$$\mathbb{E}[XY^2] = p(n+1) \left(\frac{(1-p)(2n+1)}{6} + \frac{pn(n+1)}{4} \right)$$

? Exercice 2

En utilisant le théorème de transfert, montrer que dans l'exemple précédent,

$$\mathbb{E}[Y] = \frac{n+1}{2} p \quad \text{ainsi que} \quad \mathbb{E}[Y^2] = \frac{p(n+1)}{6} (3 + 2pn - 2p).$$

(On pourra remarquer que $Y = X^0 Y$.) En déduire $V(Y)$.

II-2 Covariance

🌿 Définition

Soit (X, Y) un couple de v.a.d. Si la variable aléatoire $(X - E(X))(Y - E(Y))$ admet une espérance, on appelle *covariance du couple (X, Y)* la valeur

$$\text{Cov}(X, Y) = \mathbb{E}[(X - E(X))(Y - E(Y))]$$

En général, on n'utilise pas cette formule pour trouver la covariance en pratique. En effet, la formule peut nettement se simplifier (et peut même faire office de définition)

Théorème 113 (Formule de Huyghens-Koenig)

Soit (X, Y) un couple de v.a.d. Si $\mathbb{E}[X], \mathbb{E}[Y]$ et $\mathbb{E}[XY]$ existent, alors le couple (X, Y) admet une covariance qui peut être calculée grâce à la formule

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$$

Démonstration :

★ Existence de $\text{Cov}(X, Y)$:

Il s'agit de montrer que $\mathbb{E}[(X - E(X))(Y - E(Y))]$ existe :

$$\begin{aligned} |(X - E(X))(Y - E(Y))| &= |XY + \mathbb{E}[XY] - X\mathbb{E}[X] - Y\mathbb{E}[Y]| \\ &\leq |XY| + \mathbb{E}[|XY|] + |X|\mathbb{E}[|X|] + |Y|\mathbb{E}[|Y|] \end{aligned}$$

Or, par hypothèse, l'espérance de $|XY|$, X et Y existent, ce qui signifie que l'espérance de $(X - E(X))(Y - E(Y))$ existe.

★ Valeur de $\text{Cov}(X, Y)$:

Il suffit de développer l'espérance par linéarité de l'espérance. □

On peut même encore un peu simplifier les hypothèses de la formule de H-K dans certains cas :

Propriété 114

Soit (X, Y) un couple de v.a. discrètes. Si les variables X et Y admettent un moment d'ordre 2, alors l'espérance $\mathbb{E}[XY]$ existe.

Démonstration :

XY est une variable aléatoire définie sur le même espace probabilisé que X et Y . Or, On rappelle que, pour tout $x, y \in \mathbb{R}$, on a

$$x^2 + y^2 - 2xy = \underbrace{(x - y)^2}_{\geq 0}$$

ce qui donne

$$xy \leq \frac{1}{2} (x^2 + y^2)$$

Appliqué à $|X|$ et $|Y|$, on obtient

$$|XY| \leq \frac{1}{2} (X^2 + Y^2)$$

Par hypothèse, X et Y admettent un moment d'ordre 2, ainsi, $\mathbb{E}[X^2]$ et $\mathbb{E}[Y^2]$ existent, d'où $\mathbb{E}[X^2 + Y^2]$ existe.

Par comparaison de variables, l'espérance de XY existe. □

■ Exemple 8 :

On effectue une série infinie de lancers indépendants d'un dé équilibré et on note A le rang d'apparition du premier 1, et B le rang d'apparition du premier 2. Montrons que $\mathbb{E}[AB]$ existe.

On sait que $A, B \hookrightarrow \mathcal{G}(\frac{1}{6})$. On en déduit que A et B admettent des moments d'ordre 2, ainsi on peut dire que $\mathbb{E}[AB]$ existe.

(Mais on n'a pas la valeur pour autant... Le calcul est HP mais pour information, on a $\mathbb{E}[AB] = \frac{6-2p}{p^2}$)

Corollaire

Soit (X, Y) un couple de v.a.d. Si les variables X et Y admettent un moment d'ordre 2, alors le couple (X, Y) admet une covariance et

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

Démonstration :

Les hypothèses permettent d'appliquer le théorème qui précise que l'espérance de XY existe. Ensuite, le théorème de Koenig Huyghens permet de conclure. □

■ Exemple 9 :

Si A et B sont les moments de première apparition respectivement de 1 et de 2 dans une série infinie de lancers, cherchons $Cov(A, B)$ en admettant que l'on connaît $\mathbb{E}[AB]$:

D'après le début de l'exemple déjà traité précédemment, on sait que $A, B \hookrightarrow \mathcal{G}(\frac{1}{6})$ et admettent chacune un moment d'ordre 2. Ainsi, $\mathbb{E}[AB]$, $Cov(A, B)$, $\mathbb{E}[A]$ et $\mathbb{E}[B]$ existent, avec comme formule

$$Cov(A, B) = \mathbb{E}[AB] - \mathbb{E}[A]\mathbb{E}[B]$$

Après calcul de $\mathbb{E}[AB]$ (admis) on trouverait (avec $p = \frac{1}{6}$)

$$\mathbb{E}[XY] = \frac{6-2p}{p^2}$$

De plus, on sait que $\mathbb{E}[A] = \mathbb{E}[B] = \frac{1}{p} = 6$. D'où

$$Cov(A, B) = \frac{6-2p}{p^2} - \frac{1}{p^2} = \frac{5-2p}{p^2} = 132$$



Remarque :

Si les variables sont finies, alors la covariance existe toujours.

? Exercice 3

On tire une boule au hasard dans une urne contenant n boules identiques numérotées de 1 à n . On note X le résultat. On tire ensuite X fois de manière indépendante sur une cible avec une probabilité d'atteindre son objectif de p . On note Y le nombre de succès. Déterminer $Cov(X, Y)$.

Solution

On sait d'après la formule de H-K, que

$$Cov(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

D'après le théorème de transfert, on sait que

$$\begin{aligned} \mathbb{E}[XY] &= \sum_{i=1}^n \sum_{j=0}^n ijP(X=i, Y=j) = \sum_{i=1}^n \sum_{j=0}^i ijP(X=i, Y=j) \\ &= \sum_{i=1}^n \sum_{j=0}^n ijP(X=i)P_{X=i}(Y=j) = \frac{1}{n} \sum_{i=1}^n i \underbrace{\left(\sum_{j=0}^i j \binom{j}{i} p^j (1-p)^{i-j} \right)}_{\mathbb{E}[Y/X=i]} \end{aligned}$$

Or,

$$\mathbb{E}[Y/X=i] = ip$$

d'où

$$\mathbb{E}[XY] = \frac{1}{n} \sum_{i=1}^n i^2 p = \frac{p}{n} \frac{n(n+1)(2n+1)}{6} = p \frac{(n+1)(2n+1)}{6}$$

Ainsi

$$\begin{aligned} Cov(X, Y) &= p \frac{(n+1)(2n+1)}{6} - \frac{n+1}{2} \mathbb{E}[Y] = p \frac{(n+1)(2n+1)}{6} - p \left(\frac{n+1}{2} \right)^2 \\ &= p \frac{(n+1)}{12} (2(2n+1) - 3(n+1)) \end{aligned}$$

i.e.

$$Cov(X, Y) = p \frac{(n+1)(n-1)}{12}$$

Propriété 115

Soient X, Y des v.a.d. définies sur un même espace probabilisé admettant chacune un moment d'ordre 2. On a

- $Cov(X, X) = V(X)$
- $Cov(X, Y) = Cov(Y, X)$

Propriété 116 (bilinéarité de la covariance)

Soient X, Y des v.a.d définies sur un même espace probabilisé admettant chacune un moment d'ordre 2. On a

- $Cov(\cdot, Y) : X \mapsto Cov(X, Y)$ est linéaire au sens suivant :
Si X, X' sont deux v.a.d. telles que $\mathbb{E}[XY]$ et $\mathbb{E}[X'Y]$ existent, alors
 $\forall \alpha, \beta \in \mathbb{R}, Cov(\alpha X + \beta X', Y) = \alpha Cov(X, Y) + \beta Cov(X', Y)$
- $Cov(X, \cdot) : Y \mapsto Cov(X, Y)$ est linéaire au sens suivant :
Si Y, Y' sont deux v.a.d. telles que $\mathbb{E}[XY]$ et $\mathbb{E}[XY']$ existent, alors
 $\forall \alpha, \beta \in \mathbb{R}, Cov(X, \alpha Y + \beta Y') = \alpha Cov(X, Y) + \beta Cov(X, Y')$

■ Exemple 10 :

Soient X, Y deux variables discrètes admettant la même variance. Montrons que $Cov(X + Y, X - Y) = 0$:

Par bilinéarité de la covariance, on a

$$\begin{aligned} Cov(X + Y, X - Y) &= Cov(X, X - Y) + Cov(Y, X - Y) \\ &= \underbrace{Cov(X, X)}_{V(X)} - Cov(X, Y) + Cov(Y, X) - \underbrace{Cov(Y, Y)}_{V(Y)} \\ &= V(X) - V(Y) \quad \text{par symétrie de la covariance} \\ &= 0 \end{aligned}$$

Théorème 117

Soient X, Y des v.a.d. définies sur un même espace probabilisé admettant chacune un moment d'ordre 2. On a

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

Démonstration :

On applique les définitions et la linéarité de l'espérance. \square



Remarque :

Ce théorème permet aussi dans certains cas de calculer $Cov(X, Y)$ avec la formule :

$$Cov(X, Y) = \frac{1}{2}(V(X + Y) - V(X) - V(Y))$$

■ Exemple 11 :

On effectue n lancers indépendants d'une même dé équilibré. On note X le nombre de fois où on obtient 1 et Y le nombre de fois où on obtient 2. On cherche à trouver $Cov(X, Y)$.

Tout d'abord, on remarque que X et Y sont deux variables finies telles que

$$X, Y \hookrightarrow \mathcal{B}(n, p), \quad \text{avec } p = \frac{1}{6}$$

Ceci garantit l'existence des moments d'ordre 2 de X, Y et valide donc la formule

$$Cov(X, Y) = \frac{1}{2}(V(X + Y) - V(X) - V(Y))$$

De plus, la variable $X + Y$ correspond au nombre total de fois où on rencontre 1 ou 2. Or, comme on ne peut pas avoir en même temps 1 et 2, ceci est en réalité, le nombre de réalisation du succès "obtenir 1 ou 2" dans une série de n tentatives indépendantes. Ainsi, on sait que

$$X + Y \hookrightarrow \mathcal{B}(n, \alpha) \quad \text{avec } \alpha = \frac{1}{3} = 2p$$

D'où

$$Cov(X, Y) = \frac{1}{2}(n\alpha(1 - \alpha) - 2np(1 - p)) = -np^2 = -\frac{n}{36}$$

Corollaire

Soient X, Y des v.a.d. définies sur un même espace probabilisé admettant chacune un moment d'ordre 2. On a

$$|Cov(X, Y)| \leq \sigma(X)\sigma(Y)$$

Démonstration :

Observons le polynôme en λ : $V(Y + \lambda X) = V(Y) + \lambda^2 V(X) + 2\lambda Cov(X, Y)$.

La variance étant toujours positive, c'est un polynôme n'admettant qu'au maximum une racine, c'est-à-dire de discriminant négatif.

Ainsi, $0 \geq \Delta = 4Cov(X, Y)^2 - 4V(X)V(Y)$, D'où

$$Cov(X, Y)^2 \leq V(X)V(Y)$$

ce qui donne le résultat annoncé. \square

II-3 Coefficient de corrélation linéaire

Définition

Soient X, Y des v.a.d. sur un même espace probabilisé, admettant chacune un moment d'ordre 2 **non nul**. On appelle *coefficient de corrélation linéaire* de X et Y la valeur

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

Exemple 12 :

Revenons à l'exemple des tirs sur une cible. (On admettra que $V(X) = \frac{n^2-1}{12}$.) On a alors, grâce aux calculs effectués (dans les exemples et exercices précédents)

$$r(X, Y) = \frac{p \frac{(n+1)(n-1)}{12}}{\sqrt{\frac{n^2-1}{12}} \sqrt{\frac{p(n+1)}{12} (6 + pn - 7p)}} = \sqrt{\frac{p(n-1)}{pn + 6 - 7p}}$$

Propriété 118

Soient X, Y des v.a.d.

- sur un même espace probabilisé,
- admettant chacune un moment d'ordre 2 **non nul**,

$$-1 \leq r(X, Y) \leq 1$$

Propriété 119

Soient X, Y des v.a.d.

- sur un même espace probabilisé,
- admettant chacune un moment d'ordre 2 **non nul**,

alors

$$|r(X, Y)| = 1 \Leftrightarrow \text{il existe } a \in \mathbb{R}^*, b \in \mathbb{R} \text{ tels que } P(Y = aX + b) = 1$$

Démonstration :

En reprenant la démonstration du corollaire précédent (partie 2a), on constate que

$$|r(X, Y)| = 1 \Leftrightarrow Cov(X, Y)^2 \leq V(X)V(Y) \Leftrightarrow \Delta = 0$$

Or, $\Delta = 0$ équivaut exactement à dire que le polynôme en λ : $V(X + \lambda Y) = V(X) + \lambda^2 V(Y) + 2\lambda Cov(X, Y)$ admet une racine (double). On note λ_0 cette racine. Ainsi,

$$|r(X, Y)| = 1 \Leftrightarrow V(Y + \lambda_0 X) = 0$$

Or, une variable admet une variance nulle ssi elle est constante presque sûrement, c'est-à-dire s'il existe $b \in \mathbb{R}$ tel que

$$P(Y + \lambda_0 X = b) = 1$$

On note que $\lambda_0 \neq 0$ car sinon on aurait $V(Y) = 0$ ce qui est contraire à l'énoncé. En conclusion, avec les notations précédentes

$$|r(X, Y)| = 1 \Leftrightarrow P(Y = \underbrace{-\lambda_0}_{a} X + b) = 1$$

□

II-4 Indépendance

Définition

Étant donné un couple de variables aléatoires discrètes (X, Y) sur un espace probabilisé $(\Omega, \mathcal{P}(\Omega), p)$, on dit que X et Y sont *indépendantes* si

$$p((X = x) \cap (Y = y)) = p(X = x)p(Y = y) \quad \forall x \in \text{Supp}(X), y \in \text{Supp}(Y)$$

Contre-Exemple(s) :

On lance un dé une infinité de fois. On reprend les variables aléatoires X et Y donnant le rang de première et deuxième apparition de 6. Alors

$$P(X = 5 \cap Y = 3) = 0 \neq \underbrace{P(X = 5)}_{\neq 0} \underbrace{P(Y = 3)}_{\neq 0}$$

Les deux variables aléatoires ne peuvent donc pas être indépendantes.

? Exercice 4

On lance une infinité de fois deux dés de manière indépendante. On note X le rang de première apparition de 6 sur le premier dé et Y le rang de première apparition d'un double. (Y vaut 0 s'il n'y a jamais de double.) Montrer que X et Y sont indépendantes.

Solution

X et Y suivent une loi géométrique, avec

$$X \hookrightarrow \mathcal{G}\left(\frac{1}{6}\right) \quad \text{et} \quad Y \hookrightarrow \mathcal{G}\left(\frac{6}{36}\right) = \mathcal{G}\left(\frac{1}{6}\right)$$

Vérifions que

$$P(X = i \cap Y = j) = P(X = i)P(Y = j) \quad \forall i, j \in \mathbb{N}^*$$

il y a 25 couples qui ne contiennent ni un 6 sur le premier dé, ni un double. Ainsi,

• Si $0 < i < j$:

$$P(X = i \cap Y = j) = \underbrace{\left(\frac{25}{36}\right)^{i-1}}_{\text{ni 6 ni double jusqu'à } i / 6 \text{ à } i \text{ non double}} \underbrace{\frac{5}{36}}_{\text{pas de double}} \underbrace{\left(\frac{5}{6}\right)^{j-i-1}}_{\text{pas de double jusqu'à } j-1} \underbrace{\frac{1}{6}}_{\text{double à } j}$$

réorganisé autrement, on trouve bien

$$P(X = i \cap Y = j) = \left(\frac{5}{6}\right)^{i-1} \left(\frac{5}{6}\right)^{i-1} \frac{1}{6} \frac{5}{6} \left(\frac{5}{6}\right)^{j-i-1} \frac{1}{6} = \left(\frac{5}{6}\right)^{i-1} \frac{1}{6} \left(\frac{5}{6}\right)^{j-1} \frac{1}{6} = P(X = i)P(Y = j)$$

• Si $0 < j < i$:

$$P(X = i \cap Y = j) = \underbrace{\left(\frac{25}{36}\right)^{j-1}}_{\text{ni 6 ni double jusqu'à } j / \text{double à } j, \text{ distinct de 6}} \underbrace{\frac{5}{36}}_{\text{pas de 6}} \underbrace{\left(\frac{5}{6}\right)^{j-i-1}}_{\text{pas de 6 jusqu'à } j-1} \underbrace{\frac{1}{6}}_{6 \text{ à } j}$$

De même qu'avant, on trouve

$$P(X = i \cap Y = j) = P(X = i)P(Y = j)$$

• Si $0 < i = j$:

$$P(X = i \cap Y = j) = \underbrace{\left(\frac{25}{36}\right)^{j-1}}_{\text{ni 6 ni double jusqu'à } j} \underbrace{\frac{1}{36}}_{\text{double 6 à } j} = \left(\frac{5}{6}\right)^{j-1} \frac{1}{6} \left(\frac{5}{6}\right)^{j-1} \frac{1}{6}$$

Or, $i = j$, on retrouve donc bien

$$P(X = i \cap Y = j) = P(X = i)P(Y = j)$$

Conclusion, X et Y sont bien indépendantes.

Propriété 120

X et Y sont indépendantes ssi l'une des assertions suivantes est vérifiée

- $X_{/Y=y}$ a même loi que X pour tout $y \in \text{Supp}(Y)$ de probabilité non nulle.
- la loi de $X_{/Y=y}$ est indépendante de y .

Démonstration :

• Première assertion :

pour tout $x \in \text{Supp}(X)$, on a

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

Ainsi, si les variables $X_{/Y=y}$ et X ont même loi, on obtient

$P(X = x \cap Y = y) = P(X = x)P(Y = y)$, donc X et Y sont indépendantes

et si les variables X et Y sont indépendantes, alors par simplification de

$$P(X = x)P(Y = y) = P_{Y=y}(X = x)P(Y = y)$$

on obtient

$$P(X = x) = P_{Y=y}(X = x) \text{ donc } X_{/Y=y} \text{ et } X \text{ ont même loi.}$$

• Deuxième assertion :

Si X et Y sont indépendantes, alors clairement $P_{Y=y}(X = x) = \frac{P(X=x \cap Y=y)}{P(Y=y)} = P(X = x)$ est indépendante de la valeur y de Y .

Si $P_{Y=y}(X = x)$ est indépendant de y , notons $u_i = P_{Y=y}(X = x_i)$ pour tout i et $y \in \text{Supp}(Y)$. Alors, pour tout i , on a

$$P(X = x_i, Y = y_j) = u_i P(Y = y_j)$$

En sommant sur $j \in \mathbb{N}$, on obtient par formule des probas totales :

$$P(X = x_i) = u_i$$

et donc finalement

$$\square \quad P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j) \quad \text{i.e. } X \text{ et } Y \text{ ind.}$$

■ Exemple 13 :

Dans l'exemple de X et Y resp. les 1^{ères} et 2^{èmes} apparition du 6 dans une série infinie de lancers indépendants, on a vu que pour tout $n \in \mathbb{N}^*$, on a

$$X_{/(Y=n)} \hookrightarrow \mathcal{U}_{\{1, \dots, n\}}$$

La loi de $X_{/(Y=n)}$ dépend donc de la valeur de Y , ce qui signifie que les variables X et Y ne sont pas indépendantes. (On s'en doutait un peu!!)

■ Exemple 14 :

Si on a des variables X, Y telles que pour tout $n \in \text{Supp}(Y)$,

$$X_{/Y=n} \hookrightarrow \mathcal{G}\left(\frac{1}{2}\right)$$

Alors X et Y sont indépendantes et de surcroît, $X \hookrightarrow \mathcal{G}\left(\frac{1}{2}\right)$.

⚠ Remarque :

| Le résultat précédent est évidemment également vrai en échangeant X et Y .

Lemme 121

Soit (X, Y) un couple de v.a.d. indépendantes, alors,

- pour tous $a, b \in \mathbb{R}$, $X - a$ et $Y - b$ sont indépendantes.
- mieux, pour toutes fonctions réelles g, h , les v.a.d. $g(X)$ et $h(Y)$ sont indépendantes.

Démonstration :

- la première est conséquence de la deuxième.
- On note $\text{Supp}(X) = \{x_i \mid i \in \mathbb{N}\}$ et $\text{Supp}(Y) = \{y_j \mid j \in \mathbb{N}\}$. Alors

$$\text{Supp}(g(X)) = \{g(x_i) \mid i \in \mathbb{N}\} \quad \text{Supp}(h(Y)) = \{h(y_j) \mid j \in \mathbb{N}\}$$

Ainsi, pour tout $k \in \text{Supp}(g(X)), l \in \text{Supp}(h(Y))$, on a

$$\begin{aligned} P(g(X) = k \cap h(Y) = l) &= P(X \in g^{-1}(k) \cap Y \in h^{-1}(l)) \\ &= P(X \in g^{-1}(k))P(Y \in h^{-1}(l)) \\ &= P(g(X) = k)P(h(Y) = l) \end{aligned}$$

□

Théorème 122

Soit (X, Y) un couple de v.a.d.

- Si $\mathbb{E}[X]$ et $\mathbb{E}[Y]$ existent
- Si X et Y sont indépendantes,

alors $\mathbb{E}[XY]$ existe et

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$

Démonstration : admise. □

⚠ LA RÉCIPROQUE DE L'INDÉPENDANCE EST FAUSSE !

Exemple : Soit X la variable aléatoire de loi uniforme sur $\{-1, 0, 1\}$.

$$p(X = -1) = p(X = 0) = p(X = 1) = \frac{1}{3}.$$

On pose $Y = X^2$, d'où

$$\text{Supp}(Y) = \{0, 1\}, \quad p(Y = 0) = \frac{1}{3}, \quad p(Y = 1) = \frac{2}{3}$$

Y et X sont clairement non indépendants! (pour preuve, on peut par exemple vérifier que $p(X = 0 \cap Y = 0) \neq p(X = 0)p(Y = 0)$).

Or, notons que $XY = X$, ainsi,

$$\mathbb{E}[XY] = \mathbb{E}[X] = 0 = \mathbb{E}[X]\mathbb{E}[Y].$$

Corollaire

Soit (X, Y) un couple de v.a.d.

- Si $\mathbb{E}[X]$ et $\mathbb{E}[Y]$ existent
- Si X et Y sont indépendantes, alors $\text{Cov}(X, Y)$ existe et

$$\text{Cov}(X, Y) = 0$$

III Convolution

III-1 Généralités

Dans cette section, on se donne toujours X et Y deux variables aléatoires discrètes et on note

$$\text{Supp}(X) = \{x_i \mid i \in \mathbb{N}\} \quad \text{et} \quad \text{Supp}(Y) = \{y_j \mid j \in \mathbb{N}\}$$

où les x_i (puis les y_j) sont deux à deux distincts de probabilité nulle.

But : Le but de ce paragraphe est de pouvoir déterminer la loi de la v.a.d. $X + Y$.

Stratégie : Utiliser la partition d'événements $\{(Y = y_j)\}_{j \in \mathbb{N}}$ ou $\{(X = x_i)\}_{i \in \mathbb{N}}$

En effet, si $x \in \mathbb{R}$, on peut écrire :

$$\begin{aligned} P(X + Y = x) &= P\left(\bigsqcup_{i \in \mathbb{N}} (X + Y = x) \cap (X = x_i)\right) = P\left(\bigsqcup_{i \in \mathbb{N}} (x_i + Y = x) \cap (X = x_i)\right) \\ &= P\left(\bigsqcup_{i \in \mathbb{N}} (Y = x - x_i) \cap (X = x_i)\right) = \sum_{i=0}^{+\infty} P\left((Y = x - x_i) \cap (X = x_i)\right) \end{aligned}$$

Par exemple, si X et Y sont indépendantes, alors

$$P(X + Y = x) = \sum_{i=0}^{+\infty} P(Y = x - x_i)P(X = x_i)$$

Définition et Proposition

Soit (X, Y) un couple de v.a.d. de loi respectives $\mathcal{L}(X)$ et $\mathcal{L}(Y)$. On appelle *convolution* des lois de X et Y on note $\mathcal{L}(X) \star \mathcal{L}(Y) : \mathbb{R} \rightarrow \mathbb{R}$ l'application définie par

$$\mathcal{L}(X) \star \mathcal{L}(Y)(x) = \sum_{i=0}^{+\infty} P(X = x_i)P(Y = x - x_i)$$

où la série $\sum_{i \in \mathbb{N}} P(X = x_i)P(Y = x - x_i)$ est convergente

Démonstration :

La série $\sum_{i \in \mathbb{N}} P(X = x_i)P(Y = x - x_i)$ est à termes positifs et on a

$$\forall i \in \mathbb{N}, \quad 0 \leq P(X = x_i)P(Y = x - x_i) \leq P(X = x_i)$$

où $P(X = x_i)$ est le terme général d'une série convergente. Ainsi, par le théorème de comparaison des séries à termes positifs, on a la convergence de $\sum_{i \in \mathbb{N}} P(X = x_i)P(Y = x - x_i)$

Démonstration :

Comme $\mathbb{E}[X]$ et $\mathbb{E}[Y]$ existent, alors $\mathbb{E}[(X - \mathbb{E}[X])]$ et $\mathbb{E}[(Y - \mathbb{E}[Y])]$ existent. De plus, comme X et Y sont indépendantes, alors $X - \mathbb{E}[X]$ et $Y - \mathbb{E}[Y]$ sont également indépendantes. (lemme).

D'après le théorème précédent, on sait donc que $\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$ existe, ce qui assure l'existence de la covariance.

Par simple calcul, on obtient encore une fois

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

ce qui vaut 0 par indépendance de X et Y . \square

Corollaire

Soit (X, Y) un couple de v.a.d.

- Si X et Y admettent un moment d'ordre 2,
 - Si X et Y sont indépendantes,
- alors $V(X + Y)$ existe et

$$V(X + Y) = V(X) + V(Y)$$

Démonstration :

On utilise la décomposition de $V(X + Y)$ grâce à la covariance :

$$V(X + Y) = V(X) + V(Y) + 2\text{Cov}(X, Y)$$

puis $\text{Cov}(X, Y) = 0$ par le corollaire précédent. \square

Corollaire

Soient X, Y des v.a.d.

- sur un même espace probabilisé,
- admettant chacune un moment d'ordre 2 **non nul**,
- **indépendantes**,

alors $r(X, Y) = 0$.

Démonstration :

Si X et Y sont indépendantes, alors la covariance est nulle! \square

(et même, de plus $0 \leq \sum_{i=0}^{+\infty} P(X = x_i)P(Y = x - x_i) \leq \sum_{i=0}^{+\infty} P(X = x_i) = 1$.) \square

Théorème 123

Soit (X, Y) un couple de v.a.d. **indépendantes**, alors la loi de $X + Y$ est définie par

$$\mathcal{L}(X) \star \mathcal{L}(Y).$$

Démonstration :

La démonstration est le raisonnement effectué juste avant le théorème. \square

Cas particuliers :

- Si X, Y sont des variables aléatoires discrètes indépendantes prenant des valeurs dans \mathbb{Z} :

Alors :

- ★ $Supp(X + Y) \subset \mathbb{Z}$, d'où, si $x \in \mathbb{R} - \mathbb{Z}$, on a $P(X + Y = x) = 0$.
- ★ Si $k \in \mathbb{Z}$, on a

$$P(X + Y = k) = \sum_{j=-\infty}^{+\infty} P(X = k - j)P(Y = j)$$

- Si X, Y sont des variables aléatoires discrètes indépendantes prenant des valeurs dans \mathbb{N} :

Alors :

- ★ $Supp(X + Y) \subset \mathbb{Z}$, d'où, si $x \in \mathbb{R} - \mathbb{N}$, on a $P(X + Y = k) = 0$
- ★ Si $k \in \mathbb{N}$, on a

$$\begin{aligned} P(X + Y = k) &= \sum_{j=0}^{+\infty} \underbrace{P(X = k - j)P(Y = j)}_{=0 \text{ si } k-j < 0} \\ &= \sum_{j=0}^k P(X = k - j)P(Y = j) \end{aligned}$$

III-2 Exemples

III.2-a) Convolution de Loïs binomiales

Théorème 124

Soient $n, m \in \mathbb{N}$ et $p \in [0; 1]$. Alors $\mathcal{B}(n, p) \star \mathcal{B}(m, p) = \mathcal{B}(n + m, p)$.

Ou autrement dit

Théorème 125

Soient $n, m \in \mathbb{N}$ et $p \in [0; 1]$. Si X et Y sont deux v.a.d. indépendantes telles que

$$X \hookrightarrow \mathcal{B}(n, p) \quad \text{et} \quad Y \hookrightarrow \mathcal{B}(m, p)$$

alors

$$X + Y \hookrightarrow \mathcal{B}(n + m, p)$$

Démonstration :

- Par le calcul : On se donne $X \hookrightarrow \mathcal{B}(n, p)$ et $Y \hookrightarrow \mathcal{B}(m, p)$ deux v.a.d. indépendantes $Supp(X + Y) = \{0, 1, \dots, n + m\}$. Ainsi,

$$\begin{aligned} \star \quad \text{Si } x \notin Supp(X + Y), \text{ on a } \quad P(X + Y = x) &= 0 \\ \star \quad \text{Si } x \in Supp(X + Y), \text{ on a } \quad P(X + Y = x) &= \sum_{j=0}^{+\infty} \underbrace{P(X = x - j)P(Y = j)}_{=0 \text{ si } x-j < 0} \\ &= \sum_{j=0}^x P(X = x - j)P(Y = j) \end{aligned}$$

On rappelle que si on n'a pas $0 \leq K \leq N$, la quantité $\binom{N}{K}$ est nulle. En utilisant cette convention, on peut écrire

$$\begin{aligned} P(X + Y = x) &= \sum_{j=0}^x \binom{n}{x-j} p^{x-j} (1-p)^{n-x+j} \binom{m}{j} p^j (1-p)^{m-j} \\ &= \sum_{j=0}^x \binom{n}{x-j} \binom{m}{j} p^x (1-p)^{n+m-x} = p^x (1-p)^{n+m-x} \underbrace{\sum_{j=0}^x \binom{n}{x-j} \binom{m}{j}}_{= \binom{n+m}{x}} \quad (\text{Vandermonde}) \\ &= \binom{n+m}{x} p^x (1-p)^{n+m-x} \end{aligned}$$

- On aurait également pu établir ce résultat de manière en expliquant que $X \hookrightarrow \mathcal{B}(n, p)$ modélise par exemple n expériences de Bernoulli indépendantes, dont le succès a comme probabilité p .

De même, $Y \hookrightarrow \mathcal{B}(m, p)$ indépendante de X modélise les mêmes expériences de Bernoulli, mais m fois.

Ainsi, "en tout", $X + Y$ modélise $n + m$ expériences de Bernoulli de probabilité de succès p . \square

III.2-b) Convolution de loïs de Poisson

Théorème 126

Soient $\lambda, \mu > 0$. Alors $\mathcal{P}(\lambda) \star \mathcal{P}(\mu) = \mathcal{P}(\lambda + \mu)$.

autrement dit

Corollaire

Soient $\lambda, \mu > 0$. Si X et Y sont deux v.a.d. indépendantes telles que

$$X \hookrightarrow \mathcal{P}(\lambda) \quad \text{et} \quad Y \hookrightarrow \mathcal{P}(\mu)$$

alors

$$X + Y \hookrightarrow \mathcal{P}(\lambda + \mu)$$

Démonstration :

On se donne X et Y sont deux v.a.d. indépendantes telles que

$$X \hookrightarrow \mathcal{P}(\lambda) \quad \text{et} \quad Y \hookrightarrow \mathcal{P}(\mu)$$

alors

- $\text{Supp}(X + Y) = \mathbb{N}$.

★ Si $x \notin \text{Supp}(X + Y) = \mathbb{N}$, on a

$$P(X + Y = x) = 0$$

★ Si $x \in \text{Supp}(X + Y) = \mathbb{N}$, on a

$$\begin{aligned} P(X + Y = x) &= \sum_{j=0}^{+\infty} \underbrace{P(X = x - j) P(Y = j)}_{=0 \text{ si } x-j < 0} \\ &= \sum_{j=0}^x P(X = x - j) P(Y = j) \\ &= \sum_{j=0}^x e^{-\lambda} \frac{\lambda^{x-j}}{(x-j)!} e^{-\mu} \frac{\mu^j}{j!} \\ &= e^{-\lambda-\mu} \sum_{j=0}^x \frac{\lambda^{x-j}}{(x-j)!} \frac{\mu^j}{j!} \\ &= \frac{e^{-\lambda-\mu}}{x!} \sum_{j=0}^x \binom{x}{j} \lambda^{x-j} \mu^j \\ &= \frac{e^{-(\lambda+\mu)}}{x!} (\lambda + \mu)^x \end{aligned}$$

On a donc bien $X + Y \hookrightarrow \mathcal{P}(\lambda + \mu)$ □

Dans ce chapitre, on pose $\mathbb{K} = \mathbb{R}$ ou \mathbb{C} ainsi que E et F deux \mathbb{K} espaces vectoriels.

I Généralités

I-1 Notion d'application linéaire

Définition

Soit $L : E \rightarrow F$ une application. On dit que L est *linéaire* si **les trois** conditions suivantes sont vérifiées :

- E, F sont deux espaces vectoriels sur \mathbb{K}
- $\forall u, v \in E, \quad L(u + v) = L(u) + L(v)$
- $\forall \alpha \in \mathbb{K}, \forall u \in E \quad L(\alpha u) = \alpha L(u)$

On note $\mathcal{L}(E; F)$ l'ensemble de toutes les applications linéaires de E dans F .

■ Exemple 1 (utilisable dans les exercices) :

Si E est un espace vectoriel, l'application identité $id : E \rightarrow E, u \mapsto u$ est linéaire.

En effet, $\forall u, v \in E, \alpha \in \mathbb{K}, id(u + v) = u + v = id(u) + id(v), id(\alpha u) = \alpha u = \alpha id(u)$

■ Exemple 2 :

$$L_1 : \mathbb{R}^2 \rightarrow \mathbb{R}^3 \quad \text{est linéaire car :}$$

$$(x, y) \rightarrow (2x, 3x - y, y - x)$$

\mathbb{R}^2 est un \mathbb{R} -ev. Soient alors $X = (x_1, x_2), Y = (y_1, y_2) \in \mathbb{R}^2$, et $\alpha \in \mathbb{R}$.

$$\begin{aligned} \text{On a } L(X + Y) &= L\left(\underbrace{(x_1 + x_2)}_x, \underbrace{(y_1 + y_2)}_y\right) = (2x, 3x - y, y - x) \\ &= (2x_1 + 2x_2, 3x_1 + 3x_2 - y_1 - y_2, y_1 + y_2 - x_1 - x_2) = L(X) + L(Y) \end{aligned}$$

$$L(\alpha X) = L\left(\underbrace{(\alpha x_1)}_x, \underbrace{(\alpha y_1)}_y\right) = (2\alpha x_1, 3\alpha x_1 - \alpha y_1, \alpha y_1 - \alpha x_1) = \alpha L(X)$$

■ Exemple 3 (utilisable dans les exercices) :

Si $E = \mathbb{R}^n; F = \mathbb{R}^m; A \in \mathcal{M}_{m,n}(\mathbb{K})$ et $L_1 : E \rightarrow F$ alors L_1 est linéaire.

$$X \mapsto AX$$

Application : Dans l'exemple 1, on peut écrire : $L_1 \left(\begin{pmatrix} x \\ y \end{pmatrix} \right) = \begin{pmatrix} 2 & 3 & -1 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$.

Elle est donc directement linéaire.

Proposition 127

Si $L : E \rightarrow F$ est une application linéaire, alors $L(0_E) = 0_F$

Démonstration :

$$L(0_E) = L\left(\underbrace{0}_\alpha \times \underbrace{0_E}_r\right) = 0 \times L(0_E) = 0_F \quad \square$$

Théorème 128 (Caractérisation des applications linéaires :)

Soient E et F deux \mathbb{K} espaces vectoriels et $L : E \rightarrow F$ une application. L est linéaire si et seulement si **l'une des** conditions suivantes est vérifiée :

- $\forall \alpha, \beta \in \mathbb{K}, u, v \in E \quad L(\alpha u + \beta v) = \alpha L(u) + \beta L(v)$
- $\forall \beta \in \mathbb{K}, u, v \in E \quad L(u + \beta v) = L(u) + \beta L(v)$

Démonstration :

• Si L est linéaire, alors

$$\forall \alpha, \beta \in \mathbb{K}, u, v \in E \quad L\left(\underbrace{\alpha u}_{\in E} + \underbrace{\beta v}_{\in E}\right) = L(\alpha u) + L(\beta v) = \alpha L(u) + \beta L(v)$$

• C'est i avec $\alpha = 1$.

• Si i ou ii est vérifiée, montrons que L est linéaire.

★ Si i est vérifiée :

Soient $u, v \in E$. On prend $\alpha = \beta = 1$ et on obtient le i de la définition. Soit $\alpha \in \mathbb{K}$ et $u \in E$. On prend $v = 0$ et on obtient le ii de la définition.

★ Si ii est vérifiée :

Soient $u, v \in E$. On prend $\beta = 1$ et on obtient le i de la définition. Pour le ii de la définition : Soit $\alpha \in \mathbb{K}$ et $u \in E$. Alors

$$\square \quad L(\alpha u) = L(0_E + \alpha u) = \underbrace{L(0_E)}_0 + \alpha L(u) = \alpha L(u)$$

■ Exemple 4 (utilisable dans les exercices) :

$$D : C^1(\mathbb{R}) \rightarrow C^0(\mathbb{R}) \quad \text{est linéaire :}$$

$$f \mapsto f'$$

$C^1(\mathbb{R})$ et $C^0(\mathbb{R})$ sont deux \mathbb{R} -ev.

Soient $f, g \in C^1(\mathbb{R})$ et $\beta \in \mathbb{R}$. On a

$$D(f + \beta g) = (f + \beta g)' = f' + \beta g' = D(f) + \beta D(g)$$

D est donc linéaire par théorème de caractérisation des applications linéaires.

■ Contre-Exemple(s) :

$$\begin{aligned} L: \mathcal{C}(\mathbb{R}) &\rightarrow \mathbb{R} \\ f &\rightarrow f(0)^2 \end{aligned} \quad \text{n'est pas une application linéaire.}$$

En effet, $\mathcal{C}(\mathbb{R})$ est bien un espace vectoriel, mais par exemple pour $\cos \in \mathcal{C}(\mathbb{R})$, $L(-\cos) = \cos^2(0) = 1 \neq -L(\cos)$.

 Définition

On se donne une application linéaire $L : E \rightarrow F$.

- Si $L(E) \subset E$, on dit que L est un *endomorphisme* de E . On note $\mathcal{L}(E)$ à la place de $\mathcal{L}(E; E)$.
- Si L est bijective, on dit que L est un *isomorphisme*. (E peut être différent de F !)
- On dit que deux espaces vectoriels E et F sont *isomorphes* s'il existe un isomorphisme $\varphi : E \rightarrow F$.
- Une application linéaire $L : E \rightarrow E$ est appelée *automorphisme* si c'est un endomorphisme bijectif. On note $Aut(E)$ ou $GL(E)$ l'ensemble des automorphismes de E .

■ Exemple 5 :

L'application linéaire (à prouver) "primitive qui s'annule en a "

$$\begin{aligned} L_a: \mathcal{C}^0(\mathbb{R}) &\rightarrow \mathcal{C}^1(\mathbb{R}) \\ f &\mapsto F: x \mapsto \int_a^x f \end{aligned}$$

- est un endomorphisme de $\mathcal{C}^0(\mathbb{R})$ car : pour tout $f \in \mathcal{C}^0(\mathbb{R})$, on a $L(f) \in \mathcal{C}^1(\mathbb{R}) \subset \mathcal{C}^0(\mathbb{R})$
- n'est pas un automorphisme de $\mathcal{C}^0(\mathbb{R})$ car : l'application $L : \mathcal{C}^0(\mathbb{R}) \rightarrow \mathcal{C}^0(\mathbb{R})$ n'est pas surjective; on rappelle par exemple que les fonctions continues ne sont pas toutes dérivables...
- n'est pas un isomorphisme de $\mathcal{C}^0(\mathbb{R})$ dans $\mathcal{C}^1(\mathbb{R})$: (là encore, non surjective car toute image de L s'annule en a , ce qui n'est pas le cas d'une fonction quelconque dans $\mathcal{C}^1(\mathbb{R})$.)

■ Exemple 6 :

$$\begin{aligned} \text{L'application } L: \mathbb{R}_3[X] &\rightarrow \mathbb{R}[X] \quad \text{est un endomorphisme.} \\ P &\mapsto XP' - 2P \end{aligned}$$

En effet, elle est **linéaire** (à faire) et de plus, pour tout $P \in \mathbb{R}_3[X]$, on a $\deg L(P) = \deg(XP' + X) \leq \max(\deg(XP'), \deg P) = \deg P \leq 3$. Ainsi,

$$L(P) \in \mathbb{R}_3[X]$$

et c'est donc bien un endomorphisme

■ Contre-Exemple(s) :

$$\begin{aligned} \text{Soit } n \in \mathbb{N}^*. \text{ L'application } L: \mathbb{R}_n[X] &\rightarrow \mathbb{R}[X] \quad \text{est une application li-} \\ P &\mapsto X^2P' - P \end{aligned}$$

néaire, mais pas un endomorphisme.

En effet, on a par exemple $X^n \in \mathbb{R}_n[X]$, mais $L(X^n) = nX^{n+1} - X^n \notin \mathbb{R}_n[X]$.

■ Contre-Exemple(s) :

$$\begin{aligned} \text{L'application "dérivée" } d: \mathcal{C}^1(\mathbb{R}) &\rightarrow \mathcal{C}^0(\mathbb{R}) \quad \text{n'est pas un isomorphisme.} \\ f &\mapsto f' \end{aligned}$$

En effet, l'application n'est pas injective. Par exemple, les deux applications $f : x \mapsto x+1$ et $g : x \mapsto x+2$ dans $\mathcal{C}^1(\mathbb{R})$ ont la même dérivée (donc la même image par L).

I-2 Opérations sur les applications linéaires

Proposition 129 (Combinaisons d'A.L.)

Si E et F sont deux \mathbb{K} -espaces vectoriels, l'ensemble $\mathcal{L}(E, F)$ est \mathbb{K} -un espace vectoriel.

Traduction : Toute combinaison linéaire finie d'applications linéaires est une application linéaire.

Démonstration :

$\mathcal{L}(E; F) \subset \mathcal{F}(E; F)$, où F est un espace vectoriel. Or, $\mathcal{F}(E; F)$ est un espace vectoriel, il suffit donc d'utiliser le théorème de caractérisation des sev.

• $\mathcal{L}(E; F)$ est non vide, car l'application nulle, notée ici $\bar{0}$, est linéaire :

$$\begin{aligned} \bar{0}(\alpha u + \beta v) &= 0_F \\ &= 0_F + 0_F \\ &= \alpha \bar{0}(u) + \beta \bar{0}(v) \end{aligned}$$

• Conservation des combinaisons linéaires :

On pose $f, g \in \mathcal{L}(E; F)$ et $\lambda \in \mathbb{K}$. Montrons que $f + \lambda g \in \mathcal{L}(E; F)$:
Soient $u, v \in E$ et $\alpha, \mu \in \mathbb{K}$. On a

$$\begin{aligned} (f + \lambda g)(\alpha u + \beta v) &= f(\alpha u + \beta v) + \lambda g(\alpha u + \beta v) \\ &= (\alpha f(u) + \beta f(v)) + \lambda(\alpha g(u) + \beta g(v)) \\ &= \alpha(f(u) + \lambda g(u)) + \mu(f(v) + \lambda g(v)) \\ &= \alpha(f + \lambda g)(u) + \beta(f + \lambda g)(v) \end{aligned}$$

D'où $f + \lambda g \in \mathcal{L}(E, F)$ \square

■ Exemple 7 :

$$\begin{aligned} L: C^1(\mathbb{R}) &\rightarrow C^0(\mathbb{R}) \quad \text{est linéaire :} \\ f &\mapsto f + 2f' \end{aligned}$$

Tout d'abord, $C^1(\mathbb{R})$ et $C^0(\mathbb{R})$ sont deux \mathbb{R} -ev.
On remarque que

$$L = id + 2d$$

où $id: C^1(\mathbb{R}) \rightarrow C^1(\mathbb{R}) \subset C^0(\mathbb{R})$ est l'application identité
 $f \mapsto f$

et $d: C^1(\mathbb{R}) \rightarrow C^0(\mathbb{R})$ sont deux applications linéaires
 $f \mapsto f'$

Ainsi, L est bien une application linéaire, par combinaison linéaire d'applications linéaires.

Proposition 130 (Composition d'A.L.)

Soient E, F, G des \mathbb{K} -espaces vectoriels, alors

$$f \in \mathcal{L}(E, F), g \in \mathcal{L}(F, G) \implies g \circ f \in \mathcal{L}(E, G)$$

Démonstration :

E et G sont des \mathbb{K} e.v. Soient $u, v \in E$ et $\beta \in \mathbb{K}$. Alors

$$g \circ f(u + \beta v) \stackrel{\text{linéarité de } f}{=} g(f(u) + \beta f(v)) \stackrel{\text{linéarité de } g}{=} g(f(u)) + \beta g(f(v))$$

Ainsi, $g \circ f$ est bien linéaire par TCAL. \square

■ Exemple 8 :

Si

$$f: P \in \mathbb{R}_2[X] \mapsto 2P - P'' \in \mathbb{R}[X] \quad \text{et} \quad g: P \in \mathbb{R}[X] \mapsto P(1)$$

alors f, g sont deux applications linéaires (à faire) et donc $f \circ g$ également.
On peut également l'observer si on fait le calcul :

$$f: P \in \mathbb{R}_2[X] \mapsto 2P(1) - P''(1) \in \mathbb{R}$$

qui est bien linéaire.

⚠️ **PRODUIT NON LINÉAIRE**

Si $f: E \rightarrow F$ et $g: E \rightarrow F$ sont des applications linéaires, **leur produit** h telle que $h(u) = f(u)g(u)$ pour tout $u \in E$, **n'est pas (jamais) une application linéaire** si elle est non nulle.

Soit par exemple x tel que $h(x) \neq 0$. Alors

$$h(2x) = f(2x)g(2x) = 2 \times 2 f(x)g(x) = 4h(x) \neq 2h(x)$$

📖 Notation :

Dans le cas des applications linéaires, on note la plupart du temps gf à la place de $g \circ f$. **Attention**, ce n'est pas la multiplication de deux fonctions, mais bien la composition.

Ainsi, dans le cas d'un endomorphisme f , on notera f^n au lieu de $\underbrace{f \circ \dots \circ f}_{n \text{ fois}}$.

(Ces notations sont dues à la relation entre applications linéaires et matrices.)

I-3 Lien entre matrices et applications linéaires en dimension finie

I.3-a) Matrice d'une application linéaire en dimension finie

Dans ce paragraphe, on suppose que E et F sont de dimension finie.

📖 Définition

Soient E, F deux espaces vectoriels de bases respectives $\mathcal{B}_E = (e_1, \dots, e_n)$ et $\mathcal{B}_F = (f_1, \dots, f_m)$. Pour une application $f \in \mathcal{L}(E, F)$, on appelle **matrice de f dans les bases $\mathcal{B}_E, \mathcal{B}_F$** la matrice donnant les coordonnées des $f(e_i)$ en fonction des f_j .

$$M_{\mathcal{B}_F, \mathcal{B}_E}(f) = \begin{matrix} & f(e_1) & \dots & f(e_n) \\ \begin{matrix} f_1 \\ \vdots \\ f_m \end{matrix} & \left(\begin{array}{ccc} * & \dots & * \\ \vdots & & \vdots \\ * & \dots & * \end{array} \right) \end{matrix}$$

 Notation :

Si $E = F$ et que $\mathcal{B}_E = \mathcal{B}_F$, on parle simplement de *la matrice de f dans la base \mathcal{B}* , où $\mathcal{B} = \mathcal{B}_E$ et on note $M_{\mathcal{B}}(f)$ au lieu de $M_{\mathcal{B},\mathcal{B}}(f)$.

Proposition 131

Soient E, F deux espaces vectoriels de bases respectives \mathcal{B}_E et \mathcal{B}_F . Soient $f \in \mathcal{L}(E, F)$ et $u \in E$. Alors

$$M_{\mathcal{B}_F}(f(u)) = M(f)_{\mathcal{B}_F, \mathcal{B}_E} M_{\mathcal{B}_E}(u)$$

Démonstration :

On note $\mathcal{B}_E = (e_1, \dots, e_n)$ et $u = \alpha_1 e_1 + \dots + \alpha_n e_n$. Alors $f(u) = \alpha_1 f(e_1) + \dots + \alpha_n f(e_n)$.

Si on note C_i la $i^{\text{ème}}$ colonne de $A = M(f)_{\mathcal{B}_F, \mathcal{B}_E}$, on a par définition

$$C_i = M_{\mathcal{B}_F}(f(e_i)).$$

Par calcul matriciel, on obtient alors

$$A \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = \alpha_1 C_1 + \dots + \alpha_n C_n = M_{\mathcal{B}_F}(\alpha_1 f(e_1) + \dots + \alpha_n f(e_n)) = M_{\mathcal{B}_F}(f(u))$$

□

■ Exemple 9 :

Soit $L_0 : \mathbb{R}_2[X] \rightarrow \mathbb{R}_3[X]$

$P \mapsto \int_0^X P$: primitive de P qui s'annule en 0

• Matrice dans les bases canoniques :

La matrice de L_0 dans les bases $(1, X, X^2), (1, X, X^2, X^3)$ est

$$\begin{matrix} & L_0(1) & L_0(X) & L_0(X^2) \\ \begin{matrix} 1 \\ X \\ X^2 \\ X^3 \end{matrix} & \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix} \end{matrix}$$

car $L_0(1) = X$

$$L_0(X) = \frac{X^2}{2}$$

$$L_0(X^2) = \frac{X^3}{3}$$

• Image d'un vecteur : $L_0(1 + 2X - X^2) = X + X^2 - \frac{1}{3}X^3$ car

$$U = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \\ -\frac{1}{3} \end{pmatrix}$$

• Matrice de L_0 dans d'autres bases :

La matrice de L dans les bases $(1, X, X^2), (1, X - 1, X^2 - X, X^3)$ est

$$\begin{matrix} & L_0(1) & L_0(X) & L_0(X^2) \\ \begin{matrix} 1 \\ X-1 \\ X^2-X \\ X^3 \end{matrix} & \begin{pmatrix} 1 & \frac{1}{2} & 0 \\ 1 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix} \end{matrix}$$

car $L_0(1) = X = (X - 1) + 1$

$$L_0(X) = \frac{X^2}{2} = \frac{1}{2}((X^2 - X) + X) = \frac{1}{2}(X^2 - X) + \frac{1}{2}(X - 1) + \frac{1}{2}$$

$$L_0(X^2) = \frac{X^3}{3}$$

Proposition 132

Soient E, F, G des ev de dimension finie, $\mathcal{B}_E, \mathcal{B}_F, \mathcal{B}_G$ des bases respectives de E, F, G .

- Si $f, g \in \mathcal{L}(E, F), \alpha, \beta \in \mathbb{K}$. Alors

$$M_{\mathcal{B}_F, \mathcal{B}_E}(\alpha f + \beta g) = \alpha M_{\mathcal{B}_F, \mathcal{B}_E}(f) + \beta M_{\mathcal{B}_F, \mathcal{B}_E}(g)$$

- Si $f \in \mathcal{L}(E, F)$ et $g \in \mathcal{L}(F, G)$, alors

$$M_{\mathcal{B}_G, \mathcal{B}_E}(gf) = M_{\mathcal{B}_G, \mathcal{B}_F}(g)M_{\mathcal{B}_F, \mathcal{B}_E}(f)$$

- Soit $f \in \mathcal{L}(E, F)$. Alors f est inversible ssi $M_{\mathcal{B}_F, \mathcal{B}_E}(f)$ est inversible et dans ce cas

$$M_{\mathcal{B}_E, \mathcal{B}_F}(f^{-1}) = (M_{\mathcal{B}_F, \mathcal{B}_E}(f))^{-1}$$

Démonstration :

i) et ii) admis. iii) est conséquence de ii). □

■ Exemple 10 :

Soit l'application $T : \mathbb{R}_2[X] \rightarrow \mathbb{R}_3[X]$

$P \mapsto 2P +$ primitive de P qui s'annule en 0

On a alors $T = 2id + L_0$ où id est l'application identité et L_0 l'application de l'exemple précédent. Ainsi, dans les bases canoniques respectives $\mathcal{B}_2 = (1, X, X^2), \mathcal{B}_3 = (1, X, X^2, X^3)$ de $\mathbb{R}_2[X]$ et $\mathbb{R}_3[X]$

$$M_{\mathcal{B}_2, \mathcal{B}_3}(T) = 2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 2 & 0 \\ 0 & \frac{1}{2} & 2 \\ 0 & 0 & \frac{1}{3} \end{pmatrix}$$



Remarque :

Dans le but de savoir retrouver plus facilement et retenir les différentes formules de ce chapitre, on se propose de schématiser une application linéaire $L : E \rightarrow F$ de la manière suivante :

$$\begin{array}{ccc} F & \xleftarrow{L} & E \\ \text{base } \mathcal{B}_F & & \text{base } \mathcal{B}_E \end{array}$$

(Oui, c'est écrit de la droite vers la gauche, au lieu du contraire habituellement pour les fonctions !) De cette manière, **c'est écrit dans le même sens que les formules de composition et les formules matricielles :**

$$g \circ f : \begin{array}{ccc} G & \xleftarrow{g} & F \\ \text{base } \mathcal{B}_G & & \text{base } \mathcal{B}_F \end{array} \xleftarrow{f} \begin{array}{ccc} E & & \\ & & \text{base } \mathcal{B}_E \end{array} \quad \text{décrit} \quad M_{\mathcal{B}_G, \mathcal{B}_E}(f \circ g) = M_{\mathcal{B}_G, \mathcal{B}_F}(g) M_{\mathcal{B}_F, \mathcal{B}_E}(f)$$

et

$$\begin{array}{ccc} F & \xleftarrow{f} & E \\ \text{base } \mathcal{B}_F & & \text{base } \mathcal{B}_E \end{array} \quad \text{décrit} \quad M_{\mathcal{B}_F}(f(u)) = M_{\mathcal{B}_F, \mathcal{B}_E}(f) M_{\mathcal{B}_E}(u)$$

$$f(u) \quad \leftarrow \quad u$$

I.3-b) Changement de base dans un eV en tant qu'application linéaire

Rappelons que pour changer de base "à l'intérieur" d'un même espace vectoriel E , on dispose de la matrice de passage :

$$P_{\mathcal{B}, \mathcal{B}'} = M_{\mathcal{B}}(\mathcal{B}') = \begin{array}{ccc} & e'_1 & \dots & e'_n \\ e_1 & \left(\begin{array}{ccc} \alpha_{1,1} & \dots & \alpha_{1,n} \\ \vdots & & \\ \alpha_{n,1} & \dots & \alpha_{n,n} \end{array} \right) \\ & e_n & & \end{array}$$

avec $\mathcal{B} = (e_1, \dots, e_n)$, $\mathcal{B}' = (e'_1, \dots, e'_n)$ deux bases de E .

Explications :

En regardant de plus près, on observe qu'en fait **cette matrice est celle de l'application "identité"** dans la base de départ \mathcal{B}' et base d'arrivée \mathcal{B} :

$$\begin{array}{ccc} id : & E & \leftarrow & E \\ & \text{base } \mathcal{B}' & & \text{base } \mathcal{B} \end{array}$$

i.e.

$$M_{\mathcal{B},\mathcal{B}'}(id) = M_{\mathcal{B}}(\mathcal{B}') = P_{\mathcal{B},\mathcal{B}'}$$

On observe alors que les formules des images par une AL appliquée à la matrice identité donne :

$$M_{\mathcal{B}}(u) = M_{\mathcal{B},\mathcal{B}'}(id)M_{\mathcal{B}'}(u)$$

i.e.

$$M_{\mathcal{B}}(u) = M_{\mathcal{B}}(\mathcal{B}')M_{\mathcal{B}'}(u) = P_{\mathcal{B},\mathcal{B}'}M_{\mathcal{B}'}(u)$$

ce qui démontre immédiatement la formule de changement de base annoncée dans le cours sur les espaces vectoriels! Quant à l'autre formule annoncée dans ledit chapitre et non démontrée,

$$P_{\mathcal{B},\mathcal{C}}P_{\mathcal{C},\mathcal{D}} = P_{\mathcal{B},\mathcal{D}}$$

la démonstration est tout simplement

$$M_{\mathcal{B},\mathcal{C}}(id)M_{\mathcal{C},\mathcal{D}}(id) = M_{\mathcal{B},\mathcal{D}}(id \circ id) = M_{\mathcal{B},\mathcal{D}}(id)$$

(Formule dont on déduit également le fait que $P_{\mathcal{B},\mathcal{B}'}^{-1} = P_{\mathcal{B}',\mathcal{B}}$.)

I.3-c) Changement de base dans "E" et/ou "F" pour des AL

Si on peut changer de base à l'intérieur d'un espace vectoriel, on peut également changer de base lorsque l'on souhaite exprimer matriciellement une application linéaire. En effet, il n'est pas rare de devoir adapter les bases choisies pour l'expression de $M(f)$ suivant les conditions du problème (souvent pour avoir la matrice la plus "simple" possible) Pour ce faire, on peut utiliser (et mixer) les formules existantes pour tout faire en une seule fois :

On se donne ici une application linéaire $f : E \rightarrow F$ ainsi que deux bases $\mathcal{B}, \mathcal{B}'$ de E et deux bases $\mathcal{G}, \mathcal{G}'$ de F .

On suppose disposer de la matrice $M = M_{\mathcal{B},\mathcal{B}'}(f)$ et on cherche $M_{\mathcal{G},\mathcal{G}'}(f)$.

Méthode 1 : Changement de base "à vue"

On note $\mathcal{B}' = (e'_1, \dots, e'_n)$ et $\mathcal{G}' = (f'_1, \dots, f'_m)$.

On cherche directement toutes les images $f(e'_i)$ dans la base \mathcal{G}' à vue.

■ Exemple 11 :

Reprenons l'exemple de $L_0 : \mathbb{R}_2[X] \rightarrow \mathbb{R}_3[X]$

$P \mapsto$ primitive de P qui s'annule en 0

Dans les bases $\mathcal{B} = (1, X, X^2)$, $\mathcal{G} = (1, X, X^2, X^3)$ la matrice de L_0 est $\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix}$.

On pose $\mathcal{B}' = (\underbrace{1}_{e'_1}, \underbrace{X+1}_{e'_2}, \underbrace{X^2-X}_{e'_3})$, $\mathcal{G}' = (\underbrace{1}_{f'_1}, \underbrace{X-1}_{f'_2}, \underbrace{X^2-X}_{f'_3}, \underbrace{X^3}_{f'_4})$.

Alors,

$$\begin{cases} f(e'_1) = f(1) = X = X - 1 + 1 = f'_2 + f'_1 \\ f(e'_2) = f(X+1) = f(X) + f(1) = \frac{1}{2}X^2 + X = \frac{1}{2}(X^2 - X) + \frac{1}{2}X + X = \frac{1}{2}f'_3 - \frac{1}{2}(f'_2 + f'_1) + X \\ f(e'_3) = f(X^2 - X) = f(X^2) - f(X) = \frac{1}{3}X^3 - \frac{1}{2}X^2 = \frac{1}{3}f'_4 - \frac{1}{2}(X^2 - X) - \frac{1}{2}X \\ \qquad \qquad \qquad = \frac{1}{3}f'_4 - \frac{1}{2}f'_3 - \frac{1}{2}(f'_2 + f'_1) \end{cases}$$

D'où

$$M_{\mathcal{G}',\mathcal{B}'}(L_0) = \begin{pmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & \frac{1}{3} \end{pmatrix}$$

Méthode 2 : Changement de base avec formule matricielle

On utilise directement la formule matricielle :

Théorème 133 de changement de base

Soient E un espace vectoriel de bases \mathcal{B} et \mathcal{B}' et F un espace vectoriel de bases respectives $\mathcal{G}, \mathcal{G}'$. Soit $f \in \mathcal{L}(E, F)$, alors

$$M(f)_{\mathcal{G}',\mathcal{B}'} = M_{\mathcal{G}'}(\mathcal{G}) M_{\mathcal{G},\mathcal{B}}(f) M_{\mathcal{B}}(\mathcal{B}')$$

Démonstration :

Traduisons la situation à l'aide d'un schéma :

$$\begin{array}{ccccccc} f : & F & \xleftarrow{id} & F & \xleftarrow{f} & E & \xleftarrow{id} & E \\ & \text{base } \mathcal{G}' & & \text{base } \mathcal{G} & & \text{base } \mathcal{B} & & \text{base } \mathcal{B}' \end{array}$$

La formule de composition de matrices donne

$$M(f)_{\mathcal{G}',\mathcal{B}'} = M_{\mathcal{G}',\mathcal{G}}(id) M(f)_{\mathcal{G},\mathcal{B}} M_{\mathcal{B},\mathcal{B}'}(id)$$

ce qui correspond exactement à la formule de changement de bases annoncée dans le théorème. \square

■ Exemple 12 :

Reprenons l'exemple de $L_0 : \mathbb{R}_2[X] \rightarrow \mathbb{R}_3[X]$.
 $P \mapsto$ primitive de P qui s'annule en 0

Dans les bases $\mathcal{B} = (1, X, X^2)$, $\mathcal{G} = (1, X, X^2, X^3)$ la matrice de L_0 est $\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix}$.

On pose $\mathcal{G}' = (1, X - 1, X^2 - X, X^3)$. D'où $M_{\mathcal{G}'}(\mathcal{G}) = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$

Les opérations $C_2 \leftarrow C_2 + C_1$ puis $C_3 \leftarrow C_3 + C_2$ permettent de calculer

$$M_{\mathcal{G}'}(\mathcal{G}) = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Alors,

$$M_{\mathcal{G}',\mathcal{B}}(L_0) = M_{\mathcal{G}'}(\mathcal{G})M_{\mathcal{G},\mathcal{B}}(L_0) = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{2} & 0 \\ 1 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix}$$

 Définition

Soient M et M' deux matrices. S'il existe une matrice P inversible telle que

$$M' = P^{-1}MP \quad (\text{ou } M' = PMP^{-1})$$

alors on dit que M et M' sont *semblables*.

Corollaire

En particulier, pour le cas d'un endomorphisme $f \in \mathcal{L}(E)$, si \mathcal{B} et \mathcal{B}' sont deux bases de E , on a

$$M(f)_{\mathcal{B}'} = M_{\mathcal{B}'}(\mathcal{B}) M_{\mathcal{B}}(f) M_{\mathcal{B}}(\mathcal{B}')$$

ou autrement dit,

$$M' = P^{-1}MP$$

où P est la matrice de changement de base $P_{\mathcal{B},\mathcal{B}'}$, $M = M_{\mathcal{B}}(f)$ et $M' = M(f)_{\mathcal{B}'}$.

II Image, noyau, injectivité, surjectivité

II-1 Image (d'un sev ou de E)

On rappelle que pour tout sous-ensemble A de E et une fonction f quelconque, on note

$$f(A) = \{f(x) \mid x \in A\}$$

Voyons ce qui se passe si f est linéaire et que A est plus précisément un sev. En premier lieu, si c'est un sev de dimension finie :

Propriété 134

Soient $f \in \mathcal{L}(E, F)$ et H un sev de E avec $H = \text{Vect}(v_1, \dots, v_n)$ (la famille (v_1, \dots, v_n) n'étant pas forcément libre...), alors

$$f(H) = f(\text{Vect}(v_1, \dots, v_n)) = \text{Vect}(f(v_1), \dots, f(v_n))$$

Démonstration :

Supposons que $H = \text{Vect}(v_1, \dots, v_n)$.

• Montrons que $f(H) \subset \text{Vect}(f(v_1), \dots, f(v_n))$:

Soit $y \in f(H)$. Alors, il existe $x \in H$ tel que $f(x) = y$. Or, comme $H = \text{Vect}(v_1, \dots, v_n)$, il existe $\alpha_1, \dots, \alpha_n \in \mathbb{K}$ tels que

$$x = \alpha_1 v_1 + \dots + \alpha_n v_n$$

d'où

$$y = f(x) = f(\alpha_1 v_1 + \dots + \alpha_n v_n) = \alpha_1 f(v_1) + \dots + \alpha_n f(v_n) \in \text{Vect}(f(v_1), \dots, f(v_n))$$

• Montrons que $\text{Vect}(f(v_1), \dots, f(v_n)) \subset f(H)$:

"démonstration 1" : on a $f(v_1), \dots, f(v_n) \in f(H)$ et on verra dans la généralisation que $f(H)$ est un espace vectoriel. Par définition du Vect (plus petit espace vectoriel), on a donc bien $\text{Vect}(f(v_1), \dots, f(v_n)) \subset f(H)$.

"démonstration 2" : Soit $y \in \text{Vect}(f(v_1), \dots, f(v_n))$. Il existe alors $\alpha_1, \dots, \alpha_n \in \mathbb{K}$ tels que

$$y = \alpha_1 f(v_1) + \dots + \alpha_n f(v_n) = f(\underbrace{\alpha_1 v_1 + \dots + \alpha_n v_n}_{\text{on note ceci } x})$$

Or $x \in H$, donc $y \in f(H)$.

* Conclusion : Par double inclusion, on a bien $f(H) = \text{Vect}(f(v_1), \dots, f(v_n))$. \square

■ Exemple 13 :

Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2; (x, y) \mapsto (2x + y, -x)$ qui est une application linéaire. On pose $H = \text{Vect}(1, 1)$. Alors

$$f(H) = \text{Vect}(f(1, 1)) = \text{Vect}((3, -1))$$

Pour des sev quelconques :

Proposition 135

Soient $f \in \mathcal{L}(E, F)$ et H un sous espace vectoriel de E , alors $f(H)$ est un sev de F .

Démonstration :

Supposons que H soit un sev de E . Montrons que $f(H)$ est un sev de F :

- $f(H)$ est non vide, car $0_E \in H$ et $0_F = f(0_E) \in f(H)$.
- Soient $\lambda, \mu \in \mathbb{K}$ et $x, y \in f(H)$. Montrons que $\lambda x + \mu y \in f(H)$:
Il existe $x', y' \in H$ tels que $f(x') = x$ et $f(y') = y$. Ainsi,

$$\lambda x + \mu y = \lambda f(x') + \mu f(y') = f(\underbrace{\lambda x' + \mu y'}_{\in H}) \in f(H)$$

Par caractérisation des sous-espaces vectoriels, $f(H)$ est bien un espace vectoriel.

□

Appliqué maintenant à $H = E$:



Définition

Soit $f \in \mathcal{L}(E, F)$. On appelle *image* de f et on note $\text{Im} f$ l'espace vectoriel $f(E)$.



Remarque :

Si E est de dimension finie, avec comme base $\mathcal{B} = (e_1, \dots, e_n)$, on sait alors que

$$\text{Im} f = \text{Vect}(f(e_1), \dots, f(e_n))$$

■ Exemple 14 :

Reprenons l'exemple précédent avec $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2; (x, y) \mapsto (2x + y, -x)$. Sa matrice dans la base canonique est

$$A = \begin{pmatrix} 2 & 1 \\ -1 & 0 \end{pmatrix}$$

On sait alors que $\text{Im} f = \text{Vect}(\text{"colonnes de } A\text{"}) = \text{Vect}\left(\begin{pmatrix} 2 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}\right)$.

Corollaire

Soit $f \in \mathcal{L}(E, F)$. Si E est de dimension finie, alors $\text{Im} f$ l'est aussi et de plus

$$\dim \text{Im} f \leq \dim E$$

■ Exemple 15 :

On considère maintenant $L_0 : \mathbb{R}_2[X] \rightarrow \mathbb{R}[X]$.

$$P \mapsto \int_0^X P(t) dt$$

On admet ici qu'elle est linéaire.

1. $\dim(\mathbb{R}_2[X]) = 3 < \infty$; mais $\mathbb{R}[X]$ n'est pas de dimension finie.
2. Malgré tout, d'après la propriété,

$$\dim \text{Im} L_0 \leq \dim(\mathbb{R}_2[X]) = 3.$$

Ce qui se vérifie car

$$\text{Im} L_0 = X\mathbb{R}_2[X] \quad (\text{exercice})$$



Définition

Si $\text{Im} f$ est un espace vectoriel de dimension finie, on appelle *rang de f* et on note $\text{rg} f$ la dimension de $\text{Im} f$. (i.e. $\text{rg} f = \dim(\text{Im} f)$.)

II-2 Noyau

Contrairement à la partie précédente, on s'intéresse maintenant aux "images réciproques", plus précisément celles de 0_F . On rappelle avec les notations utilisées, que $f(0_E) = 0_F$. En revanche, rien ne garantit qu'il n'y a pas d'autres éléments d'image nulle.



Définition

Soit $f \in \mathcal{L}(E, F)$. On appelle *noyau* de f et on note $\ker f$ l'ensemble des vecteurs d'image 0_F :

$$\ker f = \{x \in E \mid f(x) = 0_F\}$$

Le commentaire fait ci-dessus nous dit en particulier que :

Propriété 136

Si $f \in \mathcal{L}(E, F)$, alors $0_E \in \ker f$.

■ Exemple 16 :

$$\begin{aligned} \text{On considère } L : C^1(\mathbb{R}) &\rightarrow C^0(\mathbb{R}) \\ f &\mapsto f - f' \end{aligned}$$

Alors $\ker L = \text{Vect}(\exp)$:

En effet,

$$f \in \ker L \iff f - f' = 0$$

qui est une équation différentielle dont on connaît l'ensemble des solutions $\{x \mapsto ke^x, k \in \mathbb{R}\} = \text{Vect}(\exp)$.

Proposition 137

Soient $f \in \mathcal{L}(E, F)$. $\ker f$ est un sev de E .

Démonstration :

- $\ker f$ est non vide, car $0_E \in \ker f$ d'après la propriété précédente.

- Soient $\lambda, \mu \in \mathbb{K}$ et $x, y \in \ker f$. Montrons que $\lambda x + \mu y \in \ker f$, c'est-à-dire $f(\lambda x + \mu y) = 0_F$:

$$f(\lambda x + \mu y) = \underbrace{\lambda f(x)}_{=0_F} + \underbrace{\mu f(y)}_{=0_F} = 0_F$$

Par caractérisation des sous-espaces vectoriels, $\ker f$ est bien un espace vectoriel.

□

II-3 Injection, surjection

Théorème 138 de caractérisation

Soit $f \in \mathcal{L}(E, F)$.

- f est surjective si et seulement si $F = \text{Im} f$.
- f est injective si et seulement si $\ker f = \{0_E\}$.

Démonstration :

- Montrons que f est surjective si et seulement si $F = \text{Im} f$:

Il est clair que, dans tous les cas, $\text{Im} f \subset F$

On a $F = \text{Im} f \iff$ tout élément de F admet un antécédent $\iff f$ est surjective.

- Montrons que f est injective si et seulement si $\ker f = \{0_E\}$:

• Comme $\ker f$ est un sous-espace vectoriel de E , on a $\{0_E\} \subset \ker f$.

• Montrons que f est injective si et seulement si $\ker f \subset \{0_E\}$:

★ Supposons que f est injective et montrons que $\ker f \subset \{0_E\}$:

Soit $x \in \ker f$. Alors

$$f(x) = 0_F = f(0_E)$$

par injectivité, on a

$$x = 0_E$$

★ Supposons que $\ker f \subset \{0_E\}$ et montrons que f est injective :

Soient $x, y \in F$ tels que

$$f(x) = f(y)$$

par linéarité, on a

$$f(x - y) = 0_F$$

i.e.

$$x - y \in \ker f \subset \{0_E\}$$

et donc $x = y$.

□

■ Exemple 17 :

Reprenons l'exemple $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2; (x, y) \mapsto (2x + y, -x)$. Étudions son injectivité.

Etant donné qu'on a une formule, on peut calculer son noyau par exemple de la manière suivante :

$$X = (x, y) \in \ker f \iff f(X) = 0 \iff \begin{cases} 2x + y = 0 \\ -x = 0 \end{cases} \iff \begin{cases} y = 0 \\ x = 0 \end{cases}$$

Ce qui donne comme résultat

$$\ker f = \{0\}$$

ainsi, f est bien injective.

■ Exemple 18 :

Sur l'exemple précédemment traité avec $L : C^1(\mathbb{R}) \rightarrow \mathcal{C}^0(\mathbb{R})$ on avait vu que

$$f \mapsto f - f'$$

$\ker L = Vect(exp)$. L'application n'est donc pas injective.

 Remarque :

Étant donné les inclusions $\text{Im } f \subset F$ et $\{0_E\} \subset \ker f$ triviales, en pratique, on utilisera souvent le théorème suivant :

Théorème 139 *de caractérisation n° 2*

Soit $f \in \mathcal{L}(E, F)$.

- f est surjective si et seulement si $F \subset \text{Im } f$.
- f est injective si et seulement si $\ker f \subset \{0_E\}$.

 Exercice 1

Montrer que $L_0 : \mathbb{R}[X] \rightarrow \mathbb{R}[X]$ est injective (mais non bijective).
 $P \mapsto$ primitive de P qui s'annule en 0

* Montrons que L est injective :

On sait que L est une application linéaire. Alors L est injective ssi $\ker L \subset \{0\}$.
 Soit $P \in \ker L$. Alors $L(P) = 0$, Or, $L(P)$ est une primitive de P par construction. D'où, en dérivant :

$$P = (L(P))' = 0' = 0$$

D'où $\ker L \subset \{0\}$ et donc L injective.

* Montrons que L n'est pas surjective :

Il s'agit de montrer que certains éléments de $\mathbb{R}[X]$ n'ont pas d'antécédents. Supposons que $Q \in \mathbb{R}[X]$ et qu'il existe un antécédent P , c'est-à-dire $L(P) = Q$. En dérivant, on trouve $P = Q'$.

Prenons dans ce cas Q une constante non nulle (prenons par exemple 1), on obtient

$$P = Q' = 0$$

Mais dans ce cas, on avait en réalité

$$1 = Q = L(P) = L(0) = 0$$

ce qui est contradictoire. Ainsi, Q ne pouvait pas avoir d'antécédent et donc L n'est pas surjective.

 Remarque :

Une application $f \in \mathcal{L}(E; F)$ injective est un isomorphisme de E sur $\text{Im } f \subset F$.

III Applications linéaires en dimension finie

Dans cette section, on se donne E, F deux espaces vectoriels, avec E de dimension finie et $f \in \mathcal{L}(E, F)$.

III-1 Rang et matrices

On rappelle que par définition, $\text{rg } f = \dim \text{Im } f$ (valide car E de dimension finie entraîne $\text{Im } f$ nécessairement de dimension finie).

Théorème 140

Si F est de dimension finie, alors, pour toutes bases \mathcal{B} de E et \mathcal{B}' de F , on a

$$\text{rg } f = \text{rg } (M(f)_{\mathcal{B}', \mathcal{B}})$$

En particulier, on obtient que $\text{rg } (M(f)_{\mathcal{B}', \mathcal{B}})$ ne dépend pas des bases \mathcal{B}' et \mathcal{B} .

Démonstration :

On pose $\mathcal{B} = (e_1, \dots, e_n)$. Alors, la matrice $M(f)_{\mathcal{B}', \mathcal{B}}$ n'est que l'expression des vecteurs $f(e_1), \dots, f(e_n)$ dans la base \mathcal{B}' . Ainsi,

$$\text{rg } f = \dim \text{Im } f \stackrel{=}{=} \dim \text{Vect } (f(e_1), \dots, f(e_n)) = \text{rg } M(f)_{\mathcal{B}', \mathcal{B}}$$

□ $\text{Im } f = \text{Vect } (f(e_1), \dots, f(e_n))$

III-2 Image d'une base et injectivité

Théorème 141

Soit E un espace vectoriel de dimension finie, F un espace vectoriel et $f \in \mathcal{L}(E, F)$. les conditions suivantes sont équivalentes :

- f est injective
- l'image de toute base de E est une famille libre de F
- l'image d'une base de E est une famille libre de F
- $\dim E = \text{rg } f$

Démonstration :

- Montrons que $i \Rightarrow ii$:

Supposons que $f \in \mathcal{L}(E, F)$ soit injective et soit $\mathcal{B} = \{e_1, \dots, e_n\}$ une base de E .

- Montrons que $f(\mathcal{B}) = \{f(e_1), \dots, f(e_n)\}$ est une famille libre :

Soient $\lambda_1, \dots, \lambda_n \in \mathbb{K}$ tels que $\lambda_1 f(e_1) + \dots + \lambda_n f(e_n) = 0$. Par linéarité de f , on a $f(\lambda_1 e_1 + \dots + \lambda_n e_n) = 0$.

Par injectivité de f , on trouve

$$\lambda_1 e_1 + \dots + \lambda_n e_n = 0$$

puis par liberté de \mathcal{B} , on obtient $\lambda_1 = \dots = \lambda_n = 0$.

- Conclusion : $f(\mathcal{B})$ est une famille libre dans F .

- Montrons que $ii \Rightarrow iii$: trivial!

- Montrons que $iii \Rightarrow iv$:

Soit \mathcal{B} une base de E telle que $f(\mathcal{B})$ est une famille libre base de F . Donc

$$\text{rg } f = \dim \text{Im } f = \dim \text{Vect}(f(\mathcal{B})) \underbrace{=}_{f(\mathcal{B}) \text{ libre}} \text{card } f(\mathcal{B}) = \text{card } \mathcal{B} = \dim E$$

- Montrons que $iv \Rightarrow i$:

Supposons iv) et montrons i), c'est-à-dire que f est injective. D'après les propriétés des applications linéaires, cela revient à montrer que $\ker f \subset \{0_E\}$.

Soit $x \in \ker f$. On pose $\mathcal{B} = (e_1, \dots, e_n)$ une base de E . Il existe donc $\alpha_1, \dots, \alpha_n \in \mathbb{K}$ tels que

$$x = \alpha_1 e_1 + \dots + \alpha_n e_n$$

En appliquant f , on obtient, par linéarité,

$$0 = f(x) = \alpha_1 f(e_1) + \dots + \alpha_n f(e_n)$$

Par iv), la famille $\{f(e_1), \dots, f(e_n)\}$ est libre. D'où $\alpha_1 = \dots = \alpha_n = 0$ et donc $x = 0$.

L'application f est donc injective. \square

■ Exemple 19 :

Soit $L_0 : \mathbb{R}_2[X] \rightarrow \mathbb{R}_3[X]$

$P \mapsto$ primitive de P qui s'annule en 0

On admet ici qu'elle est bien définie et linéaire. Voyons qu'elle est injective encore d'une autre manière par-rapport au cas $L_0 : \mathbb{R}[X] \rightarrow \mathbb{R}[X]$ traité précédemment.

On pose la base canonique de $\mathbb{R}_2[X] : \mathcal{B} = (1, X, X^2)$. On observe alors que son image par L est

$$L_0(\mathcal{B}) = (f(1), f(X), f(X^2)) = \left(X, \frac{1}{2}X^2, \frac{1}{3}X^3\right)$$

Ceci est une famille libre de $\mathbb{R}_3[X]$ (car c'est une famille de polynômes non nuls de degrés échelonnés). L_0 est donc injective. (" $iii \Rightarrow i$ ") du théorème)



Remarque :

Généralement, la technique de l'exemple précédent n'est pas la méthode la plus pertinente pour démontrer l'injectivité. La technique du noyau s'avère souvent plus efficace.

Corollaire

Soit E un espace vectoriel de dimension finie, F un espace vectoriel et $f \in \mathcal{L}(E, F)$. Alors les conditions suivantes sont équivalentes :

- f est un isomorphisme ;
- l'image de toute base de E est une base de F .
- l'image d'une base de E est une base de F .

Démonstration :

• On rappelle que, d'après une proposition précédente, $f(\mathcal{B}) = \{f(e_1), \dots, f(e_n)\}$ est une famille génératrice de $\text{Im } f$.

- Montrons que $i \Rightarrow ii$: Supposons que f est un isomorphisme. Alors :
 - comme f est injective, d'après le théorème précédent, l'image d'une base de E est une base de $\text{Im } f$.
 - comme f est surjective, $\text{Im } f = F$. L'image d'une base de E est une base de $\text{Im } f = F$.
- Montrons que $ii \Rightarrow iii$: trivial!
- Montrons que $iii \Rightarrow i$: Soit \mathcal{B} une base telle que $f(\mathcal{B})$ est une base de F .
- Montrons que f est surjective : . Alors,

$$F = \text{Vect } f(\mathcal{B}) = \text{Im } f$$

f est donc surjective.

• Montrons que f est injective : Comme $F = \text{Im } f$, alors l'image d'une base de E est une base de $\text{Im } f$. D'après le théorème précédent, f est donc injective.

□

■ Exemple 20 :

Revoyons que $L_0 : \mathbb{R}_2[X] \rightarrow \mathbb{R}_3[X]$ n'est pas un isomprphisme :

$$P \mapsto \int_0^X P$$

En effet, l'image de la base $(1, X, X^2)$ est $(f(1), f(X), f(X^2)) = (X, \frac{1}{2}X^2, \frac{1}{3}X^3)$, qui n'est pas une base de $\mathbb{R}_3[X]$ car c'est une famille de cardinal 3 dans $\mathbb{R}_3[X]$ de dimension 4.

Corollaire

Soient E, F deux espaces vectoriels de dimension finie, de bases respectives \mathcal{B}, \mathcal{G} et $f \in \mathcal{L}(E, F)$.

Si f est un isomorphisme, alors toute matrice $M_{\mathcal{B}, \mathcal{G}}(f)$ est carrée.

Démonstration :

On note $n = \dim E$ et $\mathcal{F} = (f(e_1), \dots, f(e_n))$ où $\mathcal{B} = (e_1, \dots, e_n)$.
D'après le corollaire précédent, on sait que \mathcal{F} est une base de F , donc $\dim F = n$.
Comme

$$\dim E = \dim F$$

la matrice est donc carrée. \square

Corollaire

Soient E, F deux espaces vectoriels de dimension finie, de bases respectives \mathcal{B}, \mathcal{G} et soit $f \in \mathcal{L}(E, F)$. Alors

f est un isomorphisme ssi $M_{\mathcal{B}, \mathcal{G}}(f)$ est inversible.

Démonstration :

On note $\mathcal{F} = \{f(e_1), \dots, f(e_n)\}$ où $\mathcal{B} = e_1, \dots, e_n$.

- Supposons que f est un isomorphisme.

Alors, d'après le corollaire précédent, la matrice $M = M_{\mathcal{B}, \mathcal{G}}(f)$ est carrée. Ses vecteurs colonnes qui forment la famille \mathcal{F} sont libres car l'image \mathcal{F} de la base \mathcal{B} est une base de F (corollaire d'avant).

- Supposons que la matrice est inversible.

Alors les vecteurs colonnes \mathcal{F} sont libres et de cardinal $n = \dim F$. Autrement dit, ils forment une base de F . D'après le cours précédent (corollaire), comme l'image \mathcal{F} d'une base \mathcal{B} de E par f est une base de F , alors c'est un isomorphisme.

\square

III-3 Théorème du rang et conséquences

Théorème 142 du rang

Supposons que E soit un espace vectoriel de dimension finie, que F soit un espace vectoriel et $f \in \mathcal{L}(E, F)$. Alors

$$\dim E = \dim \ker f + \operatorname{rg} f$$

Démonstration :

$\ker f$ est un sous espace de E de dimension finie. On note $\mathcal{B}_k = (e_1, \dots, e_r)$ une base de $\ker f$. D'après le théorème de la base incomplète, on sait que l'on peut compléter \mathcal{B}_k en une base $\mathcal{B} = (e_1, \dots, e_n)$ de E .

Notons $G = \operatorname{Vect}(e_{r+1}, \dots, e_n)$ et f_G l'application linéaire

$$f_G : G \rightarrow F$$

$$x \mapsto f(x)$$

- Montrons que f_G est injective :

Soit $x \in \ker f_G$. Alors $x \in G$ et $0_F = f_G(x) = f(x)$, d'où $x \in \ker f \cap G = \{0_E\}$. Ainsi, $x = 0_E$ et donc $\ker f_G \subset \{0_E\}$. Comme f_G est injective, on obtient en particulier $\dim G = \dim \operatorname{Im} f_G$, c'est-à-dire

autrement dit,

$$n - r = \dim \operatorname{Im} f_G,$$

$$\dim E - \dim \ker f = \dim \operatorname{Im} f_G$$

- Montrons que $\operatorname{Im} f_G = \operatorname{Im} f$:

Par définition de f_G , il est clair que $\operatorname{Im} f_G \subset \operatorname{Im} f$.

Montrons que $\operatorname{Im} f \subset \operatorname{Im} f_G$: Soit $y \in \operatorname{Im} f$. Il existe alors $x \in E$ tel que $f(x) = y$. Or, comme \mathcal{N} est une base de E , il existe $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ tels que

$$x = \underbrace{\alpha_1 e_1 + \dots + \alpha_r e_r}_{\in \ker f} + \underbrace{\alpha_{r+1} e_{r+1} + \dots + \alpha_n e_n}_{\in G}$$

D'où

$$y = f(x) = \underbrace{f(a)}_{=0_F} + \underbrace{f(\alpha_{r+1} e_{r+1} + \dots + \alpha_n e_n)}_{g \in G}$$

autrement dit, il existe $g \in G$ tel que $y = f(g)$. Ainsi, $\operatorname{Im} f \subset \operatorname{Im} f_G$.

Cette égalité implique clairement que $\dim \operatorname{Im} f = \dim \operatorname{Im} f_G$.

- Conclusion : Les deux égalités de dimension obtenues en encadré donnent clairement le résultat annoncé. \square

Théorème 143

Soient E, F deux espaces vectoriels de même dimension et $f \in \mathcal{L}(E, F)$. On a

$$f \text{ injective} \iff f \text{ bijective} \iff f \text{ surjective}$$

Démonstration :

$$f \text{ injective} \iff \ker f = \{0_E\}$$

$$\iff \dim(\ker f = \{0_E\}) = 0$$

$$\iff \dim E = \operatorname{rg} f \quad (\text{Théorème du rang})$$

$$\iff \operatorname{Im} f = E \quad (\operatorname{Im} f \subset F \text{ et égalité des dimensions})$$

$$\iff f \text{ surjective}$$

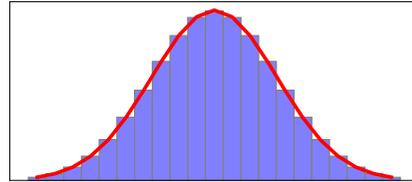
Ainsi, si f est injective, alors elle est surjective et donc bijective.

Si f est bijective, alors elle est injective et donc bijective. \square

I Densité

Commentaires :

Nous avons vu, dans le cadre des variables aléatoires finies (par exemple pour la loi binomiale), qu'en établissant, dans certains cas, l'histogramme des probabilités d'une variable $X \hookrightarrow B(n, p)$ (ci-contre $n = 100, p = 0,5$), on obtient un graphique pouvant être approximativement délimitée par une courbe (cf ci-contre).



La courbe donnée en exemple ci-dessus (qui donne les "bords" de l'histogramme) admet donc une aire égale environ égale ici à la "somme totale de l'aire des rectangles", c'est-à-dire, la somme de toutes les probabilités : 1. De même, la probabilité d'obtenir par exemple 4 succès pourrait être environ obtenue par une intégrale par un morceau d'aire sous la courbe (entre 2 bornes).

Concernant les variables discrètes, une telle courbe n'est qu'approximative, mais il existe d'autres variables, dont le comportement est plus exactement décrit par ce type d'objet (que nous appellerons des densités). C'est ce que nous verrons dans ce chapitre.

I-1 Définitions et généralités

Définition

On appelle $f : \mathbb{R} \rightarrow \mathbb{R}$ une *densité* si f est une fonction

- positive,
- continue (sauf peut être en un nombre fini de points),
- d'intégrale convergente sur \mathbb{R} avec $\int_{-\infty}^{+\infty} f(t) dt = 1$.

Remarque :

Notons qu'en terme de calcul, si f est continue sauf en un nombre fini de points et si est nulle en dehors d'un intervalle $]a, b[$, l'étude de l'intégrale $\int_{-\infty}^{+\infty} f$ est réduite à l'étude de $\int_a^b f$.

En effet, l'étude de $\int_{-\infty}^a f = \int_{-\infty}^a 0$ ainsi que $\int_b^{+\infty} f = \int_b^{+\infty} 0$ est triviale. Ces deux intégrales sont convergentes et valent toutes deux 0. Ainsi, $\int_{-\infty}^{+\infty} f$ converge ssi $\int_a^b f$ converge et les deux integrales ont même valeur.

Exemple 1 :

La fonction $f : x \in \mathbb{R} \mapsto \begin{cases} e^{-x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$ est une densité :

En effet, elle est continue sur \mathbb{R}^* , positive sur \mathbb{R} et de plus, on sait que

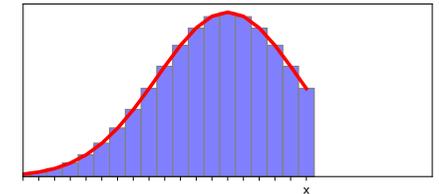
$$\int_{-\infty}^{+\infty} f = \int_0^{+\infty} e^{-t} dt = 1$$

Commentaires :

Nous allons commencer à voir comment ces fonctions dites "densités" vont en effet pouvoir décrire en des lois de probabilités et comment nous pouvons nous en servir précisément.

Or, on rappelle que toute loi de variable aléatoire peut être décrite par une fonction de répartition et réciproquement. Nous allons donc établir un lien entre "densité" et "fonction de répartition".

Sur le schéma ci-dessous, on rappelle que la valeur de la fonction de répartition $F(x)$ de la loi donnée correspond à l'aire des "barres" :



Théorème 144

Soit f une densité.

La fonction $F : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction de répartition car :

$$x \mapsto \int_{-\infty}^x f(t) dt$$

- $0 \leq F(x) \leq 1 \quad \forall x \in \mathbb{R}$.
- F est croissante.
- F est continue (donc continue à droite) en tout point de \mathbb{R}
- $\lim_{x \rightarrow -\infty} F(x) = 0 ; \lim_{x \rightarrow +\infty} F(x) = 1$

Démonstration :

Pour commencer, comme f est intégrable sur \mathbb{R} , cela signifie que $\int_{-\infty}^x f(t) dt$ est bien définie pour tout $x \in \mathbb{R}$.

ii) Montrons que F est croissante :

f est positive, donc, si $u \leq v$, par convergence des différentes intégrales, la relation de Chasles nous dit que

$$F(v) = \int_{-\infty}^v f(t) dt = \underbrace{\int_{-\infty}^u f(t) dt}_{=F(u)} + \underbrace{\int_u^v f(t) dt}_{\geq 0 \text{ (bornes croissantes)}}$$

D'où

$$F(u) \leq F(v)$$

ce qui donne la croissance de F .

iv) Montrons que $\lim_{x \rightarrow -\infty} F(x) = 0$; $\lim_{x \rightarrow +\infty} F(x) = 1$:

D'une part, par hypothèse, $\lim_{x \rightarrow +\infty} F(x) = \int_{-\infty}^{+\infty} f(t) dt = 1$.

D'autre part, on a

$$F(x) = \int_{-\infty}^x f(t) dt = \int_{-\infty}^0 f(t) dt + \underbrace{\int_0^x f(t) dt}_{\xrightarrow{x \rightarrow -\infty} \int_0^{-\infty} f(t) dt = - \int_{-\infty}^0 f(t) dt}$$

D'où $\lim_{x \rightarrow -\infty} F(x) = 0$.

i) Montrons que $0 \leq F(x) \leq 1 \quad \forall x \in \mathbb{R}$: Conséquence immédiate de ii) et iv).

iii) La continuité : Pour tous $a, x \in \mathbb{R}$, par convergence et relation de Chasles, on a

$$|F(x) - F(a)| = \left| \int_a^x f(t) dt \right|$$

Soit $\Delta = \{a_1, \dots, a_n\}$ sont les points de discontinuité éventuels de f , classés par ordre croissant : $a_1 < \dots < a_n$. Pour les besoins d'écriture, on écrit $a_{n+1} = +\infty$

· Si $a \notin \Delta$, la limite quand $x \rightarrow a$ ne pose pas de problème, puisque, pour x proche de a , c'est une intégrale classique. La limite vaut 0.

· Si $a = a_i \in \Delta$, pour x proche de a , c'est une intégrale impropre dont le seul problème est en a . On a

$$|F(x) - F(a)| = \left| \int_{a_i}^x f(t) dt \right| = \left| \int_{a_i}^{a_{i+1}} f(t) dt - \int_x^{a_{i+1}} f(t) dt \right|$$

C'est maintenant par hypothèse de convergence qu'elle tend vers 0 quand $x \rightarrow a = a_i$. □

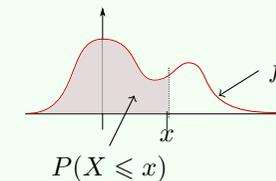
Définition

On dit que la variable aléatoire X "est de densité f " ou "admet une densité f " si :

■ f est une densité

■ $\forall x \in \mathbb{R}, \quad P(X \leq x) = \int_{-\infty}^x f$

i.e. $F_X : x \in \mathbb{R} \mapsto \int_{-\infty}^x f(t) dt$ est une fonction de répartition de X .



Remarque :

Il n'y a pas unicité de la densité possible pour une variable aléatoire. En effet, celle-ci est définie à un nombre fini de points près. En effet, si on modifie la valeur d'un nombre fini de points de f , ceci ne change pas la valeur de l'intégrale. Néanmoins, on fait généralement toujours en sorte que la densité soit "la plus continue possible" afin de ne pas compliquer les situations.

■ **Exemple 2 :**

Loi uniforme sur $]0; 1]$:

Supposons que X désigne un nombre pris au hasard dans l'intervalle $]0; 1]$.

Rappelons sa fonction de répartition $F_X : F_X(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ x & \text{si } 0 < x \leq 1 \\ 1 & \text{si } x > 1 \end{cases}$

Posons f telle que $f(t) = \begin{cases} 0 & \text{si } t < 0 \\ 1 & \text{si } 0 \leq t \leq 1 \\ 0 & \text{si } t > 1 \end{cases}$

★ Densité : f est une densité (laissé en exercice)

★ Fonction de répartition : Vérifions que, pour tout $x \in \mathbb{R}, \int_{-\infty}^x f = F_X(x)$:

Si $x < 0$: $\int_{-\infty}^x 0 = 0 = F_X(x)$

Si $x \in [0, 1]$: $\int_{-\infty}^x f = \int_0^x 1 = x = F_X(x)$

Si $x > 1$: $\int_{-\infty}^x f = \int_0^1 1 = 1 = F_X(x)$

Ainsi, $F_X(x) = \int_{-\infty}^x f$ pour tout $x \in \mathbb{R}$ et ainsi, X est de densité f .

Théorème 145

Pour toute densité f , il existe une variable aléatoire X telle que X soit une variable de densité f .

Démonstration : admise. \square

Sans préciser la densité :



Définition

On dit qu'une variable aléatoire X "est à densité" ou "admet une densité" s'il existe une densité f telle que X est de densité f .

■ Exemple 3 :

La variable aléatoire qui donne un nombre aléatoire dans l'intervalle $]0, 1]$ est à densité.

I-2 Si je sais qu'une variable admet une densité

Propriété 146

Si X est une variable aléatoire à densité, alors F_X est continue.

Démonstration :

C'est la propriété iii) du théorème de la page 145! \square



CONFUSIONS

Il s'agit là bien de la fonction de répartition et non de la densité, qui elle, peut avoir des points de discontinuité.

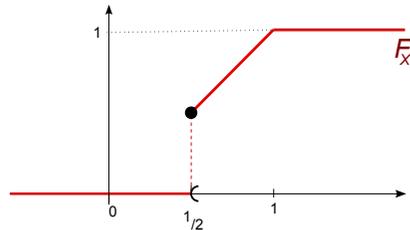
Corollaire

Une variable aléatoire finie ne peut jamais être à densité.

Démonstration :

La fonction de répartition d'une variable aléatoire finie est en escalier... \square

■ Exemple 4 Variable aléatoire qui n'est ni finie ni à densité :



On considère les deux expériences suivantes :

expérience(1) :
donne avec certitude le nombre $\frac{1}{2}$.

expérience(2) :
donne un nombre au hasard entre $\frac{1}{2}$ et 1.

On choisit au hasard d'appliquer expérience(1) ou expérience(2) et on note X le résultat obtenu.

La fonction de répartition est ci-dessus (on pourra éventuellement la prouver un peu plus tard, quand on aura parlé des variables uniformes sur des intervalles quelconques).

La variable n'est pas à densité (fonction de répartition non continue ...)

Corollaire

Si X est une variable aléatoire à densité, $P(X = x) = 0$ pour tout $x \in \mathbb{R}$.

Démonstration :

Nous avons vu précédemment que

$$P(X = x) = P(X \leq x) - P(X < x) = F_X(x) - P(X < x)$$

et que de ce fait, $P(X = x)$ correspondait à la "hauteur de discontinuité" de F_X en x . S'il n'y a pas de discontinuité, ceci vaut donc 0. Soit $x \in \mathbb{R}$. Par propriété des probabilités, on a toujours

$$P(X = x) = P(X \leq x) - P(X < x) = F_X(x) - P(X < x)$$

Or, nous avons observé que

$$P(X < x) = \lim_{t \rightarrow x, t < x} P(X \leq t) = \lim_{t \rightarrow x, t < x} F_X(t)$$

Ainsi, par continuité de F_X , on obtient

$$P(X < x) = F_X(x) = P(X \leq x).$$

D'où

$$P(X = x) = 0 \quad \square$$

Corollaire

Si X est une va de densité f , de probabilité p et de fonction de répartition F_X , alors, pour tous $a, b \in \overline{\mathbb{R}}$, avec la notation $F(+\infty) = 1$ et $F(-\infty) = 0$:

- $P(]-\infty; x]) = \int_{-\infty}^x f(t) dt$ pour tout $x \in \mathbb{R}$.
- $P(a \leq X \leq b) = \int_a^b f(t) dt$ pour tout $a, b \in \mathbb{R}$.
- $P(X = a) = 0$ pour tout $a \in \mathbb{R}$.
- $\int_a^b f(t) dt = F_X(b) - F_X(a) = P(a < X \leq b) = P(a \leq X \leq b) = P(a \leq X < b) = P(a < X < b)$

Démonstration : exercice. \square

Proposition 147

On se donne X une variable aléatoire de densité f , alors, en tout point x où f est continue,

$$\begin{cases} F_X \text{ est } \mathcal{C}^1 \text{ en } x \\ F'_X(x) = f(x) \quad (\text{i.e. } F_X \text{ est une primitive de } f.) \end{cases}$$

Démonstration :

Posons $F : x \mapsto \int_{-\infty}^x f(t) dt$ en les points cités. Soit donc $x \in \mathbb{R}$ tel que f soit continue en x .

On note $n \in \mathbb{N}^*$ et $\Delta = \{a_1, \dots, a_n\}$ l'ensemble fini des points de discontinuité de f , numérotés de façon à avoir $a_i < a_{i+1}$. On note $a_0 = -\infty$ et $a_{n+1} = +\infty$. Alors il existe $i \in \llbracket 0, n \rrbracket$ tel que

$$a_i < x < a_{i+1}$$

On pose

$$\alpha \in]a_i, a_{i+1}[\quad \text{une valeur constante quelconque.}$$

On a alors

$$F(x) = \int_{-\infty}^{\alpha} f + \int_{\alpha}^x f = F(\alpha) + \int_{\alpha}^x f$$

Comme f est continue sur $]a_i, a_{i+1}[$ avec $\alpha, x \in]a_i, a_{i+1}[$, on sait que $\varphi : x \in]a_i, a_{i+1}[\mapsto \int_{\alpha}^x f$ est une primitive de f sur $]a_i, a_{i+1}[$ (et donc en particulier \mathcal{C}^1 .)

Comme, $F(\alpha)$ est une constante par-rapport à x , par somme, F est \mathcal{C}^1 en x avec

$$F'(x) = f(x)$$

□

■ Exemple 5 :

Soit X une variable aléatoire de densité $f : t \mapsto \frac{1}{2}e^{-|t|}$. Déterminons la fonction de répartition de X .

On note que f est continue sur \mathbb{R} . On note F_X la fonction de répartition de X sur \mathbb{R} . Comme F est une primitive de f sur \mathbb{R} et que l'on a

$$f : t \mapsto \begin{cases} \frac{1}{2}e^{-t} & \text{si } t \geq 0 \\ \frac{1}{2}e^t & \text{si } t < 0 \end{cases}$$

Par primitive (sur les intervalles ouverts...!), on sait qu'il existe a, b tels que

$$F_X : t \mapsto \begin{cases} -\frac{1}{2}e^{-t} + a & \text{si } t > 0 \\ \frac{1}{2}e^t + b & \text{si } t < 0 \end{cases}$$

Or, la fonction de répartition doit vérifier $\lim_{-\infty} F_X = 0$, ce dont on déduit, par passage à la limite dans la formule de F_X :

$$b = 0$$

De même,

$$\lim_{+\infty} F_X = 1 \quad \Rightarrow \quad a = 1$$

On a donc

$$F_X : t \mapsto \begin{cases} 1 - \frac{1}{2}e^{-t} & \text{si } t > 0 \\ \frac{1}{2}e^t & \text{si } t < 0 \end{cases}$$

La fonction F_X étant nécessairement continue à droite (et même continue sur \mathbb{R} car X est à densité), on peut tout simplement poser

$$F_X : t \mapsto \begin{cases} 1 - \frac{1}{2}e^{-t} & \text{si } t \geq 0 \\ \frac{1}{2}e^t & \text{si } t < 0 \end{cases}$$

? Exercice 1

Soit X une variable aléatoire de densité $f : t \mapsto \frac{1}{\pi(1+t^2)}$. Déterminer la fonction de répartition de X .

Solution

On note que f est continue sur \mathbb{R} . En notant F_X la fonction de répartition de X , on sait que

$$F'_X(x) = f(x) = \frac{1}{\pi(1+x^2)} \quad \forall t \in \mathbb{R}$$

On note F_X la fonction de répartition de X sur \mathbb{R} . Comme F est une primitive de f sur \mathbb{R} , par primitive, On sait alors qu'il existe a tel que

$$F_X(x) = \frac{1}{\pi} \arctan x + a$$

Or, la fonction de répartition doit vérifier $\lim_{-\infty} F_X = 0$. ce dont on déduit, par passage à la limite dans la formule de F_X :

$$a = \frac{1}{2}$$

Corollaire

Soit X une variable aléatoire de densité f et $a, b \in \bar{\mathbb{R}}$ ($a < b$). Alors,

$$Supp(X) \subset [a, b] \iff f = 0 \text{ en tout point de continuité dans } \mathbb{R} \setminus [a, b]$$

⚠ Remarque :

Dans les ouvrages de probabilités, l'écriture " $Supp(X) \subset [a, b]$ " est généralement remplacée par

$$a \leq X \leq b \quad p.s.$$

où "p.s." signifie "presque sûrement", ce qui signifie que la probabilité de l'événement " $a \leq X \leq b$ " est 1 ; autrement dit, qu'on est quasiment certain que X sera borné par a et b . En pratique, par abus de notation, on se passe souvent de la notation "p.s." pour ne dire que $a \leq X \leq b$.

Démonstration :

- Supposons que la variable X soit bornée par a, b , i.e. $Supp(X) \subset [a, b]$.

Alors, en notant F_X la fonction de répartition de X , on a

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & \text{si } x < a \\ 1 & \text{si } x \geq b \end{cases}$$

Or, X est une variable à densité et F_X est \mathcal{C}^1 sur $] -\infty; a[\cup] b; +\infty[$. On peut donc en déduire qu'en tout point de continuité de f en dehors de $[a, b]$, on a

$$f(x) = F'_X(x) = \begin{cases} 0 & \text{si } x < a \\ 0 & \text{si } x > b \end{cases}$$

- Supposons que $f = 0$ sur $] -\infty; a[\cup] b; +\infty[$.

On note F_X la fonction de répartition de X . f étant continue sur $] -\infty; a[\cup] b; +\infty[$, on sait que F_X est \mathcal{C}^1 sur $] -\infty; a[\cup] b; +\infty[$. Ainsi, par primitive de la fonction nulle, il existe deux constantes $\alpha, \beta \in \mathbb{R}$ telles que

$$F_X(x) = P(X \leq x) = \begin{cases} \alpha & \text{si } x < a \\ \beta & \text{si } x > b \end{cases}$$

Les limites de F_X en $-\infty$ et $+\infty$ donnent alors $\alpha = 0$ et $\beta = 1$. Par continuité de la fonction de répartition, on en déduit que

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & \text{si } x \leq a \\ 1 & \text{si } x \geq b \end{cases}$$

Ainsi,

$$P(a \leq X \leq b) = F_X(b) - F_X(a) = 1 - 0 = 1$$

La démonstration pour f non continue en dehors de $[a, b]$ est admise. \square

■ Exemple 6 :

Reprenons la densité

$$f(t) = \begin{cases} 0 & \text{si } t < 0 \\ 1 & \text{si } 0 \leq t \leq 1 \\ 0 & \text{si } t > 1 \end{cases}$$

et déterminons cette fois-ci la fonction de répartition F de X de densité f grâce aux techniques que nous venons de voir :

Tout d'abord, nous constatons sur la fonction densité f , qui est nulle sur $\mathbb{R} - [0, 1]$, que

$$Supp(X) \subset [0, 1]$$

et donc

$$F(t) = \begin{cases} 0 & \text{si } t < 0 \\ 1 & \text{si } t \geq 1 \end{cases}$$

Si $t \in]0, 1[$, f étant continue sur cet intervalle, on sait que $F'(t) = f(t)$. Ainsi, il existe $\alpha \in \mathbb{R}$ tel que

$$F(t) = t + \alpha$$

La continuité de F en 0 nous donne alors

$$0 = \lim_{0^-} F = F(0) = \lim_{0^+} F = \alpha$$

En conclusion, on a encore une fois

$$F(t) = \begin{cases} 0 & \text{si } t < 0 \\ t & \text{si } 0 \leq t \leq 1 \\ 1 & \text{si } t \geq 1 \end{cases}$$

I-3 Si je veux montrer qu'une variable admet une densité

On se demande ici comment déterminer si une variable est à densité, et de surcroît, déterminer cette densité.

Proposition 148

On se donne X une variable aléatoire de fonction de répartition F . Si F est

- continue sur tout \mathbb{R}
- et de classe \mathcal{C}^1 sauf en un nombre fini de points a_1, \dots, a_n , alors X est une va à densité de densité f telle que

$$f(x) = F'(x) \quad \forall x \neq a_1, \dots, a_n$$

(les valeurs de $f(a_1), \dots, f(a_n) \geq 0$ pouvant être choisies arbitrairement.)

BIEN PENSER À TOUTES LES CONDITIONS

➤ On ne peut pas retirer la condition "continue sur tout \mathbb{R} ".

Démonstration :

f est une densité de X ssi

- f est continue sauf en un nombre fini de points et positive.
- $\int_{-\infty}^{+\infty} f$ est convergente de valeur 1
- $F(x) = \int_{-\infty}^x f$ pour tout x .

• Cas de i :

f est définie par $f(x) = F'(x) \quad \forall x \neq a_1, \dots, a_n$ avec F et de classe \mathcal{C}^1 sauf en un nombre fini de points a_1, \dots, a_n . Ainsi, par hypothèse, f est continue sauf éventuellement en a_1, \dots, a_n par définition.

De plus, F étant croissante, $f = F'$ est bien positive pour tout $x \neq a_1, \dots, a_n$. Les valeurs de $f(a_1), \dots, f(a_n)$ sont ensuite choisies arbitrairement.

• Cas de ii :

Quitte à renommer les points, on peut supposer que $a_1 < \dots < a_n$. On note $a_0 = -\infty$ et $a_{n+1} = +\infty$

On note également

$$F(a_0) = \lim_{t \rightarrow a_0} F(t) = \lim_{t \rightarrow -\infty} F(t) = 0$$

et

$$F(a_{n+1}) = \lim_{t \rightarrow a_{n+1}} F(t) = 1$$

★ Étude de $\int_{a_i}^{a_{i+1}} f(t) dt$ pour $i = 0, \dots, n$:

On pose $f(x) = F'(x) \quad \forall x \neq a_1, \dots, a_n$ et $f(a_1), \dots, f(a_n)$ quelconques. Alors, sur chaque intervalle $]a_i, a_{i+1}[$, F est une primitive de f . Aussi, on a

$$\int_u^v f(t) dt = F(u) - F(v) \quad \forall u, v \in]a_i, a_{i+1}[$$

Il y a donc deux problèmes de convergence éventuels (a_i et a_{i+1}). Comme F est continue, les limites en a_i et a_{i+1} existent, ainsi, $\int_{a_i}^{a_{i+1}} f(t) dt$ est convergente et vaut, par passage à la limite en $u \rightarrow a_i$ et $v \rightarrow a_{i+1}$:

$$\int_{a_i}^{a_{i+1}} f(t) dt = F(a_{i+1}) - F(a_i) \quad \text{car } F \text{ continue}$$

Chaque morceau $\int_{a_i}^{a_{i+1}} f$ étant convergent, l'intégrale $\int_{-\infty}^{+\infty} f$ est convergent, avec

$$\int_{-\infty}^{+\infty} f = \sum_{i=0}^n \int_{a_i}^{a_{i+1}} f = \sum_{i=0}^n (F(a_{i+1}) - F(a_i)) = F(a_{n+1}) - F(a_0) = 1 - 0 = 1$$

En conclusion, f est bien une densité.

• Cas de iii

D'après ce qui précède, on sait que $\int_{-\infty}^x f$ converge. Soit $x \in \mathbb{R}$. Comme

$\bigcup_{i=0}^n]a_i, a_{i+1}[=]-\infty, +\infty[$, il existe $i \in \{0, \dots, n+1\}$ tel que $x \in]a_i, a_{i+1}[$. Alors

$$\int_{-\infty}^x f(t) dt = \int_{-\infty}^{a_1} f + \sum_{k=1}^{i-1} \int_{a_k}^{a_{k+1}} f + \int_{a_i}^x f$$

Or, on a vu ci-dessus que $\int_{a_k}^{a_{k+1}} f = F(a_{k+1}) - F(a_k)$ ainsi que $\int_{-\infty}^{a_1} f = F(a_1)$.

Le même principe que dans ce qui précède permet également de montrer que

$$\int_{a_{i-1}}^x f = F(x) - F(a_{i-1})$$

D'où $\int_{-\infty}^x f(t) dt = F(a_1) + \sum_{k=1}^{i-1} (F(a_{k+1}) - F(a_k)) + F(x) - F(a_{i-1}) = F(x)$

• En conclusion

on vient de montrer que F était bien associée à la fonction densité f . \square

■ Exemple 7 :

Soit X une variable aléatoire de fonction de répartition F_X définie par

$$F(t) = \begin{cases} 0 & \text{si } t < 1 \\ \ln t & \text{si } t \in [1, e[\\ 1 & \text{si } t \geq e \end{cases}$$

Dire si X est à densité et si c'est le cas, la déterminer.

• X est à densité :

On observe que F est une fonction de répartition continue sur $] -\infty, 1[, [1, e[$ et $[e, +\infty[$. En 1^- : $\lim_{t \rightarrow 1^-} F = 0 = \ln 1 = F(1)$. Ainsi, F est continue en 1^- (et en 1^+) d'après la remarque précédente, donc continue en 1.

En e^- : $\lim_{t \rightarrow e^-} F = \ln e = 1 = F(e)$. Ainsi, F est continue en e^- (et en e^+) d'après la remarque précédente, donc continue en e .

En conclusion, F est continue sur \mathbb{R} .

De plus, F est \mathcal{C}^1 sur \mathbb{R} sauf éventuellement en un nombre fini de points : 1 et e .

On peut donc dire que

X est une variable aléatoire à densité.

- Calcul de la densité :

On note f_X une densité de X . On a alors, pour tout $t \neq 1, e$,

$$f_X(t) = F'_X(t) = \begin{cases} 0 & \text{si } t < 1 \\ \frac{1}{t} & \text{si } t \in]1, e[\\ 0 & \text{si } t > e \end{cases}$$

On peut ensuite choisir arbitrairement les valeurs de $f(1)$ et $f(e)$. On propose par exemple

$$f_X(t) = \begin{cases} 0 & \text{si } t \leq 1 \\ \frac{1}{t} & \text{si } t \in [1, e] \\ 0 & \text{si } t > e \end{cases}$$



Définition

Si une variable aléatoire X est à densité, *donner la loi de X* signifie "prouver qu'elle admet une densité et la donner".

? Exercice 2

Reprendre l'exemple de la variable X qui suit une loi uniforme donnée par la fonction de répartition $F : x \mapsto \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } 0 \leq x < 1 \\ 1 & \text{si } x > 1 \end{cases}$. Montrer (cette fois-ci à l'aide de la technique précédente), que X est une variable à densité, puis déterminer (à nouveau) sa densité.

Solution

On a précédemment montré que X était à densité, notée ici f , avec

$$f(t) = \begin{cases} 0 & \text{si } t < 0 \\ 1 & \text{si } 0 \leq t \leq 1 \\ 0 & \text{si } t > 1 \end{cases}$$

II Moments, Espérance et variance

Dans tout ce paragraphe, sauf mention faite du contraire, on supposera que X est une variable aléatoire à densité.

🌿 Définition

Soit X une variable aléatoire réelle de densité f . On dit que X *admet une espérance* (ou que $\mathbb{E}[X]$ existe) si $\int_{-\infty}^{+\infty} |x|f(x) dx$ existe. On appelle alors espérance de X et on note $\mathbb{E}(X)$ la valeur

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

■ Exemple 8 :

Si X est une variable aléatoire de densité $f : t \mapsto \frac{1}{\pi(1+t^2)}$, alors elle n'admet pas d'espérance.

En effet, on observe que

$$|t|f(t) \underset{+\infty}{\sim} \frac{1}{\pi t} \geq 0$$

Or, l'intégrale $\int_1^{+\infty} \frac{1}{\pi t} dt$ diverge, donc il en va de même de $\int_1^{+\infty} |t|f(t)$, et donc de $\int_{-\infty}^{+\infty} |t|f(t)$.

🌿 Définition

Si X est une var à densité admettant une espérance $\mathbb{E}(X) = 0$, on dit qu'elle est *centrée*.

Propriété 149

On suppose que X est une var à densité qui admet une espérance. Si $X \geq 0$, alors $\mathbb{E}(X) \geq 0$.

Démonstration : exercice. \square

⚠️ Remarque :

Il n'existe pas de variable **positive** et à densité d'espérance nulle :

En effet, supposons que X soit positive et à densité (on la suppose continue. Alors elle admet une densité f telle que $f = 0$ sur $]-\infty; 0[$. Alors

$$0 = \mathbb{E}[X] = \int_0^{+\infty} \underbrace{xf(x)}_{\geq 0} dx$$

Ainsi, $xf(x)$ est une fonction nulle sur $]0; +\infty[$ (sauf éventuellement en un nombre fini de points). On suppose ici pour faciliter l'écriture qu'elle est nulle partout sur $]0; +\infty[$. On en déduit, par primitive, que

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 & \text{si } x > 0 \end{cases}$$

ce qui nous donne une fonction de répartition non continue sur \mathbb{R} , ce qui est impossible pour une variable à densité...

Théorème 150 ("linéarité")

On suppose que X et Y sont deux variables aléatoires admettant chacune une espérance.

- Alors $\mathbb{E}(X + Y)$ existe et $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$;
- Pour tous $a, b \in \mathbb{R}$, $aX + b$ est une variable à densité et $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$

Démonstration :

(premier point admis) et deuxième en exercice. \square



GARE AUX AFFIRMATIONS HATIVES !

On ne dit pas que $X + Y$ est forcément une variable à densité, mais seulement que l'espérance existe :

Par exemple, si X est une variable à densité, $-X$ également, mais

$$X + (-X) = 0$$

C'est une variable constante ; donc finie et non à densité.

Théorème 151 de transfert

Si X est une variable aléatoire réelle de densité f , ainsi qu'un intervalle I tel que $\varphi : I \rightarrow \mathbb{R}$ une fonction continue (sauf en un nombre fini de points). Alors

$\varphi(X)$ admet une espérance ssi

$$\left\{ \begin{array}{l} \text{il existe un intervalle ouvert }]a, b[\text{ tel que } \text{Supp}(X) \subset]a, b[\subset I \\ \text{et } \int_a^b |\varphi(x)| f(x) dx \text{ converge} \end{array} \right.$$

Dans ce cas,

$$\mathbb{E}(\varphi \circ X) = \int_a^b \varphi(x) f(x) dx$$

Démonstration :

- Cas $\varphi \in \mathcal{C}^1$ bijective croissante. Densité de $\varphi \circ X$:

$$P(\varphi(X) \leq x) = P(X \leq \varphi^{-1}(x)) = \int_{-\infty}^{\varphi^{-1}(x)} f(t) dt$$

On pose $t = \varphi(u)$, alors $dt = \varphi'(u) du$, d'où

$$\int_{-\infty}^{\varphi^{-1}(x)} f(t) dt = \int_{-\infty}^x f(\varphi(u)) \varphi'(u) du$$

Une densité de $\varphi \circ X$ est donc $t \mapsto f(\varphi(t)) \varphi'(t) \geq 0$.

$\varphi \circ X$ admet donc une espérance ssi

$$\int_{-\infty}^{+\infty} |x| f(\varphi(x)) \varphi'(x) dx \text{ existe}$$

Le changement de variable dans le sens inverse donne le résultat escompté.

- les autres cas sont admis. \square



Remarque :

Le théorème de transfert nous dit en particulier que **l'existence** de l'espérance d'une v.a. X est équivalente à l'existence de celle de $|X|$. De plus, on a la propriété suivante :

■ Exemple 9 :

Soit X une variable de densité f définie sur \mathbb{R} par $f(t) = \frac{1}{2}e^{-|t|}$. Alors la variable $X^3 \ln(1 + |X|)$ admet une espérance :

En effet, d'après le théorème de transfert, $X^3 \ln(1 + |X|)$ admet une espérance ssi $\frac{1}{2} \int_{-\infty}^{+\infty} |t^3 \ln(1 + |t|)| e^{-|t|} dt$ converge, ce qui ici, par parité de la fonction (et ensuite remplacement de la valeur absolue dans l'intégrale), revient à la convergence de $I = \int_0^{+\infty} t^3 \ln(1 + t) e^{-t} dt$.

La fonction dans l'intégrale est continue sur $[0, +\infty[$, et on a

$$0 \leq t^3 \ln(1 + t) e^{-t} \leq t^3 t e^{-t} = t^4 e^{-t} \quad \forall t \geq 0 \quad (\text{exo})$$

Or, d'après le cours, $\int_0^{+\infty} t^4 e^{-t} dt$ converge. Par théorème de comparaison des intégrales de fonctions positives, on sait alors que I converge et que donc l'espérance existe.

Propriété 152

Si X admet une espérance, alors $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$

Démonstration :

D'après le théorème de transfert, $|X|$ admet une espérance si et seulement si $\int_{-\infty}^{+\infty} |x| f(x) dx$ converge. Or, c'est exactement la même condition que pour l'existence de l'espérance de X .

De plus, toujours par ce même théorème, cette valeur correspond à $\mathbb{E}[|X|]$. Voyons maintenant la comparaison des valeurs :

Pour tout $x \in \mathbb{R}$, on a $-|x| \leq x \leq |x|$

ainsi, comme f est positive, $-|x|f(x) \leq xf(x) \leq |x|f(x)$

D'où, par convergence de toutes les intégrales, $-\mathbb{E}[|X|] \leq \mathbb{E}[X] \leq \mathbb{E}[|X|]$. \square

■ Exemple 10 :

Dans l'exemple précédent, étant donné que

$$|X^3 \ln(1 + |X|)| \leq X^4 \quad (\text{exo})$$

et que $\mathbb{E}[X^4]$ existe et vaut $\frac{4!}{1^4} = 24$ (exo) on en déduit que

$$|\mathbb{E}[X^3 \ln(1 + |X|)]| \leq 24$$

Commentaires :

| Le théorème de transfert légitime également la définition suivante :



Définition

Si X^n admet une espérance (i.e. si $\int_{-\infty}^{+\infty} |x|^n f(x) dx < \infty$), on dit que le moment d'ordre n existe et on appelle $\mathbb{E}(X^n) = \int_{-\infty}^{+\infty} x^n f(x) dx$ le *moment d'ordre n* de la variable X

Théorème 153

Si X est une variable aléatoire bornée à densité, alors elle admet des moments de tout ordre.

Démonstration :

Soit M un majorant de $|X|$ et f la densité de X . Alors, pour tout $n \in \mathbb{N}$ Comme on a $Supp(X) \subset [-M, M]$, la densité f est nulle en dehors de $[-M, M]$.

Ainsi, la convergence de $\int_{-\infty}^{+\infty} |x|^n f(x) dx$ équivaut à celle de $\int_{-M}^M |x|^n f(x) dx$.

Or, comme pour tout $x \in \mathbb{R}$ on a

$$|x|^n f(x) \leq M^n f(x)$$

on en déduit, comme $\int_{-M}^M f$ converge, par théorème de comparaison des intégrales,

que $\int_{-M}^M |x|^n f(x) dx$ converge et donc que le moment d'ordre n existe. \square

? Exercice 3

Soit X une variable aléatoire de fonction de répartition F_X définie par

$$F(t) = \begin{cases} 0 & \text{si } t < 1 \\ \frac{e^t - e}{e^e - e} \ln t & \text{si } t \in [1, e] \\ 1 & \text{si } t > e \end{cases}$$

Montrer que X admet des moments de tout ordre.

Solution

La variable est à densité et bornée car $1 \leq X \leq e$ p.s., donc admet des moments de tout ordre.

Proposition 154

Soit $n \in \mathbb{N}$. Si X^n admet une espérance, alors X^m admet également une espérance pour tout $0 \leq m \leq n$.



Définition

Si X^2 est intégrable, la quantité

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

existe et est appelée *variance* de X .

On appelle également $\sigma(X) = \sqrt{V(X)}$ l'*écart-type* de X .

Propriété 155

Soient X admettant un moment d'ordre 2 et $a, b \in \mathbb{R}$. Alors, on a

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

Démonstration : exercice. \square

Propriété 156

Soient X et Y deux variables aléatoires définies sur un même espace probabilisé.

H1 : Si les moments d'ordre 2 de X et Y existent

H2 : Si les variables X et Y sont indépendantes,

alors

$$V(X + Y) = V(X) + V(Y)$$

Démonstration : exercice. \square

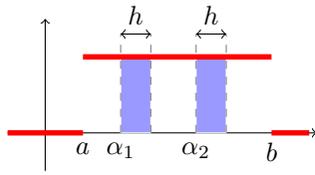
III Exemples fondamentaux

III-1 Loi uniforme

Commentaires :

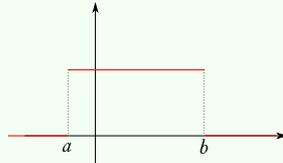
On cherche à décrire la loi d'une variable donnant un nombre au hasard dans un intervalle $[a, b]$. On sait qu'une densité existe pour l'intervalle $]0, 1[$ et on imagine donc qu'il en existe une sur $[a, b]$. Dans ce cas :

- $Supp(X) \subset [a, b]$ donc la densité est nulle en dehors de $[a, b]$.
- Pour h constant, les probabilités du type $P(\alpha < X \leq \alpha + h)$ (aire sous la courbe densité entre α et $\alpha + h$) avec $a \leq \alpha < \alpha + h$ étant toutes identiques (par choix au hasard), on imagine que s'il y a une densité, elle doit nécessairement être constante.



Définition et Proposition

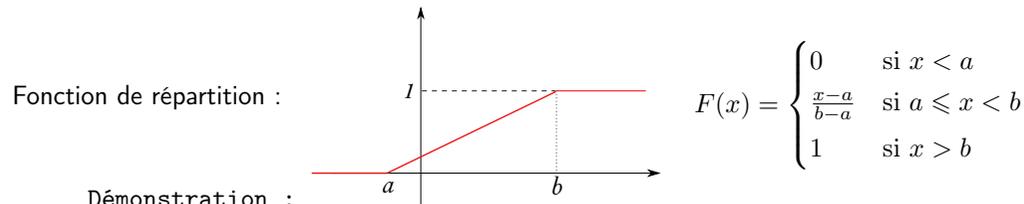
Soit X une variable aléatoire réelle. On dit qu'elle suit une *loi uniforme* sur $[a, b]$ ($a < b$) si elle admet comme densité la fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ telle que

$$f(t) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(t) = \begin{cases} 0 & \text{si } t < a \\ \frac{1}{b-a} & \text{si } a \leq t \leq b \\ 0 & \text{si } t > b \end{cases}$$


On note $\mathcal{L}(X) = \mathcal{U}_{[a,b]}$ ou $X \hookrightarrow \mathcal{U}_{[a,b]}$.

Démonstration :

On vérifie que c'est bien une densité \square



Démonstration :

On note F la fonction de répartition de X . La fonction densité est continue sauf en 0 et 1. Ainsi, F est \mathcal{C}^1 sur \mathbb{R} sauf en 0 et 1, avec F une primitive de f . Ainsi, il existe $\alpha, \beta, \gamma \in \mathbb{R}$ tels que :

$$F_X(x) = \begin{cases} \alpha & \text{si } x < a \\ \frac{x-a}{b-a} + \beta & \text{si } a < x < b \\ \gamma & \text{si } x > b \end{cases}$$

Les limites de F en $\pm\infty$ donnent $\alpha = 0$ et $\gamma = 1$. F est la fonction de répartition d'une variable à densité, ainsi, F est continue. La continuité en a donne alors

D'où

$$\beta = -\frac{a}{b-a}$$

$$F(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{x-a}{b-a} & \text{si } a \leq x < b \\ 1 & \text{si } x > b \end{cases}$$

Pour finir, la continuité en b permet de mettre une inégalité large sur $a \leq x \leq b$ ou $x \geq b$. \square

Commentaires :

X suivant une loi uniforme sur $[a, b]$ peut être le résultat d'un tirage au hasard d'un nombre entre a et b .

Remarque :

On peut généraliser cette définition à toute loi uniforme sur $[a, b[$ ou $]a, b]$, ou encore $]a, b[$ en adaptant f , mais en réalité, c'est inutile, car les fonctions de répartition sont en réalité égales.

Proposition 157

Si $X \rightsquigarrow \mathcal{U}_{[a,b]}$, alors les moments de tout ordre existent.

En particulier,

$$\mathbb{E}(X) = \frac{a+b}{2} \text{ et } V(X) = \frac{(b-a)^2}{12}.$$

Démonstration :

On note $f = \frac{1}{b-a} \mathbb{1}_{[a,b]}$ la densité de X .

$$\int_{-\infty}^{+\infty} |x^n| f(x) dx = \frac{1}{b-a} \int_a^b |x^n| dx \text{ existe}$$

alors $|X^n|$ est intégrable pour tout $n \in \mathbb{N}$.

• Espérance :

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f(x) dx = \frac{1}{b-a} \int_a^b x dx = \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b = \frac{a+b}{2}$$

• Variance :

$$\mathbb{E}(X^2) = \int_{-\infty}^{+\infty} x^2 f(x) dx = \frac{1}{b-a} \int_a^b x^2 dx = \frac{1}{b-a} \left[\frac{x^3}{3} \right]_a^b = \frac{1}{b-a} \frac{b^3 - a^3}{3} = \frac{a^2 + ab + b^2}{3}$$

D'où

$$V(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{a^2 + ab + b^2}{3} - \frac{a^2 + b^2 + 2ab}{4} = \frac{a^2 - 2ab + b^2}{12}$$

\square

Remarque : En particulier, si $X \rightsquigarrow \mathcal{U}_{[0,1]}$, on a $\mathbb{E}(X) = \frac{1}{2}$ et $V(X) = \frac{1}{12}$.

Proposition 158

Soient $a < b \in \mathbb{R}$. Alors X suit une loi uniforme sur $[0, 1]$ si et seulement si $Y = (b-a)X + a$ suit une loi uniforme sur $[a, b]$.

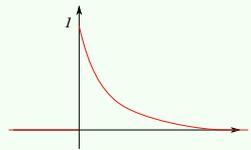
Démonstration : exercice. \square

III-2 Loi exponentielle

Définition et Proposition

Soient $\lambda > 0$ et X une variable aléatoire réelle. On dit qu'elle suit une *loi exponentielle* de paramètre λ si elle admet comme densité la fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ telle que

$$f(t) = \lambda e^{-\lambda t} \mathbf{1}_{[0, +\infty[}(t) = \begin{cases} 0 & \text{si } t < 0 \\ \lambda e^{-\lambda t} & \text{si } t \geq 0 \end{cases}$$

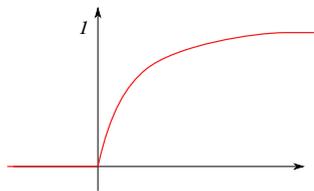


On note $\mathcal{L}(X) = \varepsilon(\lambda)$ ou $X \leftrightarrow \varepsilon(\lambda)$.

Démonstration :

On vérifie que c'est bien une densité \square

Fonction de répartition :



$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - e^{-\lambda x} & \text{si } x \geq 0 \end{cases}$$

Démonstration : exercice. \square

Proposition 159

Si $X \rightsquigarrow \varepsilon(\lambda)$, alors les moments de tout ordre existent.

En particulier,

$$\mathbb{E}(X) = \frac{1}{\lambda} ; \quad V(X) = \frac{1}{\lambda^2} \quad \text{et} \quad \mathbb{E}(X^n) = \frac{n!}{\lambda^n} \quad \forall n \in \mathbb{N}.$$

Démonstration :

On note $f : \mathbb{R} \rightarrow \mathbb{R}$ telle que $f(t) = \lambda e^{-\lambda t} \mathbf{1}_{[0, +\infty[}(t)$ la densité de X .

• Existence des moments : Soit $n \in \mathbb{N}$. Alors

$$\int_{-\infty}^{+\infty} |t^n| f(t) dt = \lambda \int_0^{+\infty} t^n e^{-\lambda t} dt$$

Cette intégrale est convergente (d'après les exemples usuels du chapitre précédent sur les intégrales.) On sait que plus que

$$\int_0^{+\infty} t^n e^{-\lambda t} dt = \frac{n!}{\lambda^{n+1}}$$

Comme la variable X est positive, on peut réutiliser cette valeur pour calculer

$$\mathbb{E}[X^n] = \lambda \frac{n!}{\lambda^{n+1}} = \frac{n!}{\lambda^n}$$

• Espérance :

$$\mathbb{E}(X) = I_1 = \frac{1}{\lambda} I_0 = \frac{1}{\lambda}$$

• Variance :

$$\mathbb{E}(X^2) = I_2 = \frac{2}{\lambda} I_1 = \frac{2}{\lambda^2}$$

D'où

$$V(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

\square

Propriété 160 *d'absence de mémoire*

Soit X suivant une loi $\mathcal{E}(\lambda)$. Pour tout $t, s > 0$, on a

$$P_{X>s}(X > t + s) = P(X > t)$$

Commentaires :

On peut interpréter ceci de la manière suivante :

Si X correspond au nombre de minutes d'attente avant l'arrivée d'un vendeur dans une boutique, quelqu'un qui déjà attendu pendant un temps s a exactement autant de chance d'attendre encore pendant un temps t qu'une autre personne qui vient d'arriver.

■ Exemple 11 :

Avec les notations de la propriété précédente, on a par exemple

$$P_{X>2}(X > 3) = P(X > 1)$$

Démonstration :

Par calcul :

$$P_{X>s}(X > t + s) = \frac{P(X > s \cap X > t + s)}{P(X > s)} = \frac{P(X > t + s)}{P(X > s)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = P(X > t)$$

\square

⚠ Remarque :

La seule loi discrète à avoir cette même propriété est la loi géométrique. En effet, rappelle que si $X \leftrightarrow \mathcal{G}(p)$, alors on a également Pour tout $t, s > 0$, on a

$$P_{X>s}(X > t + s) = P(X > t)$$



CONFUSION

La propriété d'absence de mémoire ne dit en aucun cas qu'il y a égalité entre $P_{X>2}(X > 3)$ et $P(X > 2)$. En effet, ceci signifierait plutôt que les événements $(X > 2)$ et $(X > 3)$ sont indépendants, ce qui n'est pas du tout le cas.

III-3 Loi normale

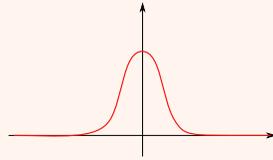
III.3-a) Loi normale centrée réduite

Théorème 161

La fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ telle que

$$f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

est une densité de probabilité



Démonstration : admise. \square

Définition

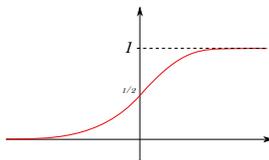
On dit qu'une variable aléatoire réelle suit une *loi normale centrée réduite* si elle admet comme densité la fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ telle que

$$f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

On note $\mathcal{L}(X) = \mathcal{N}(0, 1)$ ou $X \hookrightarrow \mathcal{N}(0, 1)$.

Notation : Dans le cas particulier de la loi normale centrée réduite, on note Φ sa fonction de répartition.

Graphes de Φ :



Proposition 162

la fonction de répartition Φ de X suivant une loi normale $\mathcal{N}(0; 1)$ vérifie :

- $\forall x \in \mathbb{R}, \Phi(-x) = 1 - \Phi(x)$; $\Phi(0) = \frac{1}{2}$
- $\forall x \geq 0, P(|X| \leq x) = 2\Phi(x) - 1$ et $P(|X| \geq x) = 2(1 - \Phi(x))$.

Démonstration : admise. \square

Proposition 163

Si $X \rightsquigarrow \mathcal{N}(0, 1)$, alors les moments de tout ordre existent.

En particulier, $\mathbb{E}(X) = 0$ et $\sigma(X) = 1$.

Démonstration :

- Montrons que X admet des moments de tous ordres :

Soit $n \in \mathbb{N}$. Montrons que l'intégrale $\int_{-\infty}^{+\infty} |x|^n e^{-\frac{x^2}{2}} dx$ converge :

La fonction $x \mapsto |x|^n e^{-\frac{x^2}{2}}$ étant continue sur \mathbb{R} , les seuls problèmes sont en $\pm\infty$. De plus, cette fonction est également paire, ce qui signifie que la convergence en $-\infty$ équivaut à la convergence en $+\infty$. Étudions donc la convergence de en $+\infty$.

Par croissance comparée, on sait que

$$\lim_{x \rightarrow +\infty} e^{\frac{x^2}{4}} \cdot |x|^n e^{-\frac{x^2}{2}} = 0$$

Il existe alors X tel que,

$$\forall x \geq X \quad 0 \leq |x|^n e^{-\frac{x^2}{2}} \leq e^{-\frac{x^2}{4}}$$

Or, l'intégrale $\int_X^{+\infty} e^{-\frac{x^2}{4}} dx$ converge. On en déduit, par comparaison d'intégrales de fonctions positives, que

$$\int_X^{+\infty} |x|^n e^{-\frac{x^2}{2}} dx \text{ converge}$$

et donc

$$\int_{-\infty}^{+\infty} |x|^n e^{-\frac{x^2}{2}} dx \text{ converge.}$$

- Si on ne veut démontrer que l'existence de l'espérance de X :

La fonction $x \mapsto |x| e^{-\frac{x^2}{2}}$ étant continue sur \mathbb{R} , les seuls problèmes sont en $\pm\infty$. De plus, cette fonction est également paire, ce qui signifie que la convergence en $-\infty$ équivaut à la convergence en $+\infty$. Ainsi elle converge ssi elle converge en $+\infty$.

Soit $t \geq 0$. On a

$$\int_0^{+\infty} |x| f(x) dx = \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} x e^{-\frac{x^2}{2}} dx = \left[-e^{-\frac{x^2}{2}} \right]_0^{+\infty} = 1 - e^{-\frac{t^2}{2}} \xrightarrow{t \rightarrow +\infty} 1$$

Ainsi, l'intégrale converge et donc $\mathbb{E}[X]$ existe.

- Calculons l'espérance de X .

Grace à la démonstration ci-dessus, on sait que $\int_{-\infty}^{+\infty} x e^{-\frac{x^2}{2}} dx$ converge (absolument). De plus, la fonction $x \mapsto x e^{-\frac{x^2}{2}}$ est impaire, ce qui entraîne

$$\int_{-\infty}^{+\infty} x e^{-\frac{x^2}{2}} dx = 0$$

et donc

$$\mathbb{E}(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{-\frac{x^2}{2}} dx = 0$$

- Calculons l'espérance de X^2 .

Grace à la démonstration ci-dessus, on sait que $\int_{-\infty}^{+\infty} x^2 e^{-\frac{x^2}{2}} dx$ converge. De plus,

la fonction $x \mapsto x e^{-\frac{x^2}{2}}$ est paire, ce qui entraîne

$$\int_{-\infty}^{+\infty} x^2 e^{-\frac{x^2}{2}} dx = 2 \int_0^{+\infty} x^2 e^{-\frac{x^2}{2}} dx$$

Effectuons une Ipp :

$$\begin{aligned} u &= x & u' &= 1 \\ v' &= x e^{-\frac{x^2}{2}} & v &= -e^{-\frac{x^2}{2}} \quad \text{où } u, v \in C^1 \end{aligned}$$

Comme $\lim_{x \rightarrow +\infty} u(x)v(x) = \lim_{x \rightarrow +\infty} -x e^{-\frac{x^2}{2}}$ existe par croissance comparée, on a

$$\int_0^{+\infty} x^2 e^{-\frac{x^2}{2}} dx = \underbrace{\left[-x e^{-\frac{x^2}{2}} \right]_0^{+\infty}}_0 + \underbrace{\int_0^{+\infty} e^{-\frac{x^2}{2}} dx}_{\frac{\sqrt{2\pi}}{2}}$$

D'où

$$\int_{-\infty}^{+\infty} x^2 e^{-\frac{x^2}{2}} dx = 1$$

• Finalement, la variance :

Le moment d'ordre 2 existe, on peut donc appliquer la formule de la variance :

$$\square \quad V(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = 1$$

————— Le calcul des valeurs précédentes avec Python :

Avec Python, on peut simuler une variable aléatoire suivant une loi normale ainsi que sa fonction de répartition ou l'inverse de sa fonction de répartition :

```
import scipy.stats as sc

va=sc.norm()      # on crée une v.a. du nom de va qui
                  # suit une loi N(0,1).

va.cdf(x)        # rend phi(x)
va.ppf(x)        # rend phi^{-1}(x)
```



Remarque :

Les abréviations ci-dessus correspondent aux termes suivants :

```
cdf      # cumulative density fonction
ppf      # percent point function
```

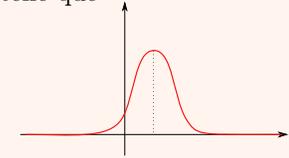
Pour le calcul des valeurs précédentes à la calculatrice, on se référera à l'annexe de fin de chapitre.

III.3-b) Loi normale

Proposition 164

Soient $\mu \in \mathbb{R}, \sigma > 0$. La fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ telle que

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}$$



est une densité de probabilité

Démonstration :

On vérifie que c'est bien une densité : rappel :

$$\int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = \sqrt{2\pi}$$

Calculons $\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt :$

Changement de variable $u = \frac{t-\mu}{\sigma}$, d'où $du = \frac{1}{\sigma} dt$, i.e. $dt = \sigma du$. D'où (exercice)

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt = \sigma \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} dt = \sigma\sqrt{2\pi}$$

□

Définition

Soit X une variable aléatoire réelle. On dit qu'elle suit une *loi normale* (ou gaussienne) de paramètres μ, σ si elle admet comme densité la fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ telle que

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}$$

On note $\mathcal{L}(X) = \mathcal{N}(\mu, \sigma^2)$ ou $X \hookrightarrow \mathcal{N}(\mu, \sigma^2)$.

Proposition 165

Si $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$, alors les moments de tout ordre existent.

En particulier,

$$\mathbb{E}(X) = \mu \text{ et } V(X) = \sigma^2.$$

Démonstration :

- Existence des moments

On note $a, b \in \mathbb{R}$ et on considère l'intégrale $\int_a^b |t|^n e^{-\frac{1}{2}(\frac{t-\mu}{\sigma})^2} dt$.

Effectuons un changement de variable

$$x = \frac{t-\mu}{\sigma}, \quad t = \sigma x + \mu \quad dt = \sigma dx$$

On obtient

$$\int_a^b |t|^n e^{-\frac{1}{2}(\frac{t-\mu}{\sigma})^2} dt = \sigma \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} |\sigma x + \mu|^n e^{-\frac{1}{2}x^2} dx$$

$$\begin{aligned} \text{Or, } 0 \leq |\sigma x + \mu|^n e^{-\frac{1}{2}x^2} &\leq (\sigma|x| + |\mu|)^n e^{-\frac{1}{2}x^2} \\ &= \sum_{k=0}^n \binom{n}{k} (\sigma|x|)^k |\mu|^{n-k} e^{-\frac{1}{2}x^2} \\ &= \sum_{k=0}^n \binom{n}{k} \sigma^k |\mu|^{n-k} |x|^k e^{-\frac{1}{2}x^2} \end{aligned}$$

Or, on sait, d'après le cas des moments d'une variable aléatoire suivant une loi normale centrée réduite, que chacune des intégrales $\int_{-\infty}^{+\infty} |x|^k e^{-\frac{1}{2}x^2} dx$ converge, ce qui donne, par multiplication par les constantes et addition d'intégrales convergentes, la convergence de $\int \sum_{k=0}^n \binom{n}{k} \sigma^k |\mu|^{n-k} |x|^k e^{-\frac{1}{2}x^2} dx$.

Par comparaison d'intégrales de fonctions positives, on en déduit la convergence de

$$\int_{-\infty}^{+\infty} |\sigma x + \mu|^n e^{-\frac{1}{2}x^2} dx$$

par passage à la limite quand $a \rightarrow -\infty$ et $b \rightarrow +\infty$, on obtient donc la convergence de $\int_{-\infty}^{+\infty} |t|^n e^{-\frac{1}{2}(\frac{t-\mu}{\sigma})^2} dt$.

- Espérance de X :

$$\text{Calcul de } \int_{-\infty}^{+\infty} \frac{t}{\sigma} e^{-\frac{1}{2}(\frac{t-\mu}{\sigma})^2} dt :$$

Servons nous du changement de variable effectué ci-dessus. On trouve, après passage à la limite,

$$\begin{aligned} \int_{-\infty}^{+\infty} t e^{-\frac{1}{2}(\frac{t-\mu}{\sigma})^2} dt &= \sigma \int_{-\infty}^{+\infty} (\sigma x + \mu) e^{-\frac{1}{2}x^2} dx \\ &\underbrace{=}_{\text{cv des int.}} \underbrace{\sigma^2 \int_{-\infty}^{+\infty} x e^{-\frac{1}{2}x^2} dx}_0 + \underbrace{\mu \sigma \int_{-\infty}^{+\infty} e^{-\frac{1}{2}x^2} dx}_{\sqrt{2\pi}} \end{aligned}$$

En conclusion,

$$\mathbb{E}(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} t e^{-\frac{1}{2}(\frac{t-\mu}{\sigma})^2} dt = \frac{1}{\sigma\sqrt{2\pi}} \sigma \mu \sqrt{2\pi} = \mu$$

- Espérance de X^2 :

Le même changement de variables, connaissant la convergence de toutes les intégrales concernées, donne

$$\begin{aligned} \mathbb{E}(X^2) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} t^2 e^{-\frac{1}{2}(\frac{t-\mu}{\sigma})^2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma x + \mu)^2 e^{-\frac{1}{2}x^2} dx \\ &= \underbrace{\sigma^2 \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 e^{-\frac{1}{2}x^2} dx}_1 + \underbrace{2\sigma\mu \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{-\frac{1}{2}x^2} dx}_0 + \underbrace{\mu^2 \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}x^2} dx}_1 \end{aligned}$$

D'où

$$\mathbb{E}(X^2) = \sigma^2 + \mu^2$$

- variance de X :

Comme le moment d'ordre 2 existe, on peut appliquer la formule :

$$\square \quad V(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \sigma^2$$

IV Opérations sur les variables aléatoires

IV-1 Multiplication et addition par un scalaire

? Exercice 4

Soit X une variable aléatoire de densité f et $b \in \mathbb{R}$. Alors, une densité de $X + b$ est $h : x \mapsto f(x - b)$

Solution

Pour tout x , $F_{X+b}(x) = P(X + b \leq x) = P(X \leq x - b) = F_X(x - b)$

On sait, de plus, que F_X est dérivable sur $\mathbb{R} - \{a_1, \dots, a_n\}$, où les a_1, \dots, a_n sont les éventuels points de discontinuité de f . Ainsi, en dérivant l'expression précédente, on trouve que, pour $\mathbb{R} - \{a_1 + b, \dots, a_n + b\}$,

$$F'_{X+b}(x) = f(x - b)$$

Ainsi, $X + b$ admet une densité, et celle-ci est

$$x \mapsto f(x - b)$$

? Exercice 5

Soit X une variable aléatoire de densité f et $a \in \mathbb{R}^*$. Alors, une densité de aX est $h : x \mapsto \frac{1}{|a|} f\left(\frac{x}{a}\right)$

Solution

- Si $a > 0$,

$$F_{aX}(x) = P(aX \leq x) = P\left(X \leq \frac{x}{a}\right) = F_X\left(\frac{x}{a}\right)$$

On sait, de plus, que F_X est dérivable sur $\mathbb{R} - \{a_1, \dots, a_n\}$, où les a_1, \dots, a_n sont les éventuels points de discontinuité de f . Ainsi, en dérivant l'expression précédente, on trouve que, pour $\mathbb{R} - \{\frac{a_1}{a}, \dots, \frac{a_n}{a}\}$,

$$F'_{aX}(x) = \frac{1}{a} f\left(\frac{x}{a}\right)$$

Ainsi, aX admet une densité, et celle-ci est

- Si $a < 0$,
à faire $x \mapsto \frac{1}{|a|} f\left(\frac{x}{a}\right)$

Corollaire 1

Soient $\mu \in \mathbb{R}$, $\sigma > 0$ et X une variable aléatoire réelle suivant une loi normale $\mathcal{N}(\mu, \sigma^2)$. On a alors

- $\forall a \in \mathbb{R}^*$, $aX \rightsquigarrow \mathcal{N}(a\mu, (a\sigma)^2)$
- $\forall b \in \mathbb{R}$, $X + b \rightsquigarrow \mathcal{N}(\mu + b, \sigma^2)$

Démonstration : exercice. \square

Corollaire 2

Soit $\lambda > 0$. Si X suit une loi exponentielle de paramètre λ , ($X \rightsquigarrow \varepsilon(\lambda)$) alors, pour tout $a > 0$,

$$aX \rightsquigarrow \varepsilon\left(\frac{\lambda}{a}\right)$$

Démonstration :

$$f_{aX}(x) = \frac{1}{|a|} f_X\left(\frac{x}{a}\right) = \frac{1}{|a|} \lambda e^{-\lambda \frac{x}{a}} = \frac{\lambda}{a} e^{-\frac{\lambda}{a} x} \quad \square$$

⚠ Remarque :

- $X + b$ ne suit pas une loi exponentielle si $b \neq 0$, car le support de $X + b$ n'est pas $[0; +\infty[$.
- De même, si $a < 0$, aX ne suit pas non plus une loi exponentielle. (Dire pourquoi.)

IV-2 Somme de va indépendantes

Théorème 166 de convolution

Soient X, Y deux variables aléatoires indépendantes de densité f et g . Alors, $X + Y$ est une variable aléatoire de densité h telle que

$$h(x) = \int_{-\infty}^{+\infty} f(t)g(x-t) dt = \int_{-\infty}^{+\infty} f(x-t)g(t) dt$$

■ Exemple 12 :

Soient X, Y deux variables aléatoires indépendantes suivant des mêmes lois uniformes sur $[0; 1]$. Déterminer la loi de $X + Y$. (*fait en classe*)

Après calcul (faits en classe) on a

$$h(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } 0 \leq x \leq 1 \\ 2-x & \text{si } 1 \leq x \leq 2 \\ 0 & \text{si } x > 2 \end{cases}$$

Corollaire

Si X_1, X_2 sont indépendantes et suivent respectivement des lois normales $\mathcal{N}(\mu_1, \sigma_1^2)$, $\mathcal{N}(\mu_2, \sigma_2^2)$, alors

$$X + Y \text{ suit une loi } \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

Démonstration :

Soit $x \in \mathbb{R}$. Calculons $h(x) = \int_{-\infty}^{+\infty} f_1(t)f_2(x-t) dt$ où f_i est une densité de X_i . Tout d'abord, soit $t \in \mathbb{R}$.

$$\begin{aligned} f_1(t)f_2(x-t) &= \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu_1}{\sigma_1}\right)^2} \frac{1}{\sigma_2\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-t-\mu_2}{\sigma_2}\right)^2} \\ &= \frac{1}{\sigma_1\sigma_2 2\pi} e^{-\frac{1}{2}\left(\left(\frac{t-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x-t-\mu_2}{\sigma_2}\right)^2\right)} \end{aligned}$$

Or, en développant et en regroupant les termes suivant t , on trouve

$$\left(\frac{t-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x-t-\mu_2}{\sigma_2}\right)^2 = at^2 + 2bt + c,$$

où

$$a = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}; \quad b = -\frac{1}{\sigma_1^2}\mu_1 - \frac{1}{\sigma_2^2}(x - \mu_2); \quad c = \frac{1}{\sigma_1^2}\mu_1^2 + \frac{1}{\sigma_2^2}(x - \mu_2)^2$$

Mettons $-\frac{1}{2}(at^2 + bt + c)$ sous forme canonique :

$$-\frac{1}{2}(at^2 + 2bt + c) = -\frac{a}{2}\left(t - \frac{b}{a}\right)^2 + \frac{b^2 - ac}{2a}$$

Dans l'intégrale, on obtient

$$h(x) = \frac{1}{\sigma_1\sigma_2 2\pi} \int_{-\infty}^{+\infty} e^{-\frac{a}{2}\left(t - \frac{b}{a}\right)^2 + \frac{b^2 - ac}{2a}} dt = \frac{1}{\sigma_1\sigma_2 2\pi} e^{-\frac{a}{2}\frac{b^2 - ac}{2a}} \int_{-\infty}^{+\infty} e^{-\frac{a}{2}\left(t - \frac{b}{a}\right)^2} dt$$

Or, reconnaissant la densité d'une loi normale $\mathcal{N}\left(\frac{b}{a}, \frac{1}{a}\right)$, on sait que

$$\int_{-\infty}^{+\infty} e^{-\frac{a}{2}\left(t - \frac{b}{a}\right)^2} dt = \sqrt{\frac{2\pi}{a}}$$

D'où

$$h(x) = \frac{\sqrt{2\pi}}{\sigma_1\sigma_2 2\pi \sqrt{a}} e^{\frac{b^2 - ac}{2a}} = \frac{1}{\sigma_1\sigma_2 \sqrt{2a\pi}} e^{\frac{b^2 - ac}{2a}}$$

• Calcul de $\sigma_1\sigma_2\sqrt{2a\pi}$:

Comme $a = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}$, on a

$$\sigma_1\sigma_2\sqrt{2a\pi} = \sqrt{2\sigma_1^2\sigma_2^2\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)\pi} = \sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}$$

• Calcul de $\frac{b^2 - ac}{a}$:

Comme

$$b = -\frac{1}{\sigma_1^2}\mu_1 - \frac{1}{\sigma_2^2}(x - \mu_2), \quad c = \frac{1}{\sigma_1^2}\mu_1^2 + \frac{1}{\sigma_2^2}(x - \mu_2)^2, \quad a = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}$$

on a

$$\begin{aligned} b^2 - ac &= \left(-\frac{1}{\sigma_1^2}\mu_1 - \frac{1}{\sigma_2^2}(x - \mu_2)\right)^2 - a\left(\frac{1}{\sigma_1^2}\mu_1^2 + \frac{1}{\sigma_2^2}(x - \mu_2)^2\right) \\ &= \underbrace{\mu_1^2\left(\frac{1}{\sigma_1^4} - \frac{a}{\sigma_1^2}\right)}_{-\frac{1}{\sigma_1^2\sigma_2^2}} + (x - \mu_2)^2 \underbrace{\left(\frac{1}{\sigma_2^4} - \frac{a}{\sigma_2^2}\right)}_{-\frac{1}{\sigma_1^2\sigma_2^2}} + 2\frac{\mu_1(x - \mu_2)}{\sigma_1^2\sigma_2^2} \\ &= \frac{1}{\sigma_1^2\sigma_2^2}(-\mu_1^2 - (x - \mu_2)^2 + 2\mu_1(x - \mu_2)) \\ &= -\frac{1}{\sigma_1^2\sigma_2^2}(x - (\mu_1 + \mu_2))^2 \end{aligned}$$

Au final, on obtient

$$\frac{b^2 - ac}{a} = \frac{-\frac{1}{\sigma_1^2\sigma_2^2}(x - (\mu_1 + \mu_2))^2}{\frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2\sigma_2^2}} = -\frac{(x - (\mu_1 + \mu_2))^2}{\sigma_1^2 + \sigma_2^2}$$

En conclusion, la densité de $X + Y$ est

$$x \mapsto \int_{-\infty}^{+\infty} f_1(t)f_2(x - t) dt = \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} e^{-\frac{(x - (\mu_1 + \mu_2))^2}{\sigma_1^2 + \sigma_2^2}}$$

qui est bien la densité d'une variable aléatoire suivant une loi normale

$$\mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

□

V Annexe : Tables et calculatrice

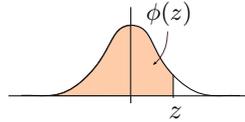
TABLE INVERSE DE LA LOI NORMALE CENTRÉE RÉDUITE

z en fonction de α tel que $\alpha = P[Z \leq z] = \Phi(z)$ pour $Z \sim \mathcal{N}(0; 1)$.

$\alpha = \Phi(z)$	0,7	0,8	0,9	0,95	0,96	0,97	0,975	0,98	0,99	0,995
z	0,524	0,842	1,282	1,645	1,751	1,881	1,960	2,054	2,326	2,576

TABLE DE LA LOI NORMALE CENTRÉE RÉDUITE

$\Phi(z) = P[Z \leq z]$ en fonction de z pour $Z \sim \mathcal{N}(0; 1)$.



	0	1	2	3	4	5	6	7	8	9
0	0,5	0,50399	0,50798	0,51197	0,51595	0,51994	0,52392	0,5279	0,53188	0,53586
0,1	0,53983	0,5438	0,54776	0,55172	0,55567	0,55962	0,56356	0,56749	0,57142	0,57535
0,2	0,57926	0,58317	0,58706	0,59095	0,59483	0,59871	0,60257	0,60642	0,61026	0,61409
0,3	0,61791	0,62172	0,62552	0,6293	0,63307	0,63683	0,64058	0,64431	0,64803	0,65173
0,4	0,65542	0,6591	0,66276	0,6664	0,67003	0,67364	0,67724	0,68082	0,68439	0,68793
0,5	0,69146	0,69497	0,69847	0,70194	0,7054	0,70884	0,71226	0,71566	0,71904	0,7224
0,6	0,72575	0,72907	0,73237	0,73565	0,73891	0,74215	0,74537	0,74857	0,75175	0,7549
0,7	0,75804	0,76115	0,76424	0,7673	0,77035	0,77337	0,77637	0,77935	0,7823	0,78524
0,8	0,78814	0,79103	0,79389	0,79673	0,79955	0,80234	0,80511	0,80785	0,81057	0,81327
0,9	0,81594	0,81859	0,82121	0,82381	0,82639	0,82894	0,83147	0,83398	0,83646	0,83891
1	0,84134	0,84375	0,84614	0,84849	0,85083	0,85314	0,85543	0,85769	0,85993	0,86214
1,1	0,86433	0,8665	0,86864	0,87076	0,87286	0,87493	0,87698	0,879	0,881	0,88298
1,2	0,88493	0,88686	0,88877	0,89065	0,89251	0,89435	0,89617	0,89796	0,89973	0,90147
1,3	0,9032	0,9049	0,90658	0,90824	0,90988	0,91149	0,91309	0,91466	0,91621	0,91774
1,4	0,91924	0,92073	0,9222	0,92364	0,92507	0,92647	0,92785	0,92922	0,93056	0,93189
1,5	0,93319	0,93448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408
1,6	0,9452	0,9463	0,94738	0,94845	0,9495	0,95053	0,95154	0,95254	0,95352	0,95449
1,7	0,95543	0,95637	0,95728	0,95818	0,95907	0,95994	0,9608	0,96164	0,96246	0,96327
1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062
1,9	0,97128	0,97193	0,97257	0,9732	0,97381	0,97441	0,975	0,97558	0,97615	0,9767
2	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,9803	0,98077	0,98124	0,98169
2,1	0,98214	0,98257	0,983	0,98341	0,98382	0,98422	0,98461	0,985	0,98537	0,98574
2,2	0,9861	0,98645	0,98679	0,98713	0,98745	0,98778	0,98809	0,9884	0,9887	0,98899
2,3	0,98928	0,98956	0,98983	0,9901	0,99036	0,99061	0,99086	0,99111	0,99134	0,99158
2,4	0,9918	0,99202	0,99224	0,99245	0,99266	0,99286	0,99305	0,99324	0,99343	0,99361
2,5	0,99379	0,99396	0,99413	0,9943	0,99446	0,99461	0,99477	0,99492	0,99506	0,9952
2,6	0,99534	0,99547	0,9956	0,99573	0,99585	0,99598	0,99609	0,99621	0,99632	0,99643
2,7	0,99653	0,99664	0,99674	0,99683	0,99693	0,99702	0,99711	0,9972	0,99728	0,99736
2,8	0,99744	0,99752	0,9976	0,99767	0,99774	0,99781	0,99788	0,99795	0,99801	0,99807
2,9	0,99813	0,99819	0,99825	0,99831	0,99836	0,99841	0,99846	0,99851	0,99856	0,99861
3	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99896	0,999
3,1	0,99903	0,99906	0,9991	0,99913	0,99916	0,99918	0,99921	0,99924	0,99926	0,99929
3,2	0,99931	0,99934	0,99936	0,99938	0,9994	0,99942	0,99944	0,99946	0,99948	0,9995
3,3	0,99952	0,99953	0,99955	0,99957	0,99958	0,9996	0,99961	0,99962	0,99964	0,99965
3,4	0,99966	0,99968	0,99969	0,9997	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976
3,5	0,99977	0,99978	0,99978	0,99979	0,9998	0,99981	0,99981	0,99982	0,99983	0,99983
3,6	0,99984	0,99985	0,99985	0,99986	0,99986	0,99987	0,99987	0,99988	0,99988	0,99989
3,7	0,99989	0,9999	0,9999	0,9999	0,99991	0,99991	0,99992	0,99992	0,99992	0,99992
3,8	0,99993	0,99993	0,99993	0,99994	0,99994	0,99994	0,99994	0,99995	0,99995	0,99995
3,9	0,99995	0,99995	0,99996	0,99996	0,99996	0,99996	0,99996	0,99997	0,99997	0,99997

LOI NORMALE ET CALCULATRICE

On suppose ici que l'on dispose d'une variable X suivant une loi normale $\mathcal{N}(\mu, \sigma^2)$, de fonction de répartition F .

Calculs de probabilités :

$$\text{Calculer } P(a < X \leq b) = F(b) - F(a)$$

Avec une TI

- Sélectionner le menu des lois de probabilités en tapant 2nd + DISTR.
- Sélectionner `normalcdf` ou `normalFRep` suivant les modèles
- Compléter les paramètres



Remarque :

À ne pas confondre avec `normalpdf` qui donne les valeurs de la fonction densité et non celles de la fonction de répartition.

Avec une Casio

- Aller dans le menu STAT
- En bas de l'écran, sélectionner DIST puis NORM et enfin Ncd
- Compléter les paramètres



Remarque :

À ne pas confondre avec `Npd` qui donne les valeurs de la fonction densité et non celles de la fonction de répartition.

Avec une NumWorks

- Aller dans le menu Probabilités
- Choisir la loi normale
- Compléter les paramètres $\mu = \dots, \sigma = \dots$
- En haut à gauche, sélectionner le type d'inégalité $P(.. \leq X \leq \dots)$
- compléter les paramètres de gauche et de droite dans $P(.. \leq X \leq ..)$

? Exercice 6

Calculer $P(-2 \leq X \leq 3)$ pour $X \leftrightarrow \mathcal{N}(1, 2)$. ($\simeq 0,9044$)

$$\text{Calculer } P(X \leq b) = F(b)$$

Avec une TI ou une Casio

L'idée est de remplacer $P(X \leq b) = F(b)$ par $P(a < x \leq b) = F(b) - F(a)$ avec a très petit, de manière à avoir $F(a) \simeq 0$ négligeable.

- ★ Pour la loi $\mathcal{N}(0, 1)$, on peut prendre $a = -4$, i.e. $P(X \leq b) \simeq P(-4 < X \leq b)$
- ★ Pour les autres lois, ne pas hésiter à prendre par exemple $a = -10^{50}$: $P(X \leq b) \simeq P(-10^{50} < X \leq b)$

Avec une NumWorks

- Aller dans le menu Probabilités
- Choisir la loi normale
- Compléter les paramètres $\mu = \dots, \sigma = \dots$
- En haut à gauche, sélectionner si nécessaire le type d'inégalité $P(X \leq \dots)$
- compléter le paramètre de droite dans $P(X \leq ..)$

? Exercice 7

Calculer $P(X \leq 3)$ pour $X \leftrightarrow \mathcal{N}(1, 2)$. ($\simeq 0.92135$)

Et quand ce sont les probabilités qui sont données :

Calculer z tel que $P(X \leq z) = \alpha$, avec α fixé.

On remarque qu'il s'agit ici d'inverser la fonction de répartition F car on cherche $z = F^{-1}(\alpha)$.

Avec une TI

- Sélectionner le menu des lois de probabilités en tapant 2nd + DISTR.
- Sélectionner `invNorm` ou `FracNormale` suivant les modèles
- Compléter les paramètres

Avec une Casio

- Aller dans le menu STAT
- En bas de l'écran, sélectionner DIST puis NORM et enfin InvN
- Compléter les paramètres

Avec une NumWorks

- Aller dans le menu **Probabilités**
- Choisir la loi normale
- Compléter les paramètres $\mu = \dots$, $\sigma = \dots$
- En haut à gauche, sélectionner si nécessaire le type d'inégalité $P(X \leq \dots)$
- compléter la probabilité z dans $P(X \leq ..) = z$

? Exercice 8

Pour $X \hookrightarrow \mathcal{N}(0, 1)$, déterminer a tel que $P(X \leq a) = 0.7$. ($a \simeq 0,52$)

Calculer z tel que $P(|X - \mu| \leq z) = \alpha$, avec α fixé.

Avec toutes les calculatrices, on peut par exemple utiliser le fait que

$$Z = \frac{X - \mu}{\sigma} \hookrightarrow \mathcal{N}(0, 1)$$

Ainsi,

$$P(|X - \mu| \leq z) = \alpha \Leftrightarrow P\left(\frac{|X - \mu|}{\sigma} \leq \frac{z}{\sigma}\right) \Leftrightarrow P\left(|Z| \leq \frac{z}{\sigma}\right) = \alpha \Leftrightarrow 2\Phi\left(\frac{z}{\sigma}\right) - 1 = \alpha \Leftrightarrow \Phi\left(\frac{z}{\sigma}\right) = \frac{\alpha + 1}{2}$$

Ainsi, **avec toutes les calculatrices**, on peut chercher $a = \frac{z}{\sigma}$ tel que $P(Z \leq a) = \frac{\alpha + 1}{2}$.

Sinon, certaines calculatrices disposent de la version "centrée" qui permettent d'obtenir directement le résultat :

Il peut par exemple s'agir de sélectionner le type d'inégalité $P(.. \leq X \leq ..) = ..$ et de remplir la probabilité. On peut obtenir directement le z ou alors plutôt

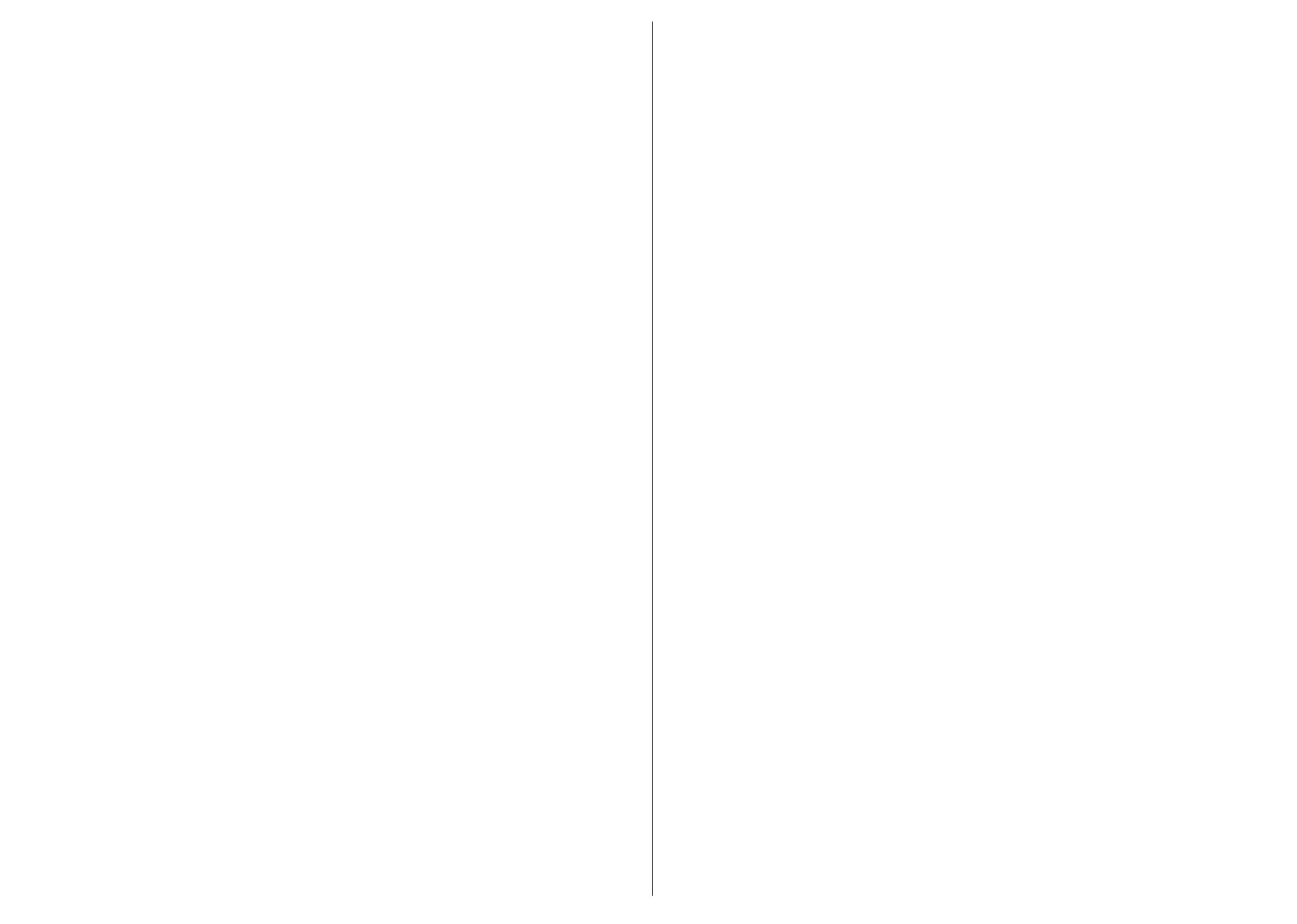
$$a = \mu - z, b = \mu + z \text{ tels que } P(a \leq X \leq b) = \alpha$$

? Exercice 9

Pour $X \hookrightarrow \mathcal{N}(0, 1)$, déterminer b tel que $P(|X| \leq b) = 0.7$. ($b \simeq 1,036$)

? Exercice 10

Pour $X \hookrightarrow \mathcal{N}(1, 4)$, déterminer c tel que $P(|X - 1| \leq c) = 0.9$. ($c \simeq 3.2897$, ce qui donnerait $P(-2.2897 \leq X \leq 4.2897) \simeq 0.9$.)



Commentaires :

On connaissait jusque là (en première année) le produit scalaire dans \mathbb{R}^2 ou \mathbb{R}^3 défini de la manière suivante par exemple si $X = (x_1, x_2), Y = (y_1, y_2) \in \mathbb{R}^2$:

$$\langle X, Y \rangle = x_1y_1 + x_2y_2$$

Son utilité était principalement (surtout dans \mathbb{R}^2) d'établir des égalités d'angles géométrique ou des orthogonalités entre deux vecteurs. On rappelle à cet effet qu'on avait

$$\langle X, Y \rangle = 0 \iff X \perp Y$$

On peut en fait généraliser cette notions à l'espace vectoriel \mathbb{R}^n . De plus, on (re)verra que ce produit scalaire peut également nous permettre :

- de faire des calculs de distance
- de diagonaliser certaines matrices dans des bases "intéressantes"
- d'inverser de manière immédiate certaines matrices
- de faire des projections orthogonales sur des sous-espaces vectoriels de \mathbb{R}^n .

I Généralités

Dans tout ce chapitre, on pose $n \in \mathbb{N}^*$ et notera \mathcal{B}_0 la base canonique de \mathbb{R}^n . Commençons par généraliser la notion de produit scalaire :

I-1 Définitions et propriétés élémentaires

Définition

Pour $u, v \in \mathbb{R}^n$ tels que $M_{\mathcal{B}_0}(u) = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, M_{\mathcal{B}_0}(v) = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n$, on appelle *produit scalaire (usuel) de u et v* le nombre

$$\langle u, v \rangle = x_1y_1 + \dots + x_ny_n$$

Commentaires :

En réalité, il existe beaucoup d'autres "produits scalaires" sur \mathbb{R}^n , d'où le nom ici rajouté de "usuel". (Ces autres produits scalaires sont notamment étudiés dans d'autres classes. Attention donc à ne pas s'éparpiller en allant voir inutilement les cours dans d'autres sections !)

Le produit scalaire usuel (celui que nous étudions) est celui utilisé le plus fréquemment dans les calculs géométriques "classiques". Étant donné que ce sera le seul étudié cette année, on pourra dans ce chapitre oublier le terme "usuel".

■ Exemple 1 :

Dans \mathbb{R}^4 , Si $u = (\underbrace{1}_{x_1}, \underbrace{0}_{x_2}, \underbrace{-1}_{x_3}, \underbrace{2}_{x_4})$ et $v = (\underbrace{2}_{y_1}, \underbrace{3}_{y_2}, \underbrace{-1}_{y_3}, \underbrace{1}_{y_4})$, on a

$$\langle u, v \rangle = 1 \times 2 + 0 \times 3 + (-1) \times (-1) + 2 \times 1 = 5$$

⚠ Remarque :

Avec les notations de la définition ci-avant, remarquons que la formule du produit scalaire " $x_1y_1 + \dots + x_ny_n$ " est également issue du **produit matriciel**

$$\underbrace{\begin{pmatrix} x_1 & \dots & x_n \end{pmatrix}}_{\text{une seule ligne, } n \text{ colonnes}} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = (x_1y_1 + \dots + x_ny_n) = \underbrace{\langle u, v \rangle}_{\text{matrice à une ligne, une colonne}}$$

Le résultat est donc une matrice à une ligne et une colonne contenant $\langle u, v \rangle$.

■ Exemple 2 :

Dans \mathbb{R}^4 , Si $u = (\underbrace{1}_{x_1}, \underbrace{0}_{x_2}, \underbrace{-1}_{x_3}, \underbrace{2}_{x_4})$ et $v = (\underbrace{2}_{y_1}, \underbrace{3}_{y_2}, \underbrace{-1}_{y_3}, \underbrace{1}_{y_4})$, on a

$$\langle u, v \rangle = (1 \ 0 \ -1 \ 2) \begin{pmatrix} 2 \\ 3 \\ -1 \\ 1 \end{pmatrix} = (1 \times 2 + 0 \times 3 + (-1) \times (-1) + 2 \times 1) = (5)$$

⚠ Remarque :

Le résultat du produit matriciel ne contenant qu'un seul élément, on fait l'amalgame entre "matrice" et "nombre" (et donc l'abus de notation) suivant :

$$\langle u, v \rangle = \langle u, v \rangle$$

D'où

$$\begin{pmatrix} x_1 & \dots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \langle u, v \rangle$$

En notant $U = M_{\mathcal{B}_0}(u), V = M_{\mathcal{B}_0}(v)$, la formule devient donc :

$${}^tUV = \langle u, v \rangle$$

où tU est le vecteur "transposé" de $U : {}^tU = (x_1 \ \dots \ x_n)$. (voir ex. ci-dessous)

■ Exemple 3 :

Dans \mathbb{R}^4 , Si $u = (\underbrace{1}_{x_1}, \underbrace{0}_{x_2}, \underbrace{-1}_{x_3}, \underbrace{2}_{x_4})$ et $v = (\underbrace{2}_{y_1}, \underbrace{3}_{y_2}, \underbrace{-1}_{y_3}, \underbrace{1}_{y_4})$, on a

$$U = \begin{pmatrix} 1 \\ 0 \\ -1 \\ 2 \end{pmatrix}, V = \begin{pmatrix} 2 \\ 3 \\ -1 \\ 1 \end{pmatrix}$$

D'où

$$\langle u, v \rangle = {}^tUV = (1 \ 0 \ -1 \ 2) \begin{pmatrix} 2 \\ 3 \\ -1 \\ 1 \end{pmatrix} = 5$$

? Exercice 1

Pour $u = (0, 3, -1) \in \mathbb{R}^3$, $v = (1, -1, 1) \in \mathbb{R}^3$, calculer $\langle u, v \rangle$ en passant par l'écriture matricielle.

Solution

$$\text{On a } \langle u, v \rangle = (0 \ 3 \ -1) \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} = -4$$

Commentaires :

Voyons maintenant les quelques propriétés principales du produit scalaire, qui vont nous servir tout au long du chapitre. Elles sont donc à maîtriser.

Proposition 167

Le produit scalaire est :

- bilinéaire ; au sens où, pour tout $u, v, w \in \mathbb{R}^n$, $\alpha \in \mathbb{R}$,

$$\langle u + \alpha v, w \rangle = \langle u, w \rangle + \alpha \langle v, w \rangle$$

et

$$\langle u, v + \alpha w \rangle = \langle u, v \rangle + \alpha \langle u, w \rangle .$$

- symétrique ; au sens où

$$\langle u, v \rangle = \langle v, u \rangle \quad \forall u, v \in \mathbb{R}^n$$

- défini positif ; au sens où

$$\langle u, u \rangle \geq 0 \quad \forall u \in \mathbb{R}^n$$

et

$$\langle u, u \rangle = 0 \Leftrightarrow u = 0$$

Démonstration :

- Montrons que le produit scalaire est bilinéaire :

Soient $u, v, w \in \mathbb{R}^n$, $\alpha \in \mathbb{R}$. On notera les coordonnées de la même manière que dans la définition. Dans ce cas, on a

$$\begin{aligned} \langle u + \alpha v, w \rangle &= {}^t(U + \alpha V) \cdot W \\ &= ({}^tU + \alpha {}^tV) \cdot W \\ &= {}^tU \cdot W + \alpha {}^tV \cdot W = \langle u, w \rangle + \alpha \langle v, w \rangle \end{aligned}$$

De même pour la deuxième égalité.

- Montrons que le produit scalaire est symétrique :

Clairement, si on échange u et v dans la formule de la définition on trouve exactement le même résultat. C'est donc bien symétrique.

- Montrons que le produit scalaire est défini positif :

$$\langle u, u \rangle = x_1^2 + \dots + x_n^2 \geq 0$$

De plus,

$$\langle u, u \rangle = 0 \Leftrightarrow x_1^2 + \dots + x_n^2 = 0$$

Or, comme le terme de gauche est une somme de nombres positifs, la seule manière d'obtenir 0 est que chaque terme soit nul. Autrement dit,

$$x_1 = \dots = x_n = 0$$

et donc

$$u = 0$$

□

⚠ Remarque :

Le terme bilinéaire vient du fait que le produit scalaire est linéaire par-rapport à chaque variable. Le comportement est celui d'un développement de parenthèses dans un produit.

? Exercice 2

Dans \mathbb{R}^n , sachant que $\|u\|^2 = 1$, $\langle u, v \rangle = 1$, $\langle v, w \rangle = -1$ et $\langle u, w \rangle = -1$, vérifier que $\langle u + w, 2v \rangle = 0$ ainsi que $\langle u + w, u + v \rangle = 0$.

Solution

$$\begin{aligned} \langle u + w, 2v \rangle &= \langle u, 2v \rangle + \langle w, 2v \rangle = 2 \langle u, v \rangle + 2 \langle w, v \rangle = 2 - 2 = 0 \\ \langle u + w, u + v \rangle &= \langle u, u \rangle + \langle u, v \rangle + \langle w, u \rangle + \langle w, v \rangle = 1 + 1 - 1 - 1 = 0 \end{aligned}$$

Définition
 On appelle *norme euclidienne* d'un vecteur u le nombre

$$\|u\| = \sqrt{\langle u, u \rangle}$$

Remarque :
 Dans la base canonique, la formule correspond à

$$\|u\| = \sqrt{\langle u, u \rangle} = \sqrt{x_1^2 + \dots + x_n^2}$$

où $u = (x_1, \dots, x_n)$

Commentaires :
 Dans \mathbb{R}^2 et \mathbb{R}^3 , on sait déjà que ceci correspond en fait à la longueur du vecteur u . C'est le même principe dans \mathbb{R}^n . Ainsi, les propriétés déjà vraies dans \mathbb{R}^2 vont se généraliser dans \mathbb{R}^n .

Propriété 168

Si $u \in \mathbb{R}^n$, on a

$$\|u\| = 0 \iff u = 0$$

Théorème 169 : Inégalité de Cauchy-Schwarz

Soient $u, v \in \mathbb{R}^n$. On a

$$|\langle u, v \rangle| \leq \|u\| \|v\|$$

avec

$$|\langle u, v \rangle| = \|u\| \|v\| \iff (u, v) \text{ proportionnels}$$

Démonstration :
 Si $v = 0$, alors l'inégalité est triviale (et c'est même une égalité) :

$$\langle u, v \rangle = 0 = \|u\| \times 0 = \|u\| \|v\|$$

Si $v \neq 0$, on pose $t \in \mathbb{R}$. On a

$$\begin{aligned} 0 \leq \|u + tv\|^2 &= \langle u + tv, u + tv \rangle \\ &= \langle u, u \rangle + t \langle u, v \rangle + t \langle v, u \rangle + t^2 \langle v, v \rangle \\ &= \|u\|^2 + 2t \langle u, v \rangle + \|v\|^2 t^2 \end{aligned}$$

L'expression obtenue est donc un polynôme en t , toujours positif, de coefficient dominant positif. Ainsi, ce polynôme n'a d'autre choix que d'avoir un discriminant négatif.

Or,

$$\begin{aligned} 0 \geq \Delta &= (2 \langle u, v \rangle)^2 - 4 \|u\|^2 \|v\|^2 \\ &= 4 (\langle u, v \rangle^2 - \|u\|^2 \|v\|^2) \end{aligned}$$

On en déduit $\langle u, v \rangle^2 \leq \|u\|^2 \|v\|^2$ et donc

$$|\langle u, v \rangle| \leq \|u\| \|v\|$$

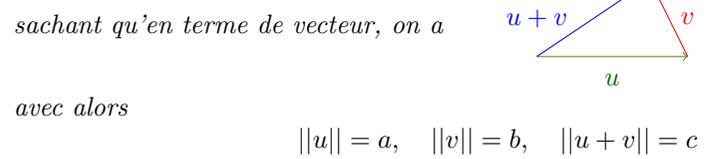
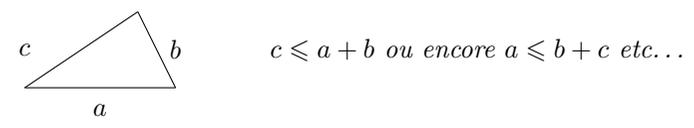
En remontant le raisonnement, on a de plus,

$$\begin{aligned} |\langle u, v \rangle| = \|u\| \|v\| &\iff \Delta = 0 \\ &\iff \text{il existe } t_0 \text{ tel que } \|u\|^2 + 2t_0 \langle u, v \rangle + \|v\|^2 t_0^2 = 0 \\ &\iff \text{il existe } t_0 \text{ tel que } \|u + t_0 v\|^2 = 0 \\ &\iff \text{il existe } t_0 \text{ tel que } u = -t_0 v \end{aligned}$$

□

Commentaires :

L'inégalité suivante prise dans \mathbb{R}^2 traduit (et justifie) en particulier le fait que dans un triangle, la somme des longueurs de deux des côtés est nécessairement supérieure à la longueur du troisième :



De manière générale :

Théorème 170 : Inégalité triangulaire

Soient $u, v \in \mathbb{R}^n$. On a

$$\|u + v\| \leq \|u\| + \|v\|$$

avec égalité si et seulement si u et v sont proportionnels dans le même sens.

Démonstration :
 Étant donné que les deux côtés de l'inégalité sont positifs on peut passer au carré :

$$\begin{aligned} \|u + v\|^2 &= \langle u + v, u + v \rangle \\ &= \langle u, u \rangle + \langle u, v \rangle + \langle v, u \rangle + \langle v, v \rangle \\ &= \|u\|^2 + \|v\|^2 + 2 \langle u, v \rangle \\ &= (\|u\| + \|v\|)^2 - 2\|u\| \|v\| + 2 \langle u, v \rangle \end{aligned}$$

Or, par inégalité de Cauchy-Schwarz, on sait que

$$|\langle u, v \rangle| \leq \|u\| \cdot \|v\|$$

d'où

$$\|u + v\|^2 \leq (\|u\| + \|v\|)^2$$

$\|u + v\|$ et $\|u\| + \|v\|$ étant positifs, on en déduit l'inégalité triangulaire.

On a de plus égalité si et seulement si $\langle u, v \rangle = \|u\| \cdot \|v\|$.

Ainsi, d'après CS, on a u et v proportionnels. Si u et v sont dans le sens opposé (et non nuls), le calcul montre que

$$\langle u, v \rangle < 0 \quad \text{et donc} \quad \langle u, v \rangle \neq \|u\| \cdot \|v\|$$

□

Commentaires :

De la démonstration du théorème précédent, on tire également que la norme de la somme de deux vecteurs peut se calculer de la manière suivante :

Propriété 171

Soient $u, v \in \mathbb{R}^n$. On a

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2 + 2\langle u, v \rangle$$

Cette propriété peut également (dans de rares cas, si on connaît la longueur des trois côtés,) servir à calculer un produit scalaire sans disposer des coordonnées :

$$\langle u, v \rangle = \frac{1}{2} (\|u + v\|^2 - \|u\|^2 - \|v\|^2)$$

? Exercice 3

Soit dans \mathbb{R}^2 un triangle ABC rectangle en A tel que $\|AC\| = 2$, $\|AB\| = 1$. Calculer $\langle \overrightarrow{CA}, \overrightarrow{CB} \rangle$.

(On remarquera que l'on sait déterminer la longueur de \overrightarrow{CB} . On prendra soin de choisir correctement ses vecteurs " u, v " afin de pouvoir appliquer la formule précédente!)

Vous devez trouver le résultat $\langle \overrightarrow{CA}, \overrightarrow{CB} \rangle = 4$.

Solution

On a $\langle \overrightarrow{CA}, \overrightarrow{CB} \rangle = -\langle \overrightarrow{AC}, \overrightarrow{CB} \rangle = -\frac{1}{2} \left(\|\underbrace{\overrightarrow{AC} + \overrightarrow{CB}}_{\overrightarrow{AB}}\|^2 - \|\overrightarrow{AC}\|^2 - \|\overrightarrow{CB}\|^2 \right)$ avec,

d'après la formule de Pythagore : $\|\overrightarrow{CB}\|^2 = \|\overrightarrow{AB}\|^2 + \|\overrightarrow{AC}\|^2$

D'où $\langle \overrightarrow{CA}, \overrightarrow{CB} \rangle = -\frac{1}{2} (-2\|\overrightarrow{AC}\|^2) = 4$

I-2 Changement de base

Commentaires :

Et si les vecteurs u et v ne sont pas donnés dans la base canonique mais dans une base autre \mathcal{B} , comment calculer $\langle u, v \rangle$? A-t-on encore le droit d'utiliser la même formule?

i.e. Si $M_{\mathcal{B}}(u) = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$, $M_{\mathcal{B}_0}(v) = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} \in \mathbb{R}^n$, comment calculer $\langle u, v \rangle$?

? Exercice 4

On se donne une base $\mathcal{B} = (\underbrace{(1, 1)}_{v_1}, \underbrace{(-1, 2)}_{v_2})$ une base de \mathbb{R}^2 ainsi que u, v exprimés dans

la base \mathcal{B} tels que $u = (-1, 1)_{\mathcal{B}}$ et $v = (a, b)_{\mathcal{B}}$. Se ramener à la base canonique pour calculer $\langle u, v \rangle$

(Vérifiez que vous trouvez $\langle u, v \rangle = 4b - a$ et non $-a + b$).

Solution

Option 1 : Une possibilité est d'exprimer u, v dans la base canonique grâce à la décomposition dans la base :

On a $u = -v_1 + v_2 = (-2, 1)$ ainsi que $v = av_1 + bv_2 = (a - b, a + 2b)$. Maintenant que les vecteurs sont exprimés dans la base canonique, on peut calculer $\langle u, v \rangle = -2 \times (a - b) + 1 \times (a + 2b) = -a + 4b$.

Option 2 : une autre possibilité est d'utiliser les matrices de passage :

La matrice de passage de \mathcal{B}_0 à \mathcal{B} est $P = M_{\mathcal{B}_0}(\mathcal{B}) = \begin{pmatrix} 1 & -1 \\ 1 & 2 \end{pmatrix}$

Ainsi,

$$M_{\mathcal{B}_0}(u) = P \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -2 \\ 1 \end{pmatrix} \quad \text{et} \quad M_{\mathcal{B}_0}(v) = P \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a - b \\ a + 2b \end{pmatrix}$$

D'où

$$\langle u, v \rangle = -2 \times (a - b) + 1 \times (a + 2b) = 4b - a$$

Commentaires :

On constate que la formule $\alpha_1\beta_1 + \dots + \alpha_n\beta_n$ n'est donc pas forcément valide pour calculer le produit scalaire quand on change de base pour exprimer les coordonnées! Oups... Cherchons alors, une formule valable dans toutes les bases!

Commençons par observer tout ceci de manière matricielle :

Si les vecteurs sont donnés dans une base nommée \mathcal{B} , on peut se ramener à la base canonique grâce à la formule classique de changement de base :

$$M_{\mathcal{B}_0}(u) = M_{\mathcal{B}_0}(\mathcal{B})M_{\mathcal{B}}(u), \quad M_{\mathcal{B}_0}(v) = M_{\mathcal{B}_0}(\mathcal{B})M_{\mathcal{B}}(v)$$

Avec les vecteurs obtenus dans la base canonique, on fait le calcul. De plus, $M_{\mathcal{B}_0}(\mathcal{B})$ est en général une matrice facilement identifiable : on n'a pas à inverser.

■ Exemple 4 :

Dans l'exercice ci-dessus : on aurait $M_{\mathcal{B}_0}(\mathcal{B}) = \begin{pmatrix} 1 & -1 \\ 1 & 2 \end{pmatrix}$

et on aurait alors

$$M_{\mathcal{B}_0}(u) = M_{\mathcal{B}_0}(\mathcal{B})M_{\mathcal{B}}(u) = \begin{pmatrix} 1 & -1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$$

$$M_{\mathcal{B}_0}(v) = M_{\mathcal{B}_0}(\mathcal{B})M_{\mathcal{B}}(v) = \begin{pmatrix} 1 & -1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a-b \\ a+2b \end{pmatrix}$$

et donc

$$\langle u, v \rangle = \begin{pmatrix} -2 & 1 \end{pmatrix} \begin{pmatrix} a-b \\ a+2b \end{pmatrix} = 4b - a$$

Notons que pour ceci, on peut également utiliser la bilinéarité du produit scalaire :

Par exemple, dans l'exercice ci-dessus : $u = -v_1 + v_2$, $v = av_1 + bv_2$ alors

$$\begin{aligned} \langle u, v \rangle &= \langle -v_1 + v_2, av_1 + bv_2 \rangle \\ &= -\langle v_1, av_1 + bv_2 \rangle + \langle v_2, av_1 + bv_2 \rangle \quad (\text{linéarité à gauche}) \\ &= -(a\langle v_1, v_1 \rangle + b\langle v_1, v_2 \rangle) + (a\langle v_2, v_1 \rangle + b\langle v_2, v_2 \rangle) \quad (\text{lin. à droite}) \\ &= -a\underbrace{\langle v_1, v_1 \rangle}_{=2} + (-b+a)\underbrace{\langle v_1, v_2 \rangle}_{=1} + b\underbrace{\langle v_2, v_2 \rangle}_{=5} \\ &= -2a + (a-b) + 5b = 4b - a \end{aligned}$$

? Exercice 5

Dans \mathbb{R}^3 , on considère la base $\mathcal{B} = (v_1, v_2, v_3) = \left(\begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix} \right)$ donnés dans la base canonique. Calculer $\langle u, v \rangle$ où $M_{\mathcal{B}}(u) = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}$ et $M_{\mathcal{B}}(v) = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$ à l'aide de la bilinéarité du produit scalaire.

(Vous devez trouver $\langle u, v \rangle = 0$.)

Solution

En écrivant

$$u = V_1 - V_2 + 2V_3 \quad \text{et} \quad v = -V_1 + V_3$$

on trouve, par bilinéarité du produit scalaire,

$$\begin{aligned} \langle u, v \rangle &= \langle V_1 - V_2 + 2V_3, v \rangle \\ &= \langle V_1, v \rangle - \langle V_2, v \rangle + 2\langle V_3, v \rangle \quad (\text{linéarité à gauche}) \\ &= (-\langle V_1, V_1 \rangle + \langle V_1, V_3 \rangle) \quad (\text{linéarité à droite avec } v = -V_1 + V_3) \\ &\quad - (-\langle V_2, V_1 \rangle + \langle V_2, V_3 \rangle) \\ &\quad + 2(-\langle V_3, V_1 \rangle + \langle V_3, V_3 \rangle) \end{aligned}$$

En calculant les produits scalaires deux à deux des vecteurs de la base, on peut conclure.

On a :

$$\begin{cases} \langle V_1, V_1 \rangle = 6 \\ \langle V_1, V_2 \rangle = -3 \\ \langle V_1, V_3 \rangle = 5 \end{cases} \quad \begin{cases} \langle V_2, V_3 \rangle = -2 \\ \langle V_3, V_3 \rangle = 6 \end{cases}$$

D'où

$$\langle u, v \rangle = 0$$

Commentaires :

Ces deux méthodes permettront toutes deux d'arriver à même une formule générale utilisant une matrice que nous allons d'abord étudier à part.

🌿 Définition

Étant donné une base $\mathcal{B} = (v_1, \dots, v_n)$ de \mathbb{R}^n . On appelle *matrice de Gram de la base \mathcal{B}* , la matrice des "produits scalaires de \mathcal{B} " ci-dessous

$$G_{\mathcal{B}} = \begin{pmatrix} \langle v_1, v_1 \rangle & \langle v_1, v_2 \rangle & \dots & \langle v_1, v_n \rangle \\ \langle v_2, v_1 \rangle & \langle v_2, v_2 \rangle & & \langle v_2, v_n \rangle \\ \vdots & & & \vdots \\ \langle v_n, v_1 \rangle & \dots & \dots & \langle v_n, v_n \rangle \end{pmatrix} = (\langle v_i, v_j \rangle)_{1 \leq i, j \leq n}$$

■ Exemple 5 :

Pour la base de l'exemple précédent $\mathcal{B} = (\underbrace{(1, 1)}_{v_1}, \underbrace{(-1, 2)}_{v_2})$ on vu dans les calculs que

$$G_{\mathcal{B}} = \begin{pmatrix} 2 & 1 \\ 1 & 5 \end{pmatrix}$$

? Exercice 6

Calculer la matrice de Gram dans l'exercice précédent pour $\mathcal{B} = (v_1, v_2, v_3) = \left(\begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix} \right)$

Solution

Des calculs de l'exercice précédent, on tire :

$$G_{\mathcal{B}} = \begin{pmatrix} 6 & -3 & 5 \\ -3 & 2 & -2 \\ 5 & -2 & 6 \end{pmatrix}$$

où il resterait à calculer $\langle v_2, v_2 \rangle$ (qui vaut $\langle v_2, v_2 \rangle = 2$). On peut donc compléter :

$$G_{\mathcal{B}} = \begin{pmatrix} 6 & -3 & 5 \\ -3 & 2 & -2 \\ 5 & -2 & 6 \end{pmatrix}$$

Commentaires :

Reprenons l'exemple précédent et voyons grâce à la bilinéarité à quoi peut bien servir cette matrice, mis à part à faire un beau tableau résumé des données !

Revenons à notre changement de base en méthode matricielle :

Pour des vecteurs donnés dans une base \mathcal{B} , on avait revu les formules :

$$M_{\mathcal{B}_0}(u) = M_{\mathcal{B}_0}(\mathcal{B})M_{\mathcal{B}}(u), \quad M_{\mathcal{B}_0}(v) = M_{\mathcal{B}_0}(\mathcal{B})M_{\mathcal{B}}(v)$$

Pour simplifier, on note :

$$U_0 = M_{\mathcal{B}_0}(u), \quad U = M_{\mathcal{B}}(u)$$

les coordonnées respectives de u dans les bases canoniques et la base \mathcal{B} , (de même pour v), puis

$$P = M_{\mathcal{B}_0}(\mathcal{B}) \quad \text{la matrice de passage}$$

Les formules deviennent

$$U_0 = PU, \quad V_0 = PV$$

Rappelons maintenant la formule matricielle du produit scalaire dans la base canonique :

$$\langle u, v \rangle = {}^tU_0V_0$$

Combinons maintenant les deux en remplaçant $U_0 = PU$ et $V_0 = PV$ dans la ligne ci-dessus :

$$\langle u, v \rangle = {}^t(PU)PV = {}^tU{}^tPPV$$

En posant $G = {}^tPP$, on obtient donc une formule sous forme de produit matriciel

$$\langle u, v \rangle = {}^t(PU)PV = {}^tUGV$$

que nous allons tester sur un exemple !

■ Exemple 6 :

Dans l'exemple déjà évoqué dans cette section, avec $\mathcal{B} = (\underbrace{(1, 1)}_{v_1}, \underbrace{(-1, 2)}_{v_2})$, ainsi que $u = (-1, 1)_{\mathcal{B}}$ et $v = (a, b)_{\mathcal{B}}$. Pour $\langle u, v \rangle$, on peut donc procéder de la manière suivante : On pose la matrice de changement de base

$$P = M_{\mathcal{B}_0}(\mathcal{B}) = \begin{pmatrix} 1 & -1 \\ 1 & 2 \end{pmatrix}$$

de transposée :

$${}^tP = \begin{pmatrix} 1 & 1 \\ -1 & 2 \end{pmatrix}$$

On calcule

$$G = {}^tP.P = \begin{pmatrix} 2 & 1 \\ 1 & 5 \end{pmatrix}$$

Avec $U = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ et $V = \begin{pmatrix} a \\ b \end{pmatrix}$, on a donc

$$\langle u, v \rangle = \begin{pmatrix} -1 & 1 \end{pmatrix} . G . \begin{pmatrix} a \\ b \end{pmatrix} = 4b - a$$

Commentaires :

Mais quel est donc le rapport avec la matrice de Gram $G_{\mathcal{B}}$ évoquée précédemment ?? Et bien, (les notations sont bien faites !!) c'est la même :

$$G_{\mathcal{B}} = {}^tPP$$

Voyons ça :

En effet, si on note V_i la $i^{\text{ème}}$ colonne de P , on sait par définition que

$$V_i = M_{\mathcal{B}_0}(v_i) \quad \text{coordonnées dans la base canonique de } v_i$$

On peut donc d'une part calculer matriciellement $\langle v_i, v_j \rangle$ avec

$$\langle v_i, v_j \rangle = {}^tV_iV_j$$

mais si on calcule le produit tPP , par formule de calcul sur les produits de matrice, tV_iV_j est exactement le coefficient dans la ligne i , colonne j de la matrice tPP . On tombe donc bien sur ${}^tPP = G_{\mathcal{B}}$

Commentaires :

Tout ceci donne donc lieu en général à la formule suivante :

Proposition 172

Si $u, v \in \mathbb{R}^n$, en notant G la matrice de Gram d'une base \mathcal{B} , alors

$$\langle u, v \rangle = {}^tM_{\mathcal{B}}(u) G_{\mathcal{B}} M_{\mathcal{B}}(v)$$

? Exercice 7

Reprenons la base de l'exercice précédent : $\mathcal{B} = \left(\begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix} \right)$.

Calculer $\langle u, v \rangle$ où $M_{\mathcal{B}}(u) = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}$ et $M_{\mathcal{B}}(v) = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$ à l'aide de la méthode matricielle en retrouvant la matrice de Gram à l'aide du calcul $G_{\mathcal{B}} = {}^t P P$.

Solution

On a $\langle u, v \rangle = {}^t M_{\mathcal{B}}(u) G_{\mathcal{B}} M_{\mathcal{B}}(v)$
avec

$$P = \begin{pmatrix} 1 & 0 & 2 \\ 2 & -1 & 1 \\ -1 & 1 & -1 \end{pmatrix} \quad \text{d'où} \quad {}^t P = \begin{pmatrix} 1 & 2 & -1 \\ 0 & -1 & 1 \\ 2 & 1 & -1 \end{pmatrix}$$

et

$$G_{\mathcal{B}} = {}^t P P = \begin{pmatrix} 1 & 2 & -1 \\ 0 & -1 & 1 \\ 2 & 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 2 \\ 2 & -1 & 1 \\ -1 & 1 & -1 \end{pmatrix} = \begin{pmatrix} 6 & -3 & 5 \\ -3 & 2 & -2 \\ 5 & -2 & 6 \end{pmatrix}$$

et donc

$$\langle u, v \rangle = (1 \quad -1 \quad 2) \begin{pmatrix} 6 & -3 & 5 \\ -3 & 2 & -2 \\ 5 & -2 & 6 \end{pmatrix} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = 0$$

Remarquons que tout ceci se retrouve par bilinéarité sur un exemple :

■ Exemple 7 :

On avait $\mathcal{B} = (\underbrace{(1, 1)}_{v_1}, \underbrace{(-1, 2)}_{v_2})$ ainsi que $M_{\mathcal{B}}(u) = (-1, 1)$ et $M_{\mathcal{B}}(v) = (a, b)$, c'est-à-dire :

$$u = -v_1 + v_2, \quad v = av_1 + bv_2$$

avec $\langle u, v \rangle = \langle -v_1 + v_2, v \rangle = -\langle v_1, v \rangle + \langle v_2, v \rangle$ (linéarité à gauche)

Matriciellement on observe alors que c'est le même calcul que

$$\langle u, v \rangle = (-1 \quad 1) \begin{pmatrix} \langle v_1, v \rangle \\ \langle v_2, v \rangle \end{pmatrix}$$

Maintenant, par linéarité à droite sur chaque terme :

$$\begin{aligned} \begin{pmatrix} \langle v_1, v \rangle \\ \langle v_2, v \rangle \end{pmatrix} &= \begin{pmatrix} \langle v_1, av_1 + bv_2 \rangle \\ \langle v_2, av_1 + bv_2 \rangle \end{pmatrix} \\ &= \begin{pmatrix} a \langle v_1, v_1 \rangle + b \langle v_1, v_2 \rangle \\ a \langle v_2, v_1 \rangle + b \langle v_2, v_2 \rangle \end{pmatrix} = \underbrace{\begin{pmatrix} \langle v_1, v_1 \rangle & \langle v_1, v_2 \rangle \\ \langle v_2, v_1 \rangle & \langle v_2, v_2 \rangle \end{pmatrix}}_{=G} \begin{pmatrix} a \\ b \end{pmatrix} \end{aligned}$$

Mis bout à bout, cela donne : $\langle u, v \rangle = (-1 \quad 1) G \begin{pmatrix} a \\ b \end{pmatrix}$

II Orthogonalité

II-1 Définitions et propriétés élémentaires

📖 Définition

Deux vecteurs $u, v \in \mathbb{R}^n$ sont dits *orthogonaux* si $\langle u, v \rangle = 0$. On note $u \perp v$

Théorème 173 de Pythagore :

Soient $u, v \in \mathbb{R}^n$. On a

$$u \perp v \iff \|u + v\|^2 = \|u\|^2 + \|v\|^2$$

Démonstration :

Da la somme de deux vecteurs, on tire l'égalité :

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2 + 2 \langle u, v \rangle$$

Ainsi,

$$\begin{aligned} u \perp v &\Leftrightarrow \langle u, v \rangle = 0 \\ &\Leftrightarrow \|u + v\|^2 = \|u\|^2 + \|v\|^2 \end{aligned}$$

□

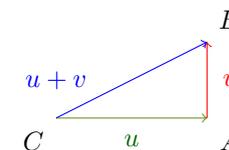
⚠️ Remarque :

Ce théorème peut se traduire également de la manière bien connue suivante : Soient $A, B, C \in \mathbb{R}^n$. On a

$$\overrightarrow{AB} \perp \overrightarrow{AC} \iff \|\overrightarrow{BC}\|^2 = \|\overrightarrow{AB}\|^2 + \|\overrightarrow{AC}\|^2$$

C'est-à-dire dans \mathbb{R}^2 ou \mathbb{R}^3 :

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2$$



📖 Définition

Une famille $(v_1, \dots, v_s) \in \mathbb{R}^n$ est appelée *base orthogonale* de l'e.v. $F \subset \mathbb{R}^n$ si :

- (v_1, \dots, v_s) est une base de F
- les vecteurs v_1, \dots, v_s sont deux à deux orthogonaux

■ Exemple 8 :

| La base canonique de \mathbb{R}^n est une base orthogonale.

? Exercice 8

Dans \mathbb{R}^3 , on pose $\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix} \right\}$. On admet que c'est une base de \mathbb{R}^3 .
Montrer que c'est une base orthogonale de \mathbb{R}^3 .

Solution

On note u, v, w respectivement, dans l'ordre cité ci-dessus les vecteurs de \mathcal{B} . On observe que $\langle u, v \rangle = \langle v, w \rangle = \langle u, w \rangle = 0$. la base \mathcal{B} de \mathbb{R}^3 est ainsi orthogonale.

🍃 Définition

Une famille $\mathcal{F} = (u_1, \dots, u_s) \in \mathbb{R}^n$ est appelée *base orthonormale* (ou *base orthonormée*) de l'e.v. $F \subset \mathbb{R}^n$ si :

- (u_1, \dots, u_s) est une base de F
- les vecteurs u_1, \dots, u_s sont deux à deux orthogonaux Dans ce cas, on dit que
- les vecteurs u_1, \dots, u_s sont tous de norme 1.

la matrice $P = \mathcal{M}_{B_0}(\mathcal{F})$ est une *matrice orthogonale*.

? Exercice 9

Dans l'exercice précédent, montrer que \mathcal{B} n'est pas une base orthonormée.

Solution

Avec les notations de l'exercice précédent, il suffit par exemple de constater que $\|u\|^2 = 3 \neq 1^2$. Ainsi, la base n'est pas orthonormée.



Remarque :

Pour transformer une base orthogonale en une base orthonormée, il suffit de diviser chaque vecteur par sa norme.

■ Exemple 9 :

Dans \mathbb{R}^3 , on a vu que la famille $\left\{ \underbrace{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}}_{v_1}, \underbrace{\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}}_{v_2}, \underbrace{\begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}}_{v_3} \right\}$ était une base orthogonale,

ainsi la famille $\left\{ \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} \right\}$ est une base orthonormée car

$$\|v_1\| = \sqrt{3}, \|v_2\| = \sqrt{2}, \|v_3\| = \sqrt{6}$$

et donc

$$\left\| \frac{1}{\sqrt{3}} v_1 \right\| = \frac{1}{\sqrt{3}} \|v_1\| = 1$$

de même

$$\left\| \frac{1}{\sqrt{2}} v_2 \right\| = 1, \quad \left\| \frac{1}{\sqrt{6}} v_3 \right\| = 1$$

? Exercice 10

Soit la base orthogonale (admis pour l'instant) de \mathbb{R}^3 notée $\mathcal{B} = \left\{ \begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix}, \begin{pmatrix} 1 \\ -2 \\ 0 \end{pmatrix}, \begin{pmatrix} 4 \\ 2 \\ 5 \end{pmatrix} \right\}$. Déterminer la base orthonormée issue de \mathcal{B} .

Solution

$$\text{On a } \mathcal{B}' = \left\{ \frac{1}{3} \begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix}, \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ -2 \\ 0 \end{pmatrix}, \frac{1}{3\sqrt{5}} \begin{pmatrix} 4 \\ 2 \\ 5 \end{pmatrix} \right\}$$

Commentaires :

Notons que dans l'exemple précédent, la matrice de Gram de la base orthogonale (v_1, v_2, v_3) en question était

$$G = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 6 \end{pmatrix}$$

La matrice est diagonale car, comme la base est orthogonale, tous les produits scalaires $\langle v_i, v_j \rangle$ sont nuls sauf les $\langle v_i, v_i \rangle$ qui sont égaux à $\|v_i\|^2$. De la même manière, la matrice de Gram de la base orthonormée $(\frac{1}{\sqrt{3}}v_1, \frac{1}{\sqrt{2}}v_2, \frac{1}{\sqrt{6}}v_3)$ est

$$G = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Ces principes restent vrais quelquesoit la famille orthogonale ou la famille orthonormée de n éléments au sens de la remarque qui va suivre :



Remarque :

Une famille de n vecteurs est orthogonale ssi sa matrice de Gram est diagonale.
 Une famille de n vecteurs est orthonormée ssi sa matrice de Gram est égale à la matrice identité.

Commentaires :

| Les deux exemples ci-dessous montrent comment utiliser la remarque dans le sens \Leftarrow :

■ Exemple 10 :

Pour la famille $\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix} \right\}$, si on note $P = \begin{pmatrix} 1 & 1 & -1 \\ 1 & 0 & 2 \\ 1 & -1 & -1 \end{pmatrix}$
 la matrice de Gram est

$${}^t P P = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 6 \end{pmatrix}$$

c'est donc une famille orthogonale.

■ Exemple 11 :

Pour la famille $\left\{ \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \frac{1}{\sqrt{6}} \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix} \right\}$, la matrice de Gram est

$${}^t P P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

c'est donc une famille orthonormée.

Commentaires :

Le sens \Rightarrow de la deuxième équivalence citée dans la remarque précédente :
 "Une famille de n vecteurs est orthonormée ssi sa matrice de Gram est égale à la matrice identité." implique notamment que, si $u, v \in \mathbb{R}^n$ et qu'on dispose d'une base \mathcal{B} orthonormée, la formule

$$\langle u, v \rangle = {}^t M_{\mathcal{B}}(u) \cdot \underbrace{G_{\mathcal{B}}}_{Id} \cdot M_{\mathcal{B}}(v) = {}^t M_{\mathcal{B}}(u) M_{\mathcal{B}}(v)$$

Autrement dit, si

$$M_{\mathcal{B}}(u) = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}, \quad M_{\mathcal{B}}(v) = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}$$

Alors

$$\langle u, v \rangle = \alpha_1 \beta_1 + \dots + \alpha_n \beta_n$$

Ceci signifie que la formule est la même que si on était dans la base canonique :

Propriété 174

Soit \mathcal{B} une base orthonormée de \mathbb{R}^n et \mathcal{B}_0 la base canonique. Soient $u, v \in \mathbb{R}^n$ tels que

$$M_{\mathcal{B}_0}(u) = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad M_{\mathcal{B}}(u) = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} \quad \text{et} \quad M_{\mathcal{B}_0}(v) = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad M_{\mathcal{B}}(v) = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}$$

Alors

$$\underbrace{\langle u, v \rangle = x_1 y_1 + \dots + x_n y_n}_{\text{formule de définition}} = \underbrace{\alpha_1 \beta_1 + \dots + \alpha_n \beta_n}_{\text{formule du commentaire précédent}}$$

ainsi que par suite :

$$\|u\|^2 = x_1^2 + \dots + x_n^2 = \alpha_1^2 + \dots + \alpha_n^2$$

Commentaires :

Cette propriété signifie que la formule du produit scalaire est applicable quelquesoit la base **orthonormée** dans laquelle on se trouve, et pas seulement dans la base canonique. Ceci permet même de ne pas connaître la base \mathcal{B} en détail !

■ Exemple 12 :

Supposons que $\mathcal{B} = (v_1, v_2, v_3, v_4)$ est une base orthonormée de \mathbb{R}^4 . Calculer le produit scalaire de $u = (1, 0, -2, 1)_{\mathcal{B}}$ et $v = (0, 2, 1, 1)_{\mathcal{B}}$.

La base \mathcal{B} étant orthonormée, on a

$$\langle u, v \rangle = 1 \times 0 + 0 \times 2 + (-2) \times 1 + 1 \times 1 = -1$$

Commentaires :

Cette propriété permet ainsi de démontrer très facilement le rapport entre angle et produit scalaire :

Corollaire

Dans \mathbb{R}^2 , si θ est l'angle (géométrique) entre les vecteurs u et v , on a bien

$$\langle u, v \rangle = \|u\| \cdot \|v\| \cdot \cos \theta$$

Démonstration :

On note $u = (a, b)$ dans la base canonique et on pose $w = (-b, a)$. Les deux vecteurs u, w étant orthogonaux et non colinéaires, on observe qu'ils forment une base orthogonale de \mathbb{R}^2 . On a donc $\mathcal{B} = \left(\frac{u}{\|u\|}, \frac{w}{\|w\|} \right)$ une base orthonormée, dans laquelle

$$M_{\mathcal{B}}(u) = (\|u\|, 0)$$

On pose $x, y \in \mathbb{R}$ tels que

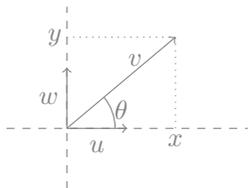
$$M_{\mathcal{B}}(v) = (x, y)$$

Comme la base \mathcal{B} est orthonormée, on sait que

$$\langle u, v \rangle = \|u\| \times x + 0 \times y$$

Or, par propriété "classique" des angles

$$x = \|v\| \cos \theta, \quad y = \|v\| \sin \theta$$



et donc

$$\square \quad \langle u, v \rangle = \|u\| \times \|v\| \cos \theta + 0 \times \|v\| \sin \theta = \|u\| \cdot \|v\| \cdot \cos \theta$$

Commentaires :

Si on observe deux vecteurs non nuls orthogonaux dans \mathbb{R}^2 , graphiquement, on voit bien qu'ils sont non colinéaires et donc forment une famille libre. Dans ce cas, dire qu'une famille est constituée de deux éléments orthogonaux suffirait à justifier que c'est une base. Nous allons voir que ce principe se généralise à \mathbb{R}^n :

Théorème 175

Si (v_1, \dots, v_r) est une famille de vecteurs non nuls deux à deux orthogonaux, alors c'est une famille libre.

Démonstration :

Soient $\lambda_1, \dots, \lambda_r \in \mathbb{R}$ tels que

$$\underbrace{\lambda_1 v_1 + \dots + \lambda_r v_r}_v = 0$$

Prenons ce vecteur et appliquons lui le produit scalaire par v_1 . Par bilinéarité du produit scalaire :

$$\langle v, v_1 \rangle = \langle \lambda_1 v_1 + \dots + \lambda_r v_r, v_1 \rangle = \lambda_1 \langle v_1, v_1 \rangle + \dots + \lambda_r \langle v_r, v_1 \rangle$$

Or, les vecteurs étant 2 à 2 orthogonaux, les produits scalaires sont tous nuls sauf $\langle v_1, v_1 \rangle$ qui vaut $\|v_1\|^2$. Ainsi,

$$\langle v, v_1 \rangle = \lambda_1 \|v_1\|^2$$

Or, le vecteur v étant nul, on a de plus

$$\langle v, v_1 \rangle = \langle \vec{0}, v_1 \rangle = 0$$

et donc

$$0 = \lambda_1 \|v_1\|^2.$$

Le vecteur v_1 étant non nul, on obtient alors

$$\lambda_1 = 0$$

En poursuivant de proche en proche avec $\langle v, v_2 \rangle$, puis $\langle v, v_3 \rangle$, on finit par montrer que tous les λ_i sont nuls. Ainsi, la famille (v_1, \dots, v_r) est libre. \square

Corollaire

Si (v_1, \dots, v_s) est une famille de vecteurs non nuls deux à deux orthogonaux dans $F \subset \mathbb{R}^n$ tel que $\dim F = s$, alors c'est une base de F .

Démonstration :

D'après le théorème précédent, la famille est libre. Comme elle est de cardinal s dans F de dimension s , c'est bien une base de F . \square

■ Exemple 13 :

Reprenons la famille $\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix} \right\}$. On demande de montrer que c'est une base de \mathbb{R}^3 .

Notons P la matrice de coordonnées de \mathcal{B} dans la base canonique. Alors

$$P = \begin{pmatrix} 1 & 1 & -1 \\ 1 & 0 & 2 \\ 1 & -1 & -1 \end{pmatrix}$$

On calcule la matrice de Gram de P : ${}^t P P = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 6 \end{pmatrix}$

La matrice est diagonale. On en déduit que c'est une famille orthogonale, de plus on a trois éléments dans \mathcal{B} , c'est donc une base de \mathbb{R}^3 .

? Exercice 11

Montrer que $\mathcal{B} = \left\{ \begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix}, \begin{pmatrix} 1 \\ -2 \\ 0 \end{pmatrix}, \begin{pmatrix} 4 \\ 2 \\ 5 \end{pmatrix} \right\}$ est une base orthogonale de \mathbb{R}^3 .

Solution

En posant $P = \begin{pmatrix} 2 & 1 & 4 \\ 1 & -2 & 2 \\ -2 & 0 & 5 \end{pmatrix}$, on a ${}^t P P = \begin{pmatrix} 9 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 45 \end{pmatrix}$

La matrice est diagonale, donc la base est orthogonale.

Commentaires :

Une conséquence de tout ceci est qu'il existe certaines matrices pour lesquelles l'inversibilité est immédiate et pour lesquelles le calcul de l'inverse n'est qu'une formalité :

Corollaire

On se donne P une matrice carrée.

- Si les vecteurs colonnes de P forment une famille de vecteurs 2 à 2 orthogonaux (i.e. tPP est diagonale), alors :
 P est inversible
- Si de surcroît les vecteurs sont orthonormés (i.e. ${}^tPP = Id$), alors
 P est inversible et $P^{-1} = {}^tP$

■ Exemple 14 :

Montrons que la matrice $P = \begin{pmatrix} 1 & 1 & -1 \\ 1 & 0 & 2 \\ 1 & -1 & -1 \end{pmatrix}$ est inversible.

La matrice de Gram des vecteurs colonnes est

$$G = {}^tPP = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 6 \end{pmatrix}$$

Cette matrice étant diagonale, on en déduit que les vecteurs de P forment une famille libre et que P est inversible. (Mais ATTENTION, on n'a pas l'inverse pour autant!) L'inverse de P est en fait

$$P^{-1} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & 0 & 1 \\ \frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} \end{pmatrix}$$

■ Exemple 15 :

Montrons que la matrice $P = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \end{pmatrix}$ est inversible.

C'est une base orthonormée, donc P est inversible et $P^{-1} = {}^tP$

II-2 Traduction des formules de changement de base dans une base orthonormée

Commentaires :

Si on se donne une base $\mathcal{B} = (v_1, \dots, v_n)$, on rappelle que la formule générale de changement de base pour un vecteur u donné dans la base \mathcal{B}_0 (par exemple la base canonique) est

$$M_{\mathcal{B}}(u) = M_{\mathcal{B}}(\mathcal{B}_0)M_{\mathcal{B}_0}(u)$$

où l'on sait que

$$M_{\mathcal{B}}(\mathcal{B}_0) = \underbrace{M_{\mathcal{B}_0}(\mathcal{B})}^{\text{notée } P} = P^{-1}$$

Nous avons vu que si la base était orthonormée, alors

$$P^{-1} = {}^tP$$

La formule se transforme donc en

$$M_{\mathcal{B}}(u) = {}^tPM_{\mathcal{B}_0}(u)$$

? Exercice 12

Soit $\mathcal{B} = \left(\frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \frac{1}{\sqrt{6}} \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix} \right)$.

Déterminer dans cette base orthonormée les coordonnées du vecteur $u = (1, -1, 1)$ à l'aide de la remarque précédente.

Solution

Vous devez trouver $u = \frac{1}{\sqrt{3}}(1, 0, -2\sqrt{2})_{\mathcal{B}}$.

Commentaires :

Il existe également une autre manière de procéder directement avec les produits scalaires, qui va nous permettre un peu plus tard dans le chapitre de faire des projections orthogonales :

Propriété 176

Si (u_1, \dots, u_s) est une base orthonormée de $F \subset \mathbb{R}^n$, alors, pour tout $u \in F$, on a

$$u = \langle u, u_1 \rangle u_1 + \dots + \langle u, u_s \rangle u_s$$

Démonstration :

Comme (u_1, \dots, u_s) est une base, il existe forcément des réels $\alpha_1, \dots, \alpha_s$ tels que

$$u = \alpha_1 u_1 + \dots + \alpha_s u_s$$

Or, par bilinéarité du produit scalaire, on a

$$\langle u, u_1 \rangle = \alpha_1 \underbrace{\langle u_1, u_1 \rangle}_{=1} + \alpha_2 \underbrace{\langle u_2, u_1 \rangle}_{=0} + \dots + \alpha_s \underbrace{\langle u_s, u_1 \rangle}_{=0} = \alpha_1$$

de même, en calculant $\langle u, u_2 \rangle, \dots, \langle u, u_n \rangle$, on trouve

$$\alpha_i = \langle u, u_i \rangle \quad \forall i = 1, \dots, n$$

□

Commentaires :

$$\text{Ceci signifie qu'on a "directement" } M_{\mathcal{B}}(u) = \begin{pmatrix} \langle u, u_1 \rangle \\ \vdots \\ \langle u, u_n \rangle \end{pmatrix}$$

où on a noté $\mathcal{B} = (u_1, \dots, u_n)$.

■ **Exemple 16 :**

$$\text{Soit } \mathcal{B} = \left\{ \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \frac{1}{\sqrt{6}} \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix} \right\}.$$

Déterminer dans cette base les coordonnées du vecteur $u = (1, -1, 1)$ à l'aide de la propriété précédente.

On sait déjà que la base est orthonormée. On peut donc appliquer la formule de la propriété précédente, sous réserve de calculer les 3 produits scalaires en question. Si on note U, U_i les vecteurs représentant respectivement u, u_i dans la base canonique, on a

$$\langle u, u_1 \rangle = {}^t U \cdot U_1 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{3}}$$

et

$$\langle u, u_2 \rangle = 0 \quad ; \quad \langle u, u_3 \rangle = -\frac{4}{\sqrt{6}} = -\frac{2\sqrt{2}}{\sqrt{3}}$$

En conclusion,

$$u = \frac{1}{\sqrt{3}} (u_1 - 2\sqrt{2}u_3)$$

? **Exercice 13**

Dans la base orthonormée $\mathcal{B} = \left\{ \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\}$, déterminer les coordonnées de $u = (1, 1, 1)$.

Solution

Vous devez trouver $u = \left(\frac{1}{\sqrt{3}}, \frac{4}{\sqrt{6}}, 0 \right)_{\mathcal{B}}$



Remarque :

Les trois produits scalaires de l'exemple précédent ont été obtenus en calculant ${}^t U \cdot U_i$. Rappelons que le produit scalaire est symétrique et ainsi que

$${}^t U \cdot U_i = \langle u, u_i \rangle = \langle u_i, u \rangle = {}^t U_i \cdot U$$

Les trois lignes du vecteur colonne $M_{\mathcal{B}}(u) = \begin{pmatrix} \langle u, u_1 \rangle \\ \langle u, u_2 \rangle \\ \langle u, u_3 \rangle \end{pmatrix}$ sont donc issues du calcul

de ${}^t U_i \cdot U$. C'est exactement la même chose que de dire que la colonne précédente a été obtenue en une seule fois par le calcul

$$M_{\mathcal{B}}(u) = {}^t P \cdot U$$

ce qui revient donc exactement au même que de dire que

$$M_{\mathcal{B}}(u) = {}^t P M_{\mathcal{B}_0}(u)$$

ce qui était la formule annoncée au début du paragraphe. (Ouf!!)

Moralité : pour ce calcul là, on peut choisir indifféremment l'une ou l'autre technique, il n'y a pas de différence de temps de calcul.

Commentaires :

Que faire maintenant si la base n'est pas orthonormée, mais "simplement" orthogonale. On a vu précédemment que toute base orthogonale pouvait donner lieu à une autre base orthonormée en divisant chaque vecteur par sa norme. Si la famille (v_1, \dots, v_n) est alors une base orthogonale, ceci signifie que $\left(\frac{v_1}{\|v_1\|}, \dots, \frac{v_n}{\|v_n\|} \right)$ est une base orthonormée. On peut facilement passer de l'une à l'autre de la façon suivante : Si on note $u_i = \frac{v_i}{\|v_i\|}$, la propriété précédente nous dit que la formule de décomposition dans la base (u_1, \dots, u_n) donne :

$$u = \langle u, u_1 \rangle u_1 + \dots + \langle u, u_n \rangle u_n$$

Or, on a $u_i = \frac{v_i}{\|v_i\|}$, d'où, en remplaçant dans la formule ci-dessus :

$$u = \langle u, \frac{v_1}{\|v_1\|} \rangle \frac{v_1}{\|v_1\|} + \dots + \langle u, \frac{v_n}{\|v_n\|} \rangle \frac{v_n}{\|v_n\|}$$

mais les $\|v_i\|$ ne sont que des scalaires, ils peuvent donc être sortis du produit scalaire :

$$u = \langle u, v_1 \rangle \frac{1}{\|v_1\|} \frac{v_1}{\|v_1\|} + \dots + \langle u, v_n \rangle \frac{1}{\|v_n\|} \frac{v_n}{\|v_n\|}$$

ce qui, en simplifié, donne

$$u = \underbrace{\langle u, v_1 \rangle}_{\alpha_1} \frac{v_1}{\|v_1\|^2} + \dots + \underbrace{\langle u, v_n \rangle}_{\alpha_n} \frac{v_n}{\|v_n\|^2}$$

Ainsi, les $\alpha_1, \dots, \alpha_n$ sont les coefficients de u dans la base (v_1, \dots, v_n) , alors que les $\langle u, u_i \rangle$ sont les coefficients de u dans la base (u_1, \dots, u_n) . Nous venons en fait de faire la démonstration du corollaire ci-dessous :

Corollaire

Si (v_1, \dots, v_s) est une base orthogonale de $F \subset \mathbb{R}^n$, alors, pour tout $u \in F$, on a

$$u = \frac{\langle u, v_1 \rangle}{\|v_1\|^2} v_1 + \dots + \frac{\langle u, v_s \rangle}{\|v_s\|^2} v_s$$

Démonstration :

cf. commentaires ci-dessus. \square

■ Exemple 17 :

Soit $\mathcal{B} = \left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix} \right) = (v_1, v_2, v_3)$.

Déterminer dans cette base les coordonnées du vecteur $u = (1, -1, 1)$

Ayant déjà fait le calcul, on sait que \mathcal{B} est une base orthogonale. Ainsi,

$$u = \langle u, v_1 \rangle \frac{v_1}{\|v_1\|^2} + \langle u, v_2 \rangle \frac{v_2}{\|v_2\|^2} + \langle u, v_3 \rangle \frac{v_3}{\|v_3\|^2}$$

Or,

$$\langle u, v_1 \rangle = {}^t M_{\mathcal{B}_0}(u) \cdot M_{\mathcal{B}_0}(v_1) = (1 \quad -1 \quad 1) \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 1$$

puis de la même manière

$$\langle u, v_2 \rangle = 0 \quad ; \quad \langle u, v_3 \rangle = -4$$

En conclusion,

$$u = \frac{v_1}{3} - 4 \cdot \frac{v_3}{6} = \frac{1}{3}(v_1 - 2v_3)$$



Remarque :

Pour cette base là, on ne PEUT PAS appliquer de formule du type ${}^t P U$ car dans la formule du changement de base, la matrice P n'est PAS d'inverse ${}^t P$.

? Exercice 14

Dans la base orthogonale $\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\}$, déterminer les coordonnées de $u = (1, 1, 1)$.

Solution

Vous devez trouver $u = (\frac{1}{3}, \frac{2}{3}, 0)_{\mathcal{B}}$

II-3 Diagonalisation dans une base orthonormée

Commentaires :

Cette partie va concerner la diagonalisation de matrice. Nous allons voir que dans certains cas de matrices, on peut systématiquement obtenir une matrice diagonale, de surcroît dans une base très pratique.

Théorème 177 spectral

Toute matrice A **symétrique réelle** est diagonalisable dans une base orthonormée de vecteurs propres au sens suivant : il existe une matrice diagonale réelle D et une matrice P de vecteurs propres orthonormés et deux à deux orthogonaux telle que

$$A = P D P^{-1} \quad \text{avec } P^{-1} = {}^t P$$

■ Exemple 18 :

Soit $M = \begin{pmatrix} 1 & 2 & -1 \\ 2 & -1 & 1 \\ -1 & 1 & 2 \end{pmatrix}$. Sans faire de calcul, on voit que cette matrice est symétrique et RÉELLE, on sait alors que cette matrice est diagonalisable et qu'il existe une base orthonormée de vecteurs propres. (Mais en revanche, on ne sait pas lesquels)

On peut le vérifier par calcul :

$$Sp(M) = \{2, \sqrt{7}, -\sqrt{7}\}$$

et les espaces propres sont :

$$E_2 = Vect \left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right) \quad ; \quad E_{\sqrt{7}} = Vect \left(\begin{pmatrix} 1 - \sqrt{7} \\ \sqrt{7} - 3 \\ 2 \end{pmatrix} \right) \quad ; \quad E_{-\sqrt{7}} = Vect \left(\begin{pmatrix} 1 + \sqrt{7} \\ -\sqrt{7} - 3 \\ 2 \end{pmatrix} \right)$$

Les trois vecteurs sont bien 2 à 2 orthogonaux.

Corollaire

Les espaces propres d'une matrice **symétrique réelle** sont deux à deux orthogonaux.

⚠ IMPORTANCE DES HYPOTHÈSES !

Les matrices diagonalisables qui ne sont pas symétriques ou non réelles, ne sont pas nécessairement diagonalisables dans une base orthonormée (ni même orthogonale.)

■ Contre-Exemple(s) :

La matrice réelle $M = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & -2 \\ 3 & 6 & -5 \end{pmatrix}$ **non symétrique** est diagonalisable de valeur

propre 0, -4, 4 dans une base de vecteurs propres $\left(\begin{pmatrix} 1 \\ -1/2 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ -3 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right)$.

(Faire le calcul pour s'entraîner)

Mais cette base n'est pas du tout orthogonale et il n'existe pas d'autre base de vecteurs propres pouvant être diagonale. (à méditer)...

⚠ ET SI LA MATRICE N'EST PAS RÉELLE ?

> De la même façon, l'hypothèse "matrice réelle" est importante également :

■ Contre-Exemple(s) :

Prenons la matrice **non réelle** mais **symétrique** $M = \begin{pmatrix} 2 & i \\ i & 0 \end{pmatrix}$ et cherchons des valeurs propres :

(Petite astuce au passage :) Comme il s'agit d'une matrice d'ordre 2, on peut utiliser le déterminant de la manière suivante :

$$\begin{aligned} \lambda \text{ vap} &\Leftrightarrow \text{rg}(M - \lambda Id) < 2 \\ &\Leftrightarrow M - \lambda Id \text{ non inversible} \Leftrightarrow \det(M - \lambda Id) = 0 \\ &\Leftrightarrow (\lambda - 1)^2 = 0 \end{aligned}$$

Il n'y a donc qu'une seule valeur propre, c'est 1. Or, on a déjà vu lors d'un exercice précédent que ce type de configuration ne pouvait donner lieu à une matrice diagonalisable.

En conclusion, M est une matrice symétrique mais non diagonalisable...

III Distance et projections orthogonales

Commentaires :

Cette dernière partie du cours va nous donner des formules qui permettent de calculer des coordonnées de projections ainsi que des distances dans des espaces \mathbb{R}^n de manière immédiate grâce à des produits scalaires.

📖 Définition

On appelle **distance** entre les deux vecteurs (ou points) $u, v \in \mathbb{R}^n$ le nombre $\|u - v\|$.

📖 Définition et Proposition

Si A est une partie non vide de \mathbb{R}^n et $M \in \mathbb{R}^n$. Il existe un nombre d tel que

$$\{\|M - a\| \mid a \in A\} \quad \text{soit minimal}$$

On appelle alors **distance de M à A** le nombre

$$d(M, A) = \inf\{\|M - a\| \mid a \in A\}$$

Démonstration :

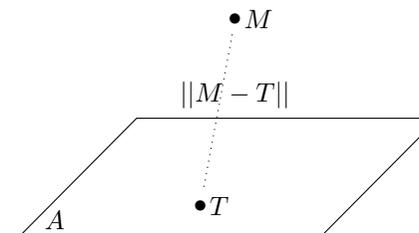
Cette démonstration repose simplement sur le fait que

$$\emptyset \neq \{\|M - a\| \mid a \in A\} \subset [0; +\infty[$$

Or, tout sous-ensemble non vide et minoré de \mathbb{R} admet une borne inférieure. \square

Explications sur un exemple :

Prenons par exemple un plan A dans \mathbb{R}^3 , M un point de \mathbb{R}^3 non compris dans A . On regarde toutes les distances $\|M - T\|$ avec T qui varie dans A :



On prend ensuite la plus petite de ces valeurs, et on affirme alors que c'est la distance entre M et le plan A .

Commentaires :

Voyons dans la suite comment obtenir des formules pour cette distance si A est un sous-espace vectoriel de \mathbb{R}^n . Pour commencer, on va la caractériser par des projections orthogonales. Il faut donc commencer par définir ce qu'est une "projection orthogonale".

Définition
 Un vecteur u est *orthogonal à un espace* F si u est orthogonal à tout vecteur de F .
 On note $u \perp F$.

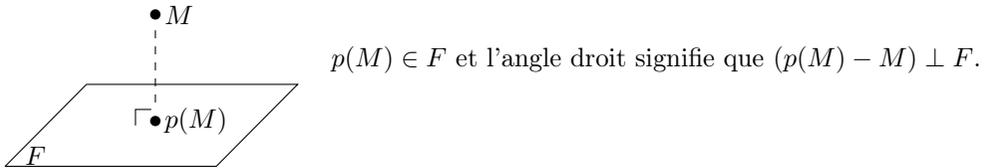
Définition
 Pour tout sev $F \subset \mathbb{R}^n$, on note $F^\perp = \{u \in \mathbb{R}^n \mid u \perp F\}$, i.e. $F^\perp = \{u \in \mathbb{R}^n \mid u \perp v \forall v \in F\}$. On appelle cet espace *l'orthogonal de* F .

Définition
 On appelle *projection orthogonale sur un sous-espace vectoriel* F de \mathbb{R}^n un endomorphisme p de \mathbb{R}^n tel que

$$\begin{cases} \bullet p(u) \in F & \forall u \in \mathbb{R}^n \\ \bullet p(u) - u \perp F & \forall u \in \mathbb{R}^n \end{cases}$$

où $p(u) - u \perp F$ signifie que $p(u) - u \perp v$ pour tout $v \in F$.

Explications sur un exemple :
 Reprenons par exemple un plan F dans \mathbb{R}^3 , M un point de \mathbb{R}^3 non compris dans F . Graphiquement, la projection orthogonale se présente comme suit :



Théorème 178

Si F est un sous-espace vectoriel F de \mathbb{R}^n , il existe une unique projection orthogonale sur F .

Démonstration : admise. \square

Remarque :

Si $u \in F$, alors $p(u) = u$:

Graphiquement, sur le schéma précédent, on peut bien observer ce phénomène : si M est dans le plan A , alors $p(M)$ est confondu avec M .
 De manière générale : si $u \in F$, alors $\langle u - u, v \rangle = 0 \quad \forall v \in F$, d'où $u - u \perp F$.
 Les conditions sont donc remplies pour que $u = p(u)$ par définition de la projection et par unicité de la projection.

Corollaire

Pour tout sev $F \subset \mathbb{R}^n$ et tout $x \in \mathbb{R}^n$, il existe une unique décomposition $(x_F, x_{F^\perp}) \in F \times F^\perp$ telle que

$$x = x_F + x_{F^\perp}$$

(x_F est la projection orthogonale de x sur F .)

Démonstration :
 c'est la traduction immédiate de la définition et du théorème d'unicité précédent.
 \square

Remarque :

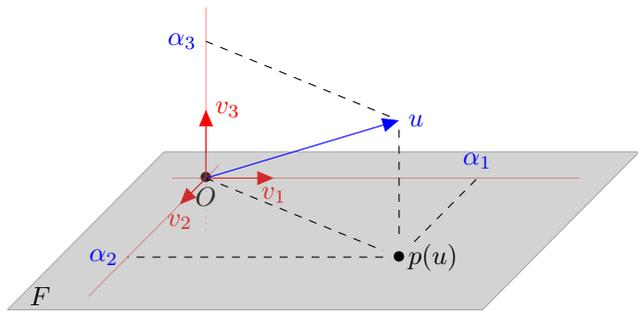
Si un vecteur u est exprimé dans une base orthogonale relative à l'espace F sur lequel on projette, alors l'expression de la projection est triviale :

Propriété 179

Soit $\mathcal{B} = (v_1, \dots, v_n)$ une base orthogonale de \mathbb{R}^n et $F = Vect(v_1, \dots, v_s)$.
 Si p est une projection orthogonale sur F et que $M_{\mathcal{B}}(u) = (\alpha_1, \dots, \alpha_n)$, alors

$$p(u) = \alpha_1 v_1 + \dots + \alpha_s v_s, \quad \text{i.e.} \quad M_{\mathcal{B}}(p(u)) = (\alpha_1, \dots, \alpha_s, 0, \dots, 0).$$

Explications sur un exemple : Reprenons par exemple sev de dimension 2 F dans \mathbb{R}^3 , u un vecteur de \mathbb{R}^3 non compris dans F . Graphiquement, dans le repère orthogonal (v_1, v_2, v_3) , où $F = Vect(v_1, v_2)$, la projection orthogonale se présente comme suit :



où on observe bien que

$$u = \alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3$$

et ensuite que sa projection est

$$p(u) = \alpha_1 v_1 + \alpha_2 v_2$$

on a donc simplement retiré la partie en v_3 pour obtenir $p(u)$ à partir de u .

Démonstration :

Par hypothèse, on a

$$u = \underbrace{\alpha_1 v_1 + \dots + \alpha_s v_s}_{=H \in F} + \underbrace{\alpha_{s+1} v_{s+1} + \dots + \alpha_n v_n}_{\perp F}$$

Donc

$$H \in F \quad \text{et} \quad u - H \perp F$$

ce qui signifie bien par définition et unicité de la projection que

$$p(u) = H$$

□

Commentaires :

Ceci signifie en fait que pour trouver la projection orthogonale d'un vecteur sur F, on "garde les coordonnées" dans F et on "retire" celles qui sont orthogonales à F.

Quand le vecteur est exprimé dans une base adéquate, le projeté est donc immédiat (cf exemple ci-dessous)

■ **Exemple 19 :**

Soit $\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix} \right\}$ la base orthogonale (de vecteurs notés respecti-

vement (v_1, v_2, v_3)) que nous avons étudiée lors des divers exemples de ce chapitre.

On pose $A = Vect(v_1, v_2)$ (plan de \mathbb{R}^3).

Soit

$$u = (2, -1, 3)_{\mathcal{B}} = 2v_1 - v_2 + 3v_3$$

Comme la base \mathcal{B} est orthogonale, alors la projection de u sur le plan A est

$$p(u) = (2, -1, 0)_{\mathcal{B}} = 2v_1 - v_2$$

Commentaires :

| *Mais qu'advient-il si u n'est pas exprimé dans une telle base ?*

Théorème 180

Si p est une projection orthogonale sur un sev F de \mathbb{R}^n et que (u_1, \dots, u_s) est une base orthonormée de F , alors, pour tout $u \in \mathbb{R}^n$, on a

$$p(u) = \langle u, u_1 \rangle u_1 + \dots + \langle u, u_s \rangle u_s$$

Démonstration :

Soit $u \in \mathbb{R}^n$ et $H = \langle u, u_1 \rangle u_1 + \dots + \langle u, u_s \rangle u_s$. On veut montrer que $H = p(u)$, c'est-à-dire que

$$H \in F \quad \text{et} \quad H - u \perp F.$$

On a, par définition,

$$H \in Vect(u_1, \dots, u_s) = F$$

La première condition est donc vérifiée.

Vérifions maintenant que $u - H \perp F$, i.e. $\langle u - H, v \rangle = 0 \quad \forall v \in F$.

Comme $F = Vect(u_1, \dots, u_s)$, par bilinéarité du produit scalaire, ceci revient à montrer que

$$\langle u - H, u_i \rangle = 0 \quad \forall i = 1, \dots, s$$

$$\begin{aligned} \text{Or } \langle u - H, u_i \rangle &= \langle u, u_i \rangle - \langle H, u_i \rangle \\ &= \langle u, u_i \rangle - \left\langle \sum_{j=1}^s \langle u, u_j \rangle u_j, u_i \right\rangle \\ &= \langle u, u_i \rangle - \sum_{j=1}^s \langle u, u_j \rangle \underbrace{\langle u_j, u_i \rangle}_{=0 \text{ si } i \neq j} \\ &= \langle u, u_i \rangle - \langle u, u_i \rangle \underbrace{\langle u_i, u_i \rangle}_{=1} \\ &= 0 \end{aligned}$$

D'où

$$H - u \perp F$$

Par condition d'unicité sur la projection orthogonale, les conditions étant remplies, on a bien

$$\langle u, u_1 \rangle u_1 + \dots + \langle u, u_s \rangle u_s = H = p(u)$$

□

Corollaire

Si p est une projection orthogonale sur un sev F de \mathbb{R}^n et que (v_1, \dots, v_s) est une base orthogonale de F . Alors, pour tout $u \in \mathbb{R}^n$, on a

$$p(u) = \frac{\langle u, v_1 \rangle}{\|v_1\|^2} v_1 + \dots + \frac{\langle u, v_s \rangle}{\|v_s\|^2} v_s$$

■ Exemple 20 :

Soit $F = Vect \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\}$. Déterminons les coordonnées de la projection orthogonale de $u = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ sur F .

La base $\mathcal{B} = (v_1, v_2)$ de F donnée ci-dessus étant orthogonale, on utilise simplement la formule

$$p(u) = \frac{\langle u, v_1 \rangle}{\|v_1\|^2} v_1 + \frac{\langle u, v_2 \rangle}{\|v_2\|^2} v_2$$

i.e.

$$p(u) = \frac{1}{3} v_1 + \frac{1}{2} v_2$$

et donc

$$p(u) = \frac{1}{6} (5, 2, -1)$$

Commentaires :

| Voici ci-dessous quelques propriétés des projections que nous expliquerons juste après

Propriété 181

Si p est une projection orthogonale sur un sev F de \mathbb{R}^n , alors

$$p \circ p = p \quad ; \quad \mathfrak{Im} p = F \quad ; \quad p(u) = \vec{0} \quad \forall u \perp F$$

Démonstration :

- Montrons que $p \circ p = p$:

On pose $H = p(u) \in F$.

Nous avons vu précédemment que, comme $H \in F$, on a $p(H) = H$, ce qui signifie que $p \circ p(u) = p(u)$!

- Montrons que $\mathfrak{Im} p = F$:

On sait par définition que $\mathfrak{Im} p \subset F$. Il reste à montrer que $F \subset \mathfrak{Im} p$.

Soit donc $f \in F$. On sait que $p(f) = f$. Donc $f \in \mathfrak{Im} p$!

- Montrons que $p(u) = \vec{0} \quad \forall u \perp F$:

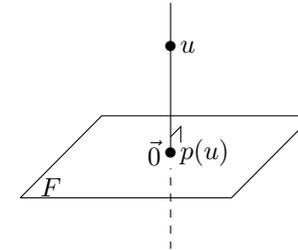
$\vec{0} \in F$ et $u - \vec{0} \perp F$, donc $p(u) = \vec{0}$ par unicité. \square

Commentaires :

- $p \circ p = p$ signifie que si on projette et qu'on recommence la projection, on ne bouge pas. En effet, $H = p(M)$ est déjà sur F , donc recommencer revient donc à dire que $p(H) = H$.

- $\mathfrak{Im} p \subset F$ car on projette sur F , mais tout élément de F peut être considéré comme une projection.

- tous les éléments orthogonaux à F seront d'image nulle :



Commentaires :

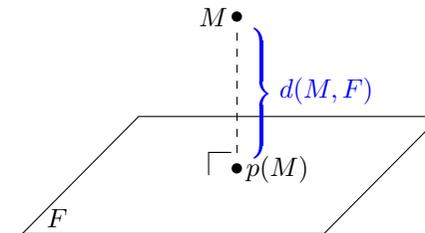
Finalemnt, voyons quel rapport a cette projection avec nos calculs de longueurs et comment nous allons pouvoir établir des formules de calcul. En premier lieu, la distance entre un point M et un sev F est donné grâce à la projection de M de la façon suivante :

Théorème 182

Si F est un sev de \mathbb{R}^n , pour tout point $M \in \mathbb{R}^n$ on a

$$d(M, F) = \|M - p(M)\|$$

où p est la projection orthonormale sur F .



Démonstration :

Montrons que $\|M - p(M)\| = \inf(\|M - y\| \mid y \in F)$.

Soit $y \in F$. On a

$$\begin{aligned} \|M - y\|^2 &= \underbrace{\|M - p(M)\|}_{\perp F}^2 + \underbrace{\|p(M) - y\|}_{\in F}^2 \\ &\stackrel{\text{Pythagore}}{=} \|M - p(M)\|^2 + \|p(M) - y\|^2 \end{aligned}$$

Ce ceci, on tire en particulier

$$\|M - y\|^2 \geq \|M - p(M)\|^2$$

d'où

$$\|M - y\| \geq \|M - p(M)\|$$

Ainsi, $\|M - p(M)\|^2$ étant constant,

$$d(M, F) \geq \|M - p(M)\|.$$

De plus, pour $y = p(M) \in F$, on trouve l'égalité. La borne inférieure est donc atteinte, ainsi

$$d(M, F) = \|M - p(M)\|$$

□

Commentaires :

Venons en maintenant à ces fameuses formules de calcul de distance données dans deux cadres : soit on dispose d'une base orthormée de F , soit seulement d'une base orthogonale.

Corollaire

Si (u_1, \dots, u_n) est une base orthonormée de \mathbb{R}^n avec (u_1, \dots, u_s) base de F , on a

$$d(M, F)^2 = \left\| \sum_{i=s+1}^n \langle M, u_i \rangle u_i \right\|^2 = \sum_{i=s+1}^n \langle M, u_i \rangle^2$$

Démonstration :

Par formule de décomposition dans la base orthogormée $\mathcal{B} = (u_1, \dots, u_n)$, on a

$$M = \langle M, u_1 \rangle u_1 + \dots + \langle M, u_n \rangle u_n$$

Par formule de projection :

$$p(M) = \langle M, u_1 \rangle u_1 + \dots + \langle M, u_s \rangle u_s$$

et donc

$$M - p(M) = \langle M, u_{s+1} \rangle u_{s+1} + \dots + \langle M, u_n \rangle u_n$$

D'où

$$\begin{aligned} d(M, F)^2 &= \|M - p(M)\|^2 \\ &= \left\| \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \langle M, u_{s+1} \rangle \\ \vdots \\ \langle M, u_n \rangle \end{pmatrix} \right\|_{\mathcal{B}}^2 \\ &= 0^2 \dots + 0^2 + \langle M, u_{s+1} \rangle^2 + \dots + \langle M, u_n \rangle^2 \\ &\quad \text{car la base } \mathcal{B} \text{ est orthonormée} \end{aligned}$$

ou alors on démontre ceci en utilisant la formule de Pythagore :

$$\begin{aligned} d(M, F)^2 &= \|M - p(M)\|^2 \\ &= \|\langle M, u_{s+1} \rangle u_{s+1} + \dots + \langle M, u_n \rangle u_n\|^2 \\ &= \|\langle M, u_{s+1} \rangle u_{s+1}\|^2 + \dots + \|\langle M, u_n \rangle u_n\|^2 \quad \text{car } \mathcal{B} \text{ est orthogonale} \\ &= \langle M, u_{s+1} \rangle^2 \underbrace{\|u_{s+1}\|^2}_{=1} + \dots + \langle M, u_n \rangle^2 \underbrace{\|u_n\|^2}_{=1} \\ &= \langle M, u_{s+1} \rangle^2 + \dots + \langle M, u_n \rangle^2 \end{aligned}$$

□

■ Exemple 21 :

Soit $F = \text{Vect}((1, 1, 1), (1, 0, -1))$. Déterminer $d(M, F)$ avec $M = (2, 3, -1)$.

Les vecteurs $(1, 1, 1), (1, 0, -1)$ sont orthogonaux. Une base orthogonale de l'espace est par exemple $\mathcal{B} = ((1, 1, 1), (1, 0, -1), (-1, 2, -1))$ (déjà vu précédemment) et une base

orthonormée est $\mathcal{B} = \left(\frac{1}{\sqrt{3}}(1, 1, 1), \frac{1}{\sqrt{2}}(1, 0, -1), \frac{1}{\sqrt{6}}(-1, 2, -1) \right)$ Ainsi,

$$d(M, F)^2 = \langle M, u_3 \rangle^2 = \left(\frac{5}{\sqrt{6}} \right)^2$$

et donc

$$d(M, F) = \frac{5}{\sqrt{6}}$$

Corollaire

Si (v_1, \dots, v_n) est une base orthogonale de \mathbb{R}^n avec (v_1, \dots, v_s) base de F , on a

$$d(M, F)^2 = \left\| \sum_{i=s+1}^n \langle M, \frac{v_i}{\|v_i\|} \rangle \frac{v_i}{\|v_i\|} \right\|^2 = \sum_{i=s+1}^n \langle M, \frac{v_i}{\|v_i\|} \rangle^2 = \sum_{i=s+1}^n \frac{\langle M, v_i \rangle^2}{\|v_i\|^2}$$

■ Exemple 22 :

Soit $F = \text{Vect}((1, 1, 1), (1, 0, -1))$. Déterminer $d(M, F)$ avec $M = (2, 3, -1)$.

Les vecteurs $(1, 1, 1), (1, 0, -1)$ sont orthogonaux. Une base orthogonale de l'espace est par exemple $\mathcal{B} = (\underbrace{(1, 1, 1)}_{v_1}, \underbrace{(1, 0, -1)}_{v_2}, \underbrace{(-1, 2, -1)}_{v_3})$ (déjà vu précédemment) Ainsi,

$$d(M, F)^2 = \frac{\langle M, v_3 \rangle^2}{\|v_3\|^2} = \frac{5^2}{6}$$

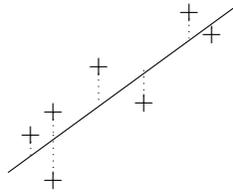
et donc

$$d(M, F) = \frac{5}{\sqrt{6}}$$

Cas particulier : Principe des moindres carrés et projection

On considère un nuage de points $\{(x_i, y_i)\}_{i=1, \dots, n}$. On note $X = (x_1, \dots, x_n) \in \mathbb{R}^n$ et $Y = (y_1, \dots, y_n) \in \mathbb{R}^n$.

On rappelle que le principe des moindres carrés consiste à déterminer une droite (d'équation $y = ax + b$) se rapprochant le plus possible du nuage au sens "vertical", c'est-à-dire de trouver des coefficients a, b tels que $\sum_{k=1}^n \|y_i - ax_i - b\|^2$ soit le plus petit possible :



(i.e. on veut ici minimiser globalement les normes des parties en pointillés.)

On remarque que si on pose $U = (1 \dots, 1) \in \mathbb{R}^n$, pour $v = aX + bU$, on a

$$\|Y - v\|^2 = \sum_{k=1}^n \|y_i - ax_i - b\|^2$$

La problématique consiste donc en réalité à trouver le point $v \in F = \text{Vect}(X, U)$ tel que $\|Y - v\|^2$ soit le plus petit possible. D'après le théorème précédent, c'est la projection de Y sur F . Autrement dit :

La droite est optimale ssi $aX + bU = p_{\text{Vect}(X, U)}(Y)$

? Exercice 15

On considère le nuage de points associé aux données suivantes :

X	1	2	-2	1	-3
Y	0	-1	3	2	-2

1. Montrer que $\mathcal{B}' = \{(1, \dots, 1), (6, 11, -9, 6, -14)\}$, est une base orthogonale de $F = \text{Vect}(X, U)$ (où on utilise les notations de l'explication ci-dessus).
2. Déterminer les coefficients de la droite de régression associée aux données par la méthode des moindres carrés grâce à la méthode citée ci-dessus.

Solution

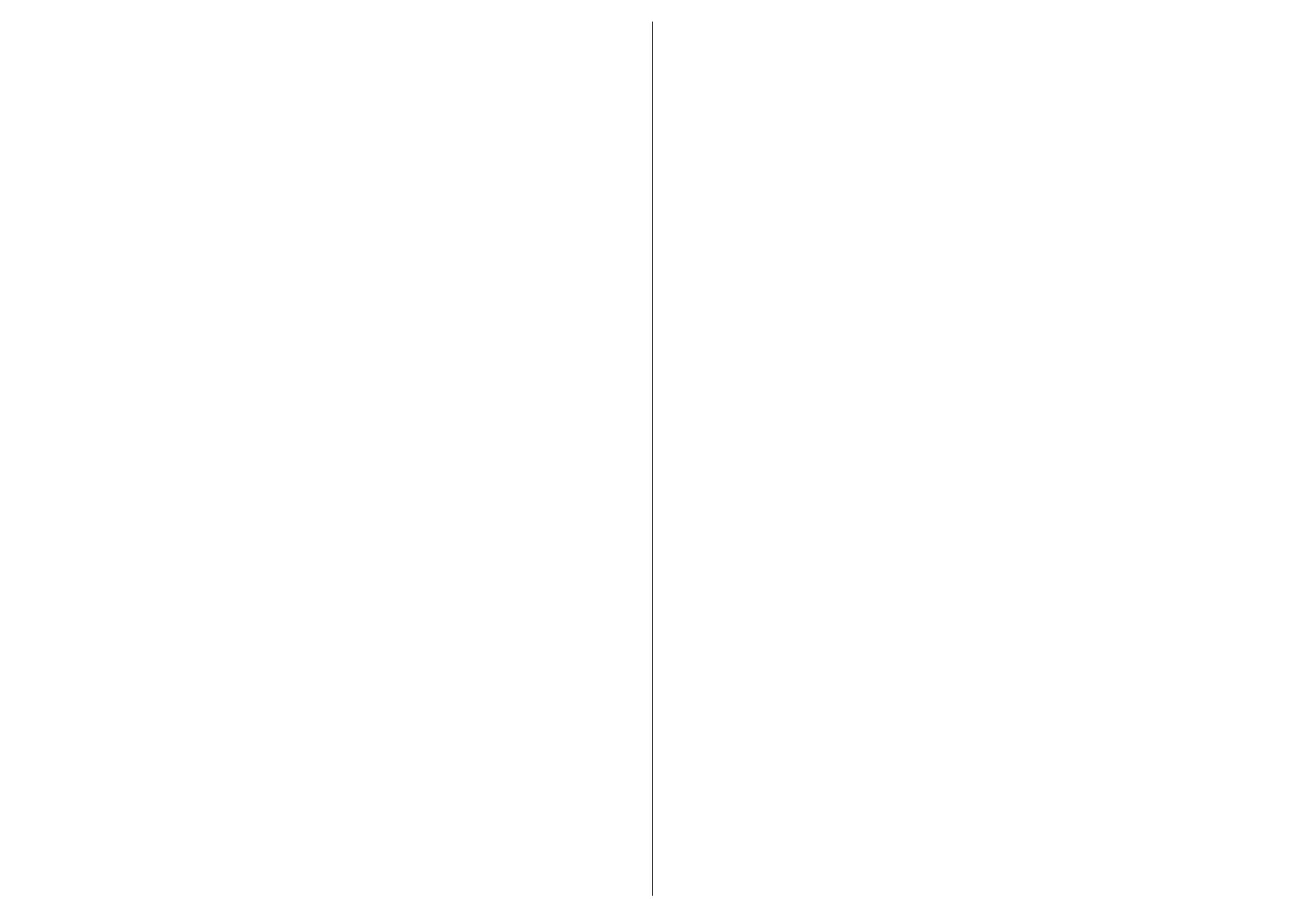
Notons $V_1 = (6, 11, -9, 6, -14)$ et $X = (1, 2, -2, 1, -3)$.

1. On montre facilement que $\langle U, V_1 \rangle = 0$. On constate également que $V_1 = U + 5X \in F$ où F est de dimension 2. Ainsi, \mathcal{B}' est bien une base orthogonale de F .
2. On sait que a, b correspondent au vecteur

$$aX + bU = p_F(Y) = \frac{\langle Y, U \rangle}{\|U\|^2} U + \frac{\langle Y, V_1 \rangle}{\|V_1\|^2} V_1 = \frac{2}{5} U + \frac{2}{470} V_1 = \left(\frac{2}{5} + \frac{1}{235}\right) U + \frac{1}{235} \cdot 5X$$

Or a est le coefficient en X et b le coefficient en U , d'où

$$a = \frac{5}{235} = \frac{1}{47} \simeq 0,02, \quad b = \frac{95}{235} = \frac{19}{47} \simeq 0,4$$



Rappel : Lien entre matrice et application linéaire

• Toute application linéaire $f : E \rightarrow F$ admet une matrice dans n'importe quelles bases \mathcal{B} de E et \mathcal{G} de F . (On la note $M_{\mathcal{B},\mathcal{G}}(f)$.)

Dans le cas d'un endomorphisme, on utilise généralement la même base au départ et à l'arrivée. Dans la base \mathcal{B} , on note alors $M_{\mathcal{B}}(f)$. La matrice est alors carrée.

• À partir de toute matrice $M \in \mathcal{M}_{n,m}(\mathbb{K})$, en choisissant E et F des \mathbb{K} -espaces vectoriels de dimensions respectives m et n ainsi que des bases respectives \mathcal{B}, \mathcal{G} de E et G , on peut construire une unique application linéaire $f : E \rightarrow F$ de matrice M dans les bases en question en définissant :

$$M_{\mathcal{G},\mathcal{B}}(f) = M = \mathcal{G} \left\{ \begin{array}{c} \overbrace{f(b_1) \dots f(b_m)}^{''f(\mathcal{B})''} \\ g_1 \begin{pmatrix} m_{1,1} & \dots & m_{1,m} \\ \vdots & \dots & \vdots \\ m_{n,1} & \dots & m_{n,m} \end{pmatrix} \\ \vdots \\ g_n \end{array} \right.$$

En particulier, si $M \in \mathcal{M}_n(\mathbb{K})$ est une matrice carrée, elle peut être considérée comme la matrice d'un endomorphisme de E de dimension n .

I Valeur propre et vecteur propre

Dans tout ce chapitre, on considèrera E un espace vectoriel sur $\mathbb{K}(= \mathbb{R} \text{ ou } \mathbb{C})$ de dimension n , \mathcal{B} une base de E et $f \in \mathcal{L}(E)$ un endomorphisme de E .

 **Notation :**
Si \mathcal{B} est une base de E , pour simplifier les notations, on notera souvent $M_{\mathcal{B}}$ au lieu de $M_{\mathcal{B}}(f)$.

I-1 Valeur propre d'un endomorphisme

Commentaires :

Les matrices les plus pratiques à utiliser, tant dans le calcul du rang que le calcul des puissances sont les matrices triangulaires, voir même mieux, les matrices diagonales. Le but de ce chapitre est déterminer, si possible, une base dans laquelle la matrice de l'endomorphisme sera diagonale (ou du moins "la plus diagonale possible")

 **Remarque :**

On cherche à "annuler" une bonne partie des coefficients non diagonaux. Or, si on note $\mathcal{B} = (e_1, \dots, e_n)$, on obtient l'équivalence suivante :

$$M_{\mathcal{B}}(f) = e_i \left(\begin{array}{ccc} & f(e_i) & \\ & 0 & \\ & \vdots & \\ & 0 & \dots \\ & \lambda & \\ & 0 & \\ & \vdots & \\ & 0 & \end{array} \right) \iff f(e_i) = \lambda e_i$$

Le but est donc de trouver un maximum de vecteurs non nuls linéairement indépendants tels que $f(v)$ et v soient colinéaires (i.e. il existe $\lambda \in \mathbb{K}$ tel que $f(v) = \lambda v$.)

 **Définition**

On dit que $\lambda \in \mathbb{K}$ est une *valeur propre* (*vap*) de l'endomorphisme f s'il existe $v \neq 0$ tel que $f(v) = \lambda v$. L'ensemble des valeurs propres d'un endomorphisme f est appelé *spectre* de f et noté $Sp(f)$.

■ **Exemples :**

Supposons $n = 2$ et $\mathcal{B} = \{e_1, e_2\}$ une base de E ;

1 ■ Si $M_{\mathcal{B}}(f) = \begin{pmatrix} 2 & 1 \\ 0 & -1 \end{pmatrix}$, alors 2 est valeur propre car

$$f(e_1) = 2 e_1$$

2 ■ Si $M_{\mathcal{B}}(f) = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, alors 0 est valeur propre car

$$f(e_2 - e_1) = \vec{0} = \mathbf{0} \times \underbrace{(e_2 - e_1)}_{\neq 0}$$

I-2 Vecteurs propres et espaces propres d'un endomorphisme

 **Définition**

Si $\lambda \in \mathbb{K}$ est une *valeur propre* de l'endomorphisme f , on appelle v un *vecteur propre* (*vep*) de la valeur propre λ tout vecteur v **non nul** tel que $f(v) = \lambda v$. L'ensemble $E_{\lambda}(f) = \{v \in E \mid f(v) = \lambda v\}$ s'appelle *l'espace propre* associé à la valeur propre λ .

■ Exemples :

Supposons $n = 2$ et $=\{e_1, e_2\}$ une base de E ;

3 ■ Si $M_{\mathcal{B}}(f) = \begin{pmatrix} 2 & 1 \\ 0 & -1 \end{pmatrix}$, alors $e_1 \neq 0$ est vecteur propre de valeur propre 2 car $f(e_1) = 2e_1$

4 ■ Si $M_{\mathcal{B}}(f) = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, alors $v = e_2 - e_1 \neq 0$ et vecteur propre de valeur propre 0 car $f(v) = \vec{0} = 0 \times v$

Remarques :

- 1) λ est une valeur propre de f si et seulement si $E_{\lambda}(f) \neq \{0_E\}$. (c'est la définition d'une valeur propre.)
- 2) $0_E \in E_{\lambda}(f)$. Les vecteurs propres de $vap \lambda$ sont les éléments non nuls de $E_{\lambda}(f)$.
- 3) $E_{\lambda} = \ker(f - \lambda Id)$, ce qui entraîne immédiatement que E_{λ} est un \mathbb{K} -e.v. et qu'un **vecteur propre n'est pas unique!** (Par exemple, si v est un vecteur propre de λ , tout αv , avec $\alpha \in \mathbb{K}$ est un vecteur propre de λ .)
- 4) Si v doit être non nul, il se peut par contre que $\lambda = 0$. Dans ce cas, on a $E_0(f) = \ker f$.
- 5) Les valeurs propres et les vecteurs propres ne dépendent pas d'une base.

Notation : S'il n'y a pas de confusion possible, on notera E_{λ} au lieu de $E_{\lambda}(f)$.

I-3 **Stratégies de calcul des valeurs propres d'un endomorphisme**

Tout repose sur les équivalences suivantes :

Propriété 183 *fondamentale*

λ est une valeur propre	\iff	$E_{\lambda} \neq \{0_E\}$
	\iff	$\ker(f - \lambda Id) \neq \{0_E\}$
	\iff	$f - \lambda Id$ est non injective
	\iff	$f - \lambda Id$ est non surjective
	\iff	$\text{rg}(f - \lambda Id) < n = \dim E$

Application : Cette propriété permet de déterminer plusieurs méthodes pour trouver l'ensemble des valeurs propres de f .

Méthode 1 : On détermine l'ensemble des valeurs λ telles que $\text{rg}(f - \lambda Id) < n$.

Ceci se fait généralement à l'aide d'une matrice de f dans une base.

■ Exemple 5 :

Supposons que E soit de dimension 3. Soit f un endomorphisme de E de matrice

$M = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & -2 \\ 3 & 6 & -5 \end{pmatrix}$ dans une base \mathcal{B} . Alors, $\text{rg}(f - \lambda Id) = \text{rg}(M - \lambda I_3)$. Or,

$$M - \lambda I_3 = \begin{pmatrix} 1-\lambda & 2 & 1 \\ 2 & 4-\lambda & -2 \\ 3 & 6 & -5-\lambda \end{pmatrix}$$

Première stratégie possible :

"Sans se poser de questions" : faire un **pivot de Gauss** pour arriver à une matrice triangulaire. Il faut néanmoins choisir stratégiquement ses pivots, mais attention, toujours **non nuls!**

$$M - \lambda I_3 = \begin{pmatrix} 1-\lambda & 2 & 1 \\ 2 & 4-\lambda & -2 \\ 3 & 6 & -5-\lambda \end{pmatrix}$$

on annule les coeff : $L_2 \leftarrow L_2 + 2L_1$; $L_3 \leftarrow L_3 + (5 + \lambda)L_1$

$$\sim \begin{pmatrix} 1-\lambda & 2 & 1 \\ 4-2\lambda & 8-\lambda & 0 \\ -\lambda^2-4\lambda+8 & 16+2\lambda & 0 \end{pmatrix}$$

Le nouveau pivot devrait être en $C^2; L_2$, mais l'expression peut s'annuler.

À ce stade là, on a donc deux possibilités :

★ continuité de la méthode : "je ne me pose pas de questions" :

On prend comme pivot la quantité $8 - \lambda$ (coefficient du milieu). Néanmoins, il faut s'assurer qu'il est non nul. Il faut donc séparer les cas :

Si $\lambda = 8$:

$$M - \lambda I_3 \sim \begin{pmatrix} -7 & 2 & 1 \\ -12 & 0 & 0 \\ -88 & 32 & 0 \end{pmatrix} \sim \begin{pmatrix} -7 & 2 & 1 \\ -88 & 32 & 0 \\ -12 & 0 & 0 \end{pmatrix}$$

c'est une matrice de rang 3. $\lambda = 8$ n'est donc pas valeur propre.

Si $\lambda \neq 8$: on choisit comme pivot $8 - \lambda$:

$$M - \lambda I_3 \sim \begin{pmatrix} 1 - \lambda & 2 & 1 \\ 4 - 2\lambda & \boxed{8 - \lambda} & 0 \\ -\lambda^2 - 4\lambda + 8 & 16 + 2\lambda & 0 \end{pmatrix}$$

opération : $L_3 \leftarrow (8 - \lambda)L_3 - (16 + 2\lambda)L_2$

$$\sim \begin{pmatrix} 1 - \lambda & 2 & 1 \\ 4 - 2\lambda & \boxed{8 - \lambda} & 0 \\ A & 0 & 0 \end{pmatrix}$$

où $A = (8 - \lambda)(-\lambda^2 - 4\lambda + 8) - (16 + 2\lambda)(4 - 2\lambda)$

$$= \lambda^3 + \lambda^2 \underbrace{(-8 + 4 + 4)}_{=0} + \lambda \underbrace{(-4 \times 8 - 8 + 2 \times 16 + 2 \times 4)}_{=-16} + \underbrace{8^2 - 4 \times 16}_{=0}$$

$$= \lambda^3 - 16\lambda = \lambda(\lambda^2 - 16) = \boxed{\lambda(\lambda - 4)(\lambda + 4)}$$

Ainsi

$$\boxed{Sp(f) = \{0, -4, 4\}}.$$

★ ou bien "je me débarrasse du problème de pivot nul" :

à partir de $\begin{pmatrix} 1 - \lambda & 2 & 1 \\ 4 - 2\lambda & 8 - \lambda & 0 \\ -\lambda^2 - 4\lambda + 8 & 16 + 2\lambda & 0 \end{pmatrix}$, on retire les λ du coefficient diagonal : $L_2 \leftarrow 2L_2 + L_3$

$$M - \lambda I_3 \sim \begin{pmatrix} 1 - \lambda & 2 & 1 \\ 2(4 - 2\lambda) - \lambda^2 - 4\lambda + 8 & 32 & 0 \\ -\lambda^2 - 4\lambda + 8 & 16 + 2\lambda & 0 \end{pmatrix} = \begin{pmatrix} 1 - \lambda & 2 & 1 \\ -\lambda^2 - 8\lambda + 16 & 32 & 0 \\ -\lambda^2 - 4\lambda + 8 & 16 + 2\lambda & 0 \end{pmatrix}$$

On simplifie la deuxième colonne par 2

$$\sim \begin{pmatrix} 1 - \lambda & 1 & 1 \\ -\lambda^2 - 8\lambda + 16 & 16 & 0 \\ -\lambda^2 - 4\lambda + 8 & 8 + \lambda & 0 \end{pmatrix} \quad \text{puis } L_3 : 16L_3 - (8 + \lambda)L_2$$

$$\sim \begin{pmatrix} 1 - \lambda & 1 & 1 \\ -\lambda^2 - 8\lambda + 16 & 16 & 0 \\ A & 0 & 0 \end{pmatrix}$$

où $A = 16(-\lambda^2 - 4\lambda + 8) - (8 + \lambda)(-\lambda^2 - 8\lambda + 16)$

$$= 16(-\lambda^2 - 4\lambda + 8) - (-\lambda^3 + \lambda^2 \underbrace{(-8 - 8)}_{-16} + \lambda \underbrace{(-8 \cdot 8 + 16)}_{-64 + 16 = -48}) + 8 \cdot 16$$

$$= \lambda^3 + \lambda^2 \underbrace{(-16 + 16)}_{-16} + \lambda \underbrace{(-4 \cdot 16 + 48)}_{-16} = \lambda^3 - 16\lambda = \lambda(\lambda^2 - 16) = \boxed{\lambda(\lambda - 4)(\lambda + 4)}$$

Ainsi

$$\boxed{Sp(f) = \{0, -4, 4\}}.$$

Avantage : avec ces méthodes, on obtient toujours une équation à résoudre.

Inconvénient : il n'est pas toujours simple de la résoudre ...

Deuxième stratégie possible :

Tenter de mettre en valeur par de simples **combinaisons de lignes ou de colonnes** une colonne ou une ligne entière ayant des coefficients proportionnels.

$$M - \lambda I_3 = \begin{pmatrix} 1 - \lambda & 2 & 1 \\ 2 & 4 - \lambda & -2 \\ 3 & 6 & -5 - \lambda \end{pmatrix} \sim \begin{pmatrix} 4 - \lambda & 2 & 1 \\ 4 - \lambda & 4 - \lambda & -2 \\ 4 - \lambda & 6 & -5 - \lambda \end{pmatrix}$$

$C_1 \leftarrow C_1 + C_2 + C_3$

on annule les coefficients de la première colonne :

$$L_2 \rightarrow L_2 - L_1 \quad L_3 \rightarrow L_3 - L_1$$

$$\sim \begin{pmatrix} 4 - \lambda & 2 & 1 \\ 0 & 2 - \lambda & -3 \\ 0 & \boxed{4} & -6 - \lambda \end{pmatrix}$$

pas de simplification évidente, alors on revient au pivot de Gauss

on place sur la diagonale un élément non nul de manière certaine, on échange

$$M - \lambda I_3 \sim \begin{pmatrix} 4 - \lambda & 2 & 1 \\ 0 & \boxed{4} & -6 - \lambda \\ 0 & 2 - \lambda & -3 \end{pmatrix}$$

et c'est parti pour le pivot ! $L_3 \rightarrow 4L_3 - (2 - \lambda)L_2$

$$M - \lambda I_3 \sim \begin{pmatrix} 4 - \lambda & 2 & 1 \\ 0 & 4 & -6 - \lambda \\ 0 & 0 & \underbrace{-3 \times 4 - (2 - \lambda)(-6 - \lambda)}_{-\lambda^2 - 4\lambda} \end{pmatrix} = \begin{pmatrix} 4 - \lambda & 2 & 1 \\ 0 & 4 & -6 - \lambda \\ 0 & 0 & -\lambda(\lambda + 4) \end{pmatrix}$$

On trouve de même que

$$\boxed{Sp(f) = \{0, -4, 4\}}$$

Avantage : Au final le polynôme dont il faut trouver les racines n'est ici "que" de degré 2 au lieu d'être de degré 3 dans l'autre méthode...

Inconvénient : Il n'est pas toujours évident de trouver des combinaisons simples à effectuer, mais il peut valoir la peine de prendre quelques minutes pour y réfléchir. Surtout si la matrice est de grande taille.

Méthode 2 : par résolution de système : "recherche" de $\ker(f - \lambda Id)$:

On se place dans une base \mathcal{B} , dans laquelle on dispose de la matrice M de f . On pose $v = \begin{pmatrix} x \\ y \\ z \end{pmatrix}_{\mathcal{B}}$. On cherche les λ pour lesquels $\vec{0}$ n'est pas la seule solution du système.

Par exemple, toujours pour $M = M_{\mathcal{B}}(f) = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & -2 \\ 3 & 6 & -5 \end{pmatrix}$. Alors

$$(f - \lambda Id)(v) = 0 \Leftrightarrow \begin{cases} (1 - \lambda)x + 2y + \boxed{z} = 0 \\ 2x + (4 - \lambda)y - 2z = 0 \\ 3x + 6y - (5 + \lambda)z = 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} (1 - \lambda)x + 2y + z = 0 \\ (4 - 2\lambda)x + (8 - \lambda)y = 0 & L_2 \leftarrow L_2 + 2L_1 \\ (3 + (5 + \lambda)(1 - \lambda))x + (6 + 2(5 + \lambda))y = 0 & L_3 \leftarrow L_3 + (5 + \lambda)L_1 \end{cases}$$

$$\Leftrightarrow \begin{cases} (1 - \lambda)x + 2y + z = 0 \\ (4 - 2\lambda)x + (8 - \lambda)y = 0 \\ (8 - 4\lambda - \lambda^2)x + (16 + 2\lambda)y = 0 \end{cases}$$

Recherche d'un pivot non nul : $L_2 \leftarrow 2L_2 + L_3$

$$\Leftrightarrow \begin{cases} (1 - \lambda)x + 2y + z = 0 \\ (-\lambda^2 - 8\lambda + 16)x + \boxed{32}y = 0 \\ (8 - 4\lambda - \lambda^2)x + (16 + 2\lambda)y = 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} (1 - \lambda)x + 2y + z = 0 \\ (-\lambda^2 - 8\lambda + 16)x + 32y = 0 \\ Ax = 0 & L_3 \leftarrow 16L_3 - (8 + \lambda)L_2 \end{cases}$$

où $A = 16(-\lambda^2 - 4\lambda + 8) - (8 + \lambda)(-\lambda^2 - 8\lambda + 16) = \lambda(\lambda - 4)(\lambda + 4)$

Le système admet une unique solution si et seulement si $A = \lambda(\lambda - 4)(\lambda + 4) \neq 0$. On a donc

$$\lambda \text{ vap} \Leftrightarrow \lambda \in \ker(f - \lambda Id) \Leftrightarrow \boxed{\lambda \in \{0, -4, 4\}}$$

Avantage : Méthode efficace à tous les coups. (Méthode très similaire au pivot de Gauss dans la matrice $M - \lambda Id$.) Permet de calculer ensuite rapidement les espaces propres...

Inconvénient : Pas très esthétique et un peu rébarbatif. De plus, comme dans la méthode du pivot de Gauss, l'équation à résoudre n'est pas toujours très simple.

I-4 Stratégies de calcul des espaces propres d'un endomorphisme

Pour trouver les espaces propres de l'endomorphisme f :

On résoud $(f - \lambda)(v) = 0$, où $\lambda \in Sp(f)$

Ici encore, plusieurs stratégies possibles *qui dépendent principalement de la méthode utilisée pour trouver les vap.* :

- Si on a déterminé les valeurs propres à l'aide du rang de la matrice :

on peut

- Résoudre $MX = \lambda X$ pour chaque valeur propre (où $M = M(f)$.) (long!)
- Pour ne pas avoir à refaire intégralement l'ensemble des calculs pour chaque valeur propre : on note S la matrice la plus simple obtenue à partir des simplifications **par lignes uniquement** à partir de la matrice $M(f) - \lambda Id$. On pourra alors résoudre plutôt $SX = 0$ au lieu de $(M(f) - \lambda Id)X = 0$.

- Si on a utilisé la deuxième méthode pour déterminer les valeurs propres :

Il suffit de résoudre le système restant pour les valeurs propres trouvées.

■ Exemple 6 :

Les espaces propres de l'endomorphisme dans l'exemple précédent sont :

$$E_{-4} = \text{Vect} \begin{pmatrix} 1 \\ -1 \\ -3 \end{pmatrix} \quad E_0 = \text{Vect} \begin{pmatrix} 1 \\ -1/2 \\ 0 \end{pmatrix} \quad E_4 = \text{Vect} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Voyons ceci par deux des stratégies proposées.

Avec la stratégie 1 : en utilisant la matrice $M - \lambda Id$ simplifiée :

La matrice la plus simple que nous avons obtenue par manipulation de lignes à partir de $M - \lambda Id$ (matrice appelée *réduite de Gauss*) est

$$M - \lambda I \sim S = \begin{pmatrix} 1 - \lambda & 2 & 1 \\ -\lambda^2 - 8\lambda + 16 & 32 & 0 \\ -\lambda^2 - 4\lambda + 8 & 16 + 2\lambda & 0 \end{pmatrix}$$

Le système à résoudre est donc

$$S = \begin{cases} (1 - \lambda)x + 2y + z = 0 \\ (-\lambda^2 - 8\lambda + 16)x + 32y = 0 \\ (-\lambda^2 - 4\lambda + 8)x + (16 + 2\lambda)y = 0 \end{cases}$$

pour chacun des $\lambda \in Sp(f) = \{0, -4, 4\}$.

★ Si $\lambda = 0$:

$$S = \begin{cases} x + 2y + z = 0 \\ 16x + 32y = 0 \\ 8x + 16y = 0 \end{cases} \iff \begin{cases} x + 2y + z = 0 \\ x + 2y = 0 \end{cases} \iff \begin{cases} z = 0 \\ x + 2y = 0 \end{cases}$$

D'où $E_0 = \text{Vect} \begin{pmatrix} 1 \\ -1/2 \\ 0 \end{pmatrix}$

★ Si $\lambda = 4$:

$$S = \begin{cases} -3x + 2y + z = 0 \\ -32x + 32y = 0 \\ -24x + 24y = 0 \end{cases} \iff \begin{cases} -3x + 2y + z = 0 \\ x = y \end{cases} \iff \begin{cases} x = z \\ x = y \end{cases}$$

D'où $E_4 = \text{Vect} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$

★ De même si $\lambda = -4 \dots$

Stratégie 2 : en utilisant le système d'équations grâce auquel on a obtenu $Sp(f)$:

Rappelons que si $v = \begin{pmatrix} x \\ y \\ z \end{pmatrix}_B$, on a

$$(f - \lambda Id)(v) = 0 \iff \begin{cases} (1 - \lambda)x + 2y + z = 0 \\ (-\lambda^2 - 8\lambda + 16)x + 32y = 0 \\ \lambda(\lambda - 4)(\lambda + 4)x = 0 \end{cases}$$

Ainsi, si $\lambda \in Sp(f) = \{0; -4; 4\}$, la dernière ligne s'annule et le système se résoud simplement.

★ Si $\lambda = 4$: $(f - \lambda Id)(v) = 0 \iff \begin{cases} -3x + 2y + z = 0 \\ -32x + 32y = 0 \end{cases} \iff \begin{cases} -x + z = 0 \\ x = y \end{cases}$

D'où $E_4 = \text{Vect} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$

★ Si $\lambda = -4$: $(f - \lambda Id)(v) = 0 \iff \begin{cases} 5x + 2y + z = 0 \\ 32x + 32y = 0 \end{cases} \iff \begin{cases} 3x + z = 0 \\ x = -y \end{cases}$

D'où $E_{-4} = \text{Vect} \begin{pmatrix} 1 \\ -1 \\ -3 \end{pmatrix}$

Si seules les dimensions de l'espace propre nous intéressent :

Il se peut que l'énoncé ne nécessite pas le calcul exact de l'espace propre associé à la *vap* λ , mais seulement la connaissance de la dimension de E_λ . Dans ce cas, notons que

$$\dim E_\lambda = \dim \ker(f - \lambda Id) = n - \text{rg}(f - \lambda Id)$$

Si on a utilisé la méthode 1 pour trouver les valeurs propres, ce rang est obtenu directement sur la matrice réduite de $M - \lambda Id$.

■ Exemple 7 :

Avec $M = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & -2 \\ 3 & 6 & -5 \end{pmatrix}$, on a trouvé une réduite : $\begin{pmatrix} 4 - \lambda & 2 & 1 \\ 0 & 4 & -6 - \lambda \\ 0 & 0 & -\lambda(\lambda + 4) \end{pmatrix}$.

Pour la valeur propre 4, on a donc

$$\text{rg}(M - 4Id) = \text{rg} \begin{pmatrix} 0 & 2 & 1 \\ 0 & 4 & -10 \\ 0 & 0 & 0 \end{pmatrix} = 2$$

D'où

$$\dim E_4 = 3 - 2 = 1$$

Remarque : Dans le cas de cette matrice, on aurait pu trouver les dimensions associées d'une manière encore plus rapide. C'est ce que nous verrons entre autres par la suite.

I-5 Valeurs propres d'une matrice

Définition

Soit $M \in \mathcal{M}_n(\mathbb{K})$ une matrice carrée. On appelle λ une *valeur propre* de M s'il existe un vecteur $X \in \mathbb{K}^n$ **non nul** tel que $MX = \lambda X$. On note $Sp(M)$ et on appelle *spectre de M* l'ensemble des valeurs propres de M .

Définition

Si λ est une valeur propre de M , alors on note $E_\lambda(M) = \{X \in \mathbb{K}^n \mid MX = \lambda X\}$ et on appelle *vecteur propre* de valeur propre λ de M n'importe quel élément **non nul** de $E_\lambda(M)$. On appelle $E_\lambda(M)$ l'*espace propre* de la valeur propre λ .

■ Exemple 8 :

Avec $M = \begin{pmatrix} 2 & -1 \\ 0 & 1 \end{pmatrix}$ et $V = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, on a $MV = 2V$ et donc V est vecteur propre de valeur propre 2 de la matrice M .



Remarque :

Les vecteurs propres de matrice sont toujours des éléments de l'espace vectoriel \mathbb{K}^n (et non des polynômes ou autre...)

Propriété 184

Si $f : E \rightarrow E$ est un endomorphisme associé à M tel que $M = M_{\mathcal{B}}(f)$, alors

- X est un vecteur propre de M de vap λ si et seulement si le vecteur v de coordonnées X dans la base \mathcal{B} est un vecteur propre de f , de vap λ .
- λ est une valeur propre de M si et seulement si c'est une valeur propre de f .

Démonstration :

• i) Soit X un vecteur de \mathbb{K}^n et $v = X_{\mathcal{B}}$. Alors

$$\begin{aligned} X \text{ est un vecteur propre de } M &\iff MX = \lambda X \\ &\iff M_{\mathcal{B}}(f)X = \lambda X \\ &\iff f(v) = \lambda v \\ &\iff v \text{ est un vecteur propre de } f \text{ de vap } \lambda. \end{aligned}$$

• ii) C'est une conséquence directe de i). \square

Exemple 9 :

Supposons donné l'endomorphisme $u : \mathbb{R}_2[X] \rightarrow \mathbb{R}_2[X]; P \mapsto XP' - 2P$. Dans la base canonique $\mathcal{B} = (1, X, X^2)$ de $\mathbb{R}_2[X]$, on obtient la matrice de u suivante :

$$M = \mathcal{M}_{\mathcal{B}}(u) = \begin{pmatrix} -2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

On observe également directement sur la matrice que $v_1 = 1, v_2 = X, v_3 = X^2$ sont trois vep de l'endomorphisme u (de vap. resp. $-2, -1, 0$). En posant alors

$$V_i = \mathcal{M}_{\mathcal{B}}(v_i)$$

on obtient $V_1 = (1, 0, 0), V_2 = (0, 1, 0), V_3 = (0, 0, 1)$ et on peut (directement par calcul matriciel) observer qu'en effet, V_1, V_2, V_3 sont vecteurs propres de la matrice (resp. associés aux valeurs propres $-2, -1, 0$).

Moralité de la propriété : parler de valeur propre ("et vecteur propre") d'une matrice ou de l'endomorphisme associé est exactement la même chose. Les mêmes propriétés s'appliquent.

Corollaire n°1

L'ensemble des valeurs propres d'une matrice triangulaire est l'ensemble de tous les coefficients diagonaux.

Démonstration :

Quitte à ce que les coefficients * ci-dessous soient nuls ou à transposer la matrice, on ne perd rien à supposer qu'elle est triangulaire supérieure. On pose

$$T = \begin{pmatrix} \lambda_1 & * & \dots & * \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \dots & 0 & \lambda_n \end{pmatrix}$$

λ est une valeur propre de M si et seulement si $\text{rg}(M - \lambda Id) < n$. Or $D - \lambda Id =$

$$\begin{pmatrix} \lambda_1 - \lambda & * & \dots & * \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \dots & 0 & \lambda_n - \lambda \end{pmatrix}$$

qui est de rang $< n$ si et seulement si l'un des coefficients diagonaux s'annule, autrement dit, si et seulement si $\lambda \in \{\lambda_1, \dots, \lambda_n\}$, d'où $Sp(T) = \{\lambda_1, \dots, \lambda_n\}$ \square

Corollaire n°2

Deux matrices semblables ont les mêmes valeurs propres.

Démonstration :

Deux matrices semblables sont deux matrices d'un même endomorphisme exprimées dans deux bases différentes. \square

Corollaire n°3

Si M est semblable à une matrice triangulaire ou diagonale T , alors $Sp(M)$ est constitué de tous les coefficients diagonaux de M .

Démonstration :

Le spectre d'une matrice triangulaire ou diagonale est constitué de l'ensemble de tous ses coefficients diagonaux et, d'après le corollaire précédent, $Sp(M) = Sp(T)$. \square



Remarque :

Pour un même vecteur propre $v \in E$ d'un endomorphisme u , en écrivant les matrices M et M' dans 2 bases différentes (i.e. M et M' semblables), les vecteurs propres de M et M' associés à v sont reliés :

Propriété 185

Soient $M, M' \in \mathcal{M}_n(\mathbb{K})$ deux matrices semblables, $\lambda \in \mathbb{K}$ et $X \in \mathbb{K}^n$. Alors,

$$MX = \lambda X \Leftrightarrow M'X' = \lambda X'$$

où $X' = P^{-1}X$, si $M' = P^{-1}MP$. (ce qui correspond au changement de base.)

■ Exemple 10 :

Soit $\mathcal{B} = (e_1, e_2)$ la base canonique de \mathbb{R}^2 . On pose le changement de base

$$P = M_{\mathcal{B}}(\mathcal{B}') = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \quad \text{pour lequel } P^{-1} = M_{\mathcal{B}'}(\mathcal{B}) = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}.$$

$$\text{Si } M = \begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix}, \text{ on a } M' = P^{-1}MP = \begin{pmatrix} 6 & 6 \\ -3 & -3 \end{pmatrix}$$

Vecteur propres de M :

$$V_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad V_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (\text{vap resp. } 3, 0)$$

D'où les vecteurs propres de M' :

$$V'_1 = P^{-1}V_1 = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}, \quad V'_2 = P^{-1}V_2 = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

où V_i, V'_i sont donc les 2 écritures différentes de e_i dans les bases \mathcal{B} et \mathcal{B}' :

$$V_1 = M_{\mathcal{B}}(e_1), \quad V'_1 = M_{\mathcal{B}'}(e_1), \quad V_2 = M_{\mathcal{B}}(e_2), \quad V'_2 = M_{\mathcal{B}'}(e_2)$$

On peut notamment vérifier que

$$MV_1 = 3V_1, \quad MV_2 = 0V_2, \quad M'V'_1 = 3V'_1, \quad M'V'_2 = 0V'_2$$

II Diagonalisation

II-1 Définitions

🌿 Définition

Soit f un endomorphisme (resp. $M \in \mathcal{M}_n(\mathbb{K})$ une matrice carrée). On dit que f (resp. M) est *diagonalisable* si elle admet (resp. si elle est semblable à) une matrice diagonale.

Rappel : Deux matrices semblables sont deux matrices d'un même endomorphisme dans deux bases différentes.

Corollaire

Supposons que M soit la matrice associée à $f \in \mathcal{L}(E)$ dans une base \mathcal{B} .

M est diagonalisable \Leftrightarrow il existe D diagonale telle que M et D soient deux matrices d'un même endomorphisme

\Leftrightarrow il existe une base \mathcal{B}' t.q. $M_{\mathcal{B}'}(f)$ soit diagonale.

\Leftrightarrow f est diagonalisable.

⚠️ Remarque :

De cette propriété, on déduit le langage (un peu abusif) pour une matrice M d'être "diagonalisable dans une base \mathcal{B}' ".

II-2 Nombre de valeurs propres / dimension des espaces propres

Théorème 186

Soient $\lambda_1, \dots, \lambda_d$ des valeurs propres deux à deux distinctes d'un endomorphisme f , et v_1, \dots, v_d des vecteurs propres (non nuls) de valeurs propres respectives $\lambda_1, \dots, \lambda_d$, alors la famille $\{v_1, \dots, v_d\}$ est libre.

Démonstration :

(Par récurrence sur d .)

Soit \mathcal{H}_d : (Si $\lambda_1, \dots, \lambda_d$ sont des valeurs propres deux à deux distinctes d'un endomorphisme, toute famille (v_1, \dots, v_d) de vecteurs propres associés est libre.)

★ Initialisation : triviale $d = 1$. (Une famille d'un élément non nul est libre.)

★ Hérédité : Supposons \mathcal{H}_{d-1} vraie pour $d \in \mathbb{N}^*$ et montrons que \mathcal{H}_d est vraie.

Supposons qu'il existe $\alpha_1, \dots, \alpha_d \in \mathbb{K}$ tels que $\alpha_1 v_1 + \dots + \alpha_d v_d = 0$.

Par application de f , on obtient

$$\begin{aligned} \alpha_1 v_1 + \dots + \alpha_d v_d &= 0 \\ \alpha_1 \lambda_1 v_1 + \dots + \alpha_d \lambda_d v_d &= 0 \end{aligned}$$

Pour $L_2 - \lambda_1 L_1$, on obtient

$$\alpha_2(\lambda_2 - \lambda_1)v_2 + \dots + \alpha_d(\lambda_d - \lambda_1)v_d = 0$$

Par hypothèse de récurrence, comme les λ_i sont deux à deux distincts, on trouve $\alpha_2 = \dots = \alpha_d = 0$. Le cas α_1 s'en déduit en réinjectant les résultats dans la première ligne. \square

Corollaire

Soit $f \in \mathcal{L}(E)$, (resp. $M \in \mathcal{M}_n(\mathbb{K})$) où E est un espace vectoriel de dimension n .

- f (resp. M) admet au maximum n valeurs propres distinctes.
- Si f (resp. M) admet n valeurs propres distinctes, alors f (resp. M) est diagonalisable.

Démonstration :

- Supposons que f (ou M) admette d valeurs propres distinctes $\lambda_1, \dots, \lambda_d$. Il existe alors d vecteurs propres v_1, \dots, v_d de valeurs propres respectives $\lambda_1, \dots, \lambda_d$. La propriété nous indique que la famille $\{v_1, \dots, v_d\}$ est libre. Ainsi,

$$d = \text{card}\{v_1, \dots, v_d\} \leq \dim E = n$$

- On note $\{v_1, \dots, v_n\}$ une famille de vecteurs propres associés aux n valeurs propres distinctes. D'après la propriété précédente, la famille est libre de cardinal $n = \dim E$. C'est donc une base. Dans cette base, f (resp. M) est diagonale ... par construction!

□

⚠ Remarque :

Si le nombre de valeurs propres distinctes est différent de $n = \text{card } E$, on ne peut rien dire a priori, mais rien n'est désespéré...!

- Exemple 11 :

Matrice diagonalisable : $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$ (car déjà diagonale)

- Exemple 12 :

Matrice non diagonalisable : $M = \begin{pmatrix} a & 0 & 0 \\ 0 & a & 1 \\ 0 & 0 & a \end{pmatrix}$:

On a trivialement $Sp(M) = \{a\}$. Ainsi, si M était diagonalisable, on aurait M semblable à $a Id$, d'où l'existence de P inversible telle que

$$M = P a Id P^{-1} = a P P^{-1} = a Id$$

Or $M \neq a Id$, d'où la contradiction.

Théorème 187 de juxtaposition des bases 1 :

Supposons donné :

- $\lambda_1, \dots, \lambda_d \in \mathbb{K}$ des valeurs propres 2 à 2 **distinctes** de f (resp. M)
- $\mathcal{B}_1, \dots, \mathcal{B}_d$ des familles libres respectivement dans $E_{\lambda_1}, \dots, E_{\lambda_d}$

Alors $\mathcal{B}_1 \cup \dots \cup \mathcal{B}_d$ est libre.

Démonstration :

Notons $\underbrace{v_1, \dots, v_{s_1}}_{\text{dans } E_{\lambda_1}}, \underbrace{v_{s_1+1}, \dots, v_{s_2}}_{\text{dans } E_{\lambda_2}}, \dots, \underbrace{v_{s_{d-1}}, \dots, v_{s_d}}_{\text{dans } E_{\lambda_d}}$.

Montrons que (v_1, \dots, v_{s_d}) est une famille libre par récurrence sur $d \in \mathbb{N}^*$:

★ Initialisation : Si $d = 1$, la famille (v_1, \dots, v_{s_1}) est libre par hypothèse.

★ Hérédité : De $d-1$ à d . Supposons la propriété vraie pour $d-1 \in \mathbb{N}^*$ et montrons qu'elle est vraie pour d .

On pose $\alpha_1, \dots, \alpha_{s_d} \in \mathbb{K}$ tels que

$$\alpha_1 v_1 + \dots + \alpha_{s_d} v_{s_d} = 0 \quad (*_1)$$

on applique f et on obtient

$$\underbrace{\alpha_1 \lambda_1 v_1 + \dots + \alpha_{s_1} \lambda_1 v_{s_1}}_{\text{vap } \lambda_1} + \dots + \underbrace{\alpha_{s_{d-1}+1} \lambda_d v_{s_{d-1}+1} + \dots + \alpha_{s_d} \lambda_d v_{s_d}}_{\text{vap } \lambda_d} = 0 \quad (*_2)$$

On fait $(*_2) - \lambda_d(*_1)$ et on obtient

$$\underbrace{\alpha_1 (\lambda_1 - \lambda_d) v_1 + \dots + \alpha_{s_1} (\lambda_1 - \lambda_d) v_{s_1}}_{\text{vap } \lambda_1} + \dots + \underbrace{\alpha_{s_{d-2}+1} (\lambda_{d-1} - \lambda_d) v_{s_{d-2}+1} + \dots + \alpha_{s_d} (\lambda_{d-1} - \lambda_d) v_{s_d}}_{\text{vap } \lambda_d} = 0$$

Par HR, on obtient la liberté de $(v_1, \dots, v_{s_{d-1}})$, d'où, comme les λ_i sont 2 à 2 distincts,

$$\alpha_1, \dots, \alpha_{s_{d-1}} = 0$$

Il reste

$$\alpha_{s_{d-1}+1} v_{s_{d-1}+1} + \dots + \alpha_{s_d} v_{s_d} = 0$$

mais la famille $(v_{s_{d-1}+1}, \dots, v_{s_d})$ étant une famille libre, les α_i restant sont tous nuls.

★ En conclusion, la famille (v_1, \dots, v_{s_d}) est une famille libre.

□

On en déduit le théorème suivant :

Théorème 188 de juxtaposition des bases 2 :

Supposons donné :

- $\lambda_1, \dots, \lambda_d \in \mathbb{K}$ des valeurs propres 2 à 2 **distinctes** de f (resp. M)
- $\mathcal{B}_1, \dots, \mathcal{B}_d$ des familles libres respectivement dans $E_{\lambda_1}, \dots, E_{\lambda_d}$
- $\text{card}(\mathcal{B}_1) + \dots + \text{card}(\mathcal{B}_d) \geq n$

Alors

- $\text{card}(\mathcal{B}_1) + \dots + \text{card}(\mathcal{B}_d) = n$
- $\mathcal{B} = \mathcal{B}_1 \cup \dots \cup \mathcal{B}_d$ est une base de E
- f (resp. M) est diagonalisable dans la base \mathcal{B} .

De plus,

- \mathcal{B}_i est finalement une base de E_{λ_i} (i.e. $\text{card}(\mathcal{B}_i) = \dim E_{\lambda_i}$)
- il n'existe pas d'autre valeur propre de f (resp. M .)

Démonstration :

démontré uniquement pour le cas de f , le cas de M étant une conséquence immédiate.

- $\text{card}(\mathcal{B}_1) + \dots + \text{card}(\mathcal{B}_d) = n$:

D'après le théorème précédent, \mathcal{B} est une famille libre. On a de plus,

$$\text{card}(\mathcal{B}) = \text{card}(\mathcal{B}_1) + \dots + \text{card}(\mathcal{B}_d) \geq n$$

Or, une famille libre ne peut pas avoir plus de n éléments, d'où $\text{card}(\mathcal{B}_1) + \dots + \text{card}(\mathcal{B}_d) = n$.

- \mathcal{B} est une base de E :

C'est une famille libre avec $n = \dim E$ éléments, donc c'est une base de E .

- f est diagonalisable dans \mathcal{B} :

C'est immédiat, car dans la base \mathcal{B} qui est un ensemble de vecteurs propres, la matrice de f est diagonale.

- \mathcal{B}_i est une base de E_{λ_i} :

On sait que $\mathcal{B}_i \subset E_{\lambda_i}$ et $\text{card}(\mathcal{B}_i) \leq \dim E_{\lambda_i}$ pour tout $i = 1, \dots, d$.

s'il existe k tel que $\text{card}(\mathcal{B}_k) < \dim E_{\lambda_k}$, (quitte à réorganiser le vap, on peut supposer que $k = d$, alors il existe une base \mathcal{B}'_k de E_{λ_k} de cardinal

$$\text{card}(\mathcal{B}'_k) > \text{card}(\mathcal{B}_k).$$

Par le théorème précédent, la famille $(\mathcal{B}_1, \dots, \mathcal{B}_{d-1}, \mathcal{B}'_d)$ est libre, mais

$$\text{card}(\mathcal{B}_1) + \dots + \text{card}(\mathcal{B}_{d-1}) + \text{card}(\mathcal{B}'_d) > \text{card}(\mathcal{B}_1) + \dots + \text{card}(\mathcal{B}_d) = n$$

ce qui est donc impossible car une famille libre ne peut avoir plus de n éléments. Ainsi, on a bien $\text{card}(\mathcal{B}_i) = \dim E_{\lambda_i}$ pour tout i et la famille est alors bien une base de E_{λ_i} .

- Il n'existe pas d'autre vap :

S'il existait une valeur propre supplémentaire, cela reviendrait encore une fois à rajouter une famille libre à la famille \mathcal{B} déjà existante, le tout étant alors encore libre par le théorème précédent, ce qui est impossible. \square

? Exercice 1

Montrer que la matrice $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix}$ est diagonalisable.

Solution

Notons $\mathcal{B} = (e_1, e_2, e_3)$ la base canonique de \mathbb{R}^3 et M la matrice de l'énoncé. M est triangulaire, on sait donc que ses valeurs propres sont ses éléments diagonaux :

$$Sp(M) = \{1, 2\}$$

On observe immédiatement que e_1, e_2 sont des vecteurs propres de M de valeur propre 1. On pose $\mathcal{B}_1 = (e_1, e_2)$ et on a alors

$$\mathcal{B}_1 \subset E_1 \text{ une famille libre.}$$

De plus, on sait qu'il existe au moins un vecteur propre v dans E_2 . On note alors $\mathcal{B}_2 = \{v\}$. On a

$$\mathcal{B}_2 \subset E_2.$$

D'après le théorème de juxtaposition des bases, on sait alors que $\mathcal{B} = (e_1, e_2, v)$ est une famille libre dans \mathbb{R}^3 de cardinal ≥ 3 . C'est donc une base de vecteurs propres de \mathbb{R}^3 et la matrice M est bien diagonalisable.

et dans les deux sens finalement, si on est certain d'avoir tout le spectre :

Théorème 189

On se donne un endomorphisme $f \in \mathcal{L}(E)$ (resp. une matrice $M \in \mathcal{M}_n(\mathbb{K})$) On note $n = \dim E$ et $\lambda_1, \dots, \lambda_d \in \mathbb{K}$ l'ensemble des valeurs propres **distinctes** de f (resp. M). Alors

$$\begin{aligned} f \text{ (resp. } M) \text{ est diagonalisable} &\iff \dim E_{\lambda_1} + \dots + \dim E_{\lambda_d} \geq n \\ &\iff \dim E_{\lambda_1} + \dots + \dim E_{\lambda_d} = n \end{aligned}$$

Démonstration :

On démontrera ceci dans le cadre de l'application f . Il ne reste que le premier \Rightarrow à démontrer par rapport au corollaire précédent.

Comme f est diagonalisable, alors il existe une base $\mathcal{B} = (v_1, \dots, v_n)$ dans laquelle la matrice M de f est diagonale. c'est-à-dire que chacun des éléments de la base

est un vecteur propre. Quitte à réorganiser les vecteurs, on peut supposer qu'on a

$$\underbrace{v_1, \dots, v_{s_1}}_{vap \lambda_1}, \underbrace{v_{s_1+1}, \dots, v_{s_2}}_{vap \lambda_2}, \dots, \underbrace{v_{s_{d-1}+1}, \dots, v_n}_{vap \lambda_d}$$

i.e. $v_1, \dots, v_{s_1} \in E_{\lambda_1}$, $v_{s_1+1}, \dots, v_{s_2} \in E_{\lambda_2}$, $v_{s_{d-1}+1}, \dots, v_n \in E_{\lambda_d}$.

d'où

$$\text{Vect}(v_1, \dots, v_{s_1}) \subset E_{\lambda_1}, \quad \text{Vect}(v_{s_1+1}, \dots, v_{s_2}) \subset E_{\lambda_2}, \quad \dots \quad \text{Vect}(v_{s_{d-1}+1}, \dots, v_n) \subset E_{\lambda_d}$$

Comme toute sous-famille d'une base est une famille libre, on trouve

$$s_1 \leq \dim E_{\lambda_1}, \quad s_2 - s_1 \leq \dim E_{\lambda_2}, \quad \dots \quad s_{d-1} - s_{d-2} \leq \dim E_{\lambda_{d-1}}, \quad n - s_{d-1} \leq \dim E_{\lambda_d}$$

d'où

$$n = s_1 + (s_2 - s_1) + \dots + (n - s_{d-1}) \leq \dim E_{\lambda_1} + \dots + \dim E_{\lambda_d}$$

□

II-3 Diagonalisabilité sur \mathbb{R} ou \mathbb{C} ?

? Exercice 2

Montrer que si $M \in \mathcal{M}_n(\mathbb{R})$ admet une valeur propre complexe non réelle, alors elle ne peut pas être semblable à une matrice diagonale réelle.

Solution

Supposons que $\lambda \in \mathbb{C} - \mathbb{R}$ soit une valeur propre de M . D'après une propriété précédente, quelqu' soit la matrice diagonale en question, l'ensemble des valeurs propres figurent toutes sur la diagonale. L'une d'entre elle étant complexe, elle ne peut être réelle.

🌿 Définition

Soit $M \in \mathcal{M}_n(\mathbb{R})$. On dit que M est *diagonalisable sur \mathbb{R}* (resp. \mathbb{C}) si M est diagonalisable et que toutes ses valeurs propres sont dans \mathbb{R} (resp. \mathbb{C}).

■ Exemple 13 :

La matrice $M = \begin{pmatrix} 2 & -1 & 1 \\ 2 & -1 & 0 \\ -1 & 1 & 0 \end{pmatrix}$ est diagonalisable sur \mathbb{C} , mais pas sur \mathbb{R} .

$$M - \lambda Id = \begin{pmatrix} 2-\lambda & -1 & 1 \\ 2 & -1-\lambda & 0 \\ -1 & 1 & -\lambda \end{pmatrix}$$

Les valeurs propres sont donc

$$\begin{array}{l} L_1 \leftarrow \tilde{L}_1 + L_3 \\ C_3 \leftarrow \tilde{C}_3 - C_1 \\ C_2 \leftarrow \tilde{C}_2 \\ L_2 \leftarrow \tilde{L}_2 + (1+\lambda)L_3 \end{array} \begin{pmatrix} 1-\lambda & 0 & 1-\lambda \\ 2 & -1-\lambda & 0 \\ 1-\lambda & 0 & 0 \\ 1-\lambda & 0 & 0 \\ 2 & -2 & -1-\lambda \\ -1 & 1-\lambda & \boxed{1} \\ 1-\lambda & 0 & 0 \\ * & -1-\lambda^2 & 0 \\ * & * & 1 \end{pmatrix}$$

les valeurs qui annulent $1 - \lambda$ et $-1 - \lambda^2$, autrement dit

$$Sp(M) = \{1, i, -i\}$$

Il y a trois valeurs propres distinctes et l'ordre de M est trois. Ainsi, M est diagonalisable dans \mathbb{C} , mais pas sur \mathbb{R} !

Propriété 190

Soit $M \in \mathcal{M}_n(\mathbb{R})$. λ est valeur propre de M (de vecteur propre X), ssi $\bar{\lambda}$ est valeur propre (de vecteur propre \bar{X}).

Démonstration :

$$MX = \lambda X \iff \overline{MX} = \overline{\lambda X}$$

$$\iff \overline{M} \overline{X} = \bar{\lambda} \overline{X}$$

Or, comme M est réelle, on a $\overline{M} = M$, d'où $M \overline{X} = \bar{\lambda} \overline{X}$ □

⚠ Remarque :

Cette propriété permet entre autres, si on a deux valeurs propres $\lambda, \bar{\lambda}$, de ne calculer qu'un seul espace propre. L'autre s'en déduit simplement par conjugaison.

Commentaires :

La diagonalisabilité va nous servir, par exemple, à calculer les puissances $n^{\text{ème}}$ de matrices. Il ne faut donc pas s'arrêter à la diagonalisation dans \mathbb{R} et ne pas hésiter à se servir des valeurs propres complexes!

Contre exemple : SI M n'est pas réelle, la propriété ne fonctionne pas.

Exemple : Si $M = \begin{pmatrix} 1 & 1 \\ 0 & i \end{pmatrix}$, alors $Sp(M) = \{1, i\}$ et \bar{i} n'est pas vap.

Et avec diagonale $i; -i$ vecteurs propres

II-4 Formule de changement de base

On note $M \in \mathcal{M}_n(\mathbb{K})$. On suppose ici M diagonalisable. On note

$$P = \begin{pmatrix} v_1 & \dots & v_n \\ \star & & \star \\ \vdots & & \vdots \\ \star & & \star \end{pmatrix} \quad \text{une matrice de vecteurs propres de } M$$

et

$$D = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_n \end{pmatrix} \quad \text{la matrice de valeurs propres de } M \text{ associée}$$

où λ_i est la valeur propre associée à v_i . Notons qu'ici, on ne suppose pas forcément que les λ_i sont deux à deux distincts.

Alors on a

$$M = PDP^{-1}$$

C'est en effet la formule de changement de base classique issue de

$$\underbrace{M_{\mathcal{B}}}_{M} = \underbrace{M_{\mathcal{B}'}(\mathcal{B})}_{P} \underbrace{M_{\mathcal{B}'}}_D \underbrace{M_{\mathcal{B}}(\mathcal{B}')}_{P^{-1}}$$

où \mathcal{B} est la base en cours et \mathcal{B}' la base de vecteurs propres.

III Applications : puissances $n^{\text{èmes}}$

Si M est diagonalisable, avec les notations de la partie précédente, on a $M = PDP^{-1}$. Ainsi, si l'on souhaite calculer la puissance $n^{\text{ème}}$ de M , on a

$$M^n = \overbrace{(PDP^{-1})(PDP^{-1}) \dots (PDP^{-1})}^{n \text{ fois}} = PD^n P^{-1}$$

Or, D^n est facile à calculer, puisque

$$D^n = \begin{pmatrix} \lambda_1^n & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_n^n \end{pmatrix}$$

C'est cette stratégie que nous allons tenter d'utiliser dans la suite.

III-1 Suites récurrentes linéaires

On souhaite déterminer de manière exacte l'expression de u_n sachant que

$$u_{n+d} = a_0 u_n + a_1 u_{n+1} + \dots + a_{d-1} u_{n+d-1}$$

■ Exemple 14 :

$$u_{n+3} = -u_{n+2} - u_{n+1} - u_n.$$

• Mise en forme du problème sous forme de matrice :

On pose $U_n = \begin{pmatrix} u_{n+2} \\ u_{n+1} \\ u_n \end{pmatrix}$, alors

$$U_{n+1} = \begin{pmatrix} u_{n+3} \\ u_{n+2} \\ u_{n+1} \end{pmatrix} = \begin{pmatrix} -u_{n+2} - u_{n+1} - u_n \\ u_{n+2} \\ u_{n+1} \end{pmatrix} = \begin{pmatrix} -1 & -1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} u_{n+2} \\ u_{n+1} \\ u_n \end{pmatrix}$$

i.e. $U_{n+1} = AU_n$, où $A = \begin{pmatrix} -1 & -1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$. On en déduit que

$$U_n = A^n U_0$$

Il est donc "très utile" de pouvoir calculer A^n , ou au moins la dernière ligne de A^n , grâce à laquelle on trouvera u_n en fonction de u_2, u_1, u_0 , puisque

$$\begin{pmatrix} u_{n+2} \\ u_{n+1} \\ u_n \end{pmatrix} = U_n = A^n \begin{pmatrix} u_2 \\ u_1 \\ u_0 \end{pmatrix}.$$

• Diagonalisation de A ?

Les valeurs propres de A sont $-i, i$ et -1 , avec

$$E_i = \text{Vect} \begin{pmatrix} 1 \\ -i \\ -1 \end{pmatrix}, \quad E_{-i} = \text{Vect} \begin{pmatrix} 1 \\ i \\ -1 \end{pmatrix}, \quad E_{-1} = \text{Vect} \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$$

Alors $A = PDP^{-1}$ où $D = \begin{pmatrix} -1 & 0 & 0 \\ 0 & i & 0 \\ 0 & 0 & -i \end{pmatrix}$ et $P = \begin{pmatrix} 1 & 1 & 1 \\ -1 & -i & i \\ 1 & -1 & -1 \end{pmatrix}$

• A^n

Le calcul de P^{-1} donne $P^{-1} = \frac{1}{4} \begin{pmatrix} 2 & 0 & 2 \\ 1+i & 2i & -1+i \\ 1-i & -2i & -1-i \end{pmatrix}$. Ainsi,

$$A^n = PD^n P^{-1}$$

Tout ce qui nous intéresse est la dernière ligne :

$$A^n = \frac{1}{4} \begin{pmatrix} * & * & * \\ * & * & * \\ 2(-1)^n - 2Re(i^n(1+i)) & 2Re(-2i i^n) & 2(-1)^n - 2Re(i^n(-1+i)) \end{pmatrix}$$

i.e.

$$A^{2n} = \frac{1}{2} \begin{pmatrix} * & * & * \\ * & * & * \\ 1 - (-1)^n & 0 & 1 + (-1)^n \end{pmatrix}$$

$$A^{2n+1} = \frac{1}{2} \begin{pmatrix} * & * & * \\ * & * & * \\ -1 - (-1)^{n+1} & 2(-1)^n & -1 - (-1)^{n+1} \end{pmatrix}$$

Ainsi,

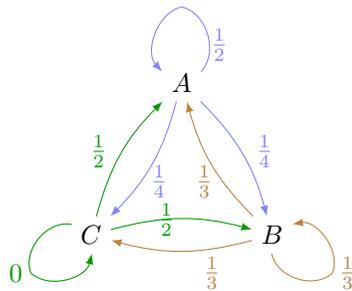
$$u_{2n} = \frac{1}{2} ((1 - (-1)^n)u_2 + (1 + (-1)^n)u_0) \quad \forall n \in \mathbb{N}$$

$$u_{2n+1} = \frac{1}{2} ((-1 - (-1)^{n+1})u_2 + 2(-1)^n u_1 + (-1 - (-1)^{n+1})u_0) \quad \forall n \in \mathbb{N}$$

III-2 Sauts aléatoires par un exemple

On suppose qu'une puce se déplace aléatoirement sur les sommets d'un triangle ABC . À chaque saut, les probabilités sont les suivantes :

- Si la puce est sur A :
la probabilité de rester sur A sur est de $1/2$
la probabilité d'aller sur B est de $1/4$
la probabilité d'aller sur C est de $1/4$
- Si la puce est sur B :
la probabilité d'aller sur A sur est de $1/3$
la probabilité de rester sur B est de $1/3$
la probabilité d'aller sur C est de $1/3$
- Si la puce est sur C :
la probabilité d'aller sur A sur est de $1/2$
la probabilité d'aller sur B est de $1/2$
la probabilité de rester sur C est de 0



Si on note $a_n, (resp. b_n, c_n)$ les probabilités de se trouve sur le point A (resp. B, C) juste après avoir effectué le $n^{\text{ème}}$ saut, on souhaite calculer ces valeurs en fonction de a_0, b_0, c_0 .

Si on note M_n le point sur lequel on est juste après le $n^{\text{ème}}$ saut

$$\begin{cases} a_{n+1} = P_{M_n=A}(M_{n+1}=A)a_n + P_{M_n=B}(M_{n+1}=A)b_n + P_{M_n=C}(M_{n+1}=A)c_n \\ b_{n+1} = P_{M_n=A}(M_{n+1}=B)a_n + P_{M_n=B}(M_{n+1}=B)b_n + P_{M_n=C}(M_{n+1}=B)c_n \\ c_{n+1} = P_{M_n=A}(M_{n+1}=C)a_n + P_{M_n=B}(M_{n+1}=C)b_n + P_{M_n=C}(M_{n+1}=C)c_n \end{cases}$$

Ce qui peut s'exprimer sous forme matricielle

$$U_{n+1} = H_n U_n,$$

où $U_n = \begin{pmatrix} a_n \\ b_n \\ c_n \end{pmatrix}$ et $H_n = \begin{pmatrix} P_{M_n=A}(M_{n+1}=A) & P_{M_n=B}(M_{n+1}=A) & P_{M_n=C}(M_{n+1}=A) \\ P_{M_n=A}(M_{n+1}=B) & P_{M_n=B}(M_{n+1}=B) & P_{M_n=C}(M_{n+1}=B) \\ P_{M_n=A}(M_{n+1}=C) & P_{M_n=B}(M_{n+1}=C) & P_{M_n=C}(M_{n+1}=C) \end{pmatrix}$

ce qui, avec les valeurs de l'énoncé ici, donne

$$H_n = \begin{pmatrix} 1/2 & 1/3 & 1/2 \\ 1/4 & 1/3 & 1/2 \\ 1/4 & 1/3 & 0 \end{pmatrix} \quad \forall n \in \mathbb{N}$$

H_n étant constante en fonction de n , on note $H = H_n$ pour tout $n \in \mathbb{N}$ et on obtient

$$U_n = H^n U_0$$

Il reste à calculer H^n en essayant de diagonaliser !

- Diagonalisation de H : (exercice)

les valeurs propres sont

$$1, -\frac{1}{12}(\sqrt{7}+1), \frac{1}{12}(\sqrt{7}-1)$$

et des vecteurs propres associés sont

$$\begin{pmatrix} 4 \\ 3 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 + \sqrt{7} \\ -3 - \sqrt{7} \end{pmatrix}, \begin{pmatrix} 2 \\ 1 - \sqrt{7} \\ \sqrt{7} - 3 \end{pmatrix}$$

Il y a trois valeurs propres pour une matrice d'ordre 3. Elle est diagonalisable, avec

$$H^n = P D^n P^{-1}$$

où

$$P = \begin{pmatrix} 4 & 2 & 2 \\ 3 & 1 + \sqrt{7} & 1 - \sqrt{7} \\ 2 & -3 - \sqrt{7} & \sqrt{7} - 3 \end{pmatrix}$$

on a

$$P^{-1} = \frac{1}{36} \begin{pmatrix} 4 & 4 & 4 \\ 5 - \frac{11}{\sqrt{7}} & \frac{16}{\sqrt{7}} - 4 & -\frac{2}{\sqrt{7}} - 4 \\ 5 + \frac{11}{\sqrt{7}} & -\frac{16}{\sqrt{7}} - 4 & \frac{2}{\sqrt{7}} - 4 \end{pmatrix}$$

Alors

$$H^n = \frac{1}{36} \begin{pmatrix} 16 + 10\lambda_n + \frac{22}{\sqrt{7}}\mu_n & 16 - 8\lambda_n - \frac{32}{\sqrt{7}}\mu_n & 16 - 8\lambda_n + \frac{4}{\sqrt{7}}\mu_n \\ 12 - 6\lambda_n - \frac{24}{\sqrt{7}}\mu_n & 12 + 12\lambda_n + \frac{12}{\sqrt{7}}\mu_n & 12 - 6\lambda_n + \frac{30}{\sqrt{7}}\mu_n \\ 8 - 4\lambda_n + \frac{2}{\sqrt{7}}\mu_n & 8 - 4\lambda_n + \frac{20}{\sqrt{7}}\mu_n & 8 + 14\lambda_n - \frac{34}{\sqrt{7}}\mu_n \end{pmatrix}$$

$$\text{où } \lambda_n = \left(-\frac{1+\sqrt{7}}{12}\right)^n + \left(\frac{\sqrt{7}-1}{12}\right)^n \text{ et } \mu_n = \left(\frac{\sqrt{7}-1}{12}\right)^n - \left(-\frac{1+\sqrt{7}}{12}\right)^n$$

Conclusion :

Connaissant H^n , on peut déterminer les probabilités d'être sur n'importe quel point à l'instant n . Par exemple, si la puce part du point A , alors on a

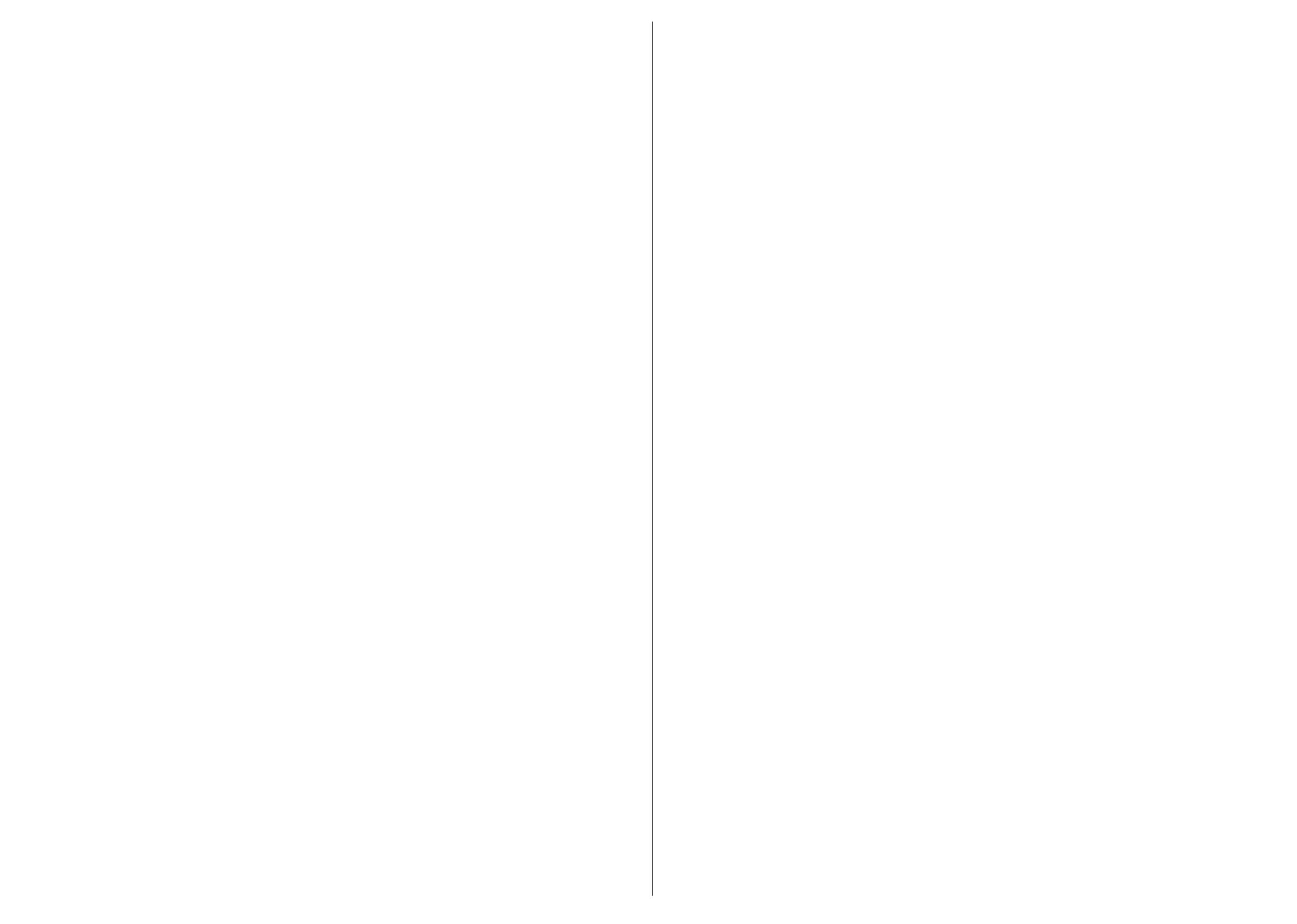
$$U_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

On a

$$\begin{aligned} U_n = H^n U_0 &= \frac{1}{36} \begin{pmatrix} 16 + 10\lambda_n + \frac{22}{\sqrt{7}}\mu_n & 16 - 8\lambda_n - \frac{32}{\sqrt{7}}\mu_n & 16 - 8\lambda_n + \frac{4}{\sqrt{7}}\mu_n \\ 12 - 6\lambda_n - \frac{24}{\sqrt{7}}\mu_n & 12 + 12\lambda_n + \frac{12}{\sqrt{7}}\mu_n & 12 - 6\lambda_n + \frac{30}{\sqrt{7}}\mu_n \\ 8 - 4\lambda_n + \frac{2}{\sqrt{7}}\mu_n & 8 - 4\lambda_n + \frac{20}{\sqrt{7}}\mu_n & 8 + 14\lambda_n - \frac{34}{\sqrt{7}}\mu_n \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \\ &= \frac{1}{36} \begin{pmatrix} 16 + 10\lambda_n + \frac{22}{\sqrt{7}}\mu_n \\ 12 - 6\lambda_n - \frac{24}{\sqrt{7}}\mu_n \\ 8 - 4\lambda_n + \frac{2}{\sqrt{7}}\mu_n \end{pmatrix} = \frac{1}{18} \begin{pmatrix} 8 + 5\lambda_n + \frac{11}{\sqrt{7}}\mu_n \\ 6 - 3\lambda_n - \frac{12}{\sqrt{7}}\mu_n \\ 4 - 2\lambda_n + \frac{1}{\sqrt{7}}\mu_n \end{pmatrix} \end{aligned}$$

Ou autrement dit, si la puce part de A , la probabilité après le $n^{\text{ème}}$ saut d'être

$$\begin{array}{l|l} \text{en } A & \left. \begin{array}{l} \frac{4}{9} + \frac{5}{18} \left(\left(-\frac{1+\sqrt{7}}{12}\right)^n + \left(\frac{\sqrt{7}-1}{12}\right)^n \right) + \frac{11}{18\sqrt{7}} \left(\left(\frac{\sqrt{7}-1}{12}\right)^n - \left(-\frac{1+\sqrt{7}}{12}\right)^n \right) \\ \frac{1}{3} - \frac{1}{6} \left(\left(-\frac{1+\sqrt{7}}{12}\right)^n + \left(\frac{\sqrt{7}-1}{12}\right)^n \right) - \frac{2}{3\sqrt{7}} \left(\left(\frac{\sqrt{7}-1}{12}\right)^n - \left(-\frac{1+\sqrt{7}}{12}\right)^n \right) \\ \frac{2}{9} - \frac{1}{9} \left(\left(-\frac{1+\sqrt{7}}{12}\right)^n + \left(\frac{\sqrt{7}-1}{12}\right)^n \right) + \frac{1}{18\sqrt{7}} \left(\left(\frac{\sqrt{7}-1}{12}\right)^n - \left(-\frac{1+\sqrt{7}}{12}\right)^n \right) \end{array} \right. \\ \text{en } B & \text{est de} \\ \text{en } C & \end{array}$$



Commentaires :

Globalement, en statistiques, on peut chercher à faire plusieurs types de choses :

1. **un résumé des phénomènes observés** (moyenne d'un échantillon, médiane, etc...) Ce sont les statistiques dites "*descriptives*";
2. **Utiliser la théorie pour prévoir les observations** (fréquence d'un événement, moyenne d'un échantillon, etc...). C'est ce qu'on a fait par exemple dans les chapitres précédents de probabilités;
3. **Se servir des observations pour estimer les données théoriques** (probabilité, moyenne, etc...). Ce sont les statistiques dites "*inférentielles*";
4. **Utiliser les observations pour élaborer des modèles théoriques et "prédire l'avenir!"**. Ce sont les statistiques dites "*prédictives*".

(1) : Le principe des statistiques descriptives est, comme son nom l'indique, de "décrire" les phénomènes constatés sur l'ensemble de la population en les résumant avec différents outils potentiels de description (mode, moyenne, écart-type, médiane, quantiles, etc...) On appelle ceci généralement des **résultats "empiriques"**, c'est-à-dire ceux qui sont réellement obtenus en collectant les données sur la population.

(2) : Le principe des probabilités de manière générale est d'établir un comportement typique partant de données théoriques. Elles permettent entre autre de :

- prévoir un comportement général
- prévoir les résultats empiriques que l'on est censé obtenir
- d'évaluer dans une certaine mesure l'écart à ce comportement. (On étudiera dans ce chapitre divers résultats dans ce sens.)

(3) : L'étude des probabilités nous permet de plus d'étudier des comportements dits "limites", qui se produisent lors d'un grand nombre de répétitions d'une même expérience. Les outils et attentes sont multiples, mais dans le cadre des statistiques inférentielles, nous étudierons par exemple dans ce chapitre une approche des "tests de conformité", qui permettent dans certains cas, de prévoir si un échantillon a une chance de faire partie d'une population donnée.

(4) : Dès lors, en revanche, que l'on souhaite faire des prévisions sur les réalisations futures de la variable, il faut aller un peu plus loin dans l'analyse. En effet, on a généralement à faire à des quantités dont on ne connaît pas les caractéristiques théoriques (espérance, variance, ou loi). Or, ce sont des informations capitales si on veut pouvoir établir des tendances et faire des prévisions avec plus ou moins de précision. C'est là que les statistiques prédictives prennent le relais, mais mis à part des cas d'études très simples et relativement naïfs, ceci ne fait pas partie du champs d'étude de notre chapitre.

Notation :

Dans tout le chapitre, sauf précision, le terme *variable aléatoire* désignera une variable aléatoire réelle finie, ou discrète, ou à densité. De plus les suites de variables aléatoires (X_n) seront systématiquement construites sur un même espace probabilisé. La notation F_T désignera la fonction de répartition de la variable aléatoire T et de même, la notation f_T désignera la densité de T si celle-ci existe.

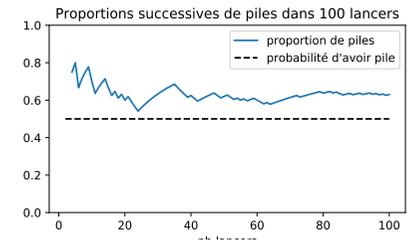
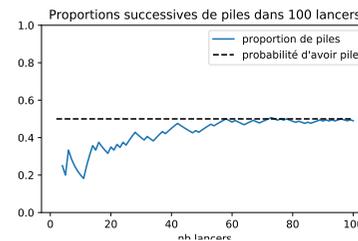
I Loi faible des grands nombres et approximation de la moyenne, d'une probabilité et de la variance

I-1 Problématique et vocabulaire sur un exemple :

Exemple 1 :

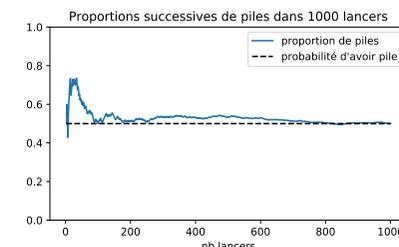
On s'intéresse à une pièce **déséquilibrée** et on se demande quelle est la probabilité d'obtenir Pile.

L'instinct nous dit que si on lance un grand nombre n de fois la pièce, la proportion observée de "Piles" sur une suite de lancers devrait être assez proche de la probabilité théorique qui serait ici $\frac{1}{2}$. Observons des résultats obtenus sur des simulations graphiquement grâce à Python :



La graphique de gauche nous confirme l'instinct. Néanmoins, le graphique de droite semble ne pas corroborer cette affirmation, étant donné que les proportions semblent encore relativement "loin" de la probabilité $\frac{1}{2}$!

Néanmoins, on peut observer qu'en augmentant le nombre de tentatives, le phénomène fini inexorablement par se produire :



Il nous reste donc à démontrer tout ceci. De plus, la suite du chapitre nous permettra de caractériser le caractère plus ou moins "proche", en fonction du nombre de tentatives!

Mettons en place un peu de vocabulaire afin de simplifier les énoncés qui vont suivre :

Définition
Si X est une variable aléatoire, on appelle *n-échantillon* de X une n -liste (X_1, \dots, X_n) de variables aléatoires mutuellement indépendantes et de même loi que X .

■ **Exemple 2 :**

Soit X le résultat d'un lancer de dés. Si on fait n lancers et qu'on note X_1, \dots, X_n les résultats des n lancers successifs, (X_1, \dots, X_n) est un n -échantillon de X . Un résultat de ce n -échantillon avec $n = 5$ pourrait donc être $(1, 6, 4, 2, 4)$.

Ce résultat désigne le fait que le premier lancer donne 1, le deuxième lancer donne 6, etc. . .

Dans la suite, si X est une variable aléatoire, on posera donc (X_1, \dots, X_n) un n -échantillon de X .

Commentaires :

Afin de prouver nos dires sur l'exemple 1, il nous faut tout d'abord transformer un calcul de proportion en un calcul de moyenne. L'exemple ci-dessous nous montre comment :

■ **Exemple 3 :**

Soit X la variable de Bernoulli donnant le nombre de succès "obtenir Pile" dans un seul lancer d'une pièce non nécessairement équilibrée :

$$X = \begin{cases} 1 & \text{si on obtient Pile} \\ 0 & \text{sinon} \end{cases}, \quad \text{avec } X \hookrightarrow \mathcal{B}\left(\frac{1}{2}\right)$$

Alors, dans une série de n lancers, si on note X_i la valeur de X obtenue pour le $i^{\text{ème}}$ lancer, le nombre total de Piles correspond à $X_1 + \dots + X_n$, et la proportion de Piles dans l'échantillon est

$$\frac{X_1 + \dots + X_n}{n}$$

Par exemple, si $n = 5$, voici un exemple de n -échantillon possible :

$$\left(\underbrace{0}_{X_1}, \underbrace{1}_{X_2}, \underbrace{0}_{X_3}, \underbrace{0}_{X_4}, \underbrace{1}_{X_5} \right) \text{ désigne } F P F F F P$$

Ici, le nombre de Piles obtenus est $X_1 + \dots + X_n = 2$ et $\frac{2}{5} = \frac{X_1 + \dots + X_n}{n}$ est la proportion de Piles obtenus.

De plus, $\mathbb{E}[X] = P(\text{"obtenir un Pile"})$. On la note

Ainsi, justifier la véracité de l'intuition de l'exemple 1 reviendrait donc à démontrer l'approximation suivante :

$$\underbrace{\mathbb{E}[X]}_{\substack{\text{i.e. } P(\text{"obtenir un Pile"}) \\ \text{(valeur théorique)}}} \simeq \underbrace{\frac{X_1 + \dots + X_n}{n}}_{\substack{\text{proportion de Piles} \\ \text{(valeur effective)}}$$

Commentaires :

On différencie ainsi les paramètres théoriques (que l'on cherche souvent à déterminer) et les résultats empiriques observés. On distinguera donc en particulier :

- la moyenne théorique $\mathbb{E}[X]$ et la moyenne empirique notée $\overline{X}_n = \frac{X_1 + \dots + X_n}{n}$.
- la variance théorique $V(X)$ et la variance empirique $S_n^2 = \frac{1}{n} \sum_{i=1}^n ((X_i - \overline{X}_n)^2)$.
- la loi et l'histogramme d'un n -échantillon (point et explications sur lesquels nous reviendrons plus tard)

————— En Python :

À partir d'une liste de données du nom de **Donnees**, pour obtenir les caractéristiques de base d'un échantillon avec Python, les commandes issues de la bibliothèque **numpy** sont les suivantes :

```
# moyenne empirique  $\overline{X}_n$  :
np.mean(Donnees)
# variance empirique  $S_n^2$  :
np.var(Donnees)
```

On peut également obtenir les médianes, quartiles, etc. . . avec d'autres commandes.

I-2 Inégalités

Cette partie va permettre de mettre en place les résultats préliminaires permettant la conjecture avancée dans la partie précédente. (Re)-voyons un premier résultat général permettant de décrire le comportement théorique d'une variable :

Lemme 191 (Inégalité de Markov)

Si X est une variable aléatoire réelle **positive** admettant une espérance, alors elle vérifie l'inégalité

$$P(X \geq a) \leq \frac{E(X)}{a} \quad \forall a > 0$$

Remarque :

Ce résultat confirme en particulier que la probabilité que X prenne des valeurs très grandes est forcément petite. (Ceci est d'ailleurs trivial sur les variables finies...)

Démonstration :

• Cas d'une variable discrète : On note $Supp(X) = \{x_i \mid i \in \mathbb{N}\}$ les valeurs de X . (Toutes positives par hypothèse.) Alors,

$$\begin{aligned}
 E(X) &= \sum_{i=0}^{+\infty} x_i P(X = x_i) = \sum_{\substack{i=0 \\ x_i < a}}^{+\infty} \underbrace{x_i P(X = x_i)}_{\geq 0} + \sum_{\substack{i=0 \\ x_i \geq a}}^{+\infty} x_i P(X = x_i) \\
 &\geq \sum_{\substack{i=0 \\ x_i \geq a}}^{+\infty} \underbrace{x_i}_{\geq a} P(X = x_i) \geq a \sum_{\substack{i=0 \\ x_i \geq a}}^{+\infty} P(X = x_i) = a P(X \geq a)
 \end{aligned}$$

• Cas d'une variable à densité : On note f une densité de X . Notons que, comme X est positive, on peut supposer que f est nulle sur $]-\infty; 0[$.

$$\begin{aligned}
 E(X) &= \int_{-\infty}^{+\infty} x f(x) dx = \int_0^{+\infty} x f(x) dx \quad (X \geq 0) \\
 &= \int_0^a \underbrace{x f(x)}_{\geq 0} dx + \int_a^{+\infty} \underbrace{x}_{\geq a} f(x) dx \\
 &\geq \int_0^{+\infty} a f(x) dx = a \int_a^{+\infty} f(x) dx \geq a P(X \geq a)
 \end{aligned}$$

□

Passons maintenant à l'écart à la moyenne. On peut la majorer grâce à une formule simple :

Théorème 192 (Inégalité de Bienaymé-Tchebychev)

Si X est une variable aléatoire admettant un moment d'ordre 2, alors elle vérifie l'inégalité

$$P(|X - E(X)| \geq \epsilon) \leq \frac{V(X)}{\epsilon^2} \quad \forall \epsilon > 0$$

Démonstration :

On pose $Y = |X - E(X)|^2$. Y est une variable aléatoire positive. Comme X admet un moment d'ordre 2, alors $E(|X - E(X)|^2)$ existe (c'est la variance de X) et donc, d'après l'inégalité de la proposition précédente, on a

$$P(|X - E(X)|^2 \geq \epsilon^2) \leq \frac{E(|X - E(X)|^2)}{\epsilon^2} \quad \forall \epsilon > 0$$

i.e.
$$P(|X - E(X)| \geq \epsilon) \leq \frac{V(X)}{\epsilon^2} \quad \forall \epsilon > 0$$

□

Commentaires :

On constate que plus ϵ est grand ou plus la variance est petite, plus la probabilité devient faible. Autrement dit, ceci confirme notre éventuelle intuition que X ne peut s'éloigner de manière trop importante de son espérance qu'avec une probabilité faible. Cette inégalité servira par exemple à démontrer le théorème de la loi faible des grands nombres qui suit dans le paragraphe suivant.

I-3 Loi faible : approximation de la moyenne

Commentaires :

Le théorème qui suit est fondamental en statistiques (même s'il n'est pas le seul !). Il permet de commencer à prouver que l'intuition est bien suivie d'un résultat théorique avéré : le fait qu'une moyenne empirique se rapproche de la moyenne théorique et en conséquence, que la proportion se rapproche en effet inexorablement du paramètre théorique (exemple 1)

Théorème 193 (Loi faible des grands nombres)

Soit (X_1, \dots, X_n) un n -échantillon d'une variable aléatoire X d'espérance μ admettant une variance.

Alors la moyenne empirique $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$ vérifie

$$\lim_{n \rightarrow +\infty} P\left(|\bar{X}_n - \underbrace{\mu}_{=E[X]=E(\bar{X}_n)}| \geq \epsilon\right) = 0 \quad \forall \epsilon > 0$$

Démonstration :

C'est une application du théorème de Bienaymé-Tchebychev. On note

$$S_n = X_1 + \dots + X_n.$$

Par linéarité de E , on a $E(S_n) = n\mu$, d'où

$$E(\bar{X}_n) = \mu.$$

Comme les variables sont non corrélées, on a également

$$V(S_n) = V(X_1) + \dots + V(X_n) = n\sigma^2,$$

d'où $V(\bar{X}_n) = V\left(\frac{S_n}{n}\right) = \frac{1}{n}\sigma^2.$

D'après l'inégalité de Bienaymé-Tchebychev, on a

$$P(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}$$

Le passage à la limite achève la démonstration. □

Commentaires :

La loi faible des grands nombres signifie qu'en quelque sorte, \overline{X}_n converge vers son espérance, d'où,

$$\text{pour } n \text{ grand, } \overline{X}_n \simeq \mathbb{E}[X]$$

On note de plus dans la démonstration que

$$\mathbb{E}[\overline{X}_n] = \mu = \mathbb{E}[X]$$

et on peut montrer que (exercice)

$$V(\overline{X}_n - \mu) \xrightarrow{n \rightarrow +\infty} 0$$

ce qui confirme que pour n grand, \overline{X}_n s'éloigne très peu de μ .

I-4 Application : approximation des probabilités

La loi faible vaut en particulier pour des variables X_k qui suivent une loi de Bernoulli. Dans ce cas, on rappelle que

$$\text{Si } X \hookrightarrow \mathcal{B}(p), \text{ alors } \mu = \mathbb{E}[X] = p$$

Ainsi, le théorème de la loi faible se traduit comme ci-dessous :

Théorème 194 (de Bernoulli)

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires **indépendantes** suivant une même loi de Bernoulli $\mathcal{B}(p)$. Alors, si on pose $\overline{X}_n = \frac{X_1 + \dots + X_n}{n}$, on a

$$\lim_{n \rightarrow +\infty} P(|\overline{X}_n - p| \geq \epsilon) = 0 \quad \forall \epsilon > 0$$

Démonstration :

C'est exactement la loi faible appliquée aux variables de Bernoulli.

En effet, avec les notations de la loi faible, on a $\mu = \mathbb{E}[X] = p$. \square



Remarque :

Traduit en "variable binomiale", cela donne :

Si $(S_n)_{n \in \mathbb{N}^*}$ est une suite de variable suivant respectivement une loi binomiale $\mathcal{B}(n, p)$, alors

$$\lim_{n \rightarrow +\infty} P\left(\left|\frac{S_n}{n} - p\right| \geq \epsilon\right) = 0 \quad \forall \epsilon > 0$$

Commentaires :

Cet exemple justifie en particulier le fait suivant :

$$\text{Si } X \hookrightarrow \mathcal{B}(p), \text{ alors } \overline{X}_n \simeq p \text{ pour } n \text{ grand.}$$

i.e. : si on répète un grand nombre de fois une épreuve de Bernoulli $\mathcal{B}(p)$,

la fréquence du nombre de succès se rapproche fatalement de la probabilité de succès p .

(cf exemple sur les n lancers de pièce déséquilibrée du début du chapitre.) La théorie confirme donc encore une fois l'intuition.

I-5 Application : approximation de la variance

On s'intéresse à la variance empirique $S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X}_n)^2$ d'un échantillon dont on ne connaît pas la variance théorique. On aimerait justifier que

$$S_n^2 \text{ est une approximation de } \sigma^2.$$

Nous allons justifier ce résultat en 2 étapes.

La loi faible donne directement le résultat suivant :

Propriété 195

Si X est une variable admettant un moment d'ordre 4, alors, pour $Y = (X - \mu)^2$, on a

$$\lim_{n \rightarrow +\infty} P(|\overline{Y}_n - \sigma^2| \geq \epsilon) = 0 \quad \forall \epsilon > 0$$

d'où, pour n grand :

$$\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \simeq V(X).$$



Remarque :

On observe d'ailleurs la relation (vraie dès que X admet un moment d'ordre 2) :

$$\mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2\right] = \sigma^2 = V(X)$$

il suffit, pour le prouver, d'appliquer la linéarité de l'espérance à la somme pour l'obtenir, en sachant que $V(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$.

Commentaires :

Sachant que d'après la loi faible, on a pour n grand : $\mu \simeq \overline{X}_n$, si on admet que l'on peut faire les deux approximations successives (attention, la somme a un grand nombre de termes, on pourrait donc peut être avoir beaucoup de perte de précision avec cette approximation...), on pourrait dire que

$$\sigma^2 \simeq \overline{Y}_n \simeq S_n^2$$

ainsi

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X}_n)^2 \simeq \sigma^2.$$

Néanmoins, contrairement à ce qui se passe pour $\mu = \mathbb{E}[\overline{X}_n]$, on a $\mathbb{E}[S_n^2] \neq \sigma^2$. En effet :

Propriété 196

Sous réserve d'existence (X admettant au moins un moment d'ordre 4), on a

$$\mathbb{E}[S_n^2] = \left(\frac{n-1}{n}\right) \sigma^2$$

Démonstration :

$$\begin{aligned} \mathbb{E}[S_n^2] &= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n (X_i - \overline{X}_n)^2\right] = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n (X_i - \mu + \mu - \overline{X}_n)^2\right] \\ &= \frac{1}{n} \sum_{i=1}^n \underbrace{\mathbb{E}[(X_i - \mu)^2]}_{=\sigma^2} + \frac{1}{n} \sum_{i=1}^n \underbrace{\mathbb{E}[(\mu - \overline{X}_n)^2]}_{V(\overline{X}_n) = \frac{\sigma^2}{n}} - 2 \underbrace{\mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n (X_i - \mu)(\overline{X}_n - \mu)\right]}_{\alpha} \\ &= \frac{1}{n} \cdot n \cdot \sigma^2 + \frac{1}{n} \cdot n \cdot \frac{\sigma^2}{n} - 2\alpha \end{aligned}$$

Or,

$$\begin{aligned} \alpha &= \mathbb{E}\left[(\overline{X}_n - \mu) \left(\frac{1}{n} \sum_{i=1}^n (X_i - \mu)\right)\right] = \mathbb{E}\left[(\overline{X}_n - \mu) \left(\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \cdot n\mu\right)\right] \\ &= -\mathbb{E}[(\overline{X}_n - \mu)^2] = -V(\overline{X}_n) = \frac{\sigma^2}{n} \end{aligned}$$

Ainsi, en réinjectant dans la somme plus haut, on obtient :

$$\square \quad \mathbb{E}[S_n^2] = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n} \sigma^2$$

Commentaires :

De la propriété précédente, on tire également

$$\mathbb{E}[S_n^2] \xrightarrow{n \rightarrow +\infty} \sigma^2$$

De plus, on peut montrer (exercice), que $V(S_n^2 - \sigma^2) \xrightarrow{n \rightarrow +\infty} 0$

Ainsi, si n est grand, la variable S_n^2 ne varie que très peu autour de σ^2 , i.e.

$$\text{si } n \text{ est grand, } S_n^2 \simeq \sigma^2$$

⚠ Remarque :

Du résultat précédent, on tire également une autre valeur approchée de σ^2 :

$$\widehat{S}_n^2 = \frac{n}{n-1} S_n^2 = \frac{1}{n-1} \sum_{i=1}^n ((X_i - \overline{X}_n)^2)$$

C'est ce qu'on appelle la "variance corrigée".

Pour information, votre calculatrice vous propose généralement de calculer soit l'écart-type, soit l'écart-type corrigé (souvent notés resp. σ et $\widehat{\sigma}$). Remarquez que la variance corrigée est plus grande que la variance car $\frac{n}{n-1} > 1$, alors $\widehat{S}_n^2 = \frac{n}{n-1} S_n^2 > S_n^2$.

Nous avons vu dans cette partie comment approcher quelques paramètres élémentaires (probabilité, espérance, variance) si on est assuré d'avoir un échantillon de grande taille de notre variable. Nous allons maintenant aborder des résultats plus précis concernant non seulement certains paramètres, mais plus généralement la loi des "variables limites".

II Convergence en loi et variables à supports entiers

Dans cette partie, on va mettre en place des stratégies pour savoir dans quelle mesure deux variables X, Y "ont des lois approximativement identiques."

Dans le cadre des variables discrètes, on pourrait estimer que les lois sont approximativement identiques (on note $\mathcal{L}(X) \simeq \mathcal{L}(Y)$) si

$$\text{Supp}(X) = \text{Supp}(Y) \quad \text{et} \quad P(X = k) \simeq P(Y = k) \quad \forall k \in \text{Supp}(X) = \text{Supp}(Y)$$

Mais qu'en est-il par exemple, si X suit une loi uniforme sur $\llbracket 1, 6 \rrbracket$ et Y suit elle aussi une loi uniforme, mais sur $\{1 - 10^{-4}, \dots, 6 - 10^{-4}\}$ (valeurs extrêmement proches). On a plutôt

$$P(X = k) = \frac{1}{6} = P(Y = k') \quad \forall k \in \llbracket 1, 6 \rrbracket \text{ et } k' \simeq k$$

Là aussi on aurait envie de dire que $\mathcal{L}(X) \simeq \mathcal{L}(Y)$ non ?

Qu'en est-il de plus dans le cadre des variables à densité ? On peut se placer du point de vu de l'histogramme de la loi et penser que deux lois de variables sont proches ssi leurs histogrammes le sont, ce qui, en terme de densité, signifierait

$$f_X(x) \simeq f_Y(x) \quad \forall x \in \mathbb{R}$$

Pour conclure, le dénominateur commun à ces toutes ces approches consiste surtout à penser que les lois deux variables sont proches si

$$P(a < X \leq b) \simeq P(a < Y \leq b) \quad \forall a, b \in \mathbb{R} \quad (a < b)$$

i.e. pour synthétiser :

$$F_X(x) \simeq F_Y(x) \quad \forall x \in \mathbb{R}$$

II-1 Convergence en loi : cas général

Généralement, on a besoin d'approcher une loi par une autre si les données de l'une sont plus compliquées ou longues à calculer que l'autre. Par exemple, si une variable est le résultat d'un grand nombre n de répétitions d'une même expérience. On aura envie de dire que "plus n est grand, plus la loi se rapproche d'une autre". Ceci se traduit par exemple par la notion de "convergence en loi" ci-dessous :

Définition

Soit $(Y_i)_{i \in \mathbb{N}}$ une suite de variables aléatoires de fonctions de répartition respectives F_{Y_i} , $\forall i \in \mathbb{N}$. On dit que (Y_n) converge en loi vers Y et on note $Y_n \xrightarrow{\mathcal{L}} Y$ si

$$\lim_{n \rightarrow +\infty} F_{Y_n}(u) = F_Y(u) \quad \forall u \in \mathbb{R} \text{ où } F_Y \text{ est continue en } u.$$

Ceci traduit en particulier que pour n grand, on a pour tout u :

$$P(Y_n \leq u) \simeq P(Y \leq u)$$

et donc que, pour tout $a, b \in \mathbb{R} \quad (a < b)$:

$$P(a < Y_n \leq b) \simeq P(a < Y \leq b)$$

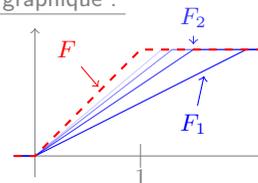
i.e. on peut approcher les probabilités concernant Y_n par celles concernant Y avec un n suffisamment grand.

Exemple 4 :

On considère (X_n) la suite de v.a. telle que $X_n \hookrightarrow \mathcal{U}_{[0; 1 + \frac{1}{n}]}$.

Alors $(X_n) \xrightarrow{\mathcal{L}} X \hookrightarrow \mathcal{U}_{[0; 1]}$:

Version graphique :



où F_i est la fonction de répartition de X_i .

Version calculatoire : On a $F_n(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{x}{1 + \frac{1}{n}} & \text{si } 0 \leq x \leq 1 + \frac{1}{n} \\ 1 & \text{si } x \geq 1 + \frac{1}{n} \end{cases}$

Ainsi, à x fixé, on a :

★ Si $x < 0$,

$$F_n(x) = 0 \quad \text{et donc} \quad \lim_{n \rightarrow +\infty} F_n(x) = 0$$

★ Si $0 \leq x \leq 1$, alors $0 \leq x \leq 1 + \frac{1}{n} \quad \forall n \in \mathbb{N}$ donc $F_n(x) = \frac{x}{1 + \frac{1}{n}}$ et donc

$$\lim_{n \rightarrow +\infty} F_n(x) = x$$

★ Si $x > 1$, alors, pour n assez grand, on a $1 + \frac{1}{n} < x$ et donc

$$F_n(x) = 1 \quad \text{puis} \quad \lim_{n \rightarrow +\infty} F_n(x) = 1$$

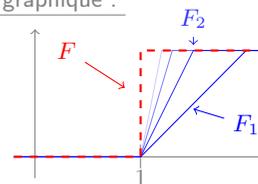
Problème des points de discontinuité de F_Y sur un exemple :

Exemple 5 :

On considère (X_n) la suite de v.a. telle que X_n suit une loi uniforme sur $[1; 1 + \frac{1}{n}]$.

Alors (X_n) converge en loi vers la variable X constante égale à 1 p.s.

Version graphique :



où F_i est la fonction de répartition de X_i .

Version calculatoire : On a $F_n(x) = \begin{cases} 0 & \text{si } x < 1 \\ \frac{x-1}{\frac{1}{n}} & \text{si } 1 \leq x \leq 1 + \frac{1}{n} \\ 1 & \text{si } x \geq 1 + \frac{1}{n} \end{cases}$

Ainsi, à x fixé, on a :

★ Si $x < 1$,

$$F_n(x) = 0 \quad \text{et donc} \quad \lim_{n \rightarrow +\infty} F_n(x) = 0$$

★ Si $x > 1$, alors, pour n assez grand, on a $1 + \frac{1}{n} < x$ et donc

$$F_n(x) = 1 \quad \text{puis} \quad \lim_{n \rightarrow +\infty} F_n(x) = 1$$

★ Si $x = 1$, on remarque que

$$F_n(x) = 0 \quad \forall n \in \mathbb{N}^*$$

alors

$$\lim_{n \rightarrow +\infty} F_n(x) = \begin{cases} 0 & \text{si } x \leq 1 \\ 1 & \text{si } x > 1 \end{cases}$$

mais la fonction ainsi définie n'est pas une fonction de répartition. (Non continue à droite.)

Fort heureusement, ce n'est pas un problème, puisqu'on a a quand même

$$\lim_{n \rightarrow +\infty} F_n(x) = F(x) \quad \forall x \neq 1$$

Notation :

On notera également que pour n grand, si Y_n converge en loi vers Y , alors Y_n suit "à peu près" la loi de Y . On notera ce dernier point dans le cours $Y_n \overset{\sim}{\rightarrow} \mathcal{L}(Y)$.

II-2 Convergence en loi : cas des supports dans \mathbb{N}

Dans le cas des variables discrètes, plus précisément à valeurs entières, la problématique de la fonction de répartition peut être simplifiée :

II.2-a) En général

Proposition 197

Avec les notations de la définition, si $Supp(Y_i) \subset \mathbb{N} \quad \forall i \in \mathbb{N}$ et $Supp(Y) \subset \mathbb{N}$, alors $Y_n \xrightarrow{\mathcal{L}} Y$ ssi

$$P(Y_n = k) \xrightarrow{n \rightarrow +\infty} P(Y = k) \quad \forall k \in \mathbb{N}$$

Démonstration :

- Supposons qu'il y ait convergence en loi.

Comme les supports de Y_n et Y sont entiers, on a :

$$P(Y_n = k) = F_{Y_n} \left(k + \frac{1}{2} \right) - F_{Y_n} \left(k - \frac{1}{2} \right)$$

Or, comme Y est à support entier, F_Y est continue en $(k + \frac{1}{2})$ et $(k - \frac{1}{2})$. Ainsi

$$P(Y_n = k) \xrightarrow{n \rightarrow +\infty} F_Y \left(k + \frac{1}{2} \right) - F_Y \left(k - \frac{1}{2} \right) = P(Y = k)$$

- Supposons que les différentes probabilités convergent.

Soit $x \in \mathbb{R} - \mathbb{N}$. Alors

$$\begin{aligned} F_{Y_n}(x) &= P(Y_n \leq x) = P(Y_n \leq \lfloor x \rfloor) \quad \text{car } Supp(Y_n) \subset \mathbb{N} \\ &= \sum_{k=0}^{\lfloor x \rfloor} P(Y_n = k) \xrightarrow{n \rightarrow +\infty} \sum_{k=0}^{\lfloor x \rfloor} P(Y = k) \\ &= F_Y(\lfloor x \rfloor) = F_Y(x) \quad \text{car } Supp(Y) \subset \mathbb{N} \end{aligned}$$

Ainsi, (Y_n) converge en loi vers Y . \square

II.2-b) Approximation d'une loi binomiale par une loi de Poisson

Théorème 198

Si $(X_n)_{n \in \mathbb{N}^*}$ est une suite de variables aléatoires suivant respectivement une loi $\mathcal{B} \left(n, \frac{\lambda}{n} \right)$, avec $\lambda > 0$, alors elle converge en loi vers une variable suivant la loi de Poisson $\mathcal{P}(\lambda)$, i.e.

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

Démonstration :

Soit $k \in \mathbb{N}$ fixé, $n \geq k$ et $\lambda \in]0; +\infty[$. On a

$$\begin{aligned} P(X_n = k) &= \binom{n}{k} \left(\frac{\lambda}{n} \right)^k \left(1 - \frac{\lambda}{n} \right)^{n-k} = \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n} \right)^k \left(1 - \frac{\lambda}{n} \right)^{n-k} \\ &= \underbrace{\frac{\lambda^k}{k!} \frac{n!}{(n-k)!}}_{A_n} \underbrace{\frac{1}{n^k} \left(1 - \frac{\lambda}{n} \right)^{n-k}}_{B_n} \end{aligned}$$

$$\text{Or,} \quad A_n = \frac{n(n-1) \dots (n-k+1)}{n^k} = \frac{n}{n} \frac{n-1}{n} \dots \frac{n-k+1}{n} \xrightarrow{n \rightarrow +\infty} 1$$

$$\text{et} \quad B_n = \left(1 - \frac{\lambda}{n} \right)^{n-k} = e^{(n-k) \ln \left(1 - \frac{\lambda}{n} \right)}$$

$$\text{Or,} \quad (n-k) \ln \left(1 - \frac{\lambda}{n} \right) \underset{n \rightarrow +\infty}{\sim} -(n-k) \frac{\lambda}{n} \underset{n \rightarrow +\infty}{\sim} -\lambda$$

$$\text{D'où } B_n \xrightarrow{n \rightarrow +\infty} e^{-\lambda} \text{ puis la limite annoncée : } \lim_{n \rightarrow +\infty} P(X_n = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad \square$$

Commentaires :

Généralement les lois binomiales se présentent sous la forme $\mathcal{B}(n, p)$ et non $\mathcal{B}(n, \frac{\lambda}{n})$. Dans les notations du théorème, si n est grand, on aurait donc $\frac{\lambda}{n} = p$ très petit. Le théorème dit donc que si n est grand, $\mathcal{B}(n, p)$ se rapproche de $\mathcal{P}(\lambda)$ avec $\lambda = np$. (D'où : On appelle quelquefois la loi de Poisson la "loi des événements rares".)

En pratique, on estime que, dès que $n \geq 30$ et $p \leq 0,1$, on peut approcher la loi binomiale $\mathcal{B}(n, p)$ par la loi de Poisson $\mathcal{P}(np)$.

■ Exemple 6 :

On considère une variable aléatoire X suivant une loi $\mathcal{B}(50; 0,05)$. On souhaite calculer $P(X = 3)$.

• Calcul exact :

$$P(X = 3) = \binom{50}{3} (0,05)^3 (0,95)^{47} = \frac{50 \times 49 \times 48}{3 \times 2} (0,05)^3 (0,95)^{47} \simeq 0,2199$$

• Calcul approché : On a

$$n = 50 \geq 30 \text{ et } p \leq 0,1.$$

On peut donc approcher $\mathcal{B}(50; p)$ par la loi $\mathcal{P}(np) = \mathcal{P}(2,5)$

D'où

$$P(X = 3) \simeq e^{-2,5} \frac{2,5^3}{3 \times 2} \simeq 0,214$$

Intérêt 1 : On élimine des calculs de factoriels et de puissances $(0,05)^3 (0,95)^{47}$ pouvant donner lieu à des erreurs d'approximation : (Par exemple, en disant $(0,05)^3 \simeq 0$, donc $P(X = 3) \simeq 0$... ce qui est donc faux ici.)

Intérêt 2 : Par ordinateur, on gagne de la vitesse en éliminant un grand nombre de multiplications, (néanmoins au profit d'un calcul d'exponentiel)

? Exercice 1

Supposons donnée une urne contenant deux variétés de boules (rouges et vertes) en quantité totale N . On note V le nombre de boules vertes. On y fait un tirage d'un nombre $n \leq N$ de boules et on veut compter Y_N le nombre de boules vertes.

1. Montrer que

$$P(Y_N = k) = \frac{\binom{V}{k} \binom{N-V}{n-k}}{\binom{N}{n}} \quad \forall k \leq n$$

2. Montrer que si $N \rightarrow +\infty$, à proportion $p = \frac{V}{N}$ de boules vertes constante,

$$P(Y_N = k) \underset{N \rightarrow +\infty}{\sim} \binom{n}{k} p^k (1-p)^{n-k} \quad \forall k \leq n$$

3. En déduire que Y_N converge en loi vers une loi binomiale $\mathcal{B}(n, p)$.

Interprétation : un tirage sans remise se rapproche d'un tirage avec remise dans une très grande urne.

Solution

1. On se donne l'univers des possibilités équiprobables de tirages constitué des poignées de n éléments parmi les N de l'urne. Ainsi, il y a $\binom{N}{n}$ tirages possibles au total. L'événement $(Y_N = k)$ signifie qu'on a tiré une poignée avec k boules rouges : $\binom{k}{N}$ possibilités et en parallèle $N - k$ boules vertes : $\binom{N-k}{n-k}$, d'où la probabilité annoncée.

2. Si $N \rightarrow +\infty$, avec une proportion p de boules vertes constantes, on a nécessairement $V \rightarrow +\infty$ et dépasse nécessairement n au bout d'un certain temps. On peut donc supposer être dans ce cas de figure. Dans ce cas, d'après 1) :

$$\begin{aligned} P(Y_N = k) &= \frac{V!}{k!(V-k)!} \frac{(N-V)!}{(n-k)!(N-V-(n-k))!} \frac{n!(N-n)!}{N!} \\ &= \frac{n!}{k!(n-k)!} \frac{V!}{(V-k)!} \frac{(N-V)!}{(N-V-(n-k))!} \frac{(N-n)!}{N!} \quad (\text{en réorganisant les termes}) \\ &= \binom{n}{k} \frac{\overbrace{V(V-1)\dots(V-k+1)}^{k \text{ termes}} \overbrace{(N-V)(N-V-1)\dots(N-V-(n-k)+1)}^{n-k \text{ termes}}}{\underbrace{N(N-1)\dots(N-n+1)}_{n \text{ termes}}} \end{aligned}$$

Le nombre de termes k , k et n étant constants, on obtient :

$$P(Y_N = k) \sim \binom{n}{k} \frac{V^k (N-V)^{n-k}}{N^n} = \binom{n}{k} \frac{V^k (N-V)^{n-k}}{N^k N^{n-k}} = \binom{n}{k} \left(\frac{V}{N} \right)^k \left(\frac{N-V}{N} \right)^{n-k}$$

3. Comme le terme $\binom{n}{k} p^k (1-p)^{n-k}$ est constant par rapport à N , on a bien

$$\lim_{\substack{N \rightarrow +\infty \\ N, p \in \mathbb{N}}} P(Y_N = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \forall k \leq n$$

ce qui est la convergence en loi annoncée.

Commentaires :

Que se passe-t-il dans les autres cas ? Et bien il existe des résultats sur les variables de type $\overline{X_n}$ extrêmement utiles. Ils s'intègrent dans un théorème s'intitulant "théorème de la limite centrée" ou encore "th. central limite" que l'on retrouve dans notre programme sous deux formes que l'on découvrira ci-dessous.

III Théorème central limite : deux formes

III-1 Si on connaît σ^2

III.1-a) Le théorème ; première forme

Rappelons la définition suivante :

Définition

Soit Y une variable aléatoire réelle admettant une variance non nulle. Alors, lorsque l'on note

$$\mu = E(Y), \quad \sigma^2 = V(Y) \quad \text{et} \quad Y^* = \frac{Y - \mu}{\sigma}$$

on appelle Y^* la *variable centrée réduite* associée à Y . (car $E(Y) = 0$ et $V(Y) = 1$).

Commentaires :

Le théorème ci-dessous va établir de quelle manière on pourra supposer que la loi de \bar{X}_n^* peut être remplacée par la loi $\mathcal{N}(0, 1)$, quel que soit la variable X de départ !! Ce résultat est donc extrêmement puissant...

Théorème 199 central limite (ou de la limite centrée) ; première forme

Soit (X_1, \dots, X_n) un n -échantillon d'une variable aléatoire X de variance non nulle. On note

$$\mu = E(X), \quad \sigma^2 = V(X), \quad \bar{X}_n = \frac{X_1 + \dots + X_n}{n} \quad \text{et} \quad \bar{X}_n^* = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

on obtient que

$$\bar{X}_n^* \text{ converge en loi vers } Y \leftrightarrow \mathcal{N}(0, 1)$$

i.e. pour tout $a, b \in \mathbb{R}$ avec $a < b$,

$$\lim_{n \rightarrow +\infty} P(a < \bar{X}_n^* \leq b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2} dt = \Phi(b) - \Phi(a)$$

où Φ est la fonction de répartition associée à la loi $\mathcal{N}(0, 1)$.

Démonstration :

admise \square

Remarque :

On rappelle que $\bar{X}_n^* = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ est la loi centrée réduite de \bar{X}_n car

$$\begin{aligned} \mathbb{E}[\bar{X}_n] &= \frac{1}{n} \sum_{k=1}^n \mathbb{E}[X_k] = \frac{1}{n} \sum_{k=1}^n \mu = \frac{n}{n} \mu = \mu \\ \mathbb{V}[\bar{X}_n] &\stackrel{\text{indépendance des } X_i}{=} \underbrace{\frac{1}{n^2} \sum_{k=1}^n \mathbb{V}[X_k]}_{\text{indépendance des } X_i} = \frac{1}{n^2} \sum_{k=1}^n \sigma^2 = \frac{n}{n^2} \sigma^2 = \frac{1}{n} \sigma^2 \end{aligned}$$

i.e. :

$$\forall n \in \mathbb{N}, \quad \mathbb{E}[\bar{X}_n^*] = 0, \quad \text{et} \quad V(\bar{X}_n^*) = 1$$

Il est donc cohérent que la loi de la variable \bar{X}_n^* s'approche d'une loi d'espérance 0 et de variance 1

Remarque :

Le TCL ayant comme conséquence le fait que pour n grand,

$$\mathcal{L}(\bar{X}_n^*) \simeq \mathcal{N}(0, 1),$$

par les propriétés des lois normales, on pourrait dire également que

$$\mathcal{L}(\bar{X}_n) \simeq \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

ou encore $\mathcal{L}\left(\sum_{k=1}^n X_k\right) \simeq \mathcal{N}(n\mu, n\sigma^2)$

mais **ATTENTION**, on ne dit certainement pas que chacun des X_n suit approximativement une loi normale. ça peut en effet être complètement faux...

Exemple 7 :

On choisit 500 fois au hasard un nombre compris entre 0 et 1. Quelle est la probabilité approximative que la somme de ces nombres soit comprise entre 240 (strictement) et 260 ?

On introduit les variables aléatoires X_i correspondant au nombre obtenu au $i^{\text{ème}}$ choix. Alors, (X_1, \dots, X_{500}) est un 500-échantillon de $X \leftrightarrow \mathcal{U}([0; 1])$

En posant

$$T_n = X_1 + \dots + X_{500},$$

la question revient à chercher $P(240 < T_n \leq 260)$.

Résolution exacte ?

Ne connaissant pas la loi de T_n , on pourrait peut être la calculer, mais ceci signifie qu'il faudrait appliquer la formule de convolution 499 fois. Qui se lance ... ?

Résolution approchée : (à l'aide du TCL) Posons

$$\mu = E(X_1) = 1/2 \quad \text{et} \quad \sigma^2 = \frac{1}{12}$$

on a alors, $\frac{T_n/n - \mu}{\sigma/\sqrt{n}} \rightsquigarrow \mathcal{N}(0, 1)$ d'où

$$\begin{aligned} P(240 < T_n \leq 260) &= P\left(\frac{240/n - \mu}{\sigma/\sqrt{n}} < \frac{T_n/n - \mu}{\sigma/\sqrt{n}} \leq \frac{260/n - \mu}{\sigma/\sqrt{n}}\right) \\ &= P\left(\underbrace{\frac{240/500 - 0,5}{1/\sqrt{500} \times 12}}_{\alpha \simeq -1,55} < \frac{T_n/n - \mu}{\sigma/\sqrt{n}} \leq \underbrace{\frac{260/500 - 0,5}{1/\sqrt{500} \times 12}}_{\beta \simeq 1,55}\right) \end{aligned}$$

D'après le TCL, on peut donc estimer que

$$P(240 < T_n \leq 260) \simeq \phi(1,55) - \phi(-1,55) = 2\phi(1,55) - 1 \simeq 0.879$$

III.1-b) Appl.1 : approximation d'une loi binomiale par une loi normale

Commentaires :

Voyons le cas particulier de X qui suit une loi de Bernoulli de paramètre p , où on sait alors que $T_n = X_1 + \dots + X_n$ suit une loi binomiale $\mathcal{B}(n, p)$

Théorème 200 de Moivre-Laplace

Soit $(T_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires qui suivent respectivement une loi $\mathcal{B}(n, p)$, où $p \in]0; 1[$. Alors,

$$T_n^* \xrightarrow{\mathcal{L}} Y \hookrightarrow \mathcal{N}(0, 1)$$

i.e., pour tous les $a, b \in \mathbb{R}$ où $a < b$, on a

$$P\left(a < \underbrace{\frac{T_n - np}{\sqrt{np(1-p)}}}_{T_n^*} \leq b\right) \xrightarrow{n \rightarrow +\infty} \phi(b) - \phi(a) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{t^2}{2}} dt$$

Démonstration :

C'est tout simplement la traduction du théorème central limite à une suite de variables de Bernoulli. En effet, on pose (X_1, \dots, X_n) un n -échantillon d'une variable $X \hookrightarrow \mathcal{B}(p)$, de manière à ce que $T_n = X_1 + \dots + X_n$. Alors, d'après le TCL, on

$$P(a < X_n^* \leq b) \xrightarrow{n \rightarrow +\infty} \Phi(b) - \Phi(a)$$

$$\text{Or } X_n^* = \frac{T_n/n - p}{\sigma/\sqrt{n}} = \frac{T_n - np}{\sqrt{n}\sigma} = \frac{T_n - np}{\sqrt{np(1-p)}} \quad \square$$

Commentaires :

En traduisant ce théorème, ceci signifie que la loi de T_n^* peut être approchée par une loi $\mathcal{N}(0; 1)$ si n est "grand". Or,

$$T_n = \sigma\sqrt{n} \cdot T_n^* + n\mu, \quad \text{avec } \sigma^2 = p(1-p)$$

Donc la loi de T_n (i.e. $\mathcal{B}(n, p)$) peut être approchée par la loi $\mathcal{N}\left(\underbrace{np}_{\mathbb{E}[T_n]}; \underbrace{n\sigma^2}_{\mathbb{V}[T_n]}\right)$

Dans la pratique, on estime que l'on peut approcher la loi de T_n^* par la loi $\mathcal{N}(0, 1)$ ou $\mathcal{B}(n, p)$ par la loi $\mathcal{N}(np, np(1-p))$ dès que $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$.

■ Exemple 8 :

On lance un dé équilibré 100 fois. On souhaite approximer la probabilité pour que le nombre de 3 soit compris entre 20 et 30 (au sens large).

On note T_n la variable aléatoire donnant le nombre de 3 dans la série de lancers. Alors

$$T_n \rightsquigarrow \mathcal{B}(n, p), \quad \text{où } p = 1/6.$$

On a $n \geq 30$, $np = 100/6 (\simeq 17) \geq 5$ et $n(1-p) = 100 \times 5/6 (\simeq 83) \geq 5$.

D'après le théorème de Moivre-Laplace (ou d'après le TCL), la loi de T_n peut être approchée par $\mathcal{N}(np, np(1-p))$. Or :

$$P(20 \leq T_n \leq 30) = P(19 < T_n \leq 30)$$

Si votre calculatrice le permet, vous pouvez donc dès à présent faire ce calcul. Sinon, on passe à la loi normale centrée réduite en constatant que $T_n^* = \frac{T_n - np}{\sqrt{np(1-p)}}$ peut être considérée comme suivant par la loi normale centrée réduite $\mathcal{N}(0; 1)$. D'où

$$\begin{aligned} P(20 \leq T_n \leq 30) &= P\left(\frac{19 - np}{\sqrt{np(1-p)}} < T_n^* \leq \frac{30 - np}{\sqrt{np(1-p)}}\right) \\ &= P\left(\underbrace{\frac{19 - 100/6}{\sqrt{100 \times 5/36}}}_{\alpha \simeq 0,626} < T_n^* \leq \underbrace{\frac{30 - 100/6}{\sqrt{100 \times 5/36}}}_{\beta \simeq 3,578}\right) \\ &\left(\simeq P(\alpha < Z \leq \beta) \quad \text{où } Z \rightsquigarrow \mathcal{N}(0; 1) \right) \end{aligned}$$

Ainsi, d'après les tables de la loi normale (ou avec votre calculatrice)

$$P(\alpha < T_n^* \leq \beta) \simeq \phi(\beta) - \phi(\alpha) \simeq \phi(3,578) - \phi(0,626) \simeq 0,9998 - 0,7357 \simeq 0,2651$$

On peut également formuler ceci légèrement différemment à l'aide cette fois ci des proportions :

Corollaire (thm de Moivre-Laplace, 2^{ème} version)

Soit $(F_n)_{n \in \mathbb{N}^*}$ La fréquence d'obtention d'un événement A de probabilité p , dans une suite de n répétitions identiques et indépendantes d'une même expérience. alors, quand

$$n \geq 30, np \geq 5 \text{ et } n(1-p) \geq 5,$$

on estime que

$$\mathcal{L}\left(\frac{F_n - p}{\sqrt{\frac{p(1-p)}{n}}}\right) \simeq \mathcal{N}(0, 1), \quad \text{i.e.} \quad \mathcal{L}(F_n) \simeq \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$$

■ Exemple 9 :

On lance une pièce équilibrée $n = 100$ fois. On cherche à savoir quelle sera la probabilité d'avoir au moins 1/4 des tirages qui soient des 1.

On note F_n la proportion de 1 dans la série de lancés. Alors, d'après le théorème de Moivre Laplace, comme on lance n fois de manière identique et indépendante, que la probabilité d'obtenir 1 à chaque lancer est $p = \frac{1}{6}$, que pour finir

$$n \geq 30, \quad np \simeq 17 \geq 5, \quad n(1-p) \simeq 84 \geq 5$$

On en déduit qu'on peut approcher la loi de F_n par $\mathcal{N}\left(p, \frac{p(1-p)}{n}\right) = \mathcal{N}\left(\frac{1}{6}, \frac{5/36}{100}\right)$. D'où, directement à la calculatrice (ou en passant à la loi normale centrée réduite)

$$P(0.25 \leq F_n) = 1 - P(F_n < 0.25) \simeq 0.0127$$

Très faible! Mieux vaut ne pas compter dessus...

Correction de continuité

Lorsque l'on approche une loi discrète par une loi continue, on a un problème d'approximation du type suivant : Si X suit une loi discrète (par exemple entière), on a par exemple

$$P(10 \leq X \leq 15) = P(9,9 < X \leq 15,3) = \dots$$

Évidemment, l'approximation s'en trouve donc légèrement modifiée. On souhaite alors équilibrer l'erreur obtenue "de chaque côté de X ". La solution la moins douloureuse en général consiste donc à introduire une *correction de continuité*, c'est-à-dire, si n et m sont deux entiers, on écrira

$$P(n \leq X \leq m) = P(n - 0,5 < X \leq m + 0,5)$$

Cependant, cette manipulation n'est pas exigible au concours. Dans le cas ci-dessus, toute valeur issue de $P(10 \leq X \leq 15) = P(9,9 < X \leq 15,3) = P(9 < X < 16) = \dots$ sera acceptée.

■ Exemple 10 :

On reprend l'exemple des 100 lancés de dés en utilisant une correction de continuité.

$$\begin{aligned} P(20 \leq T_n \leq 30) &= P(19,5 < T_n \leq 30,5) = P\left(\frac{19,5 - np}{\sqrt{np(1-p)}} < T_n^* \leq \frac{30,5 - np}{\sqrt{np(1-p)}}\right) \\ &= P\left(\frac{19,5 - 100/6}{\sqrt{100 \times 5/36}} < T_n^* \leq \frac{30,5 - 100/6}{\sqrt{100 \times 5/36}}\right) \\ &\left(\simeq P(\alpha < Z \leq \beta) \quad \text{où } Z \sim \mathcal{N}(0; 1) \right) \end{aligned}$$

Ainsi, d'après les tables de la loi normale,

$$P(\alpha < Y_n^* \leq \beta) \simeq \phi(\beta) - \phi(\alpha) \simeq \phi(3,712) - \phi(0,760) \simeq 0,9999 - 0,7764 \simeq 0,2235$$

À titre d'information :

- le calcul exact sur la loi binomiale effectué par ordinateur donne environ 0,2195.
- Sans correction de continuité :

$$P(20 \leq T_n \leq 30) \simeq 0.185 \quad ; \quad P(19 < T_n \leq 30) = 0.265 \quad ; \quad P(20 \leq T_n < 31) = 0.185$$

III.1-c) Appl 2 : approximation d'une loi de Poisson par une loi normale

Théorème 201

Soit $(T_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires qui suivent respectivement une loi $\mathcal{P}(np)$, où $p > 0$. Alors,

$$T_n^* \xrightarrow{\mathcal{L}} Y \hookrightarrow \mathcal{N}(0, 1)$$

i.e., pour tous les $a, b \in \bar{\mathbb{R}}$ où $a < b$, on a

$$P(a < T_n^* \leq b) \xrightarrow{n \rightarrow +\infty} \phi(b) - \phi(a).$$

Démonstration :

Encore une fois, ce n'est que l'application du TCL à la suite $(T_n)_{n \in \mathbb{N}^*}$.

En effet, on peut écrire que pour tout $n \in \mathbb{N}$ $T_n = X_1 + \dots + X_n$, où $(X_n)_{n \in \mathbb{N}^*}$ est une suite de variables aléatoires mutuellement indépendantes, de même loi de Poisson $\mathcal{P}(\lambda)$ (et donc de variance $\sigma^2 = \lambda$ non nulle.) \square

- En pratique, on estime que l'on peut approcher la loi de T_n^* par $\mathcal{N}(0; 1)$ si $np \geq 18$.
- Pour $\lambda \geq 18$, ceci signifie également que la loi $\mathcal{P}(\lambda)$ peut être approchée par $\mathcal{N}(\lambda, \lambda)$.

■ Exemple 11 :

On pose $n = 40$ et $\lambda = 20$. On suppose que X suit une loi

$$\mathcal{P}(20).$$

On cherche $P(X \leq 17)$.

Dans notre calculatrice ou Python, on peut obtenir que

$$P(X \leq 17) \simeq 0.2970$$

L'approximation par loi normale donne :

$$\begin{aligned} P(X \leq 17) &= P(X \leq 17,5) && \text{(correction de continuité)} \\ &= P\left(\frac{X - np}{\sqrt{np}} \leq \frac{17,5 - np}{\sqrt{np}}\right) = P\left(\frac{X - np}{\sqrt{np}} \leq \underbrace{\frac{17,5 - 20}{\sqrt{20}}}_{\alpha \simeq -0.5590}\right) \\ &\simeq 0.2881 \end{aligned}$$

III.1-d) Les approximations en bref

Condition	On peut approcher	par
$n \geq 30$ et $p \leq 0,1$	$\mathcal{B}(n, p)$	$\mathcal{P}(np)$
$n \geq 30$, $np \geq 5$ et $n(1 - p) \geq 5$	$\mathcal{B}(n, p)$	$\mathcal{N}(np, np(1 - p))$
$\lambda \geq 18$	$\mathcal{P}(\lambda)$	$\mathcal{N}(\lambda, \lambda)$.

Commentaires :

Astuce pour se souvenir des paramètres à appliquer dans les approximations :

Les lois ci-dessus par lesquelles on approche admettent toutes des paramètres que l'on peut retrouver grâce à leur espérance et variance. (Par exemple, pour $\mathcal{P}(np)$, son espérance est np .)

Pour retrouver ces paramètres, il suffit d'aller les chercher dans la variable de départ. Dans l'ordre ci-dessus :

$$\begin{aligned} \underbrace{\mathcal{B}(n, p)}_{\text{espérance } np} &\longrightarrow \underbrace{\mathcal{P}(np)}_{\text{paramètre} = \text{espérance } np} \\ \underbrace{\mathcal{B}(n, p)}_{\text{espérance } np, \text{ variance } np(1-p)} &\longrightarrow \underbrace{\mathcal{N}(np, np(1-p))}_{\text{espérance } np, \text{ variance } np(1-p)} \\ \underbrace{\mathcal{P}(\lambda)}_{\text{espérance} = \text{variance} = \lambda} &\longrightarrow \underbrace{\mathcal{N}(\lambda, \lambda)}_{\text{espérance} = \text{variance} = \lambda} \end{aligned}$$

⚠ Remarque :

Si la situation le permet, il est également possible d'enchaîner les approximations. Par exemple, (*cas "extrême"*) : On pioche n boules dans une très grande urne avec remise et on compte X le nombre de "bonnes" boules. si $n \geq 30$ et $p \leq 0,1$:

$\mathcal{B}(n, p)$ peut être approchée par $\mathcal{P}(np)$.

Rien ne s'oppose à ce qu'on ait également $np \geq 18$. Ainsi,

$\mathcal{P}(np)$ peut être approchée par $\mathcal{N}(np, np)$.

Au final, on peut donc supposer que

X suit une loi $\mathcal{N}(np, np)$

On ne s'attendra néanmoins pas à ce que les résultats soient très précis. Par exemple, si on a également $n(1 - p) \geq 5$, alors il vaut mieux passer directement de la loi binomiale à la loi normale, car dans ce cas, on tombe sur $\mathcal{N}(np, np(1 - p))$, dont la variance est plus petite que $\mathcal{N}(np, np)$.

III-2 Si on ne connaît pas σ^2 : Deuxième forme du TCL

Commentaires :

La première version du TCL utilise l'espérance μ et la variance σ de la variable X . Or, dans de multiples cas, contrairement aux exemples de la partie précédente où on connaît σ , on ne dispose pas nécessairement de toutes les données théoriques lorsqu'on étudie un caractère sur un échantillon de population. D'après la première partie, on sait néanmoins que l'on peut approcher l'espérance et la variance à l'aide (respectivement) de \overline{X}_n et S_n^2 (notation de la partie I.)

On va alors remarquer que le TCL est encore valable en remplaçant simplement σ par S_n :

Théorème 202 TCL (deuxième forme)

Soit (X_1, \dots, X_n) un n -échantillon d'une variable aléatoire X . On note

$$\mu = E(X), \quad \overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i, \quad \text{et} \quad S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X}_n)^2 = \overline{X_n^2} - \overline{X}_n^2$$

on obtient que

$$\frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \text{ converge en loi vers } Y \hookrightarrow \mathcal{N}(0, 1)$$

i.e., pour tous $a, b \in \mathbb{R}$ avec $a < b$,

$$\lim_{n \rightarrow +\infty} P\left(a < \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2} dt = \phi(b) - \phi(a)$$

où ϕ est la fonction de répartition associée à la loi $\mathcal{N}(0; 1)$.

⚠ Remarque :

Par rapport à la première version, on a remplacé σ par S_n . Ceci revient à remplacer

$$\overline{X_n} * \text{ par } \frac{\overline{X_n} - \mu}{S_n/\sqrt{n}}.$$

Remarquez que $\overline{X_n} *$ était la "centrée réduite" de $\overline{X_n}$, ce qui n'est plus le cas de $\frac{\overline{X_n} - \mu}{S_n/\sqrt{n}}$. Le TCL est toutefois encore valable mais néanmoins, il y a fort à parier que les approximations soient moins bonnes!

? Exercice 2

On a planté deux types de graines A et B pour lesquels on note respectivement $\mu_A, \sigma_A, \mu_B, \sigma_B$ leur moyennes et écart-type théoriques (inconnus) au jour 100 de leur croissance. Après mesure sur un échantillon de 50 graines chacune, notant m_A, s_A, m_B, s_B leur moyennes et écart-type empirique au jour 100 de leur croissance, on trouve

$$m_A = 15,8, \quad s_A = 2, \quad m_B = 17, \quad s_B = 3$$

1. Déterminer, pour la variété A , une valeur a telle que

$$P\left(m_A - a \frac{s_A}{\sqrt{n}} < \mu_A < m_A + a \frac{s_A}{\sqrt{n}}\right) \simeq 0.95$$

2. En déduire un intervalle I_A tel que

$$P(\mu_A \in I_A) = 0.95$$

3. Trouver de même un intervalle I_B correspondant à

$$P(\mu_B \in I_B) = 0.95$$

4. Peut-on répondre à la question "est-ce que $\mu_A = \mu_B$?" avec une probabilité raisonnable 0.95 d'avoir raison?

Solution

On observe que

$$P\left(m_A - a \frac{s_A}{\sqrt{n}} < \mu_A < m_A + a \frac{s_A}{\sqrt{n}}\right) = P\left(-a \leq \frac{m_A - \mu_A}{s_A/\sqrt{n}} \leq a\right)$$

D'après la deuxième forme du TCL, on sait que

$$P\left(-a \leq \frac{m_A - \mu_A}{s_A/\sqrt{n}} \leq a\right) \simeq \Phi(a) - \Phi(-a) = 2\Phi(a) - 1$$

Ainsi,

$$P\left(m_A - a \frac{s_A}{\sqrt{n}} < \mu_A < m_A + a \frac{s_A}{\sqrt{n}}\right) = 0,95 \Leftrightarrow 2\Phi(a) - 1 \simeq 0.95 \Leftrightarrow a \simeq 1,96$$

On trouve alors l'intervalle

$$I_A \simeq [15,25; 16,35]$$

Les mêmes calculs avec les valeurs de B donnent

$$I_B \simeq [16.17; 17.83]$$

Ce qui signifie que

$$\mu_A \in I_A, \quad \mu_B \in I_B, \quad I_A \cap I_B \neq \emptyset$$

On pourrait très bien avoir $\mu_A = \mu_B$ mais on n'est pas certain! **on ne peut donc pas conclure...**

Corollaire *Application à la loi binomiale*

Si $T_n \hookrightarrow \mathcal{B}(n, p)$. Alors, pour n grand, si on note r la proportion empirique de succès, on peut dire estimer que

$$\frac{T_n - np}{\sqrt{nr(1-r)}} \hookrightarrow \mathcal{N}(0, 1)$$

Démonstration :

Comme déjà utilisé précédemment, T_n peut s'écrire $T_n = X_1 + \dots + X_n$ où $X_i \hookrightarrow \mathcal{B}(p)$ qui rend 1 si un succès arrive au rang i .

Ainsi, on sait d'après le TCL (deuxième forme), comme $\mathbb{E}[X_i] = p$, on peut estimer que

$$\frac{\overline{X_n} - p}{S_n/\sqrt{n}} \hookrightarrow \mathcal{N}(0, 1)$$

En multipliant par n le numérateur et dénominateur :

$$\frac{T_n - np}{S_n \sqrt{n}} \hookrightarrow \mathcal{N}(0, 1)$$

Par définition dans le TCL de la variance empirique :

$$S_n^2 = \overline{X_n^2} - \overline{X_n}^2$$

Or

$$\overline{X_n^2} = \left(\frac{T_n}{n}\right)^2 = p'^2 \quad \text{et} \quad X_i^2 = X_i$$

d'où

$$\overline{X_n^2} = \overline{X_n} = \frac{T_n}{n} = p'$$

Ainsi

$$S_n^2 = p' - p'^2 = p'(1 - p')$$

ce qui donne la conclusion annoncée. \square

Commentaires :

Afin de retenir cette formule, on peut remarquer que la variance correspond exactement à la variance empirique.

Dans la pratique, on estime qu'on peut approcher la loi $\mathcal{B}(n, p)$ par la loi $\mathcal{N}(np, nr(1-r))$ dès que $n \geq 30$, $nr \geq 10$, $n(1-r) \geq 10$,
(Conditions plus contraignantes que pour une approximation grâce au TCL1)



Remarque :

Comme dans la première version, (thm de Moivre Laplace) on peut traduire ceci en version "calcul de fréquence" en posant $F_n = \frac{T_n}{n}$, qui représente la fréquence d'obtention d'un événement A dans une série de n répétitions indépendantes d'une même expérience. Ainsi, on pourrait estimer que dans les conditions pré-citées, on aurait

$$\mathcal{L}\left(\frac{F_n - p}{\sqrt{\frac{r(1-r)}{n}}}\right) \simeq \mathcal{N}(0, 1), \quad \text{i.e.} \quad \mathcal{L}(F_n) \simeq \mathcal{N}\left(p, \frac{r(1-r)}{n}\right)$$

ce qui est particulièrement utile quand on veut estimer des probabilités!

? Exercice 3

Un institut de sondage a été contacté afin de tenter de prévoir le gagnant avant la fin de la totalité du dépouillement, le soir du second tour des élections présidentielles. Les bureaux de vote ferment normalement à 19h, mais certains ont une dérogation préfectorale afin d'ouvrir jusqu'à 20h. À cette heure là, l'institut dispose donc déjà d'une grande partie des données déjà dépouillées. Les journaux ont donc pris l'habitude d'annoncer un potentiel gagnant dès l'ouverture du journal de 20h.

Pour ce faire, on note p la proportion finale de français ayant voté pour le candidat A , n le nombre de votes déjà connus et T_n le nombre de français ayant voté pour A sur les n connus.

1. Soit r la proportion empirique de votes pour A qu'on estime $> 10\%$. Déterminer x tel que

$$P\left(\left|\frac{T_n - np}{\sqrt{nr(1-r)}}\right| \leq x\right) \simeq 0,99$$

2. À 19h10, on effectue d'abord une première estimation après $n = 2000$ bulletins déjà dépouillés. On obtient 1019 bulletins en faveur de A . Déterminer la proportion empirique de succès. A-t-on envie de déclarer la candidat A gagnant ?

3. Déterminer maintenant un intervalle dans lequel se trouve p avec une probabilité de 99%. Peut-on donc véritablement affirmer que A sera gagnant ?

4. Vers 19h30, on dispose déjà de $n = 200\,000$ résultats. On obtient maintenant 102\,234 bulletins en faveur de A . Quelle est la margeur d'erreur à 99% en annonçant que le candidat A a une probabilité r de l'emporter ? Peut-on enfin raisonnablement déclarer A gagnant ?

5. Est-ce encore le cas avec une probabilité de 0,999 ?

Solution

1. D'après le TCL, comme n est très grand ($n \geq 30$, $nr \geq 10$, $n(1-r) \geq 10$), on peut estimer que la variable dans la probabilité suit une loi normale $\mathcal{N}(0, 1)$. Ainsi, on obtiendra

$$x \simeq 2.5758$$

2. On a ici $r = \frac{T_n}{n} = \frac{1019}{2000} = 0.5095 > 0.5$. On pourrait penser que A sera gagnant.

3. D'après le calcul de la question 1, après en isolant p au centre de l'inégalité, on obtient que

$$P\left(r - x \frac{\sqrt{r(1-r)}}{\sqrt{n}} \leq p \leq r + x \frac{\sqrt{r(1-r)}}{\sqrt{n}}\right) \simeq 0,99$$

Ce qui, avec les chiffres annoncés ici donne l'intervalle approximatif demandé

$$[0.4983; 0.5207]$$

Ainsi, on peut très bien trouver en réalité $p < 0,5$. On ne peut pas affirmer que p sera gagnant.

4. La proportion empirique de succès est maintenant

$$r = 0.51117$$

avec une marge d'erreur de

$$x \frac{\sqrt{r(1-r)}}{\sqrt{n}} \simeq 0,00112$$

Ainsi, on est certain à 99% que le candidat A a une probabilité de remporter les élections supérieure à $r - 0,00112 \simeq 0,51 > 0.5$. On peut maintenant envisager de le déclarer gagnant.

5. On obtient cette fois-ci $x \simeq 3.29053$. L'intervalle devient donc $[0.5100, 0.5123]$. C'est donc encore le cas!

IV Introduction aux tests de conformité

Commentaires :

Le TCL a de multiples autres utilités issues de l'approximation par une loi normale. Il permet par exemple

- de prévoir dans quels intervalles on va trouver les données empiriques ;
- ou au contraire, d'extrapoler des intervalles dans lesquelles se situent les données théoriques à partir d'échantillons (ex. exercice de la partie précédente).

Il existe d'énormes variétés de procédures que l'on peut mettre en place pour tester un grand nombre de choses, mais à votre programme se trouve explicitement les "tests de conformité", qui sont liés au premier point.

IV-1 Principe d'un test de conformité sur un exemple

Globalement, un test de conformité consiste à comparer un échantillon avec la population théorique de référence et à tenter d'émettre un avis sur le fait que l'échantillon corresponde ou non à ladite population théorique.

La procédure est la suivante :

- On pose deux hypothèses opposées (classiquement nommées H_0 et H_1) afin de tester laquelle est la bonne (nous préciserons ceci plus tard) ;
- On effectue certains calculs avec les données dont on dispose ;
- On tente d'apporter une conclusion à la problématique.

■ Exemple 12 :

On considère deux variétés de fraises différentes, conditionnées sur deux palettes distinctes pour leur vente. Le responsable de l'étiquetage a peur d'avoir confondu les deux palettes. Il sait néanmoins que les deux variétés ont des poids moyens significativement différents et que sur une des deux palettes, il y a des gariguettes dont il connaît le poids moyen par fruit : 16g.

Sa seule manière d'identifier les palettes est de faire un prélèvement d'échantillon (suffisamment conséquent).

On pose alors les hypothèses :

H_0 : "La palette n° 1 est celle des gariguettes" (*hypothèse précise*)

H_1 : "La palette n° 1 n'est pas celle des gariguettes" (*hypothèse contraire*)

Sur la palette testée, il trouve un poids moyen empirique de 16,5g.

Est-ce ou non la palette de gariguette ? La suite de cette partie consiste à trouver des outils donnant des éléments de réponse à cette question.

Commentaires :

Nous verrons dans cette partie que le travail de test de conformité se rapproche d'un raisonnement par l'absurde. Les calculs nous permettant (avec nos connaissances actuelles en Bcpst) de donner un élément de réponse sont issus du TCL (1^{ère} ou 2^{ème} forme selon besoin).

On rappelle ainsi que, si n est grand

$$\overline{X}_n^* \xrightarrow{\mathcal{L}} Z \hookrightarrow \mathcal{N}(0,1)$$

et de manière plus large

$$\frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \xrightarrow{\mathcal{L}} Z \hookrightarrow \mathcal{N}(0,1)$$

Notre but est donc de ramener chaque problématique à un problème de moyenne afin d'utiliser le TCL pour établir des résultats.

? Exercice 4

Revenons à notre exemple des 2 palettes de fraises. Supposons que les poids des gariguettes ait un écart type de $\sigma = 2g$ et que le responsable de l'étiquetage ait prélevé aléatoirement $n = 100$ fraises sur la palette n° 1. On note X_i le poids de la fraise i et on pose l'hypothèse

H_0 : "La palette n° 1 est celle des gariguettes"

1. Sous l'hypothèse H_0 , quelle est la valeur de a qui correspond à la probabilité $P\left(\left|\frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}\right| \leq a\right) \simeq 0.95$

En déduire un intervalle I pour lequel $P(\overline{X}_n \in I) \simeq 0.95$.

2. La valeur moyenne empirique effectivement observée sur l'échantillon prélevée est $m_A = 16,5$. Permet-elle de savoir quelle est la palette de gariguette ?

Solution

1. Pour la probabilité 0.95, D'après d'après le TCL, avec $2\Phi(a) - 1 \simeq 0.95$, la calculatrice nous donne

$$\Phi(a) = \frac{1 + 0,95}{2}, \quad \text{d'où } a \simeq 1,96.$$

Ainsi, en développant : l'intervalle I est le suivant :

$$I \simeq \left[\mu - a \frac{\sigma}{\sqrt{n}}; \mu + a \frac{\sigma}{\sqrt{n}} \right] \simeq [15.61; 16.39]$$

2. Comme $16,5 \notin I$, on peut en déduire avec une probabilité d'au moins 95% d'avoir raison, que l'on peut rejeter l'hypothèse H_0 . Ce n'est pas la palette de gariguette. Par élimination, la palette de gariguettes est donc la deuxième.

Explications sur l'exercice et l'exemple :

Posons $\mu = 16$ le poids théorique des gariguettes et μ_2 le poids théorique correspondant à la palette étudiée. Nous souhaitons en fait conclure ici sur les hypothèses

H_0 : "c'est la palette de gariguette", H_1 : "ce n'est pas la palette de gariguette" que l'on remplace ici par

$$H_0 : " \mu = \mu_2 ", \quad H_1 : " \mu \neq \mu_2 "$$

Dans le cadre de l'exercice précédent, on a fait un **un raisonnement par l'absurde**, en supposant \mathcal{H}_0 vraie. Sous cette hypothèse, on a trouvé :

"C'est la palette de gariguette" \Rightarrow "95% des valeurs possibles de \overline{X}_n sont dans I "

Or, m_A est une valeur (celle effectivement obtenue) de la variable \overline{X}_n et on a trouvé

$$m_A = 16,5 \notin I.$$

Comme on trouve finalement une valeur m_A de \overline{X}_n qui est très peu probable (moins de 5% de chance de tomber en dehors de I), on estime avoir une **contradiction**. Ainsi, l'hypothèse initiale \mathcal{H}_0 est fautive. Ainsi, on en déduit que **ce n'est pas la palette escomptée**, H_1 est vraie.

Commentaires :

Qu'en est-il si en revanche la valeur empirique de \overline{X}_n est dans l'intervalle I ? Que peut-on en déduire ?

■ Exemple 13 :

Reprenons l'exercice précédent sur les fraises avec $m_A = 16,3$. Peut-on conclure sur le fait que ce soit ou non la palette de gariguettes ?

On avait $I = [15.61; 16.39]$ Cette fois-ci, on a $16,3 \in I$. **On ne peut donc pas exclure que ce soit la palette de gariguette** mais peut-on affirmer en conséquence que c'est bien la palette de gariguettes ?

En fait, non. Pour comprendre ceci, relisez les explications faites un peu plus haut et souvenez-vous que dans un raisonnement par l'absurde, on ne peut conclure si on n'aboutit à une contradiction !

On ne peut donc en réalité rien décider.

⚠ PEUT-ON TROUVER UNE SOLUTION POUR DÉCIDER SI H_0 EST VRAIE ?

Avec les techniques précédentes, afin d'accepter l'hypothèse H_0 : "C'est la palette de gariguette", il nous faudrait donc **rejeter** H_1 : "Ce n'est pas la palette de gariguette". Or, le problème est que nous n'avons aucun cadre précis de calcul en partant de H_1 ! Impossible donc de fournir une formule de comparaison du type précédent. . .

En revanche, si nous avions les informations théoriques de l'autre variété de fraise, ce serait éventuellement possible (à méditer !)

⚠ Remarque :

Retour sur la signification de la probabilité $\alpha = 0.95$ du " $P(\overline{X}_n \in I) = 0.95$ " :

On comprend qu'on a

- 95% de chance que cet événement se produise
- 95% des valeurs de \overline{X}_n sont dans I
- 5% des valeurs de \overline{X}_n ne sont pas dans I .

Autrement dit, on a (au maximum) 5% de chance de se tromper quand on rejette H_0 . Si on veut réduire les chances de se tromper, il faut donc augmenter le " α ". On pourra d'ailleurs se demander pourquoi on ne pourra jamais prendre $\alpha = 1$ en obtenant un intervalle I raisonnable! . . .

? Exercice 5

Reprendre l'exercice précédent (p. 219) avec $P(\overline{X}_n \in I) \simeq 0.99$ (On veut conclure avec moins de risque de se tromper que pour l'exercice précédent où c'était 0.95). Peut-on conclure sur le fait que ce soit ou non la palette de gariguettes si $m_A = 16,5$?

Solution

Les calculs nous amènent à $a \simeq 2.576$, $I = [15.48; 16.51]$

Cette fois-ci, on a $16,5 \in I$. **On ne peut donc pas exclure que ce soit la palette de gariguette** mais on ne peut pas non plus affirmer que c'est bien la palette de gariguettes.

IV-2 Résumé de vocabulaire et principe général

On peut donc maintenant résumer le principe général d'un test de conformité en détail :

📖 Définition

On appelle *population de référence* la population générale pour laquelle on dispose des données théoriques.

📖 Notation :

On note ici \mathcal{P} une population de référence et A un échantillon d'une population \mathcal{P}_A dont on ne sait pas si elle correspond ou non à la population de référence. (ex : deux variétés de fraises peut être différentes)

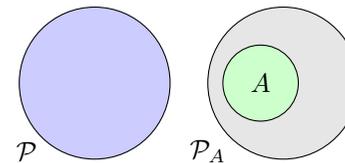
On souhaite savoir si $\mathcal{P} = \mathcal{P}_A$ en effectuant un test sur un même caractère (ex : le poids des fraises) pour lequel on note X et X_A les variables aléatoires mesurant ledit caractère sur un individu respectivement de \mathcal{P} et \mathcal{P}_A .

On estime évidemment que chaque individu est indépendant concernant le caractère étudié. On note

$$\mu = \mathbb{E}[X], \mu_A = \mathbb{E}[X_A] \text{ et } \overline{X}_n \text{ la moyenne empirique qui sera observée sur } A.$$

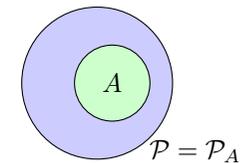
Résumé des 2 situations possibles :

A n'est pas dans \mathcal{P} ($\mathcal{P} \neq \mathcal{P}_A$) :



$$\mu \neq \mu_A$$

A est dans \mathcal{P} ($\mathcal{P} = \mathcal{P}_A$)



$$\mu = \mu_A$$

Définition
 On pose
$$H_0 : \mu = \mu_A.$$
 Cette hypothèse est appelée *l'hypothèse nulle*. On pose H_1 l'hypothèse contraire. Elle est appelée *hypothèse alternative*.

Remarque :
 H_0 traduit notre hypothèse de travail $\mathcal{P} = \mathcal{P}_A$, et c'est la seule hypothèse sous laquelle on peut effectuer nos calculs (ex : l'échantillon est dans la palette de gari-guettes)
 Il existe d'autres types d'hypothèses nulles (par exemple sur des variances $\sigma = \sigma_A$ ou autre) mais ce sera toujours une hypothèse précise avec laquelle on peut procéder à des calculs.

Raisonnement :
 Pour une probabilité donnée α (grande), on cherche un intervalle I adapté à nos hypothèses tel que
$$P(\bar{X}_n \in I) = \alpha.$$
 Si finalement la moyenne effective vaut $\bar{X}_n = m_A \notin I$, on rejette H_0 , d'où la définition suivante :

Définition
 On appelle $\mathbb{R} - I$ la *zone de rejet*.

Commentaires :
 Pour les tests de conformité, il faut donc globalement comprendre le fait suivant : quand la probabilité $P(\bar{X}_n \in I)$ est grande,

- une situation d'exclusion ($m_A \notin I$) permet de rejeter l'hypothèse.
 En effet, si l'échantillon était conforme, on devrait être dans I . On ne l'est pas, donc l'échantillon est non conforme.
- une situation d'inclusion ($m_A \in I$) ne permet pas de conclure.
- Plus on veut être certain d'avoir raison, (α augmente) moins on prend de décisions.

On va maintenant voir qu'il existe classiquement deux types d'hypothèses alternatives :

IV-3 Tests bilatéraux

Ce sont les tests les plus courants. Pour un échantillon A et une population de référence \mathcal{P} de moyenne théorique resp. μ_A et μ , ils consistent à opposer deux situations de type

$$H_0 : \mu_A = \mu \quad ; \quad H_1 : \mu_A \neq \mu$$

Pour faire les calculs, on va donc tenter de trouver des intervalles symétriques par-rapport à μ de type $I = [\mu - \dots; \mu + \dots]$ (comme dans l'exemple de la partie précédente !)
 En effet, il n'y a pas de raison de choisir une zone de rejet différente suivant qu'on soit plus grand ou plus petit que μ .

IV.3-a) Tests de conformité d'une moyenne connaissant σ^2

Propriété 203

Soit (X_1, \dots, X_n) est un n -échantillon d'une variable aléatoire X d'espérance μ et d'écart-type σ . En notant :
$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$$
 Alors, pour tout $a \geq 0$, on a
$$P\left(\left|\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}\right| < a\right) = P\left(\mu - a\frac{\sigma}{\sqrt{n}} < \bar{X}_n < \mu + a\frac{\sigma}{\sqrt{n}}\right) \xrightarrow{n \rightarrow +\infty} 2\Phi(a) - 1$$

Démonstration :
 C'est une conséquence de la première forme du théorème central limite. En effet, si $a > 0$, Posons

$$\begin{aligned} P_n &= P\left(\left|\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}\right| < a\right) \\ &= P\left(-a < \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} < a\right) \\ &= P\left(-a\frac{\sigma}{\sqrt{n}} < \bar{X}_n - \mu < a\frac{\sigma}{\sqrt{n}}\right) \\ &= P\left(\mu - a\frac{\sigma}{\sqrt{n}} < \bar{X}_n < \mu + a\frac{\sigma}{\sqrt{n}}\right) \end{aligned}$$

Or, d'après le théorème central limite, on sait que
$$P_n = P\left(-a < \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} < a\right) \xrightarrow{n \rightarrow +\infty} \Phi(a) - \Phi(-a) = 2\Phi(a) - 1$$

□

Cette propriété signifie que si n est grand, **quelle que soit** la variable X , on peut supposer que

$$P\left(\mu - a \frac{\sigma}{\sqrt{n}} < \bar{X}_n < \mu + a \frac{\sigma}{\sqrt{n}}\right) \simeq 2\Phi(a) - 1$$

Mais comment utiliser ceci ? ? Voyons ça dans l'exemple ci-dessous.

■ Exemple 14 :

Reprenons le cadre de l'exemple des palettes de gariguettes et observons ce qui a été fait :

On a noté μ_2 la valeur moyenne théorique du poids d'une fraise sur la palette étudiée, ainsi que l'hypothèse nulle

$$H_0 : \mu_2 = 16$$

et comme hypothèse alternative

$$H_1 : \mu_2 \neq 16$$

La propriété précédente nous dit alors que

$$I = \left[16 - a \frac{\sigma}{\sqrt{n}}; 16 + a \frac{\sigma}{\sqrt{n}}\right]$$

où a vérifie

$$2\Phi(a) - 1 = 0,95$$

et où $\sigma = 2$, $n = 100$. Il suffit donc de finir le calcul. On trouvait ici $a \simeq 1,96$ et

$$I = [15,61; 16,39]$$

⚠ Remarque :

Il n'est normalement pas nécessaire d'apprendre cette formule par coeur, étant donné que l'exercice vous amène généralement à la retrouver.

Cas particulier classique d'une proportion :

Comme nous l'avons vu dans ce chapitre, on rappelle qu'une probabilité p d'un événement B (donc la "proportion théorique") est approchée par la proportion empirique de réalisation dans un n échantillon d'une même variable X (avec n grand) :

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n} = \frac{\text{nombre de fois où } B \text{ se produit}}{n}$$

avec

$$X = \begin{cases} 1 & \text{si } B \text{ se réalise} \\ 0 & \text{sinon} \end{cases}$$

et où on rappelle que

$$p = \mathbb{E}[X] = \mu, \quad \sigma = \sqrt{p(1-p)}$$

La propriété permettant de faire le test est donc la suivante :

Propriété 204

Soit p la probabilité d'apparition d'un événement E dans une population de référence, et \bar{X}_n la proportion empirique d'apparition sur un échantillon de la population. Alors

$$P\left(p - a \frac{\sqrt{p(1-p)}}{\sqrt{n}} < \bar{X}_n < p + a \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) \xrightarrow{n \rightarrow +\infty} 2\Phi(a) - 1$$

En pratique, on peut dire, comme dans le théorème de Moivre-Laplace, que dès que $n \geq 30$, $np \geq 5$, $n(1-p) \geq 5$, on a

$$P\left(p - a \frac{\sqrt{p(1-p)}}{\sqrt{n}} < \bar{X}_n < p + a \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) \simeq 2\Phi(a) - 1$$

■ Exemple 15 :

Un individu souhaite savoir si son voisin a traité les pommes de son verger avec un pesticide. Il sait par expérience qu'en moyenne, les vergers non traités aux alentours fournissent un taux de 10% de fruits véreux.

Le verger étant immense, il "emprunte" 50 pommes, les coupe en 2, et constate une proportion de pommes véreuses effective de 2%. Trouver un intervalle I symétrique par-rapport à p tel que $P(\bar{X}_n \in I) = 0.95$. Peut-on en conclure si le verger est traité avec une probabilité de 0.95 ?

On pose l'hypothèse

$$H_0 : \text{"le verger n'est pas traité"}$$

(la seule hypothèse sur laquelle on peut faire des calculs avec les données que nous avons.) Sous cette hypothèse, on a

$$P\left(p - a \frac{\sqrt{p(1-p)}}{\sqrt{n}} < \bar{X}_n < p + a \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) \simeq 2\Phi(a) - 1$$

Or, ayant $n \geq 30$, $np = 50 \times 0,1 = 5 \geq 5$, $n(1-p) = 45 \geq 5$, on a

$$2\Phi(a) - 1 \simeq 0.95 \quad \Leftrightarrow \quad a \simeq 1.96$$

ainsi, un intervalle I correspondant à la demande est

$$I \simeq \left[p - a \frac{\sqrt{p(1-p)}}{\sqrt{n}}; p + a \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right] \simeq \left[0.1 - a \frac{\frac{3}{10}}{\sqrt{50}}; 0.1 + a \frac{\frac{3}{10}}{\sqrt{50}}\right] \simeq [0.017; 0.183]$$

Ici,

$$0.02 \in I$$

L'hypothèse ne peut être rejetée et ainsi :

on ne peut pas exclure le fait que le verger ne soit pas traité.

IV.3-b) Test de conformité ne connaissant pas σ^2

■ Exemple 16 :

On considère à nouveau la population de fraises gariguettes, dont le poids moyen par fruit est de $\mu = 16g$ et dont la palette a été confondue avec celle d'une autre variété. Le nouveau responsable de l'étiquetage ne connaît malheureusement pas la variance théorique liée à cette variété de fruits. Comment faire pour identifier les palettes? (*essayons de répondre à cette question ci-dessous*)

La TCL seconde forme donne lieu au résultat suivant :

Propriété 205

Soit (X_1, \dots, X_n) est un n -échantillon d'une variable aléatoire X d'espérance μ et d'écart-type σ . En notant :

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n} \quad \text{et} \quad S_n^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X}_n)^2$$

Alors, pour tout $a \geq 0$, on a

$$P\left(\left|\frac{\bar{X}_n - \mu}{S_n/\sqrt{n}}\right| < a\right) = P\left(\mu - a\frac{S_n}{\sqrt{n}} < \bar{X}_n < \mu + a\frac{S_n}{\sqrt{n}}\right) \xrightarrow{n \rightarrow +\infty} 2\Phi(a) - 1$$

Démonstration :

La démonstration est en tout point identique à celle faite à la page 221 concernant le premier test de conformité, mais en remplaçant σ par S_n . \square

? Exercice 6

Revenons à notre exemple de gariguettes. Prenant exemple sur son prédécesseur, le nouveau responsable de l'étiquetage a prélevé aléatoirement $n = 100$ fraises. Il trouve un poids moyen $\bar{X}_n = 16,5$ et une variance empirique $S_n^2 = 9$.

1. Quelle est la valeur de a qui correspond à la probabilité

$$P\left(\bar{X}_n \in \left[16 - a\frac{S_n}{\sqrt{n}}; 16 + a\frac{S_n}{\sqrt{n}}\right]\right) = 0.95$$

2. En déduire un intervalle I pour lequel $P(\bar{X}_n \in I) \simeq 0.95$.

3. La valeur moyenne observée $\bar{X}_n = 16,5$ permet-elle de savoir si la palette considérée est la palette de gariguettes?

4. Le responsable prélève maintenant un échantillon sur l'autre palette et trouve cette fois-ci un poids moyen $\bar{X}_n = 17.2$ et une variance empirique $S_n^2 = 4$. Peut-il conclure?

Solution

1. Pour la probabilité 0.95, D'après la calculatrice, on a encore

$$a \simeq 1,96.$$

2. d'après le TCL, en se plaçant dans l'hypothèse nulle \mathcal{H}_0 : "la palette est celle de gariguettes", l'intervalle I est le suivant :

$$I \simeq \left[\mu - a\frac{S_n}{\sqrt{n}}; \mu + a\frac{S_n}{\sqrt{n}}\right] = \left[16 - a\frac{\sqrt{9}}{\sqrt{100}}; 16 + a\frac{\sqrt{9}}{\sqrt{100}}\right] \simeq [15.41; 16.59]$$

3. Comme $16,5 \in I$, on ne peut rien conclure.

4. On se place dans l'hypothèse nulle \mathcal{H}_0 : "la deuxième palette est la palette de gariguettes". On obtient alors, avec les mêmes calculs, un intervalle

$$I_2 \simeq \left[\mu - a\frac{\sqrt{4}}{\sqrt{100}}; \mu + a\frac{\sqrt{4}}{\sqrt{100}}\right] = [15,6; 16.4]$$

Cette fois-ci, $17.2 \notin I_2$. On exclut donc l'hypothèse nulle au profit de l'hypothèse alternative. On peut conclure que **ce n'est pas la palette de gariguettes**. Par élimination, la première était donc la bonne!

IV-4 Tests unilatéraux

Ce sont les tests que l'on peut mettre en place si on est déjà certain que l'une des valeurs est supérieure (ou inférieure à l'autre)

Pour un échantillon A et une population de référence \mathcal{P} de moyenne théorique resp. μ_A et μ , ils consistent à opposer deux situations de type

$$H_0 : \mu_A = \mu \quad ; \quad H_1 : \mu_A > \mu$$

ou

$$H_0 : \mu_A = \mu \quad ; \quad H_1 : \mu_A < \mu$$

Pour faire nos calculs, on va donc tenter de trouver des intervalles bornés d'un seul côté par rapport à μ , de manière cohérente avec le problème : i.e. de type :

- pour le premier cas : $I =]-\infty; \mu + \beta]$ (on donne un "maximum" afin de rejeter ce qui est trop loin au dessus)
- pour le deuxième cas : $I = [\mu - \beta; +\infty[$ (on donne un "minimum" afin de rejeter ce qui est trop loin en dessous).

On a malgré tout toujours la possibilité de faire un test bilatéral comme dans la partie précédente, mais nous verrons sur un exemple quel est l'intérêt de faire un test unilatéral.

IV.4-a) On sait déjà que $\mu_A \geq \mu$; donc cas de H_1 : " $\mu_A > \mu$ "

On cherche à rejeter des intervalles de type $]\mu + \beta, +\infty[$, et donc avoir des intervalles de type

$$\overline{X}_n \in]-\infty; \mu + \beta].$$

Cas de la variance théorique connue, toujours directement par TCL :

Propriété 206

Soit (X_1, \dots, X_n) est un n -échantillon d'une variable aléatoire X d'espérance μ et d'écart-type σ . En notant : $\overline{X}_n = \frac{X_1 + \dots + X_n}{n}$, alors, pour tout $b \geq 0$, on a

$$P\left(\overline{X}_n \leq \mu + b \frac{\sigma}{\sqrt{n}}\right) = P\left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq b\right) \xrightarrow{n \rightarrow +\infty} \Phi(b)$$

? Exercice 7

Un individu souhaite savoir si son verger est contaminé par les pesticides de son voisin, qu'il soupçonne d'avoir trop largement diffusés par les airs. Il sait que la probabilité pour une pomme d'être véreuse après traitement par le pesticide en question est de $p = 1\%$.

On note p_A la proportion théorique de pommes véreuses dans son verger. On pose

$$H_0 : \text{"le verger est traité"} : "p = p_A"$$

et

$$H_1 : \text{"le verger n'est pas traité"} : "p < p_A"$$

1. Soit \overline{X}_n donnant la moyenne empirique de pommes véreuses dans son verger sur n pommes cueillies. Montrer que sous l'hypothèse H_0 , avec n suffisamment grand, on a

$$P\left(\overline{X}_n < p + b \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) \simeq \Phi(b)$$

2. En déduire un intervalle tel que $P(\overline{X}_n \in I) = 0.95$.

3. Sur les 500 pommes dans son verger, son taux effectif de pommes véreuses prélevées est de 3.5%. Peut-on en conclure si son verger a été contaminé avec une probabilité de 0.95 ?

Solution

1. On a $\overline{X}_n = \frac{T_n}{n}$, où T_n est le nombre de pommes véreuses trouvées sur les n cueillies (et indépendantes), d'où $T_n \hookrightarrow \mathcal{B}(n, p)$.

Ainsi, sous l'hypothèse H_0 , σ est connu et vaut

$$\sigma = \sqrt{p(1-p)}$$

D'après le théorème de Moivre Laplace, on sait de plus que, si $n \geq 30$, $np \geq 5$, $n(1-p) \geq 5$, on a

$$P\left(\overline{X}_n \leq \mu + b \frac{\sigma}{\sqrt{n}}\right) \simeq \Phi(b)$$

ce qui correspond à ce qui est demandé.

2. On a bien $n \geq 30$, $np \geq 5$, $n(1-p) \geq 5$. Ainsi

$$\Phi(b) \simeq 0.95 \iff b \simeq 1.645$$

ainsi, un intervalle I correspondant à la demande est

$$I \simeq \left] -\infty; p + b \frac{\sqrt{p(1-p)}}{\sqrt{n}} \right] \simeq]-\infty; 0.0173]$$

3. Ici,

$$0.035 \notin I$$

L'hypothèse H_0 est rejetée.

On en déduit que le verger n'est pas contaminé.

Cas de la variance théorique inconnue :

Propriété 207

Soit (X_1, \dots, X_n) un n -échantillon d'une variable aléatoire X d'espérance μ et d'écart-type σ . En notant

$$\overline{X}_n = \frac{X_1 + \dots + X_n}{n} \quad \text{et} \quad S_n^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \overline{X}_n)^2$$

Alors, pour tout $b \geq 0$, on a

$$P\left(\overline{X}_n \leq \mu + b \frac{S_n}{\sqrt{n}}\right) = P\left(\frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \leq b\right) \xrightarrow{n \rightarrow +\infty} \Phi(b)$$

? Exercice 8

Un ingénieur agronome souhaite savoir si le nouvel engrais qu'il envisage de produire augmente oui ou non le nombre moyen de tomate par pied. L'ancien engrais donnait un nombre moyen de $\mu = 14,38$ tomates. On note μ_A le nombre moyen de tomates avec le nouvel engrais. Il effectue le test sur un échantillon de $n = 300$ pieds. On pose

\mathcal{H}_0 : 'l'engrais n'a pas d'effet', \mathcal{H}_1 : 'l'engrais a un effet positif'

i.e.

$$\mathcal{H}_0 : \mu = \mu_A, \quad \mathcal{H}_1 : \mu_A > \mu$$

4. Soit \bar{X}_n le nombre moyen de tomate par plant ainsi que S_n l'écart type observé. Montrer que sous l'hypothèse H_0 , on a

$$P\left(\bar{X}_n \leq \mu + b \frac{S_n}{\sqrt{n}}\right) \simeq \Phi(b)$$

5. Les données effectives obtenues donnent un nombre moyen de 15,2 avec un écart-type de $S = \sqrt{3,5}$. En déduire un intervalle tel que $P(\bar{X}_n \in I) = 0.99$ et conclure.

Solution

1. Le calcul et les arguments sont exactement les mêmes que dans la propriété.
2. On obtient $b \simeq 2.326$ et

$$I \simeq]-\infty, 14.63]$$

Ainsi, $15,2 \notin I$ et on peut en effet conclure à une amélioration de l'engrais par rapport à l'ancienne version.

Commentaires généraux sur la méthode unilatérale :



Remarque :

Avec le test unilatéral, par ex. dans l'exercice sur les pesticides précédent, on trouve

$$I \simeq]-\infty; 0.033].$$

Avec un test bilatéral, i.e. $H_1 : "p \neq p_A"$ (que vous n'hésitez pas à faire pour vous entraîner), les résultats auraient été :

$$I' \simeq [-0.017; 0.037].$$

Dans ce cas, on aurait eu

$$0.035 \in I'$$

et on aurait conclu sur le fait qu'on ne peut pas conclure !

Commentaires :

En bref, le test unilatéral est plus "puissant" que le test bilatéral. En effet, comme on sait que nécessairement $\mu_A \geq \mu$, il a en réalité une zone de rejet plus "grande" (pour les cas supérieurs) :

> 0.033 pour le test unilatéral, et $> 0,037$ pour le test bilatéral

IV.4-b) Cas de $H_1 : " \mu_A < \mu "$

On inverse la procédure par-rapport à la partie précédente pour avoir des intervalles de type $\bar{X}_n \in [\mu - \beta; +\infty[$.

Cas de la variance théorique connue, toujours directement par TCL :

Propriété 208

Soit (X_1, \dots, X_n) est un n -échantillon d'une variable aléatoire X d'espérance μ et d'écart-type σ . En notant :

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$$

Alors, pour tout $b \geq 0$, on a

$$P\left(\mu - b \frac{\sigma}{\sqrt{n}} \leq \bar{X}_n\right) = P\left(b \leq \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}\right) \xrightarrow{n \rightarrow +\infty} 1 - \Phi(b)$$

Démonstration :

On utilise toujours le TCL, mais en passant à l'événement contraire :

$$P\left(b \leq \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}\right) = 1 - P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} < b\right) = 1 - \Phi(b)$$

□

? Exercice 9

Un producteur des pesticides a mis point un autre produit qu'il sait être au moins aussi efficace que la version précédente, dont on rappelle qu'elle donnait seulement $p = 1\%$ de pommes véreuses.

Il veut maintenant savoir si l'écart est significatif et si la nouvelle version est en effet plus efficace que l'autre. Il effectue donc son test dans le verger de votre voisin. . .

On note p_A la proportion théorique de pommes véreuses dans ledit jardin. On pose

$$H_0 : "le pesticide n'est pas plus efficace" : "p = p_A"$$

et

$$H_1 : "le pesticide est plus efficace" : "p > p_A"$$

1. Soit \bar{X}_n donnant la moyenne empirique de pommes véreuses dans le verger sur n pommes cueillies. Montrer que sous l'hypothèse H_0 , on a

$$P\left(p - b \frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq \bar{X}_n\right) \simeq \Phi(b)$$

2. En déduire un intervalle tel que $P(\bar{X}_n \in I) = 0.95$.

3. Il fait un test sur 500 pommes. Son taux effectif de pommes véreuses prélevées est de 0.5%. Peut-on en conclure si son pesticide est plus efficace ?

Solution

1. On a $\overline{X}_n = \frac{T_n}{n}$, où T_n est le nombre de pommes véreuses trouvées sur les n cueillies (et indépendantes), d'où $T_n \hookrightarrow \mathcal{B}(n, p)$ où $p = 0.1$.

Ainsi, sous l'hypothèse H_0 , σ est connu et vaut

$$\sigma = \sqrt{p(1-p)}$$

D'après le TCL, on sait de plus que, comme $n \geq 30$, $np \geq 5$, $n(1-p) \geq 5$,

$$P\left(\mu - b \frac{\sigma}{\sqrt{n}} \leq \overline{X}_n < +\infty\right) \simeq 1 - \Phi(-b) = \Phi(b)$$

ce qui correspond à ce qui est demandé.

2. On a

$$\Phi(b) \simeq 0.95 \Leftrightarrow b \simeq 1.645$$

ainsi, un intervalle I correspondant à la demande est

$$I \simeq \left[p - b \frac{\sqrt{p(1-p)}}{\sqrt{n}}; +\infty \right] \simeq [0.0027; +\infty[$$

3. Ici,

$$0.005 \in I$$

Ce n'est pas significatif, on ne peut pas conclure.

Cas de la variance théorique inconnue :

Propriété 209

Soit (X_1, \dots, X_n) est un n -échantillon d'une variable aléatoire X d'espérance μ et d'écart-type σ . En notant :

$$\overline{X}_n = \frac{X_1 + \dots + X_n}{n} \quad \text{et} \quad S_n^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \overline{X}_n)^2$$

Alors, pour tout $b \geq 0$, on a

$$P\left(\mu - b \frac{S_n}{\sqrt{n}} \leq \overline{X}_n\right) = P\left(b \leq \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}}\right) \xrightarrow{n \rightarrow +\infty} 1 - \Phi(b)$$

? Exercice 10

Un laboratoire souhaite savoir si le nouveau médicament qu'il envisage de distribuer contre les migraines améliore oui ou non le temps total des crises. L'ancien médicament donne une vitesse de "guérison" moyenne de $\mu = 3,25h$. On note μ_A la durée moyenne (pour l'instant inconnue) avec le nouveau médicament. Il effectue le test sur un échantillon de $n = 200$ personnes atteintes de ladite maladie. On pose

$$H_0 : \text{'le médicament n'a pas d'effet'}, \quad H_1 : \text{'le médicament a un effet positif'}$$

i.e.

$$H_0 : \mu = \mu_A, \quad H_1 : \mu_A < \mu$$

1. Soit \overline{X}_n la vitesse moyenne de guérison de n patients ainsi que S_n l'écart type observé. Montrer que sous l'hypothèse H_0 , on a

$$P\left(\overline{X}_n \geq \mu - b \frac{S_n}{\sqrt{n}}\right) \simeq \Phi(b)$$

2. Les données effectives obtenues sont une durée moyenne de crise de $2,8h$ avec un écart-type de $S = \sqrt{1,6}$. En déduire un intervalle tel que $P(\overline{X}_n \in I) = 0.99$ et conclure.

Solution

1. D'après le TCL, le calcul et les arguments sont exactement les mêmes que dans le cas où σ est connue, mais en remplaçant σ par S_n . On peut donc reprendre la même démonstration que dans l'exercice précédent.

2. On obtient $b \simeq 2.326$ et

$$I \simeq [3.042; +\infty[$$

Ainsi, $2,8 \notin I$ et on peut en effet conclure à une amélioration du médicament par rapport à l'ancien.