Mathématiques pour les études scientifiques II
LU1MA002

Faculté des Sciences — Sorbonne Université

1er semestre 2019 - Version du 10 mars 2020

Table des matières

Première partie

Calcul matriciel et algèbre linéaire dans $\mathbf{R}^n$

Chapitre 1

Matrices

1.1 Calcul matriciel

Les matrices, dont il est question ici, sont des tableaux de nombres. Nous choisissons de travailler avec des nombres réels, mais nous aurions pu aussi travailler avec des complexes. Donnons une définition précise.

definition 1.1. Soient p et q des entiers postifs. Une matrice à p lignes et q colonnes est un tableau rectangulaire de nombres réels $A = (a_{ij})$. L'indice de ligne i varie de 1 à p, l'indice de colonne j de 1 à q.

$$A = (a_{ij}) = \begin{pmatrix} a_{11} & \dots & a_{1j} & \dots & a_{1,q} \\ \vdots & & \vdots & & \vdots \\ a_{i1} & \dots & a_{ij} & \dots & a_{iq} \\ \vdots & & \vdots & & \vdots \\ a_{p1} & \dots & a_{pj} & \dots & a_{pq} \end{pmatrix}$$

Les nombres a_{ij} s'appellent les *coefficients* de la matrice.

Des exemples :

$$\begin{pmatrix} 2 & 0 \\ 3 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 \\ 3 \\ 4 \\ 5 \end{pmatrix}, \qquad \begin{pmatrix} 1 & x \\ 2+x & 3 \\ 4 & 5/3 \end{pmatrix}$$

où x est un réel.

Vocabulaire :

— Une matrice avec p lignes et q colonnes est dite de *taille* $p \times q$. Si $p = 1$, on parle de matrice *ligne*, si $q = 1$, de matrice *colonne*, si $p = q$ de matrice *carrée*.
— La *diagonale* d'une matrice carrée est l'ensemble des coefficients $a_{11}, a_{22}, \dots, a_{pp}$.
— La matrice *nulle* de taille $p \times q$ est la matrice de taille $p \times q$ dont tous les coefficients sont nuls. On la note $0_{p,q}$ ou parfois simplement 0.
— La matrice *identité* de taille $p \times p$ est la matrice dont tous les coefficients sont nuls excepté ceux de la diagonale qui sont tous égaux à 1. On la note Id_p

On note $\mathcal{M}_{p,q}$ l'ensemble des matrices avec p lignes et q colonnes.

1.1.1 *Addition et multiplication externe*

On définit l'*addition* et la *multiplication externe* :

— Si $A = (a_{ij})$ et $B = (b_{ij})$ sont deux matrices dans $\mathcal{M}_{p,q}$, alors $A + B$ est la matrice de $\mathcal{M}_{p,q}$ égale à $(a_{ij} + b_{ij})$

— Si $\lambda \in \mathbf{R}$ et $A = (a_{ij})$ est une matrice de $\mathcal{M}_{p,q}$, alors λA est la matrice de $\mathcal{M}_{p,q}$ donnée par (λa_{ij}).

Par exemple,

$$\begin{pmatrix} 1 & 1 \\ 2 & 3 \\ 1 & -1 \end{pmatrix} + \begin{pmatrix} -3 & 1 \\ 5 & -3 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} -2 & 2 \\ 7 & 0 \\ 1 & 1 \end{pmatrix}, \qquad -2 \begin{pmatrix} 1 & 1 \\ 2 & 3 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} -2 & -2 \\ -4 & -6 \\ -2 & 2 \end{pmatrix}$$

Ces deux opérations vérifient les règles usuelles de l'addition et la multiplication des nombres.

Énoncé indispensable 1 :

proposition 1.2. *Soient A, B, C trois matrices de même taille et α, β deux réels. Alors*

$$\begin{aligned} A + B &= B + A & \alpha(A + B) &= \alpha A + \alpha B \\ (A + B) + C &= A + (B + C) & (\alpha + \beta)A &= \alpha A + \beta A \\ A + 0 &= A & \alpha(\beta A) &= (\alpha\beta)A \end{aligned}$$

où 0 est la matrice nulle de même taille que A.

La deuxième propriété s'appelle l'associativité de l'addition. Puisqu'elle est vérifiée, nous écrirons $A + B + C$ sans mettre de parenthèse. Ce n'est pas nécessaire car l'ordre dans lequel nous faisons les additions ne modifie pas le résultat. C'est encore le cas lorsque nous sommons 4 matrices ou plus.

Démonstration. Vérifions la première égalité pour des matrices de taille 2×2

$$\begin{aligned} &\begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix} + \begin{pmatrix} a_2 & b_2 \\ c_2 & d_2 \end{pmatrix} = \begin{pmatrix} a_1 + a_2 & b_1 + b_2 \\ c_1 + c_2 & d_1 + d_2 \end{pmatrix} \\ &= \begin{pmatrix} a_2 + a_1 & b_2 + b_1 \\ c_2 + c_1 & d_2 + d_1 \end{pmatrix} = \begin{pmatrix} a_2 & b_2 \\ c_2 & d_2 \end{pmatrix} + \begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix} \end{aligned}$$

La preuve est exactement la même pour des matrices de taille quelconque. On remarque que l'étape importante est le fait que $a_1 + a_2 = a_2 + a_1$ et même chose pour les b_i, c_i et d_i. Ainsi nous sommes passés de la commutativité de l'addition des nombres $x + y = y + x$ à la commutativité de l'addition des matrices $A + B = B + A$.

La preuve de la deuxième égalité découle de la propriété similaire des nombres $(x + y) + z = x + (y + z)$. Par exemple si A, B et C sont les matrices de tailles 3×1

$$A = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}, \; B = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}, \; C = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}$$

nous avons

$$(A+B)+C=\begin{pmatrix}a_1+b_1\\a_2+b_2\\a_3+b_3\end{pmatrix}+C=\begin{pmatrix}(a_1+b_1)+c_1\\(a_2+b_2)+c_2\\(a_3+b_3)+c_3\end{pmatrix}=\begin{pmatrix}a_1+b_1+c_1\\a_2+b_2+c_2\\a_3+b_3+c_3\end{pmatrix}$$
$$=\begin{pmatrix}a_1+(b_1+c_1)\\a_2+(b_2+c_2)\\a_3+(b_3+c_3)\end{pmatrix}=A+\begin{pmatrix}b_1+c_1\\b_2+c_2\\b_3+c_3\end{pmatrix}=A+(B+C)$$

Les autres propriétés se vérifient par la même méthode. Par exemple, de $\alpha(x+y)=\alpha x+\alpha y$, on déduit que

$$\alpha\big(\begin{pmatrix}a_1 & a_2\end{pmatrix}+\begin{pmatrix}b_1 & b_2\end{pmatrix}\big)=\ldots=\alpha\begin{pmatrix}a_1 & a_2\end{pmatrix}+\alpha\begin{pmatrix}b_1 & b_2\end{pmatrix}$$

et de $(\alpha+\beta)x=\alpha x+\beta x$, l'on déduit

$$(\alpha+\beta)\begin{pmatrix}a & b\\c & d\\e & f\end{pmatrix}=\ldots=\alpha\begin{pmatrix}a & b\\c & d\\e & f\end{pmatrix}+\beta\begin{pmatrix}a & b\\c & d\\e & f\end{pmatrix}.$$

Nous laissons le lecteur compléter les étapes qui manquent. De même les deux dernières propriétés sont laissées en exercice, on précisera dans chaque cas la propriété des nombres réels qui justifie le calcul. □

1.1.2 *Produit de matrices*

Soient p,q,r des entiers positifs. Pour toute matrice A de taille $p\times q$ et B de taille $q\times r$, nous allons définir la matrice AB de taille $p\times r$. Ici, on constate que le nombre de colonnes de A est égal au nombre de lignes de B. De plus AB a autant de lignes que A et autant de colonnes que B. Nous commençons par un cas particulier, à savoir le produit d'une matrice ligne par une matrice colonne, c'est-à-dire $p=r=1$ et q quelconque. Nous définissons

$$\begin{pmatrix}a_1 & \ldots & a_q\end{pmatrix}\begin{pmatrix}b_1\\\vdots\\b_q\end{pmatrix}=\begin{pmatrix}a_1b_1+\ldots+a_qb_q\end{pmatrix}\in\mathcal{M}_{11}$$

Comme le résultat est une matrice de taille 1×1, on le considère souvent comme un nombre. Par exemple

$$\begin{pmatrix}0 & 1 & 2\end{pmatrix}\begin{pmatrix}3\\4\\5\end{pmatrix}=0*3+1*4+2*5=14$$

Passons au produit dans le cas général avec p,q,r des entiers positifs quelconques. Par définition, le coefficient de la i-ième ligne et j-ième colonne de AB est le produit de la i-ième ligne de A par la j-ième colonne de B.

Par exemple,

$$\begin{pmatrix} 1 & 1 \\ 2 & 3 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 0 & 1 & -1 & -2 \\ -3 & -2 & 0 & 1 \end{pmatrix} = \begin{pmatrix} -3 & -1 & -1 & -1 \\ -9 & -4 & -2 & -1 \\ 3 & 3 & -1 & -3 \end{pmatrix} \tag{1.1}$$

Détaillons une partie du calcul. On note L_1, L_2, L_3 les 3 lignes de la matrice de gauche et C_1, C_2, C_3, C_4 les colonnes de la matrice de droite. Alors les deux premiers coefficients de la première ligne du produit sont

$$L_1C_1 = \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ -3 \end{pmatrix} = -3, \qquad L_1C_2 = \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -2 \end{pmatrix} = -1.$$

Pour bien visualiser le produit des lignes de A par les colonnes de B, il peut être utile dans un premier temps de placer les matrices A, B et AB de la façon suivante :

$$\begin{array}{cc} & \begin{pmatrix} 0 & 1 & -1 & -2 \\ -3 & -2 & 0 & 1 \end{pmatrix} \\ \begin{pmatrix} 1 & 1 \\ 2 & 3 \\ 1 & -1 \end{pmatrix} & \begin{pmatrix} -3 & -1 & -1 & -1 \\ -9 & -4 & -2 & -1 \\ 3 & 3 & -1 & -3 \end{pmatrix} \end{array}$$

c'est-à-dire que la matrice A est à gauche, la matrice B au dessus et la matrice AB en bas à droite. Ainsi le coefficient de la i-ième ligne et j-ième colonne de AB est à l'intersection de la i-ième ligne L_i de A et de la j-ième colonne C_j de B. Bien entendu, ce coefficient est précisément le produit de L_i par C_j. Passés les premiers calculs, il est recommandé d'écrire le produit sous la forme (1.1) et de visualiser mentalement les produits de lignes et colonnes.

Résumons les points importants de la définition du produit.

Énoncé indispensable 2 :

1. Le produit AB est défini lorsque le nombre de colonnes de A est égal au nombre de lignes de B.
2. Le produit AB a autant de lignes que A et autant de colonnes que B.
3. Le coefficient de la i-ième ligne et j-ième colonne de AB est le produit de la i-ième ligne de A par la j-ième colonne de B.
4. Le produit d'une matrice ligne par une matrice colonne est donné par la formule
$$\begin{pmatrix} a_1 & \dots & a_q \end{pmatrix} \begin{pmatrix} b_1 \\ \vdots \\ b_q \end{pmatrix} = \begin{pmatrix} a_1b_1 + \dots + a_qb_q \end{pmatrix} \in \mathcal{M}_{11}$$

Dans la proposition suivante, nous faisons la liste des propriétés vérifiées par le produit, l'addition et la multiplication externe.

Énoncé indispensable 3 :

PROPOSITION 1.3. *Soient p, q, r et s des entiers positifs non nuls.*

1. *si $A \in \mathcal{M}_{p,q}$, $B \in \mathcal{M}_{q,r}$ et $C \in \mathcal{M}_{r,s}$, alors $(AB)C = A(BC)$.*
2. *si $A, B \in \mathcal{M}_{p,q}$ et $C \in \mathcal{M}_{q,r}$, alors $(A+B)C = AC + BC$.*
3. *si $A \in \mathcal{M}_{p,q}$ et $B, C \in \mathcal{M}_{q,r}$, alors $A(B+C) = AB + AC$.*
4. *si $\alpha \in \mathbf{R}$, $A \in \mathcal{M}_{p,q}$ et $B \in \mathcal{M}_{q,r}$, alors $\alpha(AB) = (\alpha A)B = A(\alpha B)$.*
5. *si $A \in \mathcal{M}_{p,q}$, alors $\mathrm{Id}_p\, A = A = A\, \mathrm{Id}_q$.*

Démonstration de la proposition 1.3. La première propriété, qui est l'associativité du produit, est la plus difficile à vérifier, nous y reviendrons plus tard. Pour (2), on suppose tout d'abord que A, B sont des matrices lignes et C une matrice colonne. Pour simplifier l'écriture, considérons des matrices avec trois coefficients, donc

$$A = \begin{pmatrix} a_1 & a_2 & a_3 \end{pmatrix}, \qquad B = \begin{pmatrix} b_1 & b_2 & b_3 \end{pmatrix}, \qquad C = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}$$

Alors $(A+B)C = (a_1+b_1)c_1 + (a_2+b_2)c_2 + (a_3+b_3)c_3$ et d'autre part $AC + BC = (a_1c_1 + a_2c_2 + a_3c_3) + (b_1c_1 + b_2c_2 + b_3c_3)$. Ces deux quantités sont bien les mêmes. On en déduit maintenant la propriété pour des matrices de taille quelconque : pour calculer les coefficients de $(A+B)C$ et $AB + BC$ sur la i-ième ligne, j-ième colonne, nous avons besoin de la i-ième ligne de A notée A_i, de la i-ième ligne de B notée B_i et de la j-ième colonne de C notée C_j. Alors les coefficients en question sont précisément $(A_i + B_i)C_j$ et $A_iC_j + B_iC_j$. Et nous avons vérifié précédemment que ces nombres sont égaux.

Pour vérifier les propriétés (3) et (4), on peut suivre exactement la même méthode : on suppose tout d'abord que dans chaque produit, les matrices de gauche sont des lignes et celles de droite des colonnes. On déduit ensuite le cas général de ce cas particulier.

L'égalité (5) est très importante. Elle se vérifie facilement. Par exemple pour des matrices de taille 2×2, elle s'écrit

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

La preuve de (1) est plus compliquée car on y fait le produit de 3 matrices au lieu de 2. Esquissons la méthode. On suppose dans un premier temps que A est une matrice ligne, B quelconque et C une matrice colonne. Alors $(AB)C$ et $A(BC)$ sont de taille 1×1, autrement dit des nombres. Supposons par exemple que

$$A = \begin{pmatrix} a_1 & a_2 & a_3 \end{pmatrix}, \qquad B = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \\ \mu & \nu \end{pmatrix}, \qquad C = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

Nous laissons alors le lecteur se convaincre que $(AB)C$ et $A(BC)$ sont tous deux égaux

à la somme des coefficients de la matrice

$$\begin{pmatrix} a_1\alpha c_1 & a_1\beta c_2 \\ a_2\gamma c_1 & a_2\delta c_2 \\ a_3\mu c_1 & a_3\nu c_2 \end{pmatrix}$$

Remarquer que cette matrice s'obtient en multipliant chaque ligne de B par le coefficient correspondant de A, puis chaque colonne par le coefficient correspondant de C. Une fois ceci vérifié, on en déduit le cas général avec A et C de taille quelconque. □

remarque 1.4. Il ne faudrait pas croire que le produit des matrices se comporte toujours comme celui des nombres réels. Voici trois erreurs à ne pas faire.

1. Le produit des matrices n'est pas commutatif, c'est-à-dire que le produit AB n'est pas toujours égal au produit BA. Il y a plusieurs raisons à cela. Pour commencer, il se peut qu'un seul des deux produits soit défini, par exemple :

$$\begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} -2 & -2 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 1 & -1 \end{pmatrix} \text{ n'a pas de sens.}$$

   Parfois, les deux produits sont bien définis, mais n'ont pas la même taille

$$\begin{pmatrix} 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \end{pmatrix}, \qquad \begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

   Enfin il se peut que les deux produits soient bien définis, aient même taille et qu'ils soient pourtant différents :

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

   Exceptionnellement, nous avons bien $AB = BA$, par exemple

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

2. Il est faux que $A^2 = 0$ entraine $A = 0$. Par exemple, le carré de $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ est nul.

3. La règle de simplification : ($ab = ac$ avec $a \neq 0$) $\Rightarrow b = c$ vérifiée par les nombres réels ne s'étend pas aux matrices. Par exemple

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

   et pourtant $\begin{pmatrix} 1 \\ 0 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

1.1.3 *Inverse et puissance*

Énoncé indispensable 4 :

DEFINITION 1.5. Une matrice carrée A est dite inversible si il existe une matrice carrée de même taille B qui vérifie $AB = \mathrm{Id}$ et $BA = \mathrm{Id}$. Lorsque A est inversible, la matrice B est unique et s'appelle l'inverse de A. On la note A^{-1}.

Preuve de l'unicité de B. Soient B_1 et B_2 qui vérifient $B_1A = \mathrm{Id}$ et $AB_2 = \mathrm{Id}$. Alors $B_1AB_2 = (B_1A)B_2 = \mathrm{Id}\, B_2 = B_2$. De même, $B_1AB_2 = B_1(AB_2) = B_1\,\mathrm{Id} = B_1$. Donc $B_1 = B_2$. □

EXEMPLE 1.6.

— $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ et $\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$ sont inversibles avec pour inverses $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ et $\frac{1}{2}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$ respectivement.

— $\begin{pmatrix} 1 & 2 \\ 0 & 0 \end{pmatrix}$ n'est pas inversible. En effet

$$\begin{pmatrix} 1 & 2 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a+2c & b+2d \\ 0 & 0 \end{pmatrix}$$

qui est différent de la matrice identité quel que soit le choix de a, b, c et d.

Dans la définition précédente, nous avons mis les deux conditions $AB = \mathrm{Id}$ et $BA = \mathrm{Id}$. Il se trouve que l'une implique l'autre. Cela n'est pas évident car comme nous l'avons déjà remarqué, le produit des matrices carrées de même taille n'est pas commutatif. La preuve sera donnée plus tard, cf. corollaire 1.31. La proposition suivante se vérifie facilement, elle est très importante.

Énoncé indispensable 5 :

PROPOSITION 1.7. *Soient A et B deux matrices inversibles de même taille. Alors*

— *A^{-1} est inversible, son inverse est A.*

— *AB est inversible, son inverse est $B^{-1}A^{-1}$.*

Démonstration. Le premier point est immédiat, en effet la condition $AB = BA = \mathrm{Id}$ est symétrique en A et B. Donc A est l'inverse de B si et seulement si B est l'inverse de A. Pour le second point, on calcule

$$AB(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = A\,\mathrm{Id}\,A^{-1} = AA^{-1} = \mathrm{Id}$$

et de même $(B^{-1}A^{-1})AB = \ldots = \mathrm{Id}$. □

REMARQUE 1.8.

1. Il ne faut pas écrire de fraction avec une matrice au dénominateur. En effet, même si l'on suppose B inversible, $\frac{A}{B}$ peut se lire comme AB^{-1} ou bien $B^{-1}A$, qui ne sont pas égaux en général.
2. Si $AB = AC$ avec A inversible, alors $B = C$. En effet, on multiplie l'égalité $AB = AC$ par A^{-1} à gauche, et il vient $A^{-1}AB = A^{-1}AC$, autrement dit $\mathrm{Id}\, B = \mathrm{Id}\, C$, d'où $B = C$. De même, $BA = CA$ avec A inversible entraine que $B = C$.

Définissons les puissances positives d'une matrice carrée A :

$$A^0 = \mathrm{Id}, \quad A^1 = A, \quad A^2 = AA, \qquad A^3 = AAA \quad \text{et} \quad A^n = \underbrace{A \dots A}_{n \text{ fois}} \text{ pour } n \geqslant 1.$$

On rappelle que cette écriture a du sens car le produit des matrices est associatif. Si A est inversible on définit aussi ses puissances négatives : $A^{-n} = (A^{-1})^n$ pour $n \geqslant 1$.

Énoncé indispensable 6 :

proposition 1.9. *Pour toute matrice carrée A et entiers positifs n, p, nous avons*

$$A^n A^p = A^{n+p} \qquad \text{et} \qquad (A^n)^p = A^{np}.$$

Si A est inversible, ces deux formules sont vraies pour des entiers n et p de signe quelconque.

La preuve est laissée en exercice. On utilisera bien entendu l'associativité du produit des matrices et la proposition 1.7 pour traiter les puissances négatives.

1.2 Systèmes linéaires

Nous n'allons traiter que le cas des systèmes linéaires faisant intervenir des nombres réels. On s'apercevra plus tard que ce que l'on dit dans ce cas reste valable pour des systèmes linéaires faisant intervenir des nombres complexes.

Un exemple de système linéaire est la liste d'équations

$$\begin{cases} 4x_1 + 3x_2 - 5x_3 = & 1 \\ 2x_1 - 6x_2 + 2x_3 = & 0 \end{cases}$$

Plus généralement, soient $p, q \geq 1$ des entiers.

Énoncé indispensable 7 :

Un *système linéaire* à p équations et q inconnues $x_1, \dots, x_q$ est un ensemble d'équations de la forme

$$(\mathcal{S}) \begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1q}x_q = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2q}x_q = b_2 \\ \vdots \\ a_{p1}x_1 + a_{p2}x_2 + \dots + a_{pq}x_q = b_p \end{cases}$$

Les nombres réels a_{ij} sont les coefficients du système linéaire, et les b_j sont les coefficients du second membre. Les solutions du système sont les q-uplets $(x_1, \ldots, x_q)$ de nombres réels qui vérifient toutes les équations du système.

Noter que l'on peut avoir $p > q, p = q$ ou $p < q$. Nous verrons aussi que le système peut avoir une seule, aucune, ou une infinité de solutions.

Test :

Donner des exemples de systèmes linéaires qui ne possèdent aucune solution, puis une seule solution, puis une infinité de solution, lorsque $p = q = 2$.

On peut réécrire ce système de manière plus compacte sous la forme d'une seule équation matricielle $AX = B$, où $A \in M_{p,q}(\mathbb{R})$, $X \in \mathbb{R}^q$ et $B \in \mathbb{R}^p$ sont les matrices

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1q} \\ \vdots & & \\ a_{p1} & \cdots & a_{pq} \end{pmatrix}, \qquad X = \begin{pmatrix} x_1 \\ \vdots \\ x_q \end{pmatrix}, \qquad B = \begin{pmatrix} b_1 \\ \vdots \\ b_p \end{pmatrix}.$$

Si le second membre B est nul, on dit que que le système est *homogène*. L'ensemble des solutions $\mathcal{S}_A = \{X \in \mathbb{R}^p, AX = 0\}$ d'un système homogène a une structure particulière.

Énoncé indispensable 8 :

PROPOSITION 1.10. *$\mathcal{S}_A$ est stable par somme et par multiplication externe par tout scalaire, c'est-à-dire que pour tout $X, Y \in \mathcal{S}_A$ et tout $\lambda \in \mathbf{R}$, $X + Y \in \mathcal{S}_A$ et $\lambda X \in \mathcal{S}_A$.*

Comme nous le verrons plus tard, une partie de $\mathbf{R}^q$ qui vérifie ces deux propriétés s'appelle un sous-espace vectoriel.

Considérons à présent le système $(\mathcal{S})$ avec un second membre B pas nécessairement nul. La proposition importante qui suit décrit l'ensemble $\mathcal{S} = \{X \in \mathbb{R}^p, AX = B\}$ de ses solutions lorsqu'on en connaît une, notée X_0.

PROPOSITION 1.11. *(principe de superposition) Soit $X_0 \in \mathcal{S}$. Alors les solutions du système $AX = B$ sont*

$$\mathcal{S} = \{X_0 + Y, AY = 0\}.$$

On appelle X_0 une solution particulière. On voit que toute solution est la somme de X_0 et d'une solution du système homogène *associé* $AX = 0$. On retrouve la structure de l'ensemble des solutions des équations différentielles vues au premier semestre.

Démonstration. $AX = B \iff AX = AX_0 \iff A(X - X_0) = 0$, ce qui est équivalent à demander que $Y = X - X_0$ est solution du système homogène associé. □

Énoncé indispensable 9 :

proposition 1.12. *(cas spécial d'une matrice inversible) Supposons que A soit une matrice carrée inversible. Alors le système linéaire $AX = B$ possède une solution unique, à savoir $X_0 = A^{-1}B$.*

Démonstration. En effet, si $AX = B$, alors $X = A^{-1}AX = A^{-1}B$. Et réciproquement, le vecteur colonne $X_0 = A^{-1}B$ est solution du système, puisque $AX_0 = AA^{-1}B = B$. □

Nous verrons à la fin du chapitre une réciproque de ce résultat : si pour tout second membre B, le système $AX = B$ admet une unique solution, alors A est carrée et inversible, cf. proposition 1.29 et théorème 1.30.

1.2.1 *Système triangulaires et échelonnés*

definition 1.13. Une matrice est dite triangulaire si elle est carrée et tous ses coefficients en dessous de la diagonale sont nuls.

Voici plusieurs exemples de matrices triangulaires

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

Parfois les matrices triangulaires sont appelées matrices *triangulaires supérieures* pour les distinguer des matrices *triangulaires inférieures*, qui sont les matrices dont les coefficients au dessus de la diagonale sont nuls.

Énoncé indispensable 10 :

proposition 1.14. *Soit le système $AX = B$ avec A une matrice triangulaire. Supposons que tous les coefficients de la diagonale de A sont non-nuls. Alors le système a une unique solution.*

Comme nous le verrons dans la proposition 1.33, une matrice triangulaire est inversible précisément lorsque tous les coefficients de la diagonale sont non-nuls.

Démonstration. Pour un système 3×3, nous avons

$$A = \begin{pmatrix} p_1 & q & r \\ 0 & p_2 & s \\ 0 & 0 & p_3 \end{pmatrix}, \qquad B = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

avec p_1, p_2 et p_3 non-nuls. Ces coefficients vont jouer un rôle important. On verra plus

tard qu'ils s'appellent des pivots. Le système s'écrit

$$\begin{cases} p_1x_1 & + & qx_2 & + & rx_3 & = & b_1 \\ & & p_2x_2 & + & sx_3 & = & b_2 \\ & & & & p_3x_3 & = & b_3 \end{cases}$$

Comme $p_3 \neq 0$, la troisième équation donne $x_3 = b_3/p_3$. On remplace alors x_3 par p_3/b_3 dans la deuxième équation et comme $p_2 \neq 0$, il vient

$$x_2 = (b_2 - sb_3/p_3)/p_2.$$

Enfin, on remplace dans la première équation x_2 et x_3 par les expressions précédentes, et l'on obtient x_1. Nous pourrions calculer explicitement x_1, mais le fait important est que la solution existe bien et qu'elle est uniquement déterminée. Ceci découle de la forme de l'équation $p_1x_1 + * = *$ où les $*$ représentent des nombres déterminés par les équations précédentes. Cette équation a une unique solution x_1 lorsque $p_1 \neq 0$.

Il reste à se convaincre que la méthode est générale et s'applique à des systèmes de taille plus grande. Le principe est le même : en remontant les équations, on détermine successivement chaque inconnue, de la dernière à la première. □

Lorsque la matrice est triangulaire avec des zéros sur la diagonale, le résultat précédent est faux. Par exemple

$$\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \Leftrightarrow \quad \begin{cases} x_2 = 1 \\ x_2 = 0 \end{cases}$$

qui n'a pas de solution car les deux équations se contredisent. Avec la même matrice et un second membre différent, nous obtenons bien des solutions, mais en nombre infini :

$$\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \Leftrightarrow \quad \begin{cases} x_2 = 0 \\ x_2 = 0 \end{cases} \quad \Leftrightarrow \quad x_2 = 0$$

L'ensemble des solutions est en effet $\{(t, 0),\ t \in \mathbf{R}\}$.

Nous allons à présent étudier les systèmes dits échelonnés. Leur résolution est très simple. Et comme nous le verrons dans le chapitre suivant, tout système est équivalent à un système échelonné.

Énoncé indispensable 11 :

DEFINITION 1.15. Une matrice $A \in M_{p,q}(\mathbb{R})$ est dite échelonnée à r pivots si elle a les deux propriétés suivantes

1. ses r premières lignes sont non-nulles, les lignes suivantes sont nulles.
2. pour $i = 1, \ldots, r-1$, le premier coefficient non-nul p_i de la i-ième ligne est située à gauche du premier coefficient non-nul p_{i+1} de la $(i+1)$-ième ligne.

Les coefficients $p_1, \ldots, p_r$ s'appellent les pivots de la matrice.

exemple 1.16. Les trois matrices

$$A = \begin{pmatrix} 0 & \boxed{1} & 0 & 3 \\ 0 & 0 & \boxed{5} & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, A' = \begin{pmatrix} \boxed{1} & 1 & -1 & 3 \\ 0 & 0 & \boxed{1} & 3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, A'' = \begin{pmatrix} \boxed{1} & 0 \\ 0 & \boxed{2} \end{pmatrix}$$

sont échelonnées à 2 pivots.

Test :

Précisez si les matrices sont échelonnées et, le cas échéant, identifiez les pivots et les colonnes sans pivots.

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 3 & 2 & 4 & 0 & 5 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\begin{pmatrix} 0 & 3 & 2 & 4 & 0 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Résolvons par exemple un système avec la matrice A' de l'exemple 1.16 et un second membre général

$$A' \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{pmatrix} \quad \Leftrightarrow \quad \begin{cases} x_1 + x_2 - x_3 + 3x_4 &= b_1 \\ x_3 + 3x_4 &= b_2 \\ 0 &= b_3 \\ 0 &= b_4 \end{cases}$$

Lorsque $b_3 \neq 0$ ou $b_4 \neq 0$, le système n'a pas de solution car une des deux dernières équations n'est pas vérifiée. Supposons que $b_3 = b_4 = 0$. Alors les deux dernières équations s'écrivent $0 = 0$, elles n'apportent aucune information, nous pouvons les enlever. Dans les deux équations qui restent, passons les variables x_2 et x_4 dans le membre de gauche.

$$\begin{cases} x_1 - x_3 = b_1 - x_2 - 3x_4 \\ x_3 = b_2 - 3x_4 \end{cases}$$

Nous pouvons interpréter ces équations comme un système à deux inconnues x_1, x_3 avec un second membre dépendant de paramètres x_2 et x_4. La matrice $\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$ de ce nouveau système est triangulaire, les coefficients de sa diagonale sont non-nuls. Nous pouvons donc appliquer la proposition 1.14, et conclure que pour chaque valeur de x_2, x_4, le système a une unique solution en x_1, x_3. Après quelques calculs, nous obtenons $x_3 = b_2 - 3x_4$ et $x_1 = b_1 + b_2 - x_2 - 6x_4$. L'ensemble des solutions est donc

$$\{(b_1 + b_2 - x_2 - 6x_4,\ x_2,\ b_2 - 3x_4,\ x_4) / \ x_2 \in \mathbf{R},\ x_4 \in \mathbf{R}\}$$

Nous disons qu'il est paramétré par (x_2, x_4). Et nous appelons x_2 et x_4 des variables *libres*, car leur valeur est arbitraire.

La méthode suivie s'applique à tout système échelonnée $AX = B$. Lorsque le second membre B a un coefficient non-nul dans une des lignes nulles de A, le système n'a pas de solution. Sinon, on résout en faisant passer à droite les inconnues dans des colonnes sans pivots, ce sont les variables libres. Et l'on obtient un système triangulaire, avec sur la diagonale les pivots du système initial. Nous avons donc le résultat suivant.

Énoncé indispensable 12 :

PROPOSITION 1.17. *Soit le système $AX = B$ avec p équations et q inconnues. Supposons que A est une matrice échelonnée à r pivots. Alors*

1. *si l'un des $(p - r)$ derniers coefficients de B est non-nul, le système n'a pas de solution.*
2. *sinon*
   (a) *lorsque $q = r$, il y a une unique solution.*
   (b) *lorsque $q > r$, l'ensemble des solutions est infini, il est paramétré par $q - r$ variables libres, qui sont les inconnues dont la colonne n'a pas de pivot.*

Parmi les matrices échelonnées, certaines donnent lieu à des systèmes particulièrement simple à résoudre.

DEFINITION 1.18. On dit qu'une matrice échelonnée A est *réduite* si chaque pivot p_i est égal à 1, et si chaque pivot est le seul élément non-nul de sa colonne.

En effet, considérons un système $AX = B$ avec A échelonnée réduite. Supposons qu'il existe des solutions et passons les variables libres dans les membres de droite. Alors on obtient un système triangulaire de matrice l'identité. Par exemple

$$\begin{pmatrix} 1 & 2 & 0 & 3 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 5 \\ 6 \end{pmatrix} \quad \Leftrightarrow \quad \begin{cases} x_1 + 2x_2 \quad + 3x_4 = 5 \\ \qquad\qquad x_3 + \ x_4 = 6 \end{cases}$$

$$\Leftrightarrow \quad \begin{cases} x_1 = -2x_2 - 3x_4 + 5 \\ x_2 \text{ est libre} \\ x_3 = -x_4 + 6 \\ x_4 \text{ est libre} \end{cases}$$

Le nombre r de pivots d'une matrice échelonnée de taille $p \times q$ est plus petit que p et q. Il est donc maximal lorsqu'il est égal à p ou q. Lorsque $r = q$, chaque colonne a un pivot. Lorsque $r = p$, chaque ligne a un pivot. Ceci se traduit par des propriétés d'existence ou unicité pour les solutions des systèmes de matrice A.

PROPOSITION 1.19. *Soit une matrice A de taille $p \times q$ échelonnée avec r pivots. Alors*

1. *$r = p$ ssi pour tout $B \in \mathbf{R}^p$, le système $AX = B$ admet une solution,*

2. *$r = q$ ssi le système homogène $AX = 0$ a pour unique solution $X = 0$,*
3. *$r = p = q$ ssi pour tout $B \in \mathbf{R}^p$, le système $AX = B$ admet une unique solution.*

Démonstration. Cela découle immédiatement de la proposition 1.17. Si $r = p$, A n'a pas de ligne nulle et donc $AX = B$ admet une solution. Si par contre $p > r$ et que l'on choisit pour B le vecteur dont toutes les coordonnées sont nulles sauf la dernière qui vaut 1, alors $AX = B$ n'a pas de solution. Le vecteur $X = 0$ est bien solution du système homogène $AX = 0$. D'après la deuxième partie de la proposition 1.17, cette solution est unique précisément lorsque $r = q$. On raisonne de même pour la dernière équivalence. □

1.2.2 *L'algorithme du pivot de Gauss*

Comme nous le verrons dans le chapitre suivant, tout système peut être transformé en un système équivalent dont la matrice est échelonnée. Ainsi l'étude faite dans le chapitre précédent nous donnera les solutions d'un système quelconque. Dans ce chapitre, nous présentons la méthode pour nous ramener à un système échelonnée. Pour simplifier l'exposition, nous oublions momentanément les systèmes et ne considérons que des matrices.

On définit des opérations sur les lignes des matrices. Chaque opération transforme une matrice de taille $p \times q$ en une nouvelle matrice de taille $p \times q$. Les opérations sont les suivantes :

Énoncé indispensable 13 :

1. *Dilatation*. Multiplier la i-ième ligne par un réel α non-nul. On la note $L_i \to \alpha L_i$.
2. *Transvection*. Remplacer la i-ième ligne L_i par $L_i' = L_i + \lambda L_j$ pour $j \neq i$ et λ un réel. On la note $L_i + \lambda L_j \to L_i$.
3. *Permutation*. Echanger les lignes i et j. On la note $L_i \leftrightarrow L_j$.

On les appelle *opérations élémentaires*.

La première propriété importante est que les opérations élémentaires sont réversibles.

proposition 1.20. *Si une matrice A' est obtenue à partir de A par une opération élémentaire, le contraire est vrai : A est obtenue à partir de A' par une opération élémentaire.*

Démonstration. Si on a effectué $L_i \to \alpha L_i$, on revient à la matrice initiale en effectuant $L_i \to \frac{1}{\alpha} L_i$. Si on a effectué $L_i \to L_i + \lambda L_j$, on revient à la matrice initiale en effectuant $L_i \to L_i - \lambda L_j$. Enfin, si on a effectué $L_i \leftrightarrow L_j$, on revient à la matrice initiale par la même opération. □

Le résultat fondamental est le suivant.

Énoncé indispensable 14 :

THÉORÈME 1.21. *Tout matrice peut être transformée en une matrice échelonnée réduite par un nombre fini d'opérations élémentaires.*

Démonstration. On procède en deux étapes : dans la première, on se ramene à une matrice échelonnée, dans la seconde à une matrice échelonnée réduite. Soit C la première colonne non-nulle de la matrice. Quitte à échanger les lignes, on peut supposer que le coefficient de C sur la première ligne est non-nul. Notons le p_1 et $a_2, a_3, \ldots$ les coefficients situés dessous.

$$\begin{pmatrix} 0 & p_1 & * & * & * & * \\ 0 & a_2 & * & * & * & * \\ 0 & a_3 & * & * & * & * \\ 0 & a_4 & * & * & * & * \end{pmatrix},$$

On effectue alors successivement les transvections :

- $L_2 - \frac{a_2}{p_1} L_1 \to L_2$ a pour effet de mettre un zéro sur la ligne 2, juste en dessous de p_1.
- $L_3 - \frac{a_3}{p_1} L_1 \to L_3$ a pour effet de mettre un zéro sur la ligne 3, en dessous de p_1. On continue avec les lignes suivantes $L_4, \ldots, L_p$.

Quand on a terminé cette étape, on met la ligne L_1 de côté et l'on recommence avec la matrice plus petite qui reste.

$$\begin{pmatrix} 0 & p_1 & * & * & * & * \\ 0 & 0 & * & * & * & * \\ 0 & 0 & * & * & * & * \\ 0 & 0 & * & * & * & * \end{pmatrix}.$$

On traite successivement chaque ligne jusqu'à ce qu'il ne reste que des lignes nulles. La matrice obtenue est échelonnée.

$$\begin{pmatrix} 0 & p_1 & * & * & * & * \\ 0 & 0 & p_2 & * & * & * \\ 0 & 0 & 0 & 0 & p_3 & * \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Il reste à la réduire, c'est-à-dire annuler les coefficients au dessus de chaque pivot et normaliser les pivots (les rendre égaux à 1). Pour ce faire, on commence par la dernière colonne D ayant un pivot. Par une dilatation, on normalise ce pivot, puis on annule les coefficients au dessus par des transvections successives.

$$\begin{pmatrix} 0 & p_1 & * & * & 0 & * \\ 0 & 0 & p_2 & * & 0 & * \\ 0 & 0 & 0 & 0 & 1 & * \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

On traite ainsi chaque colonne avec un pivot en remontant de droite à gauche, ce qui

nous donne notre matrice échelonnée réduite.

$$\begin{pmatrix} 0 & 1 & 0 & * & 0 & * \\ 0 & 0 & 1 & * & 0 & * \\ 0 & 0 & 0 & 0 & 1 & * \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

L'algorithme décrit a deux phases, la première descendante de gauche à droite, la seconde ascendante de droite à gauche. L'ordre des opérations est important pour ne pas défaire ce qui a été fait : une fois un coefficient annulé ou un pivot normalisé, les opérations ultérieures ne doivent plus le modifier. □

remarque 1.22. sur la preuve ci-dessus.

1. La preuve donne un algorithme, c'est-à-dire une méthode automatique qui méne au résultat voulu. Nous pouvons faire réaliser ces calculs par un ordinateur. Cet algorithme porte le nom d'algorithme de Gauss.

2. pour certains problèmes, nous avons seulement besoin de connaitre le nombre de pivots et leur position. Pour cela il suffit de réaliser la première phase de l'algorithme, la phase descendante. En effet, la deuxième phase qui consiste à réduire la matrice échelonnée ne change pas le nombre de pivot ni leur position.

3. la normalisation des pivots pourrait être faite à d'autre moment de l'algorithme. Il est même possible de ne pas le faire si l'on souhaite seulement obtenir une matrice échelonnée. Dans ce cas, les opérations de dilatations ne sont pas écessaires.

Appliquons l'algorithme de Gauss sur un exemple.

exemple 1.23. Échelonnons la matrice

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 1 \\ 2 & 2 & 2 & 1 & 3 & 4 \\ 5 & 5 & 7 & 8 & 9 & 9 \end{pmatrix}.$$

La première colonne est non-nulle, comme son premier coefficient. C'est notre premier pivot, on l'entoure.

$$\begin{pmatrix} \boxed{1} & 1 & 1 & 0 & 1 & 1 \\ 2 & 2 & 2 & 1 & 3 & 4 \\ 5 & 5 & 7 & 8 & 9 & 9 \end{pmatrix}.$$

On effectue ensuite les deux opérations $L_2 - 2L_1 \to L_2$ et $L_3 - 5L_1 \to L_3$, pour obtenir

$$\begin{pmatrix} \boxed{1} & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & 2 & 8 & 4 & 4 \end{pmatrix}.$$

La matrice plus petite, formée des deux dernières lignes, comporte deux première colonnes nulles. La troisième colonne, non-nulle, nous offre un pivot que l'on entoure. Il faut faire un échange de ligne $L_2 \leftrightarrow L_3$ pour le faire remonter. On trouve :

$$\begin{pmatrix} \boxed{1} & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & \boxed{2} & 8 & 4 & 4 \\ 0 & 0 & 0 & 1 & 1 & 2 \end{pmatrix}.$$

La matrice est alors échelonnée : toutes les lignes commencent par un pivot, que l'on entoure.

$$\begin{pmatrix} \boxed{1} & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & \boxed{2} & 8 & 4 & 4 \\ 0 & 0 & 0 & \boxed{1} & 1 & 2 \end{pmatrix}.$$

Nous pouvons à présent réduire la matrice en suivant la deuxième phase de l'algorithme de Gauss. Pour annuler le coefficient au dessus du dernier pivot, nous effectuons l'opération $L_2 \to L_2 - 8L_3$.

$$\begin{pmatrix} \boxed{1} & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & \boxed{2} & 0 & -4 & -12 \\ 0 & 0 & 0 & \boxed{1} & 1 & 2 \end{pmatrix}.$$

Nous normalisons ensuite le deuxième pivot par la dilatation $L_2 \to \frac{1}{2}L_2$

$$\begin{pmatrix} \boxed{1} & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & \boxed{1} & 0 & -2 & -6 \\ 0 & 0 & 0 & \boxed{1} & 1 & 2 \end{pmatrix}.$$

Puis nous annulons le coefficient au dessus du premier pivot par la transvection $L_1 \to L_1 - L_2$

$$\begin{pmatrix} \boxed{1} & 1 & 0 & 0 & 3 & 7 \\ 0 & 0 & \boxed{1} & 0 & -2 & -6 \\ 0 & 0 & 0 & \boxed{1} & 1 & 2 \end{pmatrix}.$$

Nous avons terminé, la matrice est échelonnée réduite. □

Nous terminons par une définition très importante.

Énoncé indispensable 15 :

DEFINITION 1.24. Le rang d'une matrice A est le nombre r de pivots d'une matrice échelonnée obtenue à partir de A par une suite d'opérations élémentaires sur les lignes.

Pour que cette définition soit acceptable, il faut s'assurer que toutes les matrices échelonnées obtenues à partir de A par des opérations élémentaires ont le même nombre de pivots. C'est bien vrai mais ce n'est pas du tout évident. Ce sera démontré plus tard, cf. corollaire 2.26.

1.2.3 *Résolution des systèmes*

Nous allons maintenant résoudre un système linéaire quelconque. La méthode va consister à effectuer des opérations sur les équations de sorte à obtenir un système

échelonné. Il convient d'être très prudent dans ces opérations, pour s'assurer que le nouveau système a exactement les mêmes solutions que le système initial.

exemple 1.25. Pour se convaincre qu'il faut faire attention, proposons une solution (erronnée) du système suivant

$$\begin{cases} 2x+2y+z=0 \\ x+2y+2z=0 \\ x+y+z=0 \end{cases}$$

On remplace la ligne L_1 par $L_1 - L_2$ pour éliminer y, L_2 par $L_2 - L_3$ pour éliminer x et L_3 par $L_3 - L_1$ pour éliminer z. On obtient

$$\begin{cases} x-z=0 \\ y+z=0 \\ -x-y=0 \end{cases}$$

Ce système équivaut à $x = -y = z$. Donc $(x, y, z) = (1, -1, 1)$ est solution. Mais ce triplet n'est pas solution du système initial... Où est l'erreur ? □

Soit une matrice A de taille $p \times q$. On associe au système $AX = B$ la matrice de taille $p \times (q+1)$ formée de A et B en dernière colonne. On la note $(A|B)$ et on l'appelle *matrice augmentée* du système. Par exemple,

$$(1.2)\quad \begin{cases} x_1 + x_2 + x_3 + x_5 = 1 \\ 2x_1 + 2x_2 + 2x_3 + x_4 + 3x_5 = 4 \\ 5x_1 + 5x_2 + 7x_3 + 8x_4 + 9x_5 = 9 \end{cases}, \qquad (A|B) = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 1 \\ 2 & 2 & 2 & 1 & 3 & 4 \\ 5 & 5 & 7 & 8 & 9 & 9 \end{pmatrix}.$$

La donnée de la matrice augmentée est équivalente à celle du système, chaque ligne correspondant à une équation du système. On parlera donc indifféremment du système ou de la matrice augmentée.

Rappelons que nous avons défini dans le chapitre précédent des opérations élémentaires sur les lignes des matrices.

Énoncé indispensable 16 :

proposition 1.26. *Soient deux systèmes $(\mathcal{S})$ et $(\mathcal{S}')$. Si l'on passe de la matrice augmentée de $(\mathcal{S})$ à celle de $(\mathcal{S}')$ par une opération élémentaire, alors $(\mathcal{S})$ et $(\mathcal{S}')$ ont même ensemble de solution, on dit qu'ils sont équivalents.*

Démonstration. On remarque que chaque solution du système initial $(\mathcal{S})$ est solution du système modifié $(\mathcal{S}')$. Comme $(\mathcal{S})$ se retrouve à partir de $(\mathcal{S}')$ par une des opérations élémentaires, cf. proposition 1.20, toute solution de $(\mathcal{S}')$ est aussi solution de $(\mathcal{S})$. □

Pour résoudre $AX = B$, on applique l'algorithme de Gauss à sa matrice augmentée $(A|B)$. Cela nous fournit un système équivalent $A'X = B'$ dont la matrice augmentée $(A'|B')$ est échelonnée. A' est alors elle-aussi échelonnée, nous pouvons donc résoudre

$A'X = B'$ comme dans le chapitre 1.2.1. Rappelons que ce système a une solution lorsque les coefficients de B' situés dans les lignes nulles de A' sont nuls. Cela revient à demander que $(A'|B')$ n'ait pas de pivot dans sa dernière colonne. Lorsque c'est le cas, on exprime les solutions en fonction des variables libres.

exemple 1.27. Considèrons le système (1.2). D'après les calculs de l'exemple 1.23, la matrice $(A|B)$ se transforme par des opérations élémentaires en la matrice

$$(A'|B') = \begin{pmatrix} \boxed{1} & 1 & 0 & 0 & 3 & 7 \\ 0 & 0 & \boxed{1} & 0 & -2 & -6 \\ 0 & 0 & 0 & \boxed{1} & 1 & 2 \end{pmatrix}.$$

qui correspond au système

$$\begin{cases} x_1 + x_2 \qquad + 3x_5 = 7 \\ \qquad\quad x_3 \quad - 2x_5 = -6 \\ \qquad\qquad x_4 + \ x_5 = 2 \end{cases}$$

Les solutions sont

$$\begin{cases} x_1 = -x_2 - 3x_5 + 7 \\ x_2 \text{ est libre} \\ x_3 = 2x_5 - 6 \\ x_4 = -x_5 + 2 \\ x_5 \text{ est libre} \end{cases}$$

Les nombre de variables libres est $2 = 5 - 3$ où 5 est le nombre de colonnes de A est 3 le nombre de pivots de A', c'est-à-dire le rang de A. □

Si la matrice est de taille $p \times q$ et de rang r, le nombre de variables libres est $q - r$. Parfois, on souhaite savoir pour quels B le système $AX = B$ a une solution. Pour cela, il suffit d'échelonner partiellement la matrice augmentée $(A|B)$ avec un B général puis d'appliquer le critère de la proposition 1.17. Regardons sur un exemple.

exemple 1.28. Soit le système

$$(1.3) \quad \begin{cases} 2x - y + 3z = a \\ -4x + 2y - 6z = b \\ 2x - y + z = c \\ 6x - 3y + 3z = d \end{cases}$$

Échelonnons les 3 premières colonnes de la matrice augmentée.

$$\begin{pmatrix} 2 & -1 & 3 & a \\ -4 & 2 & -6 & b \\ 2 & -1 & 1 & c \\ 6 & -3 & 3 & d \end{pmatrix} \rightarrow \begin{pmatrix} 2 & -1 & 3 & a \\ 0 & 0 & 0 & b+2a \\ 0 & 0 & -2 & c-a \\ 0 & 0 & -6 & d-3a \end{pmatrix}$$

$$\rightarrow \begin{pmatrix} 2 & -1 & 3 & a \\ 0 & 0 & -2 & c-a \\ 0 & 0 & 0 & b+2a \\ 0 & 0 & -6 & d-3a \end{pmatrix} \rightarrow \begin{pmatrix} 2 & -1 & 3 & a \\ 0 & 0 & -2 & c-a \\ 0 & 0 & 0 & b+2a \\ 0 & 0 & 0 & d-3c \end{pmatrix}$$

Nous arrêtons l'algorithme de Gauss. Le système (1.3) équivaut alors au système

$$\begin{cases} 2x - y + 3z &= a \\ -2z &= c-1 \\ 0 &= b+2a \\ 0 &= d-3c \end{cases}$$

qui a une solution si et seulement si les deux dernières équations sont vérifiées. □

La méthode est générale : un système $AX = B$ admet des solutions si les coefficients de B sont eux-même solution d'un système homogène à $p - r$ équations. On appelle parfois ces équations les *équations de compatibilité* car elles sont vérifiées lorsque le système initial a une solution, c'est-à-dire lorsque les équations de ce système sont compatibles entre elles.

La caractérisation des systèmes échelonnés avec un nombre maximal de pivot s'étend sans changement aux systèmes généraux.

Énoncé indispensable 17 :

proposition 1.29. *Soit une matrice A de taille $p \times q$ et de rang r. Alors*

1. *$r = p$ ssi pour tout $B \in \mathbf{R}^p$, le système $AX = B$ admet une solution,*
2. *$r = q$ ssi le système homogène $AX = 0$ a pour unique solution $X = 0$,*
3. *$r = p = q$ ssi pour tout $B \in \mathbf{R}^p$, le système $AX = B$ admet une unique solution.*

Démonstration. Soit A' la matrice échelonnée à r pivot obtenue à partir de A par des opérations élémentaires. Remarquons que ces mêmes opérations transforme la matrice augmentée $(A|0)$ en $(A'|0)$. Donc les systèmes homogènes $AX = 0$ et $A'X = 0$ sont équivalents. Donc la deuxième équivalence découle de la deuxième équivalence de la proposition 1.19.

De même, pour tout second membre B, les opérations qui transforment A en A' transforment $(A|B)$ en $(A'|B')$ avec B' un vecteur colonne. Donc les systèmes $AX = B$ et $AX = B'$ sont équivalents. Nous en déduisons que si $A'X = B'$ admet une solution pour tout B', alors $AX = B$ admet une solution pour tout B. La réciproque est vraie par le même argument car l'on passe aussi de A' à A par des opérations élémentaires, cf.

proposition 1.20. Ainsi la première équivalence découle de la première équivalence de la proposition 1.19.

La preuve de la troisième équivalence est similaire. □

Pour résumer, si A est de taille $p \times q$ et de rang r, alors

1. l'ensemble des solutions de $AX = 0$ est paramétré par $q - r$ variables libres ; et lorsque $q = r$, $X = 0$ est l'unique solution.
2. le système $AX = B$ admet (au moins) une solution lorsque B vérifie $p - r$ équations de compatibilité ; lorsque $p = r$, le système $AX = B$ a une solution quel que soit B.

De manière vague, nous pouvons mesurer la taille de l'ensemble des solutions du système homogène par le nombre de variables libres. Ainsi, plus le rang est grand, plus l'ensemble des solutions est petit. Et à l'opposé, lorsque le rang croit, le nombre d'équations de compatibilité diminue, donc l'ensemble des seconds membres tels que $AX = B$ a une solution grossit. Nous donnerons un sens précis à ces assertions plus loin : d'après le théorème 2.24, l'ensemble des solutions du système homogène $AX = 0$ est un espace vectoriel de dimension $q - r$ et l'ensemble des second membres B tels que $AX = B$ a une solution est un espace vectoriel de dimension r.

1.2.4 *Matrices inversibles*

Énoncé indispensable 18 :

THÉORÈME 1.30. *Soit A une matrice carrée de taille n. Alors*

$$A \text{ est inversible} \quad \Leftrightarrow \quad A \text{ est de rang } n.$$

Démonstration. Si A est inversible, nous avons vu dans la proposition 1.12 que le système $AX = B$ a une unique solution pour tout second membre B. Ceci implique que $n = r$ par la proposition 1.29.

La preuve de la réciproque repose sur le fait essentiel suivant : chaque opération élémentaire sur les lignes de A consiste à changer A par OA où O est une matrice inversible de taille n bien choisie. Par exemple, si l'opération est la dilatation de la i-ième ligne de rapport α, O est la matrice diagonale dont tous les coefficients de la diagonale sont égaux à 1 sauf le i-ième qui vaut α. De même, pour échanger les deux premières lignes ou bien pour remplacer L_1 par $L_1 + \lambda L_2$, on multiplie A à gauche par les matrices

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & \lambda & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad \text{si } n = 4.$$

Donc par hypothèse, il existe des matrices inversibles $O_1, \dots O_\ell$ telles que pour $O = O_1 \dots O_\ell$, la matrice OA soit échelonnée avec r pivots. Comme nous l'avons vu dans la preuve du théorème 1.21, quitte à rajouter quelques opérations, nous pouvons même supposer que OA est échelonné réduite avec le même nombre de pivot r. Si $r = n$,

nous obtenons $OA = \mathrm{Id}$. Par la proposition 1.7, O est inversible car c'est un produit de matrices inversibles. En multiplicant à gauche $OA = \mathrm{Id}$ par O^{-1}, il vient $A = O^{-1}$. La seconde partie de la proposition 1.7 montre alors que A est inversible d'inverse O. □

Énoncé indispensable 19 :

corollaire 1.31. *Soit A et A' deux matrices carrées de même taille. Alors les assertions suivantes sont équivalentes*

1. $A'A = \mathrm{Id}$
2. $AA' = \mathrm{Id}$
3. *A est inversible d'inverse A'.*

Démonstration. $(1 \Rightarrow 3)$ Supposons que $A'A = \mathrm{Id}$. Alors $X = 0$ est l'unique solution de $AX = 0$. En effet, en multipliant à gauche $AX = 0$ par A', il vient $\mathrm{Id}\, X = A'0$, autrement dit $X = 0$. D'après la deuxième équivalence de la proposition 1.29, cela montre que $r = n$. Le théorème 1.30 implique alors que A est inversible. En multipliant à droite $A'A = \mathrm{Id}$ par A^{-1}, il vient $A' = A^{-1}$.

La réciproque $(3 \Rightarrow 1)$ est claire, puisque par définition A est inversible d'inverse A' signifie que ($AA' = \mathrm{Id}$ et $A'A = \mathrm{Id}$). Donc $(1 \Leftrightarrow 3)$.

Cela montre aussi que $(2 \Leftrightarrow 3)$. En effet, A est inversible d'inverse A' ssi ($AA' = \mathrm{Id}$ et $A'A = \mathrm{Id}$) ssi A' est inversible d'inverse A. □

Expliquons comment l'on calcule l'inverse d'une matrice par l'algorithme de Gauss. Pour $i = 1, \ldots, n$, notons e_i le vecteur de $\mathbf{R}^n$ dont tous les coefficients sont nuls sauf le i-ième qui vaut 1. Supposons que A est une matrice inversible de taille n. Alors le i-ième vecteur colonne de A^{-1} est la solution du système $AX = e_i$. Cela découle de $AA^{-1} = \mathrm{Id}$ et du fait que e_i est le i-ième vecteur colonne de la matrice identité.

Donc calculer A^{-1} revient à résoudre les n systèmes $AX = e_i$, $i = 1, \ldots, n$. Au lieu d'appliquer n fois l'algorithme de Gauss, on les résout tous ensemble en appliquant l'algorithme de Gauss à la matrice A augmentée de la matrice identité.

exemple 1.32. Calculons l'inverse de $A = \begin{pmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{pmatrix}$. La matrice augmentée $(A|\,\mathrm{Id})$

est

$$\begin{pmatrix} 2 & 1 & 1 & 1 & 0 & 0 \\ 4 & -6 & 0 & 0 & 1 & 0 \\ -2 & 7 & 2 & 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 2 & 1 & 1 & 1 & 0 & 0 \\ 0 & -8 & -2 & -2 & 1 & 0 \\ 0 & 8 & 3 & 1 & 0 & 1 \end{pmatrix}$$
$$\to \begin{pmatrix} 2 & 1 & 1 & 1 & 0 & 0 \\ 0 & -8 & -2 & -2 & 1 & 0 \\ 0 & 0 & 1 & -1 & 1 & 1 \end{pmatrix}$$
$$\to \begin{pmatrix} 2 & 1 & 0 & 2 & -1 & -1 \\ 0 & -8 & 0 & -4 & 3 & 2 \\ 0 & 0 & 1 & -1 & 1 & 1 \end{pmatrix}$$
$$\to \begin{pmatrix} 2 & 0 & 0 & \frac{12}{8} & -\frac{5}{8} & -\frac{6}{8} \\ 0 & -8 & 0 & -4 & 3 & 2 \\ 0 & 0 & 1 & -1 & 1 & 1 \end{pmatrix}$$
$$\to \begin{pmatrix} 1 & 0 & 0 & \frac{12}{16} & -\frac{5}{16} & -\frac{6}{16} \\ 0 & 1 & 0 & \frac{4}{8} & -\frac{3}{8} & -\frac{2}{8} \\ 0 & 0 & 1 & -1 & 1 & 1 \end{pmatrix}$$

La matrice obtenue n'est autre que $(\mathrm{Id}\,|A^{-1})$, autrement dit

$$A^{-1} = \begin{pmatrix} \frac{12}{16} & -\frac{5}{16} & -\frac{6}{16} \\ \frac{4}{8} & -\frac{3}{8} & -\frac{2}{8} \\ -1 & 1 & 1 \end{pmatrix}$$

Énoncé indispensable 20 :

proposition 1.33. *Soit A une matrice triangulaire.*

1. *A est inversible si et seulement si tous les coefficients de sa diagonale sont non-nuls.*
2. *Si A est inversible, alors A^{-1} est aussi triangulaire.*

Démonstration. Supposons A de taille n. Si les coefficients de la diagonales sont non-nuls, A est échelonnée avec n pivots, donc inversible par le théorème 1.30. Si par contre un des coefficients est nul, alors par l'algorithme de Gauss, A est transformée en une matrice avec $r < n$ pivots. Donc A n'est pas inversible par le théorème 1.30.

Lorsque A est inversible, on peut trouver son inverse comme cela a été expliqué ci-dessus en appliquant l'algorithme de Gauss à la matrice augmentée $(A|\,\mathrm{Id})$. Cette matrice étant déjà échelonnée, nous procédons directement à la phase ascendante de la réduction. Nous affirmons qu'à toutes les étapes, la matrice sera de la forme $(B|C)$ avec B et C triangulaires. En effet, les opérations à effectuer sont des dilatations et des transvections $L_i \to L_i + \lambda L_j$ avec $i < j$. Ces opérations transforment une matrice triangulaire en une autre matrice triangulaire. □

1.3 Déterminants

Le déterminant est un nombre associé à chaque matrice carrée qui a des propriétés magiques. La théorie pour les matrices 2×2 sera traitée dans un premier temps car elle est très simple et sert de motivation pour comprendre la généralisation aux matrices de taille supérieure.

1.3.1 *Matrices* 2×2

Le déterminant de la matrice $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ est le nombre défini par la formule :

$$\det A = ad - bc. \tag{1.4}$$

Il vérifie des propriétés algébriques très intéressantes.

Énoncé indispensable 21 :

1. A est inversible ssi son déterminant est non-nul.
2. Lorsque A est inversible,
$$A^{-1} = \frac{1}{\det A} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$
3. Si B est une autre matrice 2×2,
$$\det(AB) = (\det A)(\det B)$$
et lorsque A est inversible
$$\det(A^{-1}) = (\det A)^{-1}.$$
4. Si A est inversible, la solution de $A \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix}$ est
$$x = \frac{\det \begin{pmatrix} f & b \\ g & d \end{pmatrix}}{\det A}, \qquad y = \frac{\det \begin{pmatrix} a & f \\ c & g \end{pmatrix}}{\det A}$$

Démonstration. Vérifions pour commencer que $\det(AB) = (\det A)(\det B)$.

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \qquad B = \begin{pmatrix} p & q \\ r & s \end{pmatrix} \qquad \Rightarrow \qquad AB = \begin{pmatrix} ap + br & aq + bs \\ cp + dr & cq + ds \end{pmatrix}$$

et donc

$$\begin{aligned} \det(AB) &= (ap + br)(cq + ds) - (aq + bs)(cp + dr) \\ &= apds + brcq - aqdr - bscp \end{aligned}$$

où nous avons simplifié 4 termes. D'autre part

$$\begin{aligned}(\det A)(\det B) &= (ad - bc)(ps - qr) \\ &= adps - adqr - bcps + bcqr\end{aligned}$$

et l'on retrouve bien la même expression que pour $\det(AB)$.

Remarquons que $\det(\mathrm{Id}) = 1$, donc lorsque A est inversible, $AA^{-1} = \mathrm{Id}$ entraine que $(\det A)\det(A^{-1}) = 1$ et donc $\det(A^{-1})$ est non-nul et est égal à $(\det A)^{-1}$. Par ailleurs, sans supposer A inversible, un calcul direct montre que pour $C = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$

$$AC = CA = (\det A)\,\mathrm{Id}$$

et donc lorsque $\det A \neq 0$, A est inversible d'inverse $(\det A)^{-1}C$. Nous avons montré les trois premières propriétés. Vérifions la dernière. Nous avons

$$\begin{aligned}a \det \begin{pmatrix} f & b \\ g & d \end{pmatrix} + b \det \begin{pmatrix} a & f \\ c & g \end{pmatrix} &= a(fd - bg) + b(ag - fc) = afd - bfc = f(ad - bc) \\ &= f \det A\end{aligned}$$

et de même,

$$c \det \begin{pmatrix} f & b \\ g & d \end{pmatrix} + d \det \begin{pmatrix} a & f \\ c & g \end{pmatrix} = \ldots = g \det A.$$

□

Le déterminant a aussi une expression géométrique très importante, même si nous ne l'utiliserons pas dans ce cours. Si les deux vecteurs colonnes u et v de la matrice A sont non-nuls,

$$\det A = \|u\|\,\|v\|\,\sin\theta \tag{1.5}$$

où θ est l'angle orienté formé par u et v. Ici, la convention pour l'orientation est que l'angle croit lorsque u est fixé et v se déplace dans le sens contraire des aiguilles d'une montre. Les nombres $\|u\|$ et $\|v\|$ sont les longeurs des vecteurs définies par la formule habituelle

$$\|w\| = \sqrt{x^2 + y^2} \qquad \text{si} \qquad w = \begin{pmatrix} x \\ y \end{pmatrix}.$$

Par (1.5), l'on interprète nullité, signe et valeur absolue du déterminant de la manière suivante :

1. $\det A$ est nul si et seulement si u et v sont colinéaires.
2. Lorsque u et v ne sont pas colinéaires et donc forment une base, $\det A$ est positif si et seulement si (u, v) est une base directe.
3. $|\det A|$ est l'aire du parallélogramme engendré par u et v.

Pour ces raisons, on pense souvent le déterminant des matrices 2×2 comme une aire orientée.

Démonstration. Pour montrer la formule (1.5), introduisons le vecteur v' image de v par une rotation d'angle $\pi/2$. Les coordonnées de v' sont $-d$ et b, et donc le produit scalaire

de u et v' vaut

$$\langle u, v' \rangle = -ad + bc = -\det A.$$

Par ailleurs, v et v' ont même norme et l'angle entre u et v' est $\theta + \pi/2$, donc

$$\langle u, v' \rangle = \|u\| \, \|v'\| \cos(\theta + \pi/2) = -\|u\| \, \|v\| \sin\theta.$$

En comparant avec la formule précédente, on obtient (1.5).

De (1.5), on déduit que $\det A = 0$ si $u = 0$ ou $v = 0$ ou les deux vecteurs sont non-nuls et forment un angle égal à 0 modulo π. Cela revient à dire que u et v sont colinéaires. Pour l'interprétation du signe, il suffit de se rappeler que $\sin\theta$ est positif lorsque $\theta \in]0, \pi[$, négatif lorsque $\theta \in]-\pi, 0[$. Enfin, le fait que $|\det A|$ soit l'aire du parallélogramme se déduit facilement de (1.5) par la formule "aire = produit de la base par la hauteur". □

Dans la suite de ce chapitre, nous allons généraliser ce qui précède aux matrices carrées de taille supérieure. Pour une matrice 3×3, le déterminant est défini par

$$\det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = a_{11}a_{22}a_{33} - a_{21}a_{12}a_{33} + a_{31}a_{12}a_{23} - a_{11}a_{32}a_{23} + a_{21}a_{32}a_{13} - a_{31}a_{22}a_{13} \tag{1.6}$$

et toutes les propriétés algébriques énoncées précédemment se généralisent. De plus, la valeur absolue de ce déterminant est le volume du parallélépipède engendré par les 3 vecteurs colonnes de la matrice.

Evidemment, on se demande bien pourquoi la formule (1.6) aurait de telles propriétés. L'approche que nous avons suivie pour les matrices 2×2 n'est plus du tout adaptée. Par exemple, pour vérifier la simple formule $\det(AB) = (\det A)(\det B)$, chaque coefficient de AB étant la somme de trois termes, $\det(AB)$ est la somme de $3 \times 6 = 18$ termes, chacun produit de 3 coefficients de A et 3 coefficients et B, et nous devons comparer ceci avec $(\det A)(\det B)$ qui se développe en une somme de $6 \times 6 = 36$ termes...

En fait, au lieu de travailler directement à partir de cette définition, il sera plus simple de dégager quelques propriétés qui caractérisent le déterminant et de tout en déduire. Ces propriétés nous disent comment varie le déterminant lorsque l'on transforme de certaines façons la matrice. Pour le déterminant des matrices 2×2, ces propriétés sont les suivantes :

$$\det \begin{pmatrix} c & d \\ a & b \end{pmatrix} = -\det \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \qquad \det \begin{pmatrix} \lambda a & \lambda b \\ c & d \end{pmatrix} = \lambda \det \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

$$\det \begin{pmatrix} a + a' & b + b' \\ c & d \end{pmatrix} = \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} + \det \begin{pmatrix} a' & b' \\ c & d \end{pmatrix}, \qquad \det \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 1$$

En exercice, vous pouvez montrer que le déterminant, tel que nous l'avons défini par la formule (1.4), vérifie bien ces équations, et qu'elles le caractérisent.

1.3.2 *Déterminant d'une matrice $n \times n$*

Énoncé indispensable 22 :

THÉORÈME 1.34. *Il existe une unique application* $\det : \mathcal{M}_n \to \mathbf{R}$*, appelée déterminant, qui vérifie les propriétés suivantes :*

1. **Permutation** *: si on échange deux lignes d'une matrice, son déterminant est multiplié par* -1*.*
2. **Dilatation** *: si on multiplie une ligne d'une matrice par un réel* λ*, son déterminant est multiplié par* λ*.*
3. **Additivité** *: si trois matrices* A*,* B*,* C *sont telles qu'il existe* $i \in \{1, \dots, n\}$ *pour lequel la* i*-ième ligne de* A *est la somme des* i*-ième lignes de* B *et* C *et pour tout* $j \neq i$*, les* j*-ièmes lignes de* A*,* B *et* C *sont les mêmes, alors* $\det A = \det B + \det C$*.*
4. **Normalisation** *: le déterminant de la matrice identité vaut* 1*.*

Ce théorème est (momentanément) admis. En fait, nous verrons très rapidement comment calculer le déterminant à partir de ces quatre propriétés, ce qui impliquera l'unicité de l'application déterminant. Pour l'existence, on définit le déterminant par une formule relativement compliquée, généralisant les expressions (1.4) et (1.6), que nous présenterons brièvement dans la partie 1.3.7. Mais la compréhension de cette formule suppose une certaine maturité, et on ne l'utilise que rarement, ce qui explique pourquoi nous ne la donnons pas tout de suite.

Pour que les conditions énoncées dans le théorème 1.34 soient bien claires, explicitons les pour une matrice 3×3. Par exemple, si on échange la première et troisième lignes, nous avons

$$\det \begin{pmatrix} a_3 & b_3 & c_3 \\ a_2 & b_2 & c_2 \\ a_1 & b_1 & c_1 \end{pmatrix} = -\det \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}$$

si on multiplie la deuxième ligne par $\lambda \in \mathbf{R}$,

$$\det \begin{pmatrix} a_1 & b_1 & c_1 \\ \lambda a_2 & \lambda b_2 & \lambda c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} = \lambda \det \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}$$

et la propriété d'additivité pour la troisième ligne s'écrit :

$$\det \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 + a_3' & b_3 + b_3' & c_3 + c_3' \end{pmatrix} = \det \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} + \det \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3' & b_3' & c_3' \end{pmatrix}.$$

REMARQUE 1.35.

1. le déterminant d'une matrice diagonale est égale au produit de ses coefficients diagonaux. Pour s'en convaincre, appliquer la propriété de dilatation à chaque ligne afin de se ramener au déterminant de l'identité qui vaut 1 par hypothèse.

Par exemple pour une matrice 3×3

$$\det \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix} = a \det \begin{pmatrix} 1 & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix} = ab \det \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & c \end{pmatrix}$$
$$= abc \det \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = abc$$

2. La propriété de permutation implique que le determinant d'une matrice avec deux lignes identiques est nul. En effet, on échangeant ces deux lignes, d'une part, la matrice ne change pas et donc le déterminant non plus, et d'autre part, le déterminant est multiplié par -1.

3. La propriété de dilation porte sur les lignes et non par sur la matrice elle-même. En fait, en appliquant n fois la propriété de dilatation, il vient que

$$\det(\lambda A) = \lambda^n \det A, \qquad \forall \lambda \in \mathbf{R},\ \forall A \in \mathcal{M}_n$$

car la matrice λA s'obtient en multipliant chaque ligne de A par λ.

4. La propriété d'additivité porte sur les lignes et non pas les matrices. Pour la plupart des matrices, $\det(A + B) \neq \det A + \det B$. C'est par exemple le cas pour $A = B = \mathrm{Id}$ dès que $n > 1$. En effet, $\mathrm{Id} + \mathrm{Id} = 2\,\mathrm{Id}$ entraine par la remarque précédente que $\det(\mathrm{Id} + \mathrm{Id}) = 2^n \det(\mathrm{Id}) = 2^n$ tandis que $\det(\mathrm{Id}) + \det(\mathrm{Id}) = 2$.

5. la propriété de multiplicativité vaut aussi pour $\lambda = 0$. Cela implique que le déterminant d'une matrice avec une ligne nulle est zéro.

1.3.3 *Réduction de Gauss*

La première façon de calculer le déterminant est par l'algorithme de Gauss. Rappelons que cet algorithme est basé sur trois opérations élémentaires sur les lignes qui sont les dilatations, les permutations et les transvections. Nous savons déjà comment se modifie le déterminant lorsqu'on effectue une dilatation ou une permutation. Il reste à comprendre les transvections.

Énoncé indispensable 23 :

PROPOSITION 1.36. *Le déterminant ne change pas lorsqu'on additionne à une ligne un multiple d'une autre ligne.*

Démonstration. Vérifions le avec la première et la seconde ligne. Le raisonnement est le même en général. Nous avons en appliquant successivement la propriété d'additivité

puis de dilatation

$$\det\begin{pmatrix} L_1+\lambda L_2 \\ L_2 \\ \vdots \\ L_n \end{pmatrix} = \det\begin{pmatrix} L_1 \\ L_2 \\ \vdots \\ L_n \end{pmatrix} + \det\begin{pmatrix} \lambda L_2 \\ L_2 \\ \vdots \\ L_n \end{pmatrix} = \det\begin{pmatrix} L_1 \\ L_2 \\ \vdots \\ L_n \end{pmatrix} + \lambda \det\begin{pmatrix} L_2 \\ L_2 \\ \vdots \\ L_n \end{pmatrix} = \det\begin{pmatrix} L_1 \\ L_2 \\ \vdots \\ L_n \end{pmatrix}$$

car comme nous l'avons déjà remarqué, le déterminant d'une matrice avec deux lignes identiques est nul. □

Énoncé indispensable 24 :

théorème 1.37. *Le déterminant d'une matrice est non nul si et seulement si la matrice est inversible*

Démonstration. La remarque importante est que si l'on passe de A à B par une opération élémentaire sur les lignes, alors $\det A = 0$ si et seulement si $\det B = 0$. En effet, $\det B = -\det A$ pour une permutation, $\det B = \det A$ pour une transvection et $\det B = \lambda \det A$ pour une dilatation de rapport λ et rappelons que seules les dilatations avec un facteur non-nul font partie des opérations élémentaires.

Donc d'après le théorème 1.21, il suffit de vérifier le résultat pour une matrice échelonnée réduite E. Ici la matrice est carrée de taille n et par le théorème 1.30, E est inversible lorsqu'elle a n pivots. Dans ce cas, $E = \mathrm{Id}$, et donc $\det(E) = 1 \neq 0$. Si par contre E a moins de n pivots, une de ses lignes sera nulle. En multipliant cette ligne par 0, on ne modifie pas la matrice, ce qui implique que le déterminant est nul. □

Énoncé indispensable 25 :

proposition 1.38. *Le déterminant d'un matrice triangulaire est le produit de ses coefficients diagonaux*

Démonstration. Si l'un des coefficients diagonaux est nul, nous savons par le théorème 1.33 que la matrice n'est pas inversible et donc son déterminant est nul par le théorème 1.37. Si aucun coefficient diagonal n'est nul, par une dilatation sur chaque ligne par le facteur qui convient, on se ramène à une matrice triangulaire dont tous les coefficients de la diagonale valent 1. Il faut alors montrer que le déterminant de cette matrice est 1. Pour cela on applique la deuxième phase de l'algorithme de Gauss pour se ramener à la matrice identité. Les opérations à effectuer étant uniquement des transvections, le determinant ne change pas. □

Nous déduisons de ce qui précède l'algorithme suivant de calcul du déterminant : par la première phase de l'algorithme de Gauss, on se ramène à une matrice triangulaire. On effectue alors le produit des coefficients diagonaux.

EXEMPLE 1.39.

$$\det\begin{pmatrix}1 & -4 & 2\\ -1 & 4 & -7\\ -1 & 1 & -4\end{pmatrix} = \det\begin{pmatrix}1 & -4 & 2\\ 0 & 0 & -5\\ 0 & -3 & -2\end{pmatrix}$$
$$= -\det\begin{pmatrix}1 & -4 & 2\\ 0 & -3 & -2\\ 0 & 0 & -5\end{pmatrix} = -1(-3)(-5) = -15$$

Énoncé indispensable 26 :

THÉORÈME 1.40.

1. *Si A et B sont deux matrices de $\mathcal{M}_n$,* $\det(AB) = (\det A)(\det B)$.
2. *Si A est inversible,* $\det(A^{-1}) = (\det A)^{-1}$.

Démonstration. Remarquons tout d'abord que $\det B = 0$ entraine $\det(AB) = 0$. Si $\det B = 0$, B n'est pas inversible, donc il existe un vecteur colonne X non nul tel que $BX = 0$, donc pour ce même vecteur $(AB)X = A(BX) = 0$, donc AB n'est pas inversible. Nous avons utilisé le théorème 1.37 et la proposition 1.29.

Supposons maintenant $\det B \neq 0$ et montrons que $\det(AB) = (\det A)(\det B)$. Pour cela il suffit de montrer que la fonction

$$f : \mathcal{M}_n \to \mathbf{R}, \qquad f(A) = \frac{\det(AB)}{\det B}$$

vérifie les propriétés du théorème 1.34 qui caractérisent le déterminant. Permuter deux lignes de A puis multiplier par B revient à permuter les deux lignes correspondantes du produit AB, donc $\det(AB)$ est multiplié par -1 lorsque l'on permute deux lignes de A, et il en est de même pour $f(A)$. L'argument est le même pour les dilatations et l'additivité : à chaque fois, il revient au même d'effectuer l'opération sur A puis de multiplier à droite par B que d'effectuer directement l'opération sur le produit AB. Enfin, $\mathrm{Id}\, B = B$ entraine que $f(\mathrm{Id}) = \det B / \det B = 1$.

Pour montrer la deuxième partie, il suffit d'appliquer $\det(AB) = (\det A)(\det B)$ à $B = A^{-1}$ □

1.3.4 *Transposée*

Si $A = (a_{ij})$ est une matrice de taille $p \times q$, sa transposée est la matrice $B = (b_{ij})$ de taille $q \times p$ de coefficients $b_{ij} = a_{ji}$. On la note A^t. On remarque que les lignes/colonnes de A correspondent aux colonnes/lignes de A^t.

LEMME 1.41. *Si A est une matrice de taille $p \times q$ et B une matrice de taille $q \times r$, alors $(AB)^t = B^t A^t$.*

Démonstration. On demontre d'abord le résultat pour $p = r = 1$, c'est-à-dire A une

matrice ligne et B une matrice colonne

$$AB = \begin{pmatrix} a_1 & \dots & a_q \end{pmatrix} \begin{pmatrix} b_1 \\ \vdots \\ b_q \end{pmatrix} = a_1 b_1 + \dots + a_q b_q$$

$$= b_1 a_1 + \dots + b_q a_q = \begin{pmatrix} b_1 & \dots & b_q \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_q \end{pmatrix} = B^t A^t$$

On en déduit le cas général : $(AB)_{ij}$ = produit de la i-ième ligne de A par la j-ième colonne de B = produit de la j-ième ligne de B^t par la i-ième colonne de A^t (d'après ce qui précède) $= (B^t A^t)_{ji}$. □

Énoncé indispensable 27 :

THÉORÈME 1.42. *Si A est une matrice carrée, alors* $\det(A^t) = \det A$.

Démonstration. Notons f la fonction $\mathcal{M}_n \to \mathbf{R}$ définie par $f(A) = \det(A^t)$. Bien entendu $f(\mathrm{Id}) = 1$. Nous allons montrer que $f(A)$ est multiplié par -1 lorsque échange deux lignes de A, multiplié par λ lorsqu'on multiplie une ligne par un facteur λ, et ne change pas lorsqu'on additione à un ligne un multiple d'une autre ligne. Nous pourrons alors conclure que $f(A) = \det A$ car ces quatre propriétés caractérisent la fonction déterminant comme cela a été expliqué dans le chapitre précédent.

Pour montrer ces propriétés, rappelons que chacune de ces transformations revient à multiplier A à gauche par la matrice qui convient, cf. preuve du théorème 1.30. Par exemple, si O est la matrice diagonale avec tous ses coefficients diagonaux égaux à 1 sauf le i-ième qui vaut λ, alors OA est la matrice obtenue en multipliant la i-ième ligne de A par λ. En appliquant successivement le lemme 1.41 et le théorème 1.40, il vient $f(OA) = \det(A^t O^t) = \det(A^t)\det(O^t) = \lambda f(A)$ car le déterminant de O^t est λ. On raisonne de la même façon pour les transvections et permutations. □

La conséquence de ce résultat est que toutes les opérations que nous faisions sur les lignes peuvent être maintenant faites sur les colonnes, et elles auront le même effet sur le déterminant.

Énoncé indispensable 28 :

PROPOSITION 1.43.

1. *Lorsque l'on échange deux colonnes, le déterminant est multiplié par -1.*
2. *Lorsque l'on multiplie une colonne par un coefficient λ, le déterminant est multiplié par λ.*
3. *Lorsque l'on ajoute à une colonne un multiple d'une autre colonne, le déterminant ne change pas.*

4. *Si A, B, C sont trois matrices carrées de taille n et qu'il existe $i \in \{1, \dots, n\}$ pour lequel la i-ième colonne de A est la somme des i-ièmes colonnes de B et C tandis que pour tout $j \neq i$, les j-ième colonnes de A, B, C sont les mêmes, alors* $\det A = \det B + \det C$.

1.3.5 *Développement*

Nous allons donner une nouvelle façon de calculer le déterminant par récurrence sur la dimension. Soit A une matrice de taille $n \times n$ avec $n \geqslant 2$. Pour tout $i, j \in \{1, \dots, n\}$, notons M_{ij} la matrice obtenue en enlevant de A sa i-ième ligne et j-ième colonne. Ces matrices M_{ij} sont de taille $(n-1) \times (n-1)$. Les formules à venir donnent le déterminant de A en fonction des déterminants des M_{ij}. Il y a une formule pour chaque ligne de A et une formule pour chaque colonne de A.

Si A est une matrice 4×4, alors la formule pour la deuxième colonne est

$$\det A = -a_{12} \det M_{12} + a_{22} \det M_{22} - a_{32} \det M_{32} + a_{42} \det M_{42}.$$

On dit que l'on a *développé* A par rapport à la deuxième colonne. Expliquons le membre de droite : les coefficients a_{12}, a_{22}, a_{32} et a_{42} sont ceux de la deuxième colonne, les matrices M_{12}, M_{22}, M_{32} et M_{42} s'obtiennent à partir de A en enlevant la deuxième colonne et la première/deuxième/troisième/quatrième ligne. On remarque que le signe alterne, et que le premier signe est négatif. On retrouve ceci sur la matrice suivante

$$\begin{pmatrix} + & - & + & - \\ - & + & - & + \\ + & - & + & - \\ - & + & - & + \end{pmatrix}$$

où les signes dans la deuxième colonne sont $-$, $+$, $-$, $+$.

Toujours pour une matrice 4×4, la formule de développement par rapport à la troisième ligne est

$$\det A = a_{31} \det M_{31} - a_{32} \det M_{32} + a_{33} \det M_{33} - a_{34} \det M_{34}$$

avec ici les coefficients a_{31}, a_{32}, a_{33} et a_{34} de la troisième ligne et les matrices M_{31}, M_{32}, M_{33} et M_{34} obtenues en enlevant de A sa troisième ligne et sa première, deuxième, troisième ou quatrième colonne.

La proposition suivante donne la formule générale.

Énoncé indispensable 29 :

PROPOSITION 1.44. *Nous avons*

$$\text{pour tout } i \in \{1,\dots,n\}, \qquad \det A = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det M_{ij}$$

$$\text{pour tout } j \in \{1,\dots,n\}, \qquad \det A = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det M_{ij}$$

Donnons une idée de la preuve. Soit A une matrice carrée de taille n et A' une matrice carrée de taille $n+1$ ayant la forme

$$A' = \begin{pmatrix} 1 & * & \dots & * \\ 0 & & & \\ \vdots & & A & \\ 0 & & & \end{pmatrix}.$$

Ici les coefficients de la première colonne sont tous nuls sauf le premier qui vaut 1, les coefficients de la première ligne sont quelconques sauf le premier qui vaut 1. Les autres coefficients sont ceux de la matrice A.

LEMME 1.45. *A et A' ont même déterminant.*

Démonstration. On note $f(A) = \det A'$ et l'on montre que f vérifient les quatre propriétés qui caractérisent le déterminant, cf. théorème 1.34. Ici nous avons choisi une fois pour toutes les coefficients de la première ligne de A'. Ces quatre propriétés découlent immédiatement des quatres propriétés pour le déterminant des matrices de taille $n+1$. □

Expliquons comment en déduire la formule de développement pour une matrice 3×3 par rapport à la première colonne. Le lemme 1.45 nous donne

$$\begin{vmatrix} 1 & * & * \\ 0 & a & b \\ 0 & c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc \tag{1.7}$$

et l'on procède au calcul suivant :

$$\begin{aligned}
\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} &= \begin{vmatrix} a_1 & b_1 & c_1 \\ 0 & b_2 & c_2 \\ 0 & b_3 & c_3 \end{vmatrix} + \begin{vmatrix} 0 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ 0 & b_3 & c_3 \end{vmatrix} + \begin{vmatrix} 0 & b_1 & c_1 \\ 0 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} \\
&= a_1 \begin{vmatrix} 1 & b_1 & c_1 \\ 0 & b_2 & c_2 \\ 0 & b_3 & c_3 \end{vmatrix} + a_2 \begin{vmatrix} 0 & b_1 & c_1 \\ 1 & b_2 & c_2 \\ 0 & b_3 & c_3 \end{vmatrix} + a_3 \begin{vmatrix} 0 & b_1 & c_1 \\ 0 & b_2 & c_2 \\ 1 & b_3 & c_3 \end{vmatrix} \\
&= a_1 \begin{vmatrix} 1 & b_1 & c_1 \\ 0 & b_2 & c_2 \\ 0 & b_3 & c_3 \end{vmatrix} - a_2 \begin{vmatrix} 1 & b_2 & c_2 \\ 0 & b_1 & c_1 \\ 0 & b_3 & c_3 \end{vmatrix} + a_3 \begin{vmatrix} 1 & b_3 & c_3 \\ 0 & b_1 & c_1 \\ 0 & b_2 & c_2 \end{vmatrix} \\
&= a_1 \begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix} - a_2 \begin{vmatrix} b_1 & c_1 \\ b_3 & c_3 \end{vmatrix} + a_3 \begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix}
\end{aligned}$$

Nous avons utilisé pour commencer la propriété d'additivité et la propriété de dilatation sur la première colonne afin d'y faire apparaitre uniquement des 0 et un seul 1. Ensuite en permutant les lignes de chaque matrice, nous avons déplacé ce 1 dans la première ligne tout en gardant l'ordre des deux autres lignes. Et enfin nous avons utilisé la formule (1.7).

1.3.6 *Formules de Cramer*

Rappelons qu'un système d'équations linéaires de matrice inversible a une unique solution quel que soit le second membre. Nous savons déjà résoudre ces systèmes par l'algorithme de Gauss. Nous allons ici exprimer chaque inconnue comme le quotient de deux déterminants. A la différence de la méthode de Gauss, cela nous donnera une formule close pour la solution.

théorème 1.46. *Si A est inversible, l'unique solution de $Ax = b$ est donnée par*

$$x_1 = \frac{\det B_1}{\det A}, \quad x_2 = \frac{\det B_2}{\det A}, \quad \dots, \quad x_n = \frac{\det B_n}{\det A}$$

où pour tout i, B_i est la matrice obtenue à partir de A en remplaçant sa i-ieme colonne par le second membre b.

Démonstration. Donnons la preuve pour A une matrice 3×3. Le calcul de x_1 est basé sur l'égalité :

$$A \begin{pmatrix} x_1 & 0 & 0 \\ x_2 & 1 & 0 \\ x_3 & 0 & 1 \end{pmatrix} = \begin{pmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{pmatrix} \tag{1.8}$$

La deuxième matrice dans le produit s'obtient en remplaçant dans la matrice identité la première colonne par le vecteur x solution. La matrice du membre de droite est la

matrice B_1 introduite dans l'énoncé. L'équation (1.8) équivaut aux trois égalités

$$A\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}=\begin{pmatrix}b_1\\b_2\\b_3\end{pmatrix},\qquad A\begin{pmatrix}0\\1\\0\end{pmatrix}=\begin{pmatrix}a_{12}\\a_{22}\\a_{32}\end{pmatrix},\qquad A\begin{pmatrix}0\\0\\1\end{pmatrix}=\begin{pmatrix}a_{13}\\a_{23}\\a_{33}\end{pmatrix}$$

qui sont toutes vraies, la première étant le système $Ax = b$. Maintenant pour obtenir x_1, il suffit de prendre le déterminant dans (1.8). La deuxième matrice du produit a pour déterminant x_1 (pourquoi ?) et l'on obtient

$$(\det A)(x_1) = \det B_1$$

d'où $x_1 = (\det B_1)/(\det A)$. De même, pour calculer x_2 et x_3, on prend le déterminant dans les égalités :

$$A\begin{pmatrix}1&x_1&0\\0&x_2&0\\0&x_3&1\end{pmatrix}=\begin{pmatrix}a_{11}&b_1&a_{13}\\a_{21}&b_2&a_{23}\\a_{31}&b_3&a_{33}\end{pmatrix},\qquad A\begin{pmatrix}1&0&x_1\\0&1&x_2\\0&0&x_3\end{pmatrix}=\begin{pmatrix}a_{11}&a_{12}&b_1\\a_{21}&a_{22}&b_2\\a_{31}&a_{32}&b_3\end{pmatrix}$$

ce qui donne $(\det A)(x_2) = \det B_2$ et $(\det A)(x_3) = \det B_3$. □

Comme nous l'avions vu, la résolution d'un système inversible est essentiellement équivalente au calcul de l'inverse de la matrice du système. En effet, si A est inversible, la solution de $Ax = b$ est $b = A^{-1}x$. Réciproquement on trouve la i-ième colonne de A^{-1} en résolvant $AX = e_i$ avec e_i le i-ième vecteur de la base canonique. Donc notre dernier résultat nous calcule les coefficients de la matrice inverse comme des rapports de déterminants. Expliquons cela.

definition 1.47. Soit A une matrice carrée de taille n. Les cofacteurs de A sont les nombres

$$c_{ij} = (-1)^{i+j}\det M_{ij},\qquad i,j\in\{1,\dots,n\}$$

où M_{ij} est comme dans le chapitre 1.3.5 la matrice obtenue à partir de A en enlevant sa i-ième ligne et j-ième colonne. La matrice (c_{ij}) s'appelle la matrice des cofacteurs.

théorème 1.48. *Soit A une matrice inversible. Alors sa matrice inverse est*

$$A^{-1} = \frac{1}{\det A}C^t$$

où C est la matrice des cofacteurs.

Démonstration. Comme nous l'avons dit, cela peut se déduire du théorème 1.46. Mais nous allons donner une preuve directe. Si A est une matrice 3×3, nous avons

$$AC^t=\begin{pmatrix}a_{11}&a_{12}&a_{13}\\a_{21}&a_{22}&a_{23}\\a_{31}&a_{32}&a_{33}\end{pmatrix}\begin{pmatrix}\det M_{11}&-\det M_{21}&\det M_{31}\\-\det M_{12}&\det M_{22}&-\det M_{32}\\\det M_{13}&-\det M_{23}&\det M_{33}\end{pmatrix}$$

Le produit de la première ligne de A par la première colonne de C^t est le développement

du déterminant de A par rapport à la première ligne

$$a_{11} \det M_{11} - a_{12} \det M_{12} + a_{13} \det M_{13} = \det A.$$

De même, en développant $\det A$ par rapport à la deuxième et troisième lignes, nous obtenons que les deux autres coefficients diagonaux de AC^t sont égaux à $\det A$. Nous affirmons que tous les autres coefficients sont nuls. Par exemple, le coefficient de la première ligne, deuxième colonne de AC^t est le développement par rapport à la première ligne d'une matrice B

$$-a_{11} \det M_{21} + a_{12} \det M_{22} - a_{13} \det M_{23} = \det B$$

qui s'obtient à partir de A en remplaçant sa deuxième ligne par la première ligne. B ayant deux lignes égales, son déterminant est nul. Pour résumer nous avons :

$$AC^t = \begin{pmatrix} \det A & 0 & 0 \\ 0 & \det A & 0 \\ 0 & 0 & \det A \end{pmatrix} = (\det A)\,\mathrm{Id}$$

En raisonnant de la même façon avec des développements sur les colonnes, nous montrons que $C^t A = (\det A)\,\mathrm{Id}$, ce qui termine la preuve. Nous aurions pu aussi conclure directement par le théorème 1.31. □

1.3.7 *La formule du déterminant*

Ce chapitre est un complément. Nous allons donner l'idée de la définition du déterminant, c'est-à-dire que nous allons donner une expression pour $\det A$ qui vérifie les propriétés du théorème 1.34. Cette expression généralise la formule

$$\det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

aux matrices $n \times n$.

Rappelons qu'une *bijection* d'un ensemble F est une application de F dans F qui a la propriété que tout élément de F a un unique antécédent. Les bijections d'un ensemble fini s'appellent aussi *permutations*. Par exemple l'ensemble $\{1, 2\}$ a exactement deux permutations σ et σ' données par

$$\sigma(1) = 1, \quad \sigma(2) = 2; \qquad \sigma'(1) = 2, \quad \sigma'(2) = 1$$

L'ensemble $\{1, 2, 3\}$ a exactement six permutations

$$\begin{array}{lllllllll} \sigma_1(1) = 1, & \sigma_1(2) = 2, & \sigma_1(3) = 3; & & \sigma_2(1) = 1, & \sigma_2(2) = 3, & \sigma_2(3) = 2; \\ \sigma_3(1) = 2, & \sigma_3(2) = 1, & \sigma_3(3) = 3; & & \sigma_4(1) = 2, & \sigma_4(2) = 3, & \sigma_4(3) = 1; \\ \sigma_5(1) = 3, & \sigma_5(2) = 1, & \sigma_5(3) = 2; & & \sigma_6(1) = 3, & \sigma_6(2) = 2, & \sigma_6(3) = 1. \end{array}$$

Plus généralement, l'ensemble $\{1, \ldots, n\}$ a $n! = n(n-1)(n-2)\ldots 1$ permutations. Notons $\mathfrak{S}_n$ l'ensemble formé par ces permutations. En particulier, $\mathfrak{S}_2 = \{\sigma, \sigma'\}$ et $\mathfrak{S}_3 = \{\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6\}$.

Chacune de ces permutations a un signe, que l'on appelle la signature et que l'on définit de la façon suivante. On appelle *inversion* de $\sigma \in \mathfrak{S}_n$, un couple $i, j \in \{1, \dots, n\}$ tel que $i < j$ et $\sigma(i) > \sigma(j)$. On note $N(\sigma)$ le nombre d'inversions de σ. Par exemple, pour les permutations de $\{1, 2\}$

$$N(\sigma) = 0, \quad N(\sigma') = 1$$

et pour celles de $\{1, 2, 3\}$,

$$N(\sigma_1) = 0, \quad N(\sigma_2) = N(\sigma_3) = 1, \quad N(\sigma_4) = N(\sigma_5) = 2, \quad N(\sigma_6) = 3.$$

La signature de σ est la parité de $N(\sigma)$. Si $N(\sigma)$ est pair, la signature est $+1$, si $N(\sigma)$ est impair, la signature est -1. En général la signature vaut $(-1)^{N(\sigma)}$. On remarque que pour $\mathfrak{S}_2$ et $\mathfrak{S}_3$, la moitié des permutations sont paires, l'autre moitié impaires. Ceci est encore vrai pour $\mathfrak{S}_n$ avec $n \geqslant 4$.

théorème 1.49. *L'application déterminant de $\mathcal{M}_n$ dans $\mathbf{R}$ définie par*

$$\det A = \sum_{\sigma \in \mathfrak{S}_n} (-1)^{N(\sigma)} a_{1\sigma(1)} a_{2\sigma(2)} \dots a_{n\sigma(n)}, \qquad \forall A = (a_{ij}) \in \mathcal{M}_n$$

vérifie les quatre propriétés énoncées dans le théorème 1.34.

Les propriétés de dilatations, additivité se vérifient (relativement) facilement. Par contre la propriété de permutation est plus délicate, elle demande de bien comprendre la signature. Cela sera expliqué dans le cours de L2.

Chapitre 2

Sous-espaces vectoriels, applications linéaires

Pour $n \in \mathbb{N}^*$, on note $\mathbb{R}^n$ l'ensemble des n-uplets de réels, autrement dit

$$\mathbb{R}^n = \{(x_1, x_2, \ldots, x_{n-1}, x_n),\ x_1 \in \mathbb{R}, \ldots, x_n \in \mathbb{R}\}\ .$$

L'objectif de cette partie du cours est de généraliser pour tout $n \in \mathbb{N}^*$ des notions vues au semestre précédent dans les cas particulier du plan $\mathbb{R}^2$ et de l'espace $\mathbb{R}^3$.

On appelle un élément de $\mathbb{R}^n$ un vecteur de $\mathbb{R}^n$. Comme dans les cas du plan et de l'espace, on peut additionner des vecteurs et les multiplier par des réels (appelés des scalaires) de la façon suivante : si $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)$ sont des vecteurs de $\mathbb{R}^n$ et $\lambda \in \mathbb{R}$, alors

$$\begin{aligned} x + y &= (x_1 + y_1, \ldots, x_n + y_n) \\ \lambda x &= (\lambda x_1, \ldots, \lambda x_n)\,. \end{aligned}$$

On appelle vecteur nul de $\mathbb{R}^n$ et on note 0 le vecteur $(0, \ldots, 0)$.

2.1 Bases de $\mathbb{R}^n$

Pour $p \in \mathbb{N}^*$, considérons p vecteurs v_1, v_2, …, v_p de $\mathbb{R}^n$. On appelle le p-uplet $(v_1, \ldots, v_p)$ une *famille* de vecteurs. Nous allons définir plusieurs types de familles : les familles libres, les familles génératrices et enfin les bases.

Énoncé indispensable 30 :

DEFINITION 2.1. La famille $(v_1, \ldots, v_p)$ est dite libre si pour tout $\lambda_1, \ldots, \lambda_p \in \mathbb{R}$,

$$\lambda_1 v_1 + \cdots + \lambda_p v_p = 0 \qquad \Rightarrow \qquad \lambda_1 = \lambda_2 = \cdots = \lambda_p = 0\,.$$

De manière équivalente, on dit que les vecteurs $v_1, \ldots, v_p$ sont linéairement indépendants. Une famille de vecteurs qui n'est pas libre est dite liée.

Voici quelques exemples sous forme d'exercices pour se familiariser avec cette nouvelle notion.

exemple 2.2.

1. Une famille contenant un seul vecteur est libre si et seulement si ce vecteur est non nul.
2. Une famille de deux vecteurs est liée si et seulement si ces deux vecteurs sont colinéaires. A vous de le montrer.

   On rappelle que u et v sont colinéaires si il existe $\lambda \in \mathbf{R}$ tel que $u = \lambda v$ ou $v = \lambda u$. Il est nécessaire deux prendre en compte les deux possibilités car si $u \neq 0$ et $v = 0$, alors $u \neq \lambda v$ pour tout λ et par contre $v = 0u$.
3. Une famille qui contient le vecteur nul est liée.
4. En rajoutant des vecteurs à une famille liée, on obtient une autre famille liée. En enlevant des vecteurs à une famille libre, on obtient une autre famille libre.
5. Peut-on trouver une famille libre de trois vecteurs dans $\mathbf{R}^2$? et dans $\mathbf{R}^3$?

Énoncé indispensable 31 :

definition 2.3.

1. On dit d'un vecteur $u \in \mathbf{R}^n$ qu'il est combinaison linéaire de $v_1, \ldots, v_p$ si il existe des réels $\lambda_1, \ldots, \lambda_p$ tels que

   $$u = \lambda_1 v_1 + \ldots \lambda_p v_p.$$

   Les nombres λ_i s'appellent les coefficients de la combinaison linéaire.
2. La famille $v_1, \ldots v_p$ est dite génératrice si tout vecteur de $\mathbf{R}^n$ est combinaison linéaire des v_i.

Voici quelques exemples/exercices. On pourra comparer avec l'exemple 2.2.

exemple 2.4.

1. La famille $(1,0), (0,1)$ est une famille génératrice de $\mathbf{R}^2$. La famille $(1,0,0)$, $(0,1,0)$, $(0,0,1)$ est une famille génératrice de $\mathbf{R}^3$.
2. si on rajoute des vecteurs à une famille génératrice, on obtient une autre famille génératrice. Par conséquent, il existe des familles génératrices avec un nombre arbitrairement grand de vecteurs.
3. Peut-on trouver une famille génératrice avec un seul vecteur dans $\mathbf{R}^2$? une famille génératrice avec deux vecteurs dans $\mathbf{R}^3$?

Pour $i = 1, \ldots, n$, notons e_i le vecteur de $\mathbf{R}^n$ dont toutes les coordonnées sont nulles excepté la i-ième qui vaut 1. Soit un vecteur u de coordonnées $(x_1, \ldots, x_n)$. Remarquer que

$$u = x_1 e_1 + x_2 e_2 + \ldots + x_n e_n.$$

Par conséquent, la famille $(e_1, \ldots, e_n)$ est génératrice. Remarquons que cette famille est aussi libre. En effet, d'après l'égalité ci-dessus, la seule combinaison linéaire nulle des e_i est celle dont tous les coefficients sont nuls.

Énoncé indispensable 32 :

DEFINITION 2.5.

1. Une base est une famille libre et génératrice.
2. la famille $(e_1, \ldots, e_n)$ s'appelle la base canonique

La propriété essentielle des bases qui justifie leur introduction est la suivante.

Énoncé indispensable 33 :

PROPOSITION 2.6. *Si $(v_1, \ldots, v_p)$ est une base de $\mathbb{R}^n$, alors pour tout $u \in \mathbb{R}^n$, il existe un unique p-uplets de réels $(\lambda_1, \ldots, \lambda_p)$ tels que*

$$u = \lambda_1 v_1 + \cdots + \lambda_p v_p \,.$$

On appelle $\lambda_1, \ldots, \lambda_p$ les coordonnées de u dans la base $(v_1, \ldots, v_p)$.

Démonstration. $(v_1, \ldots, v_p)$ étant une famille génératrice, tout vecteur $u \in \mathbb{R}^n$ est une combinaison linéaire de vecteurs des v_i, donc il existe $(\lambda_1, \ldots, \lambda_p)$ tels que $u = \lambda_1 v_1 + \cdots + \lambda_p v_p$. Nous devons montrer que les coefficients λ_i sont uniques. Supposons que $u = \mu_1 v_1 + \cdots + \mu_p v_p$ avec des réels $\mu_1, \ldots, \mu_p$. Alors

$$\begin{aligned}(\lambda_1 - \mu_1)v_1 + \cdots + (\lambda_p - \mu_p)v_p \;&= \\ (\lambda_1 v_1 + \cdots + \lambda_p v_p) - (\mu_1 v_1 + \cdots + \mu_p v_p) \;&=\; u - u \;=\; 0\,.\end{aligned}$$

Or, $(v_1, \ldots, v_p)$ étant une base, c'est une famille libre et la seule combinaison linéaire nulle des v_i est celle dont tous les coefficients sont nuls, ce qui veut dire que $\lambda_1 - \mu_1 = \cdots = \lambda_p - \mu_p = 0$. Autrement dit, pour $1 \leq i \leq p$, $\lambda_i = \mu_i$. □

Ces questions d'existence et d'unicité rappellent étrangement la résolution des systèmes. Ce n'est pas un hasard. Représentons-nous les vecteurs de $\mathbf{R}^n$ comme des matrices colonnes à n coefficients. On parle de vecteurs colonnes. Alors nous pouvons exprimer les combinaisons linéaires comme des produit matriciels. Ce point de vue sera capital par la suite.

Énoncé indispensable 34 :

Soit A la matrice de taille $n \times p$ dont les colonnes sont les vecteurs $v_1, \ldots, v_p$.

Alors pour tout vecteur u

$$u = \lambda_1 v_1 + \ldots + \lambda_p v_p \quad \Leftrightarrow \quad u = A \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_p \end{pmatrix}. \tag{2.1}$$

Nous allons ainsi lire les propriétés de la famille sur la matrice. Rappelons que nous avons défini le rang de A comme étant le nombre de pivots d'une matrice échelonnée obtenue à partir de A par une suite d'opérations élémentaires sur les lignes.

Énoncé indispensable 35 :

THÉORÈME 2.7. *Si A est la matrice de vecteurs colonnes $v_1, \ldots, v_p$, alors*

1. *$(v_1, \ldots, v_p)$ est libre si et seulement si A est de rang p*
2. *$(v_1, \ldots, v_p)$ est génératrice si et seulement si A est de rang n.*
3. *$(v_1, \ldots, v_p)$ est une base si et seulement si $p = n$ et A est inversible.*

Démonstration. $(v_1, \ldots, v_p)$ est libre ssi la seule combinaison linéaire nulle des v_i a des coefficients tous nuls, autrement dit d'après la réécriture (2.1), ssi la seule solution de $AX = 0$ est $X = 0$. D'après la proposition 1.29, cela équivaut à ce que le rang de A soit égal au nombre de colonne de A, qui ici vaut p.

$(v_1, \ldots, v_p)$ est génératrice ssi tout vecteur est combinaison linéaire des v_i, autrement dit d'après (2.1), ssi le système $AX = B$ a une solution pour tout second membre B. D'après la proposition 1.29, cela équivaut à ce que le rang de A soit égal au nombre de lignes de A, qui ici vaut n.

D'après les assertions précédentes, pour que $(v_1, \ldots, v_p)$ soit une base, il faut et il suffit que $p = \operatorname{rang} A = n$, ce qui d'après le théorème 1.30 équivaut au fait que A soit carrée et inversible. □

Cela nous donne une méthode pratique pour déterminer si une famille est libre ou génératrice. Cela a aussi d'importantes conséquences théoriques. Tout d'abord, remarquons la conséquence suivante du théorème 2.7.

Énoncé indispensable 36 :

COROLLAIRE 2.8. *Toute base de $\mathbf{R}^n$ a exactement n vecteurs.*

Remarquons aussi que calculer les coordonnées dans une base revient à résoudre un systéme linéaire. Plus précisément, si $(v_1, \ldots, v_n)$ est une base de $\mathbf{R}^n$ et A la matrice associée, en multipliant par A^{-1} l'égalité matricielle (2.1), on obtient les coordonnées

$(\lambda_1, \ldots, \lambda_n)$ d'un vecteur $u \in \mathbf{R}^n$:

$$\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} = A^{-1}u.$$

Cette formule s'appelle formule du changement de base. Elle exprime les coordonnées dans la base $(v_1, \ldots, v_n)$ en fonction des coordonnées dans la base canonique. Plus généralement, on peut passser d'une base à une autre comme suit.

PROPOSITION 2.9. *Soient $(v_1, \ldots, v_n)$ et $(w_1, \ldots, w_n)$ deux bases de $\mathbb{R}^n$. Soit P la matrice (λ_{ij}) où pour tout $j = 1, \ldots, n$, $w_j = \lambda_{1j}v_1 + \ldots + \lambda_{nj}v_n$. Alors P est inversible et pour tout vecteur u,*

$$\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} = P \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}.$$

où (λ_i) sont les coordonnées de u dans la base (v_i) et (μ_i) ses coordonnées dans la base (w_i).

On appelle P la matrice de passage de la base (w_i) vers la base (v_i), car on passe des coordonnées dans (w_i) à celles dans (v_i) en multipliant par P. Par définition, les coefficients de la j-ième colonne de P sont les coordonnées de w_j dans la base (v_i).

Démonstration. Notons A et B les matrices dont les vecteurs colonnes sont $v_1, \ldots, v_n$ et $w_1, \ldots, w_n$ respectivement. D'après le théorème 2.7, ces matrices sont inversibles et comme nous l'avons vu

$$\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} = A^{-1}u, \qquad \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} = B^{-1}u.$$

Nous avons donc

$$\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} = A^{-1}u = A^{-1}BB^{-1}u = P \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}$$

où l'on pose $P := AB^{-1}$. La matrice P ainsi définie est inversible comme produit de matrices inversibles. De plus, si l'on applique cette formule à $u = w_j$ pour chaque indice j, il vient que les coefficients de la j-ième colonne de P sont les coordonnées de w_j dans la base (u_i). On vient de retrouver la définition de P donnée dans l'énoncé de la proposition. □

Une autre conséquence du théorème 2.7 est le résultat suivant

PROPOSITION 2.10.

1. *Une famille libre a au plus n vecteurs. Si une famille libre a exactement n vecteurs, c'est une base.*
2. *Une famille génératrice a au moins n vecteurs. Si une famille génératrice a exactement n vecteurs, c'est une base.*

Démonstration. Le rang d'une matrice est plus petit que son nombre de lignes, donc par la première partie du théorème 2.7, une famille libre au plus n vecteurs. Lorsqu'une famille libre a n-vecteurs, la matrice correspondante est carrée de taille n et de rang n. Par la seconde partie du théorème 2.7, la famille est aussi génératrice.

La preuve de la deuxième assertion est complètement similaire. □

2.2 Sous-espaces vectoriel de $\mathbf{R}^n$

2.2.1 *Définition et exemples*

Énoncé indispensable 37 :

definition 2.11. Un sous-espace vectoriel de $\mathbf{R}^n$ est un ensemble E de vecteurs de $\mathbf{R}^n$ non-vide qui est stable par somme et multiplication externe, c'est-à-dire que pour tout $u, v \in E$ et $\lambda \in \mathbf{R}$, $u + v \in E$ et $\lambda u \in E$.

exemple 2.12.

1. Dans le plan $\mathbf{R}^2$, toute droite passant par l'origine est un sous-espace vectoriel.
2. Dans l'espace $\mathbf{R}^3$, les droites et les plans contenant l'origine sont des sous-espaces vectoriels.
3. Dans $\mathbf{R}^2$, l'ensemble $\{(x, y);\ x \geqslant 0 \text{ et } y \geqslant 0\}$ n'est pas un sous-espace vectoriel.
4. L'ensemble des matrices 2×2 triangulaires est un sous-espace vectoriel de l'espace $\mathcal{M}_2$. Ici on considère chaque matrice de $\mathcal{M}_2$ comme un vecteur de $\mathbf{R}^4$ par

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathcal{M}_2 \to (a, b, c, d) \in \mathbf{R}^4.$$

proposition 2.13. *Soit E un sous-espace vectoriel de $\mathbf{R}^n$. Alors*

1. *le vecteur nul de $\mathbf{R}^n$ appartient à E.*
2. *Soit $v_1, \ldots, v_p$ une famille de vecteurs de E. Alors toute combinaison linéaire des v_i appartient à E.*

Démonstration. E est non vide, donc il contient un vecteur u. Alors $0_n = 0u$ appartient à E aussi. On montre la deuxième assertion par récurrence sur p. Pour $p = 1$, c'est le fait que E est stable par multiplication par tout réel. Pour l'hérédité, il suffit d'écrire que

$$\lambda_1 v_1 + \ldots + \lambda_p v_p + \lambda_{p+1} v_{p+1} = (\lambda_1 v_1 + \ldots + \lambda_p v_p) + \lambda_{p+1} v_{p+1}$$

et d'utiliser que E est stable par somme. □

exemple 2.14.

1. $\{0\}$ est un sous-espace vectoriel de $\mathbf{R}^n$, on l'appelle parfois le sous-espace réduit à 0. $\mathbf{R}^n$ est un sous-espace vectoriel de $\mathbf{R}^n$. Les sous-espaces vectoriels de $\mathbf{R}^n$ qui sont différents de $\{0\}$ et $\mathbf{R}^n$ sont dits *propres*.

2. Soit k un entier compris entre 0 et n. L'ensemble $E_{k,n}$ formés des vecteurs $(x_1, \ldots, x_n)$ de $\mathbf{R}^n$ tels que $x_{k+1} = x_{k+2} = \ldots = x_n = 0$ est un sous-espace vectoriel de $\mathbf{R}^n$. $E_{k,n}$ est paramétrée par les k premières coordonnées

$$E_{k,n} = \{(x_1, \ldots, x_k, 0, \ldots, 0) \in \mathbf{R}^n;\ (x_1, \ldots, x_k) \in \mathbf{R}^k\}.$$

On a aussi que $E_{k,n} = \mathbf{R}^k \times \{0_{n-k}\}$ où $0_{n-k} = (0, \ldots, 0) \in \mathbf{R}^{n-k}$.

Énoncé indispensable 38 :

PROPOSITION 2.15. *Soit $v_1, \ldots, v_p$ des vecteurs de $\mathbf{R}^p$. Alors l'ensemble des combinaisons linéaires de $(v_1, \ldots, v_p)$ est un sous-espace vectoriel de $\mathbf{R}^n$. On l'appelle le sous-espace engendré par la famille $(v_1, \ldots, v_p)$.*

Démonstration. Nous devons vérifier que les sommes et multiples de combinaisons linéaires des v_i sont des combinaisons linéaires des v_i. C'est immédiat :

$$(\lambda_1 v_1 + \ldots + \lambda_p v_p) + (\mu_1 v_1 + \ldots + \mu_p v_p) = (\lambda_1 + \mu_1)v_1 + \ldots + (\lambda_p + \mu_p)v_p$$

et $\lambda(\lambda_1 v_1 + \ldots + \lambda_p v_p) = (\lambda\lambda_1)v_1 + \ldots + (\lambda\lambda_p)v_p$. □

Dans l'autre sens, étant donné un sous-espace vectoriel E de $\mathbf{R}^n$, on peut se demander s'il existe des vecteurs $v_1, \ldots, v_p$ de E tels que E soit l'espace vectoriel engendré par ces vecteurs. Si c'est le cas, on dit que la famille des v_i *engendre* E. La réponse à cette question sera donné dans le chapitre suivant. Pour commencer, vous pouvez déjà vous la poser pour les exemples vus précédemment.

Une dernière famille très importante d'exemples de sous-espaces vectoriels provient des matrices.

Énoncé indispensable 39 :

Soit une matrice A de taille $p \times q$. Deux sous-espaces vectoriels sont naturellement associés à A :

1. $\ker A := \{X \in \mathbf{R}^q / \ AX = 0\}$, qui est un sous-espace vectoriel de $\mathbf{R}^q$,
2. $\mathrm{Im}\, A := \{AX / \ X \in \mathbf{R}^q\}$, qui est un sous-espace vectoriel de $\mathbf{R}^p$.

que nous appelons le *noyau* et *l'image* de A respectivement.

Le mot noyau se traduit par kernel en anglais ou kern en allemand, ce qui explique la notation ker.

Nous avons déjà rencontré ces espaces dans l'étude des systèmes linéaires. Le premier, $\ker A$, n'est autre que l'espace des solutions du système homogène $AX = 0$. Nous avions vu dans la proposition 1.10 qu'il s'agit d'un sous-espace vectoriel de $\mathbf{R}^q$. Le second, $\mathrm{Im}\, A$, est l'ensemble des second membres B tels que le système $AX = B$ admette une solution. Nous n'avons pas encore vérifié que c'est un sous-espace vectoriel. En fait, d'après (2.1), $\mathrm{Im}\, A$ est l'ensemble des combinaisons linéaires des vecteurs colonnes de

A. C'est donc un sous-espace vectoriel de $\mathbf{R}^p$ par la proposition 2.15. Le résultat étant important, énonçons-le comme une proposition.

Énoncé indispensable 40 :

PROPOSITION 2.16. Im A *est le sous-espace vectoriel de* $\mathbf{R}^p$ *engendré par les vecteurs colonnes de* A.

Grace à ce résultat, nous pouvons dire à quelle condition un vecteur appartient au sous-espace engendré par une famille donnée.

EXEMPLE 2.17. Considérons le sous-espace vectoriel E de $\mathbf{R}^4$ engendré par les vecteurs $(2,-4,2,6)$, $(-1,2-1,-3)$, $(3,-6,1,3)$. A quelle condition un vecteur (a,b,c,d) de $\mathbf{R}^4$ appartient-il à E ? Nous avons déjà répondu à cette question. En effet, $E = \operatorname{Im} A$ où

$$A = \begin{pmatrix} 2 & -1 & 3 \\ -4 & 2 & -6 \\ 2 & -1 & 1 \\ 6 & -3 & 3 \end{pmatrix}$$

et nous avons déterminé dans l'exemple 1.28, les seconds membres B tels que $AX = B$ a une solution. Nous en déduisons que (a,b,c,d) appartient à E si et seulement si $b = -2a$ et $d = 3c$. La méthode est bien entendu générale.

2.2.2 *Bases et dimension*

Dans la suite du chapitre, on se donne un sous-espace vectoriel E de $\mathbf{R}^n$.

Énoncé indispensable 41 :

DEFINITION 2.18. Une base de E est une famille libre de E qui engendre E.

Dans cette définition, la famille est libre au sens des familles de vecteurs de $\mathbf{R}^n$. Et rappelons qu'une famille engendre E si tout vecteur de E est combinaison linéaire des vecteurs de la famille. Comme pour $\mathbf{R}^n$, une base permet de définir les coordonnées d'un vecteur de E.

PROPOSITION 2.19. *Si* $(v_1,\dots,v_p)$ *est une base de* E*, alors pour tout* $u \in E$*, il existe un unique* p*-uplets de réels* $(\lambda_1,\dots,\lambda_p)$ *tels que*

$$u = \lambda_1 v_1 + \cdots + \lambda_p v_p\,.$$

On appelle $\lambda_1,\dots,\lambda_p$ *les coordonnées de* u *dans la base* $(v_1,\dots,v_p)$.

La preuve est exactement la même que celle de la proposition 2.6. Dans certains cas, il est facile de trouver une base.

EXEMPLE 2.20.

1. Si $(v_1, \ldots, v_p)$ est une famille libre de $\mathbf{R}^n$ et E est le sous-espace vectoriel engendré par les v_i, alors $(v_1, \ldots, v_p)$ est bien-entendu une base de E.
2. En particulier, si D est une droite de $\mathbf{R}^2$ ou $\mathbf{R}^3$ passant par l'origine, tout vecteur non-nul de D forme une base de D.
3. soit $(e_1, \ldots, e_n)$ la base canonique de $\mathbf{R}^n$. Alors pour tout $k = 1, \ldots, n$, $\mathbf{R}^k \times \{0_{n-k}\}$ admet pour base $(e_1, e_2, \ldots, e_k)$.

En général, ce n'est pas évident de savoir si un sous-espace vectoriel admet une base. Le résultat important est le suivant.

Énoncé indispensable 42 :

THÉORÈME 2.21. *Supposons que $E \neq \{0\}$. Alors*

1. *E admet une base.*
2. *toutes les bases de E ont le même nombre de vecteurs.*

Le nombre de vecteurs d'une base de E est appelée la dimension *de E.*

L'ensemble $\{0\}$ est bien un sous-espace vectoriel de $\mathbf{R}^n$. Par convention, sa dimension est 0. Avant la preuve, donnons un résultat facile sur la dimension.

Énoncé indispensable 43 :

PROPOSITION 2.22. *La dimension de E est comprise entre 0 et n. Si $\dim E = 0$, alors $E = \{0\}$. Si $\dim E = n$, alors $E = \mathbf{R}^n$.*

Démonstration. Une base $(v_1, \ldots, v_p)$ de E est en particulier une famille libre de $\mathbf{R}^n$, et donc $p \leqslant n$ par la proposition 2.10. Par cette même proposition, si $p = n$, alors $(v_1, \ldots, v_n)$ est une base de $\mathbf{R}^n$. Donc tout vecteur de $\mathbf{R}^n$ est combinaison linaire des v_i et donc appartient à E, autrement dit $\mathbf{R}^n \subset E$, donc $\mathbf{R}^n = E$. □

Preuve du théorème 2.21. Pour montrer l'existence d'une base, nous allons construire une famille libre maximale. La remarque importante qui fait marcher la preuve est la suivante :

> Soit une famille libre $(v_1, \ldots, v_p)$ et un vecteur u qui n'est pas combinaison linéaire des v_i. Alors $(v_1, \ldots, v_p, u)$ est libre.

Montrons ceci. On se donne une combinaison linéaire nulle $\lambda_1 v_1 + \ldots + \lambda_p v_p + \lambda u = 0$ et nous devons montrer que les coefficients λ_i et λ sont tous nuls. Si $\lambda \neq 0$, on peut écrire $u = -\frac{\lambda_1}{\lambda} v_1 - \ldots - \frac{\lambda_p}{\lambda} v_p$ ce qui contredit que u n'est pas combinaison linéaire des v_i, donc $\lambda = 0$. Mais alors $\lambda_1 v_1 + \ldots + \lambda_p v_p = 0$ et comme les v_i forment une famille libre, $\lambda_1 = \ldots = \lambda_p = 0$.

Construisons maintenons une base de E. Comme $E \neq \{0\}$, E a un vecteur non nul v_1. Etant non-nul, ce vecteur forme une famille libre et si v_1 engendre E, alors v_1 est

une base de E. Sinon, il existe un vecteur $v_2 \in E$ qui n'est pas combinaison linéaire de v_1. D'après le fait montré ci-dessus, (v_1, v_2) est libre. Si (v_1, v_2) engendre E, c'est une base de E. Sinon, il existe $v_3 \in E$ qui n'est pas combinaison linéaire de (v_1, v_2) et donc (v_1, v_2, v_3) est libre. Si cette famille engendre E, nous avons fini et sinon, on recommence. Rappelons que toute famille libre de $\mathbf{R}^n$ a au plus n vecteurs, proposition 2.10, donc nous pourrons itérer au plus n fois la construction précédente. Et la seule condition qui nous empêche d'itérer est que les v_i forment une base.

Soient maintenant deux bases $(v_1, \ldots, v_p)$ et $(w_1, \ldots, w_m)$ de E. Montrons par l'absurde que $p = m$. Supposons que $p < m$. Notons V et W les matrices de vecteurs colonnes $v_1, \ldots, v_p$ et $w_1, \ldots, w_m$ respectivement. Comme les v_i engendrent E, nous avons pour tout $j = 1, \ldots, m$,

$$w_j = a_{1j}v_1 + \ldots a_{pj}v_p, \qquad a_{1j}, \ldots, a_{pj} \in \mathbf{R}$$

autrement dit $W = VA$ où A est la matrice (a_{ij}). Nous ne savons rien sur A si ce n'est qu'elle a p lignes et m colonnes. Comme $p \leqslant m$, le rang de A est plus petit que p. Comme $p < m$, d'après la proposition 1.29, le système homogène $AX = 0$ a une solution non-nulle X. Alors $WX = 0$, et nous avons donc une combinaison linéaire nulle des w_i à coefficients non tous nuls, une contradiction avec le fait que les w_i forment une famille libre. □

La méthode utilisée pour la preuve de l'existence d'une base montre cet autre résultat important, dit de la base incomplète.

proposition 2.23. *Toute famille libre de E peut-être complétée en une base.*

Démonstration. On part d'une famille libre $(v_1, \ldots v_p)$ de E. Si elle n'engendre pas E, on peut lui rajouter un vecteur v_{p+1} de sorte que la famille reste libre. Tant que la famille n'engendre pas E, on recommence. On sait que le processus s'arrêtera car une famille libre de $\mathbf{R}^n$ a au plus n vecteurs. □

Considèrons maintenant une base $(v_1, \ldots, v_p)$ de E. En appliquant le résultat précédent à cette famille et $\mathbf{R}^n$ (qui est bien un sous-espace vectoriel de $\mathbf{R}^n$), il nous vient des vecteurs $v_{p+1}, \ldots, v_n$ de $\mathbf{R}^n$ tels que $(v_1, \ldots, v_n)$ soit une base de $\mathbf{R}^n$. Ainsi, pour tout vecteur u de $\mathbf{R}^n$, si ses coordonnées dans la base (v_i) sont $(\lambda_1, \ldots, \lambda_n)$, alors

$$u \in E \quad \Leftrightarrow \quad \lambda_{p+1} = \ldots = \lambda_n = 0.$$

Autrement dit, dans l'espace $\mathbf{R}^n \ni (\lambda_i)$ des coordonnées, le sous-espace E devient $\mathbf{R}^p \times \{0\}$. En ce sens, l'exemple 2.14 est universel.

2.2.3 *Sous-espaces vectoriels et systèmes*

Dans ce chapitre, nous allons expliquer comment l'algorithme de Gauss permet de calculer la dimension et donner une base explicite des deux sous-espaces Im A et ker A associés à une matrice A.

On suppose A de taille $p \times q$. Soit une matrice échelonnée A' obtenue à partir de A à partir par des opérations élémentaires sur les lignes. Soit r le nombre de pivot de A'. Notons $i_1, \ldots, i_r$ les indices des colonnes de A' ayant un pivot, $j_1, \ldots, j_d$ les indices des autres colonnes. Ici, $r + d = q$.

Rappelons que les solutions du système homogène $AX = 0$ sont paramétrées par les variables libres $x_{j_1}, \ldots, x_{j_d}$, c'est-à-dire que pour tout $y \in \mathbf{R}^d$, il existe une unique solution X telle que $x_{j_1} = y_1, \ldots, x_{j_d} = y_d$. On définit alors les vecteurs $u_1, \ldots, u_d$ de $\ker A$ qui correspondent à $y = (1,0,\ldots,0), (0,1,0,\ldots,0), \ldots, (0,\ldots,0,1)$.

Enfin, l'on note $v_1, \ldots, v_q$ les vecteurs colonnes de A.

Énoncé indispensable 44 :

THÉORÈME 2.24.

1. $\ker A$ *est de dimension* d *et admet pour base* $(u_1, \ldots, u_d)$.
2. $\mathrm{Im}\, A$ *est de dimension* r *et admet pour base* $(v_{i_1}, \ldots, v_{i_r})$.

EXEMPLE 2.25.

1. Soit la matrice
$$A = \begin{pmatrix} 2 & -1 & 1 & 2 \\ -8 & 4 & -6 & -4 \\ 4 & -2 & 3 & 2 \end{pmatrix}$$
En appliquant l'algorithme de Gauss, on montre qu'une forme échelonnée de cette matrice est
$$A' = \begin{pmatrix} 2 & -1 & 1 & 2 \\ 0 & 0 & -2 & 4 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$
Nous avons donc 2 pivots, 3 lignes et 4 colonnes. La dimension du noyau est alors $4 - 2 = 2$ et celle de l'image 2. Les pivots sont situés dans la première et troisième colonne. Nous en déduisons qu'une base de l'image de A est formé du premier et troisième vecteur colonne de A, à savoir : $\begin{pmatrix} 2 \\ -8 \\ 4 \end{pmatrix}, \begin{pmatrix} 1 \\ -6 \\ 3 \end{pmatrix}$.
Pour obtenir une base du noyau, nous résolvons le système $A'X = 0$ en choisissant les deux variables libres x_2, x_4 égales à $1,0$ puis $0,1$. Cela nous donne $u_1 = \begin{pmatrix} \frac{1}{2} \\ 1 \\ 0 \\ 0 \end{pmatrix}$ et $u_2 = \begin{pmatrix} -2 \\ 0 \\ 2 \\ 1 \end{pmatrix}$.
2. Soit la matrice $B = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 5 & 7 \\ 1 & 3 & 5 & 8 \end{pmatrix}$. Une réduction de Gauss mène à
$$B' = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$
Cette matrice a 3 pivots, donc le rang de B est 3, et le noyau de B a pour dimension $4 - 3 = 1$. Une base du noyau est donnée par le vecteur solution de $B'X = 0$ avec

la variable libre x_3 égale à 1, à savoir $\begin{pmatrix} 1 \\ -2 \\ 1 \\ 0 \end{pmatrix}$.

Les pivots de B' sont situés dans la première, deuxième et quatrième colonne. Une base de l'image de B est donc formé du premier, deuxième et quatrième vecteur colonne de B, à savoir $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$, $\begin{pmatrix} 2 \\ 3 \\ 3 \end{pmatrix}$, $\begin{pmatrix} 4 \\ 7 \\ 8 \end{pmatrix}$.

Démonstration. Expliquons la preuve du théorème 2.24 sur la matrice A de l'exemple 2.25, l'argument que nous allons donner est général. Le noyau de A est formé des vecteurs solutions du système homogène $AX = 0$, qui est équivalent à $A'X = 0$ où A' est la matrice échelonnée donnée précédemment. L'ensemble des solutions est paramétré par les deux variables libre x_2 et x_4 :

$$A'X = 0 \Leftrightarrow \begin{cases} 2x_1 - x_2 + x_3 + 2x_4 = 0 \\ -2x_3 + 4x_4 = 0 \end{cases} \Leftrightarrow \begin{cases} x_1 = \frac{1}{2}x_2 - 2x_4 \\ x_2 \text{ est libre} \\ x_3 = 2x_4 \\ x_4 \text{ est libre} \end{cases}$$

Par conséquent, X appartient au noyau de A si et seulement si

$$X = \begin{pmatrix} \frac{1}{2}x_2 - 2x_4 \\ x_2 \\ 2x_4 \\ x_4 \end{pmatrix} = x_2u_1 + x_4u_2 \quad \text{avec} \quad u_1 = \begin{pmatrix} \frac{1}{2} \\ 1 \\ 0 \\ 0 \end{pmatrix}, \; u_2 = \begin{pmatrix} -2 \\ 0 \\ 2 \\ 1 \end{pmatrix}.$$

De plus, (u_2, u_4) est libre car

$$\alpha u_2 + \beta u_4 = 0 \Leftrightarrow \begin{cases} \frac{1}{2}\alpha - 2\beta = 0 \\ \alpha = 0 \\ 2\beta = 0 \\ \beta = 0 \end{cases} \Leftrightarrow \alpha = \beta = 0$$

Donc le noyau de A admet bien pour base (u_1, u_2).

Pour ce qui est de l'image, les pivots de A' étant situés dans la première et troisième colonnes, le théorème 2.24 affirme qu'une base est donnée par la première et troisième colonnes de A. Justifions cela. Notons C_1, C_2, C_3 et C_4 les colonnes de A et C'_1, C'_2, C'_3 et C'_4 les colonnes de A'. Nous devons montrer que (C_1, C_3) est libre et que C_2, C_4 sont combinaisons linéaires de C_1, C_3. Pour cela, on raisonne de la façon suivante :

1. (C'_1, C'_3) est une base de l'image de A'
2. $AX = 0$ ssi $A'X = 0$. Autrement dit

$$x_1C_1 + x_2C_2 + x_3C_3 + x_4C_4 = 0 \Leftrightarrow x_1C'_1 + x_2C'_2 + x_3C'_3 + x_4C'_4 = 0$$

3. D'après le point précédent, (C'_1, C'_3) libre implique que (C_1, C_3) est aussi libre. De

même, le fait que C_2' et C_4' soient combinaisons linéaires de C_1', C_3' implique que C_2 et C_4 sont combinaisons linéaires de C_1, C_3

Nous vous laissons écrire les détails de la première et troisième étape. □

Donnons une conséquence très importante du théorème 2.24.

Énoncé indispensable 45 :

COROLLAIRE 2.26. *Si A_1' et A_2' sont deux matrices échelonnées déduites de A par des opérations élémentaires sur les lignes, alors A_1' et A_2' ont même nombre de pivots.*

En effet, Im A est défini indépendemment de toute matrice échelonnée, et il en est de même pour sa dimension, donc le nombre de pivot ne dépend pas de la matrice échelonnée. Ceci montre que la définition 1.24 du rang est cohérente. La définition "officielle" et équivalente est la suivante.

Énoncé indispensable 46 :

DEFINITION 2.27. Le rang d'une matrice A est la dimension de Im A.

2.3 Applications linéaires de $\mathbf{R}^p$ dans $\mathbf{R}^n$

Nous avons vu au premier semestre la notion de transformation linéaire du plan, par exemple les symétries et les rotations. Il s'agit ici de généraliser cette notion à des espaces de dimensions arbitraires. On se donne donc pour ce chapitre deux entiers p et n strictement positifs.

2.3.1 *Application linéaire*

Énoncé indispensable 47 :

DEFINITION 2.28. Une application $\varphi : \mathbf{R}^p \to \mathbf{R}^n$ est dite linéaire si et seulement si pour tous $u, v \in \mathbf{R}^p$ et tout $\lambda \in \mathbf{R}$, on a :

$$\varphi(u+v) = \varphi(u) + \varphi(v), \quad \text{et} \quad \varphi(\lambda u) = \lambda \varphi(u).$$

Ces deux conditions peuvent être rassemblées en une seule :

$$\varphi(u + \lambda v) = \varphi(u) + \lambda \varphi(v).$$

On note $\mathcal{L}(\mathbf{R}^p, \mathbf{R}^n)$ l'ensemble des applications linéaires de $\mathbf{R}^p$ dans $\mathbf{R}^n$.

Si $p = n$, on parle d'endormorphisme de $\mathbf{R}^p$, et on note simplement $\mathcal{L}(\mathbf{R}^p)$ l'ensemble des endomorphismes de $\mathbf{R}^p$.

exemple 2.29.

1. comme cela a été vu au premier semestre, les homothéties, rotations, projections, symétries du plan sont des applications linéaires.
2. Lorsque $p = n$, l'application identité est linéaire. Plus généralement toute homothétie de rapport $\lambda \in \mathbf{R}$, $\varphi : \mathbf{R}^p \to \mathbf{R}^p, x \mapsto \lambda x$ est linéaire.
3. Soit $\mathcal{P}_\ell$ l'ensemble des polynômes de degré au plus ℓ. On identifie $\mathcal{P}_2$ avec $\mathbf{R}^3$ en associant $P(X) = a_0 + a_1X + a_2X^2$ au vecteur $(a_0, a_1, a_2) \in \mathbf{R}^3$ et de même $\mathcal{P}_3$ avec $\mathbf{R}^4$. Les application suivantes sont linéaires :
   (a) l'intégration de $\mathcal{P}_2$ dans $\mathcal{P}_3$: $P(X) \to \int_0^X P(t)\,dt$
   $$a_0 + a_1X + a_2X^2 \longrightarrow a_0X + \frac{a_1}{2}X^2 + \frac{a_2}{3}X^3$$
   (b) la dérivation de $\mathcal{P}_3$ dans $\mathcal{P}_2$: $P(X) \to P'(X)$
   $$a_0 + a_1X + a_2X^2 + a_3X^3 \longrightarrow a_1X + 2a_2X + 3a_3X^2$$
   (c) la multiplication par $1 + 2X$ de $\mathcal{P}_2$ dans $\mathcal{P}_3$: $P(X) \to (1 + 2X)P(X)$
   $$a_0 + a_1X + a_2X^2 \longrightarrow a_0 + (a_1 + 2a_0)X + (a_2 + 2a_1)X^2 + 2a_2X^3$$
   Bien entendu, le choix du facteur multiplicatif $1 + 2X$ est arbitraire. Nous pourrions aussi étendre ces trois exemples à des polynômes de degré supérieur.

Comme dans le cas du plan, nous avons les propriétés suivantes :

Énoncé indispensable 48 :

proposition 2.30.

1. *Si φ et ψ sont deux applications linéaires de $\mathbf{R}^p$ dans $\mathbf{R}^n$, alors $\varphi + \psi : \mathbf{R}^p \to \mathbf{R}^n$ est linéaire.*
2. *Si φ est une application linéaire et λ un réel, alors $\lambda\varphi$ est linéaire.*
3. *Lorsque cela a un sens, la composée d'applications linéaires est linéaire. Précisément, si $\varphi : \mathbf{R}^p \to \mathbf{R}^n$ et $\psi : \mathbf{R}^n \to \mathbf{R}^m$ sont linéaires, alors $\psi \circ \varphi : \mathbf{R}^p \to \mathbf{R}^m$ est linéaire.*

Les applications linéaires sont en quelque sorte compatibles avec les combinaisons linéaires. En effet, si $\varphi : \mathbf{R}^p \to \mathbf{R}^n$ est une application linéaire et $u = \lambda_1 v_1 + \ldots + \lambda_m v_m$ est une combinaison linéaire dans $\mathbf{R}^p$, alors

$$\begin{aligned}
\varphi(u) &= \varphi(\lambda_1 v_1 + \ldots + \lambda_m v_m) \\
&= \lambda_1 \varphi(v_1) + \varphi(\lambda_2 v_2 + \ldots + \lambda_m v_m) \\
&= \lambda_1 \varphi(v_1) + \lambda_2 \varphi(v_2) + \varphi(\lambda_3 v_3 + \ldots + \lambda_m v_m) \\
&= \ldots \\
&= \lambda_1 \varphi(v_1) + \ldots + \lambda_m \varphi(v_m)
\end{aligned}$$

Nous en déduisons une idée maîtresse de l'algèbre linéaire : une application linéaire est caractérisée par les images des vecteurs d'une base. La formulation précise est la suivante.

Énoncé indispensable 49 :

PROPOSITION 2.31. *Soient φ et ψ deux applications linéaires de $\mathbf{R}^p$ dans $\mathbf{R}^n$. Soit $(v_1, \ldots, v_p)$ une base de $\mathbf{R}^p$. On suppose que pour tout $i = 1, \ldots, p$, $\varphi(v_i) = \psi(v_i)$. Alors $\varphi = \psi$.*

Démonstration. Tout vecteur u de $\mathbf{R}^n$ est combinaison linéaire des v_i, autrement dit $u = \lambda_1 v_1 + \ldots + \lambda_p v_p$. Nous avons alors

$$\varphi(u) = \lambda_1 \varphi(v_1) + \ldots + \lambda_p \varphi(v_p) = \lambda_1 \psi(v_1) + \ldots + \lambda_p \psi(v_p) = \psi(u).$$

□

2.3.2 *Matrices et applications linéaires*

La notion d'application linéaire peut sembler compliquée, mais en fait, comme nous allons le voir, une application linéaire n'est rien d'autre qu'une matrice. C'est-à-dire que toute application linéaire peut être représentée par une matrice et réciproquement toute matrice représente une application linéaire. Expliquons cela.

Comme dans le chapitre sur les bases et sous-espaces vectoriels, nous identifions les vecteurs de $\mathbf{R}^p$ et $\mathbf{R}^n$ avec des vecteurs colonnes à p et n lignes respectivement. Nous notons $(e_1, \ldots, e_p)$ la base canonique de $\mathbf{R}^p$.

Si A est une matrice de taille $n \times p$, nous définissons une application

$$\varphi_A : \mathbf{R}^p \to \mathbf{R}^n, \qquad \varphi_A(X) = AX \text{ pour tout } X \in \mathbf{R}^p.$$

Cette application est linéaire, en effet $A(X+Y) = AX + AY$ et $A(\lambda X) = \lambda AX$ d'après les résultats vu dans le chapitre 1.1. Donc à une matrice est associée une application linéaire. Dans l'autre sens, partant d'une application linéaire $\varphi : \mathbf{R}^p \to \mathbf{R}^n$, on définit la matrice A de taille $n \times p$ dont les vecteurs colonnes sont $\varphi(e_1), \ldots, \varphi(e_p)$.

EXEMPLE 2.32.

1. si $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}$, alors φ_A est l'application linéaire de $\mathbf{R}^2$ dans $\mathbf{R}^3$ donnée par $\varphi_A(x, y) = (x + 2y, 3x + 4y, 5x + 6y)$.
2. si φ est l'application $\mathbf{R}^3 \to \mathbf{R}^2, (x, y, z) \mapsto (x + y, x - z)$, alors $\varphi(e_1) = (1, 1)$, $\varphi(e_2) = (1, 0)$ et $\varphi(e_3) = (0, -1)$, et donc la matrice associée est $\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & -1 \end{pmatrix}$

Les deux correspondances que nous venons de définir sont inverses l'une de l'autre d'après la proposition suivante.

Énoncé indispensable 50 :

PROPOSITION 2.33.

1. *Pour toute matrice A de taille $n \times p$, $\varphi_A(e_1), \ldots, \varphi_A(e_n)$ sont les colonnes de A.*
2. *Pour toute application linéaire $\varphi : \mathbf{R}^p \to \mathbf{R}^n$, $\varphi = \varphi_A$ où A est la matrice dont les vecteurs colonnes sont $\varphi(e_1), \ldots, \varphi(e_p)$.*

Démonstration. 1. par définition $\varphi_A(e_i) = Ae_i$ qui est bien la i-ième colonne de A.

2. φ et φ_A coincident sur la base canonique, donc elles sont égale par la proposition 2.31. □

Les trois opérations que nous avons définies sur les applications linéaires, correspondent aux opérations que nous connaissons sur les matrices : la somme, la multiplication par un scalaire, et la composition qui correspond au produit.

Énoncé indispensable 51 :

PROPOSITION 2.34.

1. *Si A et B sont deux matrices de même taille et $\lambda \in \mathbf{R}$, alors $\varphi_{A+B} = \varphi_A + \varphi_B$ et $\varphi_{\lambda A} = \lambda\varphi_A$.*
2. *Si A est une matrice de taille $n \times p$ et B une matrice de taille $p \times q$, alors $\varphi_{AB} = \varphi_A \circ \varphi_B$.*

En fait, on peut aller plus loin et affirmer que les matrices ont été introduites pour avoir une représentation pratique des applications linéaires, et le produit des matrices a été défini de sorte à rendre compte de la composition des applications. Mais pour des raisons didactiques, le cours est construit dans l'autre sens : on part des matrices pour arriver aux applications linéaires.

Démonstration. Cela découle immédiatement des propriétés de la somme et du produit des matrices. Pour la première assertion,

$$\varphi_{A+B}(X) = (A+B)X = AX + BX = \varphi_A(X) + \varphi_B(X)$$
$$\varphi_{\lambda A}(X) = (\lambda A)(X) = \lambda(AX) = \lambda\varphi_A(X).$$

Pour la seconde assertion,

$$\varphi_{AB}(X) = (AB)X = A(BX) = \varphi_A(\varphi_B(X)) = (\varphi_A \circ \varphi_B)(X)$$

car le produit des matrices est associatif. □

EXEMPLE 2.35.

1. La composition de deux rotations d'angles α et β est une rotation d'angle $\alpha+\beta$. En termes de matrices, ceci équivaut à $R(\beta)R(\alpha) = R(\alpha+\beta)$ où l'on note

$$R(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

la matrice de la rotation d'angle θ.

2. Dans la troisième partie de l'exemple 2.29, l'intégration, la dérivation et la multiplication par $1+2X$ ont pour matrice respectivement

$$A = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}, \qquad C = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}$$

On calcule facilement les produits :

$$AB = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad BA = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Qu'en pensez-vous ? Peut-on vraiment dire que la primitive est l'inverse de la dérivée ?

2.3.3 *Noyau et image*

A chaque matrice, nous avons associé deux sous-espaces vectoriels, le noyau et l'image. On peut définir directement ces sous-espaces en fonction de l'application linéaire associée à la matrice.

Énoncé indispensable 52 :

DEFINITION 2.36. On définit le noyau et l' image de de $\varphi \in \mathcal{L}(\mathbf{R}^p, \mathbf{R}^n)$ par

$$\ker(\varphi) = \varphi^{-1}(\{0\}) = \{u \in \mathbf{R}^p \text{ tels que } \varphi(u) = 0\},$$

$$\mathrm{Im}(\varphi) = \varphi(\mathbf{R}^p) = \{v \in \mathbf{R}^n \text{ tels qu'il existe } u \in \mathbf{R}^p \text{ avec } \varphi(u) = v\}.$$

Comme nous l'avions vu, le noyau est un sous-espace vectoriel de $\mathbf{R}^p$, l'image un sous-espace vectoriel de $\mathbf{R}^q$. On remarque que le noyau est un sous-ensemble de l'espace de départ de l'application tandis que l'image est un sous-ensemble de l'espace d'arrivée. Le noyau et l'image nous renseignent sur le caractère injectif/surjectif de l'application. Rappelons quelques définitions.

DEFINITION 2.37. Soit une application f d'un ensemble E dans un ensemble F. On dit que

1. f est injective si et seulement si tout élément de F a au plus un antécédent par f, autrement dit

$$\forall x_1, x_2 \in E,\ f(x_1) = f(x_2) \Rightarrow x_1 = x_2$$

2. f est surjective si et seulement si tout élément de F a au moins un antécédent par f, autrement dit

$$\forall y \in F,\ \exists x \in E \text{ tel que } f(x) = y$$

3. f est bijective si et seulement si f est injective et surjective, c'est-à-dire si et seulement si tout élément de F a exactement un antécédent par f.

Énoncé indispensable 53 :

PROPOSITION 2.38.

1. *φ est injective si et seulement si* $\ker(\varphi) = \{0\}$.
2. *φ est surjective si et seulement si* $\mathrm{Im}(\varphi) = \mathbf{R}^n$.

Démonstration. Supposons φ injective, et prenons $u \in \ker(\varphi)$. Alors

$$\varphi(u) = 0 = \varphi(0)$$

et par conséquent $u = 0$. Réciproquement, supposons $\ker(\varphi) = \{0\}$. Soient u_1, u_2 dans $\mathbf{R}^p$ tels que $\varphi(u_1) = \varphi(u_2)$. Alors par linéarité, $\varphi(u_1 - u_2) = 0$, et donc $u_1 - u_2 \in \ker(\varphi)$, ce qui implique $u_1 - u_2 = 0$, autrement dit $u_1 = u_2$.

La deuxième partie de la proposition est simplement due à la définition de $\mathrm{Im}(\varphi)$ comme $\varphi(\mathbf{R}^p)$. □

Puisque le noyau et l'image de φ sont des sous-espaces vectoriels, ils ont une dimension. On peut alors mesurer le défaut d'injectivité de φ à la taille (la dimension) de son noyau : plus la dimension du noyau est grande, moins φ est injective. De même, plus la dimension de l'image est petite, moins φ est surjective. Ceci rappelle la discussion à la fin du chapitre 1.2.3 sur la résolution des systèmes linéaires.

Comme nous l'avons vu dans le chapitre 2.2.3, les dimensions de l'image et du noyau d'une matrice sont liés par une équation très simple. Avant tout, on définit le rang d'une application linéaire comme pour les matrices.

DEFINITION 2.39. Le rang de $\varphi \in \mathcal{L}(\mathbf{R}^p, \mathbf{R}^n)$ est la dimension de l'image de φ.

Énoncé indispensable 54 :

théorème 2.40. *Soit $\varphi \in \mathcal{L}(\mathbf{R}^p, \mathbf{R}^n)$. Alors on a*

$$\begin{aligned}\dim(\mathbf{R}^p) = p &= \dim(\ker(\varphi)) + \dim(\mathrm{Im}(\varphi)) \\ &= \dim(\ker(\varphi)) + \mathrm{rg}(\varphi)\end{aligned}$$

Cela a été vu dans le théorème 2.24. La preuve que nous avions donnée était basée sur l'algorithme de Gauss. Voici une preuve alternative, plus abstraite, mais aussi plus directe.

Démonstration. Lorsque φ est identiquement nulle, le résultat est vrai : le noyau de φ est alors de dimension p et son image de dimension 0. Si φ n'est pas identiquement nulle, commençons par choisir une base $(u_1, \cdots, u_d)$ de $\ker(\varphi)$. D'après le théorème de la base incomplète, cf. Proposition 2.23, on peut compléter cette base en $(u_1, \cdots, u_d, u_{d+1}, \cdots, u_p)$ base de $\mathbf{R}^p$.

Prouvons maintenant que $(\varphi(u_{d+1}), \cdots, \varphi(u_p))$ est une base de $\mathrm{Im}(\varphi)$. Pour cela, prenons un vecteur v dans $\mathrm{Im}(\varphi)$. Il est par définition l'image d'un vecteur u de $\mathbf{R}^p$: $v = \varphi(u)$. Ce vecteur u est lui-même combinaison linéaire des vecteurs de la base (u_i). Donc

$$v = \varphi(u) = \varphi\left(\sum_{i=1}^{p} \lambda_i u_i\right) = \sum_{i=1}^{p} \lambda_i \varphi(u_i) = \sum_{i=d+1}^{p} \lambda_i \varphi(u_i)$$

car $\varphi(u_1) = \cdots = \varphi(u_d) = 0$. Ceci montre que $(\varphi(u_{d+1}), \cdots, \varphi(u_p))$ engendre l'image. Montrons que cette famille est libre.

$$\sum_{i=d+1}^{p} \lambda_i \varphi(u_i) = 0 \;\Rightarrow\; \varphi\left(\sum_{i=d+1}^{p} \lambda_i u_i\right) = 0 \;\Rightarrow\; \sum_{i=d+1}^{p} \lambda_i u_i \in \ker(\varphi)$$

Comme $(u_1, \ldots, u_d)$ est une base du noyau, nous avons donc

$$\sum_{i=d+1}^{p} \lambda_i u_i = \sum_{i=1}^{d} \alpha_i u_i$$

pour des réels α_i. Mais $(u_1, \ldots, u_p)$ étant une base, cela montre que les λ_i sont tous nuls, donc la famille $(\varphi(u_{d+1}), \cdots, \varphi(u_p))$ est libre. Ceci prouve que $\mathrm{Im}(\varphi)$ est de dimension $p - d$, et donc le résultat annoncé. □

exemple 2.41.

1. Si $\varphi : \mathbf{R}^2 \to \mathbf{R}^3$, le couple $(\dim(\ker(\varphi)), \mathrm{rg}(\varphi))$ peut uniquement prendre les valeurs $(0,2)$, $(1,1)$ et $(2,0)$. En particulier, on ne peut avoir $\mathrm{rg}(\varphi) = 3$, φ ne peut être surjective.
2. Si $\varphi : \mathbf{R}^3 \to \mathbf{R}^2$, le couple $(\dim(\ker(\varphi)), \mathrm{rg}(\varphi))$ peut uniquement prendre les valeurs $(1,2)$, $(2,1)$ et $(3,0)$. En particulier, on ne peut avoir $\ker(\varphi) = \{0\}$, φ ne peut être injective.

Le théorème du rang a une conséquence importante lorsque $p = n$:

Énoncé indispensable 55 :

COROLLAIRE 2.42. *Soit $\varphi \in \mathcal{L}(\mathbf{R}^p)$.*

1. *Si φ est injective, alors φ est automatiquement surjective, donc bijective.*
2. *Si φ est surjective, alors φ est automatiquement injective, donc bijective.*

Autrement dit,

$$\varphi \text{ bijective} \Leftrightarrow \varphi \text{ injective} \Leftrightarrow \varphi \text{ surjective}$$

Démonstration.

$$\begin{aligned}
\varphi \text{ injective} &\Leftrightarrow \ker(\varphi) = \{0\} && \text{d'après la proposition 2.38} \\
&\Leftrightarrow \dim(\ker(\varphi)) = 0 && \text{d'après la proposition 2.22} \\
&\Leftrightarrow \dim(\mathrm{Im}(\varphi)) = p && \text{d'après le théorème du rang et comme } p = n \\
&\Leftrightarrow \mathrm{Im}(\varphi) = \mathbf{R}^p && \text{d'après la proposition 2.22} \\
&\Leftrightarrow \varphi \text{ surjective} && \text{d'après la proposition 2.38.}
\end{aligned}$$

□

2.4 Diagonalisation

2.4.1 *Valeurs propres et vecteurs propres*

Dans ce chapitre, on s'intéresse aux applications linéaires de $\mathbf{R}^n$ dans lui-même. Pour une telle application φ, on peut comparer un vecteur $u \in \mathbf{R}^n$ et son image $\varphi(u) \in \mathbf{R}^n$. La situation qui nous intéresse est lorsque $\varphi(u)$ est un multiple de u.

Énoncé indispensable 56 :

DEFINITION 2.43. Soit φ une application linéaire de $\mathbf{R}^n$ dans lui-même. Soit $\lambda \in \mathbf{R}$. On dit que λ est valeur propre de φ si et seulement s'il existe un vecteur **non nul** $u \in \mathbf{R}^n$ tel que

$$\varphi(u) = \lambda u$$

Un tel u est alors appelé vecteur propre associé à la valeur propre λ.

Donnons quelques exemples. Avant cela, rappelons qu'une application linéaire s'identifie à une matrice, cf. chapitre 2.3.2. Comme ici les applications linéaires ont pour espaces de départ et d'arrivée $\mathbf{R}^n$, leurs matrices sont carrées de taille n. Si A est une telle matrice, ses valeurs propres et vecteurs propres sont les valeurs propres et vecteurs propres de l'application linéaire associée. Autrement dit, u est vecteur propre de A avec pour valeur propre λ si et seulement si u est un vecteur non-nul de $\mathbf{R}^n$ et $Au = \lambda u$.

EXEMPLE 2.44.

1. La matrice $A = \begin{pmatrix} 0,8 & 0,3 \\ 0,2 & 0,7 \end{pmatrix}$ admet pour vecteurs propres $u_1 = \begin{pmatrix} 0,6 \\ 0,4 \end{pmatrix}$ et $u_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. En effet,
$$Au_1 = u_1, \qquad Au_2 = \tfrac{1}{2}u_2$$
et les valeurs propres associées sont donc 1 et $\frac{1}{2}$. Si on applique A à ces équations, il vient
$$\begin{aligned} A^2(u_1) &= A(Au_1) = A(u_1) = u_1 \\ A^2(u_2) &= A(Au_2) = A(\tfrac{1}{2}u_2) = \tfrac{1}{2}Au_2 = \tfrac{1}{4}u_2. \end{aligned}$$
ainsi u_1 et u_2 sont des vecteurs propres de A^2, les valeurs propres correspondantes sont 1 et $\frac{1}{4}$. En réitérant ce calcul, on en déduit que pour tout entier positif p,
$$A^p(u_1) = u_1, \qquad A^p(u_2) = (\tfrac{1}{2})^p u_2.$$
Ce fait est très général : si $Au = \lambda u$ avec A une matrice carrée quelconque, alors $A^p u = \lambda^p u$. Donc en élevant A à une puissance positive, le vecteur u reste un vecteur propre, la valeur propre correspondante est élevée à la puissance p.
2. La matrice $R = \begin{pmatrix} 0,5 & 0,5 \\ 0,5 & 0,5 \end{pmatrix}$ admet pour vecteurs propres
$$v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$
En effet $Rv_1 = v_1$ et $Rv_2 = 0 = 0v_2$. Remarquons que R est la matrice de la projection de $\mathbf{R}^2$ sur la droite engendrée par v_1 parallèlement à la droite engendrée par v_2. Ce fait est général : comme on l'a vu au premier semestre, si (i, j) est une base de $\mathbf{R}^2$, la projection r de $\mathbf{R}^2$ sur $\mathrm{Vect}(i)$ parallèlement à $\mathrm{Vect}(j)$ est l'application linéaire donnée par
$$r(xi + yj) = xi, \qquad \forall x, y \in \mathbf{R}.$$
En particulier, $r(i) = i$ et $r(j) = 0$, donc i et j sont vecteur propres de r, de valeurs propres 1 et 0.
3. La matrice $S = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ admet pour vecteurs propres v_1 et v_2 définis comme ci-dessus, les valeurs propres correspondantes sont 1 et -1. Notons que S est la matrice de la symétrie de $\mathbf{R}^2$ sur la droite engendrée par v_1 parallèlement à la droite engendrée par v_2. Plus généralement, si (i, j) est une base de $\mathbf{R}^2$, la symétrie s de $\mathbf{R}^2$ d'axe $\mathrm{Vect}(i)$ parallèlement à la droite $\mathrm{Vect}(j)$ admet pour vecteurs propres i et j, les valeurs propres étant 1 et -1. En effet par définition, $s(xi + yj) = xi - yj$ pour tout $x, y \in \mathbf{R}$, donc $s(i) = i$ et $s(j) = -j$.
4. Dans les deux exemples précédents, R et S partagent les mêmes vecteurs propres. La raison est que $S = 2R - \mathrm{Id}$. En effet si $Ru = \lambda u$, alors $Su = 2Ru - u = (2\lambda - 1)u$. Donc tout vecteur de propre de R est aussi un vecteur de propre de S, les valeurs propres étant reliées comme les matrices. Ceci rappelle ce que nous avions observé dans le premier exemple avec la matrice A et ses puissances. Pouvez-vous énoncer un résultat qui généralise ces deux observations ?

Dans tous ces exemples, les vecteurs propres nous ont été donnés. Comment faire pour les trouver ? On commence par les valeurs propres.

proposition 2.45. *Soit A une matrice carrée de taille n et $\lambda \in \mathbf{R}$. Alors*

$$\lambda \text{ est valeur propre de } A \quad \Leftrightarrow \quad \det(A - \lambda \operatorname{Id}) = 0$$

Démonstration. λ est valeur propre ssi il existe un vecteur non-nul X de $\mathbf{R}^n$ tel que $AX = \lambda X$ ssi le système homogène $(A - \lambda)X = 0$ admet une solution non triviale ssi $A - \lambda \operatorname{Id}$ n'est pas inversible d'après le théorème 1.30, ce qui équivaut au fait que le déterminant de $A - \lambda \operatorname{Id}$ soit nul d'après le théorème 1.37. □

exemple 2.46. Soit $A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$ et $\lambda \in \mathbf{R}$. Alors

$$\det(A - \lambda \operatorname{Id}) = \det \begin{pmatrix} 1-\lambda & 2 \\ 2 & 4-\lambda \end{pmatrix} = (1-\lambda)(4-\lambda) - 4 = \lambda^2 - 5\lambda = \lambda(\lambda - 5)$$

Bien entendu $\lambda(\lambda - 5) = 0$ si et seulement si $\lambda = 0$ ou $\lambda = 5$. Donc la matrice A admet pour valeurs propres 0 et 5.

Dans cet exemple, les valeurs propres sont les zéros d'un polynôme. C'est un fait général.

Énoncé indispensable 57 :

théorème 2.47. *Soit A une matrice carrée de taille n. Alors l'application $\chi_A : \mathbf{R} \to \mathbf{R}$ définie par $\chi_A(\lambda) = \det(A - \lambda \operatorname{Id})$ est un polynôme de degré n. On l'appelle le polynôme caractéristique de A. Les racines réelles de ce polynôme sont les valeurs propres de A.*

Démonstration. Expliquons la preuve pour une matrice 3×3. Si u_1, u_2 et u_3 sont trois vecteurs de $\mathbf{R}^3$, nous notons $f(u_1, u_2, u_3)$ le déterminant de la matrice de vecteurs colonnes u_1, u_2 et u_3. Nous devons montrer que l'application

$$\lambda \to f(u_1 - \lambda e_1, u_2 - \lambda e_2, u_3 - \lambda e_3)$$

est un polynôme de degré 3. Nous savons d'après la proposition 1.43 que f est linéaire par rapport à chacun de ses arguments. En développant successivement par rapport au premier, deuxième puis troisième argument, il vient

$$\begin{aligned}
&f(u_1 - \lambda e_1, u_2 - \lambda e_2, u_3 - \lambda e_3) \\
&= f(u_1, u_2 - \lambda e_2, u_3 - \lambda e_3) - \lambda f(e_1, u_2 - \lambda e_2, u_3 - \lambda e_3) \\
&= f(u_1, u_2, u_3 - \lambda e_3) - \lambda f(u_1, e_2, u_3 - \lambda e_3) \\
&\quad - \lambda f(e_1, u_2, u_3 - \lambda e_3) + \lambda^2 f(e_1, e_2, u_3 - \lambda e_3) \\
&= f(u_1, u_2, u_3) - \lambda f(u_1, u_2, e_3) - \lambda f(u_1, e_2, u_3) + \lambda^2 f(u_1, e_2, e_3) \\
&\quad - \lambda f(e_1, u_2, u_3) + \lambda^2 f(e_1, u_2, e_3) + \lambda^2 f(e_1, e_2, u_3) - \lambda^3 f(e_1, e_2, e_3)
\end{aligned}$$

ce qui montre le résultat. Remarquer que lorsqu'on regroupe les termes de même degré, l'on obtient la formule plus symétrique :

$$f(u_1,u_2,u_3)-\lambda(f(e_1,u_2,u_3)+f(u_1,e_2,u_3)+f(u_1,u_2,e_3))$$
$$+\lambda^2(f(u_1,e_2,e_3)+f(e_1,u_2,e_3)+f(e_1,e_2,u_3))-\lambda^3 f(e_1,e_2,e_3)$$

La preuve pour des matrices de taille plus grande est la même. □

Comme cela a été vu dans le cours du premier semestre, un polynôme de degré n a au plus n racines. D'où la conséquence très importante.

COROLLAIRE 2.48. *Une matrice carrée de taille n a au plus n valeurs propres*

Une fois que nous connaissons les valeurs propres, il est très simple de trouver les vecteurs propres correspondants. En effet, si λ est valeur propre de A, alors les vecteurs propres X associés sont les solutions non-nulles du système homogène $(A-\lambda\,\mathrm{Id})X=0$.

EXEMPLE 2.49. Reprenons la matrice A de l'exemple 2.46. Cherchons un vecteur propre associé à la valeur propre 0.

$$(A-0\,\mathrm{Id})\begin{pmatrix}x\\y\end{pmatrix}=0\Leftrightarrow\begin{pmatrix}1&2\\2&4\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}=0\Leftrightarrow\begin{cases}x+2y=0\\2x+4y=0\end{cases}\Leftrightarrow\begin{cases}x=-2y\\y\text{ est libre}\end{cases}$$

Une solution non-nulle est $\begin{pmatrix}-2\\1\end{pmatrix}$. C'est un vecteur propre de A pour la valeur propre 0.

De même les vecteurs propres associés à la valeur propre 5 sont les solutions non-nulles du système :

$$(A-5\,\mathrm{Id})\begin{pmatrix}x\\y\end{pmatrix}=0\Leftrightarrow\begin{pmatrix}-4&2\\2&-1\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}=0\Leftrightarrow\begin{cases}-4x+2y=0\\2x-y=0\end{cases}\Leftrightarrow\begin{cases}x=\frac{1}{2}y\\y\text{ est libre}\end{cases}$$

Donc un vecteur propre de A pour la valeur propre 5 est par exemple le vecteur $\begin{pmatrix}\frac{1}{2}\\1\end{pmatrix}$.

Comme nous le voyons sur cet exemple, pour une valeur propre donnée, il y a plusieurs vecteurs propres. Il n'y a rien de surprenant puisque nous savons depuis longtemps que l'ensemble des solutions d'un système homogène est un sous-espace vectoriel. Remarquons que le sous-espace vectoriel en question n'est autre que $\ker(A-\lambda\,\mathrm{Id})$. Vu son importance, on lui donne un nom.

Énoncé indispensable 58 :

DEFINITION 2.50. Si λ est valeur propre de A, le noyau de $A-\lambda\,\mathrm{Id}$ s'appelle l'espace propre de A pour la valeur propre λ. Il est composé de l'ensemble des vecteurs propres associés à λ et du vecteur nul.

exemple 2.51. Nous pouvons reformuler l'observation faite dans l'exemple 2.44 à propos des symétries et projections de $\mathbf{R}^2$. Soient D_1 et D_2 deux droites vectorielles distinctes de $\mathbf{R}^2$. Alors la projection sur D_1 parallèlement à D_2 admet pour valeurs propres 1 et 0, les espaces propres correspondants sont D_1 et D_2. La symétrie sur D_1 parallèlement à D_2 admet pour valeurs propres 1 et -1, les espaces propres correspondants sont D_1 et D_2.

2.4.2 *Diagonalisation*

Énoncé indispensable 59 :

théorème 2.52. *Soit A une matrice carrée de taille n. Alors les deux assertions suivantes sont équivalentes :*

1. *$\mathbf{R}^n$ admet une base formée de vecteurs propres de A.*
2. *il existe une matrice inversible P de taille n telle que $P^{-1}AP$ soit une matrice diagonale.*

Lorsqu'elles sont vérifiées, on dit que A est diagonalisable.

Démonstration. La preuve repose sur la troisième assertion du théorème 2.7 et la remarque suivante. Soient P et Δ deux matrices carrées de taille n, Δ étant diagonale. Notons $u_1, \ldots, u_n$ les vecteurs colonnes de P et $\lambda_1, \ldots, \lambda_n$ les coefficients diagonaux de Δ. Alors d'une part, AP est la matrice de vecteurs colonnes $Au_1, \ldots, Au_n$; d'autre part, $P\Delta$ est la matrice de vecteurs colonnes $\lambda_1 u_1, \ldots, \lambda_n u_n$. Par conséquent

$$AP = P\Delta \quad \Leftrightarrow \begin{cases} Au_1 = \lambda_1 u_1 \\ \vdots \\ A_n u_n = \lambda_n u_n \end{cases}$$

Montrons que la première assertion implique la seconde. Si chaque u_i est un vecteur propre de A de valeur propre λ_i, alors $Au_i = \lambda_i u_i$ pour tout i et donc $AP = P\Delta$. Si de plus les u_i forment une base, P est inversible d'après le théorème 2.7, et en multipliant $AP = P\Delta$ à gauche par P^{-1}, il vient $P^{-1}AP = \Delta$.

Montrons la réciproque. Si $P^{-1}AP = \Delta$, en multipliant à gauche par P, il vient $AP = P\Delta$, autrement dit $Au_i = \lambda_i u_i$ avec $u_1, \ldots, u_n$ les vecteurs colonnes de P et $\lambda_1, \ldots, \lambda_n$ les coefficients diagonaux de Δ. Les u_i étant les vecteurs colonnes d'une matrice inversible, d'après théorème 2.7 à nouveau, ils forment une base. □

Dans la preuve, il apparait le fait essentiel suivant : si A est diagonalisable, les matrices P et Δ sont reliées à la base de vecteurs propres $(u_1, \ldots, u_n)$ comme suit : les u_i sont les vecteurs colonnes de P et les valeurs propres λ_i sont les coefficients diagonaux de Δ.

exemple 2.53. Reprenons les matrices A, R et S de l'exemple 2.44.

$$P_1^{-1}AP_1 = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \qquad \text{avec } P_1 = \begin{pmatrix} 0,6 & 1 \\ 0,4 & -1 \end{pmatrix},$$
$$P_2^{-1}RP_2 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad P_2^{-1}SP_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad \text{avec } P_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Il ne faut pas croire que toute matrice est diagonalisable. Le problème se présente déjà pour des matrices 2×2.

exemple 2.54.

1. La rotation de $\mathbf{R}^2$ d'angle θ admet pour matrice $A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$. Son polynôme caractéristique est
$$P_A(\lambda) = (\lambda - \cos\theta)^2 + \sin^2\theta.$$
Lorsque θ est différent de 0 ou π modulo 2π, ce polynôme n'admet pas de racines réelles. Les rotations correspondantes n'ont donc pas de valeurs propres et ne sont pas diagonalisables.
2. La matrice $B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ admet pour polynôme caractéristique $P_B(\lambda) = \lambda^2$. Sa seule racine est 0. L'espace propre correspondant est la droite engendrée par $(1, 0)$, qui est de dimension 1. Donc une famille libre de vecteurs propres aura au plus un vecteur, donc ne pourra être une base de $\mathbf{R}^2$. Donc B n'est pas diagonalisable.
3. Plus généralement, soit une matrice C de taille 2×2 admettant une unique valeur propre. Alors C est diagonalisable si et seulement si C est diagonale. La raison est que si C est diagonalisable avec pour unique valeur propre λ, alors la matrice diagonale correspondante est $\lambda\,\mathrm{Id}$. Et en multipliant $P^{-1}CP = \lambda\,\mathrm{Id}$ à gauche par P et à droite par P^{-1}, il vient $C = P(\lambda\,\mathrm{Id})P^{-1} = \lambda PP^{-1} = \lambda\,\mathrm{Id}$.

Dans tous ces exemples, nous avions une ou zéro valeur propre. Lorsque qu'une matrice 2×2 admet deux valeurs propres, elle est toujours diagonalisale d'après la proposition suivante.

proposition 2.55. *Si A de taille 2×2 admet pour vecteurs propres u_1, u_2 de valeurs propres distinctes λ_1 et λ_2, alors (u_1, u_2) est une base de $\mathbf{R}^2$.*

Démonstration. Les espaces propres sont des sous-espace vectoriels de $\mathbf{R}^2$, il sont donc stables par homothéties, et si u_1 était un multiple de u_2, alors u_1 et u_2 auraient même valeur propre, ce qui contredit $\lambda_1 \neq \lambda_2$. Pour la même raison, u_2 n'est pas un multiple de u_1. Donc (u_1, u_2) est libre. Il s'agit donc d'une base puisque nous sommes dans $\mathbf{R}^2$. □

Énoncé indispensable 60 :

Avec l'exemple précédent et cette dernière proposition, nous pouvons donner la

méthode génerale suivante pour décider si une matrice A de taille 2×2 est diagonalisable et le cas échéant la diagonaliser.

1. calculer le polynôme caractéristique P_A et chercher ses racines réelles.
2. si P_A n'a pas de racine réelle, la matrice n'est pas diagonalisable.
3. si P_A a une unique racine réelle, A est diagonalisable si et seulement si A est diagonale.
4. si P_A a deux racines réelles distinctes λ_1 et λ_2, A est diagonalisable. Dans ce cas, on trouve deux vecteurs propres v_1, v_2 en résolvant $Av_1 = \lambda_1 v_1$ et $Av_2 = \lambda_2 v_2$. Alors (v_1, v_2) est une base propre. Et

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

où P est la matrice de vecteurs colonnes v_1, v_2.

2.4.3 *Systèmes dynamiques*

Dans ce chapitre on se donne une matrice A diagonalisable de taille n. On note $(u_1, \ldots, u_n)$ une base de vecteurs propres, $\lambda_1, \ldots, \lambda_n$ les valeurs propres correspondantes, P la matrice de vecteurs colonnes $u_1, \ldots, u_n$ et Δ la matrice diagonale de coefficients diagonaux les $\lambda_1, \ldots, \lambda_n$. Nous avons donc $P^{-1}AP = \Delta$.

proposition 2.56. *Pour tout entier positif p, nous avons*

1. *$A^p = P\Delta^p P^{-1}$ et Δ^p est la matrice diagonale de coefficients diagonaux $\lambda_1^p, \ldots, \lambda_n^p$.*
2. *si $v = \alpha_1 u_1 + \ldots + \alpha_n u_n$, alors $A^p v = \alpha_1 \lambda_1^p u_1 + \ldots + \alpha_n \lambda_n^p u_n$.*

Démonstration. Rapellons que la puissance 0-ième d'une matrice carrée est définie comme étant l'identité. Donc pour $p = 0$, $A^p = P\Delta^p P^{-1}$ s'écrit $\mathrm{Id} = P\,\mathrm{Id}\,P^{-1}$ qui est bien vrai. En multipliant $P^{-1}AP = \Delta$ à gauche par P et à droite par P^{-1}, il vient $A = P\Delta P^{-1}$. Ensuite $A^2 = P\Delta P^{-1}P\Delta P^{-1} = P\Delta^2 P^{-1}$, $A^3 = A^2 A = P\Delta^2 P^{-1}P\Delta P^{-1} = P\Delta^3 P^{-1}$ et en réitérant p fois ce calcul, il vient $A^p = P\Delta^p P^{-1}$.

Pour montrer la deuxième assertion, on établit par récurrence sur p que pour tout i, $A^p u_i = \lambda_i^p u_i$. Et on en déduit facilement le résultat. □

exemple 2.57.

1. Suivons l'évolution de la population de deux grandes villes, disons Paris et Lyon. On suppose que chaque année, 20 pour cent des parisiens déménagent à Lyon et 30 pour cents des lyonnais s'installent à Paris. Si l'on note x_k et y_k le nombre d'habitants à Paris et Lyon l'année k, nous avons $x_{k+1} = 0,8x_k + 0,3y_k$ et $y_{k+1} = 0,2x_k + 0,7y_k$. Autrement dit,

$$\begin{pmatrix} x_{k+1} \\ y_{k+1} \end{pmatrix} = A \begin{pmatrix} x_k \\ y_k \end{pmatrix} \qquad \text{avec } A = \begin{pmatrix} 0,8 & 0,3 \\ 0,2 & 0,7 \end{pmatrix}$$

Par une simple récurrence, nous en déduisons la population à l'année k en fonc-

tion de la population la première année

$$\begin{pmatrix} x_k \\ y_k \end{pmatrix} = A^k \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} \tag{2.2}$$

Nous avons diagonalisé A dans l'exemple 2.53. Nous déduisons alors de la proposition 2.56 que

$$A^k = P \begin{pmatrix} 1 & 0 \\ 0 & (\frac{1}{2})^k \end{pmatrix} P^{-1} \qquad \text{avec} \qquad P = \begin{pmatrix} 0,6 & 1 \\ 0,4 & -1 \end{pmatrix}.$$

On trouve facilement $P^{-1} = \begin{pmatrix} 1 & 1 \\ 0,4 & -0,6 \end{pmatrix}$ et on déduit de (2.2) par quelques lignes de calcul que :

$$\begin{cases} x_k = 0,6(x_0 + y_0) + (\frac{1}{2})^k (0,4x_0 - 0,6y_0) \\ y_k = 0,4(x_0 + y_0) + (\frac{1}{2})^k (-0,4x_0 + 0,6y_0) \end{cases}.$$

Il est déjà remarquable que nous ayons une formule aussi simple, mais le meilleur est à venir. Comme $(1/2)^k \to 0$ lorsque $k \to \infty$,

$$\lim_{k\to\infty} x_k = 0,6(x_0 + y_0), \qquad \lim_{k\to\infty} y_k = 0,4(x_0 + y_0)$$

autrement dit, au bout d'un temps suffisamment long, on tend vers un équilibre où 60 pour cent de la population totale vit à Paris et 40 pour cent à Lyon. Et ceci se produit quelque soit la répartition initiale.

Remarquons enfin que $\begin{pmatrix} 0,6 \\ 0,4 \end{pmatrix}$ est vecteur propre de A pour la valeur propre 1. Bien évidemment, ceci n'est pas un hasard, la suite de l'histoire est dans l'exemple 2.61.

2. La suite de Fibonacci est la suite (F_k) qui vérifie la relation de récurrence $F_{k+2} = F_{k+1} + F_k$ avec la condition initiale $F_0 = 0$, $F_1 = 1$. Elle apparait dans le Liber Abaci (Fibonacci, 1202) pour modéliser l'évolution d'une population de lapins. Chaque couple donne naissance à un autre couple chaque mois et un couple commence à se reproduire au deuxième mois après sa naissance.

   On calcule facilement les premiers termes de la suite : 0,1,1,2,3,5,8,13,21,34,55.... On ne reconnait pas d'expression particulière. Nous allons trouver F_k en diagonalisant une matrice bien choisie. Tout d'abord, nous avons

$$\begin{pmatrix} F_{k+2} \\ F_{k+1} \end{pmatrix} = B \begin{pmatrix} F_{k+1} \\ F_k \end{pmatrix} \qquad \text{avec } B = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$$

   ce qui par une simple récurrence donne

$$\begin{pmatrix} F_{k+1} \\ F_k \end{pmatrix} = B^k \begin{pmatrix} F_1 \\ F_0 \end{pmatrix} = B^k \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

Le polynôme caractéristique de B est $\lambda^2 - \lambda - 1$. Il admet pour racines

$$\lambda_1 = \frac{1+\sqrt{5}}{2} \simeq 1,618 \qquad \lambda_2 = \frac{1-\sqrt{5}}{2} \simeq -0,618$$

Donc B a deux valeur propres. Les vecteurs $u_1 = \begin{pmatrix}\lambda_1\\1\end{pmatrix}$ et $u_2 = \begin{pmatrix}\lambda_2\\1\end{pmatrix}$ sont propres de valeurs propres propres λ_1 et λ_2. Nous avons

$$\begin{pmatrix}1\\0\end{pmatrix} = \frac{1}{\lambda_1-\lambda_2}(u_1-u_2) = \frac{1}{\sqrt{5}}(u_1-u_2)$$

Donc par la deuxième assertion de la proposition 2.56, il vient

$$\begin{pmatrix}F_{k+1}\\F_k\end{pmatrix} = \frac{1}{\sqrt{5}}(\lambda_1^k u_1 - \lambda_2^k u_2)$$

La deuxième ligne de cette égalité s'écrit

$$F_k = \frac{1}{\sqrt{5}}(\lambda_1^k - \lambda_2^k) = \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^k - \left(\frac{1-\sqrt{5}}{2}\right)^k\right)$$

Cette formule apparait dans les écrits d'Euler en 1765. Curieusement, chaque F_k étant un nombre entier, les fractions et racines carrés de 5 doivent se simplifier. Le terme λ_2^k étant strictement inférieur à 1 dès que $k > 0$, F_k est la partie entière de $\frac{1}{\sqrt{5}}\lambda_1^k$. Remarquons aussi que le rapport F_{k+1}/F_k converge vers λ_1 lorsque k tend vers l'infini. Le nombre λ_1 s'appelle nombre d'or et il représenterait selon certains une proportion parfaite.

2.4.4 *Une condition suffisante pour diagonaliser*

Nous avons vu dans la proposition 2.55 qu'une condition suffisante pour qu'une matrice 2×2 soit diagonalisable est qu'elle admette deux valeurs propres distinctes. Nous allons généraliser ceci aux matrices de taille $n \times n$.

proposition 2.58. *Si $u_1, \ldots, u_k$ est une famille de vecteurs propres de A dont les valeurs propres $\lambda_1, \ldots, \lambda_k$ sont deux à deux distinctes, alors $(u_1, \ldots, u_k)$ est une famille libre.*

Démonstration. Commençons avec deux vecteurs propres u_1, u_2. Soit une combinaison linéaire nulle $\alpha_1 u_1 + \alpha_s u_2 = 0$. Alors en appliquant $A - \lambda\,\mathrm{Id}$, il vient

$$\alpha_1(\lambda_1 - \lambda)u_1 + \alpha_2(\lambda_2 - \lambda)u_2 = 0.$$

Pour $\lambda = \lambda_1$, cela se simplifie en $\alpha_2(\lambda_1 - \lambda_2)u_2 = 0$. Comme $u_2 \neq 0$ et $\lambda_1 \neq \lambda_2$, cela implique que α_2 est nul. Par le même raisonnement à partir de $\lambda = \lambda_2$, il vient $\alpha_1 = 0$. Donc la famille (u_1, u_2) est libre.

Pour une famille de k vecteurs, supposons $\alpha_1 u_1 + \ldots + \alpha_k u_k = 0$. Appliquons $A - \lambda_1\,\mathrm{Id}$, il vient

$$\alpha_1(\lambda_1 - \lambda_1)u_1 + \alpha_2(\lambda_2 - \lambda_1)u_2 + \ldots + \alpha_k(\lambda_k - \lambda_1)u_k = 0$$

ce qui se simplifie en $\alpha_2(\lambda_2 - \lambda_1)u_2 + \ldots + \alpha_k(\lambda_k - \lambda_1)u_k = 0$. De même nous appliquons $A - \lambda_2 \mathrm{Id}$ puis $A - \lambda_3 \mathrm{Id}$ jusqu'à $A - \lambda_{k-1} \mathrm{Id}$, il vient

$$\alpha_k(\lambda_k - \lambda_{k-1}) \ldots (\lambda_k - \lambda_1)u_k = 0.$$

Et comme λ_k est supposé différent de $\lambda_{k-1}, \ldots, \lambda_1$, cela donne $\alpha_k = 0$. On raisonne de même avec les autres vecteurs pour en déduire que tous les α_i sont nuls. □

théorème 2.59. *Si le polynôme caractéristique de A admet n racines réelles distinctes, alors A est diagonalisable.*

Démonstration. Les vecteurs propres associés aux n valeurs propres forment une famille libre d'après la proposition 2.58 et donc une base d'après la proposition 2.10. □

2.4.5 *Matrices et applications linéaires*

Comme nous l'avons vu dans le chapitre 2.1, lorsque l'on change de bases, les coordonnées des vecteurs sont modifiés par multiplication par une matrice inversible. Plus précisément, soient (e_i) la base canonique de $\mathbf{R}^n$, (u_i) une base quelconque de $\mathbf{R}^n$. Alors si un vecteur v a pour coordonnées (x_i) dans la base (e_i) et (y_i) dans la base (u_i), c'est-à-dire $v = x_1e_1 + \ldots + x_ne_n = y_1u_1 + \ldots + y_nu_n$, alors

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = P \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \qquad \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = P^{-1} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

où P est la matrice de vecteurs colonnes $u_1, \ldots, u_n$.

Rappelons aussi que si A est une matrice carrée de taille n, l'application linéaire associée $\varphi_A : \mathbf{R}^n \to \mathbf{R}^n$ est définie par $\varphi_A(X) = AX$. Autrement dit, on passe des coordonnées de X dans la base canonique aux coordonnées de $\varphi_A(X)$ dans la base canonique en multipliant par A.

proposition 2.60. *Si un vecteur v de $\mathbf{R}^n$ a pour coordonnées (y_i) dans la base (u_i), les coordonnées (z_i) du vecteur $\varphi_A(v)$ dans la base (u_i) sont données par*

$$\begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} = P^{-1}AP \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

En particulier, les coefficients de la j-ième colonne de $P^{-1}AP$ sont les coordonnées de $\varphi_A(u_j)$ dans la base (u_i).

Démonstration. On passe des coordonnées (y_i) aux coordonnées (x_i) en multipliant par P, puis aux coordonnées de $\varphi_A(v)$ dans la base canonique en multipliant par A, et enfin aux coordonnées (z_i) en multipliant par P^{-1} . □

On appelle $P^{-1}AP$ la matrice de φ_A dans la base (u_i). En particulier, la matrice de φ_A dans la base canonique est A.

Souvent, la résolution d'un problème en algèbre linéaire consiste à raisonner dans

une base bien choisie. Typiquement, lorsqu'on s'intéresse à une application linéaire diagonalisable de $\mathbf{R}^n$, on a intérêt à travailler dans une base propre.

exemple 2.61.

1. Soient D_1 et D_2 deux droites vectorielles distincets de $\mathbf{R}^2$. Pour définir la projection r sur D_1 parallèlement à D_2 ou la symétrie s sur D_1 parallèlement à D_2, nous avons une formule très simple en fonction des coordonnées dans la base (i, j) où i et j sont les vecteurs directeurs de D_1 et D_2 :

$$r(xi + yj) = xi, \qquad s(xi + yj) = xi - yj$$

   Et comme nous l'avons déjà remarqué, (i, j) est une base propre dans les deux cas.

2. Dans l'exemple 2.57 sur l'évolution d'une population, les coordonnées x et y de $\mathbf{R}^2$ représentent le nombre d'habitants à Paris et Lyon respectivement. L'évolution de la population se fait chaque année par multiplication par une matrice A. Les coordonnées α, β dans la base propre (u_1, u_2) sont particulièrement adaptées pour décrire cette évolution. Nous avons

$$\begin{pmatrix} x \\ y \end{pmatrix} = \alpha u_1 + \beta u_2, \qquad u_1 = \begin{pmatrix} 0,6 \\ 0,4 \end{pmatrix},\ u_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

   Un petit calcul montre que $\alpha = x + y$. Ainsi α représente la population totale. Rappelons que $\alpha u_1 = \begin{pmatrix} 0,6\alpha \\ 0,4\alpha \end{pmatrix}$ est la population d'équilibre vers laquelle on converge lorsque le nombre d'année tend vers l'infini. Comme $x = 0,6\alpha + \beta$ et $y = 0,4\alpha - \beta$, on interprète β comme l'écart à cette population d'équilibre. Si l'on note α_k, β_k les coordonnées dans la base propre de la population au bout de k années, alors les valeurs propres étant 1 et $\frac{1}{2}$, l'on a

$$\alpha_{k+1} = \alpha_k, \qquad \beta_{k+1} = \tfrac{1}{2}\beta_k$$

   c'est-à-dire que la population totale est constante, et l'écart à la population d'équilibre est une suite géométrique de raison $\frac{1}{2}$. On voit ici tout l'intérêt d'utiliser les coordonnées dans la base propre pour comprendre l'évolution de la population.

On retrouve le théorème du rang pour une matrice diagonalisable A de la façon suivante. Notons (u_i) une base propre et (λ_i) les valeurs propres correspondantes. Soit d le nombre de λ_i nuls. Quitte à réordonner les (u_i), nous avons $\lambda_i = 0$ pour $i \leqslant d$ et $\lambda_i \neq 0$ pour $i > d$.

proposition 2.62. *Le noyau et l'image de A admettent pour base respectivement $(u_i,\ i \leqslant d)$ et $(u_i,\ i > d)$. En particulier, la dimension du noyau est d, celle de l'image est $n - d$.*

Ici, par convention, une base de $\{0\}$ est la famille vide.

Démonstration. Les deux familles $(u_i,\ i \leqslant d)$ et $(u_i,\ i > d)$ sont extraites de (u_i) qui est

libre, elles sont donc libres toutes les deux. Par ailleurs si $v = \sum \alpha_i u_i$, alors

$$\varphi_A(v) = \sum_{i=1}^{n} \alpha_i \lambda_i u_i = \sum_{i>d} \alpha_i \lambda_i u_i. \tag{2.3}$$

Comme les λ_i restant sont non-nuls, nous avons $\varphi_A(v) = 0$ ssi $\alpha_i = 0$ pour tout $i > d$, ce qui montre que le noyau est engendré par les u_i, $i \leqslant d$. D'autre part, si $i > d$, $u_i = \varphi_A(\lambda_i^{-1} u_i)$ appartient à l'image et d'après (2.3), ces vecteurs engendrent l'image. □

Deuxième partie

Introduction aux probabilités

Chapitre 3

Combinatoire et dénombrement

3.1 Opérations ensemblistes

En mathématiques on n'essaie pas de définir ce qu'est un ensemble, on explique seulement quelles sont les règles pour en fabriquer. On dispose donc, dès le départ, d'une notion d'ensemble et d'appartenance à un ensemble. La notation $x \in E$ se lit "x appartient à E" ou encore "x est un élément de l'ensemble E". On dispose aussi d'un ensemble qui ne contient aucun élément : l'ensemble vide, noté $\emptyset$.

On dit qu'un ensemble A est inclus dans un ensemble E si tout élément de A appartient à E ; on note $A \subset E$. On dit aussi que A est une partie de E, ou un sous-ensemble. Les parties de E forment un ensemble qui se note $\mathcal{P}(E)$.

REMARQUE 3.1.

1. Les éléments d'un ensemble ne sont pas ordonnés : par exemple, l'ensemble $\{2,1\}$ est égal à l'ensemble $\{1,2\}$.
2. L'ensemble $\{1,2,2\}$ est aussi égal à l'ensemble $\{1,2\}$: le fait de répéter un élément dans la définition ne change rien.
3. En probabilité, l'ensemble des résultats d'une expérience aléatoire est appelé *univers*, et ses sous-ensembles sont appelés des *événements*. Par exemple, pour un lancer de dé, l'univers est $\{1,2,3,4,5,6\}$; l'événement "le dé est pair" est égal au sous-ensemble $\{2,4,6\}$.

Soient A, B deux parties d'un ensemble E. L'intersection de A et de B, notée $A \cap B$, est l'ensemble des éléments x de E qui appartiennent à la fois à A et à B :

$$x \in A \cap B \quad \Leftrightarrow \quad x \in A \text{ et } x \in B.$$

L'union de A et de B, notée $A \cup B$, est l'ensemble des éléments x de E qui appartiennent à A ou à B :

$$x \in A \cup B \quad \Leftrightarrow \quad x \in A \text{ ou } x \in B.$$

Attention, le "ou" mathématique est non exclusif : autrement dit, les éléments qui appartiennent à la fois à A et à B font partie de l'union. On peut de même définir, sans difficulté, la réunion ou l'intersection de plus de deux ensembles.

Le complémentaire de A dans E, noté $E \setminus A$, est l'ensemble des éléments de E qui n'appartiennent pas à A :

$$x \in E \setminus A \Leftrightarrow \quad x \notin A.$$

Il existe d'autres notations, comme $\mathcal{C}_E A$, A^c, ${}^c A$, et en probabilité il est souvent noté $\bar{A}$.

On dit que A et B sont disjoints s'ils n'ont aucun élément en commun, autrement dit si $A \cap B = \emptyset$. En probabilité, deux événements disjoints sont aussi appelés *incompatibles* : ils ne peuvent jamais arriver en même temps.

Test :

Que vaut $E \setminus (E \setminus A)$?

PROPOSITION 3.2.

$$E \setminus (A \cap B) = (E \setminus A) \cup (E \setminus B)$$
$$E \setminus (A \cup B) = (E \setminus A) \cap (E \setminus B).$$

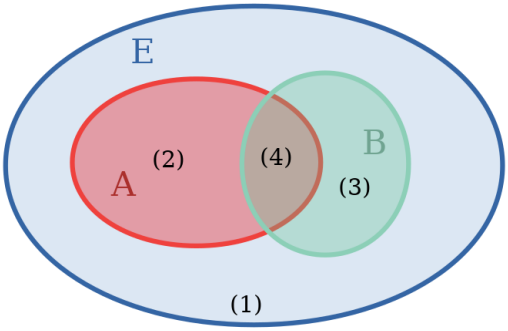


FIGURE 3.1 – On peut classer les points de E en quatre catégories : (1) n'appartenant ni à A ni à B, (2) à A mais pas à B, (3) à B mais pas à A, (4) à A et à B. Ceci correspond à une "partition" (découpage) de E en quatre parties. On obtient les égalités de la proposition en écrivant les différents ensembles comme réunion de certaines de ces quatre parties.

Preuve de la proposition (exercice). Ecrire chacun des ensembles $A, B, E \setminus A, E \setminus B$ comme réunion de certaines des quatre parties décrites dans la figure 3.1. En déduire l'expression de $E \setminus (A \cap B)$, puis de $(E \setminus A) \cup (E \setminus B)$. Conclure. Procéder de même pour montrer la seconde égalité de la proposition ci-dessus. □

En résumé, les opérations ensemblistes d'intersection, d'union et de passage au complémentaire correspondent repsectivement aux opérations logiques ET, OU et NON :

Énoncé indispensable 61 : logique et opérations ensemblistes

$$x \in A \cap B \quad \Leftrightarrow \quad x \in A \text{ et } x \in B$$

$$x \in A \cup B \quad \Leftrightarrow \quad x \in A \text{ ou } x \in B.$$

$$x \in E \setminus A \quad \Leftrightarrow \quad x \notin A.$$

3.2 Cardinal d'un ensemble

On dit qu'un ensemble E est fini s'il a un nombre fini d'éléments, qu'on peut numéroter en écrivant $E = \{x_1, \dots x_n\}$. En supposant qu'on n'a pas numéroté deux fois un même élément ($\forall i \neq j, x_i \neq x_j$), l'entier n est alors le nombre d'éléments de E, appelé le **cardinal de** E et noté $\mathrm{Card}(E)$. On utilise parfois d'autres notations, comme $|E|$ ou $\sharp E$.

Un ensemble qui n'est pas fini est dit *infini*. Un ensemble infini est **dénombrable** si on peut numéroter ses éléments en utilisant tous les nombres entiers positifs :

$$E = \{x_1, x_2, \dots\}.$$

On peut montrer par exemple que l'ensemble $\mathbb{Q}$ des fractions rationnelles est dénombrable, par contre l'ensemble de tous les nombres réels ne l'est pas (pour une preuve, voir l'argument diagonal de Cantor sur Wikipedia).

Quatre propriétés

Soient A, B deux parties d'un ensemble E.

(1) Si $A \subset B$ alors $\mathrm{Card}(A) \leq \mathrm{Card}(B)$.

(2) Si A et B sont **disjointes** alors $\mathrm{Card}(A \cup B) = \mathrm{Card}(A) + \mathrm{Card}(B)$. C'est une formule évidente mais fondamentale pour dénombrer. Plus généralement, si un ensemble E peut s'écrire comme la réunion de parties $A_1, \dots, A_n$ qui sont **deux à deux disjointes** (on dit que $A_1, \dots, A_n$ forme une *partition* de E), alors

$$\mathrm{Card}(E) = \sum_{i=1}^{n} \mathrm{Card}(A_i).$$

(3) $\mathrm{Card}(E \setminus A) = \mathrm{Card}(E) - \mathrm{Card}(A)$. En effet, E est l'union disjointe de A et de $E \setminus A$.

(4) Pour deux parties qui ne sont pas nécessairement disjointes, on a la relation suivante :

Énoncé indispensable 62 :

$$\text{Card}(A \cup B) = \text{Card}(A) + \text{Card}(B) - \text{Card}(A \cap B).$$

Démonstration. Pour démontrer cette formule, on peut à nouveau se référer à la figure 3.1 : $A \cup B$ est l'union des parties (2), (3) et (4), qui sont deux à deux disjointes. En notant $\sharp i$ le nombre d'éléments de la partie (i), on a donc

$$\text{Card}(A \cup B) = \sharp 2 + \sharp 3 + \sharp 4.$$

D'autre part, on voit sur la même figure que $\text{Card}(A) = \sharp 2 + \sharp 4$ et $\text{Card}(B) = \sharp 3 + \sharp 4$, et donc

$$\text{Card}(A \cup B) = \sharp 2 + \sharp 4 + \sharp 3 + \sharp 4.$$

Pour retrouver le cardinal de $A \cup B$ on doit donc retrancher $\sharp 4$, qui n'est rien d'autre que le cardinal de $A \cap B$, d'où la formule. □

3.3 Produit cartésien

Lorsqu'on lance successivement deux dés, le résultat de l'expérience est un couple de nombres, par exemple $(2, 1)$. Plus généralement, lorsqu'on effectue successivement deux expériences modélisées par les ensembles E et F, le résultat est un **couple** (x, y) où x est un élément de E et y un élément de F. L'ensemble de tels couples est appelé **produit cartésien de E et de F** et noté $E \times F$:

$$E \times F = \{(x, y) \mid x \in E \text{ et } y \in F\}.$$

L'ensemble $E \times E$ est aussi noté E^2.

Si E et F sont deux ensembles finis, alors

$$\text{Card}(E \times F) = \text{Card}(E) \times \text{Card}(F).$$

Démonstration. On peut démontrer cette formule de la façon suivante (illustrée dans la Figure 3.2). Ecrivons $E = \{x_1, \ldots, x_n$ avec $n = \text{Card}(E)$. On peut alors classer les éléments (x, y) de $E \times F$ selon la valeur de x, et on obtient

$$E \times F = \{x_1\} \times F \cup \cdots \cup \{x_n\} \times F$$

Or les ensembles de droite sont deux à deux disjoints, et ont le même nombre d'éléments que F : d'après le chapitre précédent, on a

$$\text{Card}(E \times F) = \sum_{i=1}^{n} \text{Card}(\{x_i\} \times F) = \sum_{i=1}^{n} \text{Card}(F) = \text{Card}(E) \times \text{Card}(F).$$

□

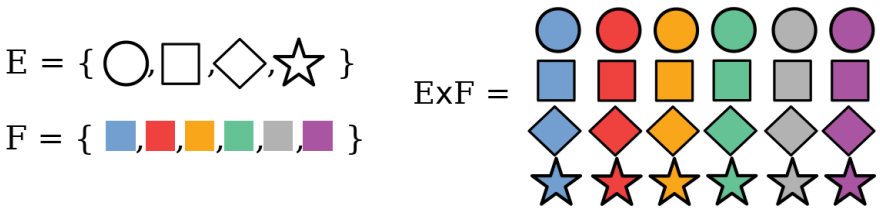


FIGURE 3.2 – Si E est un ensemble de 4 formes et F est un ensemble de 6 couleur, alors les éléments de $E \times F$ sont des couples (forme,couleur), et il y a bien $4 \times 6 = 24$ possibilités.

Plus généralement, on peut définir le produit cartésiens des ensembles $E_1, \ldots, E_n$, et A^n désigne l'ensemble des listes ordonnées $(x_1, \ldots, x_n)$ d'éléments de A, qui sont appelées des n-uplets.

REMARQUE 3.3. Attention, **l'ordre compte** : le couple $(2,1)$ est différent du couple $(1,2)$. Ne pas confondre le couple $(2,1)$ (qui est un élément du produit cartésien $\{1,...,6\} \times \{1,...,6\}$) avec l'ensemble $\{1,2\}$ (qui est un sous-ensemble à deux éléments de $\{1,...,6\}$).

Nous avons vu qu'il y a 36 couples (x,y) avec $x,y \in \{1,...,6\}$. Combien y a-t-il de parties $\{x,y\}$ avec $x,y \in \{1,...,6\}$?... Nous allons explorer ce genre de questions dans le chapitre suivant.

3.4 Dénombrement

Une urne contient 10 boules numérotées de 1 à 10. On en tire trois au hasard. Combien y a-t-il d'issues possibles ?

La réponse est : ça dépend ! Il faut préciser la question. Si on tire les boules les unes après les autres, alors le résultat est modélisé par un triplet de nombres distincts, par exemple $(3,9,5)$, qu'on appelle un *arrangement*. Si on tire les trois boules simultanément, alors le résultat est modélisé par un ensemble à trois éléments, comme $\{3,5,9\}$, qu'on appelle une *combinaison*. Dans ce chapitre nous allons voir comment dénombrer les arrangements et les combinaisons.

Test :

Y a-t-il plus d'arrangements ou de combinaisons à 3 éléments ?

3.4.1 *Arrangements*

Un arrangement à k éléments d'un ensemble E est une liste de k éléments de A, tous distincts. L'ensemble des arrangements à k élements est donc un sous-ensemble de E^k.

EXEMPLE 3.4. Si E est l'ensemble des entiers pairs compris entre 0 et 22, alors $(2,6,10)$ est un arrangement à 3 éléments de E, mais pas $(2,10,10)$ car 10 est répété. D'autre part, $(10,2,6)$ est aussi un arrangement de E, et il est distinct de $(2,6,10)$ car l'ordre compte.

Test :

Faire la liste des arrangements à 2 éléments de l'ensemble $\{1,2,3\}$. Combien y en a-t-il ? Utiliser cet exemple pour tester la formule suivante.

Énoncé indispensable 63 :

Soient k, n deux entiers tels que $1 \leq k \leq n$. Le nombre d'arrangements à k éléments d'un ensemble E à n éléments est

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

où $n! = 1 \times 2 \times \cdots \times n$ et se lit "factorielle n".

Démonstration. On donne deux preuves.

Première preuve On fabrique tous les arrangements de la façon suivante. On commence par choisir le premier terme de la liste, pour lequel on a n choix possibles. Puis on choisit le second, qui doit être différent du premier, pour lequel on a donc $(n-1)$ choix possibles. On a $(n-2)$ choix pour le troisième, et ainsi de suite jusqu'au k-ème et dernier terme pour lequel il reste $(n-k+1)$ choix. Le nombre total de choix correspond donc au nombre de feuilles d'un arbre comme décrit dans la Figure 3.3 :

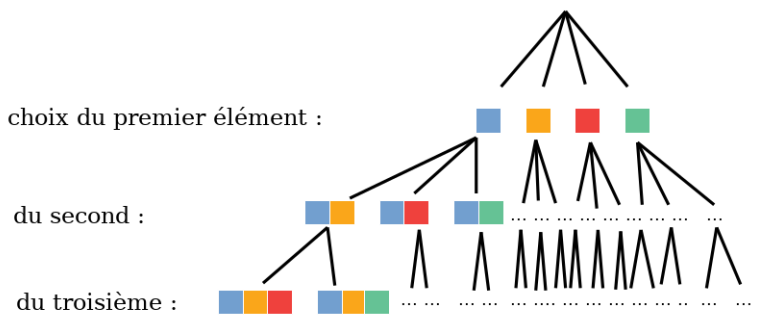


Figure 3.3 – on doit choisir un arrangement de 3 couleurs parmi 4. Pour la première couleur, on a 4 choix. Pour la seconde, puisqu'on n'a pas le droit de reprendre la même couleur, il ne reste plus que 3 choix. La troisième couleur est alors choisie dans les 2 possibilités restantes. Sur la dernière ligne, on compte donc $4 \times 3 \times 2$ feuilles, qui correspondent chacune à un arrangement.

Deuxième preuve On procède par récurrence sur l'entier n. On veut montrer la propriété suivante, pour tout entier $n \geq 1$:
(P_n) Pour tout entier k tel que $1 \leq k \leq n$, on a $A_n^k = \frac{n!}{(n-k)!}$.

Pour $n = 1$, la propriété est évidente : il y a un unique arrangement à 1 élément d'un ensemble à un élément.

Soit n un entier pour lequel la propriété (P_n) est vraie, autrement dit pour lequel la formule est vraie *pour tout entier* $k \in \{1, \dots, n\}$. Montrons la propriété (P_{n+1}). On considère donc un entier $k \in \{1, \dots, n+1\}$. Soit $E = \{x_1, \dots, x_n + 1\}$ un ensemble à $n+1$ éléments. L'ensemble $\mathcal{A}$ des arrangements à dénombrer s'écrit comme l'union des ensembles $\mathcal{A}_i$ des arrangements qui commencent par i, et ces ensembles sont deux à deux disjoints. Les éléments de $\mathcal{A}_i$ s'identifient à des arrangements à $(k-1)$ éléments de l'ensemble $E \setminus \{x_i\}$, qui a n éléments : on a donc $\mathrm{Card}(\mathcal{A}_i) = A_n^{k-1}$. En utilisant l'hypothèse de récurrence, on obtient

$$\mathrm{Card}(\mathcal{A}) = \sum_{i=1}^{n+1} \mathrm{Card}(\mathcal{A}_i) = (n+1)A_n^{k-1} = (n+1)\frac{n!}{(n-(k-1))!} = \frac{(n+1)!}{((n+1)-k)!}$$

ce qu'on voulait. □

Par exemple, $A_3^2 = \frac{3\times 2\times 1}{1} = 6$ est le nombre d'arrangements à 2 éléments d'un ensemble à 3 éléments et donne la réponse au test précédent.

Cas particulier ($k = n$). Un arrangement à n élément est une liste de *tous* les éléments de E, qu'on appelle *permutation* des éléments de E. D'après la formule précédente, il y a donc $n!$ permutations d'éléments de E.

3.4.2 *Combinaisons*

Une combinaison à k élements d'un ensemble E est une liste **non-ordonnée** de k éléments distincts de E, autrement dit un sous-ensemble de E de cardinal égal à k. L'ensemble des combinaisons à k élements est donc une partie de $\mathcal{P}(E)$ qu'on note $\mathcal{P}_k(E)$.

exemple 3.5. À la belote, qui se joue avec un jeu de 32 cartes, les joueurs reçoivent chacun 8 cartes qui constituent ce qu'on appelle leur main. Ils peuvent regarder toutes leurs cartes, et les jouer dans l'ordre qu'ils souhaitent. L'ordre dans lequel la main a été distribuée n'a donc aucune incidence. Une main est donc naturellement représentée par une combinaison de 8 cartes parmi 32, par exemple $\{1\clubsuit, 9\heartsuit, 10\heartsuit, V\spadesuit, 10\spadesuit, 1\diamondsuit, 9\clubsuit, 7\heartsuit\}$.

Énoncé indispensable 64 :

Soient k, n deux entiers tels que $0 \leq k \leq n$. Le nombre de combinaisons à k éléments d'un ensemble E à n éléments est

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

Démonstration. Nous allons démontrer cette formule à partir de la formule pour les arrangements (on verra plus bas une autre preuve basée sur le "triangle de Pascal"). Le principe de la preuve consiste à se demander combien d'arrangements (liste ordonnée) on peut fabriquer avec une combinaison (liste non ordonnée) donnée.

Plus précisément, soit E un ensemble à n éléments, et k un entier entre 1 et n. Notons $\mathcal{C}$ l'ensemble des combinaisons de k éléments de E, qu'on veut dénombrer. Soit $c = \{x_1, \ldots, x_k\}$ une combinaison. Nous avons vu que les arrangements de k éléments de c s'appellent des permutations de c, et il y en a $A_k^k = k!$. Chacun de ces arrangements est aussi un arrangement de E à k éléments. Inversement, à partir d'un arrangement de E on obtient une combinaison en "oubliant l'ordre". En classant les arrangements d'après la combinaison sous-jacente, on obtient

$$A_n^k = \sum_{c \in \mathcal{C}} A_k^k = \text{Card}(\mathcal{C}) \times A_k^k$$

d'où, en tenant compte de la formule pour les arrangements ;

$$\text{Card}(\mathcal{C}) = \frac{1}{k!} \frac{n!}{(n-k)!}$$

qui donne le résultat attendu. □

Par exemple, $\binom{3}{2} = \frac{3\times 2\times 1}{2\times 1} = 3$ est le nombre de combinaisons à 2 éléments d'un ensemble à 3 éléments. En effet, l'ensemble $\{1,2,3\}$ a bien trois sous-ensembles à deux éléments qui sont $\{1,2\}, \{1,3\}, \{2,3\}$.

3.4.3 *Calculs avec les combinaisons*

Le triangle de Pascal est une façon pratique de calculer les nombres $\binom{n}{k}$ pour les petites valeurs de n et k :

$$\begin{array}{ccccccccccccc}
 & & & & & & 1 & & & & & & \\
 & & & & & 1 & & 1 & & & & & \\
 & & & & 1 & & 2 & & 1 & & & & \\
 & & & 1 & & 3 & & 3 & & 1 & & & \\
 & & 1 & & 4 & & 6 & & 4 & & 1 & & \\
 & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & \\
1 & & & & & & \cdots & & & & & & 1
\end{array}$$

On le construit à l'aide des deux règles suivantes : (1) Chaque ligne commence et finit par un '1' ; (2) chaque terme est la somme des deux termes situés immédiatement au-dessus.

Test :

Complétez la dernière ligne du triangle à l'aide de ces deux règles.

Nous allons voir que la n ème ligne du triangle de Pascal est la suite $\binom{n}{0}, \binom{n}{1}, \ldots, \binom{n}{n}$. La clé de cette propriété est la formule suivante.

PROPOSITION 3.6. *Soient k, n deux entiers tels que $0 \leq k \leq n$. On a*

$$\binom{n+1}{k+1} = \binom{n}{k+1} + \binom{n}{k}.$$

Démonstration. Pour démontrer cette formule, on considère un ensemble E à $n+1$ éléments, on choisit l'un des éléments de E, notons-le a, et on classe les combinaisons de $k+1$ éléments de E en deux catégories, selon qu'elle contiennent ou non notre élément distingué a. On peut ainsi écrire

$$\binom{n+1}{k+1} = N + M$$

où N est le nombre de combinaisons qui ne contiennent pas a, et M le nombre de celles qui le contiennent. Les combinaisons qui ne contiennent pas a sont exactement les combinaisons de l'ensemble $E' = E \setminus \{a\}$, cet ensemble a n éléments, et on a donc $N = \binom{n}{k+1}$. Les combinaisons qui contiennent a, elles, sont exactement les réunions de $\{a\}$ avec une combinaisons à k éléments de l'ensemble E' : on a donc $M = \binom{n}{k}$. Ce qui montre la formule.

De façon condensée, en utilisant la notation pour l'ensemble des parties de cardinal fixé, on peut résumer la preuve par la formule suivante :

$$P_{k+1}(A \cup \{a\}) = \mathcal{P}_{k+1}(A) \cup \{X \cup \{a\} \mid X \in \mathcal{P}_k(A)\}$$

□

Nous pouvons alors généraliser la formule bien connue $(x+y)^2 = x^2 + 2xy + y^2$:

Énoncé indispensable 65 : Binôme de Newton

Soit n un entier positif et x, y deux nombres réels ou complexes. On a :

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

Démonstration. Nous donnons deux preuves, l'une directe et la seconde par récurrence.
(1) On peut démontrer la formule par un dénombrement direct. Par simplicité, nous expliquons seulement le cas $n = 5$. On a bien sûr

$$(x+y)^5 = (x+y)(x+y)(x+y)(x+y)(x+y)$$

et il s'agit de développer ce produit. Il apparait alors une somme dont chacun des termes est obtenu de la façon suivante : dans chacune des cinq parenthèses on choisit ou bien 'x', ou bien 'y', et ensuite on multiplie tous ces choix. Selon le nombre de 'x' et de 'y' choisis, on obtient ainsi l'un terme suivant : $y^5, xy^4, x^2y^3, x^3y^2, x^4y, x^5$. Combien de fois obtient-on un terme donné, par exemple x^2y^3 ? Ce terme apparait à chaque fois qu'on a choisi 'x' dans deux des cinq parenthèses (et 'y' dans les trois autres) : dans notre somme, il y a donc autant de fois le terme 'x^2y^3' qu'il y a de façons de choisir deux des cinq parenthèses, et ce nombre est $\binom{5}{2}$. C'est bien le coefficient attendu.
(2) Démonstration par récurrence. [1] La formule est évidente pour $n = 0$ ou 1. Soit $n \geq 1$ un entier pour lequel la formule est vraie, démontrons alors la formule au rang $n+1$.

1. Si vous n'êtes pas à l'aise avec le symbole de sommation $\sum$, lire le début de la section 3.5.

On écrit bien sûr

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

en utilisant l'hypothèse de récurrence. Puis on développe ce produit en distribuant x et y dans la somme. On obtient deux sommes, la première est obtenue en augmentant de '1' la puissance de x, il s'agit de

$$\sum_{k=0}^{n}\binom{n}{k}x^{k+1}y^{n-k} = \sum_{k'=1}^{n+1}\binom{n}{k'-1}x^{k'}y^{n-(k'-1)} = \sum_{k=1}^{n+1}\binom{n}{k-1}x^k y^{n+1-k}$$

qu'on a ré-écrit avec le changement d'indice $k' = k+1$ (voir la section 3.5). La seconde est

$$\sum_{k=0}^{n}\binom{n}{k}x^k y^{n+1-k}.$$

En additionnant ces deux sommes on obtient

$$(x+y)^{n+1} = \sum_{k=0}^{n+1}\left(\binom{n}{k-1} + \binom{n}{k}\right)x^k y^{n+1-k}$$

(il y a un petit ennui avec $k=0$ et $k=n+1$, pour que ça marche il suffit de décider que $\binom{n}{-1} = \binom{n}{n+1} = 0$). On applique maintenant la formule du triangle de Pascal, sous la forme

$$\binom{n}{k-1} + \binom{n}{k} = \binom{n+1}{k}$$

(en vérifiant à nouveau que ça marche bien pour $k=0$ et $k=n+1$) et on obtient le résultat voulu :

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

□

corollaire 3.7. *Le nombre de sous-ensembles d'un ensemble E à n est éléments est 2^n :*

$$\mathrm{Card}(\mathcal{P}(\{1,\ldots,n\})) = 2^n.$$

Démonstration. Soit E un ensemble à n éléments. Pour calculer le nombre de sous-ensembles de E, on additionne le nombre de sous-ensembles à 0 éléments, à 1 éléments, etc. jusqu'à n :

$$\mathrm{Card}(\mathcal{P}(E)) = \sum_{k=0}^{n}\mathrm{Card}(\mathcal{P}_k(E)) = \sum_{k=0}^{n}\binom{n}{k}.$$

D'un autre côté, la formule du binôme de Newton pour $x=y=1$ s'écrit

$$2^n = (1+1)^n = \sum_{k=0}^{n}\binom{n}{k}1^k 1^{n-k} = \sum_{k=0}^{n}\binom{n}{k}.$$

Et on trouve bien le résultat voulu. □

Exercice. Soit n un entier, et L_n l'ensemble des listes de 0 et de 1 de longueur 4 (par

exemple, $(0,0,1,0)$ est un élément de L_4). Dessiner les éléments de L_n sur un arbre. En déduire le cardinal de L_n. Construire une bijection entre L_n et l'ensemble $\mathcal{P}(\{1,\dots,n\})$. En déduire une nouvelle preuve de la formule Card$(\mathcal{P}(\{1,\dots,n\}) = 2^n$..

Éléments historiques 1 :

Le triangle de Pascal doit son nom au mathématicien et philosophe Blaise Pascal (1623-1662). On en trouve cependant des traces antérieures, au XI^e siècle en Chine ou au X^e siècle en Perse. D'ailleurs, il est connu en Chine sous le nom de triangle de Yang Hui, et en Italie de triangle de Tartaglia. La formule de récurrence des coefficients binomiaux, elle, était semble-t-il déjà connue en Inde au II^e siècle avant notre ère. En 1654, Pascal publie son *Traité sur le triangle arithmétique* où il recense (et surtout, démontre) de nombreuses propriétés de ce triangle, qu'il utilise ensuite pour traiter des problèmes de probabilités.

3.5 Sommes et séries

Une série est une "somme d'une infinité de termes", à laquelle on cherche à donner un sens. Dans ce chapitre nous introduisons les séries les plus simples, où les termes sont des nombres positifs ; nous en aurons besoin pour étudier les variables aléatoires qui peuvent prendre une infinité de valeurs.

3.5.1 *Symbole de sommation, changement d'indice*

Dans ce chapitre nous utiliserons intensivement le symbole de sommation

$$\sum_{n=0}^{N} a_n := a_0 + \cdots + a_N.$$

Les deux variables 'n' et 'N' qui y apparaissent ont des statuts très différents : la variable 'N' est ne variable "globale" désigne un nombre fixé, qui doit avoir été introduite auparavant. La variable 'n', par contre, ne sert qu'à l'intérieur de la somme, c'est ce qu'on appelle une variable "locale" ou "muette" ; on la choisit parmi les lettres qui n'ont pas déjà été utilisées pour des variables globales du même discours.

Test :

Parmi les quatre expressions suivantes, identifier celles qui désignent la même quantité, et celles qui n'ont pas de sens.

$$\sum_{n=0}^{N} a_n \qquad \sum_{n=0}^{N} a_k \qquad \sum_{k=0}^{N} a_k \qquad \sum_{n=0}^{n} a_n \qquad \sum_{n=p}^{q} a_n.$$

Dans chacune des sommes, dire pour chacune des variables si elle est locale ou globale.

Test :

Écrire sans le symbole de sommation

$$\sum_{k=0}^{3} a_k \qquad \sum_{k=0}^{3} a_{2k} \qquad \sum_{k=0}^{3} a_{k+1}.$$

Quelles bornes doit-on mettre dans la somme

$$\sum_{k=\cdots}^{\cdots} a_{k+1}$$

pour retrouver la première somme ?

On a souvent besoin de changer d'indice dans une somme (voir par exemple la démonstration de la formule du binôme de Newton), par exemple d'exprimer une somme du type

$$\sum_{n=p}^{q} a_{n+1}$$

comme une somme dans laquelle l'indice est 'n' et non pas '$n+1$'. Pour faire ça sans se tromper, il est recommandé de commencer par changer de nom à la variable, en posant par exemple $k = n+1$; lorsque la variable n varie entre p et q, la variable k varie entre $p+1$ et $q+1$, et la somme devient donc

$$\sum_{n=p}^{q} a_{n+1} = \sum_{k=p+1}^{q+1} a_k = \sum_{n=p+1}^{q+1} a_n.$$

Le procédé est tout à fait analogue au changement de variables dans les intégrales.

3.5.2 *Somme sur des ensembles plus généraux*

Dans les sommes précédentes, l'indice de sommation n parcourait un intervalle d'entiers. Il peut s'avérer pratique de sommer sur un ensemble I d'indices qui n'est pas forcément de cette forme. Par exemple, mettons qu'on veuille désigner la somme de tous les entiers entre 0 et 1000 qui ne sont pas des multiplies de 7. On peut noter $I = \{n \in \mathbb{N} \mid 1 \leq n \leq 1000,\ n \neq 0[7]\}$ l'ensemble de ces entiers. Leur somme s'écrit alors

$$\sum_{n \in I} n\,.$$

Alternativement, dans certains cas simples, on peut définir l'ensemble de sommation directement sous le signe somme, par exemple

$$\sum_{1 \leq n < 10,\ n \text{ pair}} \sqrt{n} = \sqrt{2} + \sqrt{4} + \sqrt{6} + \sqrt{8}$$

ou, en notant $f(n)$ le nombre de 3 que contient l'écriture décimale de l'entier n,

$$\sum_{1\leq n\leq 35,\ f(n)=1} \frac{1}{n} = \frac{1}{3}+\frac{1}{13}+\frac{1}{23}+\frac{1}{30}+\frac{1}{31}+\frac{1}{32}+\frac{1}{34}+\frac{1}{35}.$$

3.5.3 *Série à termes positifs*

Une série à termes positifs est une écriture du type

$$\sum_{n=0}^{+\infty} a_n$$

où les a_n sont des nombres réels positifs ou nuls. Par exemple, si r et x sont des nombres positifs ou nuls,

$$(1)\ \sum_{n=0}^{+\infty} n, \qquad (2)\ \sum_{n=1}^{+\infty} \frac{1}{n} \qquad (3)\ \sum_{n=0}^{+\infty} r^n, \qquad (4)\ \sum_{n=0}^{+\infty} \frac{1}{n!}x^n$$

sont des séries à termes positifs.

On aimerait attribuer une valeur numérique à ces séries, comment faire ? L'idée consiste à considérer la façon dont la somme grandit au fur et à mesure qu'on ajoute les termes. Plus précisément, on considère la suite des *sommes partielles*

$$\begin{aligned} S_0 &= a_0, \\ S_1 &= a_0+a_1, \\ S_2 &= a_0+a_1+a_2, \\ &\dots \\ S_N &= \sum_{n=0}^{N} a_n, \\ &\dots \end{aligned}$$

C'est une suite de nombres positifs qui est croissante.

DEFINITION 3.8. On dit qu'une suite réelle $(u_n)_{n\in\mathbb{N}}$ ...

1. est convergente s'il existe $u \in \mathbb{R}$ (appelé la limite de la suite) tel que pour tout $\varepsilon > 0$ il existe $n_0 \in \mathbb{N}$ tel que pour tout $n \geqslant n_0$, $|u_n - u| \leqslant \varepsilon$ (autrement dit, pour tout niveau de précision $\varepsilon > 0$, au bout d'un certain rang n_0, tous les termes de la suite sont à distance au plus ε de la limite). De façon équivalente, on dit aussi que u_n tend vers u quand n tend vers $+\infty$ (noté : $u_n \to u$ quand $n \to +\infty$).
2. tend vers $+\infty$ quand n tend vers $+\infty$ si pour tout $M \in \mathbb{R}$ il existe $n_0 \in \mathbb{N}$ tel que pour tout $n \geqslant n_0$, $u_n > M$ (autrement dit, quelque soit le seuil M, à partir d'une certain rang n_0 les termes de la suite sont tous au-dessus de M).
3. est bornée s'il existe $M > 0$ tel que $|u_n| \leqslant M$ pour tout $n \in \mathbb{N}$.

PROPOSITION 3.9. *Soit $(u_n)_{n\in\mathbb{N}}$ une suite réelle* croissante. *Alors, de deux choses l'une : soit $(u_n)_{n\in\mathbb{N}}$ est convergente, soit $u_n \to +\infty$ quand $n \to +\infty$.*

REMARQUE 3.10. Le résultat est évidemment faux si on ne suppose pas la suite croissante. Par exemple $u_n = (-1)^n$ n'est ni convergente ni tendant vers $+\infty$. De même, une suite

de termes positifs peut ne pas être bornée sans pour autant tendre vers $+\infty$, par exemple la suite définie par $u_n = n$ pour n pair et $u_n = 1$ pour n impair.

Démonstration. Si la suite n'est pas bornée, par définition cela signifie que pour tout $M > 0$, il existe $n_0 \in \mathbb{N}$ tel que $u_{n_0} > M$. Puisque d'autre part la suite est croissante, alors pour tout $n \geqslant n_0$, on a $u_n > M$. Autrement dit, pour tout $M > 0$, il existe $n_0 \in \mathbb{N}$ tel que pour tout $n \geqslant n_0$, $u_n > M$, ce qui est exactement la définition de $u_n \to +\infty$ pour $n \to +\infty$.

Inversement, supposons la suite bornée. L'ensemble de ses majorants (c'est-à-dire des réels M tels que $M > u_n$ pour tout $n \in \mathbb{N}$) est non vide et admet donc un plus petit élément m (qui est par définition $\sup\{u_n,\ n \in \mathbb{N}\}$). En particulier, $m \geqslant u_n$ pour tout $n \in \mathbb{N}$. Montrons que la suite converge vers m. On va le montrer par l'absurde : supposons que la suite ne converge pas vers m. Cela signifie qu'il existe $\varepsilon > 0$ tel que pour tout $n_0 \in \mathbb{N}$, il existe $n_1 \geqslant n_0$ tel que $|u_{n_1} - m| \geqslant \varepsilon$, et donc forcément $u_{n-1} \leqslant m - \varepsilon$ (puisque m est un majorant) et, puisque la suite est croissante et $n_1 \geqslant n_0$, $u_{n_0} \leqslant m - \varepsilon$. Autrement dit il existe $\varepsilon > 0$ tel que $u_{n_0} \leqslant m - \varepsilon$ pour tout $n_0 \in \mathbb{N}$, autrement dit tel que $m - \varepsilon$ est un majorant de la suite. C'est en contradiction avec le fait que m est, par définition, le plus petit des majorants. Donc l'hypothèse que la suite ne converge pas vers m est fausse. Autrement dit, la suite converge vers m. □

Revenons aux séries. La suite $(S_N)_{N\in\mathbb{N}}$ étant croissante, d'après ce qu'on vient de démontrer il y a donc deux possibilités. (1) Ou bien cette suite n'est pas majorée, et dans ce cas elle diverge et on renonce à attribuer une valeur numérique à la série (on peut écrire $\sum_{n=0}^{+\infty} a_n = +\infty$). (2) Ou bien cette suite converge, et on définit alors sa valeur comme la limite,

$$\sum_{n=0}^{+\infty} a_n := \lim_{N\to+\infty} \sum_{n=0}^{N} a_n.$$

Les séries données en exemple sont-elles convergentes et si oui, que valent-elles ?

(1) Pour la première, les sommes partielles valent

$$\sum_{n=0}^{N} n = \frac{N(N+1)}{2}$$

qui diverge : cette série est divergente. Plus généralement, une série dont les termes ne tendent pas vers 0 n'est jamais convergente.

(2) Pour la seconde, les termes tendent vers 0. Cependant on peut remarquer que

$$\frac{1}{3} + \frac{1}{4} > \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

$$\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} > 4 \times \frac{1}{8} = \frac{1}{2}$$

et donc

$$\sum_{n=1}^{8} \frac{1}{n} > 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} = 1 + \frac{3}{2}.$$

Le même type de minoration donnera

$$\sum_{n=1}^{16} \frac{1}{n} > 1 + \frac{4}{2}, \quad \sum_{n=1}^{32} \frac{1}{n} > 1 + \frac{5}{2}, \quad \sum_{n=1}^{64} \frac{1}{n} > 1 + \frac{6}{2}, \quad \text{etc.}$$

Pour tout entier k on peut trouver une somme partielle qui est supérieure à $1 + \frac{k}{2}$, ce qui permet de voir que la suite des sommes partielles n'est pas majorée : cette série est divergente.

(3) La série $\sum_{n=0}^{+\infty} r^n$ est appelée série géométrique. Lorsque $r \geq 1$, ses termes ne tendent pas vers 0, et la série diverge. Lorsque $r \in [0, 1[$, on a la formule explicite

$$\sum_{n=0}^{N} r^n = \frac{1 - r^{N+1}}{1 - r}$$

Puisque r est positif mais strictement plus petit que 1, la quantité r^{N+1} tend vers 0, la série est donc convergente et sa somme vaut

$$\sum_{n=0}^{+\infty} r^n = \frac{1}{1 - r}.$$

(4) Notre dernière série s'appelle série exponentielle. Elle est également convergente. Pour le voir, on considère un rang n_0 supérieur à $2x$, et on remarque qu'à partir de ce rang, chaque terme est inférieur à la moitié du terme précédent :

$$n \geq n_0 \Rightarrow \frac{1}{(n+1)!} x^{n+1} = \frac{1}{n!} x^n \times \frac{x}{n} \leq \frac{1}{n!} x^n \times \frac{1}{2}.$$

Notons alors $C = \frac{1}{n_0!} x^{n_0}$, et fixons un entier N plus grand que n_0. D'après ce qui précède, la somme des termes de n_0 à N est majorée par

$$C + \frac{1}{2}C + \frac{1}{2^2}C + \cdots + \frac{1}{2^{N-n_0}}C \leq C \sum_{n=0}^{+\infty} \frac{1}{2^n} = 2C.$$

On en déduit que notre somme est majorée (par le nombre $2C$ additionné de la somme des termes de 0 à n_0). Ceci prouve la convergence de la série.

Nous admettons le résultat remarquable suivant : la somme de cette série est égale à la fonction exponentielle.

Énoncé indispensable 66 :

Pour tout $x \geq 0$,

$$\sum_{n=0}^{+\infty} \frac{1}{n!} x^n = e^x.$$

Ce résultat est démontré dans le cours sur les séries entières (en deuxième année). On peut le comprendre, si on s'autorise à "dériver termes à termes" notre série (ce qu'il

faudrait justifier). En effet on calcule alors la dérivée de la fonction Φ qui associe à x la valeur de la série,

$$\Phi'(x) = \sum_{n=0}^{+\infty} \frac{1}{n!} n x^{n-1} = \sum_{n=1}^{+\infty} \frac{1}{(n-1)!} x^{n-1} = \sum_{k=0}^{+\infty} \frac{1}{k!} x^k = \Phi(x)$$

(on a effectué le changement d'indice $k = n + 1$). Puisque $\Phi(0) = 1$, on retrouve bien la caractérisation de l'exponentielle comme unique fonction égale à sa dérivée qui vaut 1 en 0.

Finalement, énonçons sans démonstration le résultat suivant :

proposition 3.11. *Soit* $(a_{n,k})_{n\in\mathbb{N},k\in\mathbb{N}}$ *un ensemble de réels* positifs *indexés par deux entiers. Alors*

$$\sum_{k=0}^{+\infty} \left(\sum_{n=0}^{+\infty} a_{n,k} \right) = \sum_{n=0}^{+\infty} \left(\sum_{k=0}^{+\infty} a_{n,k} \right) .$$

Autrement dit, pour calculer la somme des $a_{n,k}$, on peut d'abord sommer sur les k, puis sur les n, ou l'inverse. C'est clair pour des sommes finies mais, sans rentrer dans les détails ici, notons que c'est faux si l'on veut sommer une infinité de termes qui ne sont pas positifs (même si les deux double-sommes existent comme limite de sommes finies, elles ne sont pas nécessairement égales).

Remarquons que la proposition ne précise pas que les sommes doivent êtres finies, autrement dit il est possible que $\sum_{k=0}^{+\infty} \left(\sum_{n=0}^{+\infty} a_{n,k} \right)$ soit égal à $+\infty$, mais alors c'est également le cas de $\sum_{n=0}^{+\infty} \left(\sum_{k=0}^{+\infty} a_{n,k} \right)$. Dès que l'une des deux expression est finie, alors l'autre l'est également et elles sont égales.

Chapitre 4

Espace probabilisé

4.1 Mesures de probabilité

Les phénomènes aléatoires couvrent un champ très vaste, allant des résultats d'un tirage de pile ou face à la pluviométrie mensuelle en région Bretagne. À un phénomène donné correspondent une famille d'états (par exemple pile ou face pour le lancer d'une pièce, les entiers de 1 à 6 pour le tirage d'un dé). Une **mesure de probabilité** indique la chance qu'un ou plusieurs états se produise (par exemple $1/2$ pour pile, $1/6$ pour tirer le nombre 3).
Quel que soit le contexte, on note Ω l'ensemble dit des "états possibles", qu'on appelle l'univers. Il est plus simple et naturel de commencer par définir les probabilités lorsque Ω est un ensemble fini.

Énoncé indispensable 67 : Mesure de probabilité

definition 4.1. Pour Ω un ensemble fini d'états possibles, on appelle **mesure de probabilité** (ou loi de probabilité) sur Ω une application définie sur les sous ensembles de Ω, à valeurs dans $[0,1]$ telle que

1. $\mathbb{P}(\Omega) = 1$,

2. Additivité : pour (A_1, A_2) sous ensembles disjoints ($A_1 \cap A_2 = \emptyset$) de Ω,

$$\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2).$$

L'ensemble $(\Omega, \mathbb{P})$ est appelé **espace probabilisé**.

exemple 4.2. On prend un dé équilibré, et on considère le résultat après un ou deux lancers.
Un lancer. Après un lancer, l'ensemble des états est l'ensemble des nombres qu'on peut obtenir, donc $\{1,2,3,4,5,6\}$. On a une chance sur six de tirer chacun de ces nombres, la mesure de probabilité modélisant cela est

$$\forall\, 1 \leq i \leq 6, \qquad \mathbb{P}(\{i\}) = 1/6, \qquad \text{plus généralement } \mathbb{P}(A) = \text{card}(A)/6.$$

Deux lancers. On note les deux résultats, dans l'ordre, associés à ces deux lancers. L'ensemble des états possibles est alors $\{1, \cdots, 6\} \times \{1, \cdots, 6\}$. Il y a $6 \times 6 = 36$ résultats possibles, et la mesure de probabilité qui décrit le fait que les couples ont tous les mêmes chances d'être tirés est alors

$$\forall\, 1 \leq i, j \leq 6, \qquad \mathbb{P}(\{(i,j)\}) = 1/36.$$

remarque 4.3. Rappelons du vocabulaire introduit au chapitre précédent : les parties de l'univers sont appelés des événements, et deux événements disjoints sont dits incompatibles. Par exemple, pour un lancer de dé, les événements "le résultat est pair" et "le résultat est impair" sont incompatibles, car $\{2,4,6\} \cap \{1,3,5\} = \emptyset$. Il n'est effectivement pas possible de lancer un dé et que le résultat soit simultanément pair et impair.

proposition 4.4. *Toute mesure de probabilité satisfait*

1. $\mathbb{P}(\emptyset) = 0$,
2. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$,
3. *Pour* $(A_n)_{1 \leq n \leq N} \subset \Omega$, *avec* $A_n \cap A_m = \emptyset$ *pour tous* $n \neq m$,

$$\mathbb{P}\Big(\bigcup_{1 \leq n \leq N} A_n\Big) = \sum_{n=1}^{N} \mathbb{P}(A_n).$$

4. *Pour* (A, B) *quelconques*, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
5. *Si* $A \subset B$, $\mathbb{P}(A) \leq \mathbb{P}(B)$.

remarque 4.5. La propriété 4) peut se voir comme une généralisation de la formule $\mathrm{card}(A \cup B) = \mathrm{card}(A) + \mathrm{card}(B) - \mathrm{card}(A \cap B)$.

Démonstration.

1. La propriété d'additivité appliquée à $A_1 = A_2 = \emptyset$ (qui satisfait bien $A_1 \cap A_2 = \emptyset$) donne $\mathbb{P}(\emptyset) = 2\mathbb{P}(\emptyset)$, et donc $\mathbb{P}(\emptyset) = 0$.
2. Pour un A donné, en appliquant la propriété d'additivité pour $A_1 = A$ et $A_2 = A^c$ (à nouveau, $A \cap A^c = \emptyset$),

$$\mathbb{P}(A) + \mathbb{P}(A^c) \;=\; \mathbb{P}(A \cup A^c) \;=\; \mathbb{P}(\Omega) \;=\; 1\,.$$

3. Pour $N = 2$, c'est exactement l'additivité donc il n'y a rien à démontrer. Supposons par récurrence que le résultat est vrai pour un certain $N \geqslant 2$, et considérons une famille de $N+1$ événements $(A_i)_{1 \leqslant n \leqslant N+1}$ avec $A_n \cap A_m = \emptyset$ pour $n \neq m$. Posons $B_1 = \bigcup_{1 \leq n \leq N} A_n$ et $B_2 = A_{N+1}$. En appliquant la propriété d'attivité à B_1 et B_2 (on a bien $B_1 \cap B_2 = \emptyset$) puis l'hypothèse de récurrence,

$$\mathbb{P}\left(\bigcup_{n=1}^{N+1} A_n\right) \;=\; \mathbb{P}(B_1 \cup B_2) \;=\; \mathbb{P}(B_1) + \mathbb{P}(B_2) \;=\; \sum_{n=1}^{N} \mathbb{P}(A_n) + \mathbb{P}(A_{N+1}) \;=\; \sum_{n=1}^{N+1} \mathbb{P}(A_n)\,.$$

4. Une partition de $A \cup B$ est donnée par $C_1 = A \cap B^c$, $C_2 = A^c \cap B$ et $C_3 = A \cap B$ (on distingue trois cas : les éléments de A uniquement, de B uniquement, et simultanément de A et B). Par la proposition précédente,

$$\mathbb{P}(A \cup B) \;=\; \mathbb{P}(C_1 \cup C_2 \cup C_3) \;=\; \mathbb{P}(A \cap B^c) + \mathbb{P}(A^c \cap B) + \mathbb{P}(A \cap B)\,.$$

De même, une partition de A est donnée par $D_1 = A \cap B$ et $D_2 = A \cap B^c$, ce qui donne $\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c)$, et le même raisonnement pour B donne $\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B)$, et donc

$$\begin{aligned} \mathbb{P}(A) + \mathbb{P}(B) &= 2\mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c) + \mathbb{P}(A^c \cap B) \\ &= \mathbb{P}(A \cup B) + \mathbb{P}(A \cap B) \end{aligned}$$

ce qui est bien la relation désirée.

5. Si $A \subset B$ alors A et $B \cap A^c$ forment une partition de B, donc $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \cap A^c)$, et le résultat découle du fait que $\mathbb{P}(B \cap A^c) \geqslant 0$.

□

La définition précédente peut en fait être étendue au cas où l'univers est infini dénombrable :

DEFINITION 4.6. Étant donné Ω un ensemble d'états possibles dénombrable, on appelle **mesure de probabilité** sur Ω une fonction définie sur les sous ensembles de Ω, à valeurs dans $[0, 1]$ telle que

1. $\mathbb{P}(\Omega) = 1$,

2. Sigma-additivité : pour $(A_n)_{n\geq 0} \subset \Omega$ avec $A_n \cap A_m = \emptyset$ si $n \neq m$,

$$\mathbb{P}(\bigcup_{n\geq 0} A_n) = \sum_{n\geq 0} \mathbb{P}(A_n).$$

L'ensemble $(\Omega, \mathbb{P})$ est appelé **espace probabilisé**.

REMARQUE 4.7. Précisons le sens de la notation $\sum_{n\geq 0} \mathbb{P}(A_n)$: si on somme un nombre fini de termes c'est clair, si on en somme une infinité c'est à comprendre au sens de

$$\sum_{n\geq 0} \mathbb{P}(A_n) = \lim_{N\to\infty} \sum_{n=0}^{N} \mathbb{P}(A_n).$$

Cette limite existe bien, en effet la suite définie par $U_N = \sum_{n=0}^{N} \mathbb{P}(A_n)$ satisfait $U_{N+1} - U_N = \mathbb{P}(A_{N+1}) \geq 0$ et d'après le 5. de la propriété 4.4

$$\forall\, N,\ U_N = \mathbb{P}(\cup_{n=0}^{N} A_n) \leq \mathbb{P}(\Omega) = 1.$$

Ainsi la suite U_N est croissante, majorée, elle est donc convergente.

4.2 Probabilités conditionnelles et indépendance

Si vous tirez au hasard une personne dans un groupe de 20 (chacun ayant une chance sur 20 d'être choisi) dont 4 portent des lunettes, la personne choisie a 4 chances sur 20 d'avoir des lunettes. Supposons maintenant que, sur ces 20 personnes, 16 soient des femmes, dont 1 porte des lunettes, et 4 soient des hommes, dont 3 portent des lunettes. À nouveau, vous tirez au hasard une des 20 personnes. Vous indiquez à un ami, qui

n'était pas là au moment de l'expérience et n'a donc pas vu qui a été sélectionné, que la personne tirée au hasard est un homme. Il peut alors en déduire que la probabilité que la personne sélectionnée porte des lunettes est de $3/4$, bien plus élevée qu'avant. En obtenant une information ("la personne tirée est un homme"), les probabilités de l'événement "la personne tirée porte des lunettes" est modifié. C'est ce qui mesuré par les probabilités conditionnelles.

Énoncé indispensable 68 :

definition 4.8. Étant donné deux évènements A, B de Ω, avec $\mathbb{P}(B) > 0$, on appelle **probabilité conditionnelle** de A sachant B la quantité

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

Dans l'exemple précédent, $B =$"la personne est un homme" et $A =$"la personne porte des lunettes". Il y 4 hommes sur 20 personnes, donc $\mathbb{P}(B) = 4/20$, et 3 hommes à lunettes sur 20 personnes, donc $\mathbb{P}(A \cap B) = 3/20$. La probabilité que la personne sélectionnée porte des lunettes sachant que c'est un homme est donc $\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B) = 3/20 \times (20/4) = 3/4$.

Intuitivement, $\mathbb{P}(A|B)$ mesure la proportion de A dans B (dans l'exemple, la proportion de personnes à lunettes parmi les hommes).

Comme $\mathbb{P}(A \cap B) \leq \mathbb{P}(B)$, on a toujours $\mathbb{P}(A|B) \leq 1$, et on peut en fait vérifier :

proposition 4.9. *Pour un événement B fixé avec $\mathbb{P}(B) > 0$, l'application $A \mapsto \mathbb{P}(A|B)$ définit une nouvelle mesure de probabilité (différente de $\mathbb{P}$) sur Ω.*

Reprenons le même exemple, mais cette fois-ci, sur les 20 personnes, il y a 16 femmes dont 4 ont des lunettes et 4 hommes dont 1 a des lunettes. Maintenant, savoir que la personne sélectionnée est un homme ou une femme ne donne aucune information sur le fait qu'elle porte des lunettes, puisque la proportion est la même dans les deux cas, de $1/4$. Autrement dit, toujours en notant $B =$"la personne est un homme" et $A =$"la personne porte des lunettes", alors $\mathbb{P}(A|B) = \mathbb{P}(A)$ ce qui, par définition de la probabilité conditionnelle, est équivalent à dire que $\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)$.

Énoncé indispensable 69 : Évènements indépendants

definition 4.10. On dit que deux évènements A et B sont **indépendants** lorsque

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

Plus généralement, on dit que des événements $A_1, \dots A_n$ sont :

1. indépendants deux à deux si A_i et A_j sont indépendants pour tout couple $i \neq j$.

2. mutuellement indépendants si pour tout $J \subset \{1, \ldots, n\}$,

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

EXEMPLE 4.11.

1. Reprenons le double lancer de dés. Intuitivement, le fait de tirer un 5 avec le second lancer ne devrait pas dépendre du fait qu'on a tiré un 2 avec le premier lancer. Nommons $E5$ l'évènement "tirer un cinq au second lancer", soit $E5 = \{1,5\} \cup \{2,5\} \cup \{3,5\} \cup \{4,5\} \cup \{5,5\} \cup \{6,5\}$. Par additivité,

$$\begin{aligned} \mathbb{P}(E5) &= \mathbb{P}(\{1,5\}) + \mathbb{P}(\{2,5\}) + \mathbb{P}(\{3,5\}) + \mathbb{P}(\{4,5\}) + \mathbb{P}(\{5,5\}) + \mathbb{P}(\{6,5\}) \\ &= 6 \times 1/36 = 1/6. \end{aligned}$$

   Pour les mêmes raisons, la probabilité de l'évènement $E2$ "tirer un deux au premier lancer" est $1/6$ et on a bien

$$\mathbb{P}(\{1,2\}) = 1/36 = \mathbb{P}(E5) \times \mathbb{P}(E2).$$

2. Attention à ne pas confondre *indépendants* et *incompatibles*. Prenons l'exemple d'un unique lancer de dés, avec les événements

$$\begin{array}{rcll} A &=& \text{"le resultat est pair"} &= \{2,4,6\} \\ B &=& \text{"le résultat est au moins 3"} &= \{3,4,5,6\} \\ C &=& \text{"le résultat est impair"} &= \{1,3,5\}. \end{array}$$

   Dans ce cas, $\mathbb{P}(A) = 1/2$, $\mathbb{P}(B) = 2/3$ et $\mathbb{P}(A \cap B) = \mathbb{P}(\{2,6\}) = 1/3 = \mathbb{P}(A)\mathbb{P}(B)$, donc A et B sont indépendants. Ils ne sont pas incompatibles, puisque $A \cap B = \{4,6\}$. Au contraire, A et C sont incompatibles, et ils ne sont pas indépendants, puisque $\mathbb{P}(A \cap C) = 0 \neq \mathbb{P}(A) \times \mathbb{P}(C)$.

REMARQUE 4.12. L'indépendance mutuelle implique l'indépendance deux à deux, mais le contraire est faux. En effet, si $A_1, \ldots, A_n$ sont des événements mutuellement indépendants et si i, j sont deux indices différents entre 1 et n, alors en prenant le sous-ensemble $J = \{i, j\}$ dans la définition de l'indépendance mutuelle on obtient immédiatement que A_i et A_j sont indépendants, et comme on a pris i, j quelconques on a bien montré l'indépendance deux à deux. Pour montrer que la réciproque est fausse, construisons un contre-exemple. On lance successivement et de façon indépendante deux dés à 6 faces équilibrés et on considère les événements suivants : A ="premier dé pair", B ="second dé pair" et C ="les deux dés ont la même parité". On vérifie que

$$\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{2}, \qquad \mathbb{P}(A \cap B) = \mathbb{P}(A \cap C) = \mathbb{P}(B \cap C) = \frac{1}{4},$$

ce qui implique que les événements A, B, C sont indépendants deux à deux. En revanche,

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

les trois événements ne sont donc pas mutuellement indépendants.

Test :

On tire une carte d'un jeu de 52 cartes. Quelle est la probabilité de tirer un roi ? Un pique ? Les évènements "tirer un roi" et "tirer un pique" sont ils indépendants ?

proposition 4.13. *Si A et B sont deux événements indépendants, alors A et B^c sont indépendants, de même que A^c et B, ou que A^c et B^c.*

De même, si $A_1, \ldots, A_n$ sont des événements mutuellement indépendants, alors pour tout choix $B_1, \ldots, B_n$ avec, pour chaque i entre 1 et n, $B_i \in \{A_i, A_i^c\}$, les événements $B_1, \ldots, B_n$ sont mutuellement indépendants.

Démonstration. Rappelons qu'il est équivalent de dire que A et B sont indépendants, et que $\mathbb{P}(A|B) = \mathbb{P}(A)$. En particulier, si A et B sont indépendants, $\mathbb{P}(A^c|B) = 1 - \mathbb{P}(A|B) = 1 - \mathbb{P}(A) = \mathbb{P}(A^c)$, et donc A^c et B sont indépendants. Les autres cas sont semblables, et le cas $n \geqslant 0$ s'obtient par récurrence. □

Énoncé indispensable 70 :

Formule des probabilités totales : si $(A_i)_{1\leq i\leq n}$ sont des évènements disjoints de probabilité non nulle tels que $\cup_{i=1}^n A_i = \Omega$,

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B|A_i)\mathbb{P}(A_i).$$

De plus pour $1 \leq k \leq n$, on a la **Formule de Bayes**

$$\mathbb{P}(A_k|B) \;=\; \frac{\mathbb{P}(B|A_k)\mathbb{P}(A_k)}{\sum_{i=1}^n \mathbb{P}(B|A_i)\mathbb{P}(A_i)}$$

Démonstration. Les formules des probabilités totales et de Bayes sont des cas typiques pour lesquelles apprendre par cœur la formule (sans se tromper : où est B, où sont les A_k, où la probabilité est-elle conditionnelle, par rapport à quoi, bref) est peut-être plus difficile que d'en comprendre la démonstration (et donc de pouvoir retrouver la formule). En effet, la formule des probabilités totales est simplement la transcription d'une **disjonction de cas**. La formule de Bayes, elle, se retrouve immédiatement à partir de la formule des probabilités totales et de la définition des probabilités conditionnelles. Détaillons.

Considérons donc une partition $(A_i)_{1\leq i\leq n}$ de Ω, c'est-à-dire des événements disjoints dont l'union est Ω, et supposons que $\mathbb{P}(A_k) > 0$ pour tout k entre 1 et n (sinon l'événement A_k ne servirait à rien au sens où il aurait une probabilité nulle d'arriver).

Autrement dit, les A_k forment une disjonction de cas : nécessairement, l'un des A_k se produit, et deux évévement A_k et A_j (où $j \neq k$) ne peuvent pas se produire simultanément. On peut décomposer B en fonction de ces différents cas, c'est-à-dire écrire

$$B \;=\; B \cap \Omega \;=\; B \cap (\cup_{i=1}^{n} A_i) \;=\; \cup_{i=1}^{n} (B \cap A_i) \;.$$

Puisque les événements A_k sont incompatibles entre eux, les événements $B \cap A_k$ sont incompatibles entre eux et donc (par additivité des probabilités)

$$\mathbb{P}(B) \;=\; \mathbb{P}\left(\cup_{i=1}^{n} (B \cap A_i)\right) \;=\; \sum_{i=1}^{n} \mathbb{P}\left(B \cap A_i\right) \;.$$

Enfin, simplement en écrivant la définition de la probabilité conditionnelle, $\mathbb{P}\left(B \cap A_i\right) = \mathbb{P}\left(B|A_i\right)\mathbb{P}\left(A_i\right)$, ce qui conclut la démonstration de la formule des probabilités totales.

Maintenant, fixons-nous un indice k entre 1 et n, et démontrons la formule de Bayes. Ce qu'il faut retenir de cette formule, c'est qu'elle permet en quelque sorte "d'inverser les probabilités conditionnelles", au sens où elle exprime une quantité du genre $\mathbb{P}(A|B)$ en fonction de quantités du genre $\mathbb{P}(B|A)$. Or, la définition des probabilités conditionnelles fait apparaître $\mathbb{P}(A \cap B)$, quantité pour laquelle A et B jouent exactement le même rôle. Donc, pour démontrer la formule de Bayes, il suffit de faire apparaître cette quantité symétrique et d'intervetir les rôles de A et B. Plus précisément, la définition des probabilités conditionnelles, utilisée deux fois, nous permet de dire que

$$\mathbb{P}(A_k|B) = \frac{\mathbb{P}(A_k \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A_k)\mathbb{P}(A_k)}{\mathbb{P}(B)} \,.$$

On a bien $\mathbb{P}(A_k|B)$ à gauche et $\mathbb{P}(B|A_k)\mathbb{P}(A_k)$ à droite. Maintenant, pour retrouver la formule de Bayes, il suffit de réécrire le dénominateur $\mathbb{P}(B)$ grâce à la formule des probabilités totales :

$$\frac{\mathbb{P}(B|A_k)\mathbb{P}(A_k)}{\mathbb{P}(B)} \;=\; \frac{\mathbb{P}(B|A_k)\mathbb{P}(A_k)}{\sum_{i=1}^{n} \mathbb{P}(B|A_i)\mathbb{P}(A_i)} \,.$$

□

La formule des probabilités totales s'interprète ainsi : si vous avez $2/3$ chances de sortir si il fait beau, $1/6$ chances de sortir si il pleut, et qu'il y a une chance sur deux qu'il fasse beau, alors votre probabilité de sortir est $(2/3) \times (1/2) + (1/6) \times (1/2) = 5/12$ (sommes des probabilités conditionnelles pondérées par la probabilité de la condition).

EXEMPLE 4.14. On dispose de trois dés : un rouge, un bleu, un vert. Le rouge est équilibré, il a une chance sur 6 de faire un 6. Le bleu et le vert sont pipés, ils ont respectivement une probabilité $1/5$ et $1/4$ de faire 6. On choisit un dé au hasard (avec une chance sur 3 chacun). Quelle est la probabilité de faire 6 ? On fait une disjonction de cas selon le dé utilisé. On note A_1 ="le dé est rouge", A_2 ="le dé est bleu" et A_3 ="le dé est vert". C'est bien une partition de Ω : il y a forcément un dé tiré, et il ne peut y en avoir qu'un. On note B ="le dé fait 6". Les informations dont on dispose sont :

$$\mathbb{P}(A_1) = \mathbb{P}(A_2) = \mathbb{P}(A_3) = \frac{1}{3} \,, \qquad \mathbb{P}(B|A_1) = \frac{1}{6} \,, \qquad \mathbb{P}(B|A_2) = \frac{1}{5} \,, \qquad \mathbb{P}(B|A_3) = \frac{1}{4} \,.$$

D'après la formule des probabilités totales,

$$\mathbb{P}(B) \;=\; \frac{1}{6}\times\frac{1}{3}+\frac{1}{5}\times\frac{1}{3}+\frac{1}{4}\times\frac{1}{3} \;\simeq\; 21\%$$

remarque 4.15. La formule de Bayes est particulièrement utile dans le cas $n = 2$, lorsque les (A_i) sont simplement A et A^c, soit

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A)+\mathbb{P}(B|A^c)\mathbb{P}(A^c)}$$

exemple 4.16. Une maladie affecte la population à raison d'une personne pour 10^4. Un test de dépistage pour cette maladie est positif pour 99% des infectés, et pour $0,1\%$ des non infectés. Quelle est la probabilité d'être infecté sachant que le test est positif ?
Notons A (A^c respectivement) le fait d'être infecté (resp. non infecté), B le résultat "test positif". Les informations se traduisent par

$$\mathbb{P}(A) = 1/10^4, \qquad \mathbb{P}(A^c) = 1-1/10^4, \qquad \mathbb{P}(B|A) = 99/100, \qquad \mathbb{P}(B|A^c) = 0,1/100.$$

La formule de Bayes donne alors la probabilité $\mathbb{P}(A|B)$ d'être infecté sachant que le test est positif

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A)+\mathbb{P}(B|A^c)\mathbb{P}(A^c)} = \frac{99/100\cdot 10^{-4}}{99/100\cdot 10^{-4}+10^{-3}(1-10^{-4})} \simeq 0.09.$$

Soit 9% ! C'est le problème des faux positifs : plus un évènement est rare, plus il est difficile de le détecter de façon fiable.

Test : L'île aux chats

Une île est infestée de chats, roux et gris. 30% des chats sont mâles, les chats mâles sont roux à 60%, les chats femelles sont roux à 20%. Quelle est la probabilité qu'un chat roux soit un mâle ?

Chapitre 5

Variables aléatoires discrètes

5.1 Définition d'une variable aléatoire discrète

Dans de nombreux contextes, le résultat d'une expérience aléatoire peut mener à la définition d'un ou plusieurs nombre, aléatoires donc. Par exemple, si on jette simultanément 5 dés, on peut regarder leur somme, ou le maximum, ou le nombre de 6, etc. C'est ce qu'on appelle une variable aléatoire. Comme pour le reste de la théorie des probabilités, ces variables sont formalisées mathématiquement à l'aide des objets de la théorie des ensembles, en l'occurence les fonctions. Plus précisément :

definition 5.1. Soient Ω et D deux ensembles finis ou dénombrables, et $\mathbb{P}$ une mesure de probabilité sur Ω. On appelle variable aléatoire discrète sur Ω dans D une fonction de Ω dans D.

On considérera principalement le cas où D est un sous-ensemble fini de $\mathbb{N}$ ou bien encore le cas où $D = \mathbb{N}$. Des exemples ou exercices avec D qui est inclus dans $\mathbb{N}^n$ pour un $n \in \mathbb{N}$ plus grand strictement que 1 seront abordés (lancer simultané de plusieurs pièces ou de deux dés).

Considérons une telle application $X : \Omega \to D$. Pour chaque élément $i \in D$ on définit sa pré-image (ou image réciproque) par X comme le sous-ensemble de Ω, noté $\{X = i\}$ ou $X^{-1}(\{i\})$, donné par

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

C'est l'ensemble de tous les antécédents de l'élément i par l'application X. Cet ensemble peut éventuellement être l'ensemble vide. Plus généralement, pour tout $B \subset D$, son image réciproque par X est notée $\{X \in B\}$ ou $X^{-1}(B)$ et est définie par

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

Pour tout $B \subset D$, on a la propriété fondamentale des images réciproques :

$$X^{-1}(B) \;=\; \cup_{i \in B}\{X = i\}\,,$$

avec les sous-ensembles $\{X = i\}$ qui sont deux à deux disjoints lorsque $i \in D$ et, dans le cas discret, la réunion qui est constituée d'un nombre au plus dénombrable de sous-ensembles.

exemple 5.2 (Temps d'attente à un arrêt de bus). Un passager arrive à un arrêt de bus, il sait juste qu'un bus passe toutes les 10mn. Son *temps d'attente en minutes* peut prendre les valeurs 0 (si le bus est là quand il arrive), $1, \dots, 10$ (si le bus vient de partir quand il arrive). Supposons que le passager va être à l'heure à son rendez-vous si le bus arrive au plus dans 3 minutes, en retard sinon. On veut définir une variable aléatoire X égale à 1 si le passager est à l'heure, 0 s'il est en retard. L'espace des états est donc $\Omega = \{0, 1, 2, \cdots, 10\}$, X est à valeur dans $D = \{0, 1\}$ et on définit $X : \Omega \to D$ par :

$$X(0) = X(1) = X(2) = X(3) = 1 \quad \text{et}$$

$$X(4) = X(5) = X(6) = X(7) = X(8) = X(9) = X(10) = 0$$

Les pré-images de 0 et 1 sont donc

$$\{X = 1\} = \{0, 1, 2, 3\}\,, \qquad \{X = 0\} = \{4, 5, 6, 7, 8, 9, 10\}\,,$$

et forment une partition de Ω :

$$X^{-1}(\{0, 1\}) = \{X = 0\} \cup \{X = 1\} \quad \text{et} \quad \{X = 0\} \cap \{X = 1\} = \emptyset\,.$$

exemple 5.3 (Lancer d'un dé). On lance un dé à six faces numérotées de 1 à 6 et on gagne un point si le résultat est 5, deux points s'il est 6, zéro sinon. L'espace des états est $\Omega = \{1, 2, \cdots, 6\}$ et le nombre de points gagnés est une variable aléatoire X définie par :

$$X(1) = X(2) = X(3) = X(4) = 0\,, \qquad X(5) = 1 \quad \text{et} \quad X(6) = 2\,.$$

Test : images réciproques

Dans l'exemple 5.3, trouver D, $\{X = 0\}$, $\{X = 1\}$ et $\{X = 2\}$. Vérifier que

$$X^{-1}(\{0, 1\}) = \{X = 0\} \cup \{X = 1\} \cup \{X = 2\} \quad \text{et} \quad \{X = i\} \cap \{X = j\} = \emptyset$$

pour $i \neq j$ dans $\{0, 1, 2\}$.

Dans la suite pour simplifier les notations on écrira :

$$\mathbb{P}(\{X = i\}) = \mathbb{P}(X = i) = p_i \quad \text{pour } i \in D \quad \text{et} \quad \mathbb{P}(\{X \in B\}) = \mathbb{P}(X \in B)\,.$$

Énoncé indispensable 71 : Formule sommatoire des v.a.d.

Soit X une variable aléatoire discrète sur un espace probabilisé $(\Omega, \mathbb{P})$. Alors pour tout $B \subset D$,

$$\mathbb{P}(X \in B) = \sum_{i \in B} \mathbb{P}(X = i) = \sum_{i \in B} p_i\,.$$

En particulier $\mathbb{P}(X \in D) = \mathbb{P}(\Omega) = 1$.

REMARQUE 5.4. Quand le sous-ensemble B est de cardinal fini la somme des p_i pour $i \in B$ contient un nombre fini de termes et est donc parfaitement définie. Lorsque B n'est pas de cardinal fini, ce dernier est au plus dénombrable et on se ramène à calculer des sommes de séries à termes tous positifs ou nuls.

Démonstration. Puisque $\{X \in B\} = \cup_{i\in B}\{X = i\}$ et que $\{X = i\} \cap \{X = j\} = \emptyset$ pour $i \neq j$, les événements $\{X = i\}$ pour $i \in B$ forment une partition de $\{X \in B\}$. L'énoncé indispensable 71 est donc une conséquence directe de la propriété d'additivité (ou sigma-additivité) des probabilités (cf. chapitre précédent, l'énoncé indispensable 67 dans le cas où D est fini, et la Définition 4.6 pour le cas dénombrable). □

Étant donné une variable aléatoire X sur Ω dans D, on peut construire une application $\mathbb{Q}$ sur les parties de D à valeur dans $[0, 1]$ par

$$\mathbb{Q}(B) = \mathbb{P}(X \in B) = \sum_{i\in B} p_i$$

pour tout $B \subset D$. Il n'est pas difficile de vérifier que $\mathbb{Q}$ est alors une probabilité sur D.

DEFINITION 5.5. La mesure de probabilité $\mathbb{Q}$ est appelée la *loi de probabilité de la variable discrète* X.

EXEMPLE 5.6. Reprenons l'exemple 5.2, en supposant que les temps d'atteinte ont tous la même probabilité, c'est-à-dire que

$$p_0 = p_1 = \cdots = p_{10} = \frac{1}{11}\,.$$

La loi de X est donc la loi de probabilité $\mathbb{Q}$ sur $D = \{0, 1\}$ donnée par

$$\mathbb{Q}(1) = \mathbb{P}(X = 1) = \mathbb{P}(\{0, 1, 2, 3\}) = p_0 + p_1 + p_2 + p_3 = 4 \times \frac{1}{11} = \frac{4}{11}\ ;$$

et

$$\mathbb{Q}(0) = \mathbb{P}(X = 0) = \mathbb{P}(\{4, 5, 6, 7, 8, 9, 10\}) = p_4 + p_5 + \cdots + p_{10} = 7 \times \frac{1}{11} = \frac{7}{11}\,.$$

EXEMPLE 5.7 (Exemple fondamental de variable discrète). Dans le cas où $\Omega = D$, on peut considérer l'application identité, définie par $I_d(i) = i$ pour tout $i \in \Omega$. Dans ce cas, la loi de I_d est $\mathbb{Q} = \mathbb{P}$.

REMARQUE 5.8. Au vu de l'exemple précédent, on peut se demander à quoi sert Ω. Reprenons l'exemple de deux lancers de dés dont on regarde la somme. Plutôt que de définir $\Omega = \{1, \ldots, 6\}^2$ comme l'ensemble des couples possibles, et ensuite de définir $X : \{1, \ldots, 6\}^2 \to \{2, \ldots, 12\}$ comme la somme des deux lancers, on pourrait directement définir $\Omega = \{2, \ldots, 12\}$ et prendre $X = I_d$. Le problème, c'est que si ensuite on définit la valeur Y comme le maximum des deux lancers, et qu'on fait la même chose, il faut définir $\Omega' = \{1, \ldots, 6\}$ et $Y : \Omega' \to \Omega'$. En particulier, X et Y seront deux applications sur des ensembles différents et donc l'application $X + Y$ n'aura aucun sens, alors qu'on voudrait que cela décrive "la somme des deux lancés plus le maximum". Il vaut donc mieux garder Ω comme l'ensemble de toutes les possibilités décrites par l'expérience

aléatoire séminale, celle à partir de laquelle sont générées toutes les variables aléatoires.

D'ailleurs, comme on le verra en exercice, l'univers Ω est souvent implicite. On dit par exemples "soit X une variable aléatoire à valeur dans $D = \{1, 2, 3\}$ avec $\mathbb{P}(X = i) = i/6$ pour $i \in D$". On ne sait pas comment a été générée cette variable aléatoire (et cela importe peu). Par définition, X est donc une application d'un certain espace probabilisé $(\Omega, \mathbb{P})$ vers D, mais Ω n'interviendra jamais dans les manipulations qu'on effectuera sur X et donc on ne l'explicite pas. On peut voir Ω comme une source cachée d'aléa, qui sert successivement à générer toutes les variables aléatoires d'un même exercice.

Finalement, la notion d'indépendance vue pour les événements peut s'étendre à des variables aléatoires discrètes de la façon suivante :

definition 5.9. Deux variables aléatoires discrètes $X : \Omega \to D_1$ et $Y : \Omega \to D_2$ sont dites **indépendantes** si les événements $\{X \in A_1\}$ et $\{Y \in A_2\}$ sont indépendants pour tous sous-ensembles $A_1 \subset D_1$ et $A_2 \subset D_2$.

Remarquons que, de façon équivalente, X et Y sont indépendants si les événements $\{X = x\}$ et $\{Y = y\}$ sont indépendants pour tout $x \in D_1, y \in D_2$ (autrement dit dans la définition on peut se restreindre aux cas où $A_1 = \{x\}$ et $A_2 = \{y\}$ sont des singletons).

On reviendra plus en détail sur la notion de variables aléatoires indépendantes au chapitre 7.

5.2 Quelques lois de probabilité discrètes

5.2.1 *Loi uniforme sur un ensemble fini*

Soit Ω un ensemble fini de cardinal N non nul. On définit la mesure uniforme sur Ω comme la probabilité $\mathbb{P}$ telle que $\mathbb{P}(i) = 1/N$ pour tout $i \in \Omega$. Autrement dit, pour tout $B \subset \Omega$:

$$\mathbb{P}(B) \;=\; \sum_{i \in B} \mathbb{P}(\{i\}) \;=\; \frac{\mathrm{Card}(B)}{N} \;=\; \frac{\text{nombre de cas favorables}}{\text{nombre total de cas}}.$$

On dit que tous les éléments de Ω sont équiprobables (ils ont tous la même probabilité). On dit aussi d'une variable aléatoire dont la loi est uniforme qu'elle est uniformément distribués. La loi uniforme sur Ω est notée $\mathcal{U}(\Omega)$.

exemple 5.10. Les exemples sont nombreux : lancer de dé ($N = 6$), pile ou face ($N = 2$), la loi $\mathbb{P}$ de l'exemple 5.6 ($N = 11$ avec $\Omega = \{0, 1, \ldots, 10\}$), tirer quelqu'un au hasard dans un groupe… D'ailleurs, dans le langage courant (à éviter en mathématique !), "au hasard" signifie souvent, implicitement, "uniformément". Il faut en fait se méfier : ce n'est pas parce qu'une expérience a N issues possibles et qu'on a aucune idée des probabilités respectives des différentes issues qu'on peut supposer (par défaut) qu'elles sont équiprobables. Ainsi, de deux choses l'une : demain, soit je serai frappé par la foudre, soit je ne le serai pas. Il n'y a que ces deux possibilités, dont je ne connais pas les probabilités, mais qui ne sont vraisemblablement pas équiprobables. Plus subtile, si on lance deux dés équilibrés à 6 face, notons X le résultat du premier lancer, Y du second, et $S = X + Y$ leur somme. Alors le couple (X, Y) est uniformément distribué sur $\{1, \ldots, 6\} \times \{1, \ldots, 6\}$, tandis que S n'est pas uniformément distribué sur $\{2, \ldots, 12\}$. En

effet,

$$\mathbb{P}(S=2) \;=\; \mathbb{P}(X=1,Y=1) \;=\; \frac{1}{36}$$

tandis que

$$\mathbb{P}(S=3) \;=\; \mathbb{P}(X=1,Y=2)+\mathbb{P}(X=2,Y=1) \;=\; \frac{2}{36} \;=\; \frac{1}{18}\,.$$

5.2.2 *Loi de Bernoulli*

Soient l'ensemble $\Omega=\{0,1\}$ et un réel $p\in\,]0,1[$. La loi de Bernoulli de paramètre $p\in\,]0,1[$, notée $\mathcal{B}(p)$, est la probabilité $\mathbb{P}$ sur Ω qui vérifie :

$$\mathbb{P}(\{0\})=1-p \quad \text{et} \quad \mathbb{P}(\{1\})=p\,.$$

On a bien sûr $\mathbb{P}(\Omega)=\mathbb{P}(\{0\})+\mathbb{P}(\{1\})=1-p+p=1$.

EXEMPLE 5.11. La loi $\mathbb{Q}$ de l'exemple 5.6 suit une loi de Bernoulli de paramètre $p=4/11$. Un autre exemple est le résultat du lancer d'une pièce si on affecte, par exemple, la valeur 0 à l'observation du côté face de la pièce et la valeur 1 à l'observation du côté pile de la pièce. Si la pièce est *non biaisée* (on a autant de chance d'observer face que pile), alors $p=1/2$ tandis que si la pièce est *biaisée* $p\neq 1/2$ est donné par la probabilité de faire pile.

Éléments historiques 2 : La Dynastie Bernoulli

Il n'y eut pas un mathématicien nommé Bernoulli, mais sept ! Tout commence avec les frères Jacques (1654–1705 ; c'est lui qui donne son nom à la loi de Bernoulli) et Jean (1667–1748) Bernoulli, mathématiciens (et physiciens ; à l'époque, c'était à peu près la même chose). Le cadet, Jean, a eu trois fils, Daniel (1700–1782), Nicolas (1695–1726) et Jean II (1710–1790) Bernoulli, eux-mêmes tous mathématiciens. Jean II a ensuite deux fils mathématiciens, Jean III (1744–1807) et Jacques II (1759–1789) Bernoulli. La lignée ne s'arrête pas vraiment là puisque, entre autre, Pierre Curie (1859–1906, prix Nobel de physique en 1903 avec sa femme Marie Skłodowska-Curie et Henri Becquerel) et Pierre-Gilles de Gennes (1932–2007, prix Nobel de physique en 1991) sont tous deux des descendants directs de Jean Bernoulli.

5.2.3 *Loi binomiale*

Soient un entier naturel N non nul, un réel p de $]0,1[$ et l'ensemble $\Omega=\{0,1,2,\ldots,N\}$. La loi binomiale $\mathcal{B}(N,p)$, de paramètres $N\in\mathbb{N}^*$ et $p\in\,]0,1[$, est la probabilité $\mathbb{P}$ sur Ω définie par

$$\mathbb{P}(\{n\})=\binom{N}{n}p^n(1-p)^{N-n} \quad \text{pour } n\in\Omega\,.$$

On vérifie que $\mathbb{P}(\Omega) = \sum_{n=0}^{N} \mathbb{P}(\{n\}) = 1$ par la formule du binôme de Newton. En effet :

$$1 = (p + (1-p))^N = \sum_{n=0}^{N} \binom{N}{n} p^n (1-p)^{N-n}.$$

Énoncé indispensable 72 : lancer de N pièces

Soient $N \in \mathbb{N}^*$, $p \in]0,1[$ et une pièce ayant la probabilité p de donner la valeur pile. Si on lance N fois cette pièce, alors la variable aléatoire discrète qui compte le nombre de fois que la valeur pile est obtenue suit une loi binomiale $\mathcal{B}(N,p)$.

Démonstration. On donne deux démonstrations, la première ne s'appliquant qu'au cas particulier $p = 1/2$ (c'est-à-dire que la pièce est non biaisée : elle a autant de chances de faire pile que face).

Le cas non-biaisé. Quand on lance N fois une pièce non biaisée et que l'on note 1 pour pile et 0 pour face, le résultat est une suite de N valeurs 0 ou 1, autrement dit l'univers est $\Omega = \{0,1\}^N$. Pour une pièce non biaisée, la loi $\mathbb{P}$ correspondante est la loi uniforme sur Ω (toutes les suites ω de N valeurs 0 ou 1 sont de même probabilité et il y a 2^N réalisations possibles). On note X la variable aléatoire discrète qui compte le nombre de fois que la valeur pile est obtenue, ainsi $X(\omega)$ est la somme des N valeurs de la suite ω et X est à valeur dans $D = \{0,1,2,\ldots,N\}$.

Pour $n \in D$, déterminer $\mathbb{P}(X = n)$ nécessite de connaître le cardinal de l'ensemble $\{X = n\}$, c'est-à-dire de l'ensemble des suites à N éléments qui ont n fois la valeur 1, il y en a exactement le nombre de combinaison $\binom{N}{n}$, de sorte que :

$$\mathbb{P}(X = n) = \frac{\text{nombre de cas possibles}}{\text{nombre total de cas}} = \frac{\binom{N}{n}}{2^N} = \binom{N}{n} \times \left(\frac{1}{2}\right)^n \times \left(\frac{1}{2}\right)^{N-n}.$$

Le cas général. Comme précédemment, on lance N fois une pièce (qui a une probabilité p de faire pile) et l'on note 1 pour pile et 0 pour face. Le résultat est à nouveau une suite de N valeurs 0 ou 1, autrement dit l'univers est toujours $\Omega = \{0,1\}^N$. Cependant, cette fois, toutes les séquences ne sont pas équiprobables. Considérons une séquence $\omega = (\omega_1, \omega_2, \ldots, \omega_N)$, et calculons la probabilité d'observer cette suite de lancer. Pour cela, notons $X_i = 1$ si le i^{eme} lancer fait pile et $X_i = 0$ pour i entre 1 et N. Alors $\{X_1 = \omega_1\}, \ldots, \{X_N = \omega_N\}$ sont des événements indépendants, de sorte que

$$\mathbb{P}(\omega) = \mathbb{P}(X_1 = \omega_1,\ X_2 = \omega_2,\ \ldots,\ X_N = \omega_N) = \mathbb{P}(X_1 = \omega_1) \ldots \mathbb{P}(X_N = \omega_N).$$

De plus, les X_i suivent tous une loi $\mathcal{B}(p)$, de sorte que pour tout i entre 1 et N,

$$\mathbb{P}(X_i = \omega_i) = p^{\omega_i}(1-p)^{1-\omega_i}$$

(il suffit en effet de vérifier cette égalité pour les deux seules valeurs que peut prendre

ω_i, à savoir 0 et 1). On obtient donc

$$\mathbb{P}(\omega) \;=\; p^{\omega_1}(1-p)^{1-\omega_1}\dots p^{\omega_N}(1-p)^{1-\omega_N} \;=\; p^{\sum_{i=1}^N \omega_i}(1-p)^{N-\sum_{i=1}^N \omega_i}\,.$$

Autrement dit, si n est le nombre de 1 que contient la suite ω, alors $\mathbb{P}(\omega) = p^n(1-p)^{N-n}$. Notons S la variable aléatoire qui donne le nombre de 1 dans les N lancer. Alors

$$\mathbb{P}(S=n) \;=\; \sum_{\omega\in\Omega, S(\omega)=n} \mathbb{P}(\omega) \;=\; p^n(1-p)^{N-n} \times \mathrm{Card}\left(\{\omega\in\Omega \mid S(\omega)=n\}\right)\,.$$

Dans le cas non-biaisé, on a déjà utilisé que le nombre de suites de N valeurs 0 ou 1 qui contiennent exactement n fois la valeur 1 est $\binom{N}{n}$, et on a donc bien démontré que

$$\mathbb{P}(S=n) \;=\; \binom{N}{n} p^n(1-p)^{N-n}\,.$$

□

Test : loi binomiale

Dans l'exemple 5.2, si le passager compte le nombre de fois où le bus est arrivé en au plus 3mn sur une période de 5 jours, montrer que ce nombre suit une loi binomiale $\mathcal{B}(5, 4/11)$.

5.2.4 *Loi géométrique*

Soit $p \in]0,1[$ et $\Omega = \mathbb{N}^*$. La loi géométrique de paramètre p, notée $\mathcal{G}(p)$, est la probabilité $\mathbb{P}$ sur $\mathbb{N}^*$ donnée par

$$\mathbb{P}(\{n\}) = p(1-p)^{n-1} \quad \text{pour } n\in\Omega\,.$$

Pour vérifier que $\mathbb{P}(\Omega)=1$, on rappelle la valeur de la série géométrique : pour $r\in]0,1[$, $\sum_{n\geqslant 0} r^n = 1/(1-r)$, de sorte que

$$\mathbb{P}(\Omega) \;=\; \sum_{n\geqslant 1} p(1-p)^{n-1} \;=\; p\sum_{n\geqslant 1}(1-p)^{n-1} \;=\; p\sum_{m\geqslant 0}(1-p)^m \;=\; \frac{p}{1-(1-p)} \;=\; 1\,.$$

Énoncé indispensable 73 : Loi géométrique et lancers de pièce

Si on lance de façon répétée une pièce qui a une probabilité p de faire pile, qu'on s'arrête la première fois que la pièce tombe sur pile et qu'on note X le nombre de lancer effectués, alors $X \sim \mathcal{G}(p)$.

Démonstration. Il s'agit de vérifier que $\mathbb{P}(X=n)$ est bien donné par $p(1-p)^{n-1}$ pour tout $n\geqslant 1$. Il est en fait plus simple de calculer $\mathbb{P}(X>n)$. En effet, on fait strictement plus de n lancer si et seulement si les n premier lancers ont donné face, autrement dit l'événement $\{X>n\}$ est égal à l'événement $\{X_1=0,\dots,X_n=0\}$, où $X_i=1$ si le i^{eme} lancer donne pile et $X_i=0$ sinon. Les X_i étant distribués selon la loi $\mathcal{B}(p)$ et les lancers

étant indépendants,

$$\mathbb{P}(X > n) \;=\; \mathbb{P}(X_1 = 0, \dots, X_n = 0) \;=\; \mathbb{P}(X_1 = 0) \dots \mathbb{P}(X_n = 0) \;=\; (1-p)^n \,.$$

Puisque $\{1, \dots, n\} = \{1, \dots, n-1\} \cup \{n\}$ (l'union étant disjointe),

$$\mathbb{P}(X \leqslant n) \;=\; \mathbb{P}(X \leqslant n-1) + \mathbb{P}(X = n)$$

et donc, en utilisant que $\mathbb{P}(A) = 1 - \mathbb{P}(A^c)$ pour tout événement A,

$$\begin{aligned}\mathbb{P}(X = n) \;&=\; \mathbb{P}(X \leqslant n) - \mathbb{P}(X \leqslant n-1) \;=\; 1 - \mathbb{P}(X > n) - (1 - \mathbb{P}(X > n-1)) \\ &=\; (1-p)^{n-1} - (1-p)^n \;=\; (1-p)^{n-1}(1-(1-p)) \;=\; p(1-p)^{n-1} \,.\end{aligned}$$

CQFD. □

exemple 5.12. Au-delà des lancers de dés, la loi géométrique décrit donc le nombre de fois qu'une même expérience est répétée de façon indépendante jusqu'à donné un résultat donné. Par exemple, si quelqu'un joue tous les jours au loto, alors le nombre de jour pendant lesquels il va devoir jouer avant de gagner suit une loi géométrique de paramètre p, où p est la probabilité de gagner au loto (soit $p \simeq 0,5 \cdot 10^{-9}$).

5.2.5 *Loi de Poisson*

Soient λ un réel strictement positif et $\Omega = \mathbb{N}$. La loi de Poisson $\mathcal{P}(\lambda)$, de paramètre $\lambda \in \mathbb{R}_+^*$, est la probabilité $\mathbb{P}$ qui vérifie :

$$\mathbb{P}(\{n\}) = \frac{\lambda^n}{n!}\, e^{-\lambda} \quad \text{pour } n \in \Omega \,.$$

On a bien $\mathbb{P}(\Omega) = 1$, car

$$\sum_{n \in \mathbb{N}} \mathbb{P}(\{n\}) = \sum_{n \in \mathbb{N}} \frac{\lambda^n}{n!}\, e^{-\lambda} = e^{-\lambda} \sum_{n \in \mathbb{N}} \frac{\lambda^n}{n!} = e^{-\lambda} e^{\lambda} = 1 \,.$$

Plus généralement, pour tout $A \subset \mathbb{N}$ de cardinal fini on a :

$$\mathbb{P}(A) = \sum_{n \in A} \frac{\lambda^n}{n!}\, e^{-\lambda} \,,$$

et on admettra que le résultat reste vrai lorsque $A \subset \mathbb{N}$ n'est pas de cardinal fini.

exemple 5.13 (nombre d'événements se produisant pendant un temps donné). La loi de Poisson est en quelque sorte une version continue de la loi binomiale. Prenons par exemple le problème suivant : on veut modéliser le nombre de fois, sur un an, où un appareil est tombé en panne. Pour simplifier, on peut supposer que, chaque mois, l'appareil a une certaine probabilité p de tomber en panne, indépendemment des autres mois. Sur douze mois, le nombre de panne suit alors la loi $\mathcal{B}(12, p)$. Mais pourquoi avoir choisi le mois comme unité de temps ? On peut supposer que, chaque jour, l'appareil a une certaine probabilité p' de tomber en panne (plus petite que p : la probabilité d'être tombé en panne un jour donné est plus petite que celle d'être tombé en panne

sur un mois entier). Le nombre de pannes suit alors la loi $\mathcal{B}(365, p')$. Mais on pourrait aussi prendre comme durée de référence l'heure, ou la seconde, etc. On peut après tout mesurer avec précision le moment où la panne a eu lieu. Remarquons que le nombre de pannes ne devrait pas être affecté par la fréquence à laquelle on les mesure (sauf s'il y a deux pannes entre deux mesures, ce qui devient très improbable quand la fréquence de mesure tend vers l'infini), dit autrement ce n'est pas parce qu'on mesure de plus en plus souvent que le nombre de pannes augmente. De fait, si on ne discrétise plus du tout le temps, le nombre de panne va suivre une loi de Poisson. Cette loi modélise bien les événements rares, comme les décomposition d'atomes radioactifs ou le nombre d'accient d'un assuré avec son automobile.

Précisons le raisonnement précédent. Soient λ et T deux réels strictement positifs. On suppose qu'un type d'événements se produit de la façon suivante. Pendant un temps t petit il n'en arrive qu'un seul avec une (petite) probabilité qui vaut λt et entre deux intervalles de temps disjoints ces événements se produisent de façon indépendante. On souhaite déterminer le nombre d'événements X se produisant pendant la durée T.

Supposons que $t = T/N$ avec $N \in \mathbb{N}$ grand. On décompose l'intervalle $]0, T]$ en N intervalles de la forme $](i-1)T/N, iT/N]$ pour i entre 1 et N. Sur chacun d'eux (et indépendemment l'un par rapport à l'autre), il y a une probabilité $\lambda T/N$ qu'un évenement survienne. D'après l'énonce indispensable 72, X le nombre total d'événements qui surviennent pendant $[0, T]$ suit une loi binomiale $\mathcal{B}(N, \lambda T/N)$, c'est-à-dire qu'en notant $p = \lambda T/N$:

$$\mathbb{P}(X = n) = \binom{N}{n} p^n (1-p)^{N-n} \quad \text{pour } n \in \{0, 1, \dots, N\}.$$

Maintenant, si ce modèle n'est pas exact mais est une approximation qui devient vraie seulement en passant à la limite lorsque $N \to +\infty$, nous allons passer à la limite dans l'égalité précédente (à n fixé). En réécrivant

$$p^n = \left(\frac{\lambda T}{N}\right)^n, \qquad (1-p)^{N-n} = \exp\left((N-n)\ln(1-\lambda T/N)\right)$$

et

$$\binom{N}{n} = \frac{N(N-1)\dots(N-n+1)}{n!},$$

l'égalité précédente devient

$$\mathbb{P}(X = n) = \frac{(\lambda T)^n}{n!} \times \frac{N}{N} \times \frac{N-1}{N} \times \dots \times \frac{N-n+1}{N} \exp\left((N-n)\ln(1-\lambda T/N)\right),$$

et, à la limite $N \to +\infty$, puisque $(N-n)\ln(1-\lambda T/N) \simeq -(N-n)\lambda T/N \to -\lambda T$,

$$\lim_{N\to+\infty} \mathbb{P}(X = n) = \frac{(\lambda T)^n}{n!} e^{-\lambda T}.$$

Pour aller plus loin 1 :

Au cours de la remarque précédente, on a en fait démontré la chose suivante (en

prenant $T = 1$) :

proposition 5.14 (Convergence de la loi binomiale vers la loi de Poisson). *fixons $\lambda > 0$ et considérons pour chaque $N > \lambda$ une variable $X^{(N)}$ de loi $\mathcal{B}(N, \lambda/N)$. Soit Y une variable aléatoire de loi $\mathcal{P}(\lambda)$. Alors, pour tout $n \in \mathbb{N}$,*

$$\mathbb{P}(X^{(N)} = n) \underset{N\to+\infty}{\longrightarrow} \mathbb{P}(Y = n)\,.$$

On dit que la loi binomiale $\mathcal{B}(N, \lambda/N)$ converge vers la loi de Poisson $\mathcal{P}(\lambda)$ lorsque $N \to +\infty$.

5.3 Espérance et variance d'une variable aléatoire discrète

definition 5.15 (Espérance d'une variable discrète). Quand elle existe (c'est-à-dire, quand la somme est bien définie), c'est la *moyenne des valeurs atteintes par X pondérées par les probabilités d'atteindre ces valeurs*. Elle est notée $\mathbb{E}(X)$ et est donc donnée par :

$$\mathbb{E}(X) = \sum_{i\in D} i \times p_i\,.$$

Quand cette somme est finie on dit que *la variable X est d'espérance finie*.

remarque 5.16. Lorsque l'ensemble D est de cardinal fini, la somme définissant l'espérance de la variable X est toujours définie (c'est un barycentre). Lorsque $D = \mathbb{N}$, cette somme correspond à la somme des termes d'une série à termes positifs, cette somme est donc finie ou bien infinie. En effet, c'est la limite de la suite croissante $(S_n)_{n\in\mathbb{N}}$ où :

$$S_n = \sum_{i=0}^{n} i\, p_i\,.$$

Cette suite est convergente si et seulement si elle est bornée et sinon elle croît jusqu'à dépasser n'importe quelle valeur finie.

L'espérance d'une variable aléatoire représente une caractéristique de tendance centrale de cette variable.

exemple 5.17. Notons Y le temps total d'attente du bus dans l'exemple 5.2. Le temps moyen d'attente est :

$$\mathbb{E}(Y) = (0 + 1 + \cdots + 10) \times \frac{1}{11} = 5\text{mn}\,.$$

Test : espérance de la loi $\mathcal{B}(p)$

Montrer qu'elle vaut p.

definition 5.18 (Variance d'une variable discrète). Si X est une variable aléatoire discrète dont l'espérance est finie, alors, quand elle existe, la variance de la variable X est

l'*espérance de la variable* $Y = (X - \mathbb{E}X)^2$. Elle est notée $\mathrm{var}(X)$ et vérifie donc :

$$\mathrm{var}(X) = \sum_{i \in D} (i - \mathbb{E}X)^2 \times p_i \,.$$

Quand cette somme est finie on dit que *la variable X est de variance finie*.

REMARQUE 5.19. Lorsque l'ensemble D est de cardinal fini, la somme définissant la variance de la variable X est toujours définie. Lorsque $D = \mathbb{N}$, cette somme correspond à la somme des termes d'une série à termes positifs, cette somme est donc finie ou bien infinie.

La variance d'une variable aléatoire représente une caractéristique de dispersion autour de la caractéristique de tendance centrale qu'est l'espérance.

EXEMPLE 5.20.

1. Pour l'exemple 5.2, la variable X a pour espérance la valeur $4/11$ et pour variance la valeur $28/121$.
2. Considérons une classe de 30 étudiants qui passent un examen. Notons x_i la note de l'étudiant i pour i entre 1 et 30. Maintenant, tirons un étudiant au hasard (uniformément) et notons X sa note, qui est une variable aléatoire à valeur dans $D = \{x_1, \ldots, x_{30}\}$ (pas forcément uniformément distribuée, car il est possible que plusieurs étudiants aient la même note). L'espérance de X est alors $(x_1 + \cdots + x_{30})/30$ (qui est la moyenne de la classe à l'examen). Considérons deux situations : dans le premier cas, les 30 étudiants ont tous eu 10. Dans ce cas, $\mathbb{E}(X) = 10$ et

   $$\mathrm{var}(X) \;=\; \sum_{i=1}^{30} (x_i - \mathbb{E}(X))^2 \frac{1}{30} \;=\; \sum_{i=1}^{30} (10 - 10)^2 \frac{1}{30} \;=\; 0 \,.$$

   La variance est nulle (tout le monde est exactemet à la moyenne, il n'y a aucune variabilité). Second cas, la moitié des étudiants a eu 0 et l'autre 20. La moyenne est toujours égale à 10, ce qui montre que réduire l'ensemble des notes à leur simple moyenne ne permet pas de distinguer deux situations pourtant très différentes (dans un cas, tout le monde a moyennement réussi l'examen, dans l'autre, la moitié de la classe s'est plantée et l'autre moitié a excellé). Calculons la variance : que l'étudiant i ait eu $x_i = 0$ ou $x_i = 20$, dans les deux cas $(x_i - \mathbb{E}(X))^2 = (x_i - 10)^2 = 10^2$. Ainsi,

   $$\mathrm{var}(X) \;=\; \sum_{i=1}^{30} (x_i - \mathbb{E}(X))^2 \frac{1}{30} \;=\; \sum_{i=1}^{30} \frac{10^2}{30} \;=\; 100 \,.$$

   La variance est très grande : les notes sont très éloignées de la moyenne.

Test : variance de la loi $\mathcal{B}(p)$

Montrer qu'elle vaut $p(1-p)$.

L'espérance est la meilleure approximation par une valeur constante de la variable

X au sens des moindres carrés, ce qui signifie la chose suivante :

proposition 5.21 (Espérance et moindres carrés). *Lorsque la variable X est de variance finie, alors $\mathbb{E}(X)$ est l'unique valeur en laquelle la fonction*

$$c \mapsto \sum_{i\in D}(i-c)^2 p_i\,,$$

de $\mathbb{R}$ dans $\mathbb{R}_+$, atteint son minimum.

Démonstration. On ne présente la preuve que dans le cas où D est fini (le cas infini dénombrable est similaire mais demande de manipuler des sommes infinies). On considère la fonction f donnée par

$$f(c) = \sum_{i\in D}(i-c)^2 p_i = c^2 \sum_{i\in D} p_i - 2c \sum_{i\in D} i p_i + \sum_{i\in D} i^2 p_i \;=\; c^2 - 2\alpha c + \beta\,.$$

où l'on a utilisé que $\sum_{i\in D} p_i = 1$ et posé

$$\alpha = \sum_{i\in D} i p_i \qquad \text{et} \qquad \beta = \sum_{i\in D} i^2 p_i\,.$$

La question est donc de minimiser un trinôme en c, qui tend vers $+\infty$ en $\pm\infty$ et dont la dérivée $f'(c) = 2(c-\alpha)$ s'annule en un seul point, α, qui est donc nécessairement l'unique point où f atteint son minimum. □

Énoncé indispensable 74 : Propriétés des moyenne et variance

Soit $(\Omega, \mathbb{P})$ un espace probabilisé fini ou dénombrable, et soient X et Y deux variables aléatoires sur Ω à valeur dans $\mathbb{N}$.

1. Si X et Y sont d'espérances finies, alors $X+Y$ l'est également et
$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$
2. la variable aléatoire X est de variance finie si et seulement si la variable aléatoire X^2 est d'espérance finie. De plus, dans ce cas on a l'identité :
$$\mathrm{var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2\,.$$

Démonstration. On ne donne la démonstration que dans le cas fini. Pour $i \in \mathbb{N}$ dans l'image de X,

$$i \times p_i \;=\; i \times \mathbb{P}(X=i) \;=\; \sum_{\omega\in\Omega,\ X(\omega)=i} X(\omega)\mathbb{P}(\omega)\,.$$

On en déduit une écriture alternative pour l'espérance :

$$\mathbb{E}(X) \;=\; \sum_{i\in\mathbb{N}} i p_i \;=\; \sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\omega)\,.$$

En appliquant cette égalité à la variable aléatoire $X+Y$, on obtient

$$\begin{aligned}\mathbb{E}(X+Y) &= \sum_{\omega\in\Omega}(X(\omega)+Y(\omega))\,\mathbb{P}(\omega)\\ &= \sum_{\omega\in\Omega}X(\omega)\mathbb{P}(\omega)+\sum_{\omega\in\Omega}Y(\omega)\mathbb{P}(\omega) \;=\; \mathbb{E}(X)+\mathbb{E}(Y)\,.\end{aligned}$$

Pour démontrer la seconde affirmation, on développe le carré :

$$\begin{aligned}\mathrm{var}(X) \;=\; \sum_{i\in\mathbb{N}}(i-\mathbb{E}(X))^2p_i &= \sum_{i\in\mathbb{N}}i^2p_i-2\mathbb{E}(X)\sum_{i\in\mathbb{N}}ip_i+(\mathbb{E}(X))^2\sum_{i\in\mathbb{N}}p_i\\ &= \mathbb{E}(X^2)-2\,(\mathbb{E}(X))^2+(\mathbb{E}(X))^2\\ &= \mathbb{E}(X^2)-(\mathbb{E}(X))^2\end{aligned}$$

□

Voici les espérances et variances des lois discrètes usuelles :

PROPOSITION 5.22.

Loi	Espérance	Variance
Uniforme $\mathcal{U}(\{1,2,\dots,n\})$	$(n+1)/2$	$(n^2-1)/12$
Uniforme $\mathcal{U}(\{0,1,\dots,n\})$	$n/2$	$n(n+2)/12$
Bernoulli $\mathcal{B}(p)$	p	$p(1-p)$
Binomiale $\mathcal{B}(N,p)$	Np	$Np(1-p)$
Géométrique $\mathcal{G}(p)$	$1/p$	$(1-p)/p^2$
Poisson $\mathcal{P}(\lambda)$	λ	λ

Démonstration. Pour les lois uniformes on utilise les identités vraies pour $n\in\mathbb{N}^*$:

$$\sum_{k=1}^{n}k=n(n+1)/2 \quad\text{et}\quad \sum_{k=1}^{n}k^2=n(n+1)(2n+1)/6$$

(qu'on peut vérifier par récurrence).

Pour une variable aléatoire X suivant une loi de Bernoulli, le calcul de l'espérance est directe, et on remarque que $X^2=X$, de sorte que

$$\mathrm{var}(X) \;=\; \mathbb{E}(X^2)-(\mathbb{E}X)^2 \;=\; p-p^2\,.$$

Pour une loi binomiale, on peut utiliser le résultat des énoncés indispensables 72 et 74 en écrivant

$$X=\sum_{i=1}^{N}X_i\,,$$

où X_i est le résultat du i^{e} lancer qui suit une loi de Bernoulli (avec X_i indépendant de X_j si $i\neq j$). On obtient immédiatement l'espérance par

$$\mathbb{E}(X) \;=\; \mathbb{E}\left(\sum_{i=1}^{N}X_i\right) \;=\; \sum_{i=1}^{N}\mathbb{E}(X_i) \;=\; Np\,.$$

Pour la variance, on développe le carré,

$$X^2 = \sum_{i=1}^{N} X_i^2 + \sum_{i=1}^{N} \left(\sum_{j=1, j\neq i}^{N} X_i X_j \right) .$$

On a déjà vu que $X_i^2 = X_i$ (qui est donc d'espérance p). Si $i \neq j$, alors $X_i X_j$ vaut 0 ou 1 avec

$$\mathbb{P}(X_i X_j = 1) \;=\; \mathbb{P}(X_i = 1 \text{ et } X_j = 1) \;=\; \mathbb{P}(X_i = 1)\mathbb{P}(X_j = 1) \;=\; p^2$$

(où l'on a utilisé l'indépendance de X_i et X_j). Autrement dit, $X_i X_j$ suit la loi $\mathcal{B}(p^2)$ (d'espérance p^2). En utilisant la linéarité de l'espérance (énoncé indispensable 74),

$$\mathbb{E}(X^2) \;=\; \sum_{i=1}^{N} \mathbb{E}(X_i^2) + \sum_{i=1}^{N} \left(\sum_{j=1, j\neq i}^{N} \mathbb{E}(X_i X_j) \right) \;=\; Np + N(N-1)p^2$$

(en effet, dans la double somme en i et j, tous les termes sont égaux à p^2, il suffit juste de compter le nombre de termes : or i prend N valeurs et, pour chaque i, j prend $N-1$ valeurs (puisque la valeur i est exclue). En conclusion, pour une variable de loi $\mathcal{B}(N, p)$

$$\mathrm{var}(X) \;=\; \mathbb{E}(X^2) - (\mathbb{E}X)^2 \;=\; Np + N(N-1)p^2 - (Np)^2 \;=\; N(p - p^2) .$$

Le cas de la loi géométrique sera vu en séances de TD.
Enfin, pour une variable X de loi $\mathcal{P}(\lambda)$,

$$\mathbb{E}(X) \;=\; \sum_{n\geqslant 0} n \times \frac{\lambda^n}{n!} e^{-\lambda} \;=\; \sum_{n\geqslant 1} \frac{\lambda^n}{(n-1)!} e^{-\lambda} \;=\; \lambda e^{-\lambda} \sum_{n\geqslant 1} \frac{\lambda^{n-1}}{(n-1)!}$$

(remarquons qu'on a changé le premier indice de la somme : en effet, le terme $n \times \mathbb{P}(n)$ est nulle pour $n = 0$, on peut donc l'exclure). En faisant le changement d'indice $m = n - 1$, on trouve

$$\mathbb{E}(X) \;=\; \lambda e^{-\lambda} \sum_{m\geqslant 0} \frac{\lambda^m}{m!} \;=\; \lambda e^{-\lambda} e^{\lambda} \;=\; \lambda .$$

Le calcul de la variance est similaire. □

Chapitre 6

Variables aléatoires à densité

De nombreux phénomènes aléatoires produisent des valeurs qui ne sont pas des entiers : une température, une longueur, un angle... Autant de quantités représentées par des nombres réels quelconques. Prenons par exemple une toupie, qu'on fait tourner et dont on regarde, lorsqu'elle s'arrête, l'angle qu'elle forme avec une direction fixée au préalable. Cette angle peut prendre n'importe quelle valeur entre 0 et 2π radians, et on aimerait pouvoir dire qu'aucune direction n'est privilégiée. Pour modéliser cela, il faudrait pouvoir définir ce que signifie "tirer un nombre au hasard" dans un continuum (par exemple, l'intervalle $[0, 2\pi]$). On ne peut pas vraiment se fixer une valeur précise et se demander quelle est le probabilité de tomber exactement sur cette valeur (cette probabilité serait nulle tout le temps puisqu'il y a une infinité de valeurs possibles). On peut seulement dire qu'en coupant l'intervalle en deux moitiés de même longueur, par exemple $[0, \pi]$ et $[\pi, 2\pi]$, il devrait y avoir autant de chances d'être dans l'une que dans l'autre. En poursuivant ce raisonnement (en divisant l'intervalle en 3, 4, etc. sous-intervalles), on en déduit que, plus généralement, on voudrait dire que la probabilité de se trouver dans un secteur donné est proportionel à la longueur de ce secteur. On en arrive donc à définir la probabilité de se trouver dans n'importe quel intervalle (non réduit à un point). C'est le principe des variables aléatoires à densité.

6.1 Définitions.

definition 6.1. On appelle variable aléatoire réelle une application $X : \Omega \to D$ où Ω est un ensemble muni d'une probabilité $\mathbb{P}$ et $D \subset \mathbb{R}$.

Pour aller plus loin 2 :

Jusqu'ici on a uniquement défini les mesures de probabilités sur des ensembles finis ou dénombrables. Or, dans ce chapitre, on voudrait considérer des ensembles Ω infinis non dénombrables (typiquement $\Omega = \mathbb{R}$). Cependant, dans ce cas, la définition précédente d'une probabilité (à savoir, une application de l'ensemble des parties de Ω dans $[0, 1]$ qui satisfait certaines conditions) ne s'avère pas satisfaisante pour diverses raisons. Il faut en réalité restreindre $\mathbb{P}$ a un sous-ensemble de l'ensemble de parties de Ω (par exemple, dans $\mathbb{R}$, on pourrait dire qu'on ne définit $\mathbb{P}(A)$ que si A est un intervalle ; toutes les parties de $\mathbb{R}$ ne sont pas des intervalles).

Il y a alors des parties de Ω dont on ne peut dire la probabilité. Nous n'aborderons pas cette question dans ce cours de première année et admettrons que dans les cas que nous considérerons il n'y aura pas de problème (rappelons qu'en pratique, Ω est rarement explicité) et que toutes les propriétés des mesures de probabilités vue précédemment (Proposition 4.4, énoncé indispensable 70...) s'étendent au cas général.

definition 6.2. Soit X une variable aléatoire réelle. On appelle fonction de répartition de X la fonction $F : \mathbb{R} \to [0,1]$ définie par

$$\forall t \in \mathbb{R}, \qquad F(t) \;=\; \mathbb{P}(X \leq t)\,.$$

remarque 6.3. La fonction de répartition permet de calculer la probabilité que X se trouve dans un intervalle $]a,b]$ quels que soient les réels $a < b$. En effet, $]a,b] =]-\infty,b]\backslash]-\infty,a]$, et donc

$$\mathbb{P}(a < X \leq b) \;=\; \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a) \;=\; F(b) - F(a)\,.$$

remarque 6.4. Si $s < t$, on a l'inclusion $]-\infty,s] \subset]-\infty,t]$. Par croissance des probabilités, on en déduit que $F(s) \leqslant F(t)$. La fonction de répartition est donc toujours une fonction croissante.

definition 6.5. Soit $f : \mathbb{R} \to \mathbb{R}$ continue par morceaux. On dit que f est une densité de probabilité si elle est positive et telle que $\int_{-\infty}^{+\infty} f(t)dt = 1$.

remarque 6.6. L'intégrale $\int_{-\infty}^{+\infty} f(t)dt$ est à comprendre comme la limite de $\int_x^y f(t)dt$ lorsque $x \to -\infty$ et $y \to +\infty$. Par exemple, considérons la fonction f définie par $f(t) = 1/t^2$ pour $t \in [1,+\infty[$ et $f(t) = 0$ pour $t < 1$. Pour tous réels x, y avec $x < 0 < y$,

$$\int_x^y f(t)dt \;=\; \int_1^y \frac{1}{t^2}dt \;=\; \left[-\frac{1}{t}\right]_1^y \;=\; 1 - \frac{1}{y}\,,$$

qui tend vers 1 quand $y \to +\infty$ et $x \to -\infty$ (en fait ça ne dépend pas de x puisque f est nulle sur $]-\infty,1]$), ce qu'on écrit $\int_{-\infty}^{+\infty} f(t)dt = 1$.

Comme pour les sommes infinis, la manipulation rigoureuse d'intégrales sur des intervalles non bornés est délicate. Aussi, on prendra soin de revenir à cette définition lorsqu'il s'agira, par exemple, d'effectuer un changement de variable ou une integration par partie, c'est-à-dire qu'il conviendra de le faire sur l'intégrale sur un intervalle borné $[x,y]$ avant de faire tendre x et y vers plus ou moins l'infini.

definition 6.7. On dira qu'une variable aléatoire réelle X, de fonction de répartition F, admet une densité $f : \mathbb{R} \to \mathbb{R}_+$ si f est une densité de probabilité et si, pour tout $x \in \mathbb{R}$,

$$F(x) \;=\; \int_{-\infty}^{x} f(t)dt$$

ou, de manière équivalente, si pour tous $a < b$ réels,

$$\mathbb{P}(a < X \leq b) \;=\; \int_a^b f(t)dt\,.$$

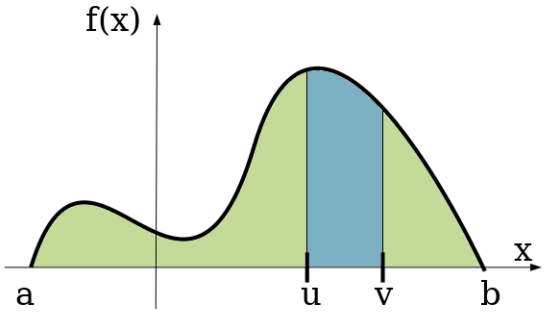


FIGURE 6.1 – Si f est la densité de probabilité d'une variable aléatoire X, la probabilité que X se trouve dans le sous-intervalle $[u, v]$ est donné par l'aire sous la courbe sur $[u, v]$.

REMARQUE 6.8. Si X admet une densité f, alors la probabilité que X soit exactement égal à une valeur fixée $a \in \mathbb{R}$ est nulle quelque soit a. En effet, pour tout $\varepsilon > 0$,

$$0 \;\leq\; \mathbb{P}(X = a) \;\leq\; \mathbb{P}(X \in]a-\varepsilon, a]) \;=\; \int_{a-\varepsilon}^{a} f(t)dt \;\underset{\varepsilon\to 0}{\longrightarrow} 0\,,$$

et donc $\mathbb{P}(X = a) = 0$. En conséquence, pour tous réels $a < b$,

$$\mathbb{P}(X \in [a, b]) = \mathbb{P}(X = a) + \mathbb{P}(X \in]a, b]) = \mathbb{P}(X \in]a, b])$$

et de même, plus généralement,

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X < b)\,.$$

6.2 Quelques densité de probabilités.

6.2.1 *Loi uniforme sur un intervalle*

Étant donnés deux réels $a < b$, on dit qu'une variable aléatoire réelle X suit la loi uniforme sur $[a, b]$ si elle admet comme densité de probabilité la fonction f définie par

$$f(x) \;=\; \begin{cases} \frac{1}{b-a} & \text{si } x \in [a, b] \\ 0 & \text{sinon.} \end{cases}$$

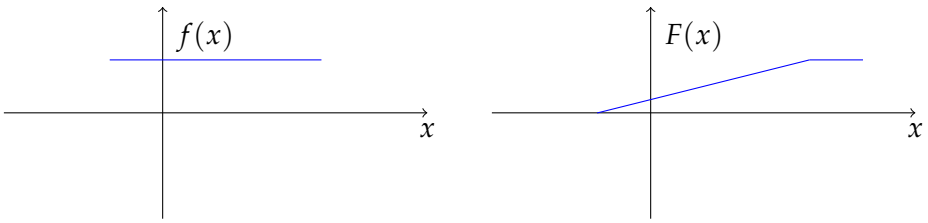
On notera cela : $X \sim \mathcal{U}([a, b])$.

Vérifions que f est bien une densité de probabilité. C'est bien une fonction positive et continue par morceaux. Il reste à voir que son intégrale est égale à 1. Puisque f est

nulle en-dehors de $[a, b]$,

$$\int_{-\infty}^{+\infty} f(t)dt \;=\; \int_{a}^{b} f(t)dt \;=\; \int_{a}^{b} \frac{1}{b-a}dt \;=\; \frac{1}{b-a}[t]_a^b \;=\; \frac{b-a}{b-a} = 1\,.$$

De même, on calcule la fonction de répartition F de X comme suit : si



exemple 6.9. La loi uniforme répond à la question : *comment tirer un nombre au hasard entre a et b ?* (en l'absence de précision, implicitement, on suppose qu'aucune valeur n'est privilégiée). Pour générer des nombres aléatoires à l'aide d'un ordinateur, par défaut, on génère des variables aléatoires de loi uniforme sur $[0, 1]$, à partir desquels peuvent ensuite être obtenues par diverses transformations des variables aléatoires distribuées selon d'autres lois. (En réalité, en toute rigueur, l'ordinateur produit des nombres dit pseudo-aléatoires qui approchent l'idéal théorique de la loi uniforme sur $[0, 1]$.)

6.2.2 *Loi exponentielle*

Étant donné $\lambda > 0$ (lettre grecque *lambda*), on dit qu'une variable aléatoire réelle X suit la loi exponentielle de paramètre λ si elle admet la densité f définie par

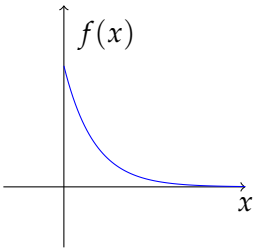
$$f(x) \;=\; \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{sinon.} \end{cases}$$

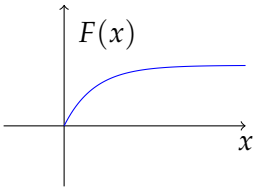
On notera cela : $X \sim \mathcal{E}(\lambda)$.

Vérifions que f est bien une densité de probabilité, et calculons la fonction de répartition associée. Puisque f est nulle sur $]-\infty, 0[$, pour tout $x \geq 0$,

$$F(x) \;=\; \int_{-\infty}^{x} f(t)dt \;=\; \int_{0}^{x} \lambda e^{-\lambda t}dt \;=\; \left[-e^{-\lambda t}\right]_0^x = 1 - e^{-\lambda x}\,,$$

qui tend bien vers 1 lorsque x tend vers $+\infty$.





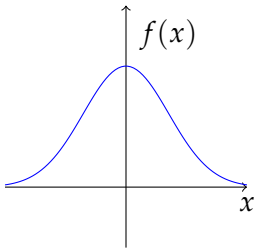
exemple 6.10. La loi exponentielle est à la loi géométrique ce que la loi de Poisson est à la loi binomiale, autrement dit une version "temps continu". En effet, la loi exponentielle modélise typiquement le temps écoulé avant qu'un événement ne survienne, mais à la différence avec la loi géométrique, le temps est mesuré continuement. Par exemple, regardons le temps que va mettre un atome radioactif à se désintégrer. On peut l'observer tous les jours et noter le nombre de jours qui s'écoulent avant la désintégration : ce sera une variable aléatoire de loi géométrique, de paramètre p la probabilité que la désintégration ait lieu chaque jour. Mais on pourrait observer l'atome toutes les heures, et noter le nombre d'heures qui s'écoulent avant la désintégration. Ce sera à nouveau une variable aléatoire de loi géométrique, mais cette fois-ci de paramètre $p/24$. Mais on pourrait également observer l'atome toutes les minutes, ou toutes les secondes, etc. Si on observe en continu l'atome et qu'on note le temps (avec une précision arbitraire) écoulé avant la désintégration, alors ce temps est une variable aléatoire de loi exponentielle (et, comme on le verra plus bas, le paramètre λ est donné par l'inverse du temps moyen avant désintégration).

6.2.3 *Loi normale (gaussienne)*

Étant donné $\mu \in \mathbb{R}$ et $\sigma > 0$ (lettres grecques *mu* et *sigma*), on dit qu'une variable aléatoire réelle X suit la loi normale (ou gaussienne) de paramètres (μ, σ^2) si elle admet la densité f définie par

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

On notera cela : $X \sim \mathcal{N}(\mu, \sigma^2)$.



Comme on le verra dans le chapitre suivant, la loi normale a un rôle crucial en probabilités du fait du théorème central limite. Dans les grandes lignes, ce dernier stipule que, lorsque de nombreux facteurs aléatoires s'additionnent, alors le résultat suit approximativement une loi normale.

La vérification que f est bien une densité de probabilité est plus délicate que pour les exemples précédents. On en donne à titre indicatif une démonstration basée sur des outils élémentaires. Commençons par traiter le cas $\mu = 0$ et $\sigma^2 = 1/2$. La fonction f étant continue et positive, il reste à vérifier qu'elle est d'intégrale 1, ce qui est équivalent à montrer que

$$\int_{-\infty}^{+\infty} e^{-x^2} dx \;=\; \sqrt{\pi}\,.$$

D'abord, remarquons que pour tout $x \in]-1, +\infty[$, $\ln(1+x) \leq x$, ce qui peut être obtenu soit en étudiant la fonction $g(x) = x - \ln(1+x)$ (elle est décroissante sur $x \in]-1, 0]$ puis croissante, et nulle en 0, donc positive), soit en déduisant de la concavité de $x \mapsto \ln(1+x)$ qu'elle est en-dessous de sa tangente en o. Cette inégalité, appliquée avec $x = -t^2/n$ et $x = t^2/n$ pour tout $n \in \mathbb{N}_*$ et tout $t \in [0, \sqrt{n}[$ implique que

$$\ln\left(1 - \frac{t^2}{n}\right) \;\leq\; -\frac{t^2}{n}\,, \qquad \text{et } \ln\left(1 + \frac{t^2}{n}\right) \;\leq\; \frac{t^2}{n}\,.$$

En multipliant ces inégalités respectivement par n et $-n$ et en prenant l'exponentielle, on obtient

$$\forall n \in \mathbb{N}_*,\ \forall t \in [0, \sqrt{n}[, \qquad \left(1 - \frac{t^2}{n}\right)^n \;\leq\; e^{-t^2} \;\leq\; \left(1 + \frac{t^2}{n}\right)^n\,,$$

et donc, pour tout $n \in \mathbb{N}_*$

$$\int_0^{\sqrt{n}} \left(1 - \frac{t^2}{n}\right)^n dt \;\leq\; \int_0^{\sqrt{n}} e^{-t^2} dt \;\leq\; \int_0^{\sqrt{n}} \left(1 + \frac{t^2}{n}\right)^n dt\,.$$

En effectuant respectivement les changements de variables $t = \sqrt{n}\cos(u)$ pour le terme de gauche et $t = \sqrt{n}\cot(u)$ dans la seconde (rappel : la fonction cotangente est définie par $\cot(u) = \cos(u)/\sin(u)$), on obtient

$$\sqrt{n}\int_0^{\frac{\pi}{2}} \sin^{2n+1}(u) du \;\leq\; \int_0^{\sqrt{n}} e^{-t^2} dt \;\leq\; \sqrt{n}\int_0^{\frac{\pi}{2}} \sin^{2n-2}(u) du\,.$$

Les intégrales des termes de gauche et de droite sont ce qu'on appelle des intégrales de Wallis, dont l'étude est un exercice classique d'intégration. Pour l'instant, admettons le résultat suivant :

$$\sqrt{n}\int_0^{\frac{\pi}{2}} \sin^n(u) du \;\underset{n\to+\infty}{\longrightarrow}\; \sqrt{\frac{\pi}{2}}\,.$$

Ainsi,

$$\begin{aligned}\sqrt{n}\int_0^{\frac{\pi}{2}}\sin^{2n+1}(u)du &= \sqrt{\frac{1}{2+\frac{1}{n}}}\times\sqrt{2n+1}\int_0^{\frac{\pi}{2}}\sin^{2n+1}(u)du\\ &\underset{n\to+\infty}{\longrightarrow} \frac{1}{\sqrt{2}}\times\sqrt{\frac{\pi}{2}} = \frac{\sqrt{\pi}}{2},\end{aligned}$$

et avec le même argument on a également

$$\sqrt{n}\int_0^{\frac{\pi}{2}}\sin^{2n-2}(u)du \underset{n\to+\infty}{\longrightarrow} \frac{\sqrt{\pi}}{2}.$$

En conséquence, par le théorème des gendarmes,

$$\int_0^{\sqrt{n}} e^{-t^2}dt \underset{n\to+\infty}{\longrightarrow} \frac{\sqrt{\pi}}{2}.$$

D'autre part, la fonction $t\mapsto e^{-t^2}$ étant paire, le changement de variable $u=-t$ donne

$$\begin{aligned}\int_{-\sqrt{n}}^{\sqrt{n}} e^{-t^2}dt &= \int_{-\sqrt{n}}^{0} e^{-t^2}dt + \int_0^{\sqrt{n}} e^{-t^2}dt\\ &= \int_0^{\sqrt{n}} e^{-u^2}du + \int_0^{\sqrt{n}} e^{-t^2}dt\\ &= 2\int_0^{\sqrt{n}} e^{-t^2}dt \underset{n\to+\infty}{\longrightarrow} 2\frac{\sqrt{\pi}}{2} = \sqrt{\pi},\end{aligned}$$

ce qui conclut le cas $\mu=0$, $\sigma^2=1/2$.

Dans le cas général, le changement de variable $y=(x-\mu)/(\sqrt{2}\sigma)$ nous ramène au cas précédent.

Pour être complet, voici la démonstration du résultat sur les intégrales de Wallis que nous avons admis. Pour tout entier n, notons $W_n=\int_0^{\pi/2}\sin^n(u)du$. En intégrant deux fois par partie, on voit que ces intégrales satisfont la relation de récurrence $W_{n+2}=(n+1)/(n+2)\times W_n$. En particulier, pour tout $n\geqslant 2$,

$$W_nW_{n+1}(n+1) = \frac{n-1}{n}W_{n-2}\frac{n}{n+1}W_{n-1}(n+1) = W_{n-2}W_{n-1}(n-1).$$

Autrement dit, la suite $n\mapsto W_nW_{n+1}(n+1)$ est constante, égale à $W_0W_1=\pi/2$. D'autre part, $n\mapsto W_n$ est une suite décroissante (car $\sin(t)\in[0,1]$ pour $t\in[0,\pi/2]$), autrement dit $R_n=W_n/W_{n+1}$ est supérieure à 1. La relation de récurrence satisfaite par les W_n implique que $R_nR_{n+1}=(n+2)/(n+1)\to 1$ quand $n\to\infty$. De même, $R_n/R_{n+2}=(n^2+2n+1)/(n^2+2n)\geqslant 1$. Les deux suites $(u_{2n})_{n\in\mathbb{N}}$ et $(u_{2n+1})_{n\in\mathbb{N}}$ sont donc décroissante et minorée par 1, donc convergent vers des limites qu'on note a et b

(qui sont nécessairement $\geqslant 1$). Ainsi, quand $n \to \infty$, $R_n R_{n+1}$ converge vers ab et vers 1, donc nécessairement $ab = 1$, et donc $a = b = 1$. En conclusion,

$$nW_n^2 \;=\; nW_nW_{n-1}/R_{n-1} \underset{n\to\infty}{\longrightarrow} \frac{\pi}{2}\,.$$

On conclut en prenant la racine carrée.

6.3 Espérance et variance.

definition 6.11. Soit X est une variable aléatoire de densité $f : \mathbb{R} \to \mathbb{R}_+$.

— On définit son espérance, notée $\mathbb{E}(X)$, par

$$\mathbb{E}(X) \;=\; \int_{-\infty}^{+\infty} x f(x)dx$$

sous réserve que cette intégrale est bien définie.

— Plus généralement, pour $g : \mathbb{R} \to \mathbb{R}$, on définit l'espérance de $g(X)$, notée $\mathbb{E}(g(X))$, par

$$\mathbb{E}\left(g(X)\right) \;=\; \int_{-\infty}^{+\infty} g(x)f(x)dx$$

si l'intégrale est bien définie.

— Si $\mathbb{E}(X)$ et $\mathbb{E}(X^2)$ sont bien définies, alors on définit la variance de X, notée $\mathrm{var}(X)$, par

$$\mathrm{var}(X) \;=\; \mathbb{E}(X^2) - \left(\mathbb{E}(X)\right)^2 \;=\; \int_{-\infty}^{+\infty} x^2 f(x)dx - \left(\int_{-\infty}^{+\infty} x f(x)dx\right)^2 .$$

Comme dans le cas de variable aléatoire discrète, l'espérance est la valeur moyenne de X, et la variance mesure l'éloignement de X à cette moyenne.

On admettra le résultat suivant :

proposition 6.12. *Soient X et Y deux variables aléatoires réelles, et $a, b \in \mathbb{R}$. Sous réserve que les espérances soient bien définies, on a les relations suivantes :*

$$\begin{aligned} \mathbb{E}(aX + bY) &= a\mathbb{E}(X) + b\mathbb{E}(Y) \\ \mathrm{var}(aX) &= a^2\mathrm{var}(X) \\ \mathrm{var}(X) &= \mathbb{E}\left(|X - \mathbb{E}(X)|^2\right) \end{aligned}$$

D'autre part, si $X \geqslant Y$, alors $\mathbb{E}(X) \geqslant \mathbb{E}(Y)$.

La dernière égalité implique en particulier que la variance est toujours positive.

Voici les espérances et variances des lois à densité usuelles :

proposition 6.13.

Loi	Espérance	Variance
Uniforme $\mathcal{U}([a,b])$	$(a+b)/2$	$(b-a)^2/12$
Exponentielle $\mathcal{E}(\lambda)$	$1/\lambda$	$1/\lambda^2$
Normale $\mathcal{N}(m,\sigma^2)$	m	σ^2

Démonstration. **Loi uniforme.** Si $X \sim \mathcal{U}([a,b])$, comme la densité est nulle en-dehors de $[a,b]$ il n'y a pas de problème de définitions des intégrales, et

$$\mathbb{E}(X) \;=\; \int_a^b x f(x)dx = \frac{1}{b-a}\left[\frac{1}{2}x^2\right]_a^b \;=\; \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}\,.$$

Pour la variance, on calcule

$$\int_a^b x^2 f(x)dx = \frac{1}{b-a}\left[\frac{1}{3}x^3\right]_a^b \;=\; \frac{b^3-a^3}{3(b-a)} = \frac{b^2+ab+a^2}{3}\,,$$

et donc

$$\begin{aligned}\mathrm{var}(X) &= \frac{b^2+ab+a^2}{3} - \left(\frac{a+b}{2}\right)^2 \\ &= \frac{4(b^2+ab+a^2)-3(a^2+2ab+b^2)}{12} \;=\; \frac{(b-a)^2}{12}\,.\end{aligned}$$

Loi exponentielle. Si $X \sim \mathcal{E}(\lambda)$, la densité étant nulle sur $]-\infty,0]$, on calcule

$$\begin{aligned}\int_{-R}^{R} x f(x)dx &= \int_0^R \lambda x e^{-\lambda x}dx \\ &\overset{ipp}{=} \left[-xe^{-\lambda x}\right]_0^R - \int_0^R \left(-e^{-\lambda x}\right)dx \\ &= -Re^{-\lambda R} + \left[-\frac{1}{\lambda}e^{-\lambda x}\right]_0^R \\ &= -Re^{-\lambda R} - \frac{1}{\lambda}e^{-\lambda R} + \frac{1}{\lambda} \;\underset{R\to+\infty}{\longrightarrow}\; \frac{1}{\lambda}\,.\end{aligned}$$

Autrement dit, $\mathbb{E}(X) = 1/\lambda$. Pour la variance, en utilisant le résultat du calcul précédent,

$$\begin{aligned}\int_0^R \lambda x^2 e^{-\lambda x}dx &\overset{ipp}{=} \left[-x^2e^{-\lambda x}\right]_0^R - \int_0^R 2x \times \left(-e^{-\lambda x}\right)dx \\ &= -R^2e^{-\lambda R} + 2\int_0^R x e^{-\lambda x}dx \;\underset{R\to+\infty}{\longrightarrow}\; \frac{2}{\lambda^2}\,,\end{aligned}$$

et donc

$$\mathrm{var}(X) \;=\; \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 \;=\; \frac{1}{\lambda^2}\,.$$

Loi normale (gaussienne). Ce cas est laissé en exercice au lecteur (indications : on peut se ramener au cas $\mu = 0$ et $\sigma^2 = 1$ par changement de variable. Ensuite, $xe^{-x^2/2}$ se primitive en $-e^{-x^2/2}$, et l'intégrale de $x^2e^{-x^2/2}$ se calcule par intégration par partie en décomposant $x^2e^{-x^2/2} = x \times xe^{-x^2/2}$).

□

Chapitre 7

Variables aléatoires indépendantes et suites de variables aléatoires

7.1 Inégalités de Markov et Bienaymé-Tchebychev

Énoncé indispensable 75 : Inégalité de Markov

Si X est une variable aléatoire réelle **positive** telle que $\mathbb{E}(X)$ est bien définie, alors pour tout $r > 0$,

$$\mathbb{P}(X \geq r) \ \leq \ \frac{\mathbb{E}(X)}{r}.$$

Démonstration. Notons g la fonction définie par $g(x) = r$ si $x \geq r$ et $g(x) = 0$ si $x < r$. Remarquons que g est une variable aléatoire discrète (elle ne prend que les valeurs r et 0), et $\{g(X) = r\} = \{X \geqslant r\}$ et $\{g(X) = 0\} = \{X < r\}$. D'autre part, $g(x) \leq x$ pour tout $x \geq 0$ et par croissance de l'espérance,

$$\mathbb{E}(X) \geq \mathbb{E}(g(X)) = 0 \times \mathbb{P}(X < r) + r \times \mathbb{P}(X \geq r)$$

et donc

$$\mathbb{P}(X \geq r) \ \leq \ \frac{1}{r}\mathbb{E}(X).$$

□

Un corollaire direct de l'inégalité de Markov est l'inégalité de Bienaymé-Tchebychev (il suffit en fait de connaître la première, et de là la seconde se retrouve immédiatement).

proposition 7.1 (Inégalité de Bienaymé-Tchebychev). *Si X est une variable aléatoire réelle telle que $\mathbb{E}(X)$ et $\mathrm{var}(X)$ sont bien définies, alors pour tout $r > 0$,*

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq r) \ \leq \ \frac{\mathrm{var}(X)}{r^2}.$$

Démonstration. En appliquant l'inégalité de Markov à la variable $Z = (X - \mathbb{E}(X))^2$, on

obtient

$$\mathbb{P}\left(|X - \mathbb{E}(X)| \geq r\right) \;=\; \mathbb{P}\left(|X - \mathbb{E}(X)|^2 \geq r^2\right) \;\leq\; \frac{\mathrm{var}(X)}{r^2}\,.$$

□

7.2 Variables aléatoires indépendantes.

definition 7.2. On dit que deux variables aléatoires réelles X et Y sont indépendantes si pour toutes fonctions f et g de $\mathbb{R}$ and $\mathbb{R}$ telles que les espérances soient bien définies,

$$\mathbb{E}\left(f(X)g(Y)\right) \;=\; \mathbb{E}\left(f(X)\right)\mathbb{E}\left(g(Y)\right)$$

remarque 7.3. Le lien avec la définition d'événements indépendants ou avec la définition de variables indépendantes dans le cas discret (définition 5.9 du chapitre 5) est le suivant. Si X et Y sont deux variables aléatoires indépendantes et que I et J sont deux intervalles de $\mathbb{R}$, alors les événements $\{X \in I\}$ et $\{Y \in J\}$ sont indépendants. On le voit en effet en considérant les fonctions f_I et g_J définies par $f_I(x) = 1$ si $x \in I$ et $f_I(x) = 0$ sinon, et de même pour g_J. Ainsi, $f_I(X)g_J(Y)$ est soit nul, soit égal à 1, et

$$\mathbb{E}\left(f_I(X)g_J(Y)\right) \;=\; \mathbb{P}(X \in I,\; Y \in J)$$

tandis que

$$\mathbb{E}\left(f_I(X)\right) \;=\; \mathbb{P}(X \in I)\,, \qquad \mathbb{E}\left(g_J(Y)\right) \;=\; \mathbb{P}(Y \in J)\,.$$

Donc l'indépendance de X et Y implique que

$$\mathbb{P}(X \in I,\; Y \in J) \;=\; \mathbb{E}\left(f_I(X)g_J(Y)\right) \;=\; \mathbb{E}\left(f_I(X)\right)\mathbb{E}\left(g_J(Y)\right) \;=\; \mathbb{P}(X \in I)\mathbb{P}(Y \in J)\,,$$

c'est-à-dire l'indépendance des événements. Pour des variables aléatoires discrètes, les deux sont équivalents, au sens où X et Y sont indépendants si et seulement si les événements $\{X = x\}$ et $\{Y = y\}$ sont indépendants pour tout x dans l'image de X et y dans l'image de Y.

exemple 7.4.

1. Soient X et Y deux variables indépendantes distribuées selon la loi uniforme sur $\{-1, 1\}$. Ainsi, $Z = XY$ peut prendre deux valeurs, 1 et -1. Calculons sa loi.

$$\begin{aligned}\mathbb{P}(Z = 1) &= \mathbb{P}(X = Y = 1) + \mathbb{P}(X = Y = -1)\\ &= \mathbb{P}(X = 1)\mathbb{P}(Y = 1) + \mathbb{P}(X = -1)\mathbb{P}(Y = -1)\\ &= \frac{1}{2}\times\frac{1}{2} + \frac{1}{2}\times\frac{1}{2} \;=\; \frac{1}{2}\,,\end{aligned}$$

et donc

$$\mathbb{P}(Z = -1) \;=\; 1 - \mathbb{P}(Z = 1) \;=\; \frac{1}{2}\,.$$

Ainsi le produit de deux variables uniformes sur $\{-1, 1\}$ *indépendantes* est égale-

ment uniforme sur $\{-1,1\}$. Ce serait faux si X et Y n'étaient pas indépendantes. Par exemple, si $Z = X$, alors X et Z sont bel et bien deux variables uniformément distribuée sur $\{-1,1\}$, pourtant $XZ = X^2$ est toujours égal à 1 (et n'est donc pas uniformément distribuée sur $\{-1,1\}$). De même, si $Z = -X$, c'est toujours une variable uniforme sur $\{-1,1\}$, mais $XZ = -X^2$ est toujours égal à -1.

2. Si on lance successivement N fois une pièce dont la probabilité de faire pile est $p \in]0,1[$ et qu'on note $X_i = 1$ si le i^{eme} lancer fait pile et $X_i = 0$ sinon, alors X_i et X_j sont indépendants si $i \neq j$. D'autre part, pour chaque i entre 1 et N, $X_i \sim \mathcal{B}(p)$. Remarquons que le nombre total de fois où l'on a obtenu pile sur les N lancers est $X = \sum_{i=1}^{N} X_i$. L'énoncé indispensable 72 peut donc se réinterpréter ainsi : *la somme de N variables de Bernoulli de paramètres p indépendantes suit une loi binomiale de paramètres* (N,p).

3. Attention, il n'est pas suffisant que $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ pour que X et Y soient indépendantes ! Voici un contre-exemple : on lance, indépendemment et l'une après l'autre, deux pièces équilibrées. On pose $X = 0$ si la première pièce fait face, $X = 1$ si la première fait pile et la seconde face, et $X = -1$ si les deux font pile. On pose $Y = 0$ si la première pièce fait pile, $Y = 1$ si la première fait face et la seconde pile, $Y = -1$ si les deux font face. Alors X et Y ont la même loi, donnée par

$$\mathbb{P}(X=0) \;=\; \frac{1}{2}, \qquad \mathbb{P}(X=1) \;=\; \mathbb{P}(X=-1) \;=\; \frac{1}{4}.$$

En particulier,

$$\mathbb{E}(X) \;=\; \frac{1}{2}\times 0 + \frac{1}{4}\times 1 + \frac{1}{4}\times(-1) \;=\; 0,$$

et de même pour Y, de sorte que $\mathbb{E}(X)\mathbb{E}(Y) = 0$. D'un autre côté, si la première pièce fait face, $X = 0$, et si elle fait pile, $Y = 0$, de sorte que dans tous les cas $XY = 0$ et donc $\mathbb{E}(XY) = 0 = \mathbb{E}(X)\mathbb{E}(Y)$. Cependant, X et Y ne sont pas indépendantes. Intuitivement, on voit bien que connaître X donne des informations sur Y : si X est nul, nécessairement Y ne l'est pas, et inversement. Plus formellement, considérons les événements $\{X = 0\}$ et $\{Y = 0\}$, qui correspondent respectivement à "la première pièce fait face" et "la première pièce fait pile". Alors

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

Les deux variables ne sont donc pas indépendantes.

REMARQUE 7.5. On dit que des variables aléatoires réelles X et Y sont **non corrélées** si $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. Comme le montre le contre-exemple précédent, des variables non corrélées ne sont pas forcément indépendantes ! La réciproque est vraie, en revanche (des variables indépendantes sont nécessairement non corrélées), puisqu'il suffit d'appliquer la définition de l'indépendance avec $f(x) = x$ et $g(y) = y$.

Énoncé indispensable 76 :

Si X et Y sont deux variables aléatoires réelles indépendantes, alors

$$\mathrm{var}(X+Y) \;=\; \mathrm{var}(X)+\mathrm{var}(Y)$$

Plus généralement, si $X_1,\dots,X_n$ sont des variables aléatoires réelles deux à deux indépendantes (c'est-à-dire telles que X_i est indépendant de X_j si $i \neq j$) alors

$$\mathrm{var}\left(\sum_{i=1}^{n} X_i\right) \;=\; \sum_{i=1}^{n} \mathrm{var}(X_i)\,.$$

Démonstration. L'indépendance de X et Y implique que $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. En développant les carrés et en utilisant la linéarité de l'espérance, on a donc

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

Le cas $n > 2$ s'obtient de la même manière, ou par récurrence. □

exemple 7.6.

1. Vérifions à la main la propriété dans un cas simple. On lance indépendemment n pièces de 1 euro et k pièces de 2 euros, toutes équilibrées ($p = 1/2$). On note X le nombre de pile parmi les pièces d'1 euro et Y le nombre de pile parmi les pièces de 2 euro. On sait donc que $X \sim \mathcal{B}(n, 1/2)$ et $Y \sim \mathcal{B}(k, 1/2)$. En même temps, si on fait abstraction de la valeur des pièces, on a lancé $n+k$ pièces équilibrées et obtenu $X+Y$ pile, donc on sait que $X+Y \sim \mathcal{B}(n+k, 1/2)$. On connaît la variance de la loi binomiale :

   $$\mathrm{var}(X+Y) \;=\; (n+k)p(1-p) \;=\; np(1-p)+kp(1-p) \;=\; \mathrm{var}(X)+\mathrm{var}(Y)\,.$$

   En fait, puisque l'on sait que la somme de n variables de loi de Bernoulli $\mathcal{B}(p)$ indépendantes suit la loi binomiale $\mathcal{B}(n,p)$, la variance de la loi binomiale s'obtient directement à partir de la variance de la loi de Bernoulli grâce à l'énoncé indispensable 76.

2. La condition d'indépendance est indispensable, le résultat est complétement faux sinon. Considérons par exemple $X \sim \mathcal{N}(0,1)$ et $Y = -X$. Alors $X+Y = 0$, donc $\mathbb{E}(X+Y) = 0 = \mathbb{E}((X+Y)^2)$ donc $\mathrm{var}(X+Y) = 0$, alors que $\mathrm{var}(Y) = (-1)^2\mathrm{var}(X) = 1$ de sorte que $\mathrm{var}(X)+\mathrm{var}(Y) = 2 \neq 0$. Évidemment, sur cet exemple, Y n'est pas indépendant de X.

7.3 Loi des Grands Nombres

Énoncé indispensable 77 : Loi (faible) des grands nombres (LGN)

On considère une suite $(X_k)_{k\in\mathbb{N}}$ de variables aléatoires réelles. On suppose que pour tout k, $\mathbb{E}(X_k)$ et $\mathrm{var}(X_k)$ sont bien définies et respectivement égaux à $\mathbb{E}(X_1)$ et $\mathrm{var}(X_1)$ (c'est-à-dire qu'elles ne dépendent pas de k), et d'autre part que pour tout $i, j \in \mathbb{N}$ avec $i \neq j$, X_i est indépendant de X_j. Alors pour tout $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}(X_1)\right| \geq \varepsilon\right) \underset{n\to+\infty}{\longrightarrow} 0\,.$$

Démonstration. Notons $Z_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Par linéarité de l'espérance (Proposition 6.12), $\mathbb{E}(Z_n) = \mathbb{E}(X_1)$, et d'après l'énoncé indispensable 76, $\mathrm{var}(Z_n) = n\mathrm{var}(X_1)$. En appliquant l'inégalité de Bienaymé-Tchebychev,

$$\begin{aligned}\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}(X_1)\right| \geq \varepsilon\right) &= \mathbb{P}\left(|Z_n - \mathbb{E}(Z_n)| \geq \varepsilon\right)\\ &\leq \frac{\mathrm{var}(X_1)}{\varepsilon^2 n} \underset{n\to+\infty}{\longrightarrow} 0\,.\end{aligned}$$

□

Intuitivement, la Loi des Grands Nombres signifie que, lorsqu'on répéte un grand nombre de fois la même expérience et qu'on fait la moyenne sur tous les essais, on s'approche de la moyenne théorique. C'est la base des **estimations statistiques**.

exemple 7.7. Une élection se profile, et on aimerait anticiper la proportion d'abstention à venir, c'est-à-dire qu'on aimerait connaître la probabilité p qu'un électeur tiré au hasard (uniformément parmi tous les électeurs) aille voter. Il y a trop d'électeurs pour leur demander à tous s'ils iront voter (on suppose pour simplifier que tous les électeurs savent déjà s'ils iront voter et qu'ils répondent franchement quand on leur pose la question), donc p est inconnue. On décide d'interroger au hasard (uniformément) $n = 1000$ électeurs (beaucoup moins que le nombre total d'électeurs ; c'est ce qu'on appelle un *échantillon*). On note $X_i = 1$ si le i^{eme} électeur ira voter, $X_i = 0$ sinon. Les X_i sont des variables de loi $\mathcal{B}(p)$ indépendantes, d'espérance p. La Loi des Grands Nombres stipule que la moyenne observée sur l'échantillon (appelé moyenne empirique), c'est-à-dire $1/n\sum_{i=1}^{n} X_n$, est probablement proche de la vraie moyenne théorique, $\mathbb{E}(X_1) = p$. Par exemple, si 28% des n sondés déclarent qu'ils iront voter, on peut *estimer* que p se situe aux alentours de 28%, et la probabilité qu'on soit loin de la réalité (si on n'a pas eu de chances, on est tombé sur un échantillon non représentatif) diminue quand le nombre de sondés n augmente.

exemple 7.8. On lance n fois le même dé et on compte le nombre de 6 obtenus, c'est-à-dire qu'on note $X_i = 1$ si le i^{eme} lancer donne 6 et $X_i = 0$ sinon, et $Z_n = \sum_{i=1}^{n} X_i/n$ la fréquence de 6 observée. On s'attend à avoir obtenu des 6 environ une fois sur 6, de sorte que Z_n devrait être de l'ordre de $1/6 \simeq 17\%$. Regardons la probabilité qu'on ait observé

moins de 10% de 6, c'est-à-dire la probabilité que $\{Z_n \leqslant 10\%\}$. Notons $\varepsilon = 1/6 - 10\%$ de sorte que, si $Z_n \leqslant 10\%$, alors $|Z_n - \mathbb{E}(X_1)| \geqslant \varepsilon$. Par croissance des probabilités,

$$\mathbb{P}(Z_n \leqslant 10\%) \;\leqslant\; \mathbb{P}(|Z_n - \mathbb{E}(X_1)| \geqslant \varepsilon) \underset{n\to+\infty}{\longrightarrow} 0$$

car $\varepsilon > 0$. Autrement dit, la probabilité d'observer une fréquence anormalement faible de 6 (moins de 10%) tend vers 0 à mesure que le nombre d'expérience augmente (cf figure 7.1).

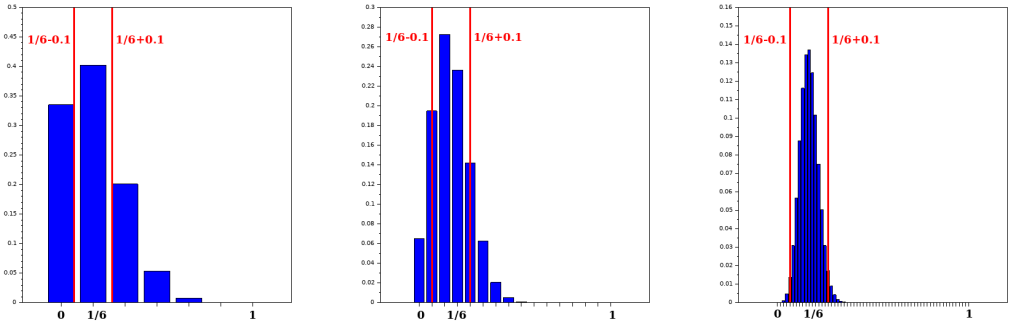


Figure 7.1 – On lance n fois un dé équilibré et l'on note Z_n la fréquence de 6 observé. Ces trois histogrammes représente la loi de Z_n pour $n = 6$ (à gauche) $n = 15$ (au milieu) et $n = 60$ (à droite). La valeur théorique attendue est $1/6$, et les lignes rouges représentent $1/6 \pm 0.1$. Autrement dit, l'aire de l'histogramme contenu entre les deux lignes représente la probabilité de se trouver entre $1/6 - 0.1$ et $1/6 + 0.1$. On voit qu'à mesure que n augmente l'histogramme se concentre vers la moyenne théorique.

remarque 7.9. En fait, si l'on regarde la démonstration de la Loi des Grands Nombres, on se rend compte qu'on a une information plus précise que simplement la convergence vers 0 : on a une borne quantitative

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}(X_1)\right| \geq \varepsilon\right) \;\leqslant\; \frac{\mathrm{var}(X_1)}{\varepsilon^2 n}\,.$$

On peut donc majorer explicitement la probabilité de voir une déviation par rapport à la moyenne théorique attendue, en fonction du nombre de lancers.

Reprenons l'exemple précédent, avec $\varepsilon = 0.1$, on obtient que

$$\mathbb{P}\left(|Z_n - 1/6| \geqslant 0.1\right) \;\leqslant\; \frac{1/6(1-1/6)}{(0.1)^2 n} \;\simeq\; \frac{14}{n}\,.$$

Si on lance par exemple $n = 100$ dés et qu'on regarde la fréquence de 6, on sait que la probabilité que cette fréquence soit plus que $1/6 + 0.1$ ou moins que $1/6 - 0.1$ est plus petite que 14%, autrement dit on a plus que 86% de chances que la fréquence soit dans l'intervalle $[1/6 - 0.1, 1/6 + 0.1]$.

exemple 7.10. Vous jouez avec votre petit cousin à pile ou face : il lance sa pièce porte-

bonheur. Si elle tombe sur pile, vous lui donnez un euro, si elle tombe sur face, il vous donne un euro. Vous avez joué 30 fois de suite et, la pièce ayant fait 20 fois pile, vous devenez soupçonneux : la pièce est-elle truquée ? Dit autrement, quelle était la probabilité que la pièce fasse plus de 20 fois pile sur les 30 coups ? D'après la Loi des Grands Nombre, si vous jouez un grand nombre de fois, le nombre moyen de pile devrait converger vers $1/2$ (si la pièce est équilibrée). Mais 30 parties n'est peut-être pas suffisant : après tout, il y a même une probabilité positive que la pièce fasse 30 fois pile d'affilée (pas une grosse probabilité ceci dit : $1/2^{30}$, de l'ordre de gagner au loto). La question est-donc : "sous l'hypothèse que la pièce est équilibrée, est-il franchement improbable qu'on observe plus de 20 pile sur les 30 lancer ?" Dans ce problème spécifique très simple on pourrait calculer explicitement cette probabilité (avec la loi binomiale), mais une méthode plus générale consiste à simplement majorer cette probabilité par l'inégalité de Bienaymé-Tchebychev (comme dans la démonstration de la Loi des Grands Nombres, ou pour les intervalles de confiance). Notons $X_i = 1$ si la pièce a fait pile au i^{eme} lancer pour i entre 1 et 30. Les X_i sont indépendants et, en supposant que la pièce est équilibrée, suivent une loi $\mathcal{B}(1/2)$. On s'attend à une fréquence moyenne de pile égale à $\mathbb{E}(X_i)$, et donc la question est d'estimer la probabilité d'observer une déviation supérieure à $\varepsilon = 2/3 - 1/2 = 1/6$ par rapport à cette valeur attendue. On a donc

$$\mathbb{P}\left(\sum_{i=1}^{30} X_i \geqslant 20\right) = \mathbb{P}\left(\frac{1}{30}\sum_{i=1}^{30} X_i \geqslant \frac{2}{3}\right) \leqslant \mathbb{P}\left(\left|\frac{1}{30}\sum_{i=1}^{30} X_i - \frac{1}{2}\right| \geqslant \varepsilon\right)$$

Si la pièce est équilibrée, la variance de X_i est $1/2(1-1/2) = 1/4$, et donc par l'inégalité de Bienaymé-Tchebychev,

$$\mathbb{P}\left(\sum_{i=1}^{30} X_i \geqslant 20\right) \leqslant \frac{\mathrm{var}(X_1)}{30\varepsilon^2} = \frac{1/4}{30 \times 1/6} = \frac{3}{10}.$$

Finalement, tout ce que vous pouvez dire à ce stade, c'est qu'il y avait moins de 3 chances sur 10 d'observer autant de fois pile. Ça reste assez élevé, en particulier si vous accusez votre cousin de tricherie la seule chose que vous savez c'est que vous avez au plus 3 chances sur 10 de vous tromper.

La conclusion aurait été différente si vous aviez fait plus de parties, mettons cent fois plus, et observé la même proportion de $2/3$ de pile. En effet, en reprenant les notations et calculs précédents,

$$\mathbb{P}\left(\sum_{i=1}^{3000} X_i \geqslant 2000\right) \leqslant \frac{\mathrm{var}(X_1)}{3000\varepsilon^2} = \frac{3}{1000}.$$

Si vous accusez votre cousin d'être un tricheur, vous avez moins de 3 chances sur 1000 de vous tromper et d'accuser à tort un innocent dont le seul crime est d'avoir eu de la chance sur ce coup-là.

Pour aller plus loin 3 :

Cet exemple est le prototype de ce qu'on appelle les **tests statistiques**, qui interviennent dans une foule de contextes où se posent des questions du type "on a fait

une étude statistique, on a eu telle observation, peut-on en tirer telle conclusion ?". Par exemple

- On a donné à n patients un traitement A, et à n autres patients un traitement B. On observe que 32% du premier groupe a guéri, contre 25% dans le second groupe. Peut-on en déduire que le traitement A est plus efficace, ou était-ce simplement l'effet du hasard ?
- On reçoit d'un fournisseur des clous qui sont censés peser en moyenne 3 grammes. On pèse au hasard n clous et on trouve une masse moyenne de 2.7 grammes. Doit-on renvoyer les clous ?
- On mesure sur n personnes l'activité électrique dans une zone donnée du cerveau pendant une conversation. On observe que, en moyenne sur tous les sujets et sur toutes les conversations, l'activité est 5% plus élevée (par rapport à l'activité moyenne sur toute la conversation) lorsque le sujet parle. Peut-on en déduire quoi que ce soit ?

7.4 Théorème central limite

On admettra sans démonstration le résultat suivant :

Énoncé indispensable 78 : Théorème central limite (TCL)

On considère une suite $(X_k)_{k\in\mathbb{N}}$ de variables aléatoires réelles. On suppose que pour tout k, $\mathbb{E}(X_k)$ et $\mathrm{var}(X_k)$ sont bien définies et respectivement égaux à $\mathbb{E}(X_1)$ et $\mathrm{var}(X_1)$ (c'est-à-dire qu'elles ne dépendent pas de k), et d'autre part que pour tout $i, j \in \mathbb{N}$ avec $i \neq j$, X_i est indépendant de X_j. Alors pour tous réels $a < b$,

$$\mathbb{P}\left(\frac{\sqrt{n}}{\sqrt{\mathrm{var}(X_1)}}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}(X_1)\right) \in [a,b]\right) \underset{n\to+\infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}dx\,.$$

remarque 7.11. C'est le théorème qui est central, pas la limite, il n'y a donc pas de e à central.

remarque 7.12. D'après la loi des grands nombres, la moyenne empirique (observée sur n expériences) $\frac{1}{n}\sum_{i=1}^{n} X_i$ converge (en un sens ; ça reste une variable aléatoire) vers la moyenne théorique $\mathbb{E}(X_1)$, autrement dit on s'attend à ce que $\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}(X_1)$ tende (en un sens) vers 0. Cependant, dans le TCL, on multiplie cette quantité par $\sqrt{n}$ qui, lui, tend vers $+\infty$. Or le produit des deux ne converge ni vers 0, ni ves $+\infty$. Autrement dit, les fluctuations des observations par rapport à la théorie, $\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}(X_1)$, sont de l'ordre de $1/\sqrt{n}$.

remarque 7.13. Remarquons que $(2\pi)^{-1/2}\int_a^b e^{-\frac{x^2}{2}}dx$ est, par définition, la probabilité qu'une variable de loi $\mathcal{N}(0,1)$ (qu'on appelle gaussienne centrée réduite, ou standard : sa moyenne est nulle et sa variance est 1) se trouve dans l'intervalle $[a,b]$. Autrement dit,

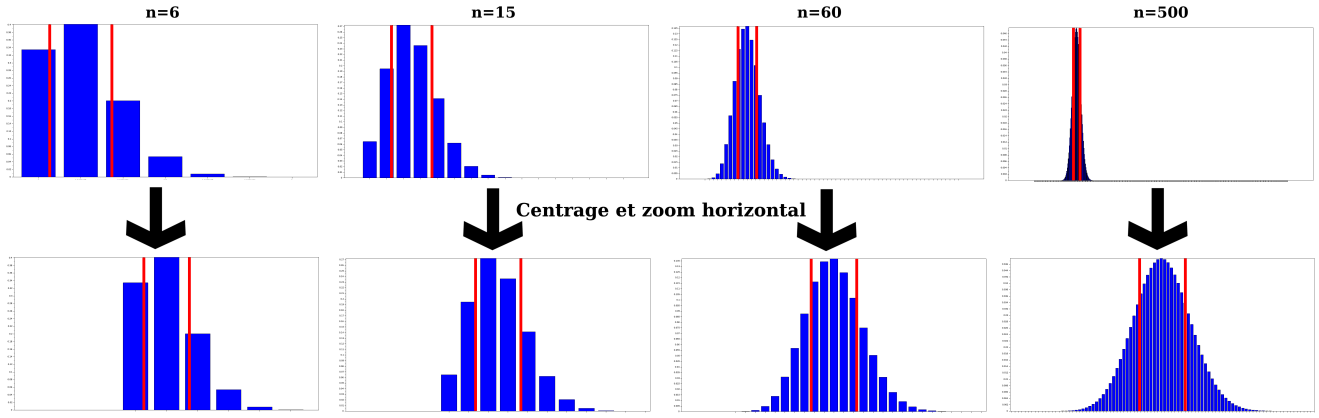


FIGURE 7.2 – La première ligne représente la même chose que la figure 7.1, sauf qu'il y a maintenant 4 valeurs différentes pour n (6, 15, 60 et 500) et, surtout, que les lignes rouges ne sont plus fixées à $1/6 \pm 0.1$, mais sont placées en $1/6 \pm 0.3/\sqrt{n}$ (elles se reserrent donc à vitesse $1/\sqrt{n}$ vers la moyenne théorique quand n augmente). La seconde ligne représente exactement les mêmes histogrammes sauf qu'ils ont été centrés (la moyenne théorique $1/6$ étant placé au milieu) et leur axe horizontal a été zoomé d'un facteur $\sqrt{n}$ de sorte que les lignes rouges (qui servent de repères) apparaissent à la même position pour les 4 valeurs de n. On voit que, ainsi renormalisé, les histogrammes tendent vers celui de la loi normale.

ce que décrit le TCL c'est que, quand n est grand, la variable

$$Y_n = \frac{\sqrt{n}}{\sqrt{\mathrm{var}(X_1)}} \left(\frac{1}{n} \sum_{i=1}^{n} X_i - \mathbb{E}(X_1) \right),$$

qui décrit les fluctuations de la moyenne empirique $\frac{1}{n}\sum_{i=1}^{n} X_i$ par rapport à la moyenne théorique $\mathbb{E}(X_1)$, est approximativement gaussienne :

$$\mathbb{P}\left(Y_n \in [a,b]\right) \underset{n\to+\infty}{\simeq} \mathbb{P}(Y \in [a,b])$$

avec $Y \sim \mathcal{N}(0,1)$. Remarquons qu'en particulier, Y et Y_n ont la même moyenne et la même variance : par linéarité de l'espérance,

$$\mathbb{E}(Y_n) = \frac{\sqrt{n}}{\sqrt{\mathrm{var}(X_1)}} \left(\mathbb{E}\left(\frac{1}{n} \sum_{i=1}^{n} X_i \right) - \mathbb{E}(X_1) \right) = \frac{\sqrt{n}}{\sqrt{\mathrm{var}(X_1)}} \left(\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left(X_i\right) - \mathbb{E}(X_1) \right) = 0$$

et d'après les propriétés de la variance (Proposition 6.12 et énoncé indispensable 76)

$$\mathrm{var}(Y_n) = \frac{n}{\mathrm{var}(X_1)} \mathrm{var}\left(\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left(X_i\right) - \mathbb{E}(X_1) \right) = \frac{n}{\mathrm{var}(X_1)} \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{var}(X_i) = 1\,.$$

De manière informelle non rigoureuse, on pourrait dire que la moyenne observée sur n expériences (pour n grand) est égal à la moyenne théorique plus une fluctuation

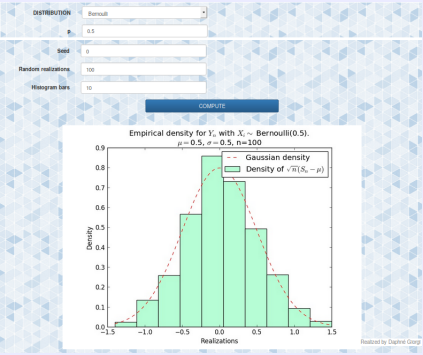
gaussienne d'ordre $1/\sqrt{n}$:

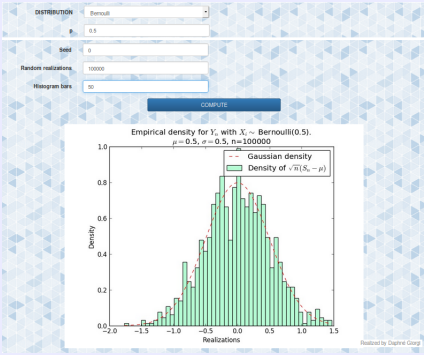
$$\frac{1}{n}\sum_{i=1}^{n} X_i \underset{n\to+\infty}{\simeq} \mathbb{E}(X_1) + \frac{\sqrt{\mathrm{var}(X_1)}}{\sqrt{n}} Y$$

avec $Y \sim \mathcal{N}(0,1)$. C'est dans le cadre de cette approximation que, dans de nombreuses situations (et ce dans toutes les sciences), on modélise des variables aléatoires par des gaussiennes. En effet, d'après le TCL, la loi gaussienne apparaît dès lors que de nombreux effets indépendants s'additionnent.

Pour aller plus loin 4 :

On trouve sur le site du Laboratoire de Probabilités, Statistiques et Modélisation (LPSM) de Sorbonne Université une page de simulations numériques, en particulier pour le Théorème Central Limite. On choisit la loi des X_i (par exemple Bernoulli de tel paramètre, ou Normale de tels paramètres, etc.), le nombre n d'expériences, et la simulation génère n variables aléatoires de la loi désirée et en représente l'histogramme (il faut préciser le nombre de barres qu'on veut pour l'histogramme) ainsi que la densité théorique de la loi gaussienne. Vous pouvez allez faire l'expérience et constater que, plus n est grand, plus l'histogramme ressemble à la densité gaussienne (il faut augmenter le nombre de bares également pour gagner en précision, tout en le gardant beaucoup plus petit que le nombre de réalisations). Vous pouvez aussi faire un tour sur les autres simulations de la page (il n'est pas toujours nécessaire de maîtriser les notions mathématiques en jeu pour trouver les résultats jolis...).





L'un des intérêts principaux du Théorème Central Limite est de fournir ce qu'on appelle des **intervalles de confiance** (plus précisément des intervalles de confiance asymptotiques, au sens où le TCL énonce un résultat pour $n \to \infty$).

exemple 7.14. Reprenons l'exemple 7.7 sur les sondages. Notons p la proportion théorique inconnue que l'on cherche à estimer, et $\hat{p}_n = 1/n \sum_{i=1}^{n} X_i$ l'estimation qu'on en fait basée sur un échantillon de n sondés. On aimerait donner une barre d'erreur à l'estimation, plus précisément on cherche une valeur δ telle que la probabilité que p soit dans l'intervalle $[\hat{p}_n - \delta, \hat{p}_n + \delta]$ soit, disons, 95% (on parlera d'un intervalle de confiance à

95%). Autrement dit, on cherche $\delta > 0$ tel que

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - p\right| > \delta\right) \;=\; 5\%\,.$$

D'après le TCL,

$$\begin{aligned}\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - p\right| > \delta\right) \quad &= \quad \mathbb{P}\left(\frac{\sqrt{n}}{\sqrt{\mathrm{var}(X_1)}}\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}(X_1)\right| > \frac{\sqrt{n}}{\sqrt{\mathrm{var}(X_1)}}\delta\right)\\ &\underset{n\to+\infty}{\simeq} \quad \mathbb{P}\left(|Y| > \frac{\sqrt{n}}{\sqrt{\mathrm{var}(X_1)}}\delta\right)\end{aligned}$$

avec $Y \sim \mathcal{N}(0,1)$. Cette loi étant explicite, on peut calculer explicitement que

$$\mathbb{P}\left(|Y| > 1,96\right) \;\simeq\; 5\%\,.$$

Autrement dit, une variable gaussienne standard a 95% de chance de se trouver dans l'intervalle $[-1.96\ ,\ 1.96]$. En supposant que la taille n de l'échantillon est suffisamment grande pour que l'approximation fournie par le TCL soit correcte[1], la barre d'erreur δ que l'on cherche est donc telle que

$$\frac{\sqrt{n}}{\sqrt{\mathrm{var}(X_1)}}\delta \;=\; 1,96\,,$$

autrement dit

$$\delta \;=\; \frac{1,96\sqrt{\mathrm{var}(X_1)}}{\sqrt{n}}\,.$$

On peut donc estimer que la valeur théorique p a environ 95% de chances de se trouver entre les valeurs $[\hat{p}_n - \delta, \hat{p}_n + \delta]$ avec ce choix de δ.

Il y a cependant un problème : on ne connaît pas $\mathrm{var}(X_1)$. En effet, pour une variable de Bernoulli, la variance est $p(1-p)$, mais p est inconnu. Une première solution serait de majorer cette variance par $1/4$ (car $p(1-p)$ est toujours plus petit que $1/4$, quelque soit $p \in [0,1]$), auquel cas on surestime la marge d'erreur (on peut dire que la probabilité d'être dans l'intervalle est *au moins* de 95%). Une autre solution utilisée en pratique est de remplacer la valeur théorique de $\mathrm{var}(X_1)$ par une approximation basée sur les observations. Par exemple, ici, d'après la loi des grands nombres, $\hat{p}_n(1-\hat{p}_n)$ devrait être proche de $p(1-p)$, donc on peut remplacer l'expression précédente de δ par

$$\delta \;\simeq\; \frac{1,96\sqrt{\hat{p}_n(1-\hat{p}_n)}}{\sqrt{n}}\,.$$

On obtient un intervalle de confiance "doublement" asymptotique, au sens où l'on fait une erreur d'une part en utilisant le TCL, et d'autre part en remplaçant la variance par

1. On dit souvent que le TCL est une bonne approximation à partir de $n \geqslant 30$, mais en réalité cela dépend de la variance des X_i. Si la variance est grande, un échantillon bien plus grand peut s'avérer nécessaire pour commencer à observer une distribution gaussienne.

une estimation qui ne tend vers la vraie valeur que pour $n \to \infty$ [2]. D'un autre côté, maintenant, tout est explicitement calculable à partir des observations : si on a mesuré $\hat{p}_n = 28\%$ d'intention d'aller voter auprès d'un échantillon de $n = 1000$ électeurs, on a donc une barre d'erreur à 95% égale à

$$\delta \simeq \frac{1,96 \times \sqrt{0,28(1-0,28}}{\sqrt{1000}} \simeq 0,03\,.$$

Un intervalle de confiance (asymptotique) à 95% est donc donné par $[0.28 - 0.03\ ,\ 0.28 + 0.03] = [0.25\ ,\ 0.32]$. On peut donc estimer, au vu des observations, que la vraie proportion p d'électeurs qui vont aller voter se situe entre $0,25$ et $0,32$ avec probabilité 95%.

Si on avait observé une proportion $\hat{p}_n = 28\%$ mais sur un échantillon plus petit de $n = 50$ sondés, la barre d'erreur aurait été

$$\delta \simeq \frac{1,96 \times \sqrt{0,28(1-0,28}}{\sqrt{100}} \simeq 0,12\,.$$

L'intervalle de confiance à 95% aurait donc été $[0.28 \pm 0.12] = [0.16\ ,\ 0.40]$. La précision se déteriore évidemment quand la taille de l'échantillon diminue.

2. Le fait que l'on puisse, dans le TCL, remplacer la variance par une estimation sans toutefois modifier la conclusion du théorème, est un résultat théorique basé sur le Lemme de Slutsky, qu'on ne présentera pas ici.