

DEVOIR SURVEILLÉ # 8

le Samedi 26/03/2022, 8h → 11h30



Consignes La qualité de la rédaction, la clarté et la précision des raisonnements, entreront pour une part importante dans l'appréciation de la copie. Les abréviations, sigles ou phrases nominales sont à proscrire. La numérotation des exercices (et des questions) doit être respectée et mise en évidence. Les résultats doivent être encadrés proprement.

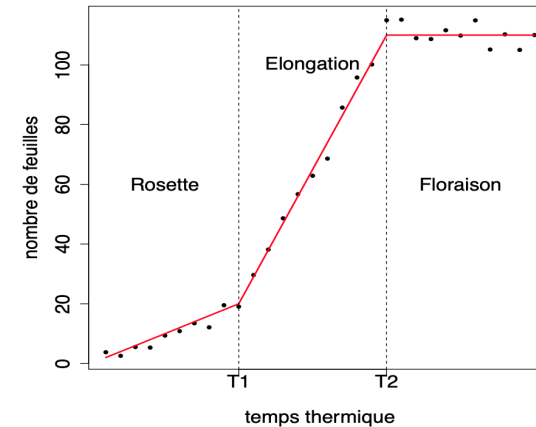
Chaque candidat est responsable de la vérification de son sujet d'épreuve : pagination et impression de chaque page. Si, au cours de l'épreuve, un candidat repère ce qui lui semble être une erreur d'énoncé, il convient de le signaler sur la copie et de poursuivre la composition en expliquant les raisons des initiatives qui ont été prises.

L'usage de la calculatrice est interdit.



Problème 1 | D'après concours Agro—Véto 2017 – Épreuve de modélisation (remaniée) (Solution 7) Dans le domaine de l'écophysiologie végétale on s'intéresse aux différentes phases du développement des plantes afin de mieux contrôler les rendements de culture. Plus précisément, on s'intéresse à l'évolution du nombre de feuilles d'une plante en fonction du "temps thermique". Ce dernier correspond à la somme cumulée des températures au cours des différents jours de l'expérience. De telles mesures permettent de mettre en évidence trois phases de développement appelées phases de rosette, d'élongation et de floraison. La modélisation la plus souvent utilisée consiste à supposer que l'évolution du nombre de feuilles en fonction du temps thermique se comporte comme une fonction affine au cours de chacune des trois phases précédentes (cf. Figure 1).

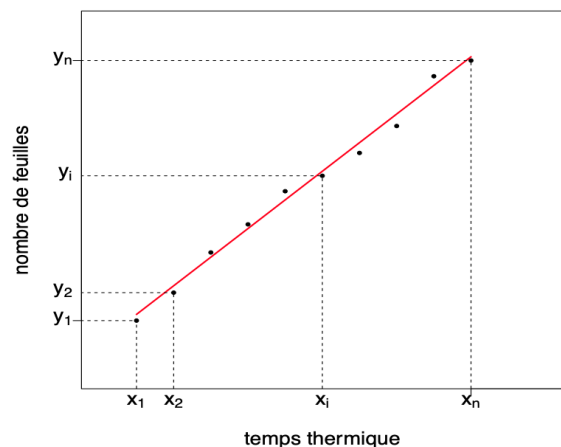
Fig. 1. : Schéma illustrant l'évolution du nombre de feuilles d'une plante en fonction du temps thermique; y sont représentés les valeurs expérimentales • et les ajustements affines dans chacune des trois phases.



L'objectif est alors de mettre en place des méthodes automatiques permettant de détecter les instants séparant d'une part les phases de rosette et d'élongation (T_1) et d'autre part les phases d'élongation et de floraison (T_2). Il est, en effet, intéressant de voir si ces instants ont tendance à changer en fonction de certaines conditions expérimentales auxquelles les plantes pourraient être soumises.

PARTIE I — SUR L'AJUSTEMENT AFFINE DÉTERMINISTE AU SENS DES MOINDRES CARRÉS Dans cette partie, on se focalise sur l'une des trois phases précédentes et on s'intéresse à l'ajustement affine que l'on obtient à partir du nuage de points $((x_1, y_1), \dots, (x_n, y_n))$, n étant un entier naturel strictement plus grand que 1, en utilisant le critère des moindres carrés où l'on supposera que les x_i sont distincts. Dans la notation (x_i, y_i) , x_i correspond au temps thermique de la i -ème observation et y_i correspond au nombre de feuilles de la i -ème observation (cf. Figure 2).

FIG. 2. : Schéma illustrant l'évolution du nombre de feuilles en fonction du temps thermique durant l'une des trois phases.




1. On note $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$. Montrer que :

$$\sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i = \sum_{i=1}^n (x_i - \bar{x})^2.$$

2. On définit la fonction $F : \mathbf{R}^2 \rightarrow \mathbf{R}$ par : $F(a, b) = \sum_{i=1}^n (y_i - bx_i - a)^2$ pour tout $(a, b) \in \mathbf{R}^2$. Montrer que l'unique point critique de F noté $(\hat{a}, \hat{b}) \in \mathbf{R}^2$ vérifie :

$$\hat{a} = \bar{y} - \hat{b} \cdot \bar{x}, \quad \hat{b} = \frac{\sum_{i=1}^n y_i (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

3. On admet que le couple (\hat{a}, \hat{b}) réalise un minimum global de F . Interpréter.

4.  On souhaite automatiser les calculs de \hat{a} et \hat{b} lorsqu'un grand nombre de valeurs (x_i, y_i) sont à disposition. On utilise Python, en considérant que les valeurs x_1, \dots, x_n sont stockées dans une liste x et que les valeurs y_1, \dots, y_n sont stockées dans une liste y .


4.1) Écrire une fonction `moy` qui prend en entrée L , une liste de nombres (non vide), et qui renvoie la moyenne des valeurs de L .

4.2) Soit L une liste de longueur $n \geq 1$. Indiquer, sans justification, la ou les réponse(s) correcte(s) :

- i) Les éléments de L sont numérotés de 0 à $n - 1$ (inclus) ;
- ii) Les éléments de L sont numérotés de 1 à n (inclus) ;
- iii) Les éléments de L sont numérotés de 0 à $n + 1$ (inclus) ;
- iv) Il y a n éléments dans L .
- v) Il y a $n + 1$ éléments dans L .

4.3) Écrire une fonction `Bchap` qui prend en entrée x et y , deux listes de nombres de même longueur $n \geq 2$, et qui renvoie la valeur de \hat{b} associée.

4.4) Écrire une fonction `Achap` qui prend en entrée x et y , deux listes de nombres de même longueur $n \geq 2$, et qui renvoie la valeur de \hat{a} associée.

5.  On dispose d'un fichier `data.txt` qui contient une série de lignes, chacune contenant deux réels (correspondant à x_i, y_i) séparés par un espace (chaîne vide). Par exemple, pour un fichier de deux lignes,

```
10 20
12 27
```

```
...
```



On souhaite dans cette question s'intéresser à l'importation des données, afin de construire les listes x , y précédentes. On rappelle les commandes ci-après.

- ▶ `f = open("monfichier.txt", "r")` ouvre le document intitulé `monfichier.txt` (en mode lecture); le fichier est alors désigné par la variable `f`.
- ▶ `f.read()` lit l'intégralité du contenu du fichier `f` et le renvoie sous forme d'une chaîne de caractères.
- ▶ `f.readline()` lit la première ligne du fichier `f` et le renvoie sous forme d'une chaîne de caractères. Ré-exécuter `f.readline()` lit alors la deuxième ligne; et ainsi de suite.
- ▶ `f.close()` ferme le fichier `f`.
- ▶ La chaîne `"\n"` (de longueur 1) désigne le passage à la ligne.

Par exemple, on pourra observer les commandes ci-après exécutées sur le fichier `data.txt` présenté ci-dessus.

```
>>> f = open("data.txt", "r")
>>> f.readline()
'10 20\n'
>>> f.readline()
'12 27'
>>> f.readline()
''
>>> f.close()
```



5.1) Écrire une fonction d'entête `extraction_ch()` qui renvoie une liste composée de chaînes de caractères : la première correspondant à la première ligne de `data.txt`, la 2ème à la seconde *etc.*. Cette fonction devra, au moins, ouvrir le fichier `data.txt`, le lire puis le fermer. Par exemple, avec le fichier `data` précédent, la fonction devra retourner la liste ci-après.

```
>>> L
['10 20\n', '12 27']
```

5.2) Compléter la fonction suivante d'entête `ch_vers_list(ch)` prenant en entrée une chaîne de caractère (de même forme que celles renvoyées par `extraction_ch`) et qui renvoie la liste de nombres correspondant; ces nombres seront de type flottant. Par exemple, `ch_vers_list('10 20')` renvoie `[10.0, 20.0]`.

```
def ch_vers_list(ch):
    L = _____
    n = len(ch)
    sh = ""
    for k in range(0, n):
        if ch[k] != " ":
            sh = sh+ch[k]
        else:
            _____
            _____
    _____
```



return L



5.3) En déduire une fonction d'en-tête `extraction_data()` qui extrait les données de `data.txt` en retournant deux listes `x`, `y` où dans `x` sont stockés les temps thermiques x_1, x_2, \dots, x_n , dans `y` sont stockés les nombres de feuilles y_1, \dots, y_n .

PARTIE II — PRISE EN COMPTE D'UN ALÉATOIRE DE VARIABILITÉ EXPÉRIMENTALE : RÉGRESSION BRUITÉE Dans cette partie, pour tenir compte de la variabilité des expérimentations, nous allons modéliser les y_i comme des réalisations de variables aléatoires Y_i définies par les n équations suivantes, n étant un entier naturel strictement plus grand que 1 :

$$Y_i = a + bx_i + \varepsilon_i, \quad 1 \leq i \leq n,$$

où a, b sont des paramètres réels inconnus, x_i est le temps thermique associé à la i -ème observation (il sera considéré comme déterministe ici c'est-à-dire non aléatoire), et les ε_i , appelées *bruits de la régression*, sont des variables aléatoires indépendantes d'espérance nulle et de variance $\sigma^2 > 0$ définies sur un même espace probabilisé $(\Omega, \mathcal{F}, \mathbf{P})$. Ceci correspond à un *modèle de régression linéaire simple*. En utilisant la même démarche que celle proposée dans la partie I, nous définissons :

$$\hat{a} = \bar{Y} - \hat{b} \cdot \bar{x}, \quad \hat{b} = \frac{\sum_{i=1}^n Y_i (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i, \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Les quantités \hat{a} et \hat{b} deviennent alors des variables aléatoires définies sur le même espace probabilisé, et donc des estimateurs de a, b .

6. Calculer $\mathbf{E}(Y_i)$ pour tout $i \in \llbracket 1, n \rrbracket$, puis $\mathbf{E}(\sum_{i=1}^n Y_i (x_i - \bar{x}))$. Pour la seconde espérance, on pourra utiliser le résultat de la question 1. même s'il n'a pas été établi.
7. En déduire que \hat{b} est un estimateur non biaisé de b .
8. Calculer $\mathbf{E}(\bar{Y})$. En déduire que \hat{a} est un estimateur non biaisé de a .

9. Montrer que :
$$\widehat{b} - b = \frac{\sum_{i=1}^n \varepsilon_i (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

10. Justifier que : $\text{Var}(\widehat{b}) = \text{Var}(\widehat{b} - b)$, puis en déduire que :

$$\text{Var}(\widehat{b}) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

11. Dans cette question, nous nous intéressons au comportement de l'estimateur \widehat{b} en fonction du nombre d'observations n pour une suite particulière de temps thermiques qui sera précisée ci-après. On le notera donc plutôt \widehat{b}_n et on rappelle que \widehat{b}_n est convergent vers b si :

$$\forall \varepsilon > 0, \quad \mathbf{P}\left(|\widehat{b}_n - b| \geq \varepsilon\right) \xrightarrow{n \rightarrow \infty} 0.$$

11.1) Calculer $\sum_{i=1}^n (x_i - \bar{x})^2$ lorsque $x_i = i$ pour tout $i \in \llbracket 1, n \rrbracket$.

11.2) Déduire en utilisant la variance de \widehat{b} calculée précédemment, que lorsque $x_i = i$ pour tout $i \in \llbracket 1, n \rrbracket$, \widehat{b}_n converge vers b .

12. Soit $\tilde{b} = \sum_{i=1}^n \mu_i Y_i$ (avec $\mu_1, \dots, \mu_n \in \mathbf{R}$) une variable aléatoire combinaison linéaire des Y_i , et telle que :

$$\sum_{i=1}^n \mu_i = 0, \quad \sum_{i=1}^n \mu_i x_i = 1.$$

On notera $\tilde{\mathcal{E}}$ l'ensemble des variables aléatoires de la forme précédente.

12.1) Justifier que : $\tilde{b} - \widehat{b} = \sum_{i=1}^n Y_i \left[\mu_i - \frac{x_i - \bar{x}}{\sum_{j=1}^n (x_j - \bar{x})^2} \right]$.

12.2) En déduire que : $\text{Var}(\tilde{b}) = \text{Var}(\tilde{b} - \widehat{b}) + \text{Var}(\widehat{b})$, puis que :

$$\text{Var}(\tilde{b}) \geq \text{Var}(\widehat{b}).$$

12.3) Montrer que : $\widehat{b} \in \tilde{\mathcal{E}}$.

12.4) Justifier l'affirmation suivante : « parmi les estimateurs sans biais de $\tilde{\mathcal{E}}$, \widehat{b} est celui de variance minimale ».

PARTIE III — CAS DE BRUITS GAUSSIENS On se place dans cette partie dans le même cadre que celui de la partie II mais on suppose de plus que les variables aléatoires ε_i sont indépendantes et identiquement distribuées de loi gaussienne commune $\mathcal{N}(0, \sigma^2)$. On rappelle que \widehat{b} est défini par :

$$\widehat{b} = \frac{\sum_{i=1}^n Y_i (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

13. Préciser la loi de Y_i pour tout $i \in \llbracket 1, n \rrbracket$. Justifier que \widehat{b} suit une loi normale dont on précisera les paramètres.

14. (**Estimation de la variance par variance empirique corrigée**) Dans la phase de floraison et uniquement dans celle-ci, une façon usuelle d'estimer σ^2 est de considérer l'estimateur suivant :

$$\widehat{\sigma^2} = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2.$$

14.1) Que vaut sûrement b dans la phase de floraison? On considèrera cette valeur de b dans la suite. Montrer que :

$$\widehat{\sigma^2} = \frac{1}{n-1} \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2, \text{ où } \bar{\varepsilon} = \frac{1}{n} \sum_{i=1}^n \varepsilon_i.$$

Jusqu'à la fin de cette question, on supposera que $n = 4$.

14.2) (**Réduction d'une matrice**) Soient

$$\tilde{\mathbf{E}} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix}, \quad \mathbf{E} = \begin{pmatrix} \varepsilon_1 - \bar{\varepsilon} \\ \varepsilon_2 - \bar{\varepsilon} \\ \varepsilon_3 - \bar{\varepsilon} \\ \varepsilon_4 - \bar{\varepsilon} \end{pmatrix}, \quad \Gamma = \begin{pmatrix} 1 - \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & 1 - \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & 1 - \frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & 1 - \frac{1}{4} \end{pmatrix}.$$

i) Montrer que : $\mathbf{E} = \Gamma \tilde{\mathbf{E}}$.

ii) Soit \mathbf{J} la matrice de taille 4×4 ne contenant que des 1. Quel est le rang de \mathbf{J} ? En déduire une valeur propre de \mathbf{J} .

iii) Montrer que le vecteur de taille 4 ne contenant que des 1 est vecteur propre de \mathbf{J} . Quelle est la valeur propre associée?

- iv) En déduire que les valeurs propres de J sont exactement 0 et 4. Est-elle diagonalisable ?
- v) Exprimer Γ en fonction de J et de I_4 , la matrice identité de taille 4×4 . En déduire les valeurs propres de la matrice Γ .
- vi) Justifier qu'il existe une matrice $D \in \mathcal{M}_4(\mathbf{R})$ diagonale, et $P \in \mathcal{M}_4(\mathbf{R})$ telles que $\Gamma = PD^T P$, ${}^T P P = I_4$. Compte-tenu des questions précédentes, préciser la forme de la matrice D .

14.3) i) Justifier que : $\sum_{i=1}^4 (\epsilon_i - \bar{\epsilon})^2 = {}^T E \cdot E$.

ii) En utilisant la relation $E = \Gamma \tilde{E}$, déduire que :

$$\sum_{i=1}^4 (\epsilon_i - \bar{\epsilon})^2 = \sum_{i=1}^4 \lambda_i^2 V_i^2,$$

où V_i désigne la i -ème composante du vecteur $V = {}^T P \tilde{E}$, et les λ_i sont les valeurs propres de Γ .

iii) Justifier que les variables aléatoires V_i suivent une loi normale. Préciser leur espérance.

14.4) On dit qu'une variable aléatoire suit une loi du $\chi^2(p)$ (loi du *chi deux* à $p \in \mathbf{N}^*$ degrés de liberté) si elle s'écrit comme une somme de p carrés de variables aléatoires suivant une loi $\mathcal{N}(0, 1)$. En admettant que les V_i sont des variables aléatoires indépendantes, et de variance σ^2 (on rappelle qu'elles sont également gaussiennes et d'espérance nulle d'après la question précédente), déduire que : $\frac{3\widehat{\sigma}^2}{\sigma^2} \hookrightarrow \chi^2(3)$.

15. Supposons maintenant que l'on a à notre disposition les observations des deux premières phases : « rosette » et « élongation ». En s'inspirant du critère des moindres carrés, proposer un critère pour estimer T_1 correspondant au temps thermique séparant les phases de rosette et d'élongation, défini dans la Figure 1.

16. Supposons plus généralement que l'on ait à notre disposition l'ensemble des observations c'est-à-dire celles correspondant aux trois phases. Proposer un critère pour estimer T_1 et T_2 correspondant aux temps thermiques séparant les différentes phases de développement définis dans la Figure 1.

PARTIE IV – DÉTECTION DES RUPTURES DE PENTE

17. Soit u une fonction connue dans $\mathcal{C}^4(\mathbf{R}, \mathbf{R})$ que l'on observe aux n temps thermiques x_1, \dots, x_n que l'on supposera deux à deux distincts. On notera u_i la valeur

de u en x_i , i.e. $u_i = u(x_i)$ pour tout $i \in \llbracket 1, n \rrbracket$. On supposera de plus que les données sont récoltées à des intervalles de temps réguliers $h > 0$ avec $h = x_{i+1} - x_i$ pour $1 \leq i \leq n - 1$.

17.1) Écrire un développement de TAYLOR-YOUNG de u à l'ordre 4 au point x_i , i.e. justifier l'existence d'un polynôme P à coefficients réels de degré 4 et d'une fonction $\epsilon : \mathbf{R} \rightarrow \mathbf{R}$ tels que :

$$u(x) = u(x_i) + P(x) + \epsilon(x)(x - x_i)^4, \quad \epsilon(x) \xrightarrow{x \rightarrow x_i} 0.$$

Préciser le polynôme P .

17.2) En appliquant le développement précédent à $x = x_{i+1}$ d'une part, et $x = x_{i-1}$ d'autre part, montrer que pour $2 \leq i \leq n - 1$:

$$u''(x_i) = \frac{u_{i-1} - 2u_i + u_{i+1}}{h^2} + \eta(h),$$

avec $\eta : \mathbf{R} \rightarrow \mathbf{R}$ une fonction à préciser.

17.3) Supposons que $u(x) = ax + b$ avec $a, b \in \mathbf{R}$ pour tout $x \in \mathbf{R}$, donc que u est une fonction affine. Montrer, par un calcul direct i.e. sans utiliser la formule de TAYLOR-YOUNG précédente, que : $u_{i-1} - 2u_i + u_{i+1} = 0$ pour tout $2 \leq i \leq n - 1$.

18. À partir de maintenant on va travailler sur un modèle simplifié où les valeurs expérimentales y_i , $1 \leq i \leq n$, sont portées par la fonction linéaire par morceaux de la Figure 1 et récoltées à des intervalles de temps réguliers d'amplitude constante $h = x_{i+1} - x_i$, $1 \leq i \leq n - 1$. Plus formellement, cela signifie qu'il existe $a_j, b_j \in \mathbf{R}$ pour $1 \leq j \leq 3$, des paramètres inconnus, tels que :

$$\begin{cases} y_i = a_1 + b_1 x_i & \text{si } x_i \leq T_1, \\ y_i = a_2 + b_2 x_i & \text{si } T_1 \leq x_i \leq T_2, \\ y_i = a_3 + b_3 x_i & \text{si } T_2 \leq x_i. \end{cases}$$

On notera aussi : $D_i = \frac{y_{i-1} - 2y_i + y_{i+1}}{h^2}$ pour $2 \leq i \leq n - 1$.

18.1) Donner la valeur de D_i lorsque $x_{i+1} \leq T_1$ en utilisant la première question de cette partie.

18.2) Proposer une méthode pour détecter les deux instants séparant les différentes phases.

19. Cette question ne nécessite pas d'avoir obtenu les résultats qui précèdent pour y répondre.

19.1) Que renvoie l'algorithme suivant, appliqué sur une liste L non vide :

Algorithme mystère

```
def Mystere(L):
    """
    L : liste non vide -> ???
    """
    n = len(L)
    p = 0
    m = L[0]
    for k in range(1, n):
        if m < L[k]:
            m = L[k]
            p = k
    return p
```

19.2) Si L est vide, indiquer avec justification la ou le(s) réponse(s) correcte(s) :

- i) Mystere(L) renvoie une erreur.
- ii) Mystere(L) renvoie le booléen False.
- iii) Mystere(L) renvoie 0.
- iv) Mystere(L) ne renvoie rien.

19.3) Dans la fonction Mystere, on peut éviter l'utilisation de la variable m (sans changer le résultat renvoyé lorsque L est non vide) : proposer, en expliquant le raisonnement suivi, une version modifiée de Mystere en ce sens. Que renvoie la fonction modifiée lorsque la liste entrée est vide ?

20. On dispose dorénavant d'une liste L de nombres, de longueur supérieure ou égale à deux, dont les éléments sont distincts deux à deux. On souhaite renvoyer les positions des deux plus grands éléments de L.

20.1) On suppose que l'on dispose d'un algorithme de tri par ordre croissant donné dans une fonction python tri. Proposer une fonction DeuxMaxTri permettant de répondre à la question, et prenant en argument une liste L : vous pourrez donc faire appel à la fonction tri dans cette fonction.

20.2) On souhaite désormais répondre sans utiliser d'algorithme de tri mais en

utilisant la fonction Mystere. On commence par calculer p1 = Mystere(L) et ensuite, on cherche le plus grand élément L[p2] de L tel que p2 est différent de p1. Recopier et compléter les trois lignes manquantes de la fonction suivante, afin qu'elle renvoie la liste [p1, p2] des positions des deux plus grands éléments de L.

Deux max

```
def DeuxMax(L):
    n = len(L)
    p1 = Mystere(L)
    p2 = 0
    for k in range(1, n):
        # compléter
        # compléter
    return [p1, p2]
```

20.3) Proposer une autre version appelée DeuxMax_bis, et ne faisant appel qu'à la fonction Mystere.

21. Dans la pratique les valeurs expérimentales sont bruitées et on a :

$$\begin{cases} y_i = a_1 + b_1 x_i + \varepsilon_i & \text{si } x_i \leq T_1, \\ y_i = a_2 + b_2 x_i + \varepsilon_i & \text{si } T_1 \leq x_i \leq T_2, \\ y_i = a_3 + b_3 x_i + \varepsilon_i & \text{si } T_2 \leq x_i, \end{cases}$$

où ε_i désigne un terme d'erreur ayant la forme particulière suivante $\varepsilon_i = (-1)^i \varepsilon$, avec $\varepsilon > 0$.

21.1) Calculer D_i dans le cas où $x_{i+1} \leq T_1$. On expliquera également pourquoi on ne peut plus appliquer directement la première question de cette partie.

21.2) En supposant ε suffisamment petit, proposer une méthode pour détecter les deux instants séparant les différentes phases.

CORRECTION

Solution (problème 1)

(Énoncé : 1)

PARTIE I — SUR L'AJUSTEMENT AFFINE DÉTERMINISTE AU SENS DES MOINDRES CARRÉS

1. Développons le carré du membre de droite :

$$\begin{aligned}\sum_{i=1}^n (x_i - \bar{x})^2 &= \sum_{i=1}^n (x_i^2 + \bar{x}^2 - 2x_i\bar{x}) \\ &= \sum_{i=1}^n x_i^2 + n\bar{x}^2 - 2\bar{x} \sum_{i=1}^n x_i \\ &= \sum_{i=1}^n x_i^2 + \bar{x} \sum_{i=1}^n x_i - 2\bar{x} \sum_{i=1}^n x_i \\ &= \sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i.\end{aligned}$$

2. Soit F la fonction définie par $F : \mathbf{R}^2 \rightarrow \mathbf{R}$ par : $F(a, b) = \sum_{i=1}^n (y_i - bx_i - a)^2$ pour tout $(a, b) \in \mathbf{R}^2$, soit $(a, b) \in \mathbf{R}^2$. On résout :

$$\begin{aligned}\begin{cases} \frac{\partial F(a,b)}{\partial a} = 0, \\ \frac{\partial F(a,b)}{\partial b} = 0 \end{cases} &\iff \begin{cases} \sum_{i=1}^n (-2)(y_i - bx_i - a) = 0, \\ \sum_{i=1}^n (-2x_i)(y_i - bx_i - a) = 0 \end{cases} \\ &\iff \begin{cases} -2n\bar{y} + 2bn\bar{x} + 2na = 0, \\ -2\sum_{i=1}^n x_i y_i + 2b\sum_{i=1}^n x_i^2 + 2an\bar{x} = 0 \end{cases} \\ &\iff \begin{cases} a + b\bar{x} = \bar{y}, \\ b\sum_{i=1}^n x_i^2 + an\bar{x} = \sum_{i=1}^n x_i y_i \end{cases} \\ &\iff \begin{cases} a + b\bar{x} = \bar{y}, \\ b = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}, \end{cases}\end{aligned}$$

où à la dernière étape nous avons réalisé l'opération $L_2 \leftarrow L_2 - (n\bar{x})L_1$.
Comme $\sum_{i=1}^n x_i y_i - n\bar{x} \cdot \bar{y} = \sum_{i=1}^n (x_i - \bar{x})^2$, nous obtenons le résultat :

l'unique solution du système précédent est (\hat{a}, \hat{b}) .

3. Comme $F(a, b)$ pour tout $(a, b) \in \mathbf{R}^2$ représente la somme des écarts au carré entre le nuage de points $((x_i, y_i))_i$ et la droite d'équation $y = bx + a$, on retrouve le fait que la droite d'équation $y = \hat{a} + \hat{b}x$ est celle qui passe au plus près du nuage au sens des moindres carrés.
4. 4.1) On initialise une variable à zéro, puis on parcourt la liste en ajoutant à la variable précédente chacun des éléments.

```
def moyenne(L):  
    """  
    Renvoie la moyenne des éléments de L  
    """  
    S = 0  
    for x in L:  
        S += x  
    return S/len(L)
```

- 4.2) Les réponses correctes sont les suivantes : 1, 4.
4.3) On utilise la fonction moyenne pour calculer efficacement les deux sommes du quotient.

```
def Bchap(x, y):  
    """  
    Renvoie l'ordonnée à l'origine de la droite des  
    moindres carrés  
    """  
    Moy_x = moyenne(x)  
    Num = len(x)*moyenne([y[i]*x[i] - y[i]*Moy_x for i in  
        range(len(x))])  
    Den = len(x)*moyenne([(x[i] - Moy_x)**2 for i in  
        range(len(x))])  
    return Num / Den
```

- 4.4) On utilise la fonction précédente, et la fonction moyenne.

```
def Achap(x, y):
    """
    Renvoie la pente de la droite des moindres carrés
    """
    Moy_x = moyenne(x)
    Moy_y = moyenne(y)
    return Moy_y - Bchap(x,y)*Moy_x
```

Par exemple, Achap([1, 3], [2, 1]) retourne 2.5.

5. >_☒

5.1)

```
def extraction_ch():
    L = []
    f = open("data.txt", "r")
    ligne = f.readline()
    while len(ligne) != 0:
        L.append(ligne)
        ligne = f.readline()
    return L
```

Ici, nous obtenons

```
>>> extraction_ch()
['10 20\n', '12 27']
```

5.2) Compléter la fonction suivante d'entête `ch_vers_list(ch)` prenant en entrée une chaîne de caractère (de même forme que celles renvoyées par `extraction_ch`) et qui renvoie la liste de nombres correspondant; ces nombres seront de type flottant. Par exemple, `ch_vers_list('10 20')` renvoie `[10.0, 20.0]`.

```
def ch_vers_list(ch):
    L = []
    n = len(ch)
    sh = ""
```

```
for k in range(0, n):
    if ch[k] != " ":
        sh = sh+ch[k]
    else:
        L.append(float(sh)) #on ajoute la tranche
        ← d'elts non vides trouvée, qui est terminée
        sh = "" # on recommence une nouvelle tranche
L.append(float(sh))
return L
```

5.3)

```
def extraction_data():
    L = extraction_ch()
    x = []
    y = []
    for ch in L:
        M = ch_vers_list(ch)
        x.append(M[0])
        y.append(M[1])
    return x, y
```

La fonction retourne bien les deux listes désirées.

```
>>> extraction_data()
([10.0, 12.0], [20.0, 27.0])
```

PARTIE II — PRISE EN COMPTE D'UN ALÉATOIRE DE VARIABILITÉ EXPÉRIMENTALE

6. Soit $i \in \llbracket 1, n \rrbracket$, $\mathbf{E}(Y_i) = \mathbf{E}(a + bx_i + \varepsilon_i) = a + bx_i + 0$ par linéarité de l'espérance et

puisque les ε_i sont centrées. Donc : $\mathbf{E}(Y_i) = a + bx_i$. De plus,

$$\begin{aligned} \mathbf{E}\left(\sum_{i=1}^n Y_i(x_i - \bar{x})\right) &= \sum_{i=1}^n \mathbf{E}(Y_i)(x_i - \bar{x}) = \sum_{i=1}^n (a + bx_i)(x_i - \bar{x}), \\ &= a \sum_{i=1}^n (x_i - \bar{x}) + b \sum_{i=1}^n x_i(x_i - \bar{x}), \\ &= a(n\bar{x} - n\bar{x}) + b \left(\sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i \right) \\ &= \boxed{b \sum_{i=1}^n (x_i - \bar{x})^2}, \end{aligned}$$

où à la dernière étape nous avons utilisé la première question.

7. Toujours par linéarité de l'espérance, nous avons :

$$\mathbf{E}(\hat{b}) = \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} \mathbf{E}\left(\sum_{i=1}^n Y_i(x_i - \bar{x})\right) = \frac{b \sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = b.$$

Dit autrement $\mathbf{E}(\hat{b}) = b$ donc \hat{b} est non biaisé.

8. On a par linéarité de l'espérance : $\mathbf{E}(\bar{Y}) = \frac{1}{n} \sum_{i=1}^n \mathbf{E}(Y_i) = \frac{1}{n} \sum_{i=1}^n (a + bx_i) = \boxed{a + b\bar{x}}$. Encore une fois par linéarité de l'espérance, nous avons :

$$\mathbf{E}(\hat{a}) = a + b\bar{x} - b\bar{x} = a.$$

Donc \hat{a} est un estimateur non biaisé de a .

9. En réduisant au même dénominateur les deux fractions, on obtient :

$$\begin{aligned} \hat{b} - b &= \frac{\sum_{i=1}^n Y_i(x_i - \bar{x}) - b \sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \frac{\sum_{i=1}^n (Y_i - b(x_i - \bar{x}))(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \frac{\sum_{i=1}^n (a + bx_i + \varepsilon_i - b(x_i - \bar{x}))(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \frac{\sum_{i=1}^n (a + \varepsilon_i + b\bar{x})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} = (a + b\bar{x}) \frac{\sum_{i=1}^n (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} + \frac{\sum_{i=1}^n \varepsilon_i(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= (a + b\bar{x}) \frac{n\bar{x} - n\bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2} + \frac{\sum_{i=1}^n \varepsilon_i(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}. \end{aligned}$$

On a bien montré la formule

$$\hat{b} - b = \frac{\sum_{i=1}^n \varepsilon_i(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

10. Par propriété de la variance (invariance par translation), on obtient : $\mathbf{Var}(\hat{b} - b) =$

$$\mathbf{Var}(\hat{b} - b + b) = \boxed{\mathbf{Var}(\hat{b})}.$$

$$\begin{aligned} \mathbf{Var}(\hat{b} - b) &= \left(\frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)^2 \mathbf{Var}\left(\sum_{i=1}^n \varepsilon_i(x_i - \bar{x})\right) \\ &= \left(\frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)^2 \sum_{i=1}^n \mathbf{Var}(\varepsilon_i)(x_i - \bar{x})^2 \\ &= \boxed{\frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}. \end{aligned}$$

11. 11.1) Supposons que $x_i = i$ pour tout $i \in \llbracket 1, n \rrbracket$. Alors $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} \sum_{i=1}^n i =$

$\frac{n+1}{2}$. Donc :

$$\begin{aligned} \sum_{i=1}^n (x_i - \bar{x})^2 &= \sum_{i=1}^n \left(i - \frac{n+1}{2} \right)^2 \\ &= \sum_{i=1}^n \left(i^2 + \left(\frac{n+1}{2} \right)^2 - (n+1)i \right) \\ &= \frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)^2}{4} - \frac{n(n+1)^2}{2} \\ &= \boxed{\frac{n(n+1)(n-1)}{12}}. \end{aligned}$$

11.2) Comme \hat{b} est d'espérance b , on peut appliquer l'inégalité de BIENAYMÉ-TCHEBYCHEV, pour tout $\varepsilon > 0$:

$$\mathbf{P}(|\hat{b} - b| \geq \varepsilon) \leq \frac{\mathbf{Var}(\hat{b})}{\varepsilon^2}.$$

Or, d'après la question précédente, $\mathbf{Var}(\hat{b}) = \frac{12\sigma^2}{n(n+1)(n-1)} \xrightarrow{n \rightarrow \infty} 0$. Donc

\hat{b} converge vers b .

12. 12.1) C'est un calcul direct : $\tilde{b} - \hat{b} = \sum_{i=1}^n \mu_i Y_i - \frac{\sum_{i=1}^n Y_i (x_i - \bar{x})}{\sum_{j=1}^n (x_j - \bar{x})^2} =$

$$\sum_{i=1}^n Y_i \left[\mu_i - \frac{x_i - \bar{x}}{\sum_{j=1}^n (x_j - \bar{x})^2} \right].$$

12.2) On en déduit alors, par indépendance des Y_i , que :

$$\begin{aligned} \mathbf{Var}(\tilde{b} - \hat{b}) &= \sum_{i=1}^n \sigma^2 \left[\mu_i - \frac{x_i - \bar{x}}{\sum_{j=1}^n (x_j - \bar{x})^2} \right]^2 \\ &= \sum_{i=1}^n \sigma^2 \left[\mu_i^2 - 2\mu_i \frac{x_i - \bar{x}}{\sum_{j=1}^n (x_j - \bar{x})^2} + \frac{(x_i - \bar{x})^2}{(\sum_{j=1}^n (x_j - \bar{x})^2)^2} \right] \\ &= \sigma^2 \left[\sum_{i=1}^n \mu_i^2 - 2 \frac{\sum_{i=1}^n \mu_i (x_i - \bar{x})}{\sum_{j=1}^n (x_j - \bar{x})^2} + \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{(\sum_{j=1}^n (x_j - \bar{x})^2)^2} \right] \end{aligned}$$

Par ailleurs, nous avons : $\sum_{i=1}^n \mu_i (x_i - \bar{x}) = \sum_{i=1}^n \mu_i x_i - \bar{x} \sum_{i=1}^n \mu_i = 1 - 0 = 1$ par hypothèse. Donc

$$\begin{aligned} \mathbf{Var}(\tilde{b} - \hat{b}) &= \sigma^2 \left[\sum_{i=1}^n \mu_i^2 - \frac{2}{\sum_{j=1}^n (x_j - \bar{x})^2} + \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{(\sum_{j=1}^n (x_j - \bar{x})^2)^2} \right] \\ &= \sigma^2 \sum_{i=1}^n \mu_i^2 + \sigma^2 \frac{\sum_{i=1}^n (x_i - \bar{x})^2 - 2 \sum_{i=1}^n (x_i - \bar{x})^2}{(\sum_{j=1}^n (x_j - \bar{x})^2)^2} \\ &= \sigma^2 \sum_{i=1}^n \mu_i^2 - \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}. \end{aligned}$$

Or, nous avons établi que $\mathbf{Var}(\hat{b}) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$, et par indépendance $\mathbf{Var}(\tilde{b}) = \sum_{i=1}^n \mu_i^2 \mathbf{Var}(Y_i) = \sigma^2 \sum_{i=1}^n \mu_i^2$. Dès lors, il vient la formule souhaitée :

$$\mathbf{Var}(\tilde{b} - \hat{b}) = \mathbf{Var}(\tilde{b}) - \mathbf{Var}(\hat{b}).$$

Puisque $\mathbf{Var}(\tilde{b} - \hat{b}) \geq 0$ par propriétés de la variance, nous obtenons immédiatement $\mathbf{Var}(\tilde{b}) \geq \mathbf{Var}(\hat{b})$.

12.3) Il s'agit de montrer plusieurs choses :

- ▶ \hat{b} est une combinaison linéaire des variables aléatoires Y_1, \dots, Y_n : c'est bien le cas en prenant $\mu_i = \frac{x_i - \bar{x}}{\sum_{j=1}^n (x_j - \bar{x})^2}$ pour tout $i \in \llbracket 1, n \rrbracket$.
- ▶ on calcule la première somme : $\sum_{i=1}^n \mu_i = \sum_{i=1}^n \frac{x_i - \bar{x}}{\sum_{j=1}^n (x_j - \bar{x})^2} = \sum_{i=1}^n \frac{n\bar{x} - n\bar{x}}{\sum_{j=1}^n (x_j - \bar{x})^2} = 0$.
- ▶ on calcule la seconde somme : $\sum_{i=1}^n \mu_i x_i = \sum_{i=1}^n \frac{x_i^2 - \bar{x}x_i}{\sum_{j=1}^n (x_j - \bar{x})^2} = \frac{\sum_{i=1}^n x_i^2 - n\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = 1$ d'après la première question.

En conclusion : $\hat{b} \in \mathcal{E}$.

12.4) L'estimateur \hat{b} est non biaisé, il appartient à \mathcal{E} (nous venons de la montrer), et il minimise la variance d'après une question précédente.

PARTIE III – CAS DE BRUITS GAUSSIENS

13. Nous avons déjà dit que \hat{b} était une combinaison linéaire de lois normales, donc par indépendance des Y_i elle suit encore une loi normale de paramètres la moyenne et la variance. Ainsi, comme $\mathbf{E}(\hat{b}) = b$ et $\mathbf{Var}(\hat{b}) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$, nous avons :

$$\hat{b} \hookrightarrow \mathcal{N}\left(b, \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right).$$

14. 14.1) Par simple lecture graphique nous avons $b = 0$ (la pente de la phase considérée est nulle). Soit $i \in \llbracket 1, n \rrbracket$, alors $Y_i - \bar{Y} = a + bx_i + \varepsilon_i - \frac{1}{n}(na + nb\bar{x} + n\bar{\varepsilon}) = b(x_i - \bar{x}) + \varepsilon_i - \bar{\varepsilon} = \varepsilon_i - \bar{\varepsilon}$ puisque $b = 0$. On obtient dès lors :

$$\widehat{\sigma^2} = \frac{1}{n-1} \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2.$$

14.2) i) C'est un calcul direct : $\Gamma \tilde{\mathbf{E}} = \begin{pmatrix} \varepsilon_1 - \frac{1}{4} \sum_{i=1}^4 \varepsilon_i \\ \varepsilon_2 - \frac{1}{4} \sum_{i=1}^4 \varepsilon_i \\ \varepsilon_3 - \frac{1}{4} \sum_{i=1}^4 \varepsilon_i \\ \varepsilon_4 - \frac{1}{4} \sum_{i=1}^4 \varepsilon_i \end{pmatrix} = \mathbf{E}.$

ii) Comme toutes les colonnes de \mathbf{J} sont identiques, nous avons $\text{Rg} \mathbf{J} = 1$, donc par le théorème du rang $\dim(\text{Ker}(\mathbf{J})) = 3 = \dim E_0(\mathbf{J})$, donc $0 \in \text{Spec} \mathbf{J}$.

iii) Notons $\mathbb{1}$ le vecteur colonne ne contenant que des 1. Alors $\mathbf{J}\mathbb{1} = 4\mathbb{1}$, donc $4 \in \text{Spec} \mathbf{J}$ et $\mathbb{1}$ est un vecteur propre associé.

iv) Nous avons $\dim E_0(\mathbf{J}) = 3$, $\dim E_4(\mathbf{J}) \geq 1$ et $\dim E_0(\mathbf{J}) + \dim E_4(\mathbf{J}) \leq 4$ donc la somme des dimensions des espaces propres est égale à 4, donc \mathbf{J} est diagonalisable, ses valeurs propres sont 0, 4.

v) Nous avons $\Gamma = \mathbf{I}_4 - \frac{1}{4}\mathbf{J}$. Comme $\mathbf{J} = \mathbf{PDP}^{-1}$ avec \mathbf{P} inversible et $\mathbf{D} = \text{Diag}(4, 0, 0, 0)$, on déduit que $\Gamma = \mathbf{PP}^{-1} - \mathbf{PDP}^{-1} = \mathbf{P} \text{Diag}(0, 1, 1, 1) \mathbf{P}^{-1}$, donc Γ est diagonalisable avec $\text{Spec} \Gamma = \{0, 1\}$.

vi) C'est une conséquence directe du théorème spectral puisque \mathbf{J} est symétrique réelle. De plus, d'après les questions précédentes, nous avons $\mathbf{D} = \text{Diag}(0, 1, 1, 1)$.

14.3) i) C'est un calcul direct.

ii) Nous avons

$$\begin{aligned} \sum_{i=1}^4 (\varepsilon_i - \bar{\varepsilon})^2 &= {}^t \mathbf{E} \mathbf{E} = {}^t (\Gamma \tilde{\mathbf{E}}) \Gamma \tilde{\mathbf{E}} = {}^t \tilde{\mathbf{E}} \left[{}^t (\mathbf{PD}^t \mathbf{P}) \mathbf{PD}^t \mathbf{P} \right] \tilde{\mathbf{E}} \\ &= {}^t \tilde{\mathbf{E}} \mathbf{PD}^2 \mathbf{P} \tilde{\mathbf{E}} \\ &= (V_1, \dots, V_4) \mathbf{D}^2 \begin{pmatrix} V_1 \\ \vdots \\ V_4 \end{pmatrix}. \end{aligned}$$

En effectuant le dernier produit matriciel, et notant $\mathbf{D}^2 =$

$$\begin{pmatrix} \lambda_1^2 & 0 & 0 & 0 \\ 0 & \lambda_2^2 & 0 & 0 \\ 0 & 0 & \lambda_3^2 & 0 \\ 0 & 0 & 0 & \lambda_4^2 \end{pmatrix}, \text{ on obtient la formule souhaitée :}$$

$$\sum_{i=1}^4 (\varepsilon_i - \bar{\varepsilon})^2 = \sum_{i=1}^4 \lambda_i^2 V_i^2.$$

iii) La variable aléatoire V_i est la i -ème coordonnée de ${}^t \mathbf{P} \tilde{\mathbf{E}}$, donc une combinaison linéaire de lois normales (les coefficients de ladite combinaison dépendent des coefficients de la matrice \mathbf{P}) indépendantes (les ε_i). Donc $V_i \hookrightarrow \mathcal{N}(\mathbf{E}(V_i), \mathbf{Var}(V_i))$. Comme les ε_i sont centrées,

$$\mathbf{E}(V_i) = 0 \text{ (par linéarité de l'espérance).}$$

iv) On suppose que $V_i \hookrightarrow \mathcal{N}(0, \sigma^2)$ et sont indépendantes. Pour $n = 4$,

$$\begin{aligned} 3 \frac{\widehat{\sigma^2}}{\sigma^2} &= 3 \frac{\frac{1}{3} \sum_{i=1}^4 (\varepsilon_i - \bar{\varepsilon})^2}{\sigma^2} \\ &= \frac{\sum_{i=1}^4 \lambda_i^2 V_i^2}{\sigma^2} = \sum_{i=1}^4 \left(\lambda_i \frac{V_i}{\sigma} \right)^2 \end{aligned}$$

Reprenons les informations sur les valeurs propres des questions précédentes : $\lambda_0 = 0, \lambda_i = 1$ pour $i \in \llbracket 2, 4 \rrbracket$. Alors

$$3 \frac{\widehat{\sigma^2}}{\sigma^2} = \sum_{i=2}^4 \left(\frac{V_i}{\sigma} \right)^2.$$

Or, $\frac{V_i}{\sigma} \hookrightarrow \mathcal{N}(0, 1)$ par propriétés de stabilité de la loi normale, donc

$$3 \frac{\widehat{\sigma^2}}{\sigma^2} \hookrightarrow \chi^2(3).$$

- 15.** On effectue une régression linéaire sur les deux premiers points sur les k premiers points, une autre sur les $n - k$ derniers pour $k \in \llbracket 1, n \rrbracket$. On obtient alors deux coefficients de régressions (mesurant la qualité) r_k^2 et R_k^2 correspondant respectivement à la première droite de régression et à la seconde. Notons $\rho_k^2 = r_k^2 + R_k^2$. On peut alors chercher l'entier k qui minimise la suite $(\rho_k^2)_k$, le réel x_k correspondant fournira donc alors une approximation du temps thermique T_1 .
- 16.** On applique le même protocole que précédemment mais en prenant un bloc de k points, de ℓ points puis de $n - k - \ell$, avec $2 \leq k, \ell \leq n - 1$. On obtient alors trois coefficients de qualité de régression, on cherche (k, ℓ) qui maximise la somme de ces trois coefficients de régression.

PARTIE IV – DÉTECTION DES RUPTURES DE PENTE.

- 17. 17.1)** C'est une conséquence de la formule de TAYLOR-YOUNG à l'ordre 4, puisque la fonction u est \mathcal{C}^4 .

$$u(x) = u(x_i) + u'(x_i)(x - x_i) + \frac{u''(x_i)}{2}(x - x_i)^2 + \frac{u'''(x_i)}{6}(x - x_i)^3 + \frac{u^{(4)}(x_i)}{4!}(x - x_i)^4 + \varepsilon(x)(x - x_i)^4.$$

On note donc

$$P(x) = u'(x_i)(x - x_i) + \frac{u''(x_i)}{2}(x - x_i)^2 + \frac{u'''(x_i)}{6}(x - x_i)^3 + \frac{u^{(4)}(x_i)}{4!}(x - x_i)^4.$$

- 17.2)** On a, toujours d'après la formule de TAYLOR-YOUNG :

$$u(x_{i+1}) = u(x_i) + u'(x_i)h + \frac{u''(x_i)}{2}h^2 + \frac{u'''(x_i)}{6}h^3 + \frac{u^{(4)}(x_i)}{4!}h^4 + \varepsilon(x_{i+1})h^4$$

$$u(x_{i-1}) = u(x_i) - u'(x_i)h + \frac{u''(x_i)}{2}h^2 - \frac{u'''(x_i)}{6}h^3 + \frac{u^{(4)}(x_i)}{4!}h^4 + \varepsilon(x_{i-1})h^4$$

Il ne reste plus qu'à faire la somme des deux égalités, on obtient l'égalité souhaitée :

$$u''(x_i) = \frac{u_{i-1} - 2u_i + u_{i+1}}{h^2} + \eta(h),$$

avec $\eta(h) = \varepsilon(x_{i+1})h^4 + \varepsilon(x_{i-1})h^4 + 2\frac{u^{(4)}(x_i)}{4!}h^4$.

- 17.3)** Soit $x \in \mathbf{R}$. On a $u(x_{i+1}) = u(x_i + h) = a(x_i + h) + b$, $u(x_i) = ax_i + b$, $u(x_{i-1}) = a(x_i - h) + b$. Donc

$$u_{i-1} - 2u_i + u_{i+1} = a(x_i - h) + b - 2(ax_i + b) + a(x_i + h) + b = 0.$$

- 18. 18.1)** Lorsque $x_{i+1} \leq T_1$, tous les points x_1, \dots, x_{i+1} précèdent le premier temps thermique T_1 , donc d'après ce qui précède $D_i = 0$.
- 18.2)** On calcule les valeurs D_1, D_2, \dots jusqu'au premier indice i_1 tel que $D_{i_1} \neq 0$. Alors $T_1 = i_1 h$, le deuxième tel indice donnera alors T_2 .
- 19. 19.1)** La fonction retourne l'indice du terme le plus grand de la liste. Si ce maximum est atteint en plusieurs indices, elle retourne le dernier indice.
- 19.2)** Elle retourne une erreur dès le départ si L est vide : à cause de l'instruction $m = L[0]$.
- 19.3)** Il suffit d'utiliser $L[p]$ directement plutôt que le stockage dans une variable.

Algorithme qui n'est plus un mystère

```
def Mystere(L):
    n = len(L)
    p = 0
    for k in range(1, n):
        if L[p] < L[k]:
            p = k
    return p
```

Si la liste est vide, alors la fonction retourne 0 : la boucle for n'est pas parcourue puisque $0 < 1$.

- 20.** On utilise la fonction de tri, on récupère ensuite simplement les deux derniers éléments. On suppose que le tri ne se fait pas en place.

```
20.1) def DeuxMaxTri(L):
    L_trie = tri_insertion(L) # pour l'exemple
    Max_1, Max_2 = L_trie[-2:]
    for i in range(len(L)):
        if L[i] == Max_1:
            pos_max_1 = i
```

```

elif L[i] == Max_2:
    pos_max_2 = i
return pos_max_1, pos_max_2
>>> DeuxMaxTri([1,5,3,2])
(2, 1)

```

20.2) On pourra aussi appeler une seconde fois la fonction Mystere.

Deux max

```

def DeuxMax(L):
    n = len(L)
    p1 = Mystere(L)
    p2 = 0
    for k in range(1, n):
        if k != p1 and L[k] > L[p2]:
            p2 = k
    return [p1, p2]
>>> DeuxMax([1,5,3,2])
[1, 2]

```

20.3)

Deux max bis

```

def DeuxMaxBis(L):
    p1 = Mystere(L)
    del L[p1]
    p2 = Mystere(L)
    if p2 > p1: # dans ce cas, ce n'est pas le bon à cause
        ↪ de la suppression (décalage)
        p2 += 1
    return [p1, p2]

```

21. 21.1) On ne peut appliquer directement la première question, puisque sur chaque sous-intervalles, les y_i ne dépendent pas des x_i de manière affine avec les mêmes coefficients (ceci est dû aux bruits que l'on introduit). Sup-

posons que $x_{i+1} \leq T_1$ pour un certain i . Alors :

$$D_i = \frac{(-1)^{i-1} - 2(-1)^i + (-1)^{i+1}}{h^2} \varepsilon = \frac{4(-1)^{i-1}}{h^2} \varepsilon.$$

21.2) Supposons que $T_2 \leq x_{i+1} > T_1$ mais $x_i \leq T_1$ alors, en utilisant comme avant $x_{i\pm 1} = x_i \pm h$, $\varepsilon_i = (-1)^i \varepsilon$, nous obtenons :

$$D_i = \frac{b_2 - b_1}{h} + 2 \frac{\varepsilon(-1)^{-1}}{h^2}.$$

Donc lorsque les points x_{i-1}, x_i, x_{i+1} (et les y associés) se trouvent sur une même pente, alors D_i est très proche de zéro, en revanche il y a un saut dès que x_{i+1} devient supérieur au premier temps thermique T_1 . Ainsi si D_{i_1} est bien différent de zéro, une bonne estimation de T_1 sera x_{i_1-1} .