

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).

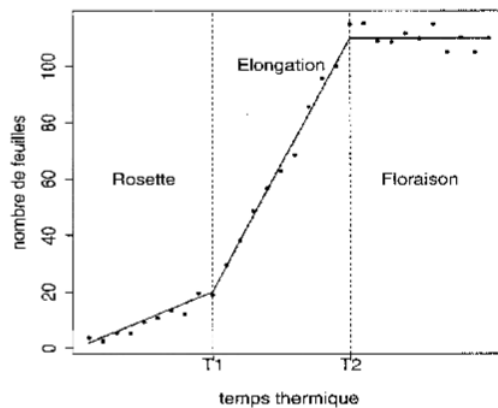


FIGURE 1 – Schéma illustrant l'évolution du nombre de feuilles d'une plante en fonction du temps thermique ; y sont représentés les valeurs expérimentales ('•') et les ajustements affines dans chacune des trois phases.

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 (T1) et d'autre part les phases d'élongation et de floraison (T2). 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 1

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 $((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n))$, 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).

1. On note $\bar{x} = (\sum_{i=1}^n x_i)/n$. Montrer que

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

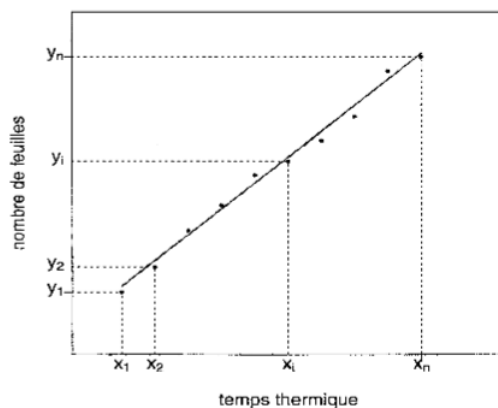


FIGURE 2 – Schéma illustrant l'évolution du nombre de feuilles en fonction du temps thermique durant l'une des trois phases.

2. On définit la fonction $F : \mathbb{R}^2 \rightarrow \mathbb{R}$ par :

$$\forall (a, b) \in \mathbb{R}^2, F(a, b) = \sum_{i=1}^n (y_i - a - bx_i)^2.$$

Montrer que le système suivant

$$\begin{cases} \frac{\partial F}{\partial a}(a, b) = 0, \\ \frac{\partial F}{\partial b}(a, b) = 0. \end{cases}$$

admet une unique solution $(\hat{a}, \hat{b}) \in \mathbb{R}^2$ où

$$\hat{a} = \bar{y} - \hat{b}\bar{x},$$

avec $\bar{y} = (\sum_{i=1}^n y_i)/n$ et

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

3. On admet que le couple (\hat{a}, \hat{b}) obtenu réalise un minimum global. Interpréter le résultat obtenu.
4. À partir des valeurs de (x_i, y_i) données dans le tableau ci-dessous, calculer à l'aide de la calculatrice les valeurs correspondantes de \hat{a} et de \hat{b} pour cet exemple.

x_i	2	4	6	8	10	12	14	16	18	20
y_i	8	13	20	27	32	38	44	50	56	62

5. 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 .

Consigne. Chaque algorithme, à écrire en Python, doit être précédé d'une phrase expliquant le raisonnement suivi pour l'écrire.

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

PARTIE 2

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 et 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 sont des variables aléatoires indépendantes d'espérance nulle et de variance σ^2 . Ceci correspond à un modèle de régression linéaire simple.

En utilisant la même démarche que celle proposée dans la partie 1, nous définissons :

$$\hat{a} = \bar{Y} - \hat{b}\bar{x},$$

où $\bar{Y} = (\sum_{i=1}^n Y_i)/n$, $\bar{x} = (\sum_{i=1}^n x_i)/n$ et

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

Le but de cette partie est d'étudier les propriétés des variables aléatoires \hat{a} et \hat{b} . Ces variables aléatoires sont appelées des estimateurs de a et b .

1. On note $\mathbb{E}(X)$ l'espérance d'une variable aléatoire X . Calculer $\mathbb{E}(Y_i)$ où $i \in \{1, \dots, n\}$.
2. Calculer $\mathbb{E}(\sum_{i=1}^n Y_i(x_i - \bar{x}))$.

3. En déduire que \widehat{b} est un estimateur sans biais de b où on dira qu'un estimateur est sans biais lorsque $\mathbb{E}(\widehat{b} - b) = 0$. On pourra utiliser le résultat de la question 1 de la partie 1 ici admis :

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

4. Calculer $\mathbb{E}(\widehat{Y})$. En déduire que \widehat{a} est un estimateur sans biais de a .
5. Montrer que

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

6. On note $\text{Var}(X)$ la variance de la variable aléatoire X . Montrer que :

$$\text{Var}(\widehat{b}) = \text{Var}(\widehat{b} - b).$$

7. En déduire que :

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

8. Dans cette question, nous nous intéressons au comportement de l'estimateur \widehat{b} en fonction du nombre d'observations n et nous le notons \widehat{b}_n . On dit que \widehat{b}_n converge en probabilité vers b lorsque :

$$\forall \varepsilon > 0, \mathbb{P}(\widehat{b}_n - b \geq \varepsilon) \rightarrow 0, \text{ quand } n \text{ tend vers l'infini.}$$

- (a) Calculer $\sum_{i=1}^n (x_i - \bar{x})^2$ lorsque $x_i = i$, où $i \in \{1, \dots, n\}$.
(b) Déduire en utilisant la question 7, que lorsque $x_i = i$, \widehat{b}_n converge en probabilité vers b .
9. Soit \widetilde{b} une variable aléatoire définie comme une combinaison linéaire des Y_i par $\widetilde{b} = \sum_{i=1}^n \mu_i Y_i$ où $\sum_{i=1}^n \mu_i = 0$ et $\sum_{i=1}^n \mu_i x_i = 1$.

- (a) Montrer que :

$$\text{Var}(\widetilde{b}) = \text{Var}(\widetilde{b} - \widehat{b}) + \text{Var}(\widehat{b}).$$

- (b) En déduire que :

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

10. Parmi les estimateurs sans biais et définis comme une combinaison linéaire des Y_i , \widehat{b} est de variance minimale. Justifier.

PARTIE 3

On se place dans cette partie dans le même cadre que celui de la partie 2 mais on suppose de plus que les variables aléatoires ε_i sont indépendantes et identiquement distribuées de loi gaussienne d'espérance nulle et de variance σ^2 notée $\mathcal{N}(0, \sigma^2)$.

1. Donner la loi de probabilité de \widehat{b} défini par

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

en utilisant les résultats des questions 3 et 7 de la partie 2 ici admis.