

Bases mathématiques pour l'économie et la gestion

Bases mathématiques Pour l'économie et la gestion

-

Table des matières

PREMIERE PARTIE : QUELQUES OUTILS

Chapitre 1 : Traitement de systèmes d'équations

- 1.1. Résolution de systèmes par substitution page 5
- 1.2. Traitement de systèmes linéaires par la méthode dite de Gauss-Jordan page 8
- Solutions des exercices (Chapitre 1) page 12

Chapitre 2 : Matrices

- 2.1. Calcul matriciel : somme, produit, multiplication par un réel page 14
- 2.2. Inversion d'une matrice par la méthode de la matrice compagnon page 15
- 2.3. Calcul du déterminant d'une matrice carrée page 16
 - 2.3.1. Pour les matrices 1×1 , 2×2 , 3×3 page 16
 - 2.3.2. Pour les matrices quelconques page 17
- 2.4. Inversion d'une matrice par la méthode de la matrice adjointe page 20
- 2.5. Résoudre un système linéaire par la méthode de Cramer page 23
- Solutions des exercices (Chapitre 2) page 25

DEUXIEME PARTIE : VARIATIONS ET CALCUL DIFFERENTIEL

Chapitre 3 : Variations

- 3.1. Modèles à 2 variables page 29
 - 3.1.1. Variation absolue (ou en valeur) page 29
 - 3.1.2. Variation relative (ou en pourcentage) page 32
- 3.2. Modèles à 3 variables ou plus page 35
- Solutions des exercices (Chapitre 3) page 38

Chapitre 4 : Dérivées partielles de fonctions de plusieurs variables	
1. Technique de dérivation	page 40
2. Application : approximation linéaire	page 41
2.1. Méthode de Newton	page 41

Chapitre 5 : Différentielles	
1. Différentier une fonction	page 49
2. Equations aux variations différentielles	page 50
3. Systèmes d'équations aux variations différentielles	page 55

TROISIEME PARTIE : OPTIMISATION

Chapitre 6 : Optimisation	
1. Recherche d'extrema de fonctions d'une variable	page 66
2. Recherche d'extrema de fonctions de plusieurs variables	page 72
3. Recherche d'extrema de fonctions de plusieurs variables sous contrainte – Méthode de Lagrange	page 81
4. Vision graphique	page 85

Chapitre 7 :

Première partie
-
Quelques outils

Chapitre 1 :

Traitement de systèmes d'équations

1.1. Résolution de systèmes par substitution

Quand on travaille avec un **système d'équations**, on envisage plusieurs équations *simultanément* (l'accolade $\{$, regroupant les équations du système, est là pour nous le rappeler). **Résoudre** un système d'équations consiste à trouver les *solutions communes* à toutes les équations du système.

Une technique de résolution d'un système : la **substitution**.

Supposons qu'on ait affaire à un système de deux équations à deux inconnues x et y .

1. Utiliser l'une des deux équations pour exprimer une inconnue, disons x en fonction de l'autre, disons y .
2. Substituer x trouvé à l'étape 1 dans l'autre équation, pour obtenir une équation en y uniquement.
3. Trouver les solutions de l'équation en y obtenue à l'étape 2.
4. Substituer les valeurs de y trouvées à l'étape 3 dans l'équation de l'étape 1 pour trouver les valeurs correspondantes de x .

La méthode « substitution » peut être étendue à des systèmes à plus de deux inconnues, ou même à des systèmes indéterminés.

Exercice résolu

Résoudre le système en (x,y) :

$$\begin{cases} x + y^2 = 6 \\ x + 2y = 3 \end{cases}$$

Etape 1 : On peut utiliser la seconde équation pour exprimer x en fonction de y :

$$x = -2y + 3$$

Etape 2 : On substitue x trouvé à l'étape 1 dans la première équation :

$$(-2y + 3) + y^2 = 6$$

C'est-à-dire

$$y^2 - 2y - 3 = 0$$

Etape 3 : On résout l'équation de l'étape 2 par rapport à y :

$$\Delta = (-2)^2 - 4 \cdot 1 \cdot (-3) = 4 + 12 = 16$$

$$y = \frac{-(-2) + \sqrt{16}}{2 \cdot 1} = 3 \quad \text{ou} \quad y = \frac{-(-2) - \sqrt{16}}{2 \cdot 1} = -1$$

Etape 4 : On substitue les valeurs de y trouvées à l'étape 3 dans l'équation de l'étape 1 pour trouver les valeurs correspondantes de x :

$$\text{Si } y = 3, \text{ alors } x = -2 \cdot (3) + 3 = -3.$$

Si $y = -1$, alors $x = -2 \cdot (-1) + 3 = 5$.
Les solutions du système d'équations sont donc $(-3,3)$ et $(5,-1)$ (on note $S = \{(-3,3),(5,-1)\}$).

Exercices

1. Résoudre le système en (x,y) :

$$\begin{cases} x \cdot y = 0 \\ x + y = 1 \end{cases}$$

2. Résoudre le système en (x,y) :

$$\begin{cases} y = 2 \\ y = -x^2 - 2x \end{cases}$$

3. Résoudre le système en (x,y) :

$$\begin{cases} y - 1 = x^2 - 2x \\ y - 2x = -3 \end{cases}$$

4. Résoudre le système en (p, q_d, q_p) :

$$\begin{cases} q_d = 11 - p^2 \\ q_p = p - 1 \\ q_p = q_d \end{cases}$$

5. Résoudre le système en (p, p^*, q_d, q_p) :

$$\begin{cases} q_d = 17 - p^2 \\ q_p = p^* - 2 \\ p^* = p - 1 \\ q_p = q_d \end{cases}$$

6. Résoudre le système en (p, p^*, q_d, q_p) :

$$\begin{cases} q_d = 11 - p^2 \\ q_p = p^* - 2 \\ p^* = \frac{3}{4} p \\ q_p = q_d \end{cases}$$

7. Résoudre le système en (λ, x_1, x_2) :

$$\begin{cases} (x_1 - 4)^2 + (x_2 - 3)^2 = 25 \\ 6 - 2\lambda(x_1 - 4) = 0 \\ 8 - 2\lambda(x_2 - 3) = 0 \end{cases}$$

8. Résoudre le système en (x,y) :

$$\begin{cases} 3^x = 81y \\ 2^x = 16y \end{cases}$$

9. Résoudre le système en (x,y) :

$$\begin{cases} x + y = 7 \\ \log(x) + \log(y) = 1 \end{cases}$$

10. Résoudre le système en (x,y) :

$$\begin{cases} y - \log_3(x+2) = 1 \\ y = \log_3(x) + \log_3(x+4) \end{cases}$$

11. Résoudre le système en (x,y,z) :

$$\begin{cases} x - y + z = 2 \\ x y z = 0 \\ 2y + z = 1 \end{cases}$$

12. Résoudre le système en (x_1,x_2) :

$$\begin{cases} x_1^2 + x_2^2 = 25 \\ x_1^2 + (x_2 - 4)^2 = 9 \end{cases}$$

13. Résoudre le système en (x_1,x_2) :

$$\begin{cases} x_1^2 + x_2^2 = 25 \\ x_1^2 + x_2 = 19 \end{cases}$$

14. Résoudre le système en (λ, x_1, x_2) :

$$\begin{cases} (x_1 - 2)^2 + x_2^2 = 4 \\ x_2 - 2\lambda(x_1 - 2) = 0 \\ x_1 - 2\lambda x_2 = 0 \end{cases}$$

15. Résoudre le système en (x,y,z) :

$$\begin{cases} 5x + 4y + 10z = 20 \\ x + 2y + 2z = 4 \\ 4x + 4y + z = 4 \end{cases}$$

16. Résoudre le système en (x,z) :

$$\begin{cases} x - 2y + 1 = 0 \\ z = 3x + 5y^2 \end{cases}$$

1.2. Traitement de systèmes linéaires par la méthode

dite de GAUSS-JORDAN

La méthode de Gauss-Jordan est une méthode qui permet de traiter un système linéaire, quel que soit le nombre d'équations et de variables que ce système contienne.

Description de la méthode de GAUSS-JORDAN

Supposons que le système à résoudre soit un système de 3 équations linéaires à 3 variables (x,y,z) .

1. Ordonner le système à résoudre
$$\begin{cases} a_{1,1} \cdot x + a_{1,2} \cdot y + a_{1,3} \cdot z = b_1 \\ a_{2,1} \cdot x + a_{2,2} \cdot y + a_{2,3} \cdot z = b_2 \\ a_{3,1} \cdot x + a_{3,2} \cdot y + a_{3,3} \cdot z = b_3 \end{cases}$$

en faisant attention à faire apparaître un réel non nul à la place du terme $a_{1,1}$.
Il faut parfois permuter deux lignes du système.

2. Noter uniquement les coefficients , dans l'ordre :

$$\begin{array}{ccc|c} a_{1,1} & a_{1,2} & a_{1,3} & b_1 \\ a_{2,1} & a_{2,2} & a_{2,3} & b_2 \\ a_{3,1} & a_{3,2} & a_{3,3} & b_3 \end{array}$$

3. Faire apparaître 0 à la place du terme $a_{2,1}$
par une combinaison linéaire utilisant uniquement la ligne 2 et la ligne 1.
Faire apparaître 0 à la place du terme $a_{3,1}$
par une combinaison linéaire utilisant uniquement la ligne 3 et la ligne 1.

On obtient
$$\begin{array}{ccc|c} * & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \end{array} .$$

4. Faire apparaître 0 à la place du terme $a_{3,2}$
par une combinaison linéaire utilisant uniquement la ligne 3 et la ligne 2.

On obtient
$$\begin{array}{ccc|c} * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \end{array} .$$

5. Faire apparaître 0 à la place du terme $a_{2,3}$
par une combinaison linéaire utilisant uniquement la ligne 2 et la ligne 3.

Faire apparaître 0 à la place du terme $a_{1,3}$
par une combinaison linéaire utilisant uniquement la ligne 1 et la ligne 3.

On obtient
$$\left| \begin{array}{ccc|c} * & * & 0 & * \\ 0 & * & 0 & * \\ 0 & 0 & * & * \end{array} \right| .$$

6. Faire apparaître 0 à la place du terme $a_{1,2}$
par une combinaison linéaire utilisant uniquement la ligne 1 et la ligne 2.

On obtient
$$\left| \begin{array}{ccc|c} * & 0 & 0 & * \\ 0 & * & 0 & * \\ 0 & 0 & * & * \end{array} \right| .$$

7. Faire apparaître 1 à la place du terme $a_{1,1}$ par division de la ligne 1.
Faire apparaître 1 à la place du terme $a_{2,2}$ par division de la ligne 2.
Faire apparaître 1 à la place du terme $a_{3,3}$ par division de la ligne 3.
... (attention : pas de division par 0)

8. Si le système admet une solution unique, on obtient
$$\left| \begin{array}{ccc|c} 1 & 0 & 0 & r \\ 0 & 1 & 0 & s \\ 0 & 0 & 1 & t \end{array} \right| ,$$

ce qui est équivalent au système
$$\begin{cases} 1.x + 0.y + 0.z = r \\ 0.x + 1.y + 0.z = s \\ 0.x + 0.y + 1.z = t \end{cases} \Leftrightarrow \begin{cases} x = r \\ y = s \\ z = t \end{cases}$$

dont la solution unique évidente est le triple (r, s, t) .

Attention :

Lors de l'application de la méthode , si on obtient

- une ligne constituée de 0 alors le système est indéterminé ;
- une ligne constituée de 0 sauf pour le terme indépendant alors le système est impossible.

Exercice résolu

Traiter ce système en (x,y,z) grâce à la méthode de « Gauss-Jordan » :

$$\begin{cases} x + 2y - z = 2 \\ 2x - y + z = 3 \\ 3x + y - z = 2 \end{cases}$$

Résolution

$$\left| \begin{array}{ccc|c} 1 & 2 & -1 & 2 \\ 2 & -1 & 1 & 3 \\ 3 & 1 & -1 & 2 \end{array} \right| \quad \begin{array}{l} L_2 \rightarrow L_2 - 2L_1 \\ L_3 \rightarrow L_3 - 3L_1 \end{array}$$

$$\left| \begin{array}{ccc|c} 1 & 2 & -1 & 2 \\ 0 & -5 & 3 & -1 \\ 0 & -5 & 2 & -4 \end{array} \right| \quad L_3 \rightarrow L_3 - L_2$$

$$\left| \begin{array}{ccc|c} 1 & 2 & -1 & 2 \\ 0 & -5 & 3 & -1 \\ 0 & 0 & -1 & -3 \end{array} \right| \quad \begin{array}{l} L_1 \rightarrow L_1 - L_3 \\ L_2 \rightarrow L_2 + 3L_3 \end{array}$$

$$\left| \begin{array}{ccc|c} 1 & 2 & 0 & 5 \\ 0 & -5 & 0 & -10 \\ 0 & 0 & -1 & -3 \end{array} \right| \quad L_1 \rightarrow 5L_1 + 2L_3$$

$$\left| \begin{array}{ccc|c} 5 & 0 & 0 & 5 \\ 0 & -5 & 0 & -10 \\ 0 & 0 & -1 & -3 \end{array} \right| \quad \begin{array}{l} L_1 \rightarrow L_1 / 5 \\ L_2 \rightarrow -L_2 / 5 \\ L_3 \rightarrow -L_3 \end{array}$$

$$\left| \begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array} \right| \quad \text{est équivalent au système} \quad \begin{cases} x = 1 \\ y = 2 \\ z = 3 \end{cases}$$

Le système admet donc une solution unique : le triple $(1,2,3)$.

On note aussi $S = \{ (1,2,3) \}$.

Exercices

Grâce à la méthode de « GAUSS-JORDAN », trouvez l'ensemble des solutions des 6 systèmes suivants.

17.
$$\begin{cases} x + 2y + 3z = 14 \\ 2x - y + z = 3 \\ 3x + 2y - 4z = -5 \end{cases}$$

18.
$$\begin{cases} x + 2y + z = 3 \\ 2x + 3y + 4z = 5 \\ 4x + 7y + 6z = 9 \end{cases}$$

$$19. \begin{cases} x + y + 2z = 3 \\ 2x + y - z - 4 = 0 \\ 3x + 2y = 7 - z \\ 10 - 3z = 3y + 4x \end{cases}$$

$$20. \begin{cases} x + 2y - z + 3t = 0 \\ 2x + y + z - t = 0 \\ 4x + 5y - z + 5t = 0 \\ 5x + 4y + z + t = 0 \end{cases}$$

$$21. \begin{cases} x - y + 2z - t = -2 \\ 2x + y + z = 4 \\ -3x - z + 2t = 5 \\ x + y + t = 8 \end{cases}$$

$$22. \begin{cases} x + 2y - z = 4 \\ x + 3y = 7 - z \\ 2x - 4z = 5 - 3y \\ 2x + 1 = 8z - y \end{cases}$$

23. Peut-on exprimer y_1 et y_2 en fonction de x ?

$$\begin{cases} 2y_1 + 3y_2 + x = 5 \\ 4y_1 + 6y_2 - x = 0 \end{cases}$$

24. Soit le système de 3 équations à 6 variables r, s, t, u, v, w :

$$\begin{cases} 2r + 2s - 2t + u - 2v = 0 \\ -r + s - 5t - 2u + 3v - w = 0 \\ 3r + s + 3t + u - v = 0 \end{cases}$$

a) Ce système permet-il de définir les variables r, s, t en fonction des variables u, v, w ?

b) Ce système permet-il de définir les variables u, v, w en fonction des variables r, s, t ?

25. Traiter analytiquement puis représenter graphiquement l'ensemble des solutions du système de 3 équations linéaires à 3 variables x, y, z suivant :

$$\begin{cases} 6x + 10y + 15z = 30 \\ 4x + 3y + 6z = 12 \\ 5y - 8z = 0 \end{cases}$$

Solution des exercices (Chapitre 1)

1. $S = \{ (0,1), (1,0) \}$
2. $S = \emptyset$
3. $S = \{ (2,1) \}$
4. $S = \{ (3,2,2), (-4,-5,-5) \}$
5. $S = \{ (4,3,1,1), (-5,-6,-8,-8) \}$
6. $S = \{ \left(\frac{13}{4}, \frac{39}{4}, \frac{7}{16}, \frac{7}{16} \right), (-4,-3,-5,-5) \}$
7. $S = \{ (-1,1,-1), (1,7,7) \}$
8. $S = \{(4,1)\}$
9. $S = \{ (5,2), (2,5) \}$
10. $S = \{ (2, 1 + \log_3(4)) \}$
11. $S = \{ (0,-1/3,5/3), (1,0,1), (5/2,1/2,0) \}$
12. $S = \{ (2.997,31/8); (-2.997,31/8) \}$
13. $S = \{ (4,3), (-4,3), (\sqrt{21},-2), (-\sqrt{21},-2) \}$
14. $S = \{ (0,0,0), \left(\frac{\sqrt{3}}{2}, 3, \sqrt{3} \right), \left(-\frac{\sqrt{3}}{2}, 3, -\sqrt{3} \right) \}$
15. $S = \{ (4/7,0,12/7) \}$
16.
$$\begin{cases} x = 2y - 1 \\ z = 5y^2 + 6y - 3 \end{cases}$$
17. $S = \{(1,2,3)\}$
18. $S = \emptyset$
19.
$$\begin{cases} x = 3z + 1 \\ y = 2 - 5z \end{cases}$$
20.
$$\begin{cases} x = \frac{5.t - 3.z}{3} \\ y = \frac{3.z - 7.t}{3} \end{cases}$$
21. $S = \{(3,-2,0,7)\}$
22.
$$\begin{cases} x = 5z - 2 \\ y = 3 - 2z \end{cases}$$

23. Non, on ne peut pas exprimer y_1 et y_2 en fonction de x .

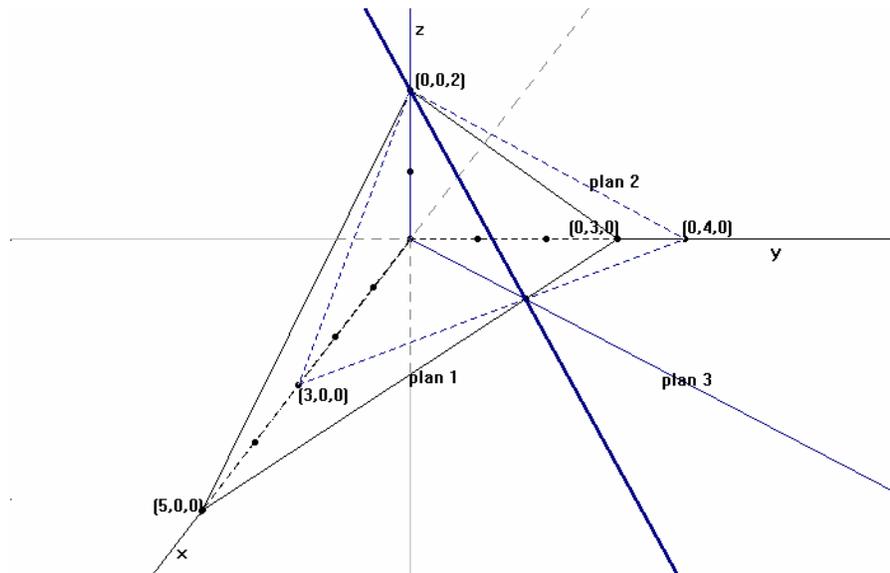
Commentaires : Et pourtant, le système n'est pas impossible... Cet exercice est là pour attirer votre attention sur le fait important suivant : en général, les économistes ne cherchent pas à connaître les solutions d'un système, mais bien à savoir s'ils peuvent **exprimer certaines variables** (les endogènes) **en fonction d'autres** (les exogènes).

24. a) Non

b) Oui

$$25. \begin{cases} x = \frac{5}{8} \cdot y \\ z = -\frac{11}{12} \cdot y + 2 \end{cases}$$

Graphiquement :



Chapitre 2 : Matrices

2.1. Calcul matriciel : somme, produit, multiplication par un réel

Exercices

Exercice 1

Soient les matrices suivantes : $A = \begin{pmatrix} 2 & 8 \\ 3 & 0 \\ 5 & 1 \end{pmatrix}$, $B = \begin{pmatrix} 2 & 0 \\ 3 & 8 \end{pmatrix}$, $C = \begin{pmatrix} 7 & 2 \\ 6 & 3 \end{pmatrix}$, $D = \begin{pmatrix} 2 & 0 & -1 \\ 1 & 1 & 1 \end{pmatrix}$.

Calculer, si possible :

1. $A+B$
2. $B+C$
3. $A+D$
4. $-A$
5. $2.B - 3.C$
6. $B.A$
7. $A.B$
8. $C.B$
9. $B.C$

Exercice 2

Si possible, effectuer les multiplications suivantes.

1. $\begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ 2 & -1 \end{pmatrix}$
2. $\begin{pmatrix} 2 & 1 \\ 6 & 0 \end{pmatrix} \cdot \begin{pmatrix} -1 \\ 4 \end{pmatrix}$
3. $(2 \ 1 \ 0) \cdot \begin{pmatrix} 4 & 0 \\ 0 & 2 \\ -1 & 1 \end{pmatrix}$
4. $\begin{pmatrix} 4 & 2 \\ 3 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
5. $\begin{pmatrix} 9 & 6 & 2 \\ 4 & 3 & 1 \end{pmatrix} \cdot \begin{pmatrix} 2 & 4 \\ 0 & 2 \end{pmatrix}$
6. $\begin{pmatrix} 2 \\ 3 \end{pmatrix} \cdot (3 \ -1)$

2.2. Inversion d'une matrice par la méthode de la matrice compagnon

Inverser la matrice $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ revient à trouver la matrice $A^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$

telle que $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \cdot \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

Pour inverser la matrice $A_{n \times n}$ par la méthode de la matrice compagnon :

1. Ecrire côte à côte la matrice A et la matrice identité I :

$$A_{n \times n} \mid I_{n \times n}$$

2. Appliquer la méthode de Gauss Jordan à la matrice A jusqu'à obtenir la matrice I . Appliquer au fur et à mesure les mêmes modifications à la matrice *compagnon*.

3. On obtient

$$I_{n \times n} \mid A^{-1}_{n \times n}$$

La matrice inverse est $A^{-1}_{n \times n}$.

Remarque 1 :

La matrice $I_{n \times n}$ est la matrice identité.

Elle comporte des 0 partout sauf sur la diagonale « descendante » où elle ne comporte que des 1.

Exemple : $I_{3 \times 3} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

Remarque 2 :

S'il est impossible de faire apparaître la matrice $I_{n \times n}$ à gauche (par exemple si on obtient une ligne entière de zéros), c'est que la matrice $A_{n \times n}$ n'est pas inversible.

Exercices

Exercice 3

Inverser les matrices suivantes par la méthode de la matrice compagnon

1. $A = \begin{pmatrix} 3 & 1 & 1 \\ 2 & 1 & 2 \\ 1 & 3 & 2 \end{pmatrix}$

$$2. B = \begin{pmatrix} 2 & 4 & 8 \\ 1 & 3 & 5 \\ 3 & 8 & 14 \end{pmatrix}$$

$$3. C = \begin{pmatrix} 1 & 1 & -1 & 1 \\ -1 & 1 & 0 & 1 \\ 1 & 2 & 3 & 0 \\ -1 & 0 & 0 & 2 \end{pmatrix}$$

2.3. Calcul du déterminant d'une matrice carrée

2.3.1. Pour les matrices 1x1, 2x2, 3x3 :

$$\text{dét}(4) = 4$$

$$\text{dét} \begin{pmatrix} 3 & 2 \\ 1 & 5 \end{pmatrix} = 3 \cdot 5 - 1 \cdot 2 = 13$$

$$\text{dét} \begin{pmatrix} 3 & 1 & 1 \\ 2 & 1 & 2 \\ 1 & 3 & 2 \end{pmatrix} \begin{matrix} 3 & 1 \\ 2 & 1 \\ 1 & 3 \end{matrix} = 3 \cdot 1 \cdot 2 + 1 \cdot 2 \cdot 1 + 1 \cdot 2 \cdot 3 - 1 \cdot 1 \cdot 1 - 3 \cdot 2 \cdot 3 - 2 \cdot 2 \cdot 1 = -9$$

Remarque :

La méthode appliquée ici (dite de **SARRUS**) n'est valable **que** pour le calcul d'un déterminant d'une matrice **3X3**.

Exercices

Exercice 4 : calculer

$$1. \text{dét} \begin{pmatrix} 2 & 4 & 8 \\ 1 & 3 & 5 \\ 3 & 8 & 14 \end{pmatrix}$$

$$2. \text{dét} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

$$3. \text{dét} \begin{pmatrix} a & 0 & 0 \\ d & b & 0 \\ e & f & c \end{pmatrix}$$

$$4. \text{dét} \begin{pmatrix} 0 & 0 & a \\ 0 & b & d \\ c & e & f \end{pmatrix}$$

2.3.2. Pour les matrices quelconques :

•

Le **mineur** associé à un terme d'une matrice A est le déterminant obtenu en supprimant dans la matrice A la ligne et la colonne contenant ce terme.

Exemple : Soit $A = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{pmatrix}$

calcul du **mineur** associé au terme $a_{1,1}$:

$$\text{dét} \begin{pmatrix} \bullet & \times & \times \\ \times & a_{2,2} & a_{2,3} \\ \times & a_{3,2} & a_{3,3} \end{pmatrix} = \text{dét} \begin{pmatrix} a_{2,2} & a_{2,3} \\ a_{3,2} & a_{3,3} \end{pmatrix} = a_{2,2} \cdot a_{3,3} - a_{3,2} \cdot a_{2,3}$$

calcul du **mineur** associé au terme $a_{1,2}$:

$$\text{dét} \begin{pmatrix} \times & \bullet & \times \\ a_{2,1} & \times & a_{2,3} \\ a_{3,1} & \times & a_{3,3} \end{pmatrix} = \text{dét} \begin{pmatrix} a_{2,1} & a_{2,3} \\ a_{3,1} & a_{3,3} \end{pmatrix} = a_{2,1} \cdot a_{3,3} - a_{3,1} \cdot a_{2,3}$$

calcul du **mineur** associé au terme $a_{1,3}$:

$$\text{dét} \begin{pmatrix} \times & \times & \bullet \\ a_{2,1} & a_{2,2} & \times \\ a_{3,1} & a_{2,3} & \times \end{pmatrix} = \text{dét} \begin{pmatrix} a_{2,1} & a_{2,2} \\ a_{3,1} & a_{2,3} \end{pmatrix} = a_{2,1} \cdot a_{2,3} - a_{3,1} \cdot a_{2,2}$$

-

Le **cofacteur** associé à un élément a_{ij} d'une matrice A est le produit du mineur associé à cet élément par $(-1)^{i+j}$.

Exemple : soit $A = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{pmatrix}$

calcul du **cofacteur** associé au terme $a_{1,1}$:

$$(-1)^{1+1} \cdot \det \begin{pmatrix} a_{2,2} & a_{2,3} \\ a_{3,2} & a_{3,3} \end{pmatrix} = (-1)^2 \cdot \det \begin{pmatrix} a_{2,2} & a_{2,3} \\ a_{3,2} & a_{3,3} \end{pmatrix} = \det \begin{pmatrix} a_{2,2} & a_{2,3} \\ a_{3,2} & a_{3,3} \end{pmatrix}$$

calcul du **cofacteur** associé au terme $a_{1,2}$:

$$(-1)^{1+2} \cdot \det \begin{pmatrix} a_{2,1} & a_{2,3} \\ a_{3,1} & a_{3,3} \end{pmatrix} = (-1)^3 \cdot \det \begin{pmatrix} a_{2,1} & a_{2,3} \\ a_{3,1} & a_{3,3} \end{pmatrix} = -\det \begin{pmatrix} a_{2,1} & a_{2,3} \\ a_{3,1} & a_{3,3} \end{pmatrix}$$

calcul du **cofacteur** associé au terme $a_{1,3}$:

$$(-1)^{1+3} \cdot \det \begin{pmatrix} a_{2,1} & a_{2,2} \\ a_{3,1} & a_{2,3} \end{pmatrix} = (-1)^4 \cdot \det \begin{pmatrix} a_{2,1} & a_{2,2} \\ a_{3,1} & a_{2,3} \end{pmatrix} = \det \begin{pmatrix} a_{2,1} & a_{2,2} \\ a_{3,1} & a_{2,3} \end{pmatrix}$$

- **Calcul d'un déterminant par la méthode des cofacteurs**

Le déterminant de la matrice A est égal à la somme des produits des éléments d'une rangée de cette matrice par les cofacteurs associés à ces éléments.

Exemple :

Calcul du déterminant de la matrice $A = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{pmatrix}$

par la méthode des cofacteurs en développant selon la première ligne de la matrice .

$$\begin{aligned} \det \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{pmatrix} &= a_{1,1} \cdot (-1)^{1+1} \cdot \det \begin{pmatrix} \bullet & \times & \times \\ \times & a_{2,2} & a_{2,3} \\ \times & a_{3,2} & a_{3,3} \end{pmatrix} \\ &+ a_{1,2} \cdot (-1)^{1+2} \cdot \det \begin{pmatrix} \times & \bullet & \times \\ a_{2,1} & \times & a_{2,3} \\ a_{3,1} & \times & a_{3,3} \end{pmatrix} \\ &+ a_{1,3} \cdot (-1)^{1+3} \cdot \det \begin{pmatrix} \times & \times & \bullet \\ a_{2,1} & a_{2,2} & \times \\ a_{3,1} & a_{3,2} & \times \end{pmatrix} \end{aligned}$$

$$\begin{aligned} \det \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{pmatrix} &= a_{1,1} \cdot (a_{2,2} \cdot a_{3,3} - a_{3,2} \cdot a_{2,3}) \\ &- a_{1,2} \cdot (a_{2,1} \cdot a_{3,3} - a_{3,1} \cdot a_{2,3}) \\ &+ a_{1,3} \cdot (a_{2,1} \cdot a_{3,2} - a_{3,1} \cdot a_{2,2}) \end{aligned}$$

• Intérêt de la méthode des cofacteurs

- 1) La règle des cofacteurs est valable pour calculer le déterminant de toute matrice **carrée** $n \times n$.
- 2) La règle des cofacteurs est très simple lorsqu'elle est appliquée à une **rangée** qui comporte **beaucoup d'éléments nuls**.

Exemple : calculer $\det \begin{pmatrix} 1 & 7 & -5 \\ 2 & 0 & 0 \\ -3 & 4 & 6 \end{pmatrix}$

La rangée comportant le plus de zéros est la deuxième ligne :

$$a_{2,1} \quad a_{2,2} \quad a_{2,3} = 2 \quad 0 \quad 0$$

La règle des cofacteurs appliquée à cette ligne donne

$$\begin{aligned}
\det \begin{pmatrix} 1 & 7 & -5 \\ 2 & 0 & 0 \\ -3 & 4 & 6 \end{pmatrix} &= 2 \cdot \text{cofacteur}(a_{2,1}) + 0 \cdot \text{cofacteur}(a_{2,2}) + 0 \cdot \text{cofacteur}(a_{2,3}) \\
&= 2 \cdot \text{cofacteur}(a_{2,1}) \\
&= 2 \cdot (-1)^{2+1} \cdot \det \begin{pmatrix} \times & 7 & -5 \\ \bullet & \times & \times \\ \times & 4 & 6 \end{pmatrix} \\
&= -2 \cdot \det \begin{pmatrix} 7 & -5 \\ 4 & 6 \end{pmatrix} \\
&= -124
\end{aligned}$$

Exercices

Exercice 5

Calculer les déterminants suivants en utilisant la méthode des cofacteurs.

1. $\det \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$

2. $\det \begin{pmatrix} 3 & 1 & 1 \\ 2 & 1 & 2 \\ 1 & 3 & 2 \end{pmatrix}$

3. $\det \begin{pmatrix} 2 & 4 & 8 \\ 1 & 3 & 5 \\ 3 & 8 & 14 \end{pmatrix}$

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

$\det \begin{pmatrix} a & 0 & 0 \\ d & b & 0 \\ e & f & c \end{pmatrix}$

6. $\det \begin{pmatrix} 0 & 0 & a \\ 0 & b & d \\ c & e & f \end{pmatrix}$

2.4. Inversion d'une matrice par la méthode de la matrice adjointe

Inverser la matrice $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ revient à trouver la matrice $A^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$

telle que $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \cdot \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

On peut inverser cette matrice en utilisant la méthode de la matrice compagne.

On peut aussi inverser cette matrice en utilisant la **méthode de la matrice adjointe**.

Inversion d'une matrice par la méthode de la matrice adjointe

1. **Calculer le déterminant** de la matrice A à inverser.

Si $\det(A) = 0$, alors la matrice A n'est pas inversible

Si $\det(A) \neq 0$, alors la matrice A est inversible

2. Ecrire la **matrice des cofacteurs**.

3. **Transposer** la matrice des cofacteurs. On obtient la **matrice adjointe** $\text{adj}(A)$.

4. La **matrice inverse** est donnée par $A^{-1} = \frac{1}{\det(A)} \cdot \text{adj}(A)$

Exercice résolu

Inverser la matrice $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ par la méthode des cofacteurs

1. **Calculer le déterminant**.

$$\det \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = -2 \neq 0 : \text{ la matrice } A \text{ est inversible.}$$

2. **Ecrire la matrice des cofacteurs**.

$$\text{Cofacteur du terme } a_{1,1} : (-1)^{1+1} \cdot \det \begin{pmatrix} \bullet & \times \\ \times & 4 \end{pmatrix} = 4$$

$$\text{Cofacteur du terme } a_{1,2} : (-1)^{1+2} \cdot \det \begin{pmatrix} \times & \bullet \\ 3 & \times \end{pmatrix} = -3$$

$$\text{Cofacteur du terme } a_{2,1} : (-1)^{2+1} \cdot \det \begin{pmatrix} \times & 2 \\ \bullet & \times \end{pmatrix} = -2$$

Cofacteur du terme $a_{2,2}$: $(-1)^{2+2} \cdot \det \begin{pmatrix} 1 & \times \\ \times & \bullet \end{pmatrix} = 1$

La matrice des cofacteurs est $\text{cft}(A) = \begin{pmatrix} 4 & -3 \\ -2 & 1 \end{pmatrix}$

3. Transposer la matrice des cofacteurs.

$${}^t \begin{pmatrix} 4 & -3 \\ -2 & 1 \end{pmatrix} = \begin{pmatrix} 4 & -2 \\ -3 & 1 \end{pmatrix} = \text{adj}(A)$$

4. Ecrire la matrice inverse.

$$\begin{aligned} A^{-1} &= \frac{1}{\det(A)} \cdot \text{adj}(A) = -\frac{1}{2} \cdot \begin{pmatrix} 4 & -2 \\ -3 & 1 \end{pmatrix} \\ &= \begin{pmatrix} -2 & 1 \\ 3 & -\frac{1}{2} \end{pmatrix} \end{aligned}$$

Exercices

Exercice 6

Pour chacune des matrices suivantes,

- déterminer si la matrice est inversible
- si oui , calculer la matrice inverse.

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

$$B = \begin{pmatrix} 2 & 4 & 8 \\ 1 & 3 & 5 \\ 3 & 8 & 14 \end{pmatrix}$$

$$C = \begin{pmatrix} 3 & 0 & 1 & 0 \\ 2 & 1 & 0 & 4 \\ 0 & 3 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

2.5. Résoudre un système linéaire par la méthode de Cramer

La méthode de Cramer permet de résoudre un système linéaire

$$\text{de } n \text{ équations à } n \text{ variables} \begin{cases} a_{1,1} \cdot z_1 + a_{1,2} \cdot z_2 + \dots + a_{1,n} \cdot z_n = b_1 \\ a_{2,1} \cdot z_1 + a_{2,2} \cdot z_2 + \dots + a_{2,n} \cdot z_n = b_2 \\ \dots \quad \dots \quad \dots \quad \dots \quad \dots \\ a_{n,1} \cdot z_1 + a_{n,2} \cdot z_2 + \dots + a_{n,n} \cdot z_n = b_n \end{cases}$$

$$\text{écrit sous forme matricielle} \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \dots & \dots & \dots & \dots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} \end{pmatrix} \cdot \begin{pmatrix} z_1 \\ z_2 \\ \dots \\ z_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_n \end{pmatrix}$$

ou encore $A_{nxn} \cdot Z_{nx1} = B_{nx1}$

Résolution d'un système linéaire par la méthode de Cramer

1. Calculer $\det(A)$;

si $\det(A) \neq 0$ alors le système possède une **solution unique**.

2. Quel que soit l'indice i , la valeur de la variable z_i est donné par :

$$\det \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & b_1 & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & b_2 & \dots & a_{2,n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{n,1} & a_{n,2} & \dots & b_n & \dots & a_{n,n} \end{pmatrix}$$

$$z_i = \frac{\det(\dots)}{\det(A)}$$

$$\det(A)$$

La $i^{\text{ème}}$ colonne de A est remplacée par la colonne B des termes indépendants

Exercice résolu

Résoudre le système $\begin{cases} x + 2.y = 5 \\ 3.x + 4.y = 11 \end{cases}$ par la méthode de Cramer.

Le système s'écrit, sous forme matricielle : $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 5 \\ 11 \end{pmatrix}$

$$A_{2 \times 2} \cdot Z_{2 \times 1} = B_{2 \times 1}$$

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

1. Calculer le déterminant de A :

$$\det A = \det \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = -2 \neq 0 \rightarrow \text{le système possède une solution unique.}$$

2. Cette solution est donnée par :

$$\text{la première variable } x = \frac{\det \begin{pmatrix} 5 & 2 \\ 11 & 4 \end{pmatrix}}{\det \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}} = \frac{-2}{-2} = 1$$

$$\text{la deuxième variable } y = \frac{\det \begin{pmatrix} 1 & 5 \\ 3 & 11 \end{pmatrix}}{\det \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}} = \frac{-4}{-2} = 2$$

La solution du système est : $\begin{cases} x = 1 \\ y = 2 \end{cases}$

Exercices

Exercice 7

Déterminer si les systèmes suivants ont une solution unique.

Si oui, utiliser la méthode de Cramer pour résoudre le système.

$$1. \begin{cases} 3x + y + z = 0 \\ 2x + y + 2z = 0 \\ x + 3y + 2z = 0 \end{cases}$$

$$2. \begin{pmatrix} 2 & 4 & 8 \\ 1 & 3 & 5 \\ 3 & 8 & 14 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 6 \end{pmatrix}$$

$$3. \begin{cases} 3x + z = 1 \\ 2x + y + 4z = -1 \\ 3y + 2z = 0 \\ t = 1 \end{cases}$$

Exercice 8

Déterminer si les systèmes suivants ont une solution unique.

Si oui, utiliser la méthode de Cramer pour résoudre le système.

$$1. \begin{cases} x - 4y = 2 \\ 2x + y = 1 \end{cases}$$

$$2. \begin{cases} 2x + 5y + 3z = 1 \\ 3x + y + 2z = 1 \\ x + 2y + z = 0 \end{cases}$$

$$3. \begin{cases} 2x + 3y - 3z = 13 \\ 5x - y + 2z = 21 \\ x + y + z = 12 \end{cases}$$

$$4. \begin{cases} x + y + z - 2 = 0 \\ 2z - y = 4 \\ 2x + 2y = 1 \end{cases}$$

Solutions des exercices (Chapitre 2)

Exercice 1

1. $A + B$ impossible vu la dimension des matrices

$$2. B + C = \begin{pmatrix} 9 & 2 \\ 9 & 11 \end{pmatrix}$$

3. $A + D$ impossible vu la dimension des matrices

$$4. -A = \begin{pmatrix} -2 & -8 \\ -3 & 0 \\ -5 & -1 \end{pmatrix}$$

$$5. 2.B - 3.C = \begin{pmatrix} -17 & -6 \\ -12 & 7 \end{pmatrix}$$

6. $B.A$ impossible vu la dimension des matrices

$$7. A.B = \begin{pmatrix} 28 & 64 \\ 6 & 0 \\ 13 & 8 \end{pmatrix}$$

$$8. C.B = \begin{pmatrix} 20 & 16 \\ 21 & 24 \end{pmatrix}$$

$$9. B.C = \begin{pmatrix} 14 & 4 \\ 69 & 30 \end{pmatrix}$$

Exercice 2

$$1. \begin{pmatrix} 2 & 1 \\ 8 & -1 \end{pmatrix}$$

$$2. \begin{pmatrix} 2 \\ -6 \end{pmatrix}$$

$$3. (8 \ 2)$$

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

5. impossible vu la dimension des matrices

$$6. \begin{pmatrix} 6 & -2 \\ 9 & -3 \end{pmatrix}$$

Exercice 3

$$1. A^{-1} = \begin{pmatrix} \frac{4}{9} & -\frac{1}{9} & -\frac{1}{9} \\ \frac{2}{9} & -\frac{5}{9} & \frac{4}{9} \\ -\frac{5}{9} & \frac{8}{9} & -\frac{1}{9} \end{pmatrix}$$

$$2. \text{ La matrice } B = \begin{pmatrix} 2 & 4 & 8 \\ 1 & 3 & 5 \\ 3 & 8 & 14 \end{pmatrix} \text{ n'est pas inversible.}$$

$$3. C^{-1} = \begin{pmatrix} \frac{3}{8} & -\frac{5}{8} & \frac{1}{8} & \frac{1}{8} \\ \frac{3}{16} & \frac{11}{16} & \frac{1}{16} & -\frac{7}{16} \\ -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{3}{16} & -\frac{5}{16} & \frac{1}{16} & \frac{9}{16} \end{pmatrix}$$

Exercice 4

1. 0
2. -2
3. $a.b.c$
4. $-a.b.c$

Exercice 5

1. -2
2. -9
3. 0
4. 12
5. $a.b.c$
6. $-a.b.c$

Exercice 6

$$1. A^{-1} = \begin{pmatrix} \frac{4}{9} & -\frac{1}{9} & -\frac{1}{9} \\ \frac{2}{9} & -\frac{5}{9} & \frac{4}{9} \\ -\frac{5}{9} & \frac{8}{9} & -\frac{1}{9} \end{pmatrix}$$

2. La matrice $B = \begin{pmatrix} 2 & 4 & 8 \\ 1 & 3 & 5 \\ 3 & 8 & 14 \end{pmatrix}$ n'est pas inversible.

3. $C^{-1} = \begin{pmatrix} \frac{1}{6} & \frac{1}{4} & -\frac{1}{12} & -1 \\ -\frac{1}{3} & \frac{1}{2} & \frac{1}{6} & -2 \\ \frac{1}{2} & -\frac{3}{4} & \frac{1}{4} & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix}$

Exercice 7

1. $\begin{cases} x=0 \\ y=0 \\ z=0 \end{cases}$

système homogène à solution unique

2. Le système n'a pas de solution unique

3. $\begin{cases} x = -\frac{13}{12} \\ y = -\frac{17}{6} \\ z = \frac{17}{4} \\ t = 1 \end{cases}$

Exercice 8

1. $\begin{cases} x = 2/3 \\ y = -1/3 \end{cases}$

2. $\begin{cases} x = -1/2 \\ y = -1/2 \\ z = 3/2 \end{cases}$

3. $\begin{cases} x = 3 \\ y = 4 \\ z = 5 \end{cases}$

4. $\begin{cases} x = 1/2 \\ y = -1 \\ z = 1/2 \end{cases}$

Chapitre 3 : Variations

Remarque préliminaire importante !!!

Dans tout ce chapitre, nous travaillerons avec des modèles mathématiques pour lesquels **il est possible d'exprimer explicitement les variables endogènes en fonctions des variables exogènes**. Les autres cas devront être traités différemment, comme nous le verrons plus tard dans ce cours.

3.1. Modèles à deux variables

Soit $y = f(x)$ un modèle explicitant la variable y en fonction de la variable x .

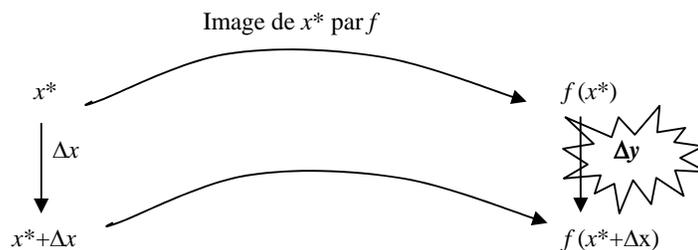
Objectif :

**étudier le comportement de la variable endogène y
au voisinage de x^* ,
valeur particulière de la variable exogène x**

(on dit aussi étudier le comportement local de la fonction f au voisinage de x^*).

3.1.1. Variation absolue (ou en valeur)

On envisage une variation en valeur de la variable x à partir de x^* . Cette variation sera positive ou négative ; on la notera Δx .



Nous allons nous intéresser à la variation Δy (induite par la variation Δx) :

comment varie y

(c'est-à-dire comment la fonction f réagit-elle)

si x^* devient $x^* + \Delta x$?

(c'est-à-dire si l'on fait subir à x une variation absolue de valeur Δx ?)

La variation absolue Δy de y

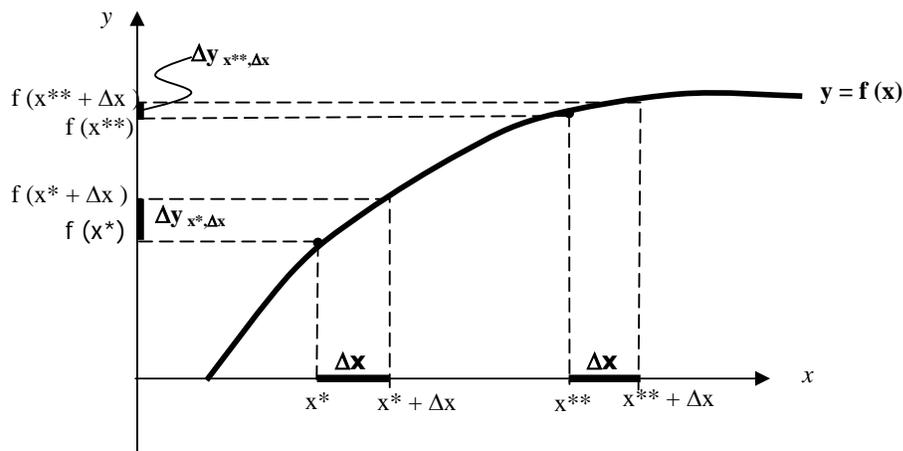
(ou encore variation en valeur de la fonction f)

est égale à :

$$\Delta y = f(x^* + \Delta x) - f(x^*)$$

La variation absolue d'une variable s'exprime dans la même unité que cette variable. Elle peut être un nombre positif ou négatif.

Attention : Pour un même Δx
mais pour deux valeurs différentes de la variable x ,
le Δy peut être différent !



DONC : Pour être précis, il faudrait faire apparaître dans la notation que Δy dépend à la fois de x^* et de Δx . Une notation du type $\Delta y(x^*, \Delta x)$ ou $\Delta y_{x^*, \Delta x}$ serait donc meilleure. Afin de ne pas alourdir la présentation, nous continuerons cependant à écrire Δy (mais il faut veiller à garder cette remarque en mémoire).

Exemple

Considérons l'expression suivante :

$$C = \underbrace{3.Q^3 - 1800.Q^2 + 360000.Q + 1000000}_{f(Q)}$$

Nous allons faire varier Q et voir ce qui se passe.

Choisissons $\Delta Q = 1$ et regardons comment C varie ...	
autour de $Q^* = 100$	autour de $Q^* = 200$
<ul style="list-style-type: none"> • Pour $Q^* = 100$, $f(Q^*) = 22000000$ • Pour $Q^* + \Delta Q = 101$, $f(Q^* + \Delta Q) = 22089103$ 	<ul style="list-style-type: none"> • Quand $Q^* = 200$, $f(Q^*) = 25000000$ • Quand $Q^* + \Delta Q = 201$, $f(Q^* + \Delta Q) = 25000003$
ΔC vaut donc...	
$\Delta C = 89103$	$\Delta C = 3$

Remarquons que pour un même ΔQ , ΔC varie d'après la valeur de Q^* !!!

• **Exercice 1**

Soit $f: \mathbb{R} \rightarrow \mathbb{R} : x \rightarrow y = x^2$.

Partant de $x^* = 1$, on donne à la variable x une variation absolue de valeur

- a) 2
- b) -1.

Quelles sont les variations absolues de la variable y correspondantes ?

Calculez aussi les rapports des variations absolues $\frac{\Delta y}{\Delta x}$ correspondants.

• **Exercice 2**

Soit $f: \mathbb{R} \rightarrow \mathbb{R} : x \rightarrow y = x^2$.

Partant de $x^* = 0$, on donne à la variable x une variation absolue de valeur

- a) 2
- b) -1.

Quelles sont les variations absolues de la variable y correspondantes ?

Calculez aussi les rapports des variations absolues $\frac{\Delta y}{\Delta x}$ correspondants.

Cas particulier : Fonction linéaire à une variable

Soit le modèle linéaire $y = a x + b$. Rappelons que la représentation graphique de ce modèle dans un repère oxy est une droite, de pente a .

Nous travaillons au voisinage du point x^* et nous recherchons la variation Δy correspondant à une variation Δx de x .

Rappel : La variation Δy de y est $\Delta y = f(x^* + \Delta x) - f(x^*)$

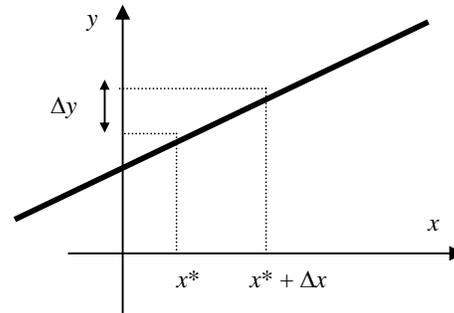
En appliquant cette formule dans ce cas particulier, on obtient :

$$\Delta y = [a(x^* + \Delta x) + b] - [a(x^*) + b] = a \cdot \Delta x.$$

Dans le cas linéaire, on a donc :

$$\Delta y = a \cdot \Delta x$$

où a est la **pente de la droite**.



Donc, si Δx est constant, Δy est constant aussi :

Δy ne dépend pas de la valeur x^* à partir de laquelle on considère la variation Δx .

De plus, on constate que le rapport $\frac{\Delta y}{\Delta x}$ est constant

et vaut la pente de la droite d'équation $y = ax + b$.

Exemple

Soit un modèle linéaire de type $y = a.x + b$:

$$C = 150000.Q + 1000000$$

où C est la variable endogène et Q la variable exogène.

Dans ce cas , $a = 150000$ et $b = 1000000$.

Choisissons une variation en valeur d'une unité de la variable Q ($\Delta Q = 1$) ; quelle que soit la valeur Q^* de Q à partir de laquelle on travaille, on obtient :

$$\begin{array}{ccc} Q^* & \rightarrow & C(Q^*) = 150000.Q^* + 1000000 \\ \Delta Q = 1 & \downarrow & \downarrow \\ Q^* + 1 & \rightarrow & C(Q^* + 1) = 150000.(Q^* + 1) + 1000000 \end{array}$$

ce qui donne

$$\Delta C = C(Q^*+1) - C(Q^*) = 150000.$$

Quel que soit Q^* , on constate donc que la variation ΔC correspondant à une variation en valeur d'une unité de Q est égale à 150000.

Observons aussi que $\Delta C = a \cdot \Delta Q$ (ce qui est caractéristique du linéaire).

3.1.2. Variation relative (ou en pourcentage)

Mathématiquement, on appelle **variation relative de la variable x**
 (en un point x^* , par rapport à une variation absolue de valeur Δx),
 le rapport $\frac{\Delta x}{x^*}$.

De la même manière, la variation relative de la variable y
 (correspondant à une variation relative Δx de la variable x à partir du point x^*)
 est le rapport $\frac{\Delta y}{y^*}$.

Les variations relatives sont des nombres sans unités, qui peuvent être positifs ou négatifs.

Signalons que la tradition est d'exprimer les variations relatives en pourcentages.
Au lieu de « variation relative », on dit parfois « (taux de) variation en pourcentage ».

Il est parfois intéressant de calculer le **rapport des variations relatives** :

$$\frac{\frac{\Delta y}{y^*}}{\frac{\Delta x}{x^*}}$$

• **Exercice 3**

- a) Augmenter la valeur de x^* de 25% revient à multiplier x^* par ... ?
- b) Diminuer la valeur de x^* de 10% revient à multiplier x^* par ... ?

• **Exercice 4**

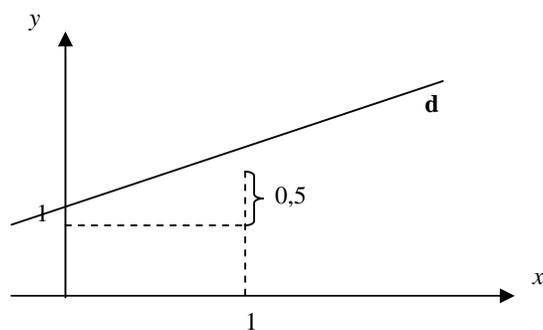
Soit $y = 3x + 2$.

- a) Quelle variation relative de la variable x correspond à une variation absolue $\Delta x = 0,5$ à partir de la valeur $x^* = 2$? (Exprimez votre réponse en %.)
- b) Calculez la variation absolue de la variable y correspondant à une variation absolue de x de 0,5 à partir de la valeur 2.
- c) Calculez la variation relative de y correspondante.
- d) Calculez le rapport des variations absolues correspondant.
- e) Calculez le rapport des variations relatives correspondant.

• **Exercice 5** :

Même question avec $x^* = 5$ et $\Delta x = -0,2$.

• **Exercice 6** :



- a) Donnez l'équation de la droite **d**, et l'expression de la fonction correspondante (dont le graphe est **d**).
- b) On donne, à partir de $x^* = 5$, une variation relative à la variable x de 20%.

Calculez le Δy correspondant, ainsi que les rapports des variations absolues et relatives.

c) Idem à partir de $x^* = 7$, avec la même variation relative de x .

d) Peut-on dire qu'une même variation relative de la variable x impliquera une même variation relative de la variable y ?

• **Exercice 7** :

Soit $f: \mathbb{R} \rightarrow \mathbb{R} : x \rightarrow x^2$.

On donne à la variable x une augmentation relative de 25% à partir de

a) $x^* = 4$

b) $x^* = 12$

c) $x^* = 20$.

Calculez les variations relatives de y correspondantes, et les rapports des variations relatives correspondants. Que constate-t-on ? Pouvait-on le prévoir ?

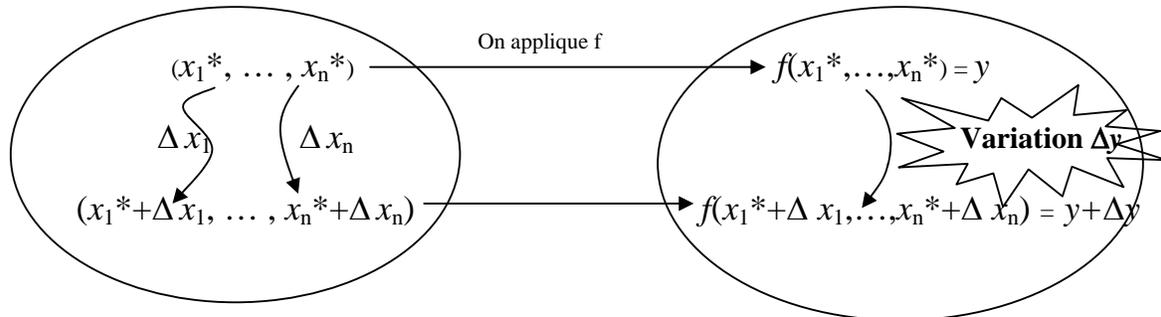
• **Exercice 8** :

Mêmes questions, en donnant à x une augmentation relative de 5%.

3.2. Modèles à trois variables ou plus

Soit $y = f(x_1, \dots, x_n)$ un modèle explicitant la variable y en fonction des variables x_1, \dots, x_n .

Question : comment varie y au voisinage de (x_1^*, \dots, x_n^*) [valeurs particulières des variables exogènes x_1, x_2, \dots, x_n] ?



Un point voisin de (x_1^*, \dots, x_n^*) peut s'écrire $f(x_1^* + \Delta x_1, \dots, x_n^* + \Delta x_n)$

où chaque Δx_i représente une variation en valeur de la variable x_i .

Aux variations en valeur $\Delta x_1, \dots, \Delta x_n$ des variables exogènes x_1, x_2, \dots, x_n [au départ de (x_1^*, \dots, x_n^*)] correspond une variation en valeur de la variable y . Elle est notée Δy . C'est à cette variation que nous allons nous intéresser. Il s'agit là d'une simple généralisation au cas de plusieurs variables exogènes de ce que nous venons de faire dans le cas où il n'y a qu'une seule variable exogène.

**Variation de valeur Δy de y
(induite par $\Delta x_1, \dots, \Delta x_n$) :**

$$\Delta y = f(x_1^* + \Delta x_1, \dots, x_n^* + \Delta x_n) - f(x_1^*, \dots, x_n^*)$$

(variation de la valeur de la fonction f)

N.B. Pour être précis, il faudrait faire apparaître dans la notation que Δy dépend de x_1^*, \dots, x_n^* et de $\Delta x_1, \dots, \Delta x_n$. Une notation du type $\Delta y(x_1^*, \dots, x_n^*, \Delta x_1, \dots, \Delta x_n)$ serait donc meilleure. Afin de ne pas alourdir la présentation, nous continuerons cependant à écrire Δy mais il faut veiller à garder cette remarque en mémoire.

Exercice résolu

Le revenu d'une compagnie est donné par l'équation : $R = 5 \cdot W^2 \cdot A^3$

Nous avons donc un modèle du type $y = f(x_1, x_2)$ [R correspond à y , W correspond à x_1 et A correspond à x_2].

W représentent les salaires payés (en millions de francs) et A les dépenses en publicité (en millions de francs).

- Si $W^* = 30$ et $A^* = 2$, alors $R^* = R(W^*, A^*) = \dots$.
- Que devient le revenu (de combien s'accroît-il ?) si W augmente de 5% (au départ de $W^* = 30$) sans modifier A ?
- Que devient le revenu si A diminue de 2% (au départ de $A^* = 2$) sans modifier W ?
- Que devient le revenu si à la fois W augmente de 5% (au départ de $W^* = 30$) et A diminue de 2% (au départ de $A^* = 2$) ?

Résolution

Quand on envisage une variation des salaires de 5% (au départ d'une valeur donnée pour les salaires), on parle de variation relative.

Mathématiquement, la variation relative en W^* est égale à $\frac{\Delta W}{W^*}$.

Dans ce cas, $W^* = 30$ et $\frac{\Delta W}{W^*} = \frac{5}{100} \Rightarrow$ la variation en valeur est : $\Delta W = \frac{5 \cdot W^*}{100} = \frac{150}{100} = 1.5$

Pour A , nous avons :

$\frac{\Delta A}{A^*} = -\frac{2}{100} \Rightarrow$ la variation en valeur est : $\Delta A = -\frac{2 \cdot A^*}{100} = -0.04$

- Notre modèle est $R = 5 \cdot W^2 \cdot A^3$
et nous travaillons au départ de $(W^*, A^*) = (30, 2)$
Nous avons $R(W^*, A^*) = 5 \cdot (30)^2 \cdot (2)^3 = 36000$.
- $R(W^* + \Delta W, A^*) = 5 \cdot (30 + 1.5)^2 \cdot (2)^3 = 39690$.
Donc la variation ΔR de R est égale à $39690 - 36000 = 3690$.
- $R(W^*, A^* + \Delta A) = 5 \cdot (30)^2 \cdot (2 - 0.04)^3 = 33882,912$.
Donc la variation ΔR de R est égale à $33882,912 - 36000 = -2117,088$.
- $R(W^* + \Delta W, A^* + \Delta A) = 5 \cdot (30 + 1.5)^2 \cdot (2 - 0.04)^3 = 37355,91048$.
Donc la variation ΔR de R est égale à $37355,91048 - 36000 = 1355,91048$.

• Exercice 9 :

Soit $y = 2x_1^2 - 3x_1x_2 + x_2^3$ et prenons $x_1^* = 40$ et $x_2^* = 30$.

- Si $\Delta x_1 = 1$ et $\Delta x_2 = -1$, que vaut Δy ?
- Si $\Delta x_1 = -1$ et $\Delta x_2 = 1$, que vaut Δy ?

Cas particulier : fonction linéaire à plusieurs variables

Soit le modèle linéaire $y = a_1 \cdot x_1 + \dots + a_n \cdot x_n + b$. Nous travaillons au voisinage du point (x_1^*, \dots, x_n^*) et nous recherchons la variation Δy correspondant à des variations Δx_1 de $x_1, \dots, \Delta x_n$ de x_n .

Rappel : La variation Δy de y est $\Delta y = f(x_1^* + \Delta x_1, \dots, x_n^* + \Delta x_n) - f(x_1^*, \dots, x_n^*)$.

Comme, dans ce cas, $f(x_1, \dots, x_n) = a_1 \cdot x_1 + \dots + a_n \cdot x_n + b$, il vient :

$$f(x_1, \dots, x_n) = a_1 \cdot x_1 + \dots + a_n \cdot x_n + b$$

$$f(x_1^* + \Delta x_1, \dots, x_n^* + \Delta x_n) = a_1 \cdot (x_1^* + \Delta x_1) + \dots + a_n \cdot (x_n^* + \Delta x_n) + b$$

Donc,

$$\Delta y = a_1 \cdot x_1^* + a_1 \cdot \Delta x_1 + \dots + a_n \cdot x_n^* + a_n \cdot \Delta x_n + b - [a_1 \cdot x_1^* + \dots + a_n \cdot x_n^* + b]$$

ce qui donne, après simplifications :

$$\Delta y = \mathbf{a_1 \cdot \Delta x_1 + \dots + a_n \cdot \Delta x_n}$$

où a_1, \dots, a_n sont **les coefficients** du modèle linéaire sur lequel on travaille au départ

(ou encore, où a_1, \dots, a_n sont **les coefficients directeurs de l'hyperplan** d'équation $y = a_1 \cdot x_1 + \dots + a_n \cdot x_n + b$).

N.B. C'est une généralisation directe de ce qui se passe pour des modèles du type $y = a \cdot x + b$: on a vu que dans ce cas, $\Delta y = a \cdot \Delta x$, où a est la pente de la droite d'équation $y = a \cdot x + b$.

• **exercice 10**

Soit $f: \mathbb{R}^3 \rightarrow \mathbb{R} : (x_1, x_2, x_3) \rightarrow y = 3x_1 - 2x_2 + x_3$.

- Calculez la variation de y correspondant à des variations $\Delta x_1 = 2/3$ et $\Delta x_2 = -1$, en partant du point $(x_1^*, x_2^*, x_3^*) = (1, 2, 3)$.
- Partant de $(x_1^*, x_2^*, x_3^*) = (1, 2, 3)$, on a effectué une variation absolue de 1 unité sur l'une des variables x_1, x_2, x_3 . La nouvelle valeur de la variable y (induite par cette variation) est de 0. Sur quelle variable a porté la variation ?

Solution des exercices (Chapitre 3)

• Exercice 1

a) On donne $f(x) = x^2$, $x^* = 1$, $\Delta x = 2$;

on a donc $f(x^*) = f(1) = 1$ et $f(x^* + \Delta x) = f(1+2) = f(3) = 9$.

La variation absolue de y est donnée par $\Delta y = f(x^* + \Delta x) - f(x^*) = 8$.

Le rapport des variations absolues est donné par $\frac{\Delta y}{\Delta x} = 4$.

b) Le rapport des variations absolues est donnée par $\frac{\Delta y}{\Delta x} = 1$.

• Exercice 2

a) Le rapport des variations absolues est donné par $\frac{\Delta y}{\Delta x} = 2$.

b) Le rapport des variations absolues est donné par $\frac{\Delta y}{\Delta x} = -1$.

• Exercice 3

a) Augmenter la valeur de x^* de 25% revient à multiplier x^* par 1,25.

b) Diminuer la valeur de x^* de 10% revient à multiplier x^* par 0,9.

• Exercice 4

La variation relative de x est de 25%.

a) La variation absolue de y est 1,5.

b) La variation relative de y est de 18,75%.

c) Le rapport des variations absolues est 3.

d) Le rapport des variations relatives est 0,75.

• Exercice 5

a) La variation relative de x est de -4% .

b) La variation absolue de y est -0,6 .

c) La variation relative de y est de -3,53% .

d) Le rapport des variations absolues est 3 .

e) Le rapport des variations relatives est 0,875 .

• Exercice 6

a) $f(x) = \frac{1}{2} \cdot x + 1$

b) Le rapport des variations relatives est 0,714285714 .

c) Le rapport des variations relatives est -1,27777 .

d) Non .

• **Exercice 7**

- a) Le rapport des variations relatives est 2,25 .
 b) Le rapport des variations relatives est 2,25 .
 c) Le rapport des variations relatives est 2,25 .

Conclusion :

La variation relative de y est constante pour une même variation relative de x quelle que soit la valeur de x^* .

• **Exercice 8**

- a) Le rapport des variations relatives est de 2,05 .
 b) Le rapport des variations relatives est de 2,05 .
 c) Le rapport des variations relatives est de 2,05 .

Remarque : On constate dans cet exercice qu'en partant d'une variation de x plus petite qu'à l'exercice 8, le rapport des variations relatives se rapproche de la valeur 2. A la limite, pour des variations de x de plus en plus petite :

$$\lim_{\Delta x \rightarrow 0} \frac{\frac{\Delta y}{y^*}}{\frac{\Delta x}{x^*}} = \frac{x^*}{y^*} \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} = \frac{x^*}{y^*} \cdot f'(x^*) = \frac{x^*}{(x^*)^2} \cdot 2x^* = 2$$

c'est l'élasticité de type point (cf. syllabus théorie, page 123-124) , qui, dans le cas d'une **fonction puissance**, est **constante** (et vaut l'exposant de la fonction)

• **Exercice 9**

- a) $\Delta y = -2416$
 b) $\Delta y = 2606$

• **Exercice 10**

a) La fonction est linéaire : $f(x_1, x_2, x_3) = 3x_1 - 2x_2 + x_3$.

Les coefficients sont $a_1 = 3$, $a_2 = -2$ et $a_3 = 1$.

On a donc : $\Delta y = 3 \cdot \frac{2}{3} - 2 \cdot (-1) + 1 \cdot 0 = 4$.

$$\begin{aligned} \text{b) } f(1+\Delta x_1, 2+\Delta x_2, 3+\Delta x_3) &= 3(1+\Delta x_1) - 2(2+\Delta x_2) + (3+\Delta x_3) \\ &= 2 + 3\Delta x_1 - 2\Delta x_2 + \Delta x_3 \\ &= 0 \end{aligned}$$

or $2 + 3\Delta x_1 - 2\Delta x_2 + \Delta x_3 = 0$ si $\Delta x_1 = 0$, $\Delta x_2 = 1$ et $\Delta x_3 = 0$.

La variation a donc porté sur la variable x_2 .

Chapitre 4 : Dérivées partielles de fonctions de plusieurs variables

4.1. Technique de dérivation

4.1.1. Dérivées partielles du premier ordre

Soit $f: \mathbb{R}^n \rightarrow \mathbb{R} : (x_1, x_2, \dots, x_n) \rightarrow f(x_1, x_2, \dots, x_n)$, une fonction de n variables.

Notation de la **dérivée partielle du premier ordre** de f par rapport à la variable x_i

- pour l'économiste : f'_{x_i}
- pour le mathématicien : $\frac{\partial f}{\partial x_i}$ ou $\frac{\partial}{\partial x_i} f$

Exercice

Exercice 1

Calculer les dérivées partielles du premier ordre de la fonction $f: \mathbb{R}^n \rightarrow \mathbb{R} : (x_1, x_2, \dots, x_n) \rightarrow f(x_1, x_2, \dots, x_n)$ dans les cas suivants

1. $f(x_1, x_2) = x_1^3 + x_2^3 - 3 x_1 x_2 + 5$
2. $f(x_1, x_2) = e^{(x_1^2 + x_2^2)}$
3. $f(x_1, x_2) = \frac{x_1}{x_2} + x_2 \ln x_1$
4. $f(x_1, x_2) = \ln (x_1 x_2)$
5. $f(x_1, x_2) = \sin x_1 + x_2 \cos (x_1^2)$
6. $f(x_1, x_2, x_3) = e^{x_1 \cdot x_2^2} + x_1 \ln (x_2 x_3^3)$

4.1.2. Dérivées partielles du second ordre

Soit $f: \mathbb{R}^n \rightarrow \mathbb{R}$, une fonction de n variables.

Notation de la **dérivée partielle du second ordre** de f par rapport à la variable x_i puis par rapport à la variable x_k

- pour l'économiste : $(f'_{x_i})_{x_k}$ ou $f''_{x_i x_k}$
- pour le mathématicien : $\frac{\partial}{\partial x_k} \left(\frac{\partial f}{\partial x_i} \right)$ ou $\frac{\partial^2 f}{\partial x_k \partial x_i}$ ou $\frac{\partial^2}{\partial x_k \partial x_i} f$

NB : $\frac{\partial^2}{\partial x_i \partial x_i} f$ est aussi noté $\frac{\partial^2}{(\partial x_i)^2} f$

Exercice

Exercice 2

Calculer les dérivées partielles du second ordre des fonctions n°1, 2 et 3 données dans l'exercice 1.

4.2. Application : approximation linéaire

4.2.1. Fonctions d'une variable (cf. cours p. 85)

Soit f une fonction de R dans R , soit x^* un élément du domaine de f .

On appelle **fonction approximation linéaire** f_{lin} de f en x^* , la fonction définie comme suit :

$$f_{lin(x^*)}(x) = f(x^*) + f'(x^*) \cdot (x - x^*)$$

Notons que $y = f(x^*) + f'(x^*) \cdot (x - x^*)$ est une équation de la **droite tangente** au graphe de f au point $(x^*, f(x^*))$.

Exercice résolu

Déterminer la fonction approximation linéaire f_{lin} de la fonction f en $x^* = 1$ si f est définie par :

$$f: R \rightarrow R : x \rightarrow f(x) = x^2 \ln x$$

Résolution

On sait que $f_{lin(x^*)}(x) = f(x^*) + f'(x^*) \cdot (x - x^*)$.

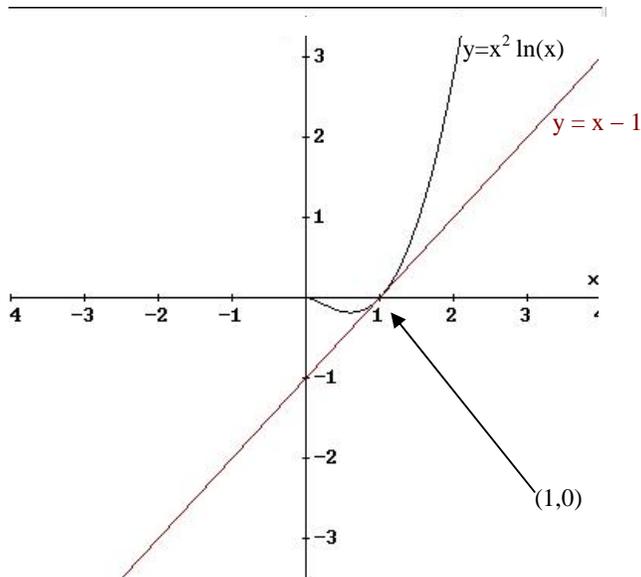
Ici,

- $x^* = 1$
- $f(x^*) = f(1) = 0$
- $f'(x) = 2x \ln x + x$
- $f'(x^*) = f'(1) = 1$

On en déduit que $f_{lin(1)}(x) = 0 + 1 \cdot (x - 1)$

ou encore $f_{lin(1)}(x) = x - 1$

Graphiquement, $y = x - 1$ est une équation de la droite tangente au graphe de f au point $(1,0)$.



4.2.2. Fonctions de plusieurs variables (cf. cours p. 96)

Soit F une fonction de \mathbb{R}^2 dans \mathbb{R} , soit (x_1^*, x_2^*) un élément du domaine de F .

On appelle **fonction approximation linéaire F_{lin} de F en (x_1^*, x_2^*)** , la fonction définie comme suit :

$$F_{lin}(x_1^*, x_2^*)(x_1, x_2) = F(x_1^*, x_2^*) + \frac{\partial F}{\partial x_1}(x_1^*, x_2^*) \cdot (x_1 - x_1^*) + \frac{\partial F}{\partial x_2}(x_1^*, x_2^*) \cdot (x_2 - x_2^*)$$

Notons que $y = F(x_1^*, x_2^*) + \frac{\partial F}{\partial x_1}(x_1^*, x_2^*) \cdot (x_1 - x_1^*) + \frac{\partial F}{\partial x_2}(x_1^*, x_2^*) \cdot (x_2 - x_2^*)$ est une équation du **plan tangent** au graphe de F au point $(x_1^*, x_2^*, F(x_1^*, x_2^*))$.

De manière générale :

Soit F une fonction de \mathbb{R}^n dans \mathbb{R} , soit (x_1^*, \dots, x_n^*) un élément du domaine de F .

On appelle **fonction approximation linéaire F_{lin} de F en (x_1^*, \dots, x_n^*)** , la fonction définie comme suit :

$$F_{lin}(x_1^*, \dots, x_n^*)(x_1, \dots, x_n) = F(x_1^*, \dots, x_n^*) + \frac{\partial F}{\partial x_1}(x_1^*, \dots, x_n^*) \cdot (x_1 - x_1^*) + \dots + \frac{\partial F}{\partial x_n}(x_1^*, \dots, x_n^*) \cdot (x_n - x_n^*)$$

Notons que $y = F(x_1^*, \dots, x_n^*) + \frac{\partial F}{\partial x_1}(x_1^*, \dots, x_n^*) \cdot (x_1 - x_1^*) + \dots + \frac{\partial F}{\partial x_n}(x_1^*, \dots, x_n^*) \cdot (x_n - x_n^*)$ est une équation de l'**hyperplan** tangent au graphe de F au point $(x_1^*, \dots, x_n^*, F(x_1^*, \dots, x_n^*))$ (plus de représentation graphique possible).

Exercice résolu

Déterminer la fonction approximation linéaire F_{lin} de la fonction F en $(1,-2)$, avec

$$F : \mathbb{R}^2 \rightarrow \mathbb{R} : (x_1, x_2) \rightarrow F(x_1, x_2) = x_2^2 \cdot e^{(2 \cdot x_1 + x_2)} + x_1^3 \cdot x_2$$

Résolution

On sait que

$$F_{lin}(x_1^*, x_2^*) (x_1, x_2) = F(x_1^*, x_2^*) + \frac{\partial F}{\partial x_1}(x_1^*, x_2^*) \cdot (x_1 - x_1^*) + \frac{\partial F}{\partial x_2}(x_1^*, x_2^*) \cdot (x_2 - x_2^*)$$

Ici,

- $(x_1^*, x_2^*) = (1, -2)$
- $F(x_1^*, x_2^*) = F(1, -2) = 2$
- $\frac{\partial F}{\partial x_1}(x_1, x_2) = 2 x_2^2 e^{(2 \cdot x_1 + x_2)} + 3 x_1^2 x_2 \quad \Rightarrow \quad \frac{\partial F}{\partial x_1}(x_1^*, x_2^*) = \frac{\partial F}{\partial x_1}(1, -2) = 2$
- $\frac{\partial F}{\partial x_2}(x_1, x_2) = 2 x_2 e^{(2 \cdot x_1 + x_2)} + x_2^2 e^{(2 \cdot x_1 + x_2)} + x_1^3 \quad \Rightarrow \quad \frac{\partial F}{\partial x_2}(x_1^*, x_2^*) = \frac{\partial F}{\partial x_2}(1, -2) = 1$

La fonction approximation linéaire F_{lin} de la fonction F en $(1,-2)$ est donc donnée par

$$F_{lin(1,-2)}(x_1, x_2) = 2 + 2 \cdot (x_1 - 1) + 1 \cdot (x_2 + 2)$$

c'est-à-dire

$$F_{lin(1,-2)}(x_1, x_2) = 2 \cdot x_1 + x_2 + 2$$

Graphiquement, $y = 2 \cdot x_1 + x_2 + 2$ est une équation du plan tangent au graphe de F au point $(1, -2, 2)$.

Exercices

Exercice 3

1. Déterminer la fonction approximation linéaire f_{lin} de la fonction f en $x^*=2$, où

$$f : \mathbb{R} \rightarrow \mathbb{R} : x \rightarrow f(x) = \frac{3 \cdot x^2 + 2}{2 \cdot x + 3}$$

2. Déterminer la fonction approximation linéaire g_{lin} de la fonction g en $x^*=0$, où

$$g : \mathbb{R} \rightarrow \mathbb{R} : x \rightarrow g(x) = e^x$$

3. Déterminer la fonction approximation linéaire G_{lin} de la fonction G en $(z_1^*, z_2^*, z_3^*) = (2, 1, 0)$, où

$$G : \mathbb{R}^3 \rightarrow \mathbb{R} : (z_1, z_2, z_3) \rightarrow G(z_1, z_2, z_3) = (z_1^2 + z_2^3) \cdot e^{z_3}$$

4. Déterminer la fonction approximation linéaire H_{lin} de la fonction H en $(x_1^*, x_2^*, x_3^*) = (3, e, 1)$, où

$$H : \mathbb{R}^3 \rightarrow \mathbb{R} : (x_1, x_2, x_3) \rightarrow H(x_1, x_2, x_3) = x_1^3 \cdot \ln(x_2^2) + x_1 x_2 x_3^3$$

4.2.3. Méthode de Newton (voir cours p. 119-123)

- A quoi sert la méthode de Newton ?

La **méthode de Newton** permet de trouver **une** solution de l'équation $f(x) = 0$ (où f est une fonction d'une variable x), ou du moins une approximation de cette solution.

En d'autres termes, la **méthode de Newton** permet de trouver un réel x^* tel que $f(x^*)=0$.

- Description de la méthode

La formule de récurrence de la méthode de Newton est donné par

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

(pour comprendre l'origine de cette formule, voir cours p. 120).

Comment l'appliquer ?

- Choisir x_0 tel que $f'(x_0) \neq 0$ (tangente au graphe de f en x_0 non horizontale).
 - Calculer $x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$
 - Calculer $x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}$
 - etc ...
- e) Si la suite de réels $x_0, x_1, x_2, \dots, x_k$ converge vers un certain réel x^* , alors x^* est une solution de l'équation $f(x) = 0$.

Exercice résolu

En utilisant la méthode de Newton et en partant de $x_0 = 0$, trouver la valeur approchée d'une solution de l'équation suivante :

$$4x^3 + 8x^2 - 31x - 35 = 0$$

Résolution

Posons $f(x) = 4x^3 + 8x^2 - 31x - 35$; donc, $f'(x) = 12x^2 + 16x - 31$.

Valeur initiale : $x_0 = 0$.

$$1^{\text{ère}} \text{ itération : } x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} = 0 - \frac{f(0)}{f'(0)} = 0 - \frac{-35}{-31} \approx -1,129032258$$

$$2^{\text{ème}} \text{ itération : } x_2 = x_1 - \frac{f(x_1)}{f'(x_1)} = -\frac{35}{31} - \frac{f\left(\frac{-35}{31}\right)}{f'\left(\frac{-35}{31}\right)} \approx -0,997518839$$

$$3^{\text{ème}} \text{ itération : } x_3 = x_2 - \frac{f(x_2)}{f'(x_2)} = -0,997518839 - \frac{f(-0,997518839)}{f'(-0,997518839)} \approx -0,9999993$$

$$4^{\text{ème}} \text{ itération : } x_4 = x_3 - \frac{f(x_3)}{f'(x_3)} = -0,9999993 - \frac{f(-0,9999993)}{f'(-0,9999993)} = -1$$

Vérification :

-1 est bien une solution de l'équation $4x^3 + 8x^2 - 31x - 35 = 0$

$$\text{car } 4.(-1)^3 + 8.(-1)^2 - 31.(-1) - 35 = 0$$

Question subsidiaire : Comment pourrait-on continuer l'exercice pour déterminer toutes les solutions de l'équation ?

Exercices

Exercice 4

En utilisant la méthode de Newton et en partant de $x_0 = 0$, trouver la valeur approchée d'une solution de chacune des équations suivantes, obtenue après 3 itérations.

a) $\frac{1}{3}q^3 + 7q^2 + 111q + 500 = 0$

b) $x^3 - 1,5x^2 - 5,5x + 3 = 0$

c) $\cos(x) = 0$

d) $e^x = x^4$

Exercice 5

Dessinez le graphe d'une fonction f telle que $f(2) = 0$ et pour laquelle, en prenant comme point de départ pour l'application de la méthode de Newton la valeur $X_0 = 1$, vous obtenez pour X_1 une solution de l'équation $f(X) = 0$.

Exercice 6

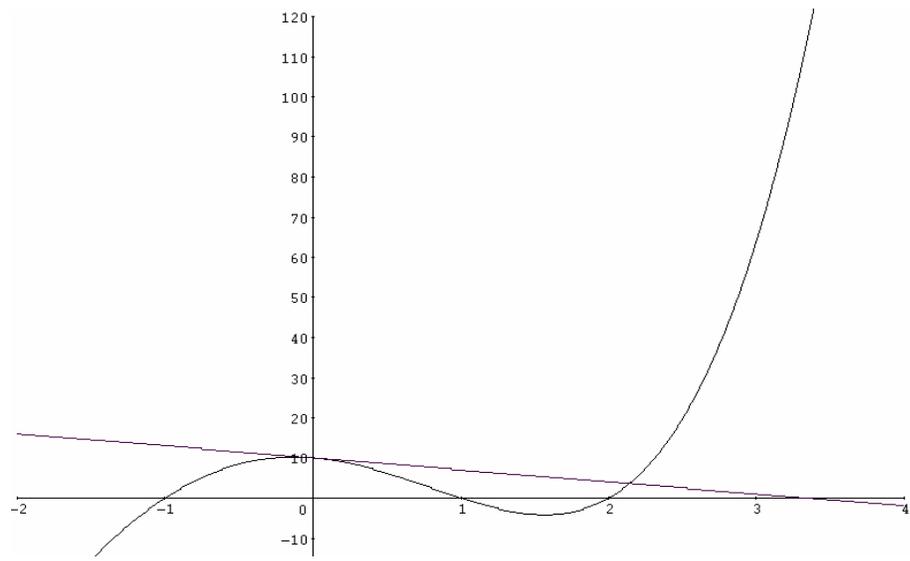
Trouver **graphiquement** (cf. graphe ci-dessous) vers quelle solution de l'équation

$$x^4 + 3x^3 - 11x^2 - 3x + 10 = 0$$

la méthode de Newton convergerait en prenant comme valeur de départ $x_0 = 0$.

Le graphe ci-dessous présente une représentation de la fonction

$$f(x) = x^4 + 3x^3 - 11x^2 - 3x + 10.$$



Solutions des exercices (Chapitre 4)

Exercice 1

$$1. \frac{\partial f}{\partial x_1}(x_1, x_2) = 3x_1^2 - 3x_2$$

$$\frac{\partial f}{\partial x_2}(x_1, x_2) = 3x_2^2 - 3x_1$$

$$2. \frac{\partial f}{\partial x_1}(x_1, x_2) = 2x_1 \cdot e^{(x_1^2+x_2^2)}$$

$$\frac{\partial f}{\partial x_2}(x_1, x_2) = 2x_2 \cdot e^{(x_1^2+x_2^2)}$$

$$3. \frac{\partial f}{\partial x_1}(x_1, x_2) = \frac{1}{x_2} + x_2 \cdot \frac{1}{x_1}$$

$$\frac{\partial f}{\partial x_2}(x_1, x_2) = -\frac{x_1}{(x_2)^2} + \ln x_1$$

$$4. \frac{\partial f}{\partial x_1}(x_1, x_2) = \frac{1}{x_1}$$

$$\frac{\partial f}{\partial x_2}(x_1, x_2) = \frac{1}{x_2}$$

$$5. \frac{\partial f}{\partial x_1}(x_1, x_2) = \cos x_1 - 2x_1 x_2 \sin(x_1^2)$$

$$\frac{\partial f}{\partial x_2}(x_1, x_2) = \cos(x_1^2)$$

$$6. \frac{\partial f}{\partial x_1}(x_1, x_2, x_3) = x_2^2 \cdot e^{x_1 \cdot x_2^2} + \ln(x_2 x_3^3)$$

$$\frac{\partial f}{\partial x_2}(x_1, x_2, x_3) = 2x_1 x_2 e^{x_1 \cdot x_2^2} + x_1 \cdot \frac{1}{x_2}$$

$$\frac{\partial f}{\partial x_3}(x_1, x_2, x_3) = 3x_1 \frac{1}{x_3}$$

Exercice 2

$$1. \frac{\partial^2 f}{(\partial x_1)^2}(x_1, x_2) = 6x_1$$

$$\frac{\partial^2 f}{\partial x_1 \cdot \partial x_2}(x_1, x_2) = -3$$

$$\frac{\partial^2 f}{(\partial x_2)^2}(x_1, x_2) = 6x_2$$

$$\frac{\partial^2 f}{\partial x_2 \cdot \partial x_1}(x_1, x_2) = -3$$

$$2. \frac{\partial^2 f}{(\partial x_1)^2}(x_1, x_2) = 2 \cdot e^{(x_1^2+x_2^2)} + 4 \cdot x_1^2 \cdot e^{(x_1^2+x_2^2)}$$

$$\frac{\partial^2 f}{\partial x_1 \cdot \partial x_2}(x_1, x_2) = 4x_1 x_2 \cdot e^{(x_1^2+x_2^2)}$$

$$\frac{\partial^2 f}{(\partial x_2)^2}(x_1, x_2) = 2 \cdot e^{(x_1^2+x_2^2)} + 4 \cdot x_2^2 \cdot e^{(x_1^2+x_2^2)}$$

$$\frac{\partial^2 f}{\partial x_2 \cdot \partial x_1}(x_1, x_2) = 4x_1 x_2 \cdot e^{(x_1^2+x_2^2)}$$

$$3. \frac{\partial^2 f}{(\partial x_1)^2}(x_1, x_2) = -\frac{x_2}{x_1^2}$$

$$\frac{\partial^2 f}{\partial x_1 \cdot \partial x_2}(x_1, x_2) = -\frac{1}{x_2^2} + \frac{1}{x_1}$$

$$\frac{\partial^2 f}{(\partial x_2)^2}(x_1, x_2) = 2 \cdot \frac{x_1}{x_2^3}$$

$$\frac{\partial^2 f}{\partial x_2 \cdot \partial x_1}(x_1, x_2) = -\frac{1}{x_2^2} + \frac{1}{x_1}$$

Exercice 3

1. $f_{lin(2)}(x) = \frac{8}{7}x - \frac{2}{7}$

2. $g_{lin(0)}(x) = x + 1$

3. $G_{lin(2,1,0)}(z_1, z_2, z_3) = 4z_1 + 3z_2 + 5z_3 - 6$

4. $H_{lin(3,e,1)}(x_1, x_2, x_3) = (54 + e) \cdot x_1 + \left(\frac{54}{e} + 3\right) \cdot x_2 + 9e x_3 - 162 - 18e$

Exercice 4

a) $q_0 = 0$

$q_1 \approx -4,5045045$

$q_2 \approx -3,616946$

$q_3 \approx -3,56640929$

Solution réelle $\approx -3,56625339$

b) $x_0 = 0$

$x_1 \approx 0,54545454$

$x_2 \approx 0,49996991757$

$x_3 \approx 0,5$

Solution réelle = 0,5

c) Impossible d'appliquer la méthode de Newton au départ de 0 car $\sin(0) = 0$

d) $x_0 = 0$

$x_1 = -1$

$x_2 \approx -0,85527976$

$x_3 \approx -0,817730774$

Solution réelle $\approx -0,8155534188$

Exercice 6

Convergence vers la valeur $x^* = 2$.

Chapitre 5 : Différentielles

5.1. Différentier une fonction

5.1.1. Notion de différentielle d'une fonction de \mathbf{R} dans \mathbf{R}

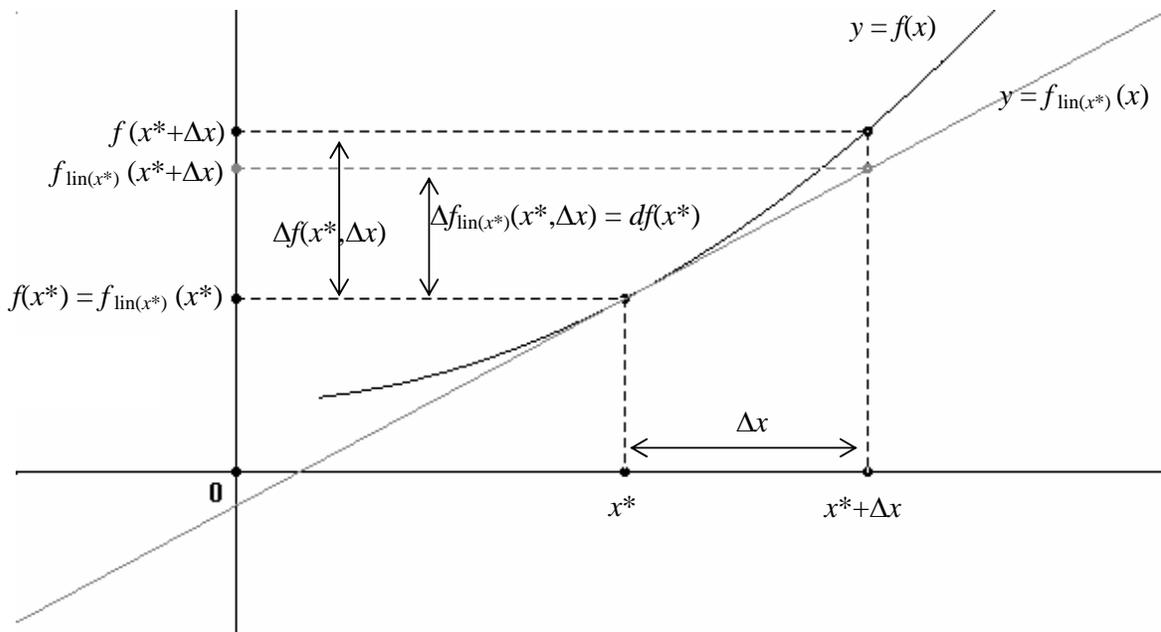
Soit f une fonction de \mathbf{R} dans \mathbf{R} continûment différentiable sur un sous-ensemble D de son domaine de définition et soit $x^* \in D$.

On appelle différentielle de f en x^* et on note $df(x^*)$ l'expression $f'(x^*) \cdot \Delta x$.

Remarques

1) $df(x^*)$ est une expression linéaire en la variation Δx de la variable x (au départ de la valeur x^*).

2) $df(x^*) = \Delta f_{\text{lin}(x^*)}(x^*, \Delta x)$ (à vérifier)



3) Pour la fonction identité $I : \mathbf{R} \rightarrow \mathbf{R} : x \rightarrow x$, on a $I'(x^*) = 1$.

Par conséquent, $dI(x^*) = 1 \cdot \Delta x = \Delta x$.

Comme la différentielle $dI(x^*)$ est conventionnellement notée dx , il vient :

$$dI(x^*) \equiv dx = \Delta x.$$

On substitue généralement dx à Δx dans l'écriture de la différentielle d'une fonction d'une variable x et on obtient alors l'expression : $df(x^*) = f'(x^*) \cdot dx$.

On voit même parfois écrit : $df(x) = f'(x) \cdot dx$

5.1.2. Notion de différentielle d'une fonction de \mathbb{R}^n dans \mathbb{R}

Soit F une fonction de \mathbb{R}^n dans \mathbb{R} continûment différentiable sur un sous-ensemble D de son domaine de définition et soit $\mathbf{x}^* = (x_1^*, x_2^*, \dots, x_n^*) \in D$.

On appelle **différentielle de F en \mathbf{x}^*** et on note $dF(\mathbf{x}^*)$ l'expression linéaire en les variations $\Delta x_1, \Delta x_2, \dots, \Delta x_n$:

$$\frac{\partial F}{\partial x_1}(\mathbf{x}^*) \cdot \Delta x_1 + \frac{\partial F}{\partial x_2}(\mathbf{x}^*) \cdot \Delta x_2 + \dots + \frac{\partial F}{\partial x_n}(\mathbf{x}^*) \cdot \Delta x_n.$$

Remarques

1) $dF(\mathbf{x}^*) = \Delta F_{\text{lin}(\mathbf{x}^*)}(\mathbf{x}^*, \Delta \mathbf{x})$

2) Pour des raisons analogues à celles expliquées dans le cas d'une fonction d'une variable, on écrit généralement :

$$dF(\mathbf{x}^*) = \frac{\partial F}{\partial x_1}(\mathbf{x}^*) \cdot dx_1 + \frac{\partial F}{\partial x_2}(\mathbf{x}^*) \cdot dx_2 + \dots + \frac{\partial F}{\partial x_n}(\mathbf{x}^*) \cdot dx_n$$

où, pour chaque valeur de i , dx_i peut être considéré

- ♦ soit comme une variation de la variable x_i au départ de la valeur x_i^* ($dx_i = x_i - x_i^*$)
- ♦ soit comme une différentielle (différentielle de la fonction identité).

Exercices

Exercice 1

Calculer la différentielle de

a) $f(x) = x^2 \cdot \ln(x)$

b) $f(x) = \sin(2x)$

c) $F(x_1, x_2) = x_1 - x_2^3$

d) $G(z_1, z_2) = (z_1 - z_2)^3$

e) $F(x, y, t, v) = x \cdot y + a \cdot t^2 - v \cdot x + 2$ (où a est un paramètre)

f) $F(x, y, t, v) = x \cdot y + a \cdot t^2 - v \cdot x + 2$ où a est un paramètre, et où $t(x, y) = e^{x \cdot y}$ et $v(y) = 2 - y$

g) $H(z) = -5$

Exercice 2

a) On donne $F : \mathbb{R}^2 \rightarrow \mathbb{R} : (x_1, x_2) \rightarrow f(x_1, x_2) = x_1^2 \cdot x_2^3$

Déterminer l'approximation de la variation de valeur de F

correspondant aux variations de valeur $\Delta x_1 = \frac{3}{10}$ et $\Delta x_2 = \frac{2}{10}$ au point $(x_1^*, x_2^*) = (3, 1)$.

b) On donne $F : \mathbb{R}^2 \rightarrow \mathbb{R} : (x_1, x_2) \rightarrow F(x_1, x_2) = x_2 \cdot e^{x_1}$

Déterminer l'approximation de la variation de valeur de F

correspondant aux variations de valeur $\Delta x_1 = \frac{1}{10}$ et $\Delta x_2 = \frac{1}{10}$ au point $(x_1^*, x_2^*) = (1, 4)$.

5.2. Equations aux variations différentielles

Soit $E(x,y) = b$, un modèle mathématique qui ne permet pas nécessairement de définir y en fonction de x .

- L' **équation linéaire « approchée »**, au voisinage de (x^*,y^*) , **du modèle** $E(x,y) = b$

est

$$\frac{\partial E}{\partial x}(x^*,y^*) \cdot (x-x^*) + \frac{\partial E}{\partial y}(x^*,y^*) \cdot (y-y^*) = 0$$

C'est l' **équation de la droite tangente**, en (x^*,y^*) , à la courbe d'équation $E(x,y) = b$.

- L' **équation aux variations différentielles**

associée au modèle $E(x,y) = b$
en sa solution (x^*,y^*)

est
$$\frac{\partial E}{\partial x}(x^*,y^*) \cdot dx + \frac{\partial E}{\partial y}(x^*,y^*) \cdot dy = 0$$

- L' **équation** $E(x,y) = b$ **définit localement**, au voisinage de (x^*,y^*) ,
la variable y en fonction de la variable x

si et seulement si

son équation linéaire « approchée », en (x^*,y^*) ,
définit (explicitement) la variable y en fonction de la variable x .

si et seulement si

son équation aux variations différentielles, en (x^*,y^*) ,
définit (explicitement) la variable dy en fonction de la variable dx .

- Si l'équation $E(x,y) = b$ définit, au voisinage de (x^*,y^*) , la fonction $y = y(x)$ (même si on ne possède pas explicitement l'expression de cette fonction) alors, le nombre dérivé de cette fonction en x^* est égal à

$$\left(\frac{d}{dy} y \right) (x^*) = \frac{dy}{dx}$$

Exercice résolu

- a) L'équation $x^2 + y^2 = 25$
définit-elle **globalement** la variable y en fonction de la variable x ?
... **autre formulation** : Peut-on exprimer explicitement la variable y en fonction de la variable x
à partir de l'équation $x^2 + y^2 = 25$?
- b) L'équation définit-elle **localement**, autour de sa solution $(x^*, y^*) = (3, 4)$, la variable y
en fonction de la variable x ?
Déterminer le nombre dérivé de la fonction ainsi définie en $(x^*, y^*) = (3, 4)$.
- c) Quelle est l'équation de la droite tangente à la courbe d'équation $x^2 + y^2 = 25$ en
 $(x^*, y^*) = (3, 4)$?
- d) L'équation définit-elle **localement**, autour de sa solution $(x^*, y^*) = (5, 0)$, la variable y
en fonction de la variable x ?
Déterminer le nombre dérivé de la fonction ainsi définie en $(x^*, y^*) = (5, 0)$.
- e) Quelle est l'équation de la droite tangente à la courbe d'équation $x^2 + y^2 = 25$ en
 $(x^*, y^*) = (5, 0)$?

Résolution :

a) Non, car $y = \sqrt{25 - x^2}$ ou $y = -\sqrt{25 - x^2}$

- b) Considérons l'équation aux variations différentielles en $(3, 4)$
associée à l'équation de départ : $\frac{\partial E}{\partial x}(3, 4) \cdot dx + \frac{\partial E}{\partial y}(3, 4) \cdot dy = 0$

$$\frac{\partial E}{\partial x}(x, y) = 2x \quad \Rightarrow \quad \frac{\partial E}{\partial x}(3, 4) = 6$$

$$\frac{\partial E}{\partial y}(x, y) = 2y \quad \Rightarrow \quad \frac{\partial E}{\partial y}(3, 4) = 8$$

On obtient $6 dx + 8 dy = 0 \quad \Rightarrow \quad dy = -\frac{3}{4} dx$

L'équation aux variations différentielles en $(3, 4)$

définit la variable dy en fonction de la variable dx .

L'équation de départ définit donc localement, au voisinage de $(3, 4)$,
la variable y en fonction de la variable x .

En $(3, 4)$, y est positif et donc $y = \sqrt{25 - x^2}$

Le nombre dérivé demandé est $\left(\frac{d}{dy} y \right) (3) = \frac{dy}{dx} = -\frac{3}{4}$

- c) L'équation de la droite tangente en $(3, 4)$,
est donnée par $\frac{\partial E}{\partial x}(3, 4) \cdot (x-3) + \frac{\partial E}{\partial y}(3, 4) \cdot (y-4) = 0$ (équation linéaire approchée)
 $6(x-3) + 8(y-4) = 0$

$$6x + 8y - 50 = 0$$

Notons que l'équation linéaire approchée en (3,4) définit (explicitement),

la variable y en fonction de la variable x : $y = -\frac{3}{4}x + \frac{25}{4}$

d) Considérons l'équation aux variations différentielles en (5,0)

associée à l'équation de départ : $\frac{\partial E}{\partial x}(5,0) \cdot dx + \frac{\partial E}{\partial y}(5,0) \cdot dy = 0$

Puisque $\frac{\partial E}{\partial x}(5,0) = 5$ et $\frac{\partial E}{\partial y}(5,0) = 0$, on obtient $5 dx + 0 dy = 0$

L'équation aux variations différentielles en (5,0) ne définit pas la variable dy en fonction de la variable dx .

L'équation de départ ne définit donc pas localement, au voisinage de (5,0), la variable y en fonction de la variable x .

e) La droite tangente en (5,0) existe et a pour équation $x = 5$ (tangente verticale).

Exercices

Exercice 3

a) L'équation $x^3 - 2x^2y + 3xy^2 - 22 = 0$ définit-elle **globalement** la variable y en fonction de la variable x ?

... *autre formulation* : Peut-on exprimer explicitement la variable y en fonction de la variable x à partir de l'équation $x^3 - 2x^2y + 3xy^2 - 22 = 0$?

b) Si non, l'équation définit-elle **localement**, autour de sa solution $(x^*, y^*) = (1, 3)$, la variable y en fonction de la variable x ?

Déterminer le nombre dérivé de la fonction ainsi définie en $(x^*, y^*) = (1, 3)$.

c) Quelle est l'équation de la droite tangente à la courbe d'équation $x^3 - 2x^2y + 3xy^2 - 22 = 0$ en $(x^*, y^*) = (1, 3)$?

Exercice 4

a) L'équation $x \cdot y^2 - 2^{x,y} - 1 = 0$

définit-elle **globalement** la variable y en fonction de la variable x ?

b) Si non, l'équation définit-elle **localement**, autour de sa solution $(x^*, y^*) = (1, 3)$, la variable y en fonction de la variable x ?

Déterminer le nombre dérivé de la fonction ainsi définie en $x^* = 1$.

c) Quelle est l'équation de la droite tangente à la courbe d'équation $x \cdot y^2 - 2^{x,y} - 1 = 0$ en $(x^*, y^*) = (1, 3)$?

d) Quelle est le nombre dérivée de y par rapport à x en $x^* = 1$?

Exercice 5 : généralisation pour un modèle à trois variables

- a) L'équation $x_1^2 \cdot x_2 \cdot y + x_1 \cdot y^2 - 2 \cdot x_2 \cdot y + 2 \cdot y^2 - 12 = 0$ définit-elle **globalement** la variable y en fonction des variables x_1 et x_2 ?
- b) Si non, l'équation définit-elle **localement**, au voisinage de sa solution $(x_1^*, x_2^*, y^*) = (1, 0, 2)$, la variable y en fonction des variables x_1 et x_2 ?
- c) Quelle est l'équation du plan tangent à la courbe d'équation $x_1^2 \cdot x_2 \cdot y + x_1 \cdot y^2 - 2 \cdot x_2 \cdot y + 2 \cdot y^2 - 12 = 0$ en $(x_1^*, x_2^*, y^*) = (1, 0, 2)$?

Exercice 6 : généralisation pour un modèle à trois variables

- a) L'équation $e^{x_1 \cdot y} \cdot x_2 + 2 \cdot x_1^2 \cdot x_2^3 - 3 \cdot y^2 - 4 \cdot x_1 \cdot x_2 - 1 = 0$ définit-elle **localement**, autour de $(x_1^*, x_2^*, y^*) = (2, 1, 0)$, la variable y en fonction des variables x_1 et x_2 ?
- b) Quelle est l'équation du plan tangent à la courbe d'équation $e^{x_1 \cdot y} \cdot x_2 + 2 \cdot x_1^2 \cdot x_2^3 - 3 \cdot y^2 - 4 \cdot x_1 \cdot x_2 - 1 = 0$ en $(x_1^*, x_2^*, y^*) = (2, 1, 0)$?

Exercice 7

Soit l'équation à 2 variables x et y : $x^2 \cdot y^2 - x^3 \cdot y + 2 \cdot x \cdot y^3 = 0$

Comme on le vérifie aisément, cette équation permet d'exprimer la variable y en fonction de la variable x localement au voisinage de sa solution $(x^*, y^*) = (2, 1)$.

- a) Déterminez le nombre dérivé de cette fonction en $x^* = 2$.
- b) Déterminez l'élasticité de type point de y par rapport à x au point (x^*, y^*) .

Exercice résolu

Le modèle
$$\begin{cases} x_1 \cdot y_1 + x_2 \cdot y_2 = 0 \\ y_1^2 + 2 \cdot x_1 \cdot x_2 + e^{y_1 \cdot y_2} = 5 \end{cases}$$

ne définit pas **globalement** les variables y_1 et y_2 en fonction des variables x_1 et x_2 .

Ce modèle définit-il **localement**, au voisinage de sa solution $(x_1^*, x_2^*, y_1^*, y_2^*) = (0, 1, 2, 0)$, les variables y_1 et y_2 en fonction des variables x_1 et x_2 ?

(C'est-à-dire le système d'équations aux variations différentielles définit-il dy_1 et dy_2 en fonction de dx_1 et dx_2 ?)

Résolution :

- NB : $(0, 1, 2, 0)$ est bien solution du modèle
$$\begin{cases} x_1 \cdot y_1 + x_2 \cdot y_2 = 0 \\ y_1^2 + 2 \cdot x_1 \cdot x_2 + e^{y_1 \cdot y_2} = 5 \end{cases}$$

- « Passage à la différentielle » en $(0, 1, 2, 0)$ (syst. d'équ. aux variations différentielles)

$$\begin{cases} y_1 \cdot dx_1 + y_2 \cdot dx_2 + x_1 \cdot dy_1 + x_2 \cdot dy_2 = 0 \\ 2 \cdot x_2 \cdot dx_1 + 2 \cdot x_1 \cdot dx_2 + (2 \cdot y_1 + y_2 \cdot e^{y_1 \cdot y_2}) \cdot dy_1 + (y_1 \cdot e^{y_1 \cdot y_2}) \cdot dy_2 = 0 \end{cases}$$

en $(0, 1, 2, 0)$:

$$\begin{cases} 2 \cdot dx_1 + 0 \cdot dx_2 + 0 \cdot dy_1 + 1 \cdot dy_2 = 0 \\ 2 \cdot dx_1 + 0 \cdot dx_2 + 4 \cdot dy_1 + 2 \cdot dy_2 = 0 \end{cases}$$

- On peut facilement exprimer dy_1 en fonction de dx_1, dx_2 et dy_2 en fonction de dx_1, dx_2

$$\begin{cases} dy_2 = -2 \cdot dx_1 \\ 4 \cdot dy_1 + 2 \cdot dy_2 = -2 \cdot dx_1 \end{cases} \quad \begin{cases} dy_2 = -2 \cdot dx_1 \\ 4 \cdot dy_1 = -2 \cdot dx_1 - 2 \cdot dy_2 \end{cases} \quad \begin{cases} dy_2 = -2 \cdot dx_1 \\ 4 \cdot dy_1 = -2 \cdot dx_1 - 2 \cdot (-2 \cdot dx_1) \end{cases}$$

$$\begin{cases} dy_2 = -2 \cdot dx_1 \\ 4 \cdot dy_1 = 2 \cdot dx_1 \end{cases}$$

$$\begin{cases} dy_2 = -2 \cdot dx_1 \\ dy_1 = \frac{1}{2} \cdot dx_1 \end{cases} \quad \text{ce qui prouve la capacité d'exprimer } dy_1 \text{ et } dy_2 \text{ en fonction de } dx_1 \text{ et } dx_2$$

En conséquence, le système de départ définit localement, au voisinage de sa solution $(0, 1, 2, 0)$, les variables y_1 et y_2 en fonction des variables x_1 et x_2 .

Exercices

Exercice 8

$$\text{Le modèle } \begin{cases} x_1^2 \cdot y_1 \cdot y_2^3 + x_1 \cdot x_2 \cdot y_2 = 15 \\ e^{x_1} \cdot \frac{y_1^2}{y_2} + x_2 \cdot y_2 \cdot (1 - x_1) = 0 \end{cases}$$

ne définit pas **globalement** les variables y_1 et y_2 en fonction des variables x_1 et x_2 .

Ce modèle définit-il **localement**, au voisinage de sa solution $(x_1^*, x_2^*, y_1^*, y_2^*) = (1, 3, 0, 5)$, les variables y_1 et y_2 en fonction des variables x_1 et x_2 ?

Exercice 9

$$\text{On donne le modèle } \begin{cases} 2 \cdot x_1 \cdot y_1 + x_2 \cdot x_3^2 \cdot y_2 - x_1 \cdot y_1 \cdot y_2^2 = 2 \\ x_1^3 \cdot y_1^2 + 3 \cdot x_1 \cdot x_2 \cdot x_3 + 4 \cdot \frac{y_1}{y_2} = -3 \end{cases}, \text{ où on considère les variables}$$

y_1 et y_2 comme endogènes et les variables x_1 , x_2 et x_3 comme exogènes.

Ce modèle définit-il **localement**, au voisinage de $(y_1^*, y_2^*, x_1^*, x_2^*, x_3^*) = (3, -2, 1, -1, 2)$ les variables y_1 et y_2 en fonction des variables x_1 , x_2 et x_3 ?

Si oui, exprimer explicitement dy_1 en fonction de dx_1 , dx_2 et dx_3
 dy_2 en fonction de dx_1 , dx_2 et dx_3 .

Exercice 10

$$\text{On donne le modèle } \begin{cases} e^{y_1 \cdot y_2} \cdot x_1 + 2 \cdot y_1^2 \cdot y_2^3 - e^{2 \cdot x_1} - 4 \cdot x_1 \cdot x_2 - 2 \cdot y_2 - 6 = 0 \\ x_1^2 \cdot x_2^3 - y_1^3 + x_2 \cdot y_2 + x_1 \cdot y_1 \cdot y_2 + 6 = 0 \end{cases}, \text{ où on}$$

considère les variables y_1 et y_2 comme endogènes et les variables x_1 et x_2 comme exogènes.

Ce modèle définit-il **localement**, au voisinage de sa solution $(x_1^*, x_2^*, y_1^*, y_2^*) = (1, 0, 2, 1)$, les variables y_1 et y_2 en fonction des variables x_1 et x_2 ?

Si oui, exprimer explicitement dy_1 en fonction de dx_1 et dx_2
 dy_2 en fonction de dx_1 et dx_2 .

Exercice 11

Le modèle
$$\begin{cases} x_1 \cdot x_2 \cdot y_1 \cdot y_2 \cdot y_3 - x_2^2 = 0 \\ x_1 \cdot y_2^2 + x_2 \cdot y_1 \cdot y_3 - 5 \cdot y_2^3 \cdot y_3^2 = 0 \\ x_1^3 \cdot x_2 \cdot y_3 - y_1^2 \cdot y_2 + 2 \cdot y_2 \cdot y_3 = 0 \end{cases}$$
 définit-il localement, autour de sa

solution $(x_1^*, x_2^*, y_1^*, y_2^*, y_3^*) = (1, 2, 2, 1, 1)$, les variables y_1, y_2 et y_3 en fonction des variables x_1 et x_2 ?

Si oui, exprimer dy_1, dy_2 et dy_3 en fonction de dx_1 et dx_2 .

Exercice 12

Le modèle
$$\begin{cases} x_2^2 \cdot y_2^3 - x_1 \cdot y_1^2 \cdot y_2 - x_1^2 \cdot x_2 + y_2^2 - 10 = 0 \\ 3 \cdot x_1^2 + 2 \cdot x_1 \cdot y_1 + y_1 \cdot y_2^3 + 5 \cdot x_2^3 - 3 \cdot y_2^2 - 13 = 0 \end{cases}$$
 définit-il localement, au

voisinage de sa solution $(x_1^*, x_2^*, y_1^*, y_2^*) = (3, 0, 1, -2)$, les variables y_1 et y_2 en fonction des variables x_1 et x_2 ?

C'est-à-dire le système d'équations aux variations différentielles définit-il dy_1 et dy_2 en fonction de dx_1 et dx_2 ?

Si oui, exprimer explicitement dy_1 en fonction de dx_1 et dx_2

dy_2 en fonction de dx_1 et dx_2 .

Exercice 13

Soit le système de 3 équations à 5 variables y_1, y_2, y_3, x_1 et x_2

$$\begin{cases} x_1 = a \cdot y_1 + f(y_3) \\ y_2 = g(y_3, x_2) \\ y_3 = h(y_1) \end{cases}$$

- où
- a est une constante réelle
 - $f(\bullet)$ et $h(\bullet)$ sont deux fonctions d'une variable de classe C_1
 - $g(\bullet, \bullet)$ est une fonction de deux variables de classe C_1
 - les variables endogènes sont y_1, y_2 et y_3
 - les variables exogènes sont x_1 et x_2 .

Soit d'autre part $(y_1^*, y_2^*, y_3^*, x_1^*, x_2^*)$ une solution de ce système.

a) Ecrire le système d'équations aux variations différentielles associé au système ci-dessus en $(y_1^*, y_2^*, y_3^*, x_1^*, x_2^*)$.

b) Est-il possible d'exprimer les variables y_1, y_2 et y_3 en fonction des variables x_1 et x_2 localement, autour de la solution $(y_1^*, y_2^*, y_3^*, x_1^*, x_2^*)$? Justifiez votre réponse.

Exercice 14

Soit le système de 2 équations à 5 variables y_1, y_2, x_1, x_2 et x_3

$$\begin{cases} a.y_1.y_2 + x_1^3.x_2.x_3^2 - (a+1).x_2 = 0 \\ 2.y_1 + a.x_1.y_2 - x_2.x_3 - 2.a.x_1 = 0 \end{cases}$$

- où
- a désigne une constante réelle
 - les variables endogènes sont y_1 et y_2
 - les variables exogènes sont x_1, x_2 et x_3 .

On vérifie aisément que $(y_1^*, y_2^*, x_1^*, x_2^*, x_3^*) = (1, 2, 1, 2, 1)$ est une solution de ce système.

a) Ecrivez le système d'équations aux variations différentielles associé au système ci-dessus en $(y_1^*, y_2^*, x_1^*, x_2^*, x_3^*) = (1, 2, 1, 2, 1)$.

b) Pour quelles valeurs de a est-il possible d'exprimer les variables endogènes en fonction des variables exogènes localement au voisinage de $(y_1^*, y_2^*, x_1^*, x_2^*, x_3^*) = (1, 2, 1, 2, 1)$? Justifiez votre réponse.

c) En donnant à la constante réelle « a » la valeur « 2 », écrivez le système aux variations différentielles de la question a) en complétant les 6 cases ci-dessous :

$$\left\{ \begin{array}{l} \boxed{} dy_1 + \boxed{} dy_2 = \boxed{} \\ \boxed{} dy_1 + \boxed{} dy_2 = \boxed{} \end{array} \right.$$

d) Au départ du système de la question c) et en utilisant la méthode de Cramer, exprimez dy_1 et dy_2 en fonction de dx_1, dx_2 et dx_3 .

e) Lorsque $a = 2$, le système

$$\begin{cases} a.y_1.y_2 + x_1^3.x_2.x_3^2 - (a+1).x_2 = 0 \\ 2.y_1 + a.x_1.y_2 - x_2.x_3 - 2.a.x_1 = 0 \end{cases}$$

permet d'exprimer les variables y_1, y_2 en fonction des variables x_1, x_2 et x_3 localement au voisinage de $(y_1^*, y_2^*, x_1^*, x_2^*, x_3^*) = (1, 2, 1, 2, 1)$.

Notons F et G les deux fonctions ainsi définies :

$$y_1 = F(x_1, x_2, x_3) \quad \text{et} \quad y_2 = G(x_1, x_2, x_3).$$

Quelle est la valeur du nombre dérivée partielle par rapport à x_1 de la fonction F en $(x_1^*, x_2^*, x_3^*) = (1, 2, 1)$? Justifiez votre réponse.

Solutions des exercices (Chapitre 4)

Exercice 1

a) $df(x) = f'(x) dx = (2.x.\ln(x) + x) dx$

b) $df(x) = 2.\cos(2x) dx$

c) $dF(x_1, x_2) = dx_1 - 3 x_2^2 dx_2$

d) $dG(z_1, z_2) = 3 (z_1 - z_2)^2 dz_1 - 3 (z_1 - z_2)^2 dz_2$

e) $dF(x, y, t, v) = (y - v) dx + x dy + 2 a t dt - x dv$

f) NB : $dt(x, y) = y e^{x.y} dx + x e^{x.y} dy$ et $dv(y) = - dy$

$$dF(x, y) = (2y - 2 + 2 a y e^{x.y}) dx + x (2 + e^{x.y}) dy$$

g) $dH(z) = 0$

Exercice 2

$$a) dF((x_1^*, x_2^*), (\Delta x_1, \Delta x_2)) = \frac{\partial F}{\partial x_1}(x_1^*, x_2^*) \cdot \Delta x_1 + \frac{\partial F}{\partial x_2}(x_1^*, x_2^*) \cdot \Delta x_2$$

$$\frac{\partial F}{\partial x_1}(x_1^*, x_2^*) = 2 x_1 x_2^3 \quad \Rightarrow \quad \frac{\partial F}{\partial x_1}(3, 1) = 6$$

$$\frac{\partial F}{\partial x_2}(x_1^*, x_2^*) = 3 x_1^2 x_2^2 \quad \Rightarrow \quad \frac{\partial F}{\partial x_2}(3, 1) = 27$$

$$\text{Finalement, } dF\left((3, 1), \left(\frac{3}{10}, \frac{2}{10}\right)\right) = 6 \cdot \frac{3}{10} + 27 \cdot \frac{2}{10} = 7,2$$

... alors que la variation en valeur vaut

$$\Delta F((x_1^*, x_2^*), (\Delta x_1, \Delta x_2)) = F(x_1^* + \Delta x_1, x_2^* + \Delta x_2) - F(x_1^*, x_2^*)$$

$$\Delta F\left((3, 1), \left(\frac{3}{10}, \frac{2}{10}\right)\right) = F\left(\frac{33}{10}, \frac{12}{10}\right) - F(3, 1) = 9,817932\dots$$

b) $dF\left((1, 4), \left(\frac{1}{10}, \frac{1}{10}\right)\right) = \frac{e}{2} \approx 1,36$

... alors que la variation en valeur vaut

$$\Delta F\left((1, 4), \left(\frac{1}{10}, \frac{1}{10}\right)\right) = F\left(\frac{11}{10}, \frac{41}{10}\right) - F(1, 4) = 1,444$$

Exercice 3

a) non

b) Equation aux variations différentielles : $18.dx + 16.dy = 0$, qui peut encore s'écrire

$$dy = -\frac{9}{8}.dx.$$

L'équation aux variations différentielles en (1,3) définit la variable dy en fonction de la variable dx .

L'équation de départ définit donc localement, au voisinage de (1,3), la variable y en fonction de la variable x (mais on ne sait pas comment, on n'a pas une écriture explicite de cette fonction).

Le nombre dérivé demandé est $\left(\frac{d}{dy} y\right)(1) = \frac{dy}{dx} = -\frac{9}{8}$

c) L'équation de la droite tangente en (1,3) est donnée par $18.x + 16.y - 66 = 0$

Exercice 4

a) non

b) Equation aux variations différentielles : $(9 - 24.\ln 2).dx + (6 - 8.\ln 2).dy = 0$, qui peut

encore s'écrire $dy = -\frac{(9 - 24.\ln 2)}{(6 - 8.\ln 2)}.dx$

L'équation aux variations différentielles en (1,3) définit la variable dy en fonction de la variable dx .

L'équation de départ définit donc localement, autour de sa solution (1,3), la variable y en fonction de la variable x (mais on ne sait pas comment, on n'a pas une écriture explicite de cette fonction).

Le nombre dérivé demandé est $\left(\frac{d}{dy} y\right)(1) = \frac{dy}{dx} = -\frac{(9 - 24.\ln 2)}{(6 - 8.\ln 2)}$.

c) L'équation de la droite tangente en (1,3) est donnée par

$$(9 - 24.\ln 2).x + (6 - 8.\ln 2).y + (-27 + 48.\ln 2) = 0$$

Exercice 5

a) non

b) Equation aux variations différentielles : $4.dx_1 - 2.dx_2 + 12.dy = 0$, qui peut encore

s'écrire $dy = -\frac{1}{3}.dx_1 + \frac{1}{6}.dx_2$

L'équation aux variations différentielles en (1,0,2) définit la variable dy en fonction des variables dx_1 et dx_2 .

L'équation de départ définit donc localement, autour de sa solution (1,0,2), la variable y en fonction des variables x_1 et x_2 (mais on ne sait pas comment, on n'a pas une écriture explicite de la fonction).

c) Le plan tangent en (1,0,2) a pour équation $2.x_1 - x_2 + 6.y - 14 = 0$

Exercice 6

a) Equation aux variations différentielles : $4 \cdot dx_1 + 17 \cdot dx_2 + 2 \cdot dy = 0$, qui peut encore s'écrire $dy = -2 \cdot dx_1 - \frac{17}{2} \cdot dx_2$

L'équation aux variations différentielles définit donc localement, autour de sa solution $(2,1,0)$, la variable dy en fonction des variables dx_1 et dx_2 .

L'équation de départ ($E(x_1, x_2, y) = 0$) définit donc, au voisinage de $(2,1,0)$, la variable y en fonction des variables x_1 et x_2 (mais on ne sait pas comment, on n'a pas une écriture explicite de la fonction).

b) Le plan tangent (au graphe des solutions de $E(x_1, x_2, y) = 0$) en $(2,1,0)$ a pour équation

$$4 \cdot x_1 + 17 \cdot x_2 + 2 \cdot y - 25 = 0$$

Exercice 7

a) Le nombre dérivé demandé est donné par $\frac{dy}{dx} = \frac{1}{2}$

b) $\varepsilon_{y/x}(x^*) = f'(x^*) \cdot \frac{x^*}{y^*} = 1$

Exercice 8

• $(1,3,0,5)$ est bien solution du modèle $\begin{cases} x_1^2 \cdot y_1 \cdot y_2^3 + x_1 \cdot x_2 \cdot y_2 = 15 \\ e^{x_1} \cdot \frac{y_1^2}{y_2} + x_2 \cdot y_2 \cdot (1 - x_1) = 0 \end{cases}$

• Système d'équations aux variations différentielles en $(1,3,0,5)$:

$$\begin{cases} 15 \cdot dx_1 + 5 \cdot dx_2 + 125 \cdot dy_1 + 3 \cdot dy_2 = 0 \\ -15 \cdot dx_1 + 0 \cdot dx_2 + 0 \cdot dy_1 + 0 \cdot dy_2 = 0 \end{cases}$$

Le système est indéterminé ; on ne peut pas écrire dy_1 et dy_2 en fonction de dx_1 et dx_2 . En conséquence, le système ne définit pas localement, au voisinage de sa solution $(1,3,5,0)$, les variables y_1 et y_2 en fonction des variables x_1 et x_2 .

Autre justification : $\begin{cases} 25 \cdot dy_1 + 3 \cdot dy_2 = -5 \cdot dx_2 \\ 0 \cdot dy_1 + 0 \cdot dy_2 = 15 \cdot dx_1 \end{cases}$ s'écrit aussi $\begin{pmatrix} 25 & 3 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} dy_1 \\ dy_2 \end{pmatrix} = \begin{pmatrix} -5 \cdot dx_2 \\ 15 \cdot dx_1 \end{pmatrix}$

Puisque $\text{dét} \begin{pmatrix} 25 & 3 \\ 0 & 0 \end{pmatrix} = 0$, on ne peut pas exprimer dy_1 et dy_2 en fonction de dx_1 et dx_2 .

Exercice 9

Système d'équations aux variations différentielles en $(3, -2, 1, -1, 2)$:

$$\begin{cases} -dy_1 + 4.dy_2 - 3.dx_1 - 4.dx_2 + 4.dx_3 = 0 \\ 4.dy_1 - 3.dy_2 + 21.dx_1 + 6.dx_2 - 3.dx_3 = 0 \end{cases}$$

Puisque $\det \begin{pmatrix} -1 & 4 \\ 4 & -3 \end{pmatrix} = -13$, on peut exprimer les « variables » dy_1, dy_2 en fonction des « variables » dx_1, dx_2, dx_3 .

On peut en déduire que les variables endogènes y_1, y_2 du système non linéaire de départ sont définies **localement**, autour de $(3, -2, 1, -1, 2)$, en fonctions des variables x_1, x_2, x_3

(mais on ne sait pas comment, on n'a pas une écriture explicite de ce modèle).

En appliquant par exemple la méthode de Cramer, on trouve

$$\begin{cases} dy_1 = -\frac{75}{13}.dx_1 - \frac{12}{13}.dx_2 \\ dy_2 = -\frac{9}{13}.dx_1 + \frac{10}{13}.dx_2 - dx_3 \end{cases}$$

Exercice 10

Système d'équations aux variations différentielles en $(1, 0, 2, 1)$:

$$\begin{cases} -e^2.dx_1 - 4.dx_2 + (8 + e^2).dy_1 + (22 + 2.e^2).dy_2 = 0 \\ 2.dx_1 + dx_2 - 11.dy_1 + 2.dy_2 = 0 \end{cases}$$

Puisque $\det \begin{pmatrix} 8 + e^2 & 22 + 2.e^2 \\ -11 & 2 \end{pmatrix} = 258 + 24.e^2$, on peut exprimer les « variables » dy_1, dy_2 en fonction des « variables » dx_1, dx_2 .

On peut en déduire que les variables endogènes y_1, y_2 du système de départ sont définies **localement**, autour de $(x_1^*, x_2^*, y_1^*, y_2^*) = (1, 0, 2, 1)$, en fonctions des variables x_1, x_2

(mais on ne sait pas comment, on n'a pas une écriture explicite de ce modèle).

$$\text{On trouve } \begin{cases} dy_1 = \frac{44 + 6.e^2}{258 + 24.e^2}.dx_1 + \frac{30 + 2.e^2}{258 + 24.e^2}.dx_2 \\ dy_2 = \frac{-16 + 9.e^2}{258 + 24.e^2}.dx_1 + \frac{36 - e^2}{258 + 24.e^2}.dx_2 \end{cases}$$

Exercice 13

$$a) \begin{cases} a \cdot dy_1 + \frac{\partial f}{\partial y_3}(y_3^*) \cdot dy_3 - dx_1 = 0 \\ dy_2 - \frac{\partial g}{\partial y_3}(y_3^*, x_2^*) \cdot dy_3 - \frac{\partial g}{\partial x_2}(y_3^*, x_2^*) \cdot dx_2 = 0 \\ \frac{\partial h}{\partial y_1}(y_1^*) \cdot dy_1 - dy_2 = 0 \end{cases}$$

b) Il est possible d'exprimer les variables dy_1 , dy_2 et dy_3 en fonction de dx_1 et dx_2 si

$$\det \begin{pmatrix} a & 0 & \frac{\partial f}{\partial y_3}(y_3^*) \\ 0 & 1 & -\frac{\partial g}{\partial y_3}(y_3^*, x_2^*) \\ \frac{\partial h}{\partial y_1}(y_1^*) & -1 & 0 \end{pmatrix} = -\frac{\partial h}{\partial y_1}(y_1^*) \cdot \frac{\partial f}{\partial y_3}(y_3^*) - a \cdot \frac{\partial g}{\partial y_3}(y_3^*, x_2^*)$$

c'est-à-dire si $-\frac{\partial h}{\partial y_1}(y_1^*) \cdot \frac{\partial f}{\partial y_3}(y_3^*) - a \cdot \frac{\partial g}{\partial y_3}(y_3^*, x_2^*) \neq 0$

Exercice 14

a)
$$\begin{cases} 2a \, dy_1 + a \, dy_2 + 6 \, dx_1 + (-a) \, dx_2 + 4 \, dx_3 = 0 \\ 2 \, dy_1 + a \, dy_2 + 0 \, dx_1 + (-1) \, dx_2 + (-2) \, dx_3 = 0 \end{cases}$$

b) Il faut que $\det \begin{pmatrix} 2a & a \\ 2 & a \end{pmatrix} \neq 0$

soit $2a^2 - 2a \neq 0$

ce qui est le cas si $a \neq 0$ ou $a \neq 1$.

c)
$$\begin{cases} \boxed{4} \, dy_1 + \boxed{2} \, dy_2 = \boxed{-6 \, dx_1 + 2 \, dx_2 - 4 \, dx_3} \\ \boxed{2} \, dy_1 + \boxed{2} \, dy_2 = \boxed{dx_2 + 2 \, dx_3} \end{cases}$$

d)
$$\begin{cases} dy_1 = -3 \, dx_1 + \frac{1}{2} \, dx_2 - 3 \, dx_3 \\ dy_2 = 3 \, dx_1 + 0 \, dx_2 + 4 \, dx_3 \end{cases}$$

e) -3

Troisième partie
-
Optimisation

Chapitre 6 : Optimisation

1. Recherche d'extrema de fonctions d'une variable

Recherche d'un extremum local de la fonction $f: R \rightarrow R: x \rightarrow f(x)$

1. Calculer la dérivée première f'
2. Résoudre l'équation $f'(x) = 0$

Les solutions x^* appartenant au domaine de définition de f sont appelés les **points stationnaires** de f .

La tangente au graphe de f en x^* est horizontale (parallèle à l'axe des abscisses).

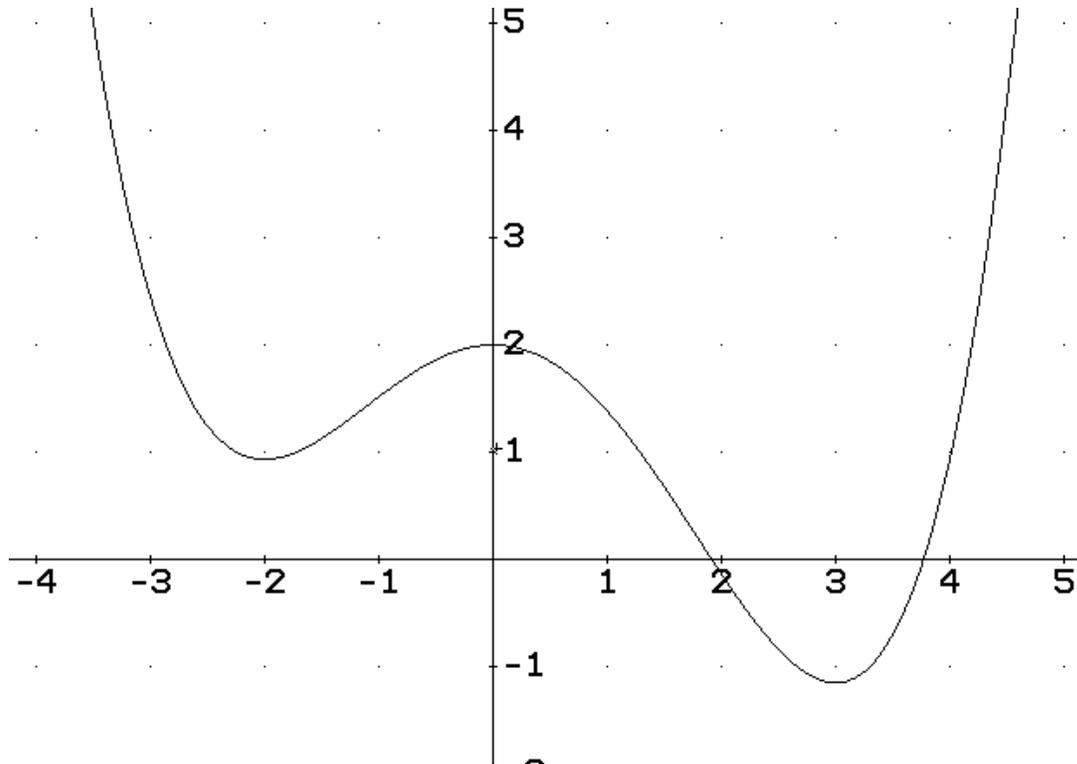
3. Calculer la dérivée seconde f''
4. En chaque point stationnaire, calculer la valeur de la dérivée seconde : $f''(x^*)$
 - Si $f''(x^*)$ est **positive** alors le point stationnaire x^* est un **minimum local**
 - Si $f''(x^*)$ est **négative** alors le point stationnaire x^* est un **maximum local**
 - Si $f''(x^*)$ est **nulle** et change de signe à gauche et à droite de x^* , alors le point stationnaire x^* est un **point d'inflexion**.

Exercice résolu

a) Rechercher les extrema éventuels de $f: R \rightarrow R: x \rightarrow f(x) = \frac{x^4}{20} - \frac{x^3}{15} - \frac{3x^2}{5} + 2$

b) En se basant sur le graphe de f , rechercher les extrema

1. sur le domaine de définition de f
2. sur l'intervalle R^+
3. sur l'intervalle R_0^+
4. sur l'intervalle $[-2,0]$
5. sur l'intervalle $] -2,0[$
6. sur l'intervalle $[-1,1]$
7. sur l'intervalle $] -1,1[$
8. sur l'intervalle $[1,2]$
9. sur l'intervalle $]1,2[$



Résolution :

$$\text{a) } f'(x) = \frac{1}{5} (x^3 - x^2 - 6x)$$

$$= \frac{1}{5} x (x^2 - x - 6)$$

$$f'(x) = 0 \Leftrightarrow x = 0 \quad \text{ou} \quad x = -2 \quad \text{ou} \quad x = 3$$

$x^* = 0$, $x^* = -2$ et $x^* = 3$ sont des **points stationnaires** (tangente horizontale)

Ces points stationnaires sont-ils des extrema locaux ?

$$f''(x) = \frac{1}{5} (3x^2 - 2x - 6)$$

- en $x^* = 0$

$f''(0) = -\frac{6}{5} < 0$: la concavité est tournée vers le bas ;

le point stationnaire $x^* = 0$ est un maximum local de f .

- en $x^* = -2$

$f''(-2) = 2 > 0$: la concavité est tournée vers le haut ;
le point stationnaire $x^* = -2$ est un minimum local de f .

- en $x^* = 3$

$f''(3) = 3 > 0$: la concavité est tournée vers le haut ;
le point stationnaire $x^* = 3$ est un minimum local de f .

b)

1. min local en $x^* = -2$, max local en $x^* = 0$ et min global en $x^* = 3$
2. max local en $x^* = 0$ et min global en $x^* = 3$
3. min global en $x^* = 3$
4. min global en $x^* = -2$, max global en $x^* = 0$
5. pas d'extremum
6. min local en $x^* = -1$ et min global en $x^* = 1$ ($f(-1) = 91/60$ et $f(1) = 83/60$) = , max global en $x^* = 0$
7. max global en $x^* = 0$
8. max global en $x^* = 1$, min global en $x^* = 2$
9. pas d'extremum

Exercices

Exercice 1

La fonction $f: R \rightarrow R: x \rightarrow f(x) = (x-1)^3 \cdot (x+1)$ possède-t-elle un extremum ?

- a) en $x^* = 0$?
- b) en $x^* = -\frac{1}{2}$?
- c) en $x^* = 1$?

Exercice 2

Rechercher les extrema éventuels de la fonction $f: R \rightarrow R: x \rightarrow f(x) = x^3 - 1$

Exercice 3

Rechercher les extrema éventuels de la fonction $f: R \rightarrow R: x \rightarrow f(x) = x^3 + x^2 - 1$

Exercice 4

Rechercher les extrema éventuels de la fonction $f: R_0^+ \rightarrow R: q \rightarrow f(q) = q^2 - 5q + 8$

Exercice 5

Rechercher les extrema éventuels de la fonction $g: R_0^+ \rightarrow R: x \rightarrow g(x) = 2x + \frac{8}{x}$

Exercice 6

Rechercher les extrema éventuels de la fonction

$$h: R \rightarrow R: x \rightarrow h(x) = \frac{1}{6}(x^3 - 6x^2 + 9x + 6)$$

Exercice 7

Rechercher les extrema éventuels de la fonction $i: R \rightarrow R: x \rightarrow i(x) = 5x^2 \cdot e^x$

Exercice 8

Rechercher les extrema éventuels de la fonction $j: R \rightarrow R: x \rightarrow j(x) = (x-1) \cdot e^x$

Exercice 9

Soient $C(q)$ le coût total d'une firme
 q la demande de son produit
 p le prix de vente de ce produit .

Si $q = 100 - p$ et $C(q) = \frac{1}{3} q^3 - 7 q^2 + 111 q + 50$,

quel est le niveau q qui maximise le profit ?

Exercice 10

Le revenu R que procure la vente d'un article est lié au prix de vente unitaire p de cet article par l'expression $R(p) = 1500 \cdot p \cdot e^{-\frac{p}{100}}$.

Déterminer (s'il en existe) une valeur de p qui maximise R .

Exercice 11

Un fabricant de lecteurs de c.d. produit q exemplaires par semaine à un coût total de $4 q^2 + 300 q + 10000$ € .

La demande hebdomadaire du produit est $q = 100 - 2 \cdot \sqrt{p}$, p étant le prix unitaire en euros.

Si le fabricant règle sa production sur la demande, quel niveau de production lui procurera un bénéfice maximum ?

Exercice 12

Une firme fabrique un nouvel alliage.

Si elle vend cet alliage p dollars la tonne , elle peut vendre $240 - p$ tonnes.

On sait aussi que le coût de production est $\frac{32000}{240 - 2 \cdot p} + 30$ dollars au total.

Combien de tonnes (q) doit-elle produire et à quel prix (p) pour maximiser son bénéfice ?

Exercice 13

La firme Channel fabrique des télévisions.

La demande journalière $q = 100 - 2 p$ si elle vend au prix unitaire p et le coût de production de q appareils est de $\frac{1}{4} q^2 + 35 q + 25$ par jour.

- Quel niveau de production maximise le bénéfice ?
- Pour quel niveau de production le coût de production unitaire est-il minimum ?

Exercice 14

L'entreprise S.E. (sans ennuis) fabrique q machine à laver mensuellement.

Elle utilise à cet effet plus ou moins de main d'œuvre (MO).

Sa fonction de production sur courte période est définie par $q = \frac{MO^2}{25} \cdot \left(3 - \frac{MO}{12}\right)$

où MO désigne le nombre de travailleurs.

Quel est l'effectif qui maximise la production ?

6.2. Recherche d'extrema de fonctions de n variables

6.2.1. Caractère défini positif, défini négatif d'une matrice carrée symétrique

Soit une matrice carrée symétrique A .

On appelle sous-matrice principale de A toute matrice obtenue en supprimant les rangées se correspondant dans les lignes et les colonnes de la matrice A .

Une sous-matrice particulière est la matrice obtenue en gardant les j premières lignes et les j premières colonnes de la matrice A .

Ces sous-matrices particulières sont symétriques et se notent $A^{(j)}$.

Exemple

Soit la matrice carrée symétrique $A = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{pmatrix}$

$A^{(1)} = (a_{1,1})$ matrice composée d'un seul élément

$$A^{(2)} = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix}$$

$$A^{(3)} = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{pmatrix} = A$$

Soit A une matrice carrée symétrique d'ordre n .

- A est définie positive ssi $\forall j \in \{1, 2, \dots, n\} : \det A^{(j)} > 0$

Par exemple, pour la matrice carrée symétrique A d'ordre 3 :

$$A \text{ est définie positive ssi } \left\{ \begin{array}{l} \det(a_{1,1}) > 0 \\ \det \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} > 0 \\ \det \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{pmatrix} > 0 \end{array} \right.$$

- A est définie négative ssi $\forall j \in \{1, 2, \dots, n\} : (-1)^j \cdot \det A^{(j)} > 0$
Autrement dit, la matrice A est définie négative ssi les $\det A^{(j)}$ sont successivement négatifs puis positifs,

en commençant par être négatif.

Par exemple, pour la matrice carrée symétrique A d'ordre 3 :

$$A \text{ est définie négative ssi } \begin{cases} \det(a_{1,1}) < 0 \\ \det \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} > 0 \\ \det \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{pmatrix} < 0 \end{cases}$$

Exercices

Exercice 15

Les matrices symétriques suivantes sont-elles définies positives, définies négatives ?

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

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

3. $\begin{pmatrix} -6 & 0 & 6 \\ 0 & -2 & 0 \\ 6 & 0 & -12 \end{pmatrix}$

4. $\begin{pmatrix} -3 & 2 \\ 2 & -1 \end{pmatrix}$

5. $\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 2 & 0 & a \end{pmatrix}$ où a est un paramètre réel

6. $\begin{pmatrix} 2 & -2 & 2 \\ -2 & 5 & 1 \\ 2 & 1 & 5 \end{pmatrix}$

6.2.2. Recherche des extrema libres d'une fonction de n variables

Soit une fonction de n variables $F : \mathbb{R}^* \rightarrow \mathbb{R} : (x_1, x_2, \dots, x_n) \rightarrow F(x_1, x_2, \dots, x_n)$

- Calculer toutes les dérivées partielles d'ordre 1 :

$$\frac{\partial F}{\partial x_1}(x_1, x_2, \dots, x_n)$$

$$\frac{\partial F}{\partial x_2}(x_1, x_2, \dots, x_n)$$

$$\dots$$

$$\frac{\partial F}{\partial x_n}(x_1, x_2, \dots, x_n)$$

- Résoudre le système

$$\begin{cases} \frac{\partial F}{\partial x_1}(x_1, x_2, \dots, x_n) = 0 \\ \frac{\partial F}{\partial x_2}(x_1, x_2, \dots, x_n) = 0 \\ \dots \\ \frac{\partial F}{\partial x_n}(x_1, x_2, \dots, x_n) = 0 \end{cases}$$

Les solutions $(x_1^*, x_2^*, \dots, x_n^*)$ de ce système sont les **points stationnaires**.

- Calculer toutes les dérivées partielles d'ordre 2.

$$HF(x_1, x_2, \dots, x_n) = \begin{pmatrix} \frac{\partial^2 F}{(\partial x_1)^2}(x_1, x_2, \dots, x_n) & \frac{\partial^2 F}{\partial x_1 \partial x_2}(x_1, x_2, \dots, x_n) & \dots & \frac{\partial^2 F}{\partial x_1 \partial x_n}(x_1, x_2, \dots, x_n) \\ \frac{\partial^2 F}{\partial x_2 \partial x_1}(x_1, x_2, \dots, x_n) & \frac{\partial^2 F}{(\partial x_2)^2}(x_1, x_2, \dots, x_n) & \dots & \frac{\partial^2 F}{\partial x_2 \partial x_n}(x_1, x_2, \dots, x_n) \\ \dots & \dots & \dots & \dots \\ \frac{\partial^2 F}{\partial x_n \partial x_1}(x_1, x_2, \dots, x_n) & \frac{\partial^2 F}{\partial x_n \partial x_2}(x_1, x_2, \dots, x_n) & \dots & \frac{\partial^2 F}{(\partial x_n)^2}(x_1, x_2, \dots, x_n) \end{pmatrix}$$

- Pour chaque point stationnaire $(x_1^*, x_2^*, \dots, x_n^*)$, écrire la matrice hessienne $HF(x_1^*, x_2^*, \dots, x_n^*)$.

- **Si $HF(x_1^*, x_2^*, \dots, x_n^*)$ est définie positive**
le point stationnaire $(x_1^*, x_2^*, \dots, x_n^*)$ est un **minimum local**.
- **Si $HF(x_1^*, x_2^*, \dots, x_n^*)$ est définie négative**
le point stationnaire $(x_1^*, x_2^*, \dots, x_n^*)$ est un **maximum local**.
- **Sinon, si $\det HF(x_1^*, x_2^*, \dots, x_n^*) \neq 0$**
le point stationnaire $(x_1^*, x_2^*, \dots, x_n^*)$ est un **point de selle**.

Exercice résolu

Optimiser la fonction f définie par $f(x_1, x_2) = 5x_1^2 + 6x_2^2 - x_1x_2$.

Résolution :

- Calculer toutes les dérivées partielles d'ordre 1.

$$\frac{\partial f}{\partial x_1}(x_1, x_2) = 10x_1 - x_2$$

$$\frac{\partial f}{\partial x_2}(x_1, x_2) = 12x_2 - x_1$$

- Résoudre le système

$$\frac{\partial f}{\partial x_1}(x_1, x_2) = 0 \quad 10x_1 - x_2 = 0 \quad x_2 = 10x_1 \quad x_2 = 10x_1 \quad x_2 = 0$$

$$\frac{\partial f}{\partial x_2}(x_1, x_2) = 0 \quad 12x_2 - x_1 = 0 \quad 12.(10x_1) - x_1 = 0 \quad 119.x_1 = 0 \quad x_1 = 0$$

La solution $(0,0)$ de ce système est le seul **point stationnaire** de la fonction.

- Calculer toutes les dérivées partielles d'ordre 2.

$$HF(x_1, x_2) = \begin{pmatrix} \frac{\partial^2 f}{(\partial x_1)^2}(x_1, x_2) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x_1, x_2) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x_1, x_2) & \frac{\partial^2 f}{(\partial x_2)^2}(x_1, x_2) \end{pmatrix}$$

$$HF(x_1, x_2) = \begin{pmatrix} 10 & -1 \\ -1 & 12 \end{pmatrix}$$

- $\det(10) = 10 > 0$

$$\det \begin{pmatrix} 10 & -1 \\ -1 & 12 \end{pmatrix} = 119 > 0$$

$HF(x_1, x_2)$ est **définie positive** :

le point stationnaire $(0,0)$ est donc un **minimum local** et $F(0,0) = 0$.

Exercices

Trouver les extrema et points de selle éventuels des fonctions suivantes

Exercice 16 : $f: \mathbb{R}^2 \rightarrow \mathbb{R} : (x_1, x_2) \rightarrow f(x_1, x_2) = x_1^2 + x_2^2$

Exercice 17 : $f: \mathbb{R}^2 \rightarrow \mathbb{R} : (x_1, x_2) \rightarrow f(x_1, x_2) = x_1^2 - x_2^2$

Exercice 18 : $f: \mathbb{R}^2 \rightarrow \mathbb{R} : (x_1, x_2) \rightarrow f(x_1, x_2) = x_1^3 - x_2^2 + 3x_1^2 + 3x_2^2 - 9x_1$

Exercice 19 : $f: \mathbb{R}^2 \rightarrow \mathbb{R} : (x_1, x_2) \rightarrow f(x_1, x_2) = x_1^2 + x_1 \cdot x_2 + x_2^2 + x_1 - 4x_2 + 4$

Exercice 20 : $f: \mathbb{R}^2 \rightarrow \mathbb{R} : (x_1, x_2) \rightarrow f(x_1, x_2) = 12x_1^2 - 24x_1 \cdot x_2 + x_2^3 + 15x_2^2 - 24x_2$

Exercice 21 : $f: \mathbb{R}^2 \rightarrow \mathbb{R} : (x_1, x_2) \rightarrow f(x_1, x_2) = 3x_1^2 + 2x_2^3 + 21x_2^2 - 12x_1x_2 - 24x_2$

Problème résolu

Une firme produit deux types de biens A et B . Son coût total de fabrication, noté C , et la demande respective des deux biens q_A et q_B sont donnés par

$$C = q_A^2 + 2q_B^2 + 10$$

$$q_A = 40 - 2p_A - p_B$$

$$q_B = 35 - p_A - p_B$$

Quels sont les niveaux de production qui maximisent le profit ?

Quels sont les prix de A et B qui suscitent une demande correspondant à ces niveaux optimaux ?

Résolution :

La fonction profit est donnée par $\Pi(q_A, q_B) = q_A \cdot p_A + q_B \cdot p_B - C(q_A, q_B)$

où $\Pi(q_A, q_B)$ ne dépend que des variables q_A et q_B

$$\text{de } \begin{cases} q_A = 40 - 2p_A - p_B \\ q_B = 35 - p_A - p_B \end{cases} \text{ on peut déduire } \begin{cases} p_A = 5 - q_A + q_B \\ p_B = 30 + q_A - 2q_B \end{cases}$$

$$\text{et donc que } \Pi(q_A, q_B) = 5q_A - 2q_A^2 + 2q_Aq_B - 4q_B^2 + 30q_B - 10$$

Maximiser le profit en fonction niveau de production :

1°) Recherche des points stationnaires de Π .

Calculer toutes les dérivées partielles d'ordre 1 de Π (dérivées premières), et résoudre

$$\text{le système } \begin{cases} \frac{\partial \Pi}{\partial q_A}(q_A, q_B) = 0 \\ \frac{\partial \Pi}{\partial q_B}(q_A, q_B) = 0 \end{cases}.$$

Les solutions (q_A^*, q_B^*) de ce système sont les **points stationnaires**.

$$\begin{cases} 5 - 4q_A + 2q_B = 0 \\ 2q_A - 8q_B + 30 = 0 \end{cases}$$

Ce système est un système linéaire.

Il peut être résolu simplement par substitution.

$$\begin{cases} q_A = \frac{5}{4} + \frac{q_B}{2} \\ 2 \cdot \left(\frac{5}{4} + \frac{q_B}{2}\right) - 8q_B + 30 = 0 \end{cases} \quad \begin{cases} q_A = \frac{5}{4} + \frac{q_B}{2} \\ q_B = \frac{65}{14} \end{cases} \quad \begin{cases} q_A = \frac{50}{14} \\ q_B = \frac{65}{14} \end{cases}$$

$\left(\frac{50}{14}, \frac{65}{14}\right)$ est donc un point stationnaire.

2°) Test des points stationnaires.

Le point stationnaire $\left(\frac{50}{14}, \frac{65}{14}\right)$ est-il un maximum local ?

- Calculer toutes les dérivées partielles d'ordre 2 (dérivées secondes).

$$H\Pi(q_A, q_B) = \begin{pmatrix} \frac{\partial^2 \Pi}{(\partial q_A)^2}(q_A, q_B) & \frac{\partial^2 \Pi}{\partial q_A \partial q_B}(q_A, q_B) \\ \frac{\partial^2 \Pi}{\partial q_B \partial q_A}(q_A, q_B) & \frac{\partial^2 \Pi}{(\partial q_B)^2}(q_A, q_B) \end{pmatrix}$$

$$H\Pi(q_A, q_B) = \begin{pmatrix} -4 & 2 \\ 2 & -8 \end{pmatrix}$$

- Pour le point stationnaire $\left(\frac{50}{14}, \frac{65}{14}\right)$: $H\Pi\left(\frac{50}{14}, \frac{65}{14}\right) = \begin{pmatrix} -4 & 2 \\ 2 & -8 \end{pmatrix}$
 $\det(-4) < 0$
 $\det\begin{pmatrix} -4 & 2 \\ 2 & -8 \end{pmatrix} > 0$
 $\rightarrow H\Pi\left(\frac{50}{14}, \frac{65}{14}\right)$ est **définie négative**.

Le point stationnaire $\left(\frac{50}{14}, \frac{65}{14}\right)$ est un **maximum local**.

Pour information : $H\Pi(q_A, q_B) = \begin{pmatrix} -4 & 2 \\ 2 & -8 \end{pmatrix}$ ne dépend pas des variables q_A et q_B .

$H\Pi(q_A, q_B) = \begin{pmatrix} -4 & 2 \\ 2 & -8 \end{pmatrix}$ est **semi-définie négative** en tout point du domaine.

Le point stationnaire $(\frac{50}{14}, \frac{65}{14})$ est un **maximum global**.

Réponse :

$q_A = \frac{50}{14}$ et $q_B = \frac{65}{14}$ maximisent le profit en fonction des niveaux de production.

Les prix associés p_A et p_B sont donnés par $p_A = 5 - q_A + q_B$ et $p_B = 30 + q_A - 2q_B$.

Les prix associés à ce niveau de production sont $p_A = \frac{85}{14}$ et $p_B = \frac{340}{14}$.

Exercice

Exercice 22

Une entreprise fabrique deux produits A et B dans les quantités q_A et q_B (en milliers). Elle les vend respectivement 12 € et 18 € l'unité. Le coût de fabrication, noté C , est donné par $2 \cdot q_A^2 + q_A \cdot q_B + 2 \cdot q_B^2$.

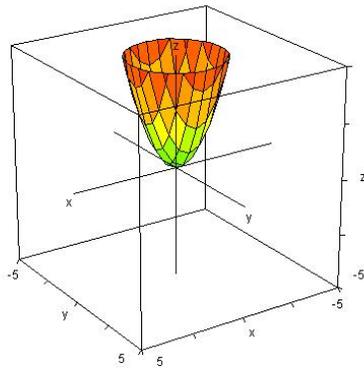
Quel est le niveau de production qui maximise le profit de cette entreprise ?

6.2.3. A propos des courbes de niveau d'une fonction de deux variables

Soit la fonction suivante :

$$f(x,y) = x^2 + y^2$$

La représentation graphique de cette fonction est un parabolôide de révolution (d'équation $z = x^2 + y^2$).



Comment représenter cet objet dans un espace à deux dimensions ?

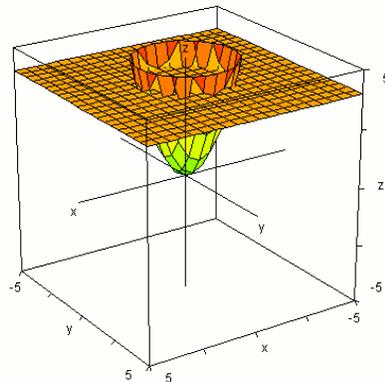
Si l'on coupe le parabolôide par un plan horizontal d'équation $z = k$ (où k est un nombre réel positif), on obtient ce qu'on appelle une courbe de niveau de la fonction.

Equation dans l'espace de la courbe de niveau à hauteur k :

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

Exemple 1

Voici une représentation à trois dimensions d'un parabolôide de révolution ($z = x^2 + y^2$) coupé par un plan horizontal à hauteur 4 :



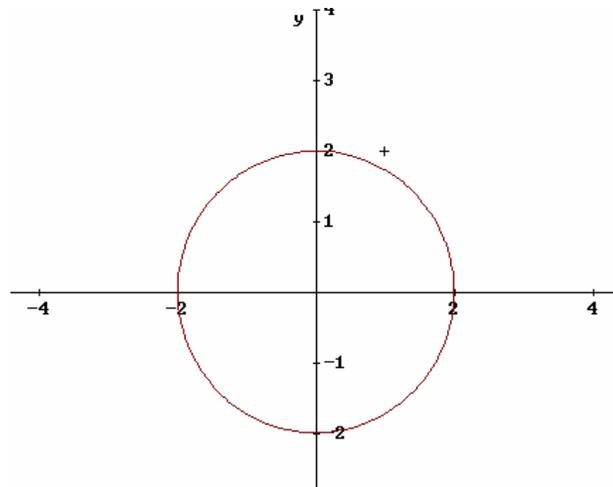
La courbe qui est l'intersection du paraboloidé et du plan a pour équation (dans l'espace) :

$$\begin{cases} z = x^2 + y^2 \\ z = 4 \end{cases}$$

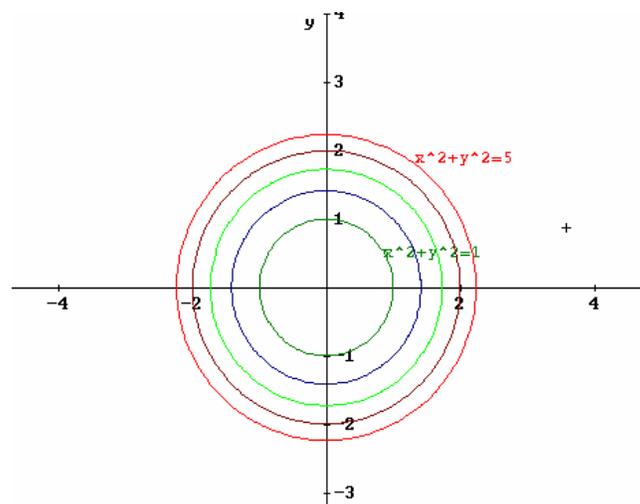
La projection de cette courbe *dans le plan* horizontal oxy est une courbe d'équation

$$x^2 + y^2 = 4 .$$

C'est un cercle, de centre (0,0) et de rayon 2 ($=\sqrt{4}$).



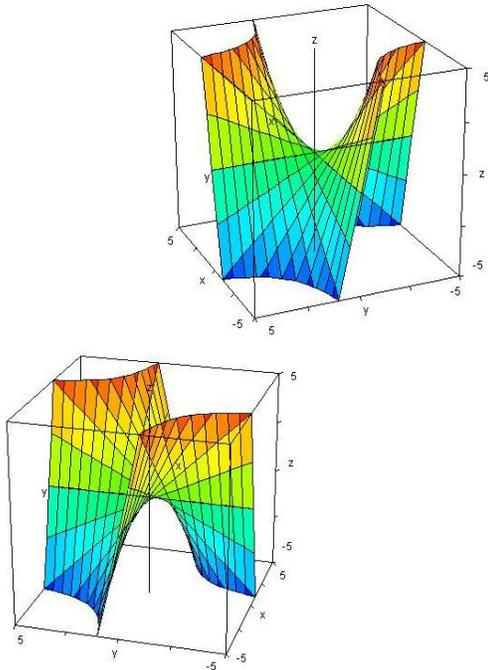
De même, si l'on coupe le paraboloidé à hauteur 0, 1, 2, 3, 4, 5, on trouve 6 courbes de niveaux du paraboloidé (la première étant réduite au point (0,0)) :



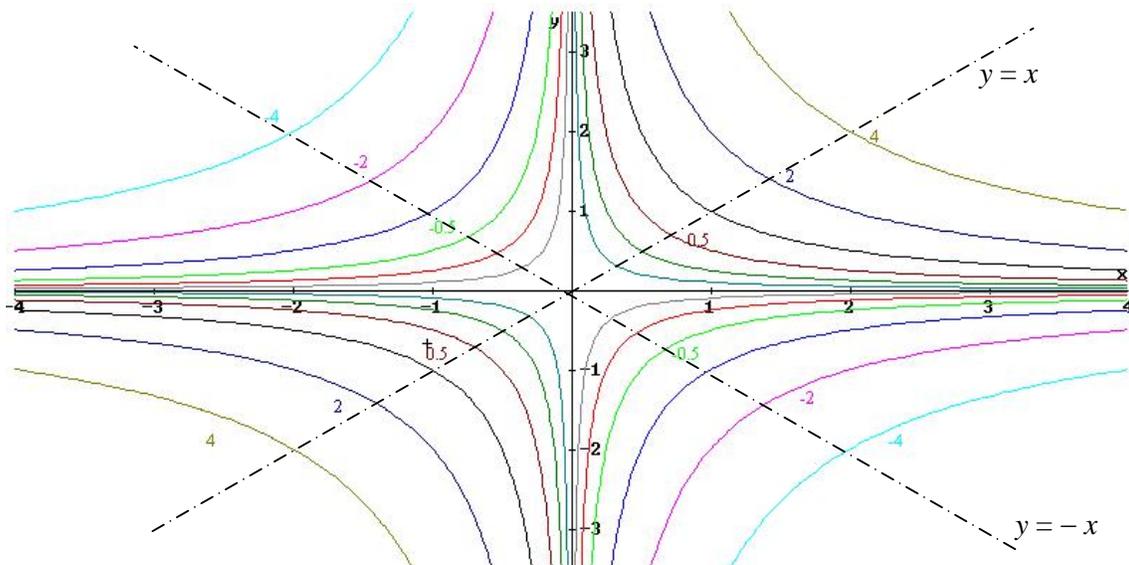
Même si on n'a pas la possibilité d'avoir une représentation graphique en 3D de la surface, on peut quand même voir, grâce aux courbes de niveau, qu'elle admet un minimum (global) en (0,0).

Exemple 2

Voici à présent une représentation à trois dimensions de ce qu'on appelle une « selle de cheval » ($z = x.y$), vue sous deux angles différents.



Ci-dessous, on trouve les courbes de niveau de cette surface (aux hauteurs $-4, -2, -1, -0.5, -0.25, -0.1, 0.1, 0.25, 0.5, 1, 2, 4$).



Le point de coordonnées $(0,0)$ est un point très particulier. En effet, si l'on suit un « chemin » imaginaire situé sur la surface à la verticale de la droite d'équation $y = x$, on constate que $(0,0)$ est un minimum (« point le plus bas ») pour ce chemin ; d'autre part, en suivant un « chemin » à la verticale de la droite d'équation $y = -x$, le point $(0,0)$ est un maximum.

Un tel point est appelé point de selle (ou encore point col, en référence aux cols des montagnes, où un tel phénomène se produit).

6.3. Recherche d'extrema de fonctions de n variables sous contrainte d'égalité – Méthode de Lagrange

Méthode :

Pour optimiser la **fonction F de 2 variables** x_1, x_2 définie par $y = F(x_1, x_2)$ soumise à la **contrainte unique** exprimée par l'égalité $g(x_1, x_2) - C = 0$.

1. Ecrire la fonction L appelée **fonction Lagrangienne**.

Cette fonction L dépend de $n+1$ variables x_1, x_2 et de la variable λ , et est définie comme suit :

$$L(\lambda, x_1, x_2) = F(x_1, x_2) - \lambda.(g(x_1, x_2) - C)$$

2. Calculer toutes les dérivées partielles d'ordre 1 de la fonction L .

3. Résoudre le système
$$\begin{cases} \frac{\partial L}{\partial \lambda}(\lambda, x_1, x_2) = 0 \\ \frac{\partial L}{\partial x_1}(\lambda, x_1, x_2) = 0 \\ \frac{\partial L}{\partial x_2}(\lambda, x_1, x_2) = 0 \end{cases}$$

Les solutions $(\lambda^*, x_1^*, x_2^*)$ de ce système sont les **points stationnaires** de L .

4. Calculer toutes les dérivées partielles d'ordre 2 (dérivées secondes).

$$HL(\lambda, x_1, x_2) = \begin{pmatrix} \frac{\partial^2 L}{(\partial \lambda)^2}(\lambda, x_1, x_2) & \frac{\partial^2 L}{\partial x_1 \cdot \partial \lambda}(\lambda, x_1, x_2) & \frac{\partial^2 L}{\partial x_2 \cdot \partial \lambda}(\lambda, x_1, x_2) \\ \frac{\partial^2 L}{\partial \lambda \cdot \partial x_1}(\lambda, x_1, x_2) & \frac{\partial^2 L}{(\partial x_1)^2}(\lambda, x_1, x_2) & \frac{\partial^2 L}{\partial x_2 \cdot \partial x_1}(\lambda, x_1, x_2) \\ \frac{\partial^2 L}{\partial \lambda \cdot \partial x_2}(\lambda, x_1, x_2) & \frac{\partial^2 L}{\partial x_1 \cdot \partial x_2}(\lambda, x_1, x_2) & \frac{\partial^2 L}{(\partial x_2)^2}(\lambda, x_1, x_2) \end{pmatrix}$$

5. Ecrire la matrice Hessienne $HL(\lambda^*, x_1^*, x_2^*)$ **pour chaque point stationnaire**.

- Si $\det(HL(\lambda, x_1^*, x_2^*)) > 0$
alors le point stationnaire $(\lambda^*, x_1^*, x_2^*)$ est un **maximum local sous contrainte**.
- Si $\det(HL(\lambda, x_1^*, x_2^*)) < 0$
alors le point stationnaire $(\lambda^*, x_1^*, x_2^*)$ est un **minimum local sous contrainte**.

Exercice résolu

Optimiser la fonction F définie par $F(x_1, x_2) = 5x_1^2 + 6x_2^2 - x_1 \cdot x_2$ soumise à la contrainte exprimée par l'égalité $x_1 + 2x_2 = 24$.

Résolution :

- La contrainte s'écrit $x_1 + 2x_2 - 24 = 0$
- Ecrire la fonction L appelée fonction Lagrangienne
 $L(\lambda, x_1, x_2) = F(x_1, x_2) - \lambda \cdot (g(x_1, x_2) - C)$
 $L(\lambda, x_1, x_2) = 5x_1^2 + 6x_2^2 - x_1 \cdot x_2 - \lambda \cdot (x_1 + 2x_2 - 24)$
- Calculer toutes les dérivées partielles d'ordre 1 de la fonction L ,

et résoudre le système
$$\begin{cases} \frac{\partial L}{\partial \lambda}(\lambda, x_1, x_2) = 0 \\ \frac{\partial L}{\partial x_1}(\lambda, x_1, x_2) = 0 \\ \frac{\partial L}{\partial x_2}(\lambda, x_1, x_2) = 0 \end{cases}$$

$$\begin{cases} -(x_1 + 2x_2 - 24) = 0 \\ 10x_1 - x_2 - \lambda = 0 \\ 12x_2 - x_1 - 2\lambda = 0 \end{cases} \Leftrightarrow \begin{cases} x_1 + 2x_2 - 24 = 0 \\ 10x_1 - x_2 = \lambda \\ 6x_2 - \frac{x_1}{2} = \lambda \end{cases} \Leftrightarrow \begin{cases} x_1 + 2x_2 - 24 = 0 \\ 10x_1 - x_2 = 6x_2 - \frac{x_1}{2} \\ 6x_2 - \frac{x_1}{2} = \lambda \end{cases}$$

$$\Leftrightarrow \begin{cases} x_1 + 2x_2 - 24 = 0 \\ \frac{21}{2}x_1 - 7x_2 = 0 \\ 6x_2 - \frac{x_1}{2} = \lambda \end{cases} \Leftrightarrow \begin{cases} x_1 + 2x_2 - 24 = 0 \\ x_2 = \frac{3}{2}x_1 \\ 6x_2 - \frac{x_1}{2} = \lambda \end{cases} \Leftrightarrow \begin{cases} x_1 + 3x_1 - 24 = 0 \\ x_2 = \frac{3}{2}x_1 \\ \lambda = 9x_1 - \frac{x_1}{2} \end{cases}$$

$$\Leftrightarrow \begin{cases} x_1 = 6 \\ x_2 = 9 \\ \lambda = 51 \end{cases} \Leftrightarrow \begin{cases} \lambda = 51 \\ x_1 = 6 \\ x_2 = 9 \end{cases}$$

La solution (51,6,9) de ce système est un point stationnaire de L ; (6,9) est un point stationnaire de F .

- Calculer la Hessienne bordée

$$HL(\lambda, x_1, x_2) = \begin{pmatrix} \frac{\partial^2 L}{(\partial \lambda)^2}(\lambda, x_1, x_2) & \frac{\partial^2 L}{\partial x_1 \cdot \partial \lambda}(\lambda, x_1, x_2) & \frac{\partial^2 L}{\partial x_2 \cdot \partial \lambda}(\lambda, x_1, x_2) \\ \frac{\partial^2 L}{\partial \lambda \cdot \partial x_1}(\lambda, x_1, x_2) & \frac{\partial^2 L}{(\partial x_1)^2}(\lambda, x_1, x_2) & \frac{\partial^2 L}{\partial x_2 \cdot \partial x_1}(\lambda, x_1, x_2) \\ \frac{\partial^2 L}{\partial \lambda \cdot \partial x_2}(\lambda, x_1, x_2) & \frac{\partial^2 L}{\partial x_1 \cdot \partial x_2}(\lambda, x_1, x_2) & \frac{\partial^2 L}{(\partial x_2)^2}(\lambda, x_1, x_2) \end{pmatrix}$$

$$HL(\lambda, x_1, x_2) = \begin{pmatrix} 0 & -1 & -2 \\ -1 & 10 & -1 \\ -2 & -1 & 12 \end{pmatrix}$$

- au point stationnaire $HL(51, 6, 9) = \begin{pmatrix} 0 & -1 & -2 \\ -1 & 10 & -1 \\ -2 & -1 & 12 \end{pmatrix}$:

$$\text{dét} \begin{pmatrix} 0 & -1 & -2 \\ -1 & 10 & -1 \\ -2 & -1 & 12 \end{pmatrix} = -56 < 0$$

Le point (6,9) est un minimum local de F sous la contrainte $x_1 + 2 \cdot x_2 = 24$ (avec $\lambda = 51$).

Exercices

Exercice 23

Optimiser la fonction F définie par $F(x_1, x_2) = x_1 \cdot x_2$ soumise à la contrainte exprimée par l'égalité $x_1 + x_2 = 6$.

Exercice 24

Optimiser la fonction F définie par $F(x_1, x_2) = x_1^2 + x_2^2$ soumise à la contrainte exprimée par l'égalité $x_1 + 4 \cdot x_2 - 2 = 0$.

Exercice 25

Optimiser la fonction F définie par $F(x_1, x_2) = 12 \cdot x_1 \cdot x_2 - 3 \cdot x_2^2 - x_1^2$ soumise à la contrainte exprimée par l'égalité $x_1 + x_2 = 16$.

Exercice 26

Optimiser la fonction F définie par $F(x_1, x_2) = 6x_1 + 8x_2 - 8$ soumise à la contrainte exprimée par l'égalité $(x_1 - 4)^2 + (x_2 - 3)^2 = 25$.

6.4. Optimisation sous contraintes, vision graphique

Exercice résolu

Soit la fonction $F(x, y) = (x - 1)^2 + (y - 1)^2$.

a) Repérez graphiquement les extrema de F .

b) Repérez graphiquement les extrema de F sous les contraintes

$$\begin{cases} 0 \leq x \leq 3 \\ 0 \leq y \leq 4 \end{cases}$$

c) Repérez graphiquement les extrema de F sous les contraintes

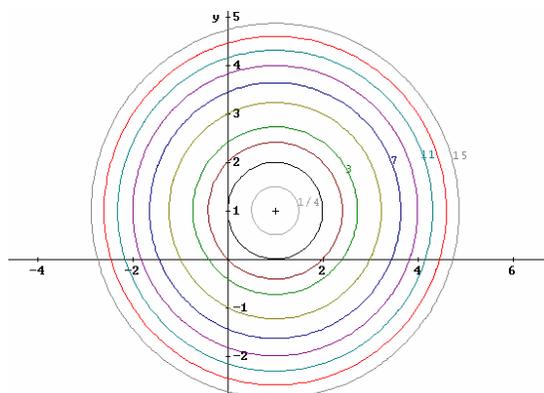
$$\begin{cases} x \geq 0, y \geq 0 \\ x + 2y \leq 4 \end{cases}$$

d) Repérez graphiquement les extrema de F sous la contrainte

$$x - y = 1.$$

Résolution :

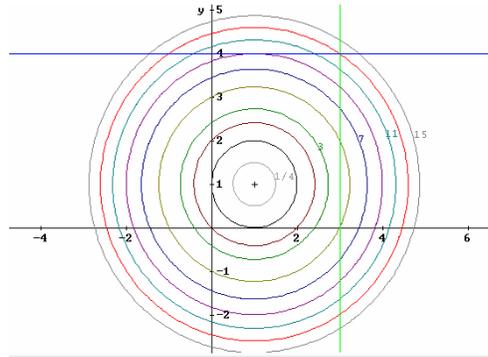
a) Repérez graphiquement les extrema de F .



F admet un minimum global en $(1,1)$.

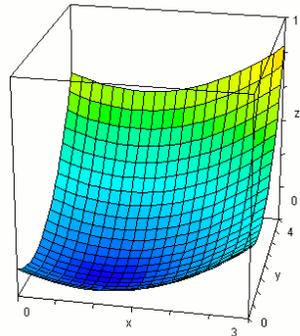
b) Repérez graphiquement les extrema de F sous les contraintes

$$\begin{cases} 0 \leq x \leq 3 \\ 0 \leq y \leq 4 \end{cases}$$



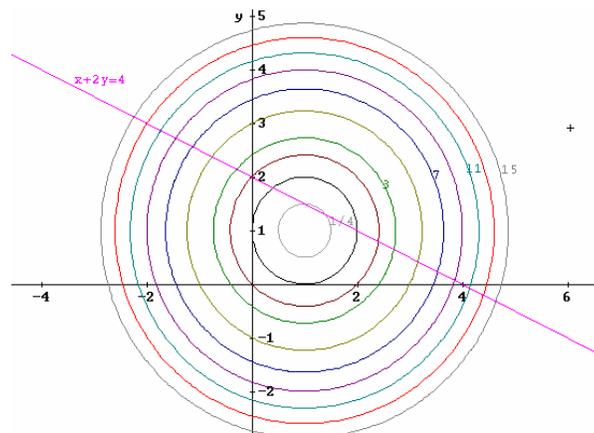
Sous les contraintes que x soit compris entre 0 et 3 et que y soit compris entre 0 et 4, F admet un minimum en $(1,1)$, et 4 extrema locaux en $(0,0)$, $(3,0)$, $(3,4)$ et $(0,4)$; parmi eux, le plus haut est $(3,4)$ (à hauteur $13 = (3-1)^2 + (4-1)^2$).

Visualisation en 3D du morceau de surface vérifiant les contraintes $0 \leq x \leq 3$ et $0 \leq y \leq 4$:



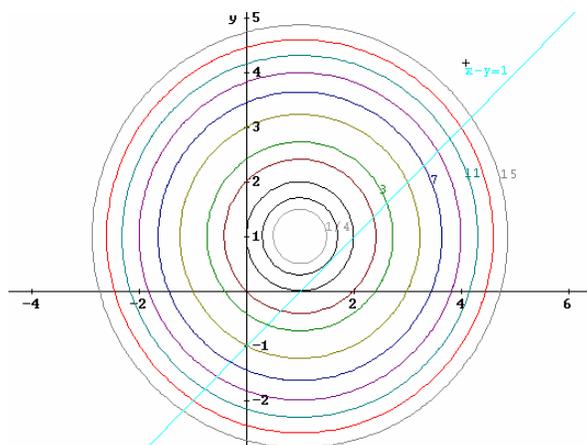
c) Repérez graphiquement les extrema de F sous les contraintes

$$\begin{cases} x \geq 0, y \geq 0 \\ x + 2y \leq 4 \end{cases}$$



Sous les contraintes ci-dessus, F admet un minimum global en $(1,1)$ et trois maxima locaux en $(0,0)$, $(4,0)$ et $(0,2)$; le plus haut de ces trois maxima est en $(4,0)$ (à hauteur $10 = (4-1)^2 + (0-1)^2$).

d) Repérez graphiquement les extrema de F sous la contrainte $x - y = 1$.



Sous la contrainte que $x - y = 1$, F admet un minimum en $(1.5 ; 0.5)$ (l'image de ce point est à hauteur $0.5 = F(1.5 ; 0.5)$).

Exercice

Exercice 27

Soit F la fonction de deux variables x_1 et x_2 définie par

$$F(x_1, x_2) = 2x_1^3 + 2x_2^3 - 9x_1^2 - 18x_2^2 + 12x_1 + 48x_2$$

(vous trouverez plus loin la représentation graphique de quelques courbes de niveau de cette fonction F).

a) Déterminez analytiquement les extrema locaux et points de selle de F .

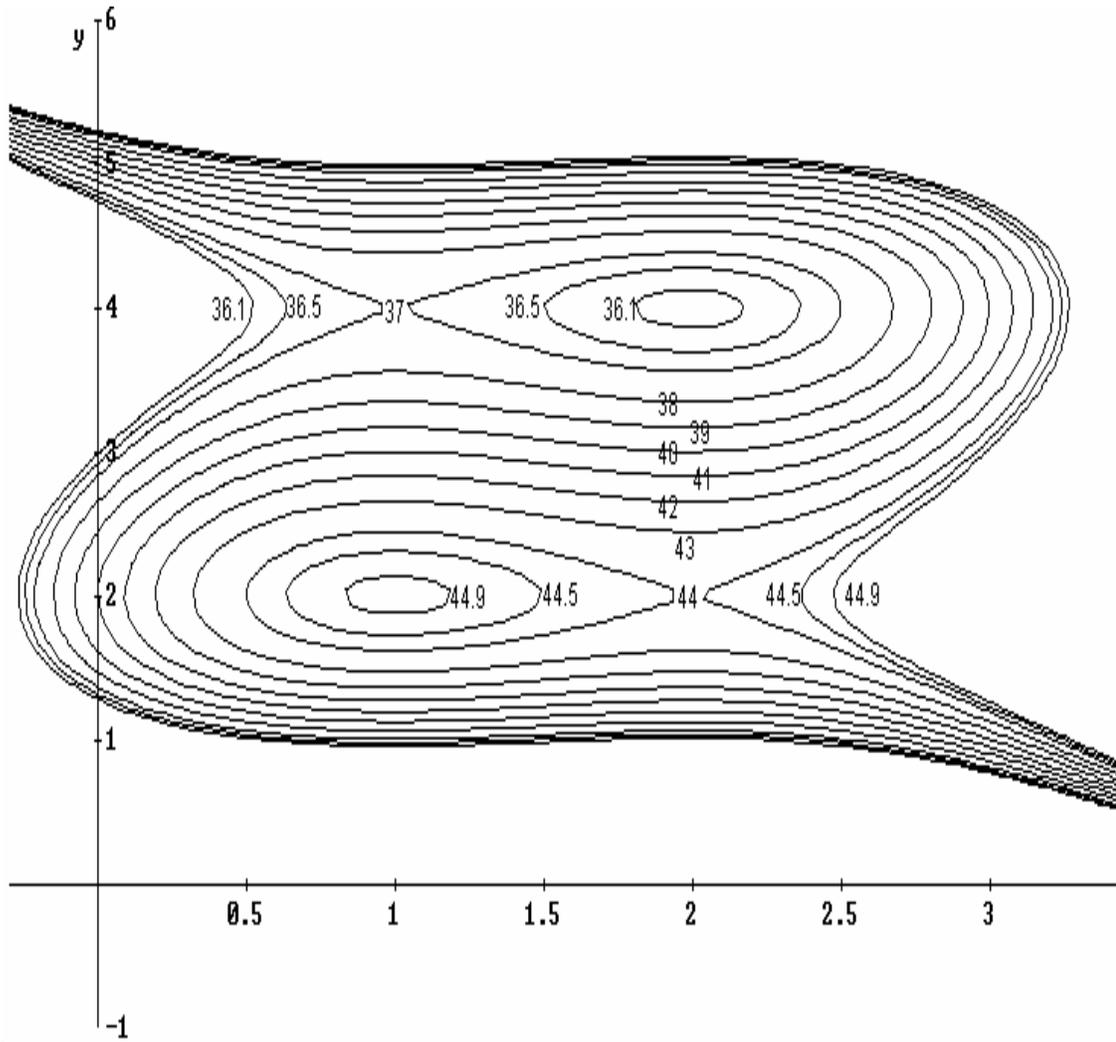
b) Déterminez (en utilisant une technique analytique ou graphique) les extrema locaux éventuels de F sous la contrainte $2x_1 - x_2 = 0$.

c) Déterminez (en utilisant une technique graphique) les extrema locaux éventuels de F sous les contraintes

$$\begin{cases} 0.5 \leq x_1 \leq 2 \\ 1.5 \leq x_2 \leq 4 \end{cases}$$

Courbes de niveau (à différentes valeurs) de la fonction F définie par

$$F(x_1, x_2) = 2x_1^3 + 2x_2^3 - 9x_1^2 - 18x_2^2 + 12x_1 + 48x_2$$

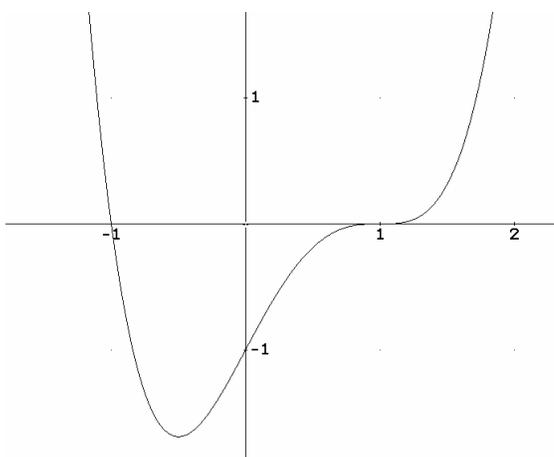


Solution des exercices (Chapitre 6)

Exercice 1

- a) $x^* = 0$ n'est pas un extremum
- b) $x^* = -\frac{1}{2}$ est un minimum local
- c) $x^* = 1$ n'est pas un extremum mais est un **point d'inflexion**

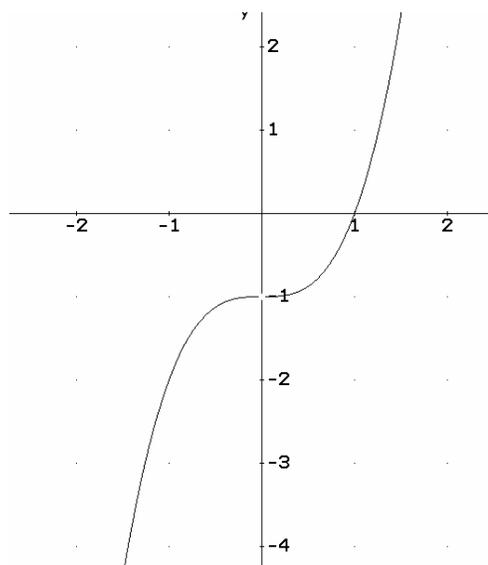
Graph de f :



Exercice 2

Le point stationnaire $x^* = 0$ est un point d'inflexion (à tangente horizontale).

Graph de f :

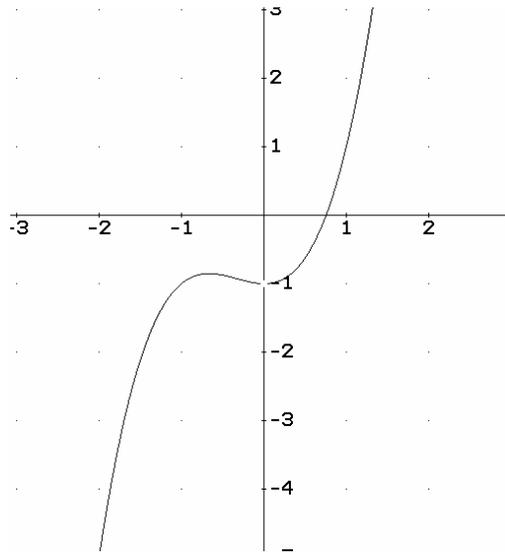


Exercice 3

Le point stationnaire $x^* = 0$ est un minimum.

Le point stationnaire $x^* = -\frac{2}{3}$ est un maximum.

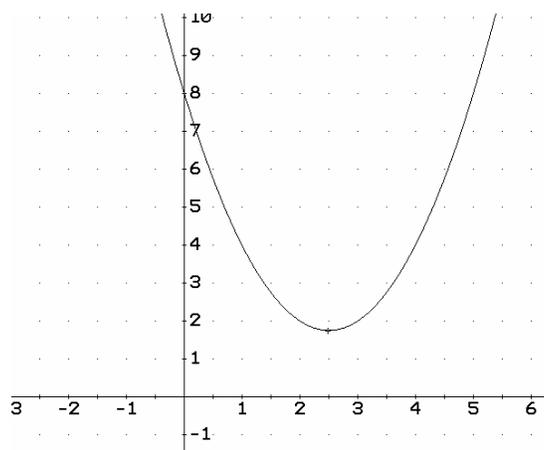
Graphe de f :



Exercice 4

Le point stationnaire $q^* = \frac{5}{2}$ est un minimum local

Graphe de f :

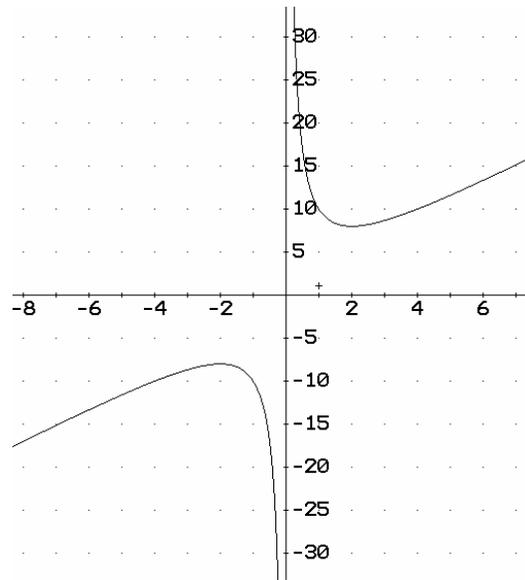


Exercice 5

Attention : $-2 \notin \mathbb{R}_0^+$; $x^* = 2$ est donc le seul point stationnaire du domaine de définition.

Le point stationnaire $x^* = 2$ est un minimum local

Graphe de f :



Exercice 6

$x^* = 1$ est un maximum local de f , $x^* = 3$ est un minimum local de f .

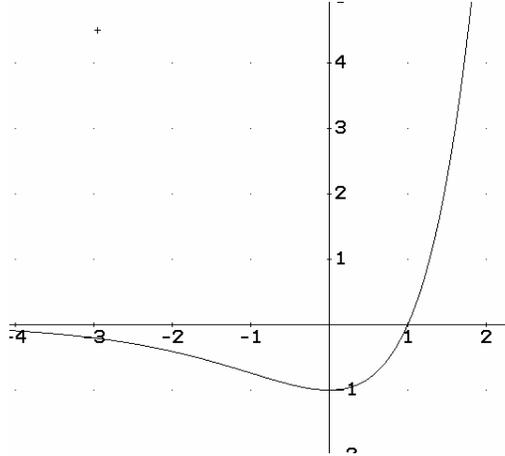
Exercice 7

$x^* = 0$ est un minimum local de f , $x^* = -2$ est un maximum local de f .

Exercice 8

Le point stationnaire $x^* = 0$ est un minimum local de f .

Graphe de f :



Exercice 9

Le profit est donné par $q \cdot p - C(q)$

La fonction profit est donnée par $P(q) = -\frac{1}{3} \cdot q^3 + 6 \cdot q^2 - 11 \cdot q - 50$

$q^* = 11$ est un maximum de P , et $q^* = 1$ est un minimum de P .

Une demande de 11 unités permet donc de maximiser le profit.

Exercice 10

$p^* = 100$ est un maximum pour R .

Exercice 11

La fonction bénéfice est donnée par $B(q) = \frac{1}{4} \cdot q^3 - 54 \cdot q^2 + 2200 \cdot q - 10000$

$q^* = 24,6$ est un maximum de B , et $q^* = 119,4$ est un minimum de B .

Une production de 25 CD par semaine procurera un bénéfice maximum.

Exercice 12

La fonction bénéfice est donnée par $B(p) = -p^2 + 240 \cdot p - \frac{32000}{240 - 2 \cdot p} - 30$

$p^* = 100$ (dollars) est un maximum de B ; dans ce cas, $q^* = 140$ (tonnes).

Exercice 13

a) La fonction bénéfice est donnée par $B(q) = -\frac{3}{4} \cdot q^2 + 15 \cdot q - 25$

$q^* = 10$ est un maximum de B .

Le niveau de production qui maximise le bénéfice est de 10 unités.

b) Le coût de production unitaire est donné par $C_u(q) = \frac{C(q)}{q}$

$q^* = 10$ est un minimum de C .

Le niveau de production qui minimise les coûts de production est de 10 unités.

Exercice 14

La production est donnée par $\frac{MO^2}{25} \cdot (3 - \frac{MO}{12})$.

$MO^* = 24$ est un maximum de P et $MO^* = 0$ est un minimum de P .

L'effectif qui maximise la production est donc de 24.

Exercice 15

1. $\det(2) = 2 > 0$ **et** $\det \begin{pmatrix} 2 & -2 \\ -2 & 4 \end{pmatrix} = 4 > 0$ **et** $\det \begin{pmatrix} 2 & -2 & 2 \\ -2 & 4 & 0 \\ 2 & 0 & 8 \end{pmatrix} = 16 > 0$

La matrice $\begin{pmatrix} 2 & -2 & 2 \\ -2 & 4 & 0 \\ 2 & 0 & 8 \end{pmatrix}$ est définie positive

2. $\det(1) = 1 > 0$ **et** $\det \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 1 > 0$ **et** $\det \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = 1 > 0$

La matrice $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ est définie positive

3. $\det(-6) = -6 < 0$ **et** $\det \begin{pmatrix} -6 & 0 \\ 0 & -2 \end{pmatrix} = 12 > 0$ **et** $\det \begin{pmatrix} -6 & 0 & 6 \\ 0 & -2 & 0 \\ 6 & 0 & -12 \end{pmatrix} = -72 < 0$

La matrice $\begin{pmatrix} -6 & 0 & 6 \\ 0 & -2 & 0 \\ 6 & 0 & -12 \end{pmatrix}$ est définie négative

4. $\det(-3) = -3 < 0$ **et** $\det \begin{pmatrix} -3 & 2 \\ 2 & -1 \end{pmatrix} = -1 < 0$

La matrice $\begin{pmatrix} -3 & 2 \\ 2 & -1 \end{pmatrix}$ n'est ni définie positive, ni définie négative

5. $\det(1) = 1 > 0$ **et** $\det \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 1 > 0$ **et** $\det \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 2 & 0 & a \end{pmatrix} = a - 4$

La matrice $\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 2 & 0 & a \end{pmatrix}$ est définie positive si $a > 4$

6. $\det(2) = 2 > 0$ **et** $\det \begin{pmatrix} 2 & -2 \\ -2 & 5 \end{pmatrix} = 6 > 0$ **et** $\det \begin{pmatrix} 2 & -2 & 2 \\ -2 & 5 & 1 \\ 2 & 1 & 5 \end{pmatrix} = 0$

La matrice $\begin{pmatrix} 2 & -2 & 2 \\ -2 & 5 & 1 \\ 2 & 1 & 5 \end{pmatrix}$ n'est ni définie positive, ni définie négative

mais est semi-définie positive.

Exercice 16 : minimum global en (0,0)

Exercice 17 : point de selle en (0,0)

Exercice 18 : minimum local en (1,0)
point de selle en (1,2) et (-3,0)
maximum local en (-3,2)

Exercice 19 : minimum global en (-2,3)

Exercice 20 : minimum local en (2,2), point de selle en (-4,-4)

Exercice 21 : minimum local en (2,1), point de selle en (-8,-4)

Exercice 22

La fonction profit est donnée par le profit est donné par $12.q_A + 18.q_B - C(q_A, q_B)$

Le profit est maximisé par le niveau de production $q_A = 2$ et $q_B = 4$

Exercice 23

(3,3) est un **maximum** local de la fonction F sous la contrainte $x_1 + x_2 = 6$ avec $\lambda = 3$.

Exercice 24

$(\frac{2}{17}, \frac{8}{17})$ est un **minimum** local de la fonction F sous la contrainte $x_1 + 4.x_2 - 2 = 0$

avec $\lambda = \frac{4}{17}$.

Exercice 25

(9,7) est un maximum local de la fonction F sous la contrainte $x_1 + x_2 = 16$ avec $\lambda = 66$.

Exercice 26

(7,7) est un maximum local de F sous contrainte avec $\lambda = 1$.

(1,-1) est un minimum local de F sous contrainte avec $\lambda = -1$.

Exercice 27

a) La fonction F admet un maximum local en (1,2), un minimum local en (2,4), et deux points de selle en (1,4) et (2,2).

b) Sous la contrainte $2x_1 - x_2 = 0$, la fonction F admet un maximum local en (1,2) et un minimum local en (2,4).

c) Max en (1,2), et min en (0.5 ;1.5), (2 ;1.5), (0.5 ;4), (2,4).

REGRESSION ORTHOGONALE DANS \mathbb{R}^2 .

I. 1. NOTION D'ESPACE VECTORIEL EUCLIDIEN.

I.1.1. Espace vectoriel \mathbb{R}^n .

Soit n un entier strictement positif et \mathbb{R} le corps des nombres réels.

L'ensemble \mathbb{R}^n des n -uples (x_1, \dots, x_n) de nombres réels est muni de sa structure usuelle d'**espace vectoriel réel**, définie par les opérations :

$$(x_1, \dots, x_n) + (x'_1, \dots, x'_n) = (x_1 + x'_1, \dots, x_n + x'_n)$$

$$\lambda (x_1, \dots, x_n) = (\lambda x_1, \dots, \lambda x_n), \forall \lambda \in \mathbb{R}.$$

Notations.

$$\begin{pmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{pmatrix}$$

On identifiera un élément $X = (x_1, \dots, x_n)$ de \mathbb{R}^n avec la matrice $X = \begin{pmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{pmatrix}$ à n lignes et 1 colonne.

La transposée de cette matrice est la matrice $X^t = \begin{pmatrix} x_1 & \dots & x_n \end{pmatrix}$ à 1 ligne et n colonnes.

Les opérations dans \mathbb{R}^n sont alors définies par des opérations sur les matrices :

Addition :

$$\begin{pmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} x'_1 \\ \vdots \\ x'_i \\ \vdots \\ x'_n \end{pmatrix} = \begin{pmatrix} x_1 + x'_1 \\ \vdots \\ x_i + x'_i \\ \vdots \\ x_n + x'_n \end{pmatrix}$$

$$\begin{pmatrix} x_1 & \dots & x_n \end{pmatrix} + \begin{pmatrix} x'_1 & \dots & x'_n \end{pmatrix} = \begin{pmatrix} x_1 + x'_1 & \dots & x_n + x'_n \end{pmatrix}$$

Multiplication par un scalaire :

$$\lambda \begin{pmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \lambda x_1 \\ \vdots \\ \lambda x_i \\ \vdots \\ \lambda x_n \end{pmatrix}$$

$$\lambda \begin{pmatrix} x_1 & \dots & x_n \end{pmatrix} = \begin{pmatrix} \lambda x_1 & \dots & \lambda x_n \end{pmatrix}$$

Dans \mathbb{R}^n , les n éléments $e_i, i \in \{1, \dots, n\}$, dont toutes les coordonnées sont nulles, sauf la i^{e} qui vaut 1, forment une base, appelée la **base canonique** de \mathbb{R}^n .

Tout élément $X = (x_1, \dots, x_n)$ de \mathbb{R}^n s'écrit de manière unique sous la forme

$$X = \sum_{i=1}^{i=n} x_i e_i$$

I.1.2. Produit scalaire dans \mathbb{R}^n .

Soit Φ une application de $\mathbb{R}^n \times \mathbb{R}^n$ dans \mathbb{R} .

On notera aussi $\langle X | \Phi | Y \rangle$ ou $\langle X | Y \rangle_{\Phi}$, le nombre réel $\Phi(X, Y)$.

I.1.2.1. Définition.

On appelle produit scalaire dans \mathbb{R}^n toute application Φ de $\mathbb{R}^n \times \mathbb{R}^n$ dans \mathbb{R} qui possède les propriétés suivantes :

a) Bilinéarité.

— Linéarité par rapport à la première variable : $\Phi(X + X', Y) = \Phi(X, Y) + \Phi(X', Y)$ et $\Phi(\lambda X, Y) = \lambda \Phi(X, Y)$, quels que soient λ dans \mathbb{R} , X, X' et Y dans \mathbb{R}^n ; cette propriété s'écrit aussi

$$\langle X + X' | \Phi | Y \rangle = \langle X | \Phi | Y \rangle + \langle X' | \Phi | Y \rangle$$

— Linéarité par rapport à la deuxième variable : $\Phi(X, Y + Y') = \Phi(X, Y) + \Phi(X, Y')$ et $\Phi(X, \lambda Y) = \lambda \Phi(X, Y)$, quels que soient λ dans \mathbb{R} , X, Y et Y' dans \mathbb{R}^n ; cette propriété s'écrit aussi

$$\langle X | \Phi | Y + Y' \rangle = \langle X | \Phi | Y \rangle + \langle X | \Phi | Y' \rangle$$

b) Symétrie.

$\Phi(X, Y) = \Phi(Y, X)$, quels que soient X et Y dans \mathbb{R}^n :

$$\langle X | \Phi | Y \rangle = \langle Y | \Phi | X \rangle$$

c) Positivité.

$\Phi(X, X)$ est un nombre réel supérieur ou égal à 0, quel que soit X dans \mathbb{R}^n :

$$\langle X | \Phi | X \rangle \geq 0$$

d) Non dégénérescence.

$\Phi(X, X) = 0$ entraîne $X = 0$:

$$\langle X | \Phi | X \rangle = 0 \Rightarrow X = 0.$$

Autrement dit, le vecteur $0 = (0, \dots, 0, \dots, 0)$ de \mathbb{R}^n est l'unique solution de l'équation $\Phi(X, X) = 0$.

On dit aussi qu'un produit scalaire sur \mathbb{R}^n est une **forme bilinéaire symétrique positive non dégénérée**.

Le mot "forme" fait simplement référence au fait que les valeurs sont des scalaires.

Lorsqu'il est muni d'un produit scalaire, \mathbb{R}^n est appelé un **espace vectoriel euclidien**.

I.1.2.2. Exemples.

a) Produit scalaire canonique.

L'application de $\mathbb{R}^n \times \mathbb{R}^n$ dans \mathbb{R} définie par :

$$((x_1, \dots, x_n), (y_1, \dots, y_n)) \mapsto \langle X | Y \rangle = {}^t X Y = \begin{pmatrix} x_1 & \dots & x_j & \dots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_j \\ \vdots \\ y_n \end{pmatrix} = \sum_{i=1}^n x_i y_i$$

est un produit scalaire sur \mathbb{R}^n qu'on appelle le **produit scalaire canonique** de \mathbb{R}^n .

En effet, les propriétés de bilinéarité, de symétrie, de positivité et de non dégénérescence sont pratiquement évidentes à vérifier.

b) Produit scalaire défini par une matrice diagonale à éléments positifs.

Considérons une matrice réelle M à n lignes et n colonnes dont tous les éléments en dehors de la diagonale principale sont nuls ($m_{ij} = 0$, quels que soient les entiers i et j dans $\{1, \dots, n\}$ avec $i \neq j$) (on dit alors que M est une **matrice diagonale**) et dont les éléments de la diagonale principale sont des **nombre réels strictement positifs** ($m_{ii} > 0$ quel que soit l'entier i dans $\{1, \dots, n\}$).

Alors l'application :

$$(X, Y) \mapsto \langle X | M | Y \rangle = {}^t X M Y = \begin{pmatrix} x_1 & \dots & x_j & \dots & x_n \end{pmatrix} M \begin{pmatrix} y_1 \\ \vdots \\ y_j \\ \vdots \\ y_n \end{pmatrix} = \sum_{ij} m_{ij} x_j y_i = \sum_i m_{ii} x_i y_i$$

est un produit scalaire sur \mathbb{R}^n . La matrice M est appelée la **matrice des poids** (les "poids" sont les éléments de la diagonale).

En effet, les propriétés de bilinéarité, de symétrie, de positivité et de non dégénérescence sont pratiquement évidentes à vérifier.

- Le produit scalaire canonique correspond au cas où la matrice M est la matrice unité I_n (tous les éléments de la diagonale sont égaux à 1 et les éléments en dehors de la diagonale sont 0) : tous les poids sont égaux à 1.

- Autre exemple : $M = D^{\frac{1}{n}} = \frac{1}{n} I_n$. Tous les poids sont égaux à $\frac{1}{n}$ et la somme des poids vaut 1.

I.1.2.3. Propriétés.

a) Matrice d'un produit scalaire.

Pour tout produit scalaire Φ sur \mathbb{R}^n , on peut écrire :

$$\Phi(X, Y) = \Phi\left(\sum_i x_i e_i, \sum_j y_j e_j\right) = \sum_{ij} \Phi(e_i, e_j) x_i y_j = \begin{pmatrix} x_1 & \dots & x_i & \dots & x_n \end{pmatrix} M_\Phi \begin{pmatrix} y_1 \\ \vdots \\ y_j \\ \vdots \\ y_n \end{pmatrix}$$

La matrice $M_\Phi = [\Phi(e_i, e_j)]$ s'appelle la **matrice du produit scalaire** Φ dans la base canonique.

Cette matrice est une matrice symétrique : $\Phi(e_i, e_j) = \Phi(e_j, e_i)$.

Les éléments de sa diagonale sont des nombres réels strictement positifs : $\Phi(e_i, e_i) > 0$.

Remarquons ces propriétés ne sont pas suffisantes : une matrice symétrique dont les éléments de la diagonale sont des nombres réels strictement positifs ne définit pas forcément un produit scalaire.

Par exemple, la matrice $\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$ a un déterminant qui vaut $-3 < 0$, donc elle possède deux valeurs propres réelles de signe opposé (3 et -1) et la forme bilinéaire $((x_1, x_2), (y_1, y_2)) \mapsto (x_1, x_2) \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ qu'elle définit n'est pas un produit scalaire car le "produit scalaire" du vecteur propre $(1, -1)$ pour la valeur propre négative, par lui-même, est un nombre réel strictement négatif $((1, -1) \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = -2)$.

La matrice $\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$ n'est donc pas la matrice d'un produit scalaire sur \mathbb{R}^2 , bien qu'elle soit symétrique et que les éléments de sa diagonale soient strictement positifs.

En réalité, pour qu'une matrice carrée symétrique réelle soit la matrice d'un produit scalaire, il faut et il suffit que toutes ses valeurs propres, qui sont toujours des nombres réels, soient strictement positives. Mais ce résultat ne sera démontré, dans sa généralité, que plus tard, en algèbre.

b) Norme d'un vecteur.

Si Φ est un produit scalaire sur \mathbb{R}^n , le nombre réel positif $\|X\|_\Phi = \sqrt{\Phi[X, X]}$ s'appelle la **Φ -norme** de X , ou **Φ -longueur** de X .

Quand il n'y a pas de confusion à craindre, on parlera simplement de norme ou de longueur, qu'on notera $\|X\|$ au lieu de $\|X\|_\Phi$.

On dit qu'un vecteur est **normé pour Φ** si sa Φ -longueur est 1.

Par exemple, dans \mathbb{R}^2 muni du produit scalaire canonique, la longueur de $X = (x_1, x_2)$ est $\|X\| = \sqrt{x_1^2 + x_2^2}$ et le vecteur $(1, 0)$ est normé.

c) Angle de deux vecteurs.

Etant donnés deux vecteurs X et Y de \mathbb{R}^n et un produit scalaire Φ sur \mathbb{R}^n , pour tout nombre réel λ , on a :

$$\begin{aligned}\Phi(X + \lambda Y, X + \lambda Y) &= \|X + \lambda Y\|_{\Phi}^2 \geq 0 \\ \lambda^2 \Phi(Y, Y) + \lambda(\Phi(Y, X) + \Phi(X, Y)) + \Phi(X, X) &\geq 0 \\ \lambda^2 \Phi(Y, Y) + 2\lambda \Phi(X, Y) + \Phi(X, X) &\geq 0 \\ \|Y\|_{\Phi}^2 \lambda^2 + 2\langle X | Y \rangle_{\Phi} \lambda + \|X\|_{\Phi}^2 &\geq 0\end{aligned}$$

Comme cette relation est vraie pour tout nombre réel λ , c'est que le discriminant de ce trinôme du deuxième degré est négatif :

$$\begin{aligned}(\langle X | Y \rangle_{\Phi})^2 - \|X\|_{\Phi}^2 \|Y\|_{\Phi}^2 &\leq 0 \\ |\langle X | Y \rangle_{\Phi}| &\leq \|X\|_{\Phi} \|Y\|_{\Phi}\end{aligned}$$

Cette inégalité, valable pour tous vecteurs X et Y de \mathbb{R}^n constitue l'**inégalité de Schwarz**. Si les deux vecteurs X et Y sont différents de 0, leur longueur n'est pas nulle, le produit de

leurs longueurs n'est pas nul, le rapport $\frac{\langle X | Y \rangle_{\Phi}}{\|X\|_{\Phi} \|Y\|_{\Phi}}$ est compris entre -1 et 1 , et il existe donc un angle compris entre 0 et π radians dont le cosinus est égal au rapport $\frac{\langle X | Y \rangle_{\Phi}}{\|X\|_{\Phi} \|Y\|_{\Phi}}$.

Par définition, cet angle unique α compris entre 0 et π , vérifiant :

$$\cos \alpha = \frac{\langle X | Y \rangle_{\Phi}}{\|X\|_{\Phi} \|Y\|_{\Phi}} = \frac{\langle X | \Phi | Y \rangle}{\|X\|_{\Phi} \|Y\|_{\Phi}}$$

est appelé l'**angle des deux vecteurs non nuls** X et Y .

d) Orthogonalité.

Etant donnés deux vecteurs X et Y de \mathbb{R}^n et un produit scalaire Φ sur \mathbb{R}^n , on dit que X et Y sont **Φ -orthogonaux** (ou simplement "orthogonaux" s'il n'y a pas de confusion à craindre) si, et seulement si, leur produit scalaire est nul :

$$\Phi(X, Y) = \langle X | Y \rangle_{\Phi} = 0$$

Exemples :

- 0 est Φ -orthogonal à tout vecteur de \mathbb{R}^n .
- L'angle de deux vecteurs non nuls Φ -orthogonaux est $\frac{\pi}{2}$.
- La base canonique de \mathbb{R}^n muni du produit scalaire canonique est formée de vecteurs normés orthogonaux deux à deux : on parle alors de **base orthonormée**.

e) Projeté orthogonal.

Soient X et Y deux vecteurs non nuls de \mathbb{R}^n et Φ un produit scalaire sur \mathbb{R}^n . Il existe un unique vecteur Z de \mathbb{R}^n , proportionnel à Y et tel que $X - Z$ soit orthogonal à Y .

Démonstration.

Pour tout vecteur Z on peut écrire :

$$\langle X - Z | Y \rangle_{\Phi} = \langle X | Y \rangle_{\Phi} - \langle Z | Y \rangle_{\Phi}$$

Si l'on prend un Z proportionnel à Y , on a $Z = a Y$, donc :

$$\langle X - Z | Y \rangle_{\Phi} = \langle X | Y \rangle_{\Phi} - a \langle Y | Y \rangle_{\Phi} = \langle X | Y \rangle_{\Phi} - a \|Y\|_{\Phi}^2.$$

Pour que $X - Z$ soit orthogonal à Y , soit $\langle X - Z | Y \rangle_{\Phi} = 0$, il faut et il suffit que l'on prenne $a = \frac{\langle X | Y \rangle_{\Phi}}{\|Y\|_{\Phi}^2}$.

L'unique vecteur $Z = \frac{\langle X | Y \rangle_{\Phi}}{\|Y\|_{\Phi}^2} Y$, proportionnel à Y et tel que $X - Z$ soit orthogonal à Y , s'appelle le **projeté orthogonal** de X sur Y .

Propriété du projeté orthogonal.

Le projeté orthogonal Z_0 de X sur Y est le vecteur Z de \mathbb{R}^n proportionnel à Y , qui minimise $\|X - Z\|_{\Phi}^2$.

Démonstration.

Soit Z un vecteur proportionnel à Y .

Soit $Z_0 = \frac{\langle X | Y \rangle_{\Phi}}{\|Y\|_{\Phi}^2} Y$ le projeté orthogonal de X sur Y .

$$\|X - Z\|_{\Phi}^2 = \|X - Z_0 + Z_0 - Z\|_{\Phi}^2.$$

Comme Z est proportionnel à Y et que Z_0 est proportionnel à Y , la différence $Z_0 - Z$ est proportionnelle à Y .

Or $X - Z_0$ est orthogonal à Y , donc $X - Z_0$ est orthogonal à $Z_0 - Z$ qui est proportionnel à Y .

Il est résulte que l'on a :

$$\|X - Z\|_{\Phi}^2 = \|X - Z_0 + Z_0 - Z\|_{\Phi}^2 = \|X - Z_0\|_{\Phi}^2 + \|Z_0 - Z\|_{\Phi}^2 \geq \|X - Z_0\|_{\Phi}^2.$$

Et cette inégalité montre que $\|X - Z\|_{\Phi}^2$ atteint son minimum lorsque $Z = Z_0$.

I.2. APPROCHE EUCLIDIENNE DE LA REGRESSION.

Considérons une variable statistique quantitative bidimensionnelle (X, Y) à valeurs dans \mathbb{R}^2 , définie dans une population Ω de taille n .

Elle est définie par l'ensemble des couples $\{(X(\omega), Y(\omega))\}_{\omega \in \Omega}$.

\mathbb{R}^2 est l'**espace des individus**.

La variable statistique est représentée par un nuage de points dans \mathbb{R}^2 et chaque point du nuage statistique représente un individu de la population Ω .

I.2.1. Espace des variables.

Les n valeurs $X(\omega)$ de X pour les n individus de la population peuvent être considérées comme les coordonnées d'un vecteur de \mathbb{R}^n .

$$\begin{pmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{pmatrix}$$

Ce vecteur est noté encore $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$.

Les n valeurs $Y(\omega)$ de Y pour les n individus de la population peuvent être considérées

comme les coordonnées d'un vecteur de \mathbb{R}^n .

$$\begin{pmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix}$$

Ce vecteur est noté encore $Y = \begin{pmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix}$.

L'espace $E = \mathbb{R}^n$ apparaît alors comme l'espace des variables.

Chaque élément de E peut être considéré comme les valeurs d'une variable statistique quantitative réelle définie sur Ω .

I.2.2. Produit scalaire.

Dans cet espace des variables, la matrice $D^{\frac{1}{n}} = \frac{1}{n} I_n$, où I_n est la matrice unité à n lignes et n colonnes, définit un **produit scalaire** :

$$\langle X | Y \rangle^{D^{\frac{1}{n}}} = \langle X | D^{\frac{1}{n}} | Y \rangle = \sum_i \frac{1}{n} x_i y_i = \frac{1}{n} \sum_i x_i y_i = \frac{1}{n} \langle X | Y \rangle$$

en notant $\langle X | Y \rangle$ le produit scalaire canonique de \mathbb{R}^n .

$$\begin{pmatrix} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

On note $\mathbf{1}_n = \begin{pmatrix} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{pmatrix}$ le vecteur dont toutes les coordonnées sont égales à 1.

On l'appelle le **vecteur unité** de \mathbb{R}^n .

On remarquera que ce vecteur unité est normé, sa longueur est $\| \mathbf{1}_n \|^{D^{\frac{1}{n}}} = \frac{1}{n} \sum_i 1 \times 1 = \frac{1}{n} \times n = 1$.

I.2.3. Moyenne d'une variable statistique.

La moyenne \bar{X} de la variable statistique X est donnée par :

$$\bar{X} = \frac{1}{n} \sum_{\omega} X(\omega) = \frac{1}{n} \sum_i x_i = \frac{1}{n} \sum_i x_i \times 1 = \langle X | D^{\frac{1}{n}} | \mathbf{1}_n \rangle = \langle X | \mathbf{1}_n \rangle^{D^{\frac{1}{n}}}$$

La moyenne de X est le produit scalaire de X par le vecteur unité $\mathbf{1}_n$.

Notons X_0 la variable centrée correspondant à X : pour chaque individu ω de la population, sa valeur est $X(\omega) - \bar{X}$:

$$X_0 = \begin{pmatrix} x_1 - \bar{X} \\ \vdots \\ x_i - \bar{X} \\ \vdots \\ x_n - \bar{X} \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{pmatrix} - \bar{X} \begin{pmatrix} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{pmatrix} = X - \bar{X} \mathbf{1}_n$$

$$X = X_0 + \bar{X} \mathbf{1}_n = X_0 + \langle X | \mathbf{1}_n \rangle^{D^{\frac{1}{n}}} \mathbf{1}_n$$

I.2.4. Variance d'une variable statistique.

$$s^2(X) = \overline{X_0^2} = \frac{1}{n} \sum_i (x_i - \bar{X})^2 = \langle X_0 | D^{\frac{1}{n}} | X_0 \rangle = \| X_0 \|^2$$

$$s^2(X) = \| X_0 \|^2$$

La variance de X est le carré de la norme de la variable centrée.

I.2.5. Covariance.

La covariance de deux variables quantitatives réelles X et Y définies sur Ω est la moyenne du produit des variables centrées :

$$\text{Cov}(X, Y) = \frac{1}{n} \sum_i (x_i - \bar{X})(y_i - \bar{Y}) = \langle X_0 | D_{\frac{1}{n}} | Y_0 \rangle = \langle X_0 | Y_0 \rangle_{D_{\frac{1}{n}}}$$

$$\boxed{\text{Cov}(X, Y) = \langle X_0 | D_{\frac{1}{n}} | Y_0 \rangle = \langle X_0 | Y_0 \rangle_{D_{\frac{1}{n}}}}$$

La covariance est le produit scalaire des variables centrées.

I.2.6. Coefficient de corrélation linéaire.

$$r_{XY} = \frac{\text{Cov}(X, Y)}{s(X) s(Y)} = \frac{\langle X_0 | D_{\frac{1}{n}} | Y_0 \rangle}{\|X_0\|_{\Phi} \|Y_0\|_{\Phi}} = \cos(X_0, Y_0)$$

$$\boxed{r_{XY} = \cos(X_0, Y_0)}$$

Le coefficient de corrélation linéaire est le cosinus de l'angle des variables centrées.

I.2.7. Prédicteur linéaire.

Soient Y la variable à expliquer, X la variable explicative, X_0 et Y_0 les variables centrées.

Le prédicteur linéaire $\Delta_{Y|X}$ est $y^* = a + b x$ ou $y^* - \bar{Y} = b(x - \bar{X})$, soit $y_0^* = b x_0$.

Il est représenté par la **droite de régression** de Y en X dans l'espace des individus.

$$\text{Le coefficient } b \text{ s'obtient par } b = \frac{\text{Cov}(X, Y)}{s^2(X)} = \frac{\text{Cov}(X, Y)}{s^2(X_0)} = \frac{\langle X_0 | Y_0 \rangle_{D_{\frac{1}{n}}}}{\|X_0\|_{D_{\frac{1}{n}}}}.$$

D'après ce qui précède (I.1.2.3.e), $b X_0 = \frac{\langle X_0 | Y_0 \rangle_{D_{\frac{1}{n}}}}{\|X_0\|_{D_{\frac{1}{n}}}} X_0$ est le projeté orthogonal de Y_0 sur X_0 , $Y_0 - b X_0$ est orthogonal à X_0 et b est la valeur qui minimise l'expression

$$S^2 = \frac{1}{n} \sum_i (Y_{0i} - b X_{0i})^2 = \|Y_0 - b X_0\|_{D_{\frac{1}{n}}}^2 = s^2(Y - bX) = s^2(Y - a - bX) = s^2(Y - Y^*) = s^2(Y_0 - Y_0^*)$$

Le prédicteur linéaire de la variable centrée Y_0 est le projeté orthogonal de Y_0 sur X_0 dans \mathbb{R}^n .
C'est la variable Y_0^* qui minimise la variance de $Y_0 - Y_0^*$.

Nous avons alors :

$$s^2(Y) = \|Y_0\|_{D_{\frac{1}{n}}}^2 = \|Y_0 - b X_0 + b X_0\|_{D_{\frac{1}{n}}}^2 = \|Y_0 - b X_0\|_{D_{\frac{1}{n}}}^2 + \|b X_0\|_{D_{\frac{1}{n}}}^2$$

$$s^2(Y) = S^2_{min} + b^2 \|X_0\|_{D_{\frac{1}{n}}}^2 = S^2_{min} + \left(\frac{\text{Cov}(X, Y)}{s^2(X)} \right)^2 s^2(X) = S^2_{min} + \left(\frac{\text{Cov}(X, Y)}{s(X) s(Y)} \right)^2 s^2(Y)$$

$$s^2(Y) = S^2_{min} + r_{XY}^2 s^2(Y).$$

Nous retrouvons la variance résiduelle S^2_{min} et la variance expliquée par la régression $r_{XY}^2 s^2(Y)$.

De façon symétrique, si X est la variable explicative et Y la variable explicative, nous aurons une expression :

$$s^2(X) = S'^2_{min} + r_{XY}^2 s^2(X).$$

avec la variance résiduelle S'^2_{min} et la variance expliquée par la régression $r_{XY}^2 s^2(X)$.

II - REGRESSION MULTIPLE.

II. 1. POSITION ET RESOLUTION DU PROBLEME.

II.1.1. Position du problème.

Considérons trois variables statistiques réelles centrées X_0, Y_0, Z_0 , définies par n triplets $(x_{0i}, y_{0i}, z_{0i}), i \in [1, n]$.

Nous considérons Z_0 comme la variable à expliquer et X_0 et Y_0 comme les variables explicatives.

Nous supposons que les observations laissent à penser que le nuage de points dans \mathbb{R}^3 pourrait être modélisé par un plan.

Le problème de la régression linéaire multiple de Z_0 en X_0 et Y_0 consiste à trouver un prédicteur

$$\hat{z}_0 = a X_0 + b Y_0$$

de Z_0 , tel que le nuage de points $(x_{0i}, y_{0i}, \hat{z}_{0i} = a x_{0i} + b y_{0i}), i \in [1, n]$, soit aussi proche possible du nuage de points $(x_{0i}, y_{0i}, z_{0i}), i \in [1, n]$, au sens des moindres carrés.

L'approche euclidienne de ce problème dans \mathbb{R}^n consiste à trouver un $\hat{z}_0 = a X_0 + b Y_0 \in \mathbb{R}^n$ tel

que $S^2 = \|Z_0 - \hat{z}_0\|_D^2$ soit minimum.

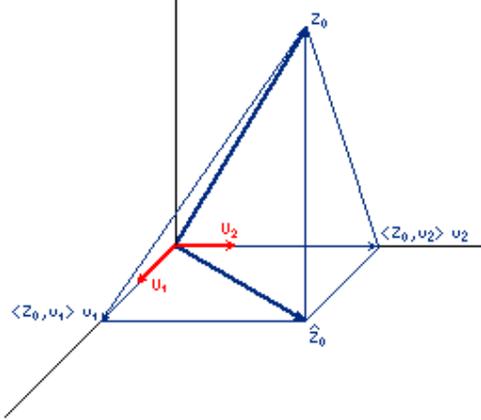
Le problème est donc de trouver, dans \mathbb{R}^n , un vecteur \hat{z}_0 du plan (= sous-espace vectoriel de dimension 2) Π défini par X_0 et Y_0 , tel que le vecteur $Z_0 - \hat{z}_0$ ait une longueur minimum (au sens du produit scalaire défini par la matrice des poids $D^{\frac{1}{2}}$).

La solution sera fournie par le projeté orthogonal \hat{z}_0 de Z_0 sur Π .

II.1.2. Projeté orthogonal sur un plan.

a) Définition.

Si nous connaissons une base orthonormée $\{u_1, u_2\}$ d'un sous-espace vectoriel Π de dimension 2, défini dans \mathbb{R}^n par les deux vecteurs X_0 et Y_0 , nous savons calculer le projeté



$$\frac{\langle Z_0 | u_1 \rangle_{D_1} u_1}{\|u_1\|_{D_1}^2}$$

orthogonal de Z_0 sur u_1 , c'est le vecteur $\langle Z_0 | u_1 \rangle_{D_1} u_1$ et

nous savons calculer aussi le projeté orthogonal $\langle Z_0 | u_2 \rangle_{D_1} u_2$ de Z_0 sur u_2 .

On appelle **projeté orthogonal** de Z_0 sur Π , l'unique vecteur \hat{Z}_0 de Π tel que $Z_0 - \hat{Z}_0$ soit orthogonal à Π .

Un tel vecteur existe et est unique.

Démonstration.

Notons \hat{Z}_0 le vecteur $\langle Z_0 | u_1 \rangle_{D_1} u_1 + \langle Z_0 | u_2 \rangle_{D_1} u_2$, somme des projetés orthogonaux de Z_0 sur les vecteurs u_1 et u_2 .

$$\begin{aligned} \langle Z_0 - \hat{Z}_0 | u_1 \rangle_{D_1} &= \langle Z_0 | u_1 \rangle_{D_1} - \langle \hat{Z}_0 | u_1 \rangle_{D_1} \\ &= \langle Z_0 | u_1 \rangle_{D_1} - \langle \langle Z_0 | u_1 \rangle_{D_1} u_1 + \langle Z_0 | u_2 \rangle_{D_1} u_2 | u_1 \rangle_{D_1} \\ &= \langle Z_0 | u_1 \rangle_{D_1} - \langle Z_0 | u_1 \rangle_{D_1} \langle u_1 | u_1 \rangle_{D_1} + \langle Z_0 | u_2 \rangle_{D_1} \langle u_2 | u_1 \rangle_{D_1} \\ &= \langle Z_0 | u_1 \rangle_{D_1} - \langle Z_0 | u_1 \rangle_{D_1} \\ &= 0 \end{aligned}$$

$$\begin{aligned} \langle Z_0 - \hat{Z}_0 | u_2 \rangle_{D_1} &= \langle Z_0 | u_2 \rangle_{D_1} - \langle \hat{Z}_0 | u_2 \rangle_{D_1} \\ &= \langle Z_0 | u_2 \rangle_{D_1} - \langle \langle Z_0 | u_1 \rangle_{D_1} u_1 + \langle Z_0 | u_2 \rangle_{D_1} u_2 | u_2 \rangle_{D_1} \\ &= \langle Z_0 | u_2 \rangle_{D_1} - \langle Z_0 | u_1 \rangle_{D_1} \langle u_1 | u_2 \rangle_{D_1} + \langle Z_0 | u_2 \rangle_{D_1} \langle u_2 | u_2 \rangle_{D_1} \\ &= \langle Z_0 | u_2 \rangle_{D_1} - \langle Z_0 | u_2 \rangle_{D_1} \\ &= 0 \end{aligned}$$

Ainsi, $Z_0 - \hat{Z}_0$ est orthogonal à u_1 et à u_2 , il est donc orthogonal à toute combinaison linéaire de u_1 et u_2 , c'est-à-dire à tout élément de Π : on dit qu'il est orthogonal à Π .

Le projeté orthogonal de \hat{Z}_0 sur u_1 est

$$\langle \hat{Z}_0 | u_1 \rangle_{D_1} u_1 = \langle Z_0 | u_1 \rangle_{D_1} u_1.$$

Le projeté orthogonal de \hat{Z}_0 sur u_2 est

$$\langle \hat{Z}_0 | u_2 \rangle_{D_1} u_2 = \langle Z_0 | u_2 \rangle_{D_1} u_2.$$

Nous pouvons donc écrire :

$$\hat{Z}_0 = \langle Z_0 | u_1 \rangle_{D_1} u_1 + \langle Z_0 | u_2 \rangle_{D_1} u_2 = \langle \hat{Z}_0 | u_1 \rangle_{D_1} u_1 + \langle \hat{Z}_0 | u_2 \rangle_{D_1} u_2.$$

Réciproquement, si Z est un vecteur de Π tel que $Z_0 - Z$ soit orthogonal à Π , nous avons :

$$Z = \langle Z | u_1 \rangle \frac{D_1}{n} u_1 + \langle Z | u_2 \rangle \frac{D_1}{n} u_2 = \langle Z_0 | u_1 \rangle \frac{D_1}{n} u_1 + \langle Z_0 | u_2 \rangle \frac{D_1}{n} u_2 = \hat{Z}_0.$$

Le vecteur :

$$\hat{Z}_0 = \langle Z_0 | u_1 \rangle \frac{D_1}{n} u_1 + \langle Z_0 | u_2 \rangle \frac{D_1}{n} u_2$$

est donc l'unique vecteur de Π tel que $Z_0 - \hat{Z}_0$ soit orthogonal à Π : c'est, par définition, le projeté orthogonal de Z_0 sur Π .

La relation :

$$\hat{Z}_0 = \langle \hat{Z}_0 | u_1 \rangle \frac{D_1}{n} u_1 + \langle \hat{Z}_0 | u_2 \rangle \frac{D_1}{n} u_2$$

signifie que le projeté orthogonal de \hat{Z}_0 sur le plan Π est \hat{Z}_0 .

b) Propriété du projeté orthogonal.

Le projeté orthogonal de Z_0 sur Π est le vecteur Z de Π , qui minimise la quantité $\| Z_0 - Z \|^2$.

Démonstration.

Soit Z un vecteur appartenant au sous-espace Π .

Soit $\hat{Z}_0 = \langle Z_0 | u_1 \rangle \frac{D_1}{n} u_1 + \langle Z_0 | u_2 \rangle \frac{D_1}{n} u_2$ le projeté orthogonal de Z_0 sur Π .

$$\| Z_0 - Z \|^2 = \| Z_0 - \hat{Z}_0 + \hat{Z}_0 - Z \|^2$$

Or $Z_0 - \hat{Z}_0$ est orthogonal à Π , donc orthogonal à tout élément de Π , donc $Z_0 - \hat{Z}_0$ est orthogonal à \hat{Z}_0 et à Z , donc aussi à $\hat{Z}_0 - Z$.

Le théorème de Pythagore s'applique :

$$\begin{aligned} \| Z_0 - \hat{Z}_0 + \hat{Z}_0 - Z \|^2 &= \| Z_0 - \hat{Z}_0 \|^2 + \| \hat{Z}_0 - Z \|^2 \\ \| Z_0 - Z \|^2 &= \| Z_0 - \hat{Z}_0 \|^2 + \| \hat{Z}_0 - Z \|^2 \end{aligned}$$

Cette relation montre que $\| Z_0 - Z \|^2$ atteint sa valeur minimum $\| Z_0 - \hat{Z}_0 \|^2$ lorsque $Z = \hat{Z}_0$.

Notre problème initial se trouve résolu :

Le prédicteur $\hat{Z}_0 = a X_0 + b Y_0$ de Z_0 qui rend minimum la quantité $S^2 = \| Z_0 - \hat{Z}_0 \|^2$ est le projeté orthogonal de Z_0 dans le plan Π défini par X_0 et Y_0 .

La seule chose qu'il nous reste à faire dans la suite, est d'explicitier ce projeté orthogonal en fonction des données (x_{0i}, y_{0i}, z_{0i}) , $i \in [1, n]$.

II.1.3. Choix d'une base orthonormée $\{ u_1, u_2 \}$.

Dans le plan Π défini par X_0 et Y_0 , nous pouvons définir un premier vecteur normé u_1 par :

$$u_1 = \frac{X_0}{\|X_0\|_{D_1}^{\frac{1}{n}}} = \frac{X_0}{s(X)}$$

On a, en effet : $s^2(X) = \|X_0\|_{D_1}^{\frac{1}{n}}$.

$$\frac{\langle Y_0 | X_0 \rangle_{D_1}^{\frac{1}{n}}}{\|X_0\|_{D_1}^{\frac{1}{n}}} X_0 \text{ et } Y_0 - \frac{\langle Y_0 | X_0 \rangle_{D_1}^{\frac{1}{n}}}{\|X_0\|_{D_1}^{\frac{1}{n}}} X_0 \text{ est orthogonal à } X_0.$$

Le projeté orthogonal de Y_0 sur X_0 est $\frac{\langle Y_0 | X_0 \rangle_{D_1}^{\frac{1}{n}}}{\|X_0\|_{D_1}^{\frac{1}{n}}} X_0$ et $Y_0 - \frac{\langle Y_0 | X_0 \rangle_{D_1}^{\frac{1}{n}}}{\|X_0\|_{D_1}^{\frac{1}{n}}} X_0$ est orthogonal à X_0 .
Le carré de sa norme est donné par :

$$\begin{aligned} \left\| Y_0 - \frac{\langle Y_0 | X_0 \rangle_{D_1}^{\frac{1}{n}}}{\|X_0\|_{D_1}^{\frac{1}{n}}} X_0 \right\|_{D_1}^2 &= \|Y_0\|_{D_1}^2 + \left(\frac{\langle Y_0 | X_0 \rangle_{D_1}^{\frac{1}{n}}}{\|X_0\|_{D_1}^{\frac{1}{n}}} \right)^2 \|X_0\|_{D_1}^2 - 2 \frac{\langle Y_0 | X_0 \rangle_{D_1}^{\frac{1}{n}}}{\|X_0\|_{D_1}^{\frac{1}{n}}} \langle Y_0 | X_0 \rangle_{D_1}^{\frac{1}{n}} \\ &= s^2(Y) - s^2(Y) \left(\frac{\text{Cov}(X, Y)}{s(X) s(Y)} \right)^2 = s^2(Y) (1 - r_{XY}^2) = \frac{s^2(X) s^2(Y) - [\text{Cov}(X, Y)]^2}{s^2(X)} \end{aligned}$$

On peut donc prendre dans le plan Π , pour vecteur normé u_2 orthogonal à u_1 , le vecteur :

$$u_2 = \frac{1}{s(X) \sqrt{1 - r_{XY}^2}} \left(Y_0 - \frac{\text{Cov}(X, Y)}{s^2(X)} X_0 \right) = \frac{s(X)}{\sqrt{s^2(X) s^2(Y) - [\text{Cov}(X, Y)]^2}} \left(Y_0 - \frac{\text{Cov}(X, Y)}{s^2(X)} X_0 \right)$$

Les vecteurs :

$$u_1 = \frac{X_0}{s(X)}$$

$$u_2 = \frac{s(X)}{\sqrt{s^2(X) s^2(Y) - [\text{Cov}(X, Y)]^2}} \left(Y_0 - \frac{\text{Cov}(X, Y)}{s^2(X)} X_0 \right)$$

forment une base orthonormée du plan Π défini par X_0 et Y_0 .

II.1.4. Calcul du projeté orthogonal de Z_0 .

Soit

$$\hat{Z}_0 = \langle Z_0 | u_1 \rangle_{D_1}^{\frac{1}{n}} u_1 + \langle Z_0 | u_2 \rangle_{D_1}^{\frac{1}{n}} u_2$$

le projeté orthogonal de Z_0 sur Π .

La première composante est le projeté orthogonal de Z_0 sur u_1 :

$$\langle Z_0 | u_1 \rangle_{D_1}^{\frac{1}{n}} u_1 = \langle Z_0 | \frac{X_0}{s(X)} \rangle_{D_1}^{\frac{1}{n}} \frac{X_0}{s(X)} = \frac{\text{Cov}(X, Z)}{s^2(X)} X_0$$

C'est aussi le projeté orthogonal de Z_0 sur X_0 .

La deuxième composante est le projeté orthogonal de Z_0 sur u_2 :

$$\begin{aligned} \langle Z_0 | u_2 \rangle_{D_1}^{\frac{1}{n}} u_2 &= \langle Z_0 | \frac{s(X)}{\sqrt{s^2(X) s^2(Y) - [\text{Cov}(X, Y)]^2}} \left(Y_0 - \frac{\text{Cov}(X, Y)}{s^2(X)} X_0 \right) \rangle_{D_1}^{\frac{1}{n}} \\ &= \frac{s(X)}{\sqrt{s^2(X) s^2(Y) - [\text{Cov}(X, Y)]^2}} \left(\langle Z_0 | Y_0 \rangle_{D_1}^{\frac{1}{n}} - \frac{\text{Cov}(X, Y)}{s^2(X)} \langle Z_0 | X_0 \rangle_{D_1}^{\frac{1}{n}} \right) \end{aligned}$$

$$\begin{aligned}
&= \frac{s^2(X)}{s^2(X)s^2(Y) - (\text{Cov}(X,Y))^2} \left(< Z_0 | Y_0 > \frac{D_{11}}{n} - \frac{\text{Cov}(X,Y)}{s^2(X)} < Z_0 | X_0 > \frac{D_{11}}{n} \right) \left(Y_0 - \frac{\text{Cov}(X,Y)}{s^2(X)} X_0 \right) \\
&= \frac{s^2(X) \text{Cov}(Z,Y) - \text{Cov}(Z,X) \text{Cov}(X,Y)}{s^2(X)s^2(Y) - (\text{Cov}(X,Y))^2} \left(Y_0 - \frac{\text{Cov}(X,Y)}{s^2(X)} X_0 \right)
\end{aligned}$$

Au total, nous obtenons :

$$\begin{aligned}
\hat{Z}_0 &= \frac{\text{Cov}(X,Z)}{s^2(X)} X_0 + \frac{s^2(X) \text{Cov}(Z,Y) - \text{Cov}(Z,X) \text{Cov}(X,Y)}{s^2(X)s^2(Y) - (\text{Cov}(X,Y))^2} \left(Y_0 - \frac{\text{Cov}(X,Y)}{s^2(X)} X_0 \right) \\
&= \frac{1}{s^2(X)} \left(\text{Cov}(X,Z) - \frac{s^2(X) \text{Cov}(Z,Y) - \text{Cov}(Z,X) \text{Cov}(X,Y)}{s^2(X)s^2(Y) - (\text{Cov}(X,Y))^2} \text{Cov}(X,Y) \right) X_0 + \\
&\quad \frac{s^2(X) \text{Cov}(Z,Y) - \text{Cov}(Z,X) \text{Cov}(X,Y)}{s^2(X)s^2(Y) - (\text{Cov}(X,Y))^2} Y_0 \\
&= \frac{s^2(Y) \text{Cov}(X,Z) - \text{Cov}(X,Y) \text{Cov}(Y,Z)}{s^2(X)s^2(Y) - (\text{Cov}(X,Y))^2} X_0 + \frac{s^2(X) \text{Cov}(Y,Z) - \text{Cov}(X,Y) \text{Cov}(X,Z)}{s^2(X)s^2(Y) - (\text{Cov}(X,Y))^2} Y_0 \\
&= \boxed{\frac{s^2(Y) \text{Cov}(X,Z) - \text{Cov}(X,Y) \text{Cov}(Y,Z)}{s^2(X)s^2(Y) - (\text{Cov}(X,Y))^2} X_0 + \frac{s^2(X) \text{Cov}(Y,Z) - \text{Cov}(X,Y) \text{Cov}(X,Z)}{s^2(X)s^2(Y) - (\text{Cov}(X,Y))^2} Y_0}
\end{aligned}$$

Cette expression est symétrique en X et Y.

On sait calculer les quantités qui interviennent dans cette expression en fonction des données

$(x_{0i}, y_{0i}, z_{0i}), i \in [1, n]$.

On commence par calculer la matrice des variances-covariances :

$$A = \frac{1}{n} \begin{pmatrix} x_{01} & \dots & x_{0n} \\ y_{01} & \dots & y_{0n} \\ z_{01} & \dots & z_{0n} \end{pmatrix} \begin{pmatrix} x_{01} & y_{01} & z_{01} \\ \vdots & \vdots & \vdots \\ x_{0n} & y_{0n} & z_{0n} \end{pmatrix} = \begin{pmatrix} s^2(X) & \text{Cov}(X,Y) & \text{Cov}(X,Z) \\ \text{Cov}(X,Y) & s^2(Y) & \text{Cov}(Y,Z) \\ \text{Cov}(X,Z) & \text{Cov}(Y,Z) & s^2(Z) \end{pmatrix}$$

Formellement, la relation $\hat{Z}_0 = \frac{s^2(Y) \text{Cov}(X,Z) - \text{Cov}(X,Y) \text{Cov}(Y,Z)}{s^2(X)s^2(Y) - (\text{Cov}(X,Y))^2} X_0 +$

$\frac{s^2(X) \text{Cov}(Y,Z) - \text{Cov}(X,Y) \text{Cov}(X,Z)}{s^2(X)s^2(Y) - (\text{Cov}(X,Y))^2} Y_0$ peut se mémoriser comme un "déterminant" :

$$\begin{vmatrix} s^2(X) & \text{Cov}(X,Y) & X_0 \\ \text{Cov}(X,Y) & s^2(Y) & Y_0 \\ \text{Cov}(X,Z) & \text{Cov}(Y,Z) & \hat{Z}_0 \end{vmatrix} = 0$$

On a remplacé la dernière colonne de la matrice des variances-covariances par $\begin{pmatrix} X_0 \\ Y_0 \\ \hat{Z}_0 \end{pmatrix}$.

II.2. COEFFICIENT DE CORRELATION MULTIPLE.

II.2.1. Définition.

Nous connaissons déjà les formules donnant les coefficients de corrélation linéaire entre deux variables :

$$r_{XY} = \frac{\text{Cov}(X, Y)}{s(X) s(Y)} = \frac{\langle X_0 | Y_0 \rangle_{D_1^n}}{s(X) s(Y)} ; r_{XZ} = \frac{\text{Cov}(X, Z)}{s(X) s(Z)} ; r_{YZ} = \frac{\text{Cov}(Y, Z)}{s(Y) s(Z)} .$$

Les coefficients de X_0 et Y_0 dans l'expression de \hat{Z}_0 deviennent :

$$\frac{s^2(Y) \text{Cov}(X, Z) - \text{Cov}(X, Y) \text{Cov}(Y, Z)}{s^2(X) s^2(Y) - (\text{Cov}(X, Y))^2} = \frac{s^2(Y) r_{XZ} s(X) s(Z) - r_{XY} s(X) s(Y) r_{YZ} s(Y) s(Z)}{s^2(X) s^2(Y) - (r_{XY} s(X) s(Y))^2} = \frac{s(X) s^2(Y) s(Z)}{s^2(X) s^2(Y)} \times \frac{r_{XZ} - r_{XY} r_{YZ}}{1 - r_{XY}^2} = \frac{s(Z)}{s(X)} \frac{r_{XZ} - r_{XY} r_{YZ}}{1 - r_{XY}^2}$$

et, en échangeant X et Y :

$$\frac{s^2(X) \text{Cov}(Y, Z) - \text{Cov}(X, Y) \text{Cov}(X, Z)}{s^2(X) s^2(Y) - (\text{Cov}(X, Y))^2} = \frac{s(Z)}{s(Y)} \frac{r_{YZ} - r_{XY} r_{XZ}}{1 - r_{XY}^2}$$

En reportant, dans l'expression de \hat{Z}_0 , les expressions obtenues pour les coefficients, on obtient :

$$\hat{Z}_0 = \frac{s(Z)}{s(X)} \frac{r_{XZ} - r_{XY} r_{YZ}}{1 - r_{XY}^2} X_0 + \frac{s(Z)}{s(Y)} \frac{r_{YZ} - r_{XY} r_{XZ}}{1 - r_{XY}^2} Y_0$$

$$\frac{\hat{Z}_0}{s(Z)} = \frac{r_{XZ} - r_{XY} r_{YZ}}{1 - r_{XY}^2} \frac{X_0}{s(X)} + \frac{r_{YZ} - r_{XY} r_{XZ}}{1 - r_{XY}^2} \frac{Y_0}{s(Y)}$$

Les vecteurs $\frac{X_0}{s(X)}$ et $\frac{Y_0}{s(Y)}$ sont normés pour le produit scalaire de \mathbb{R}^n : $\|X_0\|_{D_1^n}^2 = s^2(X)$ et $\|Y_0\|_{D_1^n}^2 = s^2(Y)$.

$$\begin{aligned} \left\| \frac{\hat{Z}_0}{s(Z)} \right\|_{D_1^n}^2 &= \frac{\|\hat{Z}_0\|_{D_1^n}^2}{s^2(Z)} = \left(\frac{r_{XZ} - r_{XY} r_{YZ}}{1 - r_{XY}^2} \right)^2 + \left(\frac{r_{YZ} - r_{XY} r_{XZ}}{1 - r_{XY}^2} \right)^2 + 2 \frac{r_{XZ} - r_{XY} r_{YZ}}{1 - r_{XY}^2} \frac{r_{YZ} - r_{XY} r_{XZ}}{1 - r_{XY}^2} \frac{\langle X_0 | Y_0 \rangle_{D_1^n}}{s(X) s(Y)} \\ &= \frac{1}{(1 - r_{XY}^2)^2} \left(r_{XZ}^2 + r_{XY}^2 r_{YZ}^2 - 2 r_{XY} r_{XZ} r_{YZ} + r_{YZ}^2 + r_{XY}^2 r_{XZ}^2 - 2 r_{XY} r_{XZ} r_{YZ} + 2 r_{XY} (r_{XZ} r_{YZ} - r_{XY} r_{XZ}^2 - r_{XY} r_{YZ}^2 + r_{XY}^2 r_{XZ} r_{YZ}) \right) \\ &= \frac{1}{(1 - r_{XY}^2)^2} \left(r_{XZ}^2 + r_{XY}^2 r_{YZ}^2 - 2 r_{XY} r_{XZ} r_{YZ} + r_{YZ}^2 + r_{XY}^2 r_{XZ}^2 - 2 r_{XY} r_{XZ} r_{YZ} + 2 r_{XY} r_{XZ} r_{YZ} - 2 r_{XY}^2 r_{XZ}^2 - 2 r_{XY}^2 r_{YZ}^2 + 2 r_{XY}^3 r_{XZ} r_{YZ} \right) \\ &= \frac{1}{(1 - r_{XY}^2)^2} \left(r_{XZ}^2 + r_{XY}^2 r_{XZ}^2 - 2 r_{XY}^2 r_{XZ}^2 + r_{YZ}^2 + r_{XY}^2 r_{YZ}^2 - 2 r_{XY}^2 r_{YZ}^2 - 2 r_{XY} r_{XZ} r_{YZ} - 2 r_{XY} r_{XZ} r_{YZ} + r_{YZ}^2 + 2 r_{XY} r_{XZ} r_{YZ} + 2 r_{XY}^3 r_{XZ} r_{YZ} \right) \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{(1-r_{XY}^2)^2} \left(r_{XZ}^2 - r_{XY}^2 r_{XZ}^2 + r_{YZ}^2 - r_{XY}^2 r_{YZ}^2 - 2 r_{XY} r_{XZ} r_{YZ} + 2 r_{XY}^3 r_{XZ} r_{YZ} \right) \\
&= \frac{1}{(1-r_{XY}^2)^2} \left(r_{XZ}^2 (1-r_{XY}^2) + r_{YZ}^2 (1-r_{XY}^2) - 2 r_{XY} r_{XZ} r_{YZ} (1-r_{XY}^2) \right) \\
&= \frac{1}{1-r_{XY}^2} \left(r_{XZ}^2 + r_{YZ}^2 - 2 r_{XY} r_{XZ} r_{YZ} \right)
\end{aligned}$$

Le coefficient :

$$R_{Z|XY} = \sqrt{\frac{1}{1-r_{XY}^2} (r_{XZ}^2 + r_{YZ}^2 - 2 r_{XY} r_{XZ} r_{YZ})}$$

s'appelle le **coefficient de corrélation linéaire multiple** de Z en X, Y.

La variance du prédicteur de Z est donnée par :

$$s^2(\hat{Z}) = \|\hat{Z}_0\|^2 \frac{D_1^2}{n} = R_{Z|XY}^2 s^2(Z)$$

II.2.2. Propriétés.

a) Validité du prédicteur de Z.

La variance de Z s'écrit :

$$s^2(Z) = s^2(Z_0) = \|Z_0\|^2 \frac{D_1^2}{n} = \|Z_0 - \hat{Z}_0 + \hat{Z}_0\|^2 \frac{D_1^2}{n} = \|Z_0 - \hat{Z}_0\|^2 \frac{D_1^2}{n} + \|\hat{Z}_0\|^2 \frac{D_1^2}{n}$$

Or $\|Z_0 - \hat{Z}_0\|^2 \frac{D_1^2}{n}$ est la valeur minimum de la quantité $S^2 = \|Z_0 - \hat{Z}\|^2 \frac{D_1^2}{n}$ pour les $\hat{Z} \in \Pi : \|Z_0 - \hat{Z}_0\|^2$

$\|\hat{Z}_0\|^2 \frac{D_1^2}{n} = S^2_{min}$, c'est la variance "**résiduelle**", donc

$$s^2(Z) = S^2_{min} + R_{Z|XY}^2 s^2(Z)$$

On retrouve la même formule de décomposition de la variance que pour la régression linéaire : la variance de Z est la somme de la variance expliquée $R_{Z|XY}^2 s^2(Z)$ par la régression linéaire multiple, et de la variance résiduelle $S^2_{min} = (1 - R_{Z|XY}^2) s^2(Z)$.

Plus le coefficient $R_{Z|XY}^2$ est proche de 1, plus la part de variance de Z expliquée par la régression linéaire multiple en X et Y est grande, donc meilleur est le prédicteur linéaire \hat{Z}_0 . La validité du prédicteur \hat{Z}_0 est mesurée par le coefficient $R_{Z|XY}^2$.

b) Calcul pratique du coefficient de corrélation linéaire multiple.

En pratique, le calcul du coefficient de corrélation linéaire multiple $R_{Z|XY}$ s'effectue de la façon suivante :

— On calcule la **matrice des corrélations** de X et Y à partir de la matrice $V_{XY} =$

$$\begin{pmatrix} \frac{x_1 - \bar{X}}{s(X)} & \frac{y_1 - \bar{Y}}{s(Y)} \\ \vdots & \vdots \\ \frac{x_n - \bar{X}}{s(X)} & \frac{y_n - \bar{Y}}{s(Y)} \end{pmatrix} \text{ des données } (X, Y) \text{ réduites :}$$

$$C_{XY} = \begin{pmatrix} 1 & r_{XY} \\ r_{XY} & 1 \end{pmatrix} = \frac{1}{n} \begin{pmatrix} \frac{x_1 - \bar{X}}{s(X)} & \dots & \frac{x_n - \bar{X}}{s(X)} \\ \frac{y_1 - \bar{Y}}{s(Y)} & \dots & \frac{y_n - \bar{Y}}{s(Y)} \end{pmatrix} \begin{pmatrix} \frac{x_1 - \bar{X}}{s(X)} & \frac{y_1 - \bar{Y}}{s(Y)} \\ \vdots & \vdots \\ \frac{x_n - \bar{X}}{s(X)} & \frac{y_n - \bar{Y}}{s(Y)} \end{pmatrix} = {}^t V_{XY} D^{\frac{1}{n}} V_{XY}.$$

— On calcule l'**inverse** de cette matrice des corrélations :

$$C_{XY}^{-1} = \frac{1}{1 - r_{XY}^2} \begin{pmatrix} 1 & -r_{XY} \\ -r_{XY} & 1 \end{pmatrix}$$

— La matrice des **coefficients de corrélation linéaire** de X et Y avec Z, peut se calculer à

partir de la matrice V_{XY} et de la variable centrée réduite $V_Z = \begin{pmatrix} \frac{z_1 - \bar{Z}}{s(Z)} \\ \vdots \\ \frac{z_n - \bar{Z}}{s(Z)} \end{pmatrix}$ par la formule :

$$\begin{pmatrix} r_{XZ} \\ r_{YZ} \end{pmatrix} = \frac{1}{n} \begin{pmatrix} \frac{x_1 - \bar{X}}{s(X)} & \dots & \frac{x_n - \bar{X}}{s(X)} \\ \frac{y_1 - \bar{Y}}{s(Y)} & \dots & \frac{y_n - \bar{Y}}{s(Y)} \end{pmatrix} \begin{pmatrix} \frac{z_1 - \bar{Z}}{s(Z)} \\ \vdots \\ \frac{z_n - \bar{Z}}{s(Z)} \end{pmatrix} = {}^t V_{XY} D^{\frac{1}{n}} V_Z.$$

— Le **coefficient de corrélation linéaire multiple** $R_{Z|XY}$ est donné par la formule :

$$R_{Z|XY}^2 = \frac{1}{1 - r_{XY}^2} \left(r_{XZ}^2 + r_{YZ}^2 - 2 r_{XY} r_{XZ} r_{YZ} \right) = (r_{XZ} \ r_{YZ}) C_{XY}^{-1} \begin{pmatrix} r_{XZ} \\ r_{YZ} \end{pmatrix}$$

formule que l'on peut écrire directement en fonction des **données centrées réduites** :

$$R_{Z|XY}^2 = \frac{1}{1 - ({}^t V_{XY} D^{\frac{1}{n}} V_{XY})^{-1} ({}^t V_{XY} D^{\frac{1}{n}} V_Z)^2} ({}^t V_{XY} D^{\frac{1}{n}} V_Z)^2$$

Remarquons, à l'usage des débutants, qu'il ne faudrait pas écrire :

$$\left({}^t V_{XY} D^{\frac{1}{n}} V_{XY} \right)^{-1} = V_{XY}^{-1} D^{\frac{1}{n}^{-1}} V_{XY}^{-1}$$

puisque la matrice V_{XY} , à n lignes et 2 colonnes, n'est pas inversible, alors que la matrice produit $C = {}^t V_{XY} D^{\frac{1}{n}} V_{XY}$, à 2 lignes et 2 colonnes, est inversible.

II.2.3. Application : technique de la régression pas à pas.

Pour connaître le rôle de chacune des variables explicatives, on calcule les coefficients de détermination r_{XZ}^2 et r_{YZ}^2 et le coefficient $R_{Z|XY}^2$.

Chacun de ces coefficients représente le pourcentage de variance de Z restitué par le prédicteur correspondant.

On conservera, pour prédicteur de Z le modèle qui restituera significativement le meilleur résultat :

$$\begin{aligned}\hat{Z}_0 &= c X_0 \\ \hat{Z}_0 &= d Y_0 \\ \hat{Z}_0 &= a X_0 + b Y_0.\end{aligned}$$

La théorie de la régression multiple que nous venons d'exposer dans le cas de deux variables explicatives peut se généraliser au cas de p variables explicatives, avec $p > 2$.

III. 1. ESPACES PROBABILISABLES.

III.1.1. Expériences et événements aléatoires.

III.1.1.1. Expérience déterministe.

Si nous mesurons la différence de potentiel U qui existe aux bornes d'un circuit résistif de résistance $R = 10$ ohms et dans lequel circule un courant d'intensité $i = 2$ ampères, le voltmètre indiquera une différence de potentiel $U = R i = 20$ volts.

Le résultat de cette expérience est parfaitement déterminé par la connaissance des paramètres : on dit que l'expérience est déterministe.

III.1.1.2. Expérience aléatoire.

Si nous lançons un dé dont les faces sont numérotées de 1 à 6, l'expérience consiste à noter le numéro de la face supérieure.

Le résultat de l'expérience n'est pas connu a priori : nous disons alors que l'expérience est aléatoire.

L'ensemble Ω de tous les résultats possibles pour une expérience aléatoire E donnée s'appelle le **référéntiel**, ou **ensemble fondamental**, de l'expérience aléatoire E .

Un élément de l'ensemble fondamental s'appelle un **résultat élémentaire**.

Nous supposons, dans un premier temps, que l'ensemble Ω des résultats élémentaires est un **ensemble fini**.

Dans l'expérience aléatoire du jet de dé, l'ensemble Ω est l'ensemble $\{1, 2, 3, 4, 5, 6\}$.

Les résultats élémentaires sont les nombres entiers 1, 2, 3, 4, 5, 6.

III.1.1.3. Algèbre d'événements.

Etant donnée une expérience aléatoire E d'ensemble fondamental Ω , nous appelons **événement aléatoire**, ou **événement**, toute partie de Ω .

Un événement est donc un élément de l'ensemble $\mathcal{P}(\Omega)$ des parties de Ω .

a) Définition.

Soit \mathcal{A} un ensemble d'événements.

\mathcal{A} est une partie de l'ensemble $\mathcal{P}(\Omega)$ des parties de l'ensemble fondamental Ω de l'expérience aléatoire E .

Nous dirons que \mathcal{A} est une **algèbre d'événements** de l'expérience aléatoire E si \mathcal{A} vérifie les propriétés suivantes :

- 1.- $\Omega \in \mathcal{A}$.
- 2.- $A \in \mathcal{A} \Rightarrow \bar{A} \in \mathcal{A}$ (\bar{A} désigne le complémentaire $\complement_{\Omega} A$ de A dans Ω).
- 3.- $A \in \mathcal{A}$ et $B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$.

On appelle **espace probabilisable** tout couple (Ω, \mathcal{A}) dans lequel

- Ω est l'ensemble fondamental d'une expérience aléatoire E ,
- \mathcal{A} une algèbre d'événements de l'expérience aléatoire E .

b) Propriétés d'une algèbre d'événements.

Si \mathcal{A} est une algèbre d'événements d'une expérience aléatoire E d'ensemble fondamental Ω , \mathcal{A} possède les propriétés suivantes.

A1.— $\emptyset \in \mathcal{A}$.

En effet, d'après la propriété 2 de la définition, le complémentaire d'un élément de \mathcal{A} est un élément de \mathcal{A} .

D'après la propriété 1 de la définition, Ω est un élément de \mathcal{A} .

Par la règle dite règle de *modus ponens*, il résulte de ce qui précède que le complémentaire de Ω est un élément de \mathcal{A} .

Or le complémentaire de Ω dans Ω est la partie vide \emptyset .
Donc la partie vide \emptyset est un élément de A .

A2.— Toute réunion finie d'éléments de A est un élément de A .

C'est vrai déjà, d'après la propriété précédente A1, pour la réunion de 0 élément de A , qui est vide.

Supposons, hypothèse de récurrence, que la réunion de toute famille de n éléments de A soit un élément de A , pour un entier $n \geq 0$.

Considérons une famille $(A_i)_{i \in \{1, \dots, n+1\}}$ de $n+1$ éléments de A .

D'après l'hypothèse de récurrence, la réunion $A_1 \cup \dots \cup A_n$ est un élément de A (cette réunion est vide si $n = 0$).

D'après la propriété 3 de la définition, comme A_{n+1} et $A_1 \cup \dots \cup A_n$ sont des éléments de A , leur réunion $A_1 \cup \dots \cup A_n \cup A_{n+1}$ est aussi un élément de A .

La propriété se trouve donc vraie pour $n+1$ dès qu'elle est vraie pour n .

D'après le principe de récurrence, la propriété est vraie pour tout entier $n \geq 0$.

A3.— Toute intersection finie d'éléments de A est un élément de A .

$$\bigcap_{i=1}^{i=n} A_i = \overline{\bigcup_{i=1}^{i=n} \overline{A_i}}$$

Si les A_i sont des éléments de A , les complémentaires $\overline{A_i}$ sont des éléments de A , d'après la propriété 2 de la définition.

$\bigcup_{i=1}^{i=n} \overline{A_i}$ est un élément de A , d'après la propriété précédente A2.

Le complémentaire $\overline{\bigcup_{i=1}^{i=n} \overline{A_i}} = \bigcap_{i=1}^{i=n} A_i$ de cet élément $\bigcup_{i=1}^{i=n} \overline{A_i}$ de A est un élément de A d'après la propriété 2 de la définition.

A4.— Evénements élémentaires.

Dans une algèbre finie d'événements A , les éléments minimaux pour l'inclusion, dans l'ensemble A^* des événements non vides, sont appelés les **événements élémentaires** de A .

Propriété.

- Tout événement est réunion des événements élémentaires qu'il contient.
- Les événements élémentaires contenus dans un événement non vide forment une partition de cet événement.

Démonstration.

1°/ Considérons un événement $A \in A$.

Soit $(A_i)_{i \in I}$ la famille des événements élémentaires de A contenus dans A .

Soit $B = \bigcup_{i \in I} A_i$ la réunion de cette famille.

B est un élément de A (propriété A2), contenu dans A .

Soit x un élément de Ω , appartenant à A et n'appartenant pas à B .

La famille des événements de A contenant x n'est pas vide, elle contient au moins A .

Soit X l'intersection de cette famille.

X est un élément de \mathbb{A} (propriété A3), contenu dans A et non vide, puisqu'il contient x .

X est un élément minimal pour l'inclusion dans \mathbb{A}^* .

En effet, si X contenait strictement un autre événement non vide Y , Y ne contiendrait pas x , puisque X est contenu dans tout événement contenant x .

$X \cap \bar{Y}$ serait alors un élément de \mathbb{A} contenant x , donc contenant X , ce qui ne peut pas être, sinon X serait contenu dans \bar{Y} , et serait donc sans point commun avec Y , contredisant le fait que Y est non vide et contenu dans X .

Mais si X est un élément minimal, donc un événement élémentaire, contenu dans A , c'est l'un des A_i .

L'élément x qui appartient à X appartient ainsi à l'un des A_i , donc à leur réunion B .

Ceci contredit notre hypothèse de choix de x , élément de Ω , appartenant à A et n'appartenant pas à B .

Il n'existe donc aucun x appartenant à A et n'appartenant pas à B : c'est que A est égal à B .

Et A est la réunion des événements élémentaires qu'il contient.

2°/ Si A n'est pas vide, il contient au moins un élément x .

Comme précédemment, l'intersection X de la famille des événements contenant x est un événement élémentaire contenant x et contenu dans A .

Donc la famille des événements élémentaires contenus dans A n'est pas vide.

A chaque élément x de A , on peut ainsi faire correspondre un événement élémentaire contenant x et contenu dans A .

Comme A est fini, la famille F des événements élémentaires ainsi définie est finie.

L'intersection de deux événements élémentaires différents est vide, sinon on pourrait trouver un événement non vide strictement plus petit que les deux autres, ce qui n'est pas possible, s'agissant d'éléments minimaux dans \mathbb{A}^* .

On peut alors extraire de la famille F une famille d'événements élémentaires distincts, et tout événement élémentaire contenu dans A est un élément de cette famille extraite.

D'après ce qui précède, A est réunion de cette famille : celle-ci forme donc une partition de A .

En particulier, Ω est réunion des événements élémentaires de \mathbb{A} , puisque c'est un événement de \mathbb{A} qui contient tous les événements élémentaires de \mathbb{A} .

c) Exemples.

1.— Algèbre grossière d'événements.

$\mathbb{A} = \{\Omega, \emptyset\}$ est une algèbre d'événements, appelée l'algèbre grossière des événements.

En effet, les trois axiomes de la définition sont trivialement vérifiés, compte tenu des relations :

$$\Omega \cup \emptyset = \Omega, \Omega \cup \Omega = \Omega, \emptyset \cup \emptyset = \emptyset, \overline{\Omega} = \emptyset, \overline{\emptyset} = \Omega.$$

Il y a un seul événement élémentaire, Ω .

2.— Algèbre de Bernoulli d'événements.

Si A est un événement non vide, $\mathbb{A} = \{\Omega, A, \bar{A}, \emptyset\}$ est une algèbre d'événements qu'on appelle une algèbre de Bernoulli d'événements.

En effet, Ω est un élément de \mathbb{A} , par définition.

La réunion de deux éléments de \mathbb{A} est un élément de \mathbb{A} :

$$\Omega \cup \Omega = \Omega \cup A = A \cup \Omega = \Omega \cup \bar{A} = \bar{A} \cup \Omega = \Omega \cup \emptyset = \emptyset \cup \Omega = A \cup \bar{A} = \bar{A} \cup A = \Omega \in \mathbb{A}.$$

$$A \cup A = A \cup \emptyset = \emptyset \cup A = A \in \mathbb{A}.$$

$$\bar{A} \cup \bar{A} = \bar{A} \cup \emptyset = \emptyset \cup \bar{A} = \bar{A} \in \mathbb{A}.$$

$$\emptyset \cup \emptyset = \emptyset \in \mathbb{A}.$$

Le complémentaire d'un élément de \mathbb{A} est un élément de \mathbb{A} .

$$\overline{\Omega} = \emptyset \in \mathbb{A}.$$

$$\bar{A} \in \mathbb{A}.$$

$$\overline{\bar{A}} = A \in \mathbb{A}.$$

$$\overline{\emptyset} = \Omega \in \mathbb{A}.$$

Les événements élémentaires sont A et \bar{A} .

Toute algèbre d'événements possédant l'événement A comme élément contient cette algèbre de Bernoulli d'événements.

Cette algèbre de Bernoulli d'événements contenant A est donc la plus petite algèbre d'événements contenant A .

On l'appelle l'algèbre d'événements engendrée par A .

3.— Ensemble des parties de l'ensemble fondamental.

L'ensemble $\mathfrak{P}(\Omega)$ des parties de Ω est une algèbre d'événements.

En effet, Ω est une partie de Ω (appelée la partie pleine).

La réunion de deux parties de Ω est une partie de Ω .

Le complémentaire d'une partie de Ω est une partie de Ω .

Les événements élémentaires sont les parties de Ω réduites à un seul élément de Ω .

Exemple : dans l'expérience aléatoire du jet de dé, les événements élémentaires sont $\{1\}$, $\{2\}$, $\{3\}$, $\{4\}$, $\{5\}$, $\{6\}$.

Tout événement est réunion des événements élémentaires qu'il contient.

Exemple : dans l'expérience aléatoire du jet de dé, l'événement "le résultat du lancer est un nombre pair" est $\{2, 4, 6\} = \{2\} \cup \{4\} \cup \{6\}$.

d) Algèbre d'événements engendrée par une partition.

Une partition finie $(A_i)_{i \in \{1, \dots, n\}}$ de Ω ne peut constituer, en elle-même, une algèbre d'événements.

En effet, la réunion de deux éléments distincts d'une partition n'est jamais un élément de la partition, puisque deux éléments distincts d'une partition sont sans point commun. Par ailleurs, la partie vide est un élément de toute algèbre d'événements et aucun élément d'une partition n'est vide.

Par contre, **les réunions de famille sont des événements** et les réunions $\bigcup_{i \in I} A_i$ des familles d'éléments de la partition, pour tous les $I \in \mathcal{P}([1, n])$, forment une algèbre d'événements, qu'on appelle **l'algèbre d'événements engendrée** par la partition.

Les événements élémentaires de l'algèbre engendrée par la partition $(A_i)_{i \in \{1, \dots, n\}}$ sont constitués des événements $(A_i)_{i \in \{1, \dots, n\}}$ de la partition.

Toute algèbre d'événements est engendrée par ses événements élémentaires : en effet, tous les événements de l'algèbre d'événements sont réunions des événements élémentaires qu'ils contiennent.

Exemple :

Etant donnés deux événements A et B , les événements $A \cap B, A \cap \bar{B}, \bar{A} \cap B, \bar{A} \cap \bar{B}$, s'ils ne sont pas vides, forment une partition de Ω et engendrent une algèbre d'événements, formée des réunions des $2^4 = 16$ familles possibles de ces événements, et dont les événements élémentaires sont $A \cap B, A \cap \bar{B}, \bar{A} \cap B, \bar{A} \cap \bar{B}$.

III. 2. PROBABILITES.

III.2.1. Définition.

Soit (Ω, \mathcal{A}) un espace probabilisable dont l'ensemble fondamental Ω est fini, de cardinal N . On appelle **probabilité** sur cet espace probabilisable, toute application P de \mathcal{A} dans l'intervalle $[0, 1]$ de \mathbb{R} , qui vérifie les deux propriétés suivantes :

P1.— $P(\Omega) = 1$

P2.— $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$

La propriété P2 s'appelle l'**axiome des probabilités totales**.

Lorsque deux événements A et B ont une intersection vide, on dit qu'ils sont **incompatibles**.

Par une récurrence immédiate, on voit que l'axiome des probabilités totales peut se généraliser à n'importe quelle famille finie $(A_i)_{i \in \{1, \dots, n\}}$ d'événements incompatibles deux à deux, c'est-à-dire vérifiant l'axiome : $(\forall i)(\forall j)(i \in [1, n] \text{ et } j \in [1, n] \text{ et } i \neq j \Rightarrow A_i \cap A_j = \emptyset)$.

$$P\left(\bigcup_{i=1}^{i=n} A_i\right) = \sum_{i=1}^{i=n} P(A_i)$$

On appelle **espace probabilisé** tout triplet (Ω, \mathcal{A}, P) dans lequel (Ω, \mathcal{A}) est un espace probabilisable et P une probabilité sur cet espace probabilisable.

III.2.2. Propriétés.

III.2.2.1. Événements contraires.

Deux événements sont dits **contraires** s'ils sont complémentaires l'un de l'autre.
Pour tout événement A , A et \bar{A} sont des événements contraires.

$$A \cap \bar{A} = \emptyset \Rightarrow P(A \cup \bar{A}) = P(A) + P(\bar{A})$$

$$A \cup \bar{A} = \Omega \Rightarrow P(\Omega) = P(A) + P(\bar{A})$$

$$P(\Omega) = 1 \Rightarrow P(A) + P(\bar{A}) = 1$$

$$P(A) + P(\bar{A}) = 1$$

III.2.2.2. Événement impossible.

Un événement est dit **impossible** si sa probabilité est nulle.
La partie vide est un événement impossible.

La propriété précédente donne, pour $A = \Omega$, $P(\Omega) + P(\emptyset) = 1$.

$$P(\Omega) = 1 \Rightarrow P(\emptyset) = 0.$$

$$P(\emptyset) = 0$$

III.2.2.3. Croissance.

Si l'événement A est contenu dans l'événement B , alors $P(A)$ est inférieure ou égale à $P(B)$.

$$A \subset B \Rightarrow P(A) \leq P(B)$$

En effet :

$$A \subset B \Leftrightarrow A \cap B = A$$

$$A \cap B = A \Rightarrow B = (B \cap A) \cup (B \cap \bar{A}) = A \cup (B \cap \bar{A}),$$

$$A \cap (B \cap \bar{A}) = \emptyset \Rightarrow P(B) = P(A \cap (B \cap \bar{A})) = P(A) + P(B \cap \bar{A})$$

$$P(B \cap \bar{A}) \in [0, 1] \text{ et } P(B) = P(A) + P(B \cap \bar{A}) \Rightarrow P(B) \geq P(A)$$

En définitive, on obtient bien : $A \subset B \Rightarrow P(A) \leq P(B)$.

Toute probabilité est une application croissante.

III.2.2.4. Probabilité d'une réunion.

$$A \cup B = A \cup (B \cap \bar{A})$$

$$A \cap (B \cap \bar{A}) = \emptyset \Rightarrow P(A \cup B) = P(A \cup (B \cap \bar{A})) = P(A) + P(B \cap \bar{A})$$

$$P(B \cap \bar{A}) = P(A \cup B) - P(A)$$

$$B = (B \cap A) \cup (B \cap \bar{A})$$

$$(B \cap A) \cap (B \cap \bar{A}) = \emptyset \Rightarrow P(B) = P((B \cap A) \cup (B \cap \bar{A})) = P(B \cap A) + P(B \cap \bar{A})$$

$$P(B \cap \bar{A}) = P(B) - P(B \cap A) = P(B) - P(A \cap B)$$

$$P(A \cup B) - P(A) = P(B) - P(A \cap B)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Généralisation : formule de Poincaré.

La formule précédente se généralise à la réunion d'une famille $(A_i)_{i \in \{1, \dots, n\}}$ de n événements, éléments de \mathcal{A} .

On intersecte les événements distincts par 1, par 2, par 3, etc., en alternant les signes :

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) + \dots + (-1)^{n+1} P(A_1 \cap \dots \cap A_n)$$

Cette formule est appelée la **formule de Poincaré**.

III.2.2.5. Événements élémentaires équiprobables dans $\mathcal{P}(\Omega)$.

Soit Ω l'ensemble fondamental, supposé fini de cardinal N , associé à une épreuve aléatoire E , $\mathcal{P}(\Omega)$ l'algèbre d'événements des parties de Ω .

Une probabilité P sur l'espace probabilisable $(\Omega, \mathcal{P}(\Omega))$ peut se définir en associant à chaque événement élémentaire $\{\omega_i\}$, $\omega_i \in \Omega$, un nombre réel p_i , avec :

$$p_i = P(\{\omega_i\}) \text{ (qu'on note aussi } P(\omega_i)\text{)}.$$

- $(\forall i)(i \in [1, N] \subset \mathbb{N} \Rightarrow p_i \in [0, 1] \subset \mathbb{R})$ [les p_i sont des nombres réels compris entre 0 et 1].

- $\sum_{i=1}^N p_i = \sum_{i=1}^N P(\{\omega_i\}) = P\left(\bigcup_{i=1}^N \{\omega_i\}\right) = P(\Omega) = 1$ [la somme des p_i est égale à 1].

Comme chaque événement est réunion des événements élémentaires qu'il contient, la probabilité d'un événement A est donnée par :

$$P(A) = P\left(\bigcup_{\omega \in \mathcal{A}} \{\omega\}\right) = \sum_{\omega \in \mathcal{A}} P(\{\omega\})$$

Un cas particulier important est celui où tous les événements élémentaires ont la même probabilité p .

La relation $\sum_{i=1}^N p_i = 1$ s'écrit $Np = 1$, soit $p = \frac{1}{N}$.

La probabilité d'un événement A qui est réalisé dans n événements élémentaires est alors donnée par :

$$P(A) = \sum_{\omega \in \mathcal{A}} P(\{\omega\}) = \text{Card}(A) \times \frac{1}{N} = \frac{n}{N}$$

La probabilité d'un événement A est le rapport entre le nombre de cas favorables n et le nombre de cas possibles N .

$$P(A) = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}}$$

III.3. PROBABILITES CONDITIONNELLES. INDEPENDANCE.

III.3.1. Définition.

Soit (Ω, \mathcal{A}, P) un espace probabilisé (Ω fini).

Soit $B \in \mathcal{A}$, un événement de probabilité différente de 0 : $P(B) \neq 0$.

Soit $A \in \mathcal{A}$, un événement.

On appelle **probabilité conditionnelle de A sachant que B est réalisé**, ou **probabilité conditionnée de A par B**, le nombre réel :

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

L'application $A \mapsto P(A|B)$ de \mathcal{A} dans \mathbb{R} définit une probabilité sur (Ω, \mathcal{A}) .

En effet :

a) Comme P est une probabilité, ses valeurs sont positives et comprises entre 0 et 1, donc, pour tout $A \in \mathcal{A}$, $P(A|B)$ est positif.

b) Comme $A \cap B$ est contenu dans B et que l'application P est croissante, $P(A \cap B)$ est inférieur

ou égal à $P(B)$, donc le rapport $\frac{P(A \cap B)}{P(B)}$ est inférieur ou égal à 1.

c) $P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$.

d) Soient A_1 et A_2 deux événements incompatibles ($A_1 \cap A_2 = \emptyset$).

$$\begin{aligned} P((A_1 \cup A_2)|B) &= \frac{P((A_1 \cup A_2) \cap B)}{P(B)} = \frac{P((A_1 \cap B) \cup (A_2 \cap B))}{P(B)} \\ (A_1 \cap B) \cap (A_2 \cap B) &\subset A_1 \cap A_2 = \emptyset \Rightarrow (A_1 \cap B) \cap (A_2 \cap B) = \emptyset \\ (A_1 \cap B) \cap (A_2 \cap B) = \emptyset &\Rightarrow P((A_1 \cap B) \cup (A_2 \cap B)) = P(A_1 \cap B) + P(A_2 \cap B) \\ P((A_1 \cup A_2)|B) &= \frac{P(A_1 \cap B)}{P(B)} + \frac{P(A_2 \cap B)}{P(B)} \\ P((A_1 \cup A_2)|B) &= P(A_1|B) + P(A_2|B). \end{aligned}$$

III.3.2. Indépendance.

III.3.2.1. Définition.

La relation $P(A|B) = \frac{P(A \cap B)}{P(B)}$ peut être écrite :

$$P(A \cap B) = P(A|B) P(B)$$

Si la probabilité de A n'est pas nulle, on peut écrire aussi :

$$P(A \cap B) = P(B|A) P(A)$$

Deux événements A et B sont dits **indépendants en probabilité**, ou, simplement, **indépendants**, s'ils vérifient la relation :

$$P(A \mid B) = P(A) P(B)$$

Pour des événements de probabilité non nulle, les propriétés suivantes sont équivalentes :

- A et B sont indépendants.
- $P(A \mid B) = P(A) P(B)$.
- $P(A \mid B) = P(A)$.
- $P(B \mid A) = P(B)$.

L'information donnée par la réalisation de A sur la réalisation de B est nulle et l'information donnée par la réalisation de B sur la réalisation de A est nulle.

Extension de la définition d'une probabilité conditionnelle.

La définition de la probabilité conditionnelle de A lorsque B est réalisé, peut s'étendre au cas où la probabilité de B est nulle.

Si $P(B) = 0$, B est indépendant de tout événement A , et tout événement A est indépendant de B , puisque $P(A \mid B) = P(A) P(B)$.

Nous dirons, dans ce cas, que $P(A \mid B) = P(A)$, par définition.

D'où la nouvelle définition :

Etant donnés deux événements A et B , on appelle **probabilité conditionnelle de A sachant que B est réalisé**, ou **probabilité conditionnée de A par B** , le nombre réel :

$$P(A \mid B) = \begin{cases} \frac{P(A \cap B)}{P(B)} & \text{lorsque } P(B) \neq 0 \\ P(A) & \text{lorsque } P(B) = 0 \end{cases}$$

Cette extension de la définition permet d'écrire la formule :

$$P(A \mid B) = P(B) P(A \mid B) = P(A) P(B \mid A)$$

dans tous les cas.

Remarque.

Il convient de ne pas confondre **indépendance et incompatibilité**.

L'indépendance est une question de probabilité : si on change la probabilité, deux événements précédemment indépendants peuvent ne plus être indépendants.

L'incompatibilité ne fait pas intervenir la probabilité : une intersection vide reste vide quelque soit la probabilité.

III.3.2.2. Généralisation.

La formule

$$P(A \mid B) = P(A) P(B \mid A)$$

peut être généralisée.

$$P(A \mid B \mid C) = P(A) P(B \mid A) P(C \mid A \mid B)$$

En effet, on peut écrire successivement :

$$P(A \mid B \mid C) = P(C \mid A \mid B) = P(A \mid B) P(C \mid A \mid B) = P(A) P(B \mid A) P(C \mid A \mid B).$$

Nous dirons que les trois événements A , B , C sont indépendants dans leur ensemble si, et seulement si, les quatre relations suivantes sont vérifiées :

$$\begin{aligned}
P(A \cap B \cap C) &= P(A) P(B) P(C) \\
P(A \cap B) &= P(A) P(B) \\
P(B \cap C) &= P(B) P(C) \\
P(C \cap A) &= P(C) P(A)
\end{aligned}$$

Plus généralement, nous dirons que n événements $(A_i)_{i \in \{1, \dots, n\}}$ sont **indépendants dans leur ensemble** si, et seulement si, pour toute partie I de l'ensemble $\{1, \dots, n\}$, la relation suivante est vérifiée :

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i)$$

Dans le cas où I est vide, l'intersection d'une famille vide est vide, et il faut prendre 0 pour produit d'une famille vide.

III.3.3. Formule de Bayes.

III.3.3.1. La formule sous sa forme générale.

Dans un espace probabilisé (Ω, \mathcal{A}, P) , on appelle **système complet d'événements** toute partition finie de Ω par des éléments de \mathcal{A} .

Soit $(B_i)_{i \in \{1, \dots, n\}}$ un système complet d'événements.

Alors, pour tout événement A de probabilité non nulle, et pour tout indice $i \in \{1, \dots, n\}$, on a la **formule de Bayes** :

$$P(B_i | A) = \frac{P(A|B_i) P(B_i)}{P(A|B_1) P(B_1) + \dots + P(A|B_n) P(B_n)}$$

Démonstration.

Comme $(B_i)_{i \in \{1, \dots, n\}}$ est une partition de Ω , on a $\Omega = \bigcup_{i=1}^n B_i$ et :

$$A = A \cap \Omega = A \cap \left(\bigcup_{i=1}^n B_i\right) = \bigcup_{i=1}^n (A \cap B_i)$$

Comme les B_i sont deux à deux incompatibles, les $A \cap B_i$, qui sont contenus dans les B_i d'indices correspondants, sont eux aussi, deux à deux incompatibles.

La probabilité de leur réunion est donc la somme de leurs probabilités :

$$P(A) = \sum_{i=1}^n P(A \cap B_i) = \sum_{i=1}^n P(A | B_i) P(B_i) = P(A | B_1) P(B_1) + \dots + P(A | B_n) P(B_n)$$

La formule $P(A \cap B_i) = P(A | B_i) P(B_i) = P(B_i | A) P(A)$ s'écrit alors :

$$P(B_i | A) [P(A | B_1) P(B_1) + \dots + P(A | B_n) P(B_n)] = P(A | B_i) P(B_i)$$

et si la probabilité de A n'est pas nulle, c'est-à-dire si l'un au moins des $P(A | B_i) P(B_i)$ n'est pas nul, la formule de Bayes s'en déduit.

III.3.3.2. Cas particulier.

Pour tout événement non vide B , B et \bar{B} forment une partition de Ω .

Dans ce cas, la formule de Bayes se réduit à :

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|\bar{B})P(\bar{B})}$$

III.3.3.3. Exemple.

On prend un dé au hasard parmi un lot de 100 dés dont on sait que 25 d'entre eux sont pipés. Pour un dé pipé, la probabilité d'obtenir un 6 est 0,5. On lance le dé choisi, on obtient 6 : quelle est la probabilité pour que le dé choisi soit pipé ?

Appelons T l'événement "le dé est pipé", S l'événement "on obtient un 6".
 T et \bar{T} forment un système complet d'événements.
 La formule de Bayes donne :

$$P(T|S) = \frac{P(S|T)P(T)}{P(S|T)P(T) + P(S|\bar{T})P(\bar{T})} = \frac{\frac{1}{2} \times \frac{1}{4}}{\frac{1}{2} \times \frac{1}{4} + \frac{1}{6} \times \frac{3}{4}} = \frac{1}{2}.$$

Remarque.

La formule de Bayes s'applique dans le cas d'une expérience aléatoire où le hasard intervient à deux reprises : une première fois dans le choix du dé, une deuxième fois dans le résultat du lancer.

Un résultat lié au premier niveau de hasard est appelé une "**cause**".

Un résultat lié au deuxième niveau de hasard est appelé une "**conséquence**".

Dans le cas présent, on a calculé la probabilité que le dé soit pipé (la cause), sachant que nous avons obtenu les 6 (la conséquence).

C'est pourquoi la formule de Bayes est aussi appelée la **formule de probabilité des causes**.

Les données du problème peuvent être présentées sous forme d'un tableau des probabilités, compte tenu de la formule $P(A \cap B) = P(B)P(A|B)$:

	Dé pipé = T	Dé non pipé = \bar{T}	Total
6 = S	$P(S T) = P(T)P(S T) = \frac{1}{4} \times \frac{1}{2}$	$P(S \bar{T}) = P(\bar{T}) \times \frac{1}{6}$	$P(S)$
non 6 = \bar{S}	$P(\bar{S} T)$	$P(\bar{S} \bar{T})$	$P(\bar{S})$
Total	$P(T) = \frac{1}{4}$	$P(\bar{T})$	$P(\Omega)$

Ces données suffisent pour remplir le tableau, par somme ou différence :

	T	\bar{T}	Total
S	$\frac{1}{8}$	$\frac{3}{4} \times \frac{1}{6} = \frac{1}{8}$	$\frac{1}{8} + \frac{1}{8} = \frac{1}{4}$
\bar{S}	$\frac{1}{4} - \frac{1}{8} = \frac{1}{8}$	$\frac{3}{4} - \frac{1}{8} = \frac{5}{8}$	$1 - \frac{1}{4} = \frac{3}{4}$
Total	$\frac{1}{4}$	$1 - \frac{1}{4} = \frac{3}{4}$	1

	S	\bar{S}	Total
T	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$
\bar{T}	$\frac{1}{8}$	$\frac{5}{8}$	$\frac{3}{4}$
Total	$\frac{1}{4}$	$\frac{3}{4}$	1

Ce tableau de probabilités conjointes remplace la formule de Bayes et permet de répondre à toutes les questions qu'on se pose sur les probabilités conditionnelles : pour calculer $P(T|S)$, on regarde dans la ligne de S , combien il y a de $T|S$.

	T	\bar{T}	Total
S	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$
\bar{S}	$\frac{1}{8}$	$\frac{5}{8}$	$\frac{3}{4}$
Total	$\frac{1}{4}$	$\frac{3}{4}$	1

$$P(T|S) = \frac{P(T \cap S)}{P(S)} = \frac{\frac{1}{8}}{\frac{1}{4}} = \frac{1}{2}$$

III. 4. VARIABLES ALEATOIRES DISCRETES.

III.4.1. Définition.

Soit (Ω, \mathcal{A}) un espace probabilisable associé à une épreuve aléatoire E .

Soit X une application de Ω dans \mathbb{R} .

Nous dirons que X est une **variable aléatoire réelle** liée à E si, et seulement si, pour tout réel x , l'image réciproque $X^{-1}]-\infty, x]$ est un élément de \mathcal{A} .

Si $X(\Omega)$ est une partie dénombrable de \mathbb{R} , nous dirons que X est une **variable aléatoire discrète**.

Une variable aléatoire discrète sera dite **finie** ou **infinie** suivant que l'ensemble de ses valeurs est fini ou infini.

Réciproquement, considérons l'ensemble fondamental Ω associé à une épreuve aléatoire E .

Soit \mathbb{R} l'ensemble des nombres réels.

Considérons une application $X : \Omega \rightarrow \mathbb{R}$.

On note $X(\Omega)$ l'ensemble des valeurs de X , qu'on suppose dénombrable.

A toute partie B de \mathbb{R} , on peut associer la partie $X^{-1}(B)$ de Ω , qui est l'ensemble, éventuellement vide, des $\omega \in \Omega$ pour lesquels $X(\omega)$ est un élément de B .

L'ensemble des parties $X^{-1}(B)$ de Ω , pour les parties B de \mathbb{R} , forme une **algèbre d'événements** \mathcal{A} .

En effet :

1°/ $X^{-1}(\mathbb{R}) = \Omega$, donc $\Omega \in \mathcal{A}$.

2°/ Soit $A \in \mathcal{A}$. Il existe une partie B de \mathbb{R} telle que $A = X^{-1}(B)$.

$\omega \in \bar{A} \Leftrightarrow \omega \notin A \Leftrightarrow X(\omega) \notin B \Leftrightarrow X(\omega) \in \bar{B} \Leftrightarrow \omega \in X^{-1}(\bar{B})$.

La relation $\bar{A} = X^{-1}(\bar{B})$ entraîne $\bar{A} \in \mathcal{A}$.

3°/ Soient A et A' des éléments de \mathcal{A} . Il existe des parties B et B' de \mathbb{R} telles que l'on ait $A = X^{-1}(B)$ et $A' = X^{-1}(B')$.

$\omega \in X^{-1}(B \cup B') \Leftrightarrow X(\omega) \in B \cup B' \Leftrightarrow X(\omega) \in B$ ou $X(\omega) \in B' \Leftrightarrow \omega \in X^{-1}(B)$ ou $\omega \in X^{-1}(B')$

$$(B') \Leftrightarrow \omega \in X^{-1}(B \cup B')$$

La relation $X^{-1}(B \cup B') = X^{-1}(B) \cup X^{-1}(B')$ entraîne $A \cup A' \in \mathcal{A}$.

Comme l'application $B \mapsto X^{-1}(B)$ est croissante, les **événements élémentaires** de cette algèbre d'événements sont les images réciproques des valeurs x_k de X .

Toute application X définie sur Ω et à valeurs dans \mathbb{R} permet donc de définir un espace probabilisable (Ω, \mathcal{A}) , sur lequel X est une variable aléatoire réelle.

III.4.2. Exemple.

Soit Ω l'ensemble fondamental associé à une épreuve aléatoire E .

Soit A une partie de Ω .

On appelle **variable indicatrice de Bernoulli** de A , la variable aléatoire 1_A définie, pour tout élément ω de Ω , par :

$$1_A(\omega) = \begin{cases} 1 & \text{lorsque } \omega \in A \\ 0 & \text{lorsque } \omega \notin A \end{cases}$$

Cette variable aléatoire n'a que deux valeurs : 0 et 1.

Les événements élémentaires de l'algèbre d'événements qu'elle définit sont :

$$1_A^{-1}(1) = A$$

$$1_A^{-1}(0) = \bar{A}$$

L'algèbre d'événements définie par la variable indicatrice de Bernoulli de A est donc l'algèbre de Bernoulli \mathcal{A} formée des éléments $\{\emptyset, A, \bar{A}, \Omega\}$ de $\mathcal{P}(\Omega)$ et l'espace probabilisable associé est (Ω, \mathcal{A}) .

III.4.3. Loi de probabilité d'une variable aléatoire.

III.4.3.1. Définition.

Soit (Ω, \mathcal{A}, P) un espace probabilisé, et X une variable aléatoire discrète.

Soit $(x_k)_{k \in I}$ l'ensemble des valeurs de X .

Soit A_k l'événement élémentaire de \mathcal{A} défini par la valeur $x_k : A_k = \{\omega \in \Omega \mid X(\omega) = x_k\} = X^{-1}(\{x_k\}) \in \mathcal{A}$.

On note P_X l'application $\mathbb{R} \rightarrow [0, 1]$ définie par :

$$P_X(x) = P(X^{-1}(\{x\}))$$

Comme les valeurs de X sont les $(x_k)_{k \in I}$, $X^{-1}(\{x\})$ est vide lorsque x n'est pas l'un des $(x_k)_{k \in I}$ et sa probabilité est nulle.

Lorsque $x = x_k$, pour un $k \in I$, on obtient :

$$P_X(x_k) = P(X^{-1}(\{x_k\})) = P(A_k) = \sum_{\omega \in A_k} P(\omega) = \sum_{X(\omega) = x_k} P(\omega)$$

Le nombre $p_k = P_X(x_k) = \sum_{X(\omega) = x_k} P(\omega)$ est souvent noté, par abus d'écriture, $P(X = x_k)$, ou $P(x_k)$.

On dit que p_k est la **probabilité de la valeur x_k** de X .

L'ensemble des couples $(x_k, p_k)_{k \in I}$ est appelé la **loi de probabilité** de la variable aléatoire X .

III.4.3.2. Propriétés.

a) Pour tout $k \in I, p_k \in [0, 1]$.

En effet, $p_k = P(X = x_k) = P(A_k)$, et les valeurs de P sont des nombres réels compris entre 0 et 1.

b) $\sum_{k \in I} p_k = 1$.

En effet, les $A_k, k \in I$, sont les événements élémentaires de l'algèbre d'événements \mathcal{A} , donc ils sont incompatibles deux à deux et leur réunion est :

$$\bigcup_{k \in I} A_k = \bigcup_{k \in I} X^{-1}(\{x_k\}) = X^{-1}(\bigcup_{k \in I} \{x_k\}) = X^{-1}(X(\Omega)) = \Omega$$

de sorte que l'on a :

$$\sum_{k \in I} p_k = \sum_{k \in I} P(A_k) = P\left(\bigcup_{k \in I} A_k\right) = P(\Omega) = 1.$$

III.4.3.3. Exemple.

Considérons la variable indicatrice de Bernoulli $\mathbb{1}_A$ d'un événement A .

Notons p la probabilité de réalisation de l'événement A , et $q = 1 - p$.

Nous avons, par définition :

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \text{lorsque } \omega \in A \\ 0 & \text{lorsque } \omega \notin A \end{cases}$$

donc :

$$p = P(A) = P(\mathbb{1}_A^{-1}(1)) = P(\mathbb{1}_A = 1) = P^{\mathbb{1}_A}(1) = P(1).$$

$$q = 1 - P(A) = P(\bar{A}) = P(\mathbb{1}_A^{-1}(0)) = P(\mathbb{1}_A = 0) = P^{\mathbb{1}_A}(0) = P(0).$$

La loi de probabilité de la variable indicatrice $\mathbb{1}_A$ est donc donnée par le tableau :

x_k	0	1
p_k	q	p

avec $p + q = 1$.

III.4.4. Fonction de répartition.

III.4.4.1. Définition.

On appelle **fonction de répartition d'une variable aléatoire** X l'application $F : \mathbb{R} \rightarrow \mathbb{R}$ définie par :

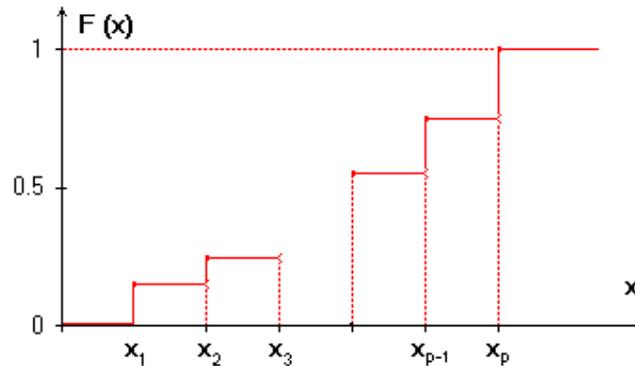
$$F(x) = P(X \leq x) = \sum_{\substack{k \in I \\ x_k \leq x}} P(X = x_k) = \sum_{\substack{k \in I \\ x_k \leq x}} p_k.$$

III.4.4.2. Propriété.

Comme les p_k sont des nombres positifs, la fonction de répartition est une application croissante, dont toutes les valeurs sont positives.

Comme la somme $\sum_{k \in I} p_k$ des p_k est égale à 1, $F(x)$ croît de 0 à 1 lorsque x varie de $-\infty$ à $+\infty$. Lorsque x atteint une valeur x_k par valeurs inférieures, $F(x)$ augmente brusquement de p_k et reste constante sur l'intervalle $[x_k, x_{k+1}[$.

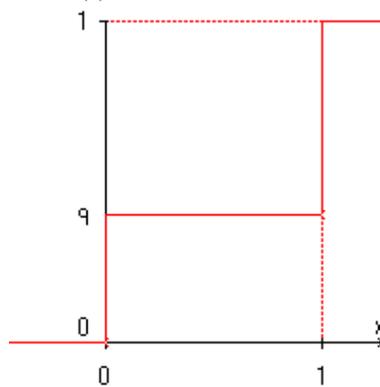
C'est une fonction en escalier, continue à droite :



III.4.4.3. Exemple.

La variable indicatrice de Bernoulli d'un événement A a pour fonction de répartition :

$$F(x) = \begin{cases} 0 & \text{lorsque } x < 0 \\ q & \text{lorsque } 0 \leq x < 1 \\ 1 & \text{lorsque } 1 \leq x \end{cases}$$



III.4.5. Espérance mathématique.

On appelle **espérance mathématique** (ou **moyenne**) d'une variable aléatoire X le nombre, quand il existe :

$$E(X) = \sum_{k \in I} p_k x_k$$

Exemple : espérance mathématique de la variable indicatrice de Bernoulli d'un événement A .

$$E(1_A) = q \times 0 + p \times 1 = p = P(A).$$

III.4.5.1. Propriétés.

a) Espérance mathématique d'une constante.

Si la variable aléatoire X n'a qu'une seule valeur $X = a$, on a $P(X = a) = 1$ et $E(X) = a P(X = a) = a$.

$$E(a) = a$$

b) Linéarité.

Soient X une variable aléatoire réelle discrète, a et b des constantes.

La variable aléatoire $Z = aX + b$ prend les valeurs $z_k = ax_k + b$ avec les probabilités :

$$P(Z = z_k) = P(X = x_k)$$

$$E(Z) = \sum_k P(X = x_k) (ax_k + b)$$

$$E(Z) = E(Z) = a \sum_k P(X = x_k) x_k + b \sum_k P(X = x_k) = a E(X) + b$$

D'où le résultat :

$$E(aX + b) = aE(X) + b$$

Plus généralement, **nous admettrons** que, pour deux variables aléatoires X et Y , et deux constantes a et b , nous avons :

$$E(aX + bY) = aE(X) + bE(Y)$$

III.4.5.2. Généralisation.

Si φ est une application de \mathbb{R} dans \mathbb{R} , $\varphi \circ X$ est une variable aléatoire qu'on note $\varphi(X)$.

L'espérance mathématique, ou moyenne, de $\varphi(X)$ est (**résultat admis**) :

$$E(\varphi(X)) = \sum_{k \in I} p_k \varphi(x_k)$$

III.4.5.3. Moments.

Pour tout entier n , l'espérance mathématique $E(X^n)$ s'appelle le **moment d'ordre n** de X , on le note m_n .

$$m_n = E(X^n) = \sum_{k \in I} p_k x_k^n$$

$$m_0 = \sum_{k \in I} p_k = 1$$

$$m_1 = \sum_{k \in I} p_k x_k = E(X)$$

$$m_2 = \sum_{k \in I} p_k x_k^2 = E(X^2)$$

III.4.5.4. Moments centrés.

Pour tout entier n , l'espérance mathématique $E((X - E(X))^n)$ s'appelle le **moment centré d'ordre n** de X , on le note μ_n .

$$\mu_n = E((X - E(X))^n) = \sum_{k \in I} p_k (x_k - E(X))^n$$

$$\mu_0 = \sum_{k \in I} p_k = 1$$

$$\mu_1 = \sum_{k \in I} p_k (x_k - E(X)) = E(X) - E(X) = 0$$

$$\mu_2 = \sum_{k \in I} p_k (x_k - E(X))^2$$

III.4.6. Variance.

Le moment centré d'ordre 2 s'appelle la **variance**. On note $Var(X)$ la variance de X .

$$Var(X) = \sum_{k \in I} p_k (x_k - E(X))^2$$

En développant le carré, on obtient :

$$\begin{aligned} Var(X) &= \sum_{k \in I} p_k (x_k^2 + (E(X))^2 - 2 x_k E(X)) \\ &= \sum_{k \in I} p_k x_k^2 + \left(\sum_{k \in I} p_k \right) (E(X))^2 - 2 \left(\sum_{k \in I} p_k x_k \right) E(X) \\ &= \sum_{k \in I} p_k x_k^2 - (E(X))^2 \end{aligned}$$

$$Var(X) = E(X^2) - (E(X))^2 = m_2 - m_1^2$$

La formule $Var(X) = E(X^2) - (E(X))^2$ est connue sous le nom de **formule de la variance**, ou **formule de Koenig**.

Elle indique que la variance est la moyenne du carré, moins le carré de la moyenne.

La racine carrée de la variance s'appelle l'**écart-type**. On $\sigma(X)$ ou σ_X l'écart-type de X .

$$\sigma_X = \sqrt{Var(X)}$$

Propriété.

$$\text{Var}(aX + b) = a^2 \text{Var}(X)$$

En effet

$$\text{Var}(aX + b) = E((aX + b - E(aX + b))^2) = E((aX + b - aE(X) - b)^2) = E(a^2(X - E(X))^2) = a^2 E((X - E(X))^2) = a^2 \text{Var}(X)$$

Exemple.

Le moment d'ordre 2 de la **variable indicatrice de Bernoulli** d'un événement A de probabilité p est :

$$m_2 = E(X^2) = q \times 0^2 + p \times 1^2 = p$$

La variance est :

$$\text{Var}(X) = m_2 - m_1^2 = p - p^2 = p(1 - p) = pq.$$

III.4.7. Fonction génératrice.

III.4.7.1. Définition.

Pour une variable aléatoire discrète X, à **valeurs entières**, on appelle **fonction génératrice** l'application g_X de \mathbb{R} dans \mathbb{R} qui, à un nombre réel u fait correspondre l'espérance mathématique de la variable aléatoire u^X :

$$g_X(u) = E(u^X) = \sum_{k \geq 0} p_k u^k = p_0 + p_1 u + p_2 u^2 + \dots$$

III.4.7.2. Propriétés.

a) Condition de normalisation.

$$g_X(1) = \sum_{k \geq 0} p_k = 1$$

b) Espérance mathématique et variance.

La dérivée de $g_X(u)$ s'obtient en dérivant terme à terme :

$$\frac{d g_X(u)}{d u} = \sum_{k \geq 1} k p_k u^{k-1} = \frac{1}{u} \sum_{k \geq 0} k p_k u^k = \frac{1}{u} E(X u^X)$$

Pour $u = 1$, cette dérivée prend la valeur :

$$\left(\frac{d g_X(u)}{d u} \right)_{u=1} = E(X)$$
$$E(X) = g'_X(1)$$

La dérivée seconde s'obtient en dérivant la dérivée première :

$$\frac{d^2 g_X(u)}{du^2} = \sum_{k \geq 2} k(k-1) p_k u^{k-2} = \frac{1}{u^2} \sum_{k \geq 0} k(k-1) p_k u^k = \frac{1}{u^2} E(X(X-1)u^X)$$

Pour $u = 1$, cette dérivée seconde prend la valeur :

$$\left(\frac{d^2 g_X(u)}{du^2} \right)_{u=1} = E(X(X-1)) = E(X^2) - E(X)$$

On en déduit :

$$E(X^2) = \left(\frac{d^2 g_X(u)}{du^2} \right)_{u=1} + E(X) = \left(\frac{d^2 g_X(u)}{du^2} \right)_{u=1} + \left(\frac{d g_X(u)}{du} \right)_{u=1} = g''_X(1) + g'_X(1)$$

$$Var(X) = E(X^2) - (E(X))^2 = g''_X(1) + g'_X(1) - (g'_X(1))^2$$

$$\boxed{\begin{aligned} Var(X) &= g''_X(1) + g'_X(1) - (g'_X(1))^2 \\ Var(X) &= E(X(X-1)) + E(X) - (E(X))^2 \end{aligned}}$$

c) Probabilités.

$P(X = k) = p_k$ est le coefficient de u^k dans le développement de $g_X(u)$ au voisinage de 0.

D'après la formule de Taylor, ce coefficient est égal à $\frac{1}{k!} \left(\frac{d^k g_X(u)}{du^k} \right)_{u=0}$

$$\boxed{P(X = k) = p_k = \frac{1}{k!} \left(\frac{d^k g_X(u)}{du^k} \right)_{u=0}}$$

III.4.7.3. Exemple.

La fonction génératrice de la **variable indicatrice de Bernoulli** d'un événement A de probabilité p est :

$$g_X(u) = E(u^X) = q + pu$$

- On a bien $g_X(1) = q + p = 1$.
- L'espérance mathématique est $E(X) = g'_X(1) = p$.
- La variance est $Var(X) = g''_X(1) + g'_X(1) - (g'_X(1))^2 = 0 + p - p^2 = p(1-p) = pq$.

III. 5. LOIS DISCRETES FINIES USUELLES.

III.5.1. Loi de Bernoulli.

III.5.1.1. Définition.

On dit qu'une variable aléatoire X est une **variable de Bernoulli de paramètre p** si sa loi de probabilité est :

x_k	0	1
p_k	q	p

avec $p + q = 1$. Cette loi est notée $B(1; p)$ et on note $X \rightsquigarrow B(1; p)$ la relation "X suit la loi de Bernoulli de paramètre p ".

III.5.1.2. Propriétés.

Si $X \rightsquigarrow B(1; p)$:

$$E(X) = p$$

$$Var(X) = p q$$

$$g_X(u) = q + p u$$

III.5.2. Loi uniforme.

III.5.2.1. Définition.

On dit qu'une variable aléatoire X suit une **loi uniforme sur $\{x_1, \dots, x_n\}$** si sa loi de probabilité est : $p_k = P(X = x_k) = \frac{1}{n}, k = 1, \dots, n$.

x_k	x_1	...	x_n
p_k	$\frac{1}{n}$...	$\frac{1}{n}$

III.5.2.2. Propriétés.

Si X suit une loi uniforme sur $\{x_1, \dots, x_n\}$:

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

$$Var(X) = \frac{1}{n} \left(\sum_{k=1}^{k=n} x_k^2 - \frac{1}{n} \left(\sum_{k=1}^{k=n} x_k \right)^2 \right)$$

III.5.2.3. Cas particulier.

Si X suit une loi uniforme sur $\{1, \dots, n\}$ (relation notée $X \rightsquigarrow U[1, n]$), les relations :

$$\sum_{k=1}^{k=n} k = \frac{n(n+1)}{2} \quad \text{et} \quad \sum_{k=1}^{k=n} k^2 = \frac{n(n+1)(2n+1)}{6}$$

entraînent :

$$E(X) = \frac{n+1}{2}$$

$$Var(X) = \frac{[n+1][2n+1]}{6} - \frac{[n+1]^2}{4} = \frac{n+1}{2} \left(\frac{2n+1}{3} - \frac{n+1}{2} \right) = \frac{n+1}{2} \times \frac{n-1}{6} = \frac{n^2-1}{12}$$

$$g_X(u) = \frac{1}{n} \sum_{k=1}^{k=n} u^k = \frac{u}{n} (1 + u + \dots + u^{n-1}) = \frac{u}{n} \frac{1-u^n}{1-u}$$

III.5.3. Loi binomiale.

III.5.3.1. Définition.

On dit qu'une variable aléatoire X suit une loi binomiale de paramètres n et p (relation notée $X \rightsquigarrow B(n; p)$) si sa loi de probabilité est donnée par :

$$P(X=k) = \binom{n}{k} p^k q^{n-k}, k=0, \dots, n.$$

avec $p + q = 1$, et où $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ est le coefficient binomial d'indices n et k .

Pour le coefficient binomial $\binom{n}{k}$, on utilise parfois l'ancienne notation C_n^k .

III.5.3.2. Propriétés.

$$g_X(u) = \sum_{k=0}^{k=n} \binom{n}{k} p^k q^{n-k} u^k = \sum_{k=0}^{k=n} \binom{n}{k} (p u)^k q^{n-k} = (p u + q)^n \text{ (formule du binôme)}$$

$$E(X) = g'_X(1) = \left(n p (q + p u)^{n-1} \right)_{u=1} = n p \text{ car } q + p = 1.$$

$$Var(X) = g''_X(1) + g'_X(1) - (g'_X(1))^2 = \left(n(n-1)p^2(q + p u)^{n-2} \right)_{u=1} + n p - n^2 p^2 = n p ((n-1)p + 1 - n p) = n p q.$$

III.5.3.3. Interprétation.

Pour $n = 1$, $B(1; p)$ est la loi de Bernoulli.

Soit E l'épreuve aléatoire consistant en n répétitions indépendantes d'une épreuve de Bernoulli, E_1, \dots, E_n , de même paramètre p .

Soit X le nombre de succès obtenus.

Les valeurs de X vont de 0 à n .

La loi de probabilité de X est une loi binomiale de paramètres n et p .

En effet, une épreuve de Bernoulli admet deux résultats possibles : l'échec, noté 0, de probabilité q , et le succès, noté 1, de probabilité p , avec $p + q = 1$.

Si E est l'épreuve aléatoire consistant en n répétitions indépendantes E_1, \dots, E_n , d'une épreuve de Bernoulli de paramètre p , un résultat possible de E est une suite de n nombres 0 ou 1, indiquant le résultat de chaque épreuve de Bernoulli.

Comme les épreuves sont indépendantes, les probabilités se multiplient et la probabilité d'un résultat particulier comportant k succès et $(n - k)$ échecs est $p^k q^{n-k}$.

Le nombre de résultats possibles comportant k succès et $(n - k)$ échecs est égal au nombre de choix possibles de la position des k succès, représentés par des 1, dans la suite des n résultats :

c'est le nombre des combinaisons des n places, k à k , soit $\binom{n}{k}$.

La probabilité d'avoir k succès, dans aussi $(n - k)$ échecs, est donc donnée par $\binom{n}{k} p^k q^{n-k}$.

En pratique, toute variable binomiale de paramètre n et p , peut toujours s'interpréter comme le **nombre de succès dans une répétition indépendante n fois d'une épreuve de Bernoulli de paramètre p** .

Remarquons aussi que si l'on note X_1, \dots, X_n , les variables de Bernoulli associées aux épreuves de Bernoulli indépendantes successives E_1, \dots, E_n , le nombre de succès X est la somme $X_1 + \dots + X_n$.

Une variable binomiale de paramètres n et p est somme de n variables de Bernoulli indépendantes de paramètre p .

Chaque variable de Bernoulli a, pour fonction génératrice, $(q + p u)$.

On a alors le résultat :

$$(q + p u)^n = g^{X_1 + \dots + X_n}(u) = g^{X_1}(u) \dots g^{X_n}(u)$$

Cette propriété caractérise les variables aléatoires dites **indépendantes** : deux variables aléatoires à valeurs entières sont indépendantes si, et seulement si, la fonction génératrice de leur somme est le produit de leurs fonctions génératrices.

III.5.3.4. Exemple : tirage avec remise.

Considérons une urne contenant N boules, dont n_1 boules blanches et n_2 boules noires, avec $N = n_1 + n_2$.

On tire une boule au hasard dans l'urne, on note sa couleur et on la remet dans l'urne (c'est ce qu'on appelle un **tirage non exhaustif**).

Si la boule est blanche, c'est le succès, noté 1, de probabilité $p = \frac{n_1}{N}$.

Si elle est noire, c'est l'échec, noté 0, de probabilité $q = \frac{n_2}{N}$.

On effectue ainsi n tirages successifs avec remise.

Au bout de n tirages, la probabilité d'avoir tiré k fois une boule blanche est donnée par la loi binomiale :

$$P(X = k) = \binom{n}{k} p^k q^{n-k},$$

où p est la proportion de boules blanches et q la proportion de boules noires.

Le modèle de l'urne avec tirages non exhaustifs peut être utilisé chaque fois que le paramètre p de la loi binomiale est un nombre rationnel.

III.5.3.5. Loi des fréquences.

Soit X une variable binomiale de paramètres n et p .

Considérons la variable $F = \frac{X}{n}$, qu'on appelle la variable fréquence.

Nous avons :

$$E(F) = E\left(\frac{X}{n}\right) = \frac{1}{n} E(X) = \frac{1}{n} n p = p.$$

$$Var(F) = Var\left(\frac{X}{n}\right) = \frac{1}{n^2} Var(X) = \frac{1}{n^2} n p q = \frac{pq}{n}.$$

III.5.3.6. Tendance vers la loi de Poisson.

Soit X_n une variable binomiale de paramètres n et p_n .

Supposons que, lorsque n augmente indéfiniment, le paramètre p_n soit tel que la limite de $n p_n$ soit un nombre réel positif λ .

Alors, pour tout k fixé, dans l'expression de la probabilité :

$$P(X_n = k) = \frac{n! (n-1)! \dots (n-k+1)!}{1 \times 2 \times \dots \times k} p_n^k (1-p_n)^{n-k}$$

$n(n-1) \dots (n-k)$ est équivalent à n^k .

$n(n-1) \dots (n-k) p_n^k$ est équivalent à $(n p_n)^k$, qui tend vers λ^k .

$(1-p_n)^{n-k}$ a pour logarithme $(n-k) \ln(1-p_n) \approx -(n-k) p_n \approx -n p_n = -\lambda$.

Donc $(1-p_n)^{n-k}$ tend vers $e^{-\lambda}$.

Finalement, $P(X_n = k)$ tend vers $e^{-\lambda} \frac{\lambda^k}{k!}$.

Une variable aléatoire prenant des valeurs entières k avec les probabilités $P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$, s'appelle une **variable de Poisson** de paramètre λ .

Sa loi de probabilité s'appelle **loi de Poisson de paramètre λ** . $X \rightsquigarrow P_\lambda$.

La loi de Poisson est aussi appelée la **loi des phénomènes rares**.

On l'obtient toujours comme approximation d'une loi binomiale lorsque p devient très petit avec une espérance $n p$ qui tend vers une constante strictement positive λ .

III.5.4. Loi hypergéométrique.

III.5.4.1. Définition.

On dit qu'une variable aléatoire X suit une **loi hypergéométrique de paramètres N, n, p** , si, et seulement si :

1°/ X prend un certain nombre de valeurs entières formant un intervalle $X(\Omega)$ contenu dans l'intervalle $[0; n]$;

2°/ $N p$ est un entier, noté N_1 ; on note alors N_2 l'entier $N - N_1$.

$$\frac{\binom{N_1}{k} \binom{N_2}{n-k}}{\binom{N}{n}}$$

3°/ Pour tout entier k appartenant à $X(\Omega)$, $P(X = k) =$

La relation " X suit une loi hypergéométrique de paramètres N, n, p " s'écrit : $X \rightsquigarrow H(N, n, p)$.

III.5.4.2. Interprétation : tirage sans remise.

Dans une urne contenant N boules dont N_1 blanches et N_2 noires, on tire au hasard n boules, sans remise (**tirage exhaustif**).

On note X le nombre de boules blanches tirées : X est une variable aléatoire dont les valeurs sont comprises entre 0 et n .

La valeur maximum de X est le plus petit des deux nombres entiers n et N_1 : on peut tirer jusqu'à n boules blanches si n est plus petit que N_1 , sinon on ne peut tirer qu'au plus N_1 boules blanches lorsque N_1 est plus petit que n .

La valeur minimum de X est le plus grand des deux entiers relatifs 0 et $n - N_2$: on peut tirer 0

boule blanche si n est inférieur au nombre N_2 de boules noires, sinon on tire au minimum $n - N_2$ boules blanches lorsque le nombre de boules tirées n est supérieur au nombre de boules noires N_2 .

$$X(\Omega) = [\text{Max}(0, n - N_2) ; \text{Min}(n, N_1)].$$

Le nombre de tirages possibles de n boules parmi N est le coefficient binomial $\binom{N}{n}$: ils sont équiprobables.

Pour un entier k appartenant à $X(\Omega)$, il y a $\binom{M_1}{k}$ façons de choisir k boules blanches parmi les N_1 boules blanches et, pour chaque choix de k boules blanches parmi les N_1 , il y a $\binom{M_2}{n-k}$ façons de choisir les $n - k$ boules noires complétant le tirage.

Le nombre de tirages de n boules contenant k boules noires est donc $\binom{M_1}{k} \binom{M_2}{n-k}$ et la probabilité d'avoir $X = k$ est donnée par le rapport :

$$P(X = k) = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}} = \frac{\binom{M_1}{k} \binom{M_2}{n-k}}{\binom{N}{n}},$$

donc X suit une **loi hypergéométrique** de paramètres $N, n, p = \frac{M_1}{N}$.

$$\frac{\binom{M_1}{k} \binom{M_2}{n-k}}{\binom{N}{n}}$$

Remarquons que l'égalité $P(X = k) = \frac{\binom{M_1}{k} \binom{M_2}{n-k}}{\binom{N}{n}}$ est valable, en fait, pour tout $k \in [0; n]$,

puisque en dehors de l'intervalle $X(\Omega)$, l'un des coefficients binomiaux $\binom{M_1}{k}$ et $\binom{M_2}{n-k}$ est nul et la probabilité correspondante est nulle.

On peut donc écrire :

$$\sum_{k=0}^{k=n} \frac{\binom{M_1}{k} \binom{M_2}{n-k}}{\binom{N}{n}} = \sum_{k=0}^{k=n} P(X = k) = 1$$

soit :

$$\sum_{k=0}^{k=n} \binom{M_1}{k} \binom{M_2}{n-k} = \binom{M_1 + M_2}{n}.$$

Cette égalité est appelée l'**identité de Vandermonde**.

III.5.4.3. Propriétés.

a) Moyenne.

$$\begin{aligned}
 E(X) &= \sum_{k=0}^{k=n} \frac{\binom{M_1}{k} \binom{M_2}{n-k}}{\binom{N}{n}} \\
 &= \sum_{k=0}^{k=n} \frac{\frac{M_1!}{k!(M_1-k)!} \times \frac{M_2!}{(n-k)!(M_2-n+k)!}}{\frac{(M_1+M_2)!}{n!(M_1+M_2-n)!}} \\
 &= \sum_{k=0}^{k=n} \frac{n!(M_1+M_2-n)!}{(M_1+M_2)!} \frac{M_1!}{k!(M_1-k)!} \frac{M_2!}{(n-k)!(M_2-n+k)!} \\
 &= \frac{n!(M_1+M_2-n)!}{(M_1+M_2)!} \sum_{k=0}^{k=n} \frac{M_1!}{k!(M_1-k)!} \frac{M_2!}{(n-k)!(M_2-n+k)!}
 \end{aligned}$$

Dans la somme, le terme correspondant à $k = 0$ est nul.

D'autre part, pour k différent de 0,

$$\begin{aligned}
 \frac{M_1!}{k!(M_1-k)!} &= N_1 \frac{(M_1-1)!}{(k-1)!(M_1-k)!} = N_1 \frac{(M_1-1)!}{(k-1)!((M_1-1)-(k-1))!} = N_1 \binom{M_1-1}{k-1} \\
 k \binom{M_1}{k} &= N_1 \binom{M_1-1}{k-1}
 \end{aligned}$$

D'où :

$$E(X) = \frac{n!(M_1+M_2-n)!}{(M_1+M_2)!} N_1 \sum_{k=1}^{k=n} \binom{M_1-1}{k-1} \binom{M_2}{n-k}$$

L'identité de Vandermonde donne, en posant $h = k - 1$, $\sum_{k=1}^{k=n} \binom{M_1-1}{k-1} \binom{M_2}{n-k} = \sum_{h=0}^{h=n-1} \binom{M_1-1}{h} \binom{M_2}{n-1-h} = \binom{M_1+M_2-1}{n-1}$.

$$E(X) = \frac{n!(M_1+M_2-n)!}{(M_1+M_2)!} N_1 \binom{M_1+M_2-1}{n-1} = \frac{n!(M_1+M_2-n)!}{(M_1+M_2)!} N_1 \frac{(M_1+M_2-1)!}{(n-1)!(M_1+M_2-1-n+1)!} = \frac{n M_1}{M_1+M_2} = n p.$$

$E(X) = n p.$

b) Variance.

La variance peut se calculer par la formule :

$$\begin{aligned}
 \text{Var}(X) &= E(X(X-1)) + E(X) - (E(X))^2. \\
 E(X(X-1)) &= \sum_{k=0}^{k=n} \frac{\binom{M_1}{k} \binom{M_2}{n-k}}{\binom{N}{n}} k(k-1) = \frac{n!(M_1+M_2-n)!}{(M_1+M_2)!} \sum_{k=0}^{k=n} \binom{M_1}{k} \binom{M_2}{n-k} k(k-1).
 \end{aligned}$$

Or on a, pour $k > 0$:

$$k \binom{M_1}{k} = N_1 \binom{M_1-1}{k-1}$$

et, pour $k > 1$:

$$\begin{aligned} & \binom{M_1-1}{k-1} = \binom{N_1-1}{k-2} \\ k \binom{M_1}{k} &= N_1 \binom{M_1-1}{k-1} = N_1 \binom{M_1-2}{k-2} \end{aligned}$$

Comme les termes correspondant à $k = 0$ et à $k = 1$, dans la somme $\sum_{k=0}^{k=n} \binom{M_1}{k} \binom{M_2}{n-k} k (k-1)$, sont nuls, il reste :

$$E(X(X-1)) = \frac{n! (M_1+M_2-n)!}{(M_1+M_2)!} \sum_{k=2}^{k=n} N_1 (N_1-1) \binom{M_1-2}{k-2} \binom{M_2}{n-k} = \frac{n! (M_1+M_2-n)!}{(M_1+M_2)!} N_1 (N_1-1) \sum_{k=2}^{k=n} \binom{M_1-2}{k-2} \binom{M_2}{n-k}$$

L'identité de Vandermonde, appliquée avec $h = k - 2$, donne :

$$\begin{aligned} & \sum_{k=2}^{k=n} \binom{M_1-2}{k-2} \binom{M_2}{n-k} = \sum_{h=0}^{h=n-2} \binom{M_1-2}{h} \binom{M_2}{n-2-h} = \binom{M_1+M_2-2}{n-2} \\ E(X(X-1)) &= \frac{n! (M_1+M_2-n)!}{(M_1+M_2)!} N_1 (N_1-1) \binom{M_1+M_2-2}{n-2} = \frac{n! (M_1+M_2-n)!}{(M_1+M_2)!} N_1 (N_1-1) \\ & \frac{(M_1+M_2-2)!}{(n-2)! (M_1+M_2-2-n+2)!} \\ &= n(n-1) \frac{M_1}{M_1+M_2} \frac{M_1-1}{M_1+M_2-1} \end{aligned}$$

$$\begin{aligned} \text{Var}(X) &= n(n-1) \frac{M_1}{M_1+M_2} \frac{M_1-1}{M_1+M_2-1} + n \frac{M_1}{M_1+M_2} - n^2 \frac{M_1^2}{(M_1+M_2)^2} \\ &= n(n-1) \frac{M_1}{M_1+M_2} \frac{M_1-1}{M_1+M_2-1} + n \frac{M_1}{M_1+M_2} \left(1 - n \frac{M_1}{M_1+M_2} \right) \\ &= n \frac{M_1}{M_1+M_2} \left((n-1) \frac{M_1-1}{M_1+M_2-1} + 1 - n \frac{M_1}{M_1+M_2} \right) \\ &= n \frac{M_1}{M_1+M_2} \frac{1}{(M_1+M_2)(M_1+M_2-1)} \left((n-1)(N_1-1)(N_1+N_2) + (N_1+N_2-1)(N_1+N_2) - n N_1 (N_1 \right. \\ & \left. + N_2 - 1) \right) \\ &= n \frac{M_1}{M_1+M_2} \frac{1}{(M_1+M_2)(M_1+M_2-1)} \left((N_1+N_2)((n-1)(N_1-1) + N_1+N_2-1 - n N_1) + n N_1 \right) \\ &= n \frac{M_1}{M_1+M_2} \frac{1}{(M_1+M_2)(M_1+M_2-1)} \left((N_1+N_2)(n N_1 - n - N_1 + 1 + N_1 + N_2 - 1 - n N_1) + n N_1 \right) \\ &= n \frac{M_1}{M_1+M_2} \frac{1}{(M_1+M_2)(M_1+M_2-1)} \left((N_1+N_2)(N_2-n) + n N_1 \right) \\ &= n \frac{M_1}{M_1+M_2} \frac{1}{(M_1+M_2)(M_1+M_2-1)} \left(N_2(N_2-n) + N_1(N_2-n) + n N_1 \right) \\ &= n \frac{M_1}{M_1+M_2} \frac{1}{(M_1+M_2)(M_1+M_2-1)} \left(N_2(N_1+N_2-n) \right) \\ &= n \frac{M_1}{M_1+M_2} \frac{M_2}{M_1+M_2} \frac{M_1+M_2-n}{M_1+M_2-1} = n p q \frac{N-n}{N-1} \end{aligned}$$

avec $q = 1 - p = \frac{M_2}{M_1+M_2}$ et $N = N_1 + N_2$.

$$\boxed{\text{Var}(X) = n p q \frac{N-n}{N-1}}$$

c) Tendence vers la loi binomiale.

Lorsqu'on compare les résultats obtenus pour la loi hypergéométrique de paramètres N, n, p , et pour la loi binomiale de paramètres n et p , on constate que :

1°/ Que les tirages soient effectués avec ou sans remise, l'espérance mathématique du nombre de boules blanches est la même, c'est np , qui ne dépend pas de N .

2°/ La variance change : le rapport $\frac{N-n}{N-1}$ est appelé rapport d'exhaustivité. Il est compris entre 0 et 1. Autrement dit, la variable hypergéométrique est moins dispersée que la variable binomiale.

3°/ Lorsque N augmente indéfiniment avec une proportion $p = \frac{M_1}{M_1+M_2} = \frac{M_1}{N}$ de boules blanches constante, le rapport :

$$\frac{\binom{M_1}{k} \binom{M_2}{n-k}}{\binom{N}{n}} = \frac{M_1(M_1-1)\dots(M_1-k+1)}{1 \times 2 \times \dots \times k} \frac{M_2(M_2-1)\dots(M_2-n+k+1)}{1 \times 2 \times \dots \times (n-k)} \frac{1 \times 2 \times \dots \times n}{N(N-1)\dots(N-n+1)}$$

est équivalent à :

$$\binom{n}{k} \frac{(M_1)^k (M_2)^{n-k}}{N^n} = \binom{n}{k} p^k q^{n-k}$$

On voit donc que la loi hypergéométrique tend vers la loi binomiale, lorsque le nombre de boules augmente indéfiniment dans l'urne en gardant une proportion constante de boules blanches.

III. 6. VARIABLES ALEATOIRES ABSOLUMENT CONTINUES.

III.6.1. Définitions.

Soit (Ω, \mathcal{A}, P) un espace probabilisé associé à une épreuve aléatoire E .

Nous étudions, dans cette section, des variables aléatoires sur (Ω, \mathcal{A}, P) susceptibles de prendre n'importe quelle valeur dans un intervalle réel, ou dans une réunion finie d'intervalles de longueurs non nulles et pour lesquelles la probabilité de prendre une valeur particulière fixée à l'avance est toujours nulle :

$$P(X = x) = 0, \text{ pour tout } x \text{ de } \mathbb{R}.$$

La **fonction de répartition** $x \mapsto F(x) = P(X \leq x)$ est alors continue.

III.6.1.1. Densité de probabilité.

a) Définition.

On appelle densité de probabilité sur \mathbb{R} , toute application $f: \mathbb{R} \rightarrow \mathbb{R}$ vérifiant :

1°/ f est positive : $f(x) \geq 0$, pour tout $x \in \mathbb{R}$.

2°/ f est continue sur \mathbb{R} sauf, éventuellement en un nombre fini de points.

3°/ L'intégrale $\int_{-\infty}^{+\infty} f(t) dt$ est convergente et vaut 1.

b) Exemples.

Exemple 1.

Soient a et b deux nombres réels, avec $a < b$.

L'application f définie par :

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{pour } a < x < b \\ 0 & \text{ailleurs} \end{cases}$$

est une densité de probabilité, appelée **densité uniforme** sur l'intervalle $[a; b]$.

Exemple 2.

Soit λ un nombre réel strictement positif.

L'application f définie par :

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{pour } 0 < x \\ 0 & \text{ailleurs} \end{cases}$$

est une densité de probabilité, appelée **densité exponentielle** de paramètre λ .

III.6.1.2. Variable aléatoire absolument continue.

Une variable aléatoire réelle X est dite **absolument continue** s'il existe une application f de \mathbb{R} dans \mathbb{R} , positive, continue sauf éventuellement en un nombre fini de points, telle que, pour tout nombre réel x , la fonction de répartition de X soit donnée par la formule :

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt.$$

L'application f est alors une densité de probabilité, appelée la **densité de probabilité de X** .

En effet, l'application f est, par définition, positive et continue sauf éventuellement en un nombre fini de points.

D'autre part, comme F est une fonction de répartition, sa limite, lorsque x tend vers $+\infty$, est 1,

de sorte que l'intégrale $\int_{-\infty}^{+\infty} f(t) dt = \lim_{x \rightarrow +\infty} \int_{-\infty}^x f(t) dt$, est convergente et vaut 1.

Le fait de changer la valeur de f en un nombre fini de points, ne change pas la valeur de l'intégrale et ne change donc pas la fonction de répartition de X .

On exprime cette propriété en disant que la densité de probabilité d'une variable aléatoire absolument continue est définie à un nombre fini de points près.

III.6.1.3. Propriété (admise).

Toute densité de probabilité sur \mathbb{R} est la densité de probabilité d'une variable aléatoire réelle absolument continue.

III.6.1.4. Loi de probabilité.

Définir la loi de probabilité d'une variable aléatoire absolument continue, c'est donner sa densité de probabilité : celle-ci s'obtient en dérivant la fonction de répartition partout où c'est possible.

En effet, d'après le théorème de dérivation d'une fonction définie par une intégrale, la dérivée de F est $f(x)$ partout où f est continue.

III.6.2. Probabilité d'un intervalle.

Soit X une variable aléatoire réelle absolument continue de densité de probabilité f . Soient a et b des nombres réels vérifiant $a < b$.

$$P(a < X \leq b) = \int_a^b f(t) dt$$

En effet :

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a) = \int_a^b f(t) dt.$$

En particulier, pour $a = x$, et $b = x + h$:

$$P(x < X \leq x + h) = F(x + h) - F(x) = \int_x^{x+h} f(t) dt.$$

Si f est continue en x , le théorème de la moyenne entraîne :

$$\lim_{h \rightarrow 0^+} \left(\frac{1}{h} \int_x^{x+h} f(t) dt \right) = f(x)$$

et l'on écrit, avec un abus d'écriture usuel en notation différentielle, pour un intervalle infiniment petit $[x; x + dx]$, de longueur dx :

$$P(x \leq X \leq x + dx) = P(x < X \leq x + dx) = f(x) dx,$$

ce qui justifie le terme de "densité de probabilité".

III.6.3. Espérance et variance.

Soit X une variable aléatoire réelle absolument continue de densité de probabilité f .

Pour un entier $k \geq 0$, on appelle **moment d'ordre k** , le nombre défini, lorsqu'il existe, par l'intégrale :

$$m_k = \int_{-\infty}^{+\infty} t^k f(t) dt.$$

L'**espérance mathématique** de X est, par définition, le nombre défini, lorsqu'il existe, par l'intégrale :

$$E(X) = m_1 = \int_{-\infty}^{+\infty} t f(t) dt.$$

La **variance** de X est, par définition, le nombre défini, lorsqu'il existe, par l'intégrale :

$$\text{Var}(X) = \int_{-\infty}^{+\infty} (t - E(X))^2 f(t) dt = E(X^2) - (E(X))^2 = m_2 - (m_1)^2$$

Lorsque la variance existe, l'**écart-type** de X est la racine carrée de la variance : $\sigma_X = \sqrt{\text{Var}(X)}$

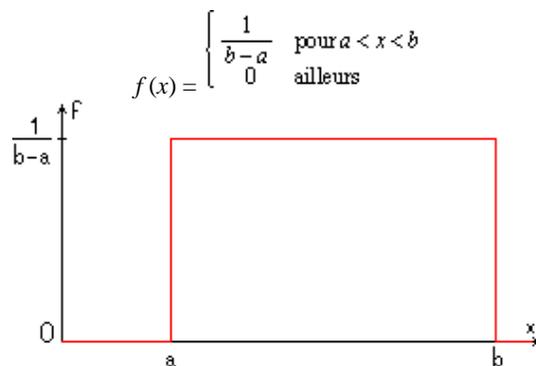
III.7. EXEMPLES : LOI UNIFORME, LOI NORMALE.

III.7.1. Loi uniforme.

III.7.1.1. Définition.

Soient a et b deux nombres réels vérifiant $a < b$.

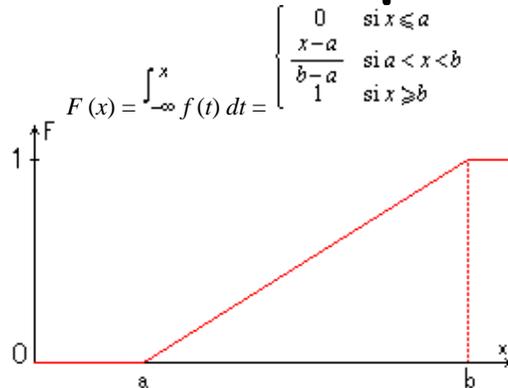
On dit qu'une variable aléatoire réelle X suit une **loi uniforme sur l'intervalle $[a; b]$** , relation notée $X \rightsquigarrow U_{[a; b]}$, si, et seulement si, elle admet pour densité de probabilité l'application f définie par :



Remarquons que la surface comprise entre la courbe représentative des variations de f et l'axe des abscisses est toujours égale à 1 puisque $\int_{-\infty}^{+\infty} f(t) dt = 1$.

III.7.1.2. Propriétés.

a) Fonction de répartition.



b) Espérance mathématique.

$$E(X) = \int_{-\infty}^{+\infty} t f(t) dt = \int_a^b t f(t) dt = \int_a^b \frac{t}{b-a} dt = \frac{1}{2} \frac{b^2 - a^2}{b-a} = \frac{a+b}{2}$$

$$E(X) = \frac{a+b}{2}$$

L'espérance mathématique d'une variable uniforme sur un intervalle est la valeur centrale.

c) Variance.

$$\begin{aligned} \text{Var}(X) &= E(X^2) - (E(X))^2 = \int_{-\infty}^{+\infty} t^2 f(t) dt - \left(\frac{a+b}{2}\right)^2 = \int_a^b \frac{t^2}{b-a} dt - \left(\frac{a+b}{2}\right)^2 = \frac{1}{3} \frac{b^3 - a^3}{b-a} - \frac{1}{4}(a^2 + 2ab + b^2) \\ &= \frac{1}{3}(b^2 + ab + a^2) - \frac{1}{4}(a^2 + 2ab + b^2) = \frac{1}{12}(4b^2 + 4ab + 4a^2 - 3a^2 - 6ab - 3b^2) = \frac{1}{12}(b^2 - 2ab + a^2) \\ &= \frac{[b-a]^2}{12} \end{aligned}$$

$$\text{Var}(X) = \frac{[b-a]^2}{12}$$

La variance d'une variable uniforme sur un intervalle est la douzième partie du carré de la longueur de l'intervalle.

III.7.2. Loi normale.

III.7.2.1. Définition.

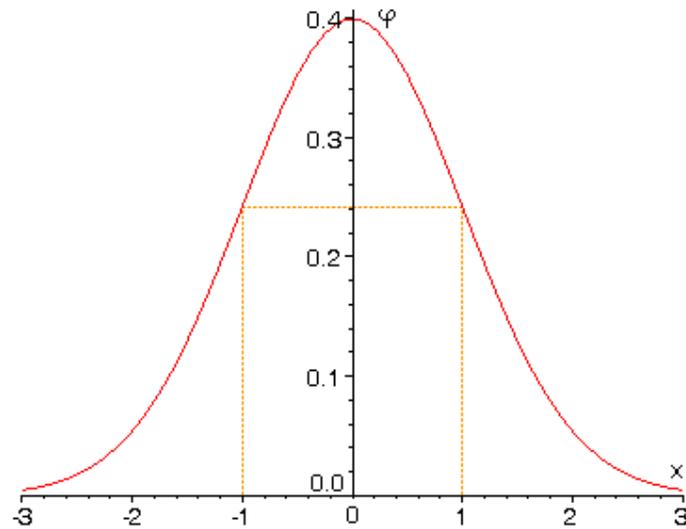
L'application $\varphi : x \mapsto \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ de \mathbb{R} dans \mathbb{R} est continue et positive.

On peut montrer que son intégrale sur \mathbb{R} vaut 1 (par changement de variables en coordonnées

polaires du plan dans l'intégrale $\iint_{\mathbb{R}^2} e^{-\frac{[x^2+y^2]}{2}} dx dy$).

C'est donc une densité de probabilité.

Sa courbe représentative a l'allure suivante :



De même, étant donnés deux nombres réels μ et σ , avec $\sigma > 0$, l'application $f: x \mapsto f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ de \mathbb{R} dans \mathbb{R} est continue et positive.

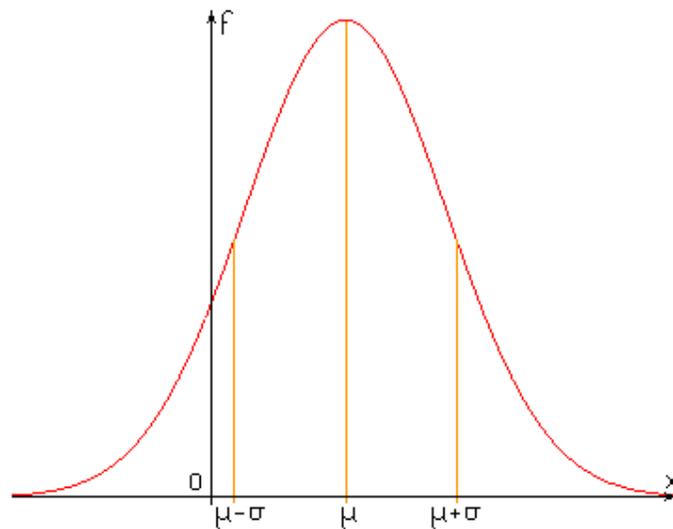
Le changement de variable $x = \mu + u\sigma$ dans l'intégrale montre que l'on a :

$$\int_{-\infty}^{+\infty} f(x) dx = \int_{-\infty}^{+\infty} \varphi(u) du = 1$$

de sorte que f est aussi une densité de probabilité.

Remarquons que f coïncide avec φ lorsque $\mu = 0$ et $\sigma = 1$.

La courbe représentative de f a l'allure suivante :



Elle est symétrique par rapport à μ et possède deux points d'inflexion, pour $x = \mu + \sigma$ et $x = \mu - \sigma$.

On appelle **variable normale de paramètres μ et σ** , toute variable aléatoire réelle absolument continue dont la densité de probabilité est donnée par :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

La relation "X est une variable normale de paramètres μ et σ " est notée $X \rightsquigarrow N(\mu ; \sigma)$, qui se lit "X suit la loi normale de paramètres μ et σ ".

On appelle **variable normale centrée réduite**, toute variable aléatoire réelle absolument continue dont la densité de probabilité est donnée par :

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

La relation "X est une variable normale centrée réduite" est la relation $X \rightsquigarrow N(0 ; 1)$, qui se lit "X suit la loi normale de paramètres 0 et 1".

La **loi normale** est aussi appelée "**loi de Gauss**", ou "**loi de Laplace-Gauss à une dimension**".

Elle a été découverte et étudiée en 1780 par Pierre Simon, marquis de Laplace.

III.7.2.2. Propriétés.

a) Espérance mathématique et variance.

La variable normale de paramètres μ et σ possède une espérance mathématique et une variance :

$$X \rightsquigarrow N(\mu ; \sigma) \Rightarrow E(X) = \mu \text{ et } Var(X) = \sigma^2.$$

b) Fonction de répartition.

Soit F la fonction de répartition de la variable normale de paramètres μ et σ , Φ la fonction de répartition de la variable normale de paramètres 0 et 1.

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

Cette relation se voit par changement de variable $t = \mu + u \sigma$ dans l'intégrale :

$$F(x) = \int_{-\infty}^x f(t) dt = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

Il résulte alors de la relation $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$ que la relation $X \rightsquigarrow N(\mu ; \sigma)$ est équivalente à la relation $U = \frac{X-\mu}{\sigma} \rightsquigarrow N(0 ; 1)$.

$$X \rightsquigarrow N(\mu ; \sigma) \Leftrightarrow U = \frac{X-\mu}{\sigma} \rightsquigarrow N(0 ; 1)$$

Autrement dit, X est une variable normale de paramètres μ et σ si, et seulement si, $U = \frac{X-\mu}{\sigma}$ est une variable normale centrée réduite.

La relation $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$ permet aussi de calculer les probabilités liées à la variable normale X de paramètres μ et σ , grâce à la table de la fonction de répartition de la variable normale centrée réduite.

III.8. LOI DES GRANDS NOMBRES.

Il y a plusieurs énoncés appelés "loi des grands nombres".

Nous énoncerons seulement celui qu'on appelle aussi la "**loi faible des grands nombres**" de **Bernoulli**.

Soit A un événement de probabilité p .

Lorsqu'on répète n fois, de façon indépendante, l'épreuve aléatoire où l'événement A est susceptible de se réaliser, la fréquence F_n de réalisation de A suit une loi donnée par la loi binomiale de paramètres n et p :

$$P(F_n = \frac{k}{n}) = \binom{n}{k} p^k q^{n-k}, k = 0, \dots, n.$$

avec $p + q = 1$, et où $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ est le coefficient binomial d'indices n et k .

L'espérance mathématique de F_n est p et sa variance est $\frac{pq}{n}$.

Lorsque n augmente indéfiniment, la variance $\frac{pq}{n}$ tend vers 0, les valeurs de F_n sont de moins en moins dispersées autour de la moyenne p et, intuitivement, la probabilité que F_n ne s'éloigne pas trop de p augmente.

La loi faible des grands nombres de Bernoulli énonce que, pour tout $\varepsilon > 0$, fixé à l'avance, la limite de la probabilité $P(|F_n - p| < \varepsilon)$ lorsque n augmente indéfiniment, est 1.

Ce résultat s'énonce aussi sous la forme :

La fréquence de l'événement A **converge en probabilité** vers la probabilité de l'événement A lorsque le nombre d'épreuves augmente indéfiniment.

Nous admettrons ce résultat, dont la démonstration précise sera donnée en deuxième année de DEUG.

La "**loi forte des grands nombres**", ou **théorème central limite**, fait, quant à elle, référence à la convergence "presque sûre", qui est la convergence uniforme de la fonction de répartition de la fréquence vers une **loi normale**.

IV. 1. GENERALITES.

IV.1.1. Introduction.

L'étude exhaustive d'un caractère donné dans une population est un **recensement**. Elle se heurte souvent à une impossibilité matérielle : coût trop élevé, ou destruction des individus étudiés.

Les méthodes d'analyse quantitative ont alors recours à la théorie des sondages, qui consiste à étudier un sous-ensemble de la population qu'on appelle un **échantillon**.

La théorie des sondages pose deux types de problèmes :

- L'échantillon doit être représentatif de la population : c'est la **théorie de l'échantillonnage**.
- Les techniques numériques utilisées sur les observations expérimentales doivent conduire à des résultats fiables, c'est-à-dire donnant une bonne représentation des paramètres inconnus de la population : c'est la **théorie de l'estimation** et des **tests**.

Les deux problèmes sont liés : la méthode d'échantillonnage utilisée a une influence sur les estimations obtenues.

En résumé, nous pouvons dire que la théorie des sondages est un outil mathématique permettant, à partir d'observations expérimentales partielles, de tenter d'atteindre une réalité inaccessible.

IV.1.2. Avantages de la méthode d'enquêtes par sondages.

La méthode d'enquêtes par sondages présente sur le recensement (lorsqu'il est possible) les avantages suivants :

1. Coût plus réduit.
2. Plus grande vitesse d'exécution (notamment pour les sondages d'opinions).
3. Plus grande fiabilité des résultats : le personnel étant plus réduit, il peut être plus qualifié.
4. Moins de risque d'erreur : le volume des données à traiter est plus faible.
5. Plus grand champ d'application, notamment dans le cas de destruction des unités testées.

IV.1.3. Etapes d'une enquête par sondage.

Pour effectuer une enquête par sondage, il est indispensable de respecter les instructions suivantes.

- Dresser une liste claire des objectifs de l'enquête.
- Etablir avec précision la population à échantillonner.
- Etablir une liste précise et courte des données à collecter.

- Définir le choix des méthodes de mesure : téléphone, convocations, visites à domicile, ...
- Etablir, lorsque c'est possible, le degré de précision désiré afin d'analyser le rapport des coûts et des avantages.
- Déterminer l'unité de l'échantillonnage : personne physique, collectivité, ...
- Etablir le plan de l'échantillonnage ou la méthode de sélection.
- Faire parfois une pré-enquête courte.
- Organiser le travail sur le terrain.
- Récueillir les données, les présenter, les synthétiser par traitement statistique.
- Conserver les données pour pouvoir les réutiliser.

IV.2. DIVERS TYPES DE SONDAGES.

Pour effectuer un sondage dans une population, c'est-à-dire pour en extraire un échantillon, deux types de méthodes sont employées : méthodes empiriques et méthodes aléatoires. Seules les méthodes aléatoires permettent d'utiliser la théorie de l'estimation.

IV.2.1. Méthodes empiriques : sondages raisonnés.

Ce sont les plus connues du grand public et les plus utilisées par les instituts de sondage d'opinion.

La précision de ces méthodes ne peut être calculée et leur réussite n'est que le résultat d'une longue pratique et de l'habileté professionnelle.

Les éléments sondés sont choisis dans la population suivant des critères fixés a priori.

IV.2.1.1. Méthode des unités types.

Elle repose sur l'idée suivante : les différentes variables attachées à un individu de la population n'étant pas indépendantes, un individu qui se trouve dans la moyenne de la population pour un certain nombre de caractères importants, sera également peu différent de la moyenne pour les autres caractères.

La méthode consiste donc à diviser la population en un certain nombre de sous-ensembles relativement homogènes et à représenter chacun d'eux par une unité-type.

On choisit donc des unités d'individus que l'on considère comme fortement représentatives de certaines catégories de population : cantons-types, bureau de vote pilotes, dont les résultats observés sur de longues périodes figurent les résultats définitifs d'une région ou d'une ville, etc.

Exemple.

L'INSEE décomposa en 1942 la France en 600 régions agricoles et, dans chaque région, désigna un canton-type.

Comme il y a en France environ 3000 cantons, la désignation de 600 cantons-types permettait de réduire d'un facteur 5 l'ampleur d'une étude des cantons.

IV.2.1.2. Méthode des quotas.

L'enquêteur prélève librement son échantillon, à condition de respecter une composition donnée à l'avance (pourcentage fixé d'agriculteurs, d'ouvriers, de cadres, etc., par exemple). Cette méthode est facile, mais aucun intervalle de confiance ne peut être donné. Elle suppose implicitement que les catégories retenues pour la détermination des quotas sont pertinentes quant à l'objet de l'étude, ce qui est bien difficile à établir. Pour diminuer l'arbitraire du choix, on impose à l'enquêteur des normes de déplacement géographique : c'est la **méthode de Politz**.

On utilise souvent des "**panels**", qui sont des échantillons permanents dont on étudie l'évolution.

Exemples.

- Panel d'audience à la télévision (médiamétrie, centres d'études d'opinion, ...).
- Panel de consommateurs (SECODIF : 4 500 ménages).
- Panel de détaillants (SOFRES).

Ces panels sont utilisés en marketing (lancement d'un produit, transfert de marques, etc.).

IV.2.2. Méthodes aléatoires.

Les éléments sondés sont extraits **au hasard** d'une liste connue a priori de la population, appelée **base de sondage**.

Exemples.

1. Liste d'immatriculation des véhicules automobiles en France.
C'est une très bonne base car elle est mise à jour régulièrement (cartes grises neuves, cartes grises à détruire).
2. Répertoire des entreprises (SIREN).
Chaque entreprise possède un numéro d'immatriculation à neuf chiffres, un nom ou raison sociale, une adresse exacte.
3. L'annuaire téléphonique est une mauvaise base de sondage car d'une part, tout individu ne possède pas obligatoirement un téléphone et, d'autre part, un individu peut posséder un téléphone et ne pas figurer sur l'annuaire (la liste rouge représente environ 8 % des abonnés et l'annuaire ne recense pas les téléphones portables, soit environ 40 % des téléphones).

Les bases de sondages sont en général établies à partir des résultats d'un recensement et elles sont corrigées périodiquement entre deux recensements.

Le tirage de l'échantillon est effectué dans la base de sondage selon des critères spécifiques à chaque méthode (plan de sondage).

Cette méthode de travail ne laisse aucune initiative aux enquêteurs : il est très simple de contrôler leur travail.

IV.2.2.1. Sondage élémentaire : échantillon aléatoire simple.

Dans un **échantillon aléatoire simple**, les éléments constituant l'échantillon sont extraits au hasard (à l'aide d'une table de nombres au hasard, par exemple) d'une liste de la population. On extrait ainsi n individus d'une population de taille N .

Le tirage peut s'effectuer avec ou sans remise, renvoyant ainsi généralement à un modèle de loi binomiale (avec remise), ou hypergéométrique (sans remise).

Si le tirage s'effectue avec remise, l'échantillon aléatoire simple est dit indépendant (EASI = **Echantillon Aléatoire Simple et Indépendant**).

La méthode permet de calculer des intervalles de confiance, comme nous le verrons plus loin.

Le rapport $f = \frac{n}{N}$ s'appelle le **taux de sondage**.

Par exemple, l'INSEE utilise des taux de sondage de l'ordre de $\frac{1}{1500}$ pour les enquêtes sur les conditions de vie des ménages.

Exemple.

Nous voulons extraire un échantillon de 8 individus dans une population formée de 437 individus.

Nous numérotons les individus de la population de 1 à 437.

Nous considérons trois colonnes consécutives d'une page de nombres au hasard : ils forment des nombres au hasard à trois chiffres.

Nous lisons ces nombres de trois chiffres en ne retenant que ceux qui sont compris entre 001 et 437.

Lorsque nous avons retenus 8 nombres, notre échantillon est constitué des 8 individus désignés dans la population par ces huit nombres.

Selon que nous effectuons un tirage avec ou sans remise, nous garderons ou écarterons un individu déjà tiré.

L'inconvénient majeur de la méthode élémentaire est son coût : les individus tirés peuvent être très éloignés géographiquement.

IV.2.2.2. Sondage stratifié.

La population étudiée Ω est partitionnée en q sous-populations $\Omega_1, \Omega_2, \dots, \Omega_q$, appelées "**strates**".

L'échantillon est constitué de la réunion de q échantillons choisis au hasard, un par strate : nous effectuons dans chaque strate un échantillonnage simple.

Exemple.

$\Omega = \{1, 2, 3, 4, 5\}, \Omega_1 = \{1, 2\}, \Omega_2 = \{3, 4, 5\}$.

Nous sélectionnons trois individus, dont un dans Ω_1 et deux dans Ω_2 .

Nous obtenons l'un des six échantillons possibles.

Cette méthode se justifie par deux raisons essentielles :

1. — L'existence d'une stratification de fait, soit pour des raisons géographiques, soit pour des raisons administratives.

Exemple 1 : enquête sur les conditions de vie pénitentiaire en France.

La population est celle des détenus en France

Les strates sont les populations de détenus dans les divers établissements pénitentiaires.

Exemple 2 : enquête sur la consommation par un organisme disposant de bureaux départementaux.

La population est celle des consommateurs français.
Les strates sont les consommateurs de chaque département.

2. — Un caractère étudié dans la population peut varier sous l'influence d'un certain nombre de facteurs.

Pour éliminer au mieux les risques de biais, nous créons des strates homogènes et, dans chacune d'elles, nous extrayons un échantillon aléatoire simple.

Exemple.

Pour étudier la consommation de tabac, si nous estimons que l'âge et le sexe sont des facteurs très influents, nous partageons la population en strates du type :

- Hommes de moins de 20 ans,
- Hommes de 20 à 30 ans,
- etc.
- Femmes de moins de 20 ans,
- Femmes de 20 à 30 ans,
- etc.

De chaque strate, nous extrayons un échantillon aléatoire simple.

IV.2.2.3. Echantillonnage systématique.

Les individus de la population Ω sont numérotés de 1 à N .

Pour sélectionner n individus, nous partageons la population en $k = \frac{N}{n}$ groupes : $\{1, \dots, k\}, \{1 + k, \dots, 2k\}, \dots, \{1 + (n-1)k, \dots, N\}$.

Nous choisissons au hasard l'individu i par les individus numérotés de 1 à k .

Nous constituons notre échantillon des individus $\{i, i + k, i + 2k, \dots, i + (n-1)k\}$.

Le choix de l'individu i détermine entièrement la constitution de l'échantillon.

Exemple.

$\Omega = \{1, \dots, 20\}, k = 4$.

Les échantillons possibles sont : $\{1, 5, 9, 13, 17\}, \{2, 6, 10, 14, 18\}, \{3, 7, 11, 15, 19\}, \{4, 8, 12, 16, 20\}$.

Cette méthode est bien adaptée à la sélection de cartes dans un fichier, ou au prélèvement de pièces dans une fabrication pour un contrôle de qualité.

Elle présente une certaine analogie avec la méthode précédente d'échantillonnage stratifié.

IV.2.2.4. Echantillonnage à plusieurs degrés.

La population Ω est divisée en sous-populations appelées unités primaires.

Chaque unité primaire est divisée en unités secondaires, etc.

Nous effectuons des tirages au hasard en cascade : nous tirons des unités primaires ; dans chaque unité primaire, nous tirons une unité secondaire, etc.

Exemple.

L'INSEE effectue des échantillonnages à quatre niveaux : départements, cantons, communes, ménages.

Cette méthode permet une exécution rapide.
Elle est économique, car elle focalise les tirages.

La méthode de tirage au hasard à chaque niveau peut varier suivant le cas, par exemple tirage proportionnel aux unités qu'il contient, ou tirage équiprobable.

Nous disons alors que nous pouvons avoir des tirages avec **probabilités inégales**.

Cas particulier : tirage par grappes.

Nous choisissons des grappes pour lesquelles nous gardons tous les "grains", ou individus.
Une "grappe" est un groupe d'individus de même nature.

Exemple : ménages d'un même immeuble.

IV.2.2.5. Conclusion.

En pratique, les diverses méthodes aléatoires peuvent être mêlées pour améliorer le rendement.

Pour chacune d'elle, nous pourrions varier les critères de tirage au hasard de chaque individu : avec remise, sans remise, avec des probabilités égales ou inégales.

IV.3. ESTIMATION DES PARAMETRES.

IV.3.1. Notion de paramètre.

Nous considérons une population Ω de taille finie N .

Dans cette population, nous étudions un caractère quantitatif réel prenant les valeurs réelles x_i , $i \in \{1, \dots, N\}$.

La fonction de répartition empirique $F_N(x)$ est une fonction en escalier.

La variable statistique représentant le caractère étudié peut être une variable quantitative discrète ou continue.

Le problème est de modéliser la fonction de répartition empirique $F_N(x)$, par la fonction de répartition $F(x)$ d'une variable aléatoire X , discrète ou continue suivant le cas, vérifiant $F(x_i) = F_N(x_i)$, $i \in \{1, \dots, N\}$.

Nous dirons que $F(x)$ définit la **loi de référence** associée à une population hypothétique infinie, dite **population de référence**.

La population Ω est appelée la **population-mère**.

La connaissance de la loi de référence du caractère étudié est d'un grand intérêt pour la déduction statistique.

Elle constitue un modèle mathématique du phénomène étudié.

Cette distribution théorique peut dépendre d'un certain nombre de paramètres inconnus.

Les sondages permettent d'estimer deux types de paramètres :

- Les paramètres propres à la population-mère : moyenne, variance, etc.
- Les paramètres propres à la loi de référence : paramètre d'une loi de Poisson, paramètres d'une loi normale, etc.

IV.3.2. Notion d'estimateur d'un paramètre de Ω .

IV.3.2.1. Estimateur et estimation ponctuelle.

Soit X un caractère quantitatif de la population Ω .

Ce caractère prend les valeurs inconnues $x_i, i \in \{1, \dots, N\}$.

Un résumé de l'ensemble des valeurs $\{x_1, \dots, x_N\}$ peut être défini par un ou plusieurs paramètres de Ω (moyenne, variance, proportion, etc.).

Soit y un tel paramètre de la population Ω .

Lorsque nous extrayons de la population un échantillon aléatoire simple E de taille n , nous pouvons calculer, avec les valeurs $\{x_1, \dots, x_n\}$ prises par X dans l'échantillon, une estimation ponctuelle de y , qui sera notée y^* .

Exemple.

Si y est la moyenne $\mu = \bar{X}$ de X , nous obtiendrons une estimation ponctuelle μ^* de la moyenne μ en prenant la moyenne arithmétique de l'échantillon :

$$\mu^* = \frac{1}{n} \sum_{i=1}^{i=n} x_i.$$

La valeur observée y^* n'est que l'une des valeurs possibles que l'on peut obtenir avec les divers échantillons possibles de taille n .

En réalité, avec une population de N individus, il y a un certain nombre, mettons k , d'échantillons possibles E_j de taille $n, j \in \{1, \dots, k\}$ (k dépend de la méthode d'échantillonnage).

Chaque échantillon possible E_j de taille n possède une certaine probabilité p_j d'être tiré.

A chaque échantillon possible E_j de taille n est associée une estimation ponctuelle y_j^* de y .

A chaque estimation ponctuelle y_j^* de y est donc associée la probabilité p_j d'être observée.

Nous pouvons alors définir une variable aléatoire \hat{y} prenant, pour chaque échantillon possible E_j de taille n , la valeur y_j^* avec la probabilité p_j .

Cette variable aléatoire \hat{y} est appelée un **estimateur** du paramètre y .

Les valeurs de \hat{y} sont les **estimations ponctuelles** de y .

La loi de probabilité de \hat{y} s'appelle la **distribution d'échantillonnage** de \hat{y} .

On appelle **fluctuation d'échantillonnage**, la variation des estimations ponctuelles de y et **aléas d'échantillonnage** les causes de ces variations.

IV.3.2.2. Caractéristiques d'un estimateur.

Il est logique de souhaiter que l'estimateur \hat{y} prenne des valeurs aussi voisines que possible de la valeur inconnue y que nous voulons estimer.

Nous sommes conduits à définir un certain nombre de qualités que doit présenter un "bon" estimateur.

a) Estimateur sans biais.

Nous dirons que \hat{y} est un estimateur sans biais du paramètre y , si, et seulement si, son espérance mathématique est y .

$$\text{sans biais} \Leftrightarrow E(\hat{y}) = y$$

Cette propriété traduit le fait qu'en moyenne, sur tous les échantillons possibles, nous retrouvons la valeur du paramètre que nous voulons estimer.

b) Estimateur robuste.

L'estimateur \hat{y} d'un paramètre y possède une variance $\sigma_{\hat{y}}^2$ qui traduit la dispersion des valeurs de \hat{y} autour de son espérance mathématique.

Cette variance dépend de la taille n de l'échantillon.

Nous dirons que \hat{y} est un estimateur robuste, ou **convergent**, de y si la limite, lorsque n tend vers N de $\sigma_{\hat{y}}^2$ est nulle.

$$\text{robuste} \Leftrightarrow \lim_{n \rightarrow N} \sigma_{\hat{y}}^2 = 0$$

Cette propriété traduit le fait suivant : si nous connaissons la valeur prise par le caractère pour tous les individus de la population, la valeur de \hat{y} est la valeur exacte y du paramètre.

Un **estimateur correct** est un estimateur sans biais et robuste.

c) Estimateur asymptotiquement gaussien.

Nous dirons qu'un estimateur \hat{y} d'un paramètre y est **asymptotiquement gaussien** si, et seulement si, il vérifie la propriété suivante :

$$\text{Lorsque } n \text{ augmente indéfiniment, la fonction de répartition de } \frac{\hat{y} - E(\hat{y})}{\sigma_{\hat{y}}} \text{ tend uniformément vers la fonction de répartition d'une variable normale centrée réduite.}$$

En pratique, dès que n est supérieur ou égal à 30, nous admettons que la fonction de

répartition de $\frac{\hat{y} - E(\hat{y})}{\sigma_{\hat{y}}}$ peut être remplacée par la fonction de répartition de la variable normale centrée réduite.

Lorsque n est suffisamment grand (en pratique $n \geq 30$), pour tout $\alpha \in [0, 1]$, le nombre réel positif u_α donné par :

$$\Phi(u_\alpha) = 1 - \frac{\alpha}{2}, \text{ où } \Phi \text{ est la fonction de répartition de la variable normale centrée réduite,}$$

vérifie :

$$P \left(\left| \frac{\hat{y} - E(\hat{y})}{\sigma_{\hat{y}}} \right| \leq u_{\alpha} \right) = 1 - \alpha.$$

En effet, comme la fonction de répartition de $\frac{\hat{y} - E(\hat{y})}{\sigma_{\hat{y}}}$ peut être remplacée par la fonction de répartition de la variable normale centrée réduite, dès que n est supérieur ou égal à 30, la symétrie de la loi normale donne :

$$P \left(\left| \frac{\hat{y} - E(\hat{y})}{\sigma_{\hat{y}}} \right| \leq u_{\alpha} \right) = \Phi(u_{\alpha}) - \Phi(-u_{\alpha}) = \Phi(u_{\alpha}) - (1 - \Phi(u_{\alpha})) = 2\Phi(u_{\alpha}) - 1 = 1 - \alpha.$$

Les valeurs de la fonction de répartition Φ sont données par des [tables](#).

Un **estimateur CAG** est un estimateur correct et asymptotiquement gaussien.

d) Amélioration d'un estimateur.

Etant donnés deux estimateurs \hat{y}_1 et \hat{y}_2 du même paramètre y , on dit que l'estimateur \hat{y}_1 est meilleur que l'estimateur \hat{y}_2 si l'espérance de $(\hat{y}_1 - y)^2$ est plus petite que l'espérance de $(\hat{y}_2 - y)^2$.

Ceci signifie simplement que l'on considère comme meilleur un estimateur dont les valeurs sont moins dispersées autour de la valeur de y .

Dans l'absolu, le meilleur estimateur d'un paramètre est celui dont pour lequel l'espérance de $(\hat{y} - y)^2$ est la plus petite possible.

Un estimateur sans biais dont la variance est minimale s'appelle un **estimateur précis**.

Pour un estimateur précis, l'espérance $E(\hat{y})$ est égale à y et la variance $\sigma_{\hat{y}}^2$ est minimale.

IV.3.3. Notion d'intervalle de confiance.

IV.3.3.1. Introduction.

Considérons un échantillon aléatoire simple E , de taille n , extrait de la population Ω (tirages au sort équiprobables, sans remise).

Dans cet échantillon, le caractère étudié prend les valeurs $\{x_1, \dots, x_n\}$.

Nous pouvons considérer la valeur prise par le caractère étudié pour l'individu i de l'échantillon comme la valeur prise par une variable aléatoire X .

L'ensemble des valeurs $\{x_1, \dots, x_n\}$ apparaît alors comme le résultat de n épreuves indépendantes sur la même variable aléatoire.

L'estimateur \hat{y} d'un paramètre y apparaît alors comme une fonction de n variables aléatoires indépendantes $X_i, i \in \{1, \dots, n\}$, de même loi de probabilité, qui est la loi de probabilité de X . X s'appelle la **variable parente**.

La connaissance de la loi de probabilité de X permet de calculer la loi de probabilité de \hat{y} .

$$\frac{\hat{y} - E(\hat{y})}{\sigma_{\hat{y}}}$$

La variable aléatoire centrée réduite $\frac{\hat{y} - E(\hat{y})}{\sigma_{\hat{y}}}$ correspondant à \hat{y} , possède une espérance mathématique nulle et une variance égale à 1.

Exemple 1.

Nous étudions la taille des individus d'une population d'effectif N .

Pour cela nous extrayons un échantillon aléatoire simple et indépendant d'effectif n .

Soit μ la moyenne de la taille des individus de la population.

Soit X la variable aléatoire "taille d'un individu" : à chaque individu de l'échantillon est associé une variable aléatoire indépendante "taille" X_i qui a la même loi de probabilité que la variable parente X .

L'estimateur

$$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^{i=n} X_i$$

de la taille moyenne μ dans la population, a, pour valeur dans l'échantillon, la moyenne arithmétique des tailles des individus de l'échantillon.

Cet estimateur possède une loi de probabilité qui peut être calculée en fonction de la loi de probabilité de X .

Exemple 2.

Soit σ^2 la variance de la taille des individus de la population.

Soit X la variable aléatoire "taille d'un individu" : à chaque individu de l'échantillon est associé une variable aléatoire indépendante "taille" X_i qui a la même loi de probabilité que la variable parente X .

L'estimateur

$$\hat{\sigma}^2 = \frac{1}{n-1} \left(\sum_{i=1}^{i=n} X_i^2 - \frac{1}{n} \left(\sum_{i=1}^{i=n} X_i \right)^2 \right)$$

de la variance σ^2 de la taille dans la population, a, pour valeur dans l'échantillon, $\frac{n}{n-1} S^2(X)$ où $S^2(X)$ est la variance des tailles des individus de l'échantillon (variance d'échantillonnage). Cet estimateur possède une loi de probabilité qui peut être calculée en fonction de la loi de probabilité de X .

IV.3.3.2. Intervalle de confiance pour les grands échantillons.

Si \hat{y} est un estimateur correct et asymptotiquement gaussien (estimateur CAG) d'un paramètre y , avec $E(\hat{y}) = y$, la relation

$$P \left(\left| \frac{\hat{y} - E(\hat{y})}{\sigma_{\hat{y}}} \right| \leq u_{\alpha} \right) = 1 - \alpha$$

s'écrit :

$$P(\hat{y} - u_{\alpha} \sigma_{\hat{y}} \leq \mu \leq \hat{y} + u_{\alpha} \sigma_{\hat{y}}) = 1 - \alpha.$$

L'événement $\hat{y} - u_{\alpha} \sigma_{\hat{y}} \leq \mu \leq \hat{y} + u_{\alpha} \sigma_{\hat{y}}$ a donc une probabilité $1 - \alpha$ de se réaliser lorsqu'on choisit au hasard un échantillon de taille $n \geq 30$.

Autrement dit, dans la population, la proportion des échantillons de taille $n \geq 30$ pour lesquels l'événement $\hat{y} - u_{\alpha} \sigma_{\hat{y}} \leq \mu \leq \hat{y} + u_{\alpha} \sigma_{\hat{y}}$ est réalisé est $1 - \alpha$.

Autrement dit encore, étant donné un échantillon de taille $n \geq 30$, choisi au hasard, la probabilité de réalisation de l'événement $\hat{y} - u_{\alpha} \sigma_{\hat{y}} \leq \mu \leq \hat{y} + u_{\alpha} \sigma_{\hat{y}}$ est $1 - \alpha$.

Or, pour un échantillon de taille n choisi au hasard, \hat{y} prend la valeur y^* et $\sigma_{\hat{y}}$ une valeur $s_{\hat{y}}$, de sorte que $\hat{y} - u_{\alpha} \sigma_{\hat{y}}$ prend une valeur

$$y_1 = y^* - u_{\alpha} s_{\hat{y}}$$

et $\hat{y} + u_{\alpha} \sigma_{\hat{y}}$ prend la valeur

$$y_2 = y^* + u_{\alpha} s_{\hat{y}}$$

L'intervalle

$$[y_1 ; y_2] = [y^* - u_{\alpha} s_{\hat{y}} ; y^* + u_{\alpha} s_{\hat{y}}]$$

dans lequel la taille n de l'échantillon est supérieure ou égale à 30 et $\Phi(u_{\alpha}) = 1 - \frac{\alpha}{2}$, s'appelle l'intervalle de confiance de y au risque α , ou **intervalle de confiance de y au niveau de confiance $1 - \alpha$** .

C'est un intervalle dans lequel la probabilité de trouver la vraie valeur de y est $1 - \alpha$.

Plus α est grand, plus l'amplitude de l'intervalle de confiance est petite, puisque Φ est une fonction croissante.

Dans la pratique, en l'absence de précision contraire, nous conviendrons de prendre $\alpha = 5 \%$.

Plus n est grand, plus la valeur de $\sigma_{\hat{y}}$ a des chances d'être proche de 0, donc plus la valeur de \hat{y} a des chances d'être proche de y .

Nous pourrons ainsi calculer la valeur de n qui permet d'avoir un intervalle de confiance d'amplitude donnée.

Les **valeurs à retenir** de la fonction de répartition de la variable aléatoire normale centrée réduite sont, pour $\Phi(u_{\alpha}) = 1 - \frac{\alpha}{2}$:

— $\Phi(1,645) = 0,950$, soit $u_{0,10} = 1,645$.

— $\Phi(1,960) = 0,975$, soit $u_{0,05} = 1,960$.

— $\Phi(2,575) = 0,995$, soit $u_{0,01} = 2,575$.

Ces valeurs donnent les intervalles de confiance aux niveaux de confiance 90 %, 95 %, 99 %.
La valeur utilisée par défaut est $u_{0,05} = 1,960$.

IV. 4. ETUDE DU SONDAGE ELEMENTAIRE.

Soit Ω une population d'effectif N dont on étudie un caractère X .

Si X est un caractère quantitatif, les paramètres qui caractérisent ce caractère sont :

$$\begin{aligned} \text{— la moyenne } \bar{X} = \mu &= \frac{1}{N} \sum_{i=1}^{i=N} x_i \\ \text{— la variance } \sigma^2 &= \frac{1}{N} \left(\sum_{i=1}^{i=N} x_i^2 - \frac{1}{N} \left(\sum_{i=1}^{i=N} x_i \right)^2 \right) \end{aligned}$$

Si X est un caractère qualitatif à deux modalités A et B , le paramètre qui caractérise X est la proportion p d'individus présentant la modalité A .

Les paramètres sont inconnus.

La théorie de l'échantillonnage a pour but de les estimer au mieux.

IV.4.1. Echantillon non exhaustif, tirage à probabilités égales.

Un tirage au hasard avec remise induit que chaque individu a une probabilité $\frac{1}{N}$ d'être tiré.

IV.4.1.1. Caractère quantitatif.

a) Loi de probabilité induite par le tirage de l'échantillon.

Le tirage avec remise, d'un individu de W , peut être représenté par une variable aléatoire parente, notée encore X , dont la loi de probabilité est définie par :

$$P(X = x_i) = \frac{1}{N}, i \in [1, N].$$

L'espérance mathématique de X est $E(X) = \sum_{i=1}^{i=N} \frac{1}{N} x_i = \frac{1}{N} \sum_{i=1}^{i=N} x_i = \mu$.

La variance de X est $Var(X) = E((X - \mu)^2) = \sigma^2$.

b) Estimateur de la moyenne de la population.

Constituer un échantillon de taille n par des tirages non exhaustifs équiprobables dans Ω , revient à définir n variables aléatoires indépendantes X_1, \dots, X_n , qui suivent toutes la même loi que X .

Soit $\{x_1, \dots, x_n\}$ la réalisation de l'échantillon E .

La moyenne arithmétique $\bar{x} = \frac{1}{n} \sum_{i=1}^{i=n} x_i$ est la réalisation par échantillonnage de la variable aléatoire

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{i=n} X_i.$$

L'espérance mathématique de l'estimateur $\hat{\mu}$ est $E(\hat{\mu}) = \frac{1}{n} \sum_{i=1}^{i=n} E(X_i) = \frac{1}{n} \times n E(X) = \mu$.

La variance de l'estimateur $\hat{\mu}$ est $\sigma_{\hat{\mu}}^2 = \text{Var}(\hat{\mu}) = \frac{1}{n^2} \sum_{i=1}^{i=n} \text{Var}(X_i) = \frac{1}{n^2} \times n \text{Var}(X) = \frac{\sigma^2}{n}$.

Par conséquent, $\hat{\mu}$ est un estimateur **sans biais** de μ ($E(\hat{\mu}) = \mu$) mais il n'est **pas robuste** ($\lim_{n \rightarrow \infty} \sigma_{\hat{\mu}}^2 = \frac{\sigma^2}{n} \neq 0$).

c) Estimateur de la variance de la population.

La variance expérimentale de l'échantillon est $s^2 = \frac{1}{n} \sum_{i=1}^{i=n} (x_i - \bar{x})^2$.

C'est la réalisation par échantillonnage de la variable aléatoire "**variance d'échantillonnage**" :

$$S^2 = \frac{1}{n} \left(\sum_{i=1}^{i=n} X_i^2 - \frac{1}{n} \left(\sum_{i=1}^{i=n} X_i \right)^2 \right) = \frac{1}{n} \sum_{i=1}^{i=n} (X_i - \hat{\mu})^2$$

L'espérance mathématique de S^2 est

$$E(S^2) = E \left(\frac{1}{n} \sum_{i=1}^{i=n} (X_i - \hat{\mu})^2 \right) = \frac{1}{n} \sum_{i=1}^{i=n} E \left((X_i - \hat{\mu})^2 \right)$$

$$E(S^2) = \frac{1}{n} \sum_{i=1}^{i=n} E \left((X_i - \mu + \mu - \hat{\mu})^2 \right)$$

$$E(S^2) = \frac{1}{n} \sum_{i=1}^{i=n} E(X_i - \mu)^2 + \frac{1}{n} \sum_{i=1}^{i=n} E(\mu - \hat{\mu})^2 + \frac{2}{n} \sum_{i=1}^{i=n} E \left((X_i - \mu)(\mu - \hat{\mu}) \right)$$

Mais on a :

$$\frac{1}{n} \sum_{i=1}^{i=n} E(X_i - \mu)^2 = \frac{1}{n} \sum_{i=1}^{i=n} E \left(X_i - E(X_i) \right)^2 = \frac{1}{n} n \text{Var}(X) = \sigma^2.$$

$$\frac{1}{n} \sum_{i=1}^{i=n} E(\mu - \hat{\mu})^2 = \frac{1}{n} \sum_{i=1}^{i=n} E \left((\hat{\mu} - E(\hat{\mu}))^2 \right) = \text{Var}(\hat{\mu}) = \frac{\sigma^2}{n}.$$

$$\frac{2}{n} \sum_{i=1}^{i=n} E \left((X_i - \mu)(\mu - \hat{\mu}) \right) = \frac{2}{n} E \left((\mu - \hat{\mu}) \sum_{i=1}^{i=n} (X_i - \mu) \right) = \frac{2}{n} E \left((\mu - \hat{\mu})(n\hat{\mu} - n\mu) \right) = -2 E \left((\hat{\mu} - \mu)^2 \right) = -2 \text{Var}(\hat{\mu}) = -2 \frac{\sigma^2}{n}.$$

Au total :

$$E(S^2) = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n} \sigma^2.$$

La variance d'échantillonnage n'est pas un estimateur sans biais de la variance σ^2 de la population : c'est un **estimateur biaisé**.

La linéarité de l'espérance mathématique montre que :

$$E\left(\frac{n}{n-1} S^2\right) = \frac{n}{n-1} E(S^2) = \sigma^2,$$

de sorte que l'estimateur :

$$\hat{\sigma}^2 = \frac{1}{n-1} \left(\sum_{i=1}^{i=n} X_i^2 - \frac{1}{n} \left(\sum_{i=1}^{i=n} X_i \right)^2 \right) = \frac{n}{n-1} S^2$$

est un estimateur **sans biais** de la variance σ^2 de la population : $E(\hat{\sigma}^2) = \sigma^2$.

IV.4.1.2. Caractère qualitatif.

Le paramètre étudié inconnu est la proportion p d'individus de la population présentant la modalité A du caractère qualitatif.

Pour chaque individu de la population, nous pouvons définir une variable aléatoire de Bernoulli, prenant la valeur 1, avec la probabilité p , si l'individu est porteur de la modalité A , 0 sinon, avec la probabilité $q = 1 - p$.

Choisir un échantillon de taille n , c'est choisir un n -uplet de variables aléatoires (X_1, \dots, X_n) de Bernoulli, indépendantes, de même paramètre p .

Soit (x_1, \dots, x_n) une réalisation de l'échantillon E .

La moyenne expérimentale $p^* = \frac{1}{n} \sum_{i=1}^{i=n} x_i$ est la réalisation par échantillonnage de la variable

aléatoire $\hat{p} = \frac{1}{n} \sum_{i=1}^{i=n} X_i$, qui représente la fréquence de la modalité A dans l'échantillon.

Son espérance mathématique est $E(\hat{p}) = \frac{1}{n} \sum_{i=1}^{i=n} E(X_i) = \frac{1}{n} \times n p = p$.

$$\hat{p} = \frac{1}{n} \sum_{i=1}^{i=n} X_i$$

est un **estimateur sans biais** de la proportion p des individus de la population présentant la modalité A du caractère étudié.

Sa variance est $Var(\hat{p}) = \frac{1}{n^2} \sum_{i=1}^{i=n} Var(X_i) = \frac{1}{n^2} \times n p (1 - p) = \frac{p(1-p)}{n}$.

Lorsque n tend vers N , cette variance ne tend pas vers 0, mais vers $\frac{p(1-p)}{N}$: l'estimateur \hat{p} de p n'est **pas un estimateur robuste**.

Pour les échantillons de grande taille ($n \geq 30$), on peut définir l'intervalle de confiance de p correspondant au risque α , par :

$$[p_1, p_2] = \left[p^* - u_\alpha \sqrt{\frac{p^*(1-p^*)}{n}} ; p^* + u_\alpha \sqrt{\frac{p^*(1-p^*)}{n}} \right]$$

avec $\Phi(u_\alpha) = 1 - \frac{\alpha}{2}$.

IV.4.2. Echantillon exhaustif, tirage à probabilités égales.

Un tirage au hasard sans remise induit que chaque échantillon de taille n a une probabilité $\frac{1}{\binom{N}{n}}$
 $= \frac{n!(N-n)!}{N!}$ d'être tiré.

IV.4.2.1. Caractère quantitatif.

a) Estimation de la moyenne.

Soit x_{ij} la réalisation du caractère X pour le j^{e} individu de l'échantillon $E_i = (X_{i1}, \dots, X_{in})$.
 La réalisation du i^{e} échantillon est un n -uplet (x_{i1}, \dots, x_{in}) .

La moyenne d'échantillonnage $\bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_{ij}$ est la réalisation d'une variable aléatoire $\frac{\Delta}{X}$ que nous allons définir.

Nous pouvons définir $\binom{N}{n}$ échantillons différents $E_i, i \in [1 ; \binom{N}{n}]$, de taille n , chacun ayant

une probabilité $p_i = \frac{1}{\binom{N}{n}} = \frac{n!(N-n)!}{N!}$ d'être tiré au hasard.

Considérons la variable aléatoire \hat{p} dont la loi de probabilité, uniforme, est définie par :

$$P\left(\frac{\Delta}{X} = \bar{x}_i\right) = p_i, i \in [1 ; \binom{N}{n}].$$

Son espérance mathématique est :

$$E\left(\frac{\Delta}{X}\right) = \sum_{i=1}^{\binom{N}{n}} p_i \bar{x}_i = \frac{1}{\binom{N}{n}} \sum_{i=1}^{\binom{N}{n}} \left(\frac{1}{n} \sum_{k=1}^n x_{ik} \right) = \frac{1}{\binom{N}{n}} \frac{1}{n} \sum_{i=1}^{\binom{N}{n}} \left(\sum_{k=1}^n x_{ik} \right).$$

La somme est une somme étendue à tous les échantillons de taille n .

Pour un k pris entre 1 et n , notons que x_{ik} est la valeur x_j du caractère X pour le k^{e} individu de l'échantillon, qui est le j^{e} individu de la population.

Cette valeur apparaît une fois dans tous les échantillons de taille n contenant cet individu de la population, mais pas forcément à la même place, c'est-à-dire pas forcément avec le même

indice k .

Or il y a $\binom{N-1}{n-1}$ échantillons de taille n contenant cet individu, de sorte que la valeur x_j de X

pour le j^{e} individu de la population, apparaît $\binom{N-1}{n-1}$ fois dans la somme $\sum_{i=1}^{\binom{M}{n}} \left(\sum_{k=1}^n x_{ik} \right)$.

Ce raisonnement est valable, bien sûr, pour tous les indices j de 1 à N .

Lorsque nous faisons la somme pour tous les échantillons de taille n , nous obtenons :

$$\sum_{i=1}^{\binom{M}{n}} \left(\sum_{k=1}^n x_{ik} \right) = \sum_{j=1}^N \binom{N-1}{n-1} x_j = \binom{N-1}{n-1} (x_1 + \dots + x_N)$$

$$E\left(\frac{\Delta}{X}\right) = \frac{1}{\binom{M}{n}} \frac{1}{n} \binom{N-1}{n-1} (x_1 + \dots + x_N) = \frac{1}{\binom{M}{n}} \frac{1}{n} \binom{N-1}{n-1} N \mu = \frac{n!(N-n)!}{N!} \frac{N}{n} \frac{(N-1)!}{(n-1)!(N-n)!} \mu = \mu$$

Moralité : la moyenne d'échantillonnage $\frac{\Delta}{X} = \frac{1}{n} \sum_{j=1}^n X_j$ est un estimateur **sans biais** de la moyenne μ du caractère X .

b) Variance de la moyenne d'échantillonnage.

La variance de $\frac{\Delta}{X}$ est donnée par $Var\left(\frac{\Delta}{X}\right) = E\left(\frac{\Delta}{X^2}\right) - \left(E\left(\frac{\Delta}{X}\right)\right)^2 = E\left(\frac{\Delta}{X^2}\right) - \mu^2$.

Calculons le terme :

$$E\left(\frac{\Delta}{X^2}\right) = \sum_{i=1}^{\binom{M}{n}} p_i \bar{x}^2$$

$$= \frac{1}{\binom{M}{n}} \sum_{i=1}^{\binom{M}{n}} \frac{1}{n^2} \left(\sum_{k=1}^n x_{ik} \right)^2 = \frac{1}{\binom{M}{n}} \frac{1}{n^2} \sum_{i=1}^{\binom{M}{n}} \left(\sum_{k=1}^n x_{ik}^2 \right)$$

$$= \frac{1}{\binom{M}{n}} \frac{1}{n^2} \left(\sum_{i=1}^{\binom{M}{n}} (x_{i1}^2 + \dots + x_{in}^2) + \sum_{i=1}^{\binom{M}{n}} \left(\sum_{\substack{j,k=1 \\ j \neq k}}^n x_{ij} x_{ik} \right) \right)$$

Pour tout individu de numéro j de Ω , il y a $\binom{N-1}{n-1}$ échantillons de taille n contenant cet

individu, de sorte que x_j^2 apparaît $\binom{N-1}{n-1}$ fois dans la somme $\sum_{i=1}^{\binom{M}{n}} (x_{i1}^2 + \dots + x_{in}^2)$.

Et ceci est vrai pour les N individus de la population.

De sorte que l'on obtient :

$$\sum_{i=1}^{\binom{M}{n}} \binom{n}{x_{i1}^2 + \dots + x_{in}^2} = \binom{N-1}{n-1} \binom{n}{x_1^2 + \dots + x_N^2} = \binom{N-1}{n-1} N \binom{n}{\sigma^2 + \mu^2} = \frac{N!}{(n-1)!(N-n)!} (\sigma^2 + \mu^2)$$

$$\sum_{i=1}^{\binom{M}{n}} \left(\sum_{\substack{j,k=1 \\ j \neq k}}^n x_{ij} x_{ik} \right)$$

Reste à calculer la somme

Dans chacun des $\binom{M}{n}$ échantillons de taille n , on forme $\frac{n(n-1)}{2}$ produits de la forme $x_{ij} x_{ik}$, avec $j \neq k$.

Dans l'ensemble des échantillons de taille n , on forme donc $\binom{M}{n} \frac{n(n-1)}{2}$ produits de deux valeurs de X différentes.

Comme il existe $\frac{M(M-1)}{2}$ produits de deux valeurs de X différentes, chacun intervient $\binom{M}{n} \frac{n(n-1)}{M(M-1)}$ fois dans la somme étendue à l'ensemble des échantillons de taille n .

On obtient donc :

$$\sum_{i=1}^{\binom{M}{n}} \left(\sum_{\substack{j,k=1 \\ j \neq k}}^n x_{ij} x_{ik} \right) = \binom{M}{n} \frac{n(n-1)}{M(M-1)} \sum_{\substack{j,k=1 \\ j \neq k}}^n x_j x_k$$

Or on peut écrire aussi :

$$\sum_{\substack{j,k=1 \\ j \neq k}}^n x_j x_k = \sum_{j=1}^n x_j \left(\sum_{\substack{k=1 \\ k \neq j}}^n x_k \right) = \left(\sum_{j=1}^n x_j \right) \left(\sum_{k=1}^n x_k \right) - \sum_{j=1}^n x_j^2$$

$$= \left(\sum_{j=1}^n x_j \right)^2 - \sum_{j=1}^n x_j^2 = (N\mu)^2 - N(\sigma^2 + \mu^2) = N((N-1)\mu^2 - \sigma^2)$$

On obtient alors :

$$\sum_{i=1}^{\binom{M}{n}} \left(\sum_{\substack{j,k=1 \\ j \neq k}}^n x_{ij} x_{ik} \right) = \binom{M}{n} \frac{n(n-1)}{M(M-1)} N((N-1)\mu^2 - \sigma^2) = \frac{N!}{n!(N-n)!} \frac{n-1}{N-1} ((N-1)\mu^2 - \sigma^2) = N \binom{N-2}{n-2} ((N-1)\mu^2 - \sigma^2)$$

$$E\left(\frac{\Delta}{X^2}\right) = \frac{1}{\binom{M}{n}} \frac{1}{n^2} \left[N \binom{N-1}{n-1} (\sigma^2 + \mu^2) + N \binom{N-2}{n-2} ((N-1)\mu^2 - \sigma^2) \right]$$

$$E\left(\frac{\Delta}{X^2}\right) = \frac{1}{\binom{M}{n}} \frac{N}{n^2} \left(\binom{N-1}{n-1} - \binom{N-2}{n-2} \right) \sigma^2 + \frac{1}{\binom{M}{n}} \frac{N}{n^2} \left(\binom{N-1}{n-1} + (N-1) \binom{N-2}{n-2} \right) \mu^2$$

$$\begin{aligned} \frac{1}{\binom{N}{n}} \frac{N}{n^2} \left(\binom{N-1}{n-1} \binom{N-2}{n-2} \right) &= \frac{n!(N-n)!}{N!} \frac{N}{n^2} \left(\frac{[N-1]!}{[n-1]!(N-n)!} - \frac{[N-2]!}{[n-2]!(N-n)!} \right) \\ &= \frac{[n-1]!(N-n)!}{[N-1]!} \frac{1}{n} \frac{[N-2]!}{[n-1]!(N-n)!} \left((N-1) - (n-1) \right) = \frac{1}{n} \frac{N-n}{N-1} \\ \frac{1}{\binom{N}{n}} \frac{N}{n^2} \left(\binom{N-1}{n-1} \binom{N-2}{n-2} \right)_{+(N-1)} &= \frac{n!(N-n)!}{N!} \frac{N}{n^2} \left(\frac{[N-1]!}{[n-1]!(N-n)!} + \frac{[N-2]!}{[n-2]!(N-n)!} \right) \\ &= \frac{[n-1]!(N-n)!}{[N-1]!} \frac{1}{n} \frac{[N-1]!}{[n-1]!(N-n)!} (1 + (n-1)) = 1 \\ E\left(\frac{\Delta}{\bar{X}}\right) &= \frac{1}{n} \frac{N-n}{N-1} \sigma^2 + \mu^2 \\ \text{Var}\left(\frac{\Delta}{\bar{X}}\right) &= E\left(\frac{\Delta}{\bar{X}}\right)^2 - \mu^2 = \frac{1}{n} \frac{N-n}{N-1} \sigma^2 \\ \boxed{\text{Var}\left(\frac{\Delta}{\bar{X}}\right) &= \frac{1}{n} \frac{N-n}{N-1} \sigma^2} \end{aligned}$$

Moralité : lorsque n tend vers N , la variance de $\frac{\Delta}{\bar{X}}$ tend vers 0, l'estimateur $\frac{\Delta}{\bar{X}}$ de μ est **robuste**.

La moyenne d'échantillonnage $\frac{\Delta}{\bar{X}} = \frac{1}{n} \sum_{j=1}^{j=n} X_{ij}$ est un estimateur sans biais et robuste, donc **correct**, de μ .

On remarquera aussi que la présence du rapport d'exhaustivité $\frac{N-n}{N-1}$, inférieur à 1, fait que la variance de $\frac{\Delta}{\bar{X}}$ est plus faible lorsque l'échantillon est exhaustif que lorsqu'il est non exhaustif : les valeurs de $\frac{\Delta}{\bar{X}}$ sont moins dispersées autour de la moyenne μ lorsque l'échantillon est exhaustif.

c) Estimation de la variance.

La variance expérimentale de l'échantillon $s^2 = \frac{1}{n} \sum_{j=1}^{j=n} (x_{ij} - \bar{x}_i)^2$ est une réalisation de la variable aléatoire :

$$S^2 = \frac{1}{n} \sum_{j=1}^{j=n} (X_{ij} - \frac{\Delta}{\bar{X}})^2 = \frac{1}{n} \left(\sum_{j=1}^{j=n} X_{ij}^2 - \frac{1}{n} \left(\sum_{j=1}^{j=n} X_{ij} \right)^2 \right)$$

L'espérance mathématique de cette variable aléatoire est ;

$$\begin{aligned} E(S^2) &= \frac{1}{n} \sum_{j=1}^{j=n} E\left((X_{ij} - \frac{\Delta}{\bar{X}})^2\right) = \frac{1}{n} \sum_{j=1}^{j=n} E\left((X_{ij} - \mu + \mu - \frac{\Delta}{\bar{X}})^2\right) \\ &= \frac{1}{n} \sum_{j=1}^{j=n} E\left((X_{ij} - \mu)^2\right) + \frac{1}{n} \sum_{j=1}^{j=n} E\left((\mu - \frac{\Delta}{\bar{X}})^2\right) - \frac{2}{n} \sum_{j=1}^{j=n} E\left((X_{ij} - \mu) \left(\frac{\Delta}{\bar{X}} - \mu\right)\right) \end{aligned}$$

Mais :

$$E((X_{ij} - \mu)^2) = E((X_{ij} - E(X_{ij}))^2) = \text{Var}(X_{ij}) = \sigma^2.$$

$$\frac{1}{n} \sum_{j=1}^{j=n} E((X_{ij} - \mu)^2) = \frac{1}{n} n \sigma^2 = \sigma^2.$$

$$\frac{1}{n} \sum_{j=1}^{j=n} E\left(\left(\mu - \frac{\Delta}{X}\right)^2\right) = \frac{1}{n} \sum_{j=1}^{j=n} \text{Var}\left(\frac{\Delta}{X}\right) = \frac{1}{n} n \text{Var}\left(\frac{\Delta}{X}\right) = \text{Var}\left(\frac{\Delta}{X}\right) = \frac{1}{n} \frac{N-n}{N-1} \sigma^2$$

$$\sum_{j=1}^{j=n} E\left(\left(X_{ij} - \mu\right) \left(\frac{\Delta}{X} - \mu\right)\right) = E\left(\left(\frac{\Delta}{X} - \mu\right) \sum_{j=1}^{j=n} (X_{ij} - \mu)\right) = E\left(\left(\frac{\Delta}{X} - \mu\right) n \left(\frac{\Delta}{X} - \mu\right)\right) = n E\left(\left(\frac{\Delta}{X} - \mu\right)^2\right) = n \text{Var}\left(\frac{\Delta}{X}\right)$$

Il reste alors :

$$E(S^2) = \sigma^2 + \frac{1}{n} \frac{N-n}{N-1} \sigma^2 - \frac{2}{n} n \text{Var}\left(\frac{\Delta}{X}\right) = \sigma^2 - \frac{1}{n} \frac{N-n}{N-1} \sigma^2 = \frac{n(N-1) - (N-n)}{n(N-1)} \sigma^2 = \frac{N[n-1]}{n(N-1)} \sigma^2$$

On voit donc que S^2 est un estimateur biaisé de σ^2 , mais que, par linéarité de l'espérance mathématique :

$$\hat{\sigma}^2 = \frac{1 - \frac{1}{N}}{1 - \frac{1}{n}} S^2 = \frac{N-1}{N} \frac{1}{n-1} \left(\sum_{j=1}^{j=n} X_{ij}^2 - \frac{1}{n} \left(\sum_{j=1}^{j=n} X_{ij} \right)^2 \right)$$

est un **estimateur sans biais de la variance σ^2** .

IV.4.2.2. Caractère qualitatif.

La fréquence d'échantillonnage $p^* = \frac{1}{n} \sum_{i=1}^{i=n} x_i$ de la modalité A du caractère qualitatif étudié est la valeur prise après échantillonnage par la variable aléatoire

$$\hat{p} = \frac{1}{n} \sum_{i=1}^{i=n} X_i.$$

Mais nous avons vu, précédemment, que l'espérance mathématique et la variance de X_i , étaient données par :

$$E(X_i) = p \\ \text{Var}(X_i) = p(1-p).$$

L'étude précédente montre que nous pouvons écrire :

$$E(\hat{p}) = p \\ \text{Var}(\hat{p}) = \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^{i=n} X_i\right) = \frac{1}{n^2} \text{Var}\left(n \frac{\Delta}{X}\right) = \text{Var}\left(\frac{\Delta}{X}\right) = \frac{1}{n} \frac{N-n}{N-1} p(1-p).$$

Ainsi, \hat{p} est un estimateur sans biais et robuste de p .

Sa réalisation $p^* = \frac{1}{n} \sum_{i=1}^{i=n} x_i$ dans un échantillon est une estimation ponctuelle sans biais de p .

Pour les grands échantillons, au niveau de confiance $1 - \alpha$, la réalisation de l'intervalle de confiance de p sera donné par $[p_1 ; p_2]$, avec

$$p_1 = p^* - u_\alpha \sqrt{\frac{N-n}{N-1}} \sqrt{\frac{p^*(1-p^*)}{n}}$$

$$p_2 = p^* + u_\alpha \sqrt{\frac{N-n}{N-1}} \sqrt{\frac{p^*(1-p^*)}{n}}$$

où u_α est défini par la relation $\Phi(u_\alpha) = 1 - \frac{\alpha}{2}$, Φ étant la fonction de répartition de la variable normale centrée réduite.

IV.4.2. Echantillon non exhaustif, tirage à probabilités inégales.

Soit $\Omega = \{\omega_1, \omega_2, \dots, \omega_N\}$ la population.

Nous étudions dans cette population un caractère quantitatif X de valeur x_j pour l'individu ω_j .

Notons p_j la probabilité de tirage de l'individu ω_j lors de la constitution de l'échantillon $\left(\sum_{j=1}^{j=N} p_j = 1 \right)$.

Tout tirage avec remise peut être schématisé par une variable aléatoire \hat{X} dont la loi de probabilité est définie par :

$$P(\hat{X} = x_j) = p_j, \forall j \in [1; N].$$

Notons :

$$\text{— } \mu = \frac{1}{N} \sum_{j=1}^{j=N} x_j, \text{ la moyenne du caractère } X \text{ dans la population.}$$

$$\text{— } \sigma^2 = \frac{1}{N} \left(\sum_{j=1}^{j=N} x_j^2 - \frac{1}{N} \left(\sum_{j=1}^{j=N} x_j \right)^2 \right), \text{ la variance de } X \text{ dans la population.}$$

Ces paramètres sont inconnus, nous cherchons à les estimer.

Nous supposons connues la taille N de la population et les probabilités p_j associées aux valeurs x_j .

Notons, pour simplifier, (x_1, \dots, x_n) la réalisation d'un échantillon.

IV.4.2.1. Estimation de la moyenne.

Considérons la variable aléatoire \hat{X} , définie par la loi de probabilité :

$$P\left(\hat{X} = \frac{x_j}{p_j}\right) = p_j, \forall j \in [1; N].$$

et soit :

$$\hat{m} = \frac{1}{Nn} \sum_{i=1}^{i=n} \hat{X}_i$$

la variable aléatoire de réalisation $m^* = \frac{1}{Nn} \sum_{i=1}^{i=n} \frac{x_i}{p_i}$ dans l'échantillon.

Nous avons :

$$E(\hat{m}) = \frac{1}{Nn} \sum_{i=1}^{i=n} E(\hat{X}_i) = \frac{1}{Nn} \sum_{i=1}^{i=n} \left(\sum_{j=1}^{j=N} \frac{x_j}{p_j} \right) = \frac{1}{Nn} \sum_{i=1}^{i=n} N\mu = \frac{1}{n} \sum_{i=1}^{i=n} \mu = \frac{1}{n} \times n \mu = \mu$$

La relation $E(\hat{m}) = \mu$ montre que la variable aléatoire \hat{m} est un **estimateur sans biais de μ** .

Sa réalisation $m^{**} = \frac{1}{Nn} \sum_{i=1}^{i=n} \frac{x_i}{p_i}$ dans l'échantillon est une estimation ponctuelle sans biais de μ .

IV.4.2.2. Variance de l'estimateur de la moyenne.

Nous avons :

$$E(\hat{X}_i) = \sum_{j=1}^{j=N} \frac{x_j}{p_j} = N\mu$$

$$E(\hat{X}_i^2) = \sum_{j=1}^{j=N} \frac{x_j^2}{p_j} = \sum_{j=1}^{j=N} \frac{x_j^2}{p_j}$$

$$Var(\hat{X}_i) = \sum_{j=1}^{j=N} \frac{x_j^2}{p_j} - N^2 \mu^2$$

Comme le tirage de l'échantillon est fait avec remise, les variables \hat{X}_i sont indépendantes, et, par conséquent :

$$Var(\hat{m}) = \left(\frac{1}{Nn} \right)^2 Var \left(\sum_{i=1}^{i=n} \hat{X}_i \right) = \frac{1}{N^2 n^2} \sum_{i=1}^{i=n} Var(\hat{X}_i)$$

$$= \frac{n}{N^2 n^2} Var(\hat{X}_i) = \frac{1}{N^2 n} Var(\hat{X}_i) = \frac{1}{N^2 n} \left(\sum_{j=1}^{j=N} \frac{x_j^2}{p_j} - N^2 \mu^2 \right)$$

$$Var(\hat{m}) = \frac{1}{N^2 n} \sum_{j=1}^{j=N} \frac{x_j^2}{p_j} - \frac{\mu^2}{n}$$

Cette variance s'exprime à l'aide de l'ensemble des valeurs x_j , inconnues, prises par le caractère X dans la population Ω .

Il serait intéressant d'en avoir une estimation à partir de la réalisation $\{x_1, \dots, x_n\}$ d'un échantillon.

IV.4.2.3. Estimation de la variance de l'estimateur de la moyenne.

Soit \hat{X}_i , la variable aléatoire définie, comme dans IV.4.2.1. par la loi de probabilité :

$$P \left(\hat{X}_i = \frac{x_j}{p_j} \right) = p_j, \forall j \in [1; N].$$

Nous avons vu que l'espérance mathématique de cette variable aléatoire était égale à $N\mu$, qu'on peut estimer par $N\hat{m}$.

Considérons la variance d'échantillonnage de la variable aléatoire \hat{X}_i , c'est la variable aléatoire :

$$\hat{s}_1^2 = \frac{1}{n} \sum_{i=1}^{i=n} (\hat{X}_i - N\hat{m})^2$$

L'espérance mathématique de \hat{s}_1^2 est :

$$\begin{aligned}
 E(\hat{s}_1^2) &= E\left(\frac{1}{n} \sum_{i=1}^{i=n} (\hat{X}_i' - N \hat{m}')^2\right) \\
 &= \frac{1}{n} \sum_{i=1}^{i=n} E\left((\hat{X}_i' - N \hat{m}')^2\right) \\
 &= \frac{1}{n} \sum_{i=1}^{i=n} E\left((\hat{X}_i' - N \mu + N \mu - N \hat{m}')^2\right) \\
 &= \frac{1}{n} \sum_{i=1}^{i=n} E\left((\hat{X}_i' - N \mu)^2\right) + \frac{1}{n} \sum_{i=1}^{i=n} E\left((N \mu - N \hat{m}')^2\right) + \frac{2}{n} \sum_{i=1}^{i=n} E\left((\hat{X}_i' - N \mu)(N \mu - N \hat{m}')\right) \\
 &= \frac{1}{n} \sum_{i=1}^{i=n} \text{Var}(\hat{X}_i') + \frac{1}{n} \sum_{i=1}^{i=n} \text{Var}(N \hat{m}') + \frac{2}{n} E\left((N \mu - N \hat{m}') \sum_{i=1}^{i=n} (\hat{X}_i' - N \mu)\right) \\
 &= \frac{1}{n} \times n \text{Var}(\hat{X}_i') + \frac{1}{n} \times n N^2 \text{Var}(\hat{m}') + \frac{2}{n} E\left((N \mu - N \hat{m}') (N n \hat{m}' - N n \mu)\right) \\
 &= \text{Var}(\hat{X}_i') + N^2 \text{Var}(\hat{m}') - \frac{2}{n} \times n N^2 \text{Var}(\hat{m}') \\
 &= \text{Var}(\hat{X}_i') - N^2 \text{Var}(\hat{m}') \\
 &= n N^2 \text{Var}(\hat{m}') - N^2 \text{Var}(\hat{m}') \\
 &= (n-1) N^2 \text{Var}(\hat{m}')
 \end{aligned}$$

La relation $E(\hat{s}_1^2) = (n-1) N^2 \text{Var}(\hat{m}')$, qui s'écrit aussi :

$$E\left(\frac{\hat{s}_1^2}{(n-1) N^2}\right) = \text{Var}(\hat{m}')$$

montre que

La variable aléatoire $\frac{\hat{s}_1^2}{(n-1) N^2}$ est un estimateur sans biais de la variance $\text{Var}(\hat{m}')$

et sa réalisation dans l'échantillon :

$$\frac{1}{n(n-1) N^2} \sum_{i=1}^{i=n} \left(\frac{x_i}{p_i} - N m'^*\right)^2 = \frac{1}{n(n-1) N^2} \left(\sum_{i=1}^{i=n} \left(\frac{x_i}{p_i}\right)^2 - \frac{1}{n} \left(\sum_{i=1}^{i=n} \frac{x_i}{p_i}\right)^2\right)$$

compte tenu de la relation $N m'^* = m'^* = \frac{1}{n} \sum_{i=1}^{i=n} \frac{x_i}{p_i}$, est une **estimation ponctuelle sans biais de la variance de \hat{m}'** .

$$\sigma_{\hat{m}'^*}^2 = \frac{1}{n(n-1) N^2} \left(\sum_{i=1}^{i=n} \left(\frac{x_i}{p_i}\right)^2 - \frac{1}{n} \left(\sum_{i=1}^{i=n} \frac{x_i}{p_i}\right)^2\right)$$

Cette estimation de la variance de \hat{m}' permet de construire, pour les grands échantillons, un **intervalle de confiance de la moyenne μ** :

$$m'^* \pm u_\alpha \sigma_{\hat{m}'^*}.$$