MTH 8302 - Modèles de Régression et d'Analyse de Variance

Série d'Exercises 1 : Préparation pour le Devoir 1 et Test t pour la Régression Linéaire

Polytechnique Montréal - Hiver 2025

Chiheb Trabelsi

February 4, 2025

Table des Matières

Problème 1 : Gradients et Hessiennes

Problème 1 : Gradients et Hessiennes

- **Rappel :**
  - Une matrice $\mathbf{A} \in \mathbb{R}^{n \times n}$ est symétrique si $\mathbf{A}^{\mathrm{T}} = \mathbf{A}$.
  - Le gradient $\nabla f(x)$ d'une fonction $f : \mathbb{R}^n \to \mathbb{R}$ est un vecteur contenant les dérivées partielles :

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix} = \left(\frac{\partial f}{\partial \mathbf{x}} \right).$$

  - La Hessienne $\nabla^2 f(\mathbf{x})$ est une matrice symétrique $n \times n$ contenant les dérivées secondes :

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_2} & \cdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_2^2} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}.$$

Règles de Dérivées Partielles :

- **Dérivée d'une forme quadratique :** Si $\mathbf{x}$ est un vecteur et $\mathbf{A}$ est une matrice symétrique, la dérivée de la forme quadratique $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$ par rapport à $\mathbf{x}$ est donnée par :

$$\frac{\partial}{d\mathbf{x}}(\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}) = 2\mathbf{A}\mathbf{x}$$

- **Dérivée d'une forme linéaire :** Pour un vecteur $\mathbf{x}$ et une matrice constante $\mathbf{A}$, la dérivée de $\mathbf{A}\mathbf{x}$ par rapport à $\mathbf{x}$ est :

$$\frac{\partial}{\partial\mathbf{x}}(\mathbf{A}\mathbf{x}) = \mathbf{A}$$

  où $\mathbf{A}$ est considéré comme une constante par rapport à $\mathbf{x}$.

- **Dérivée d'un produit de type vecteur-matrice-vecteur :** Si $\mathbf{b}$ et $\mathbf{x}$ sont des vecteurs et $\mathbf{A}$ est une matrice, alors la dérivée de $\mathbf{b}^{\mathrm{T}}\mathbf{A}\mathbf{x}$ par rapport à $\mathbf{x}$ est :

$$\frac{\partial}{\partial\mathbf{x}}(\mathbf{b}^{\mathrm{T}}\mathbf{A}\mathbf{x}) = \mathbf{A}^{\mathrm{T}}\mathbf{b}$$

Règles de Dérivation :

- **Dérivée d'une matrice dépendant d'un vecteur :** Si $\mathbf{A}$ est une matrice qui ne dépend pas du vecteur $\mathbf{x}$ et $\mathbf{A}$ n'est pas nécessairement symétrique, alors la dérivée de $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$ est :

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}) = \mathbf{x}^{\mathrm{T}}(\mathbf{A} + \mathbf{A}^{\mathrm{T}})$$

si $\mathbf{A}$ n'est pas nécessairement symétrique.

- **Dérivée du produit de deux matrices :** Si $\mathbf{A}(\mathbf{x})$ et $\mathbf{B}(\mathbf{x})$ sont des matrices dépendant du vecteur $\mathbf{x}$, alors la dérivée de leur produit $\mathbf{A}(\mathbf{x})\mathbf{B}(\mathbf{x})$ est donnée par :

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{A}(\mathbf{x})\mathbf{B}(\mathbf{x})) = \frac{\partial \mathbf{A}}{\partial \mathbf{x}}\mathbf{B} + \mathbf{A}\frac{\partial \mathbf{B}}{\partial \mathbf{x}}$$

Problème 1 : Question 1

- **Fonction :** $g(\mathbf{x}) = \log(1 + \mathbf{x}^\top \mathbf{C}\mathbf{x}) + \mathbf{d}^\top \mathbf{x}, \quad$ où $\mathbf{C}$ symmétrique.
- **Objectif :** Trouver $\nabla_{\mathbf{x}} g(\mathbf{x})$
- **Solution :**
- **Justification Détaillée :**

Problème 1 : Question 1 (Solution)

- **Fonction :** $g(\mathbf{x}) = \log(1 + \mathbf{x}^\top \mathbf{C}\mathbf{x}) + \mathbf{d}^\top \mathbf{x}$
- **Objectif :** Trouver $\nabla_{\mathbf{x}} g(\mathbf{x})$
- **Solution :** $\nabla_{\mathbf{x}} g(\mathbf{x}) = \frac{2\mathbf{C}\mathbf{x}}{1+\mathbf{x}^\top \mathbf{C}\mathbf{x}} + \mathbf{d}$
- **Justification Détaillée :**
  1. **Décomposition de la Fonction :**
     - La fonction g se compose de deux termes : un terme logarithmique $\log(1 + \mathbf{x}^\top \mathbf{C}\mathbf{x})$ et un terme linéaire $\mathbf{d}^\top \mathbf{x}$.
  2. **Dérivation du Terme Logarithmique :**
     - Utilisons la règle de la chaîne pour la dérivation du terme logarithmique. Soit $h = 1 + \mathbf{x}^\top \mathbf{C}\mathbf{x}$, la dérivée de h par rapport à $\mathbf{x}$ est $2\mathbf{C}\mathbf{x}$ si $\mathbf{C}$ est symétrique.
     - En combinant ces résultats, la contribution de la partie logarithmique au gradient est $\frac{2\mathbf{C}\mathbf{x}}{1+\mathbf{x}^\top \mathbf{C}\mathbf{x}}$.
  3. **Dérivation du Terme Linéaire :**
     - La dérivée du terme linéaire $\mathbf{d}^\top \mathbf{x}$ est simple et puisque la dérivée de $\mathbf{x}$ par rapport à lui-même est l'identité.
  4. **Combinaison des Gradients :**
     - Le gradient total de $g(\mathbf{x})$ est la somme des gradients des deux termes. Donc, $\nabla_{\mathbf{x}} g(\mathbf{x}) = \frac{2\mathbf{C}\mathbf{x}}{1+\mathbf{x}^\top \mathbf{C}\mathbf{x}} + \mathbf{d}$.

Problème 1 : Question 2

- **Fonction :** $g(\mathbf{x}) = \log(1 + \mathbf{x}^\top \mathbf{C}\mathbf{x}) + \mathbf{d}^\top \mathbf{x}, \quad$ où $\mathbf{C}$ symmétrique.
- **Objectif :** Trouver $\nabla^2_{\mathbf{x}} g(\mathbf{x})$
- **Solution :**
- **Justification Détaillée :**

Problème 1 : Question 2 (Solution)

- **Fonction :** $g(\mathbf{x}) = \log(1 + \mathbf{x}^\top \mathbf{C}\mathbf{x}) + \mathbf{d}^\top \mathbf{x}$, où $\mathbf{C}$ symmétrique.
- **Objectif :** Trouver $\nabla^2_{\mathbf{x}} g(\mathbf{x})$
- **Solution :** $\nabla^2_{\mathbf{x}} g(\mathbf{x}) = \frac{2\mathbf{C}}{1+\mathbf{x}^\top \mathbf{C}\mathbf{x}} - \frac{4\mathbf{C}\mathbf{x}\mathbf{x}^\top \mathbf{C}}{(1+\mathbf{x}^\top \mathbf{C}\mathbf{x})^2}$
- **Justification Détaillée :**
  1. **Rappel du Gradient :** Le gradient calculé précédemment est $\frac{2\mathbf{C}\mathbf{x}}{1+\mathbf{x}^\top \mathbf{C}\mathbf{x}}$.
  2. **Application de la Règle du Quotient :**
     - Pour la Hessienne, nous appliquons la règle du quotient à chaque terme du gradient. Le dénominateur commun est $(1 + \mathbf{x}^\top \mathbf{C}\mathbf{x})^2$.
     - Le numérateur implique la différentiation du produit $2\mathbf{C}\mathbf{x}$ où $\mathbf{C}$ est constant.
  3. **Différentiation du Produit :**
     - La dérivée de $2\mathbf{C}\mathbf{x}$ par rapport à $\mathbf{x}$ donne $2\mathbf{C}$, qui est constante par rapport à $\mathbf{x}$.
     - La dérivée du produit $\mathbf{x}^\top \mathbf{C}\mathbf{x}$ donne $2\mathbf{C}\mathbf{x}$ en utilisant la symétrie de $\mathbf{C}$.
     - Le produit des dérivées implique $\mathbf{C}\mathbf{x}(\mathbf{C}\mathbf{x})^\top$ soit $\mathbf{C}\mathbf{x}\mathbf{x}^\top \mathbf{C}$.
  4. **Combinaison des Résultats :**
     - Combiner tous les termes en respectant les signes et les puissances du dénominateur pour former la Hessienne finale.

Problème 1 : Question 2 (Solution)

- **Fonction :** $g(\mathbf{x}) = \log(1 + \mathbf{x}^\top \mathbf{C}\mathbf{x}) + \mathbf{d}^\top \mathbf{x}$, où $\mathbf{C}$ symmétrique.
- **Objectif :** Trouver $\nabla^2_{\mathbf{x}} g(\mathbf{x})$
- **Solution :** $\nabla^2_{\mathbf{x}} g(\mathbf{x}) = \frac{2\mathbf{C}}{1+\mathbf{x}^\top \mathbf{C}\mathbf{x}} - \frac{4(\mathbf{x}^\top \mathbf{C}\mathbf{x})\mathbf{C}\mathbf{x}}{(1+\mathbf{x}^\top \mathbf{C}\mathbf{x})^2}$
- **Justification Détaillée :**
  1. **Rappel du Gradient :** Le gradient calculé précédemment est $\frac{2\mathbf{C}\mathbf{x}}{1+\mathbf{x}^\top \mathbf{C}\mathbf{x}}$.
  2. **Application de la Règle du Quotient :**
     - Pour la Hessienne, nous appliquons la règle du quotient à chaque terme du gradient.
     - La dérivée partielle de $\mathbf{x}^\top \mathbf{C}\mathbf{x}$ est $2\mathbf{C}\mathbf{x}$, résultant de la symétrie de $\mathbf{C}$.
  3. **Différentiation du Produit :**
     - La dérivée seconde de $2\mathbf{C}\mathbf{x}$ par rapport à $\mathbf{x}$ est $2\mathbf{C}$.
     - Pour $\mathbf{x}^\top \mathbf{C}\mathbf{x}$ produit $2\mathbf{C}\mathbf{x}$, on obtient en différentiant encore $2\mathbf{C}$.
     - Finalement, la différentiation de $\mathbf{x}^\top \mathbf{C}\mathbf{x} \cdot \mathbf{x}^\top \mathbf{C}\mathbf{x}$ contribue avec $4(\mathbf{x}^\top \mathbf{C}\mathbf{x})\mathbf{C}\mathbf{x}$.
  4. **Expression Finale de la Hessienne :**

$$\nabla^2_{\mathbf{x}} g(\mathbf{x}) = \frac{2\mathbf{C}}{1 + \mathbf{x}^\top \mathbf{C}\mathbf{x}} - \frac{4(\mathbf{x}^\top \mathbf{C}\mathbf{x})\mathbf{C}\mathbf{x}}{(1 + \mathbf{x}^\top \mathbf{C}\mathbf{x})^2}$$

Problème 1 : Question 3

- **Fonction :** $h(\mathbf{x}) = e^{-\mathbf{d}^\top \mathbf{x}} + \mathbf{x}^\top \mathbf{B}\mathbf{x}$
- **Objectif :** Trouver $\nabla_{\mathbf{x}} h(\mathbf{x})$
- **Solution :**
- **Justification :**
    - Application de la règle de la chaîne pour le terme exponentiel et la dérivation directe pour le terme quadratique.

Problème 1 : Question 3 (Solution)

- **Fonction :** $h(\mathbf{x}) = e^{-d\mathbf{x}^\top} + \mathbf{x}^\top \mathbf{B}\mathbf{x}$, où $\mathbf{B}$ symétrique.
- **Objectif :** Trouver $\nabla_{\mathbf{x}} h(\mathbf{x})$
- **Solution :** $\nabla_{\mathbf{x}} h(\mathbf{x}) = -de^{-d\mathbf{x}^\top} + 2\mathbf{B}\mathbf{x}$
- **Justification :**
  - **Décomposition de la Fonction :**
    - La fonction h est composée de deux termes: un terme exponentiel $e^{-d\mathbf{x}^\top}$ et un terme quadratique $\mathbf{x}^\top \mathbf{B}\mathbf{x}$.
  - **Dérivation du Terme Exponentiel :**
    - Utilisons la règle de la chaîne pour la dérivation du terme $e^{-d\mathbf{x}^\top}$. Soit $z = -d\mathbf{x}^\top$, alors la dérivée par rapport à $\mathbf{x}$ est $-d$ multiplié par la dérivée de e^z qui est e^z, résultant en $-de^{-d\mathbf{x}^\top}$.
  - **Dérivation du Terme Quadratique :**
    - La dérivée du terme quadratique $\mathbf{x}^\top \mathbf{B}\mathbf{x}$ est directe, étant donné que $\mathbf{B}$ est symétrique. Elle est donnée par $2\mathbf{B}\mathbf{x}$, en suivant la règle standard pour la différentiation des formes quadratiques.
  - **Combinaison des Gradients :**
    - Le gradient total de $h(\mathbf{x})$ est la somme des gradients des deux termes, donc,

$$\nabla_{\mathbf{x}} h(\mathbf{x}) = -de^{-d\mathbf{x}^\top} + 2\mathbf{B}\mathbf{x}$$

Problème 1 : Question 4

- **Fonction :** Identique à la question 3.
- **Objectif :** Trouver $\nabla^2_{\mathbf{x}} h(\mathbf{x})$
- **Solution :**
- **Justification :**
  - Le calcul des dérivées secondes pour chaque terme utilise la continuité de la dérivation du terme exponentiel et la constance de $\mathbf{B}$.

Problème 1 : Question 4 (Solution)

- **Fonction :** Utilisez la même fonction que la question 3.
- **Objectif :** Trouver $\nabla^2_{\mathbf{x}} h(\mathbf{x})$
- **Solution :** $\nabla^2_{\mathbf{x}} h(\mathbf{x}) = dd^\top e^{-d\mathbf{x}^\top} + 2\mathbf{B}$
- **Justification :**
  - **Décomposition du Gradient :**
    - De la question 3, nous avons le gradient :

$$\nabla_{\mathbf{x}} h(\mathbf{x}) = -de^{-d\mathbf{x}^\top} + 2\mathbf{B}\mathbf{x}$$

  - **Dérivation du Terme Exponentiel :**
    - La dérivée seconde de $-de^{-d\mathbf{x}^\top}$ s'obtient en appliquant la règle du produit :

$$\nabla^2_{\mathbf{x}}(-de^{-d\mathbf{x}^\top}) = dd^\top e^{-d\mathbf{x}^\top}$$

  - **Dérivation du Terme Quadratique :**
    - La dérivée seconde de $2\mathbf{B}\mathbf{x}$ est simplement $2\mathbf{B}$, car $\mathbf{B}$ est une matrice constante.
  - **Combinaison des Hessiennes :**
    - En combinant les deux termes, nous obtenons la Hessienne :

$$\nabla^2_{\mathbf{x}} h(\mathbf{x}) = dd^\top e^{-d\mathbf{x}^\top} + 2\mathbf{B}$$

Problème 1 : Question 5

- **Fonction :** $j(\mathbf{x}) = \|\mathbf{C}\mathbf{x} + \mathbf{d}\|_1$
- **Objectif :** Trouver $\nabla_{\mathbf{x}} j(\mathbf{x})$
- **Solution :**
- **Justification :**

Problème 1 : Question 5 (Solution)

- **Fonction :** $j(\mathbf{x}) = \|\mathbf{C}\mathbf{x} + \mathbf{d}\|_1$
- **Objectif :** Trouver $\nabla_{\mathbf{x}} j(\mathbf{x})$
- **Solution :** $\nabla_{\mathbf{x}} j(\mathbf{x}) = \mathbf{C}^\top \operatorname{sgn}(\mathbf{C}\mathbf{x} + \mathbf{d})$
- **Justification - Définition de la Norme ℓ_1 :**
  - La norme ℓ_1 est définie par :

$$\|\mathbf{y}\|_1 = \sum_i |y_i|, \quad \text{où} \quad \mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{d}$$

- **Représentation Composante par Composante :**
  - En écrivant la norme comme une somme : $j(\mathbf{x}) = \sum_i |(\mathbf{C}\mathbf{x} + \mathbf{d})_i|$
- **Calcul du Gradient :**
  - La dérivée de $|y_i|$ est donnée par la fonction signe sgn :

$$\frac{\partial}{\partial x}|y_i| = \operatorname{sgn}(y_i) \cdot \frac{\partial y_i}{\partial x}$$

  - Or, $y_i = \mathbf{C}_i^\top \mathbf{x} + d_i$, donc sa dérivée est $\mathbf{C}_i^\top$.
  - En sommant sur toutes les composantes : $\nabla_{\mathbf{x}} j(\mathbf{x}) = \sum_i \operatorname{sgn}(y_i)\mathbf{C}_i^\top$

Problème 1 : Question 5 (Solution)

- **Expression Matricielle :** En notation matricielle compacte, cela donne : $\nabla_{\mathbf{x}} j(\mathbf{x}) = \mathbf{C}^\top \operatorname{sgn}(\mathbf{C}\mathbf{x} + \mathbf{d})$

Problème 1 : Question 6

- **Fonction :** $f(\mathbf{x}) = \|\mathbf{C}\mathbf{x} + \mathbf{d}\|_2^2$
- **Objectif :** Trouver $\nabla^2_{\mathbf{x}} f(\mathbf{x})$
- **Solution :**
- **Justification :**

Problème 1 : Question 6 (Solution)

- **Fonction :** $f(\mathbf{x}) = \|\mathbf{C}\mathbf{x} + \mathbf{d}\|_2^2$
- **Objectif :** Trouver $\nabla_{\mathbf{x}}^2 f(\mathbf{x})$
- **Solution :** $\nabla_{\mathbf{x}}^2 f(\mathbf{x}) = 2\mathbf{C}^\top \mathbf{C}$
- **Justification - Définition de la Norme ℓ_2 :**
  - La norme au carré est définie par :

$$\|\mathbf{y}\|_2^2 = \mathbf{y}^\top \mathbf{y}, \quad \text{où} \quad \mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{d}$$

- **Expansion de l'Expression :**
  - Exprimons la fonction sous une forme quadratique :

$$f(\mathbf{x}) = (\mathbf{C}\mathbf{x} + \mathbf{d})^\top (\mathbf{C}\mathbf{x} + \mathbf{d})$$

  - Développons le produit :

$$f(\mathbf{x}) = \mathbf{x}^\top \mathbf{C}^\top \mathbf{C}\mathbf{x} + 2\mathbf{d}^\top \mathbf{C}\mathbf{x} + \mathbf{d}^\top \mathbf{d}$$

Problème 1 : Question 6 (Solution)

- **Calcul du Gradient :**
  - Le gradient du terme quadratique $\mathbf{x}^\top \mathbf{A}\mathbf{x}$ (où $\mathbf{A}$ est symétrique) est $2\mathbf{A}\mathbf{x}$. Ici, $\mathbf{A} = \mathbf{C}^\top \mathbf{C}$.
  - La dérivée du terme linéaire $2\mathbf{d}^\top \mathbf{C}\mathbf{x}$ est $2\mathbf{C}^\top \mathbf{d}$.
  - En combinant ces résultats, nous obtenons :

$$\nabla_{\mathbf{x}} f(\mathbf{x}) = 2\mathbf{C}^\top \mathbf{C}\mathbf{x} + 2\mathbf{C}^\top \mathbf{d}$$

- **Calcul de la Hessienne :**
  - Le gradient étant $\nabla_{\mathbf{x}} k(\mathbf{x}) = 2\mathbf{C}^\top \mathbf{C}\mathbf{x} + 2\mathbf{C}^\top \mathbf{d}$, la dérivée seconde est obtenue en différenciant à nouveau :

$$\nabla^2_{\mathbf{x}} k(\mathbf{x}) = \nabla_{\mathbf{x}}(2\mathbf{C}^\top \mathbf{C}\mathbf{x} + 2\mathbf{C}^\top \mathbf{d})$$

  - Comme $\mathbf{C}^\top \mathbf{C}$ est une matrice constante, la dérivée donne :

$$\nabla^2_{\mathbf{x}} k(\mathbf{x}) = 2\mathbf{C}^\top \mathbf{C}$$

- **Note:** La Hessienne est une matrice constante, indiquant que la fonction $f(\mathbf{x})$ est une fonction quadratique à courbure uniforme.
- Cette Hessienne est utilisée en optimisation, notamment dans les problèmes de régression quadratique et d'apprentissage automatique.

Table des Matières

Estimation par la Méthode du Maximum de Vraisemblance et Propriétés des Estimateurs (Loi Exponentielle)

Problème 2 : Estimation par MEV pour la Loi Exp

- Soit $X_1, X_2, \ldots, X_n$ un échantillon aléatoire issu d'une **loi exponentielle** de paramètre θ, c'est-à-dire :

$$f(x;\theta) = \frac{1}{\theta} e^{-x/\theta}, \quad x > 0.$$

  - Déterminer l'estimateur du maximum de vraisemblance (EMV) du paramètre θ.
  - Vérifier si cet estimateur est biaisé.
  - Vérifier si cet estimateur est consistant (ou convergent) en calculant sa variance.

    Rappel : On appelle **erreur quadratique moyenne** la quantité :

$$EQM(\hat{\theta}_n) = \mathbb{E}\left[(\hat{\theta}_n - \theta)^2\right] = \mathrm{Var}(\hat{\theta}_n) + [\mathrm{Biais}(\hat{\theta}_n)]^2.$$

    Un estimateur est dit **consistant** si :

$$\lim_{n\to\infty} EQM(\hat{\theta}_n) = 0.$$

Problème 2 : Estimation par MEV pour la Loi Exp

- On observe les fréquences suivantes pour une variable X suivant une loi exponentielle :

x	$f_X(x)$
$(0, 1]$	10
$(1, 2]$	15
$(2, 3]$	8
$(3, 4]$	5
$(4, 5]$	2

  - Trouver l'estimateur du maximum de vraisemblance (EMV) de θ en utilisant les données observées.
  - Estimer la probabilité $P(X \leq 2)$.

Problème 2 : Estimation par MEV pour la Loi Exp

- Soit $X_1, X_2, \ldots, X_n$ un échantillon aléatoire issu d'une loi exponentielle de paramètre θ, c'est-à-dire :

$$f(x;\theta) = \frac{1}{\theta} e^{-x/\theta}, \quad x > 0.$$

- Nous allons déterminer l'estimateur du maximum de vraisemblance (EMV) du paramètre θ, puis analyser ses propriétés.

Prob 2 : Estimation par MEV pour la Loi Exp (Solution)

- La fonction de vraisemblance est donnée par :

$$L(\theta) = \prod_{i=1}^{n} f(X_i; \theta) = \prod_{i=1}^{n} \frac{1}{\theta} e^{-X_i/\theta}.$$

- En prenant le logarithme :

$$\ell(\theta) = \sum_{i=1}^{n} \left(-\ln\theta - \frac{X_i}{\theta} \right) = -n\ln\theta - \frac{1}{\theta}\sum_{i=1}^{n} X_i.$$

- On dérive par rapport à θ et on résout :

$$\frac{d\ell}{d\theta} = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} X_i = 0.$$

- En résolvant, on obtient :

$$\hat{\theta}_{EMV} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Prob 2 : Estimation par MEV pour la Loi Exp (Solution)

- Calculons l'espérance de $\hat{\theta}_{EMV}$:

$$\mathbb{E}[\hat{\theta}_{EMV}] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i].$$

- Comme $\mathbb{E}[X_i] = \theta$, on obtient :

$$\mathbb{E}[\hat{\theta}_{EMV}] = \theta.$$

- L'estimateur est donc non biaisé.

Prob 2 : Estimation par MEV pour la Loi Exp (Solution)

- La variance de $\hat{\theta}_{EMV}$ est donnée par :

$$\text{Var}(\hat{\theta}_{EMV}) = \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} \text{Var}(X_i).$$

- Comme $\text{Var}(X_i) = \theta^2$, on a :

$$\text{Var}(\hat{\theta}_{EMV}) = \frac{\theta^2}{n}.$$

- L'erreur quadratique moyenne est :

$$EQM(\hat{\theta}_{EMV}) = \text{Var}(\hat{\theta}_{EMV}) + [b(\hat{\theta}_{EMV})]^2 = \frac{\theta^2}{n}.$$

- Comme $\lim_{n\to\infty} EQM(\hat{\theta}_{EMV}) = 0$, $\hat{\theta}_{EMV}$ est consistant.

Prob 2 : Estimation par MEV pour la Loi Exp (Solution)

- On observe les fréquences suivantes :

x	$f_X(x)$
$(0,1]$	10
$(1,2]$	15
$(2,3]$	8
$(3,4]$	5
$(4,5]$	2

- L'EMV de θ est donné par :

$$\hat{\theta} = \frac{\sum x_i f_X(x_i)}{n}.$$

- En utilisant les données, nous obtenons une estimation numérique.

Prob 2 : Estimation par MEV pour la Loi Exp (Solution)

- La fonction de répartition d'une loi exponentielle est :

$$F(x) = 1 - e^{-x/\theta}.$$

- Ainsi,

$$P(X \leq 2) = 1 - e^{-2/\hat{\theta}}.$$

- En utilisant l'estimation de $\hat{\theta}$, nous obtenons une valeur numérique.

Prob 2 : Estimation par MEV pour la Loi Exp (Solution)

- On observe les fréquences suivantes :

x	$f_X(x)$
$(0,1]$	10
$(1,2]$	15
$(2,3]$	8
$(3,4]$	5
$(4,5]$	2

- Nous avons une distribution exponentielle de paramètre θ, et notre objectif est d'estimer ce paramètre à partir des données.
- Rappelons que l'estimateur du maximum de vraisemblance (EMV) pour θ est donné par la moyenne empirique :

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Prob 2 : Estimation par MEV pour la Loi Exp (Solution)

- Comme nous avons des données groupées par intervalles, nous utilisons les points centraux des intervalles pour approximer les valeurs de X_i. Ainsi, nous prenons :

$$X_1 = 0.5, \quad X_2 = 1.5, \quad X_3 = 2.5, \quad X_4 = 3.5, \quad X_5 = 4.5.$$

- En pondérant par les fréquences observées $f_X(x)$, l'EMV s'écrit :

$$\hat{\theta} = \frac{\sum x_i f_X(x_i)}{\sum f_X(x_i)}.$$

- En remplaçant par les valeurs observées :

$$\hat{\theta} = \frac{(0.5 \times 10) + (1.5 \times 15) + (2.5 \times 8) + (3.5 \times 5) + (4.5 \times 2)}{10 + 15 + 8 + 5 + 2}.$$

- En effectuant le calcul, nous obtenons une estimation numérique de $\hat{\theta} = \frac{74}{40} = 1.85$.

Prob 2 : Estimation par MEV pour la Loi Exp (Solution)

- Nous souhaitons estimer $P(X \leq 2)$ en utilisant la fonction de répartition d'une loi exponentielle.
- La fonction de répartition cumulative d'une loi exponentielle de paramètre θ est définie par :

$$F(x) = P(X \leq x) = 1 - e^{-x/\theta}.$$

- Cette formule est obtenue en intégrant la densité de la loi exponentielle :

$$P(X \leq x) = \int_0^x \frac{1}{\theta} e^{-t/\theta} dt.$$

- Une primitive de la fonction $e^{-t/\theta}$ est $-\theta e^{-t/\theta}$, ce qui nous donne :

$$P(X \leq x) = \left[-e^{-t/\theta}\right]_0^x = 1 - e^{-x/\theta}.$$

Prob 2 : Estimation par MEV pour la Loi Exp (Solution)

- En appliquant cette formule pour $x = 2$, nous obtenons :

$$P(X \leq 2) = 1 - e^{-2/\hat{\theta}}.$$

- En remplaçant $\hat{\theta}$ par son estimation obtenue précédemment, nous pouvons calculer une estimation numérique de cette probabilité.
  - $$P(X \leq 2) = 1 - e^{-1.081} \approx 1 - 0.3395 = 0.6605.$$
  - Ainsi, l'estimation de $P(X \leq 2)$ est 0.6605.

Table des Matières

Estimation par la Méthode du Maximum de Vraisemblance et Propriétés des Estimateurs (Loi Binomiale)

Problème 3 : Estimation par MEV pour la Loi Binomiale

- Soit $X_1, X_2, \ldots, X_n$ un échantillon aléatoire issu d'une **loi binomiale** de paramètres m et p, c'est-à-dire :

$$P(X = x) = \binom{m}{k} p^x (1-p)^{m-x}, \quad x = 0, 1, 2, \ldots, m.$$

  où :

  - m est le nombre d'essais indépendants.
  - p est la probabilité de succès pour un essai donné.
  - X_i représente le nombre de succès observés pour un individu dans m essais.

- Nous allons déterminer l'estimateur du maximum de vraisemblance (EMV) de p, puis analyser ses propriétés.

Problème 3 : Estimation par MEV pour la Loi Bin

- On observe les fréquences suivantes pour une variable X suivant une loi Binomiale :

x	$f_X(x)$
0	8
1	18
2	12
3	7
4	5

  - Trouver l'estimateur du maximum de vraisemblance (EMV) de $\hat{p}$ en utilisant les données observées.
  - Estimer la probabilité $P(X \leq 2)$.

Prob 3 : Estimation par MEV pour la Loi Bin (Solution)

- La fonction de vraisemblance est donnée par :

$$L(p) = \prod_{i=1}^{n} P(X_i; p) = \prod_{i=1}^{n} \binom{m}{X_i} p^{X_i} (1-p)^{m-X_i}.$$

- En prenant le logarithme :

$$\ell(p) = \sum_{i=1}^{n} \left[\ln \binom{m}{X_i} + X_i \ln p + (m - X_i) \ln(1-p) \right].$$

- En dérivant par rapport à p : $\frac{d\ell}{dp} = \sum_{i=1}^{n} \left[\frac{X_i}{p} - \frac{m-X_i}{1-p} \right]$.
- En posant cette dérivée égale à zéro : $\sum_{i=1}^{n} \left[\frac{X_i}{p} - \frac{m-X_i}{1-p} \right] = 0$.
- En résolvant pour p, on obtient :

$$\hat{p}_{\text{EMV}} = \frac{1}{nm} \sum_{i=1}^{n} X_i.$$

Prob 3 : Estimation par MEV pour la Loi Bin (Solution)

- Calculons l'espérance de $\hat{p}_{\text{EMV}}$:

$$\mathbb{E}[\hat{p}_{\text{EMV}}] = \mathbb{E}\left[\frac{1}{nm}\sum_{i=1}^{n} X_i\right] = \frac{1}{nm}\sum_{i=1}^{n}\mathbb{E}[X_i].$$

- Comme $X_i \sim \text{Bin}(m, p)$, on a $\mathbb{E}[X_i] = mp$, donc :

$$\mathbb{E}[\hat{p}_{\text{EMV}}] = \frac{1}{nm}\cdot nmp = p.$$

- L'estimateur $\hat{p}$ est donc non biaisé.

Prob 3 : Estimation par MEV pour la Loi Bin (Solution)

- La variance de $\hat{p}_{\text{EMV}}$ est donnée par :

$$\text{Var}(\hat{p}_{\text{EMV}}) = \text{Var}\left(\frac{1}{nm}\sum_{i=1}^{n} X_i\right).$$

- Comme $X_i \sim \text{Bin}(m, p)$, sa variance est :

$$\text{Var}(X_i) = mp(1-p).$$

- Donc :

$$\text{Var}(\hat{p}_{\text{EMV}}) = \frac{1}{n^2m^2}\sum_{i=1}^{n} mp(1-p) = \frac{p(1-p)}{nm}.$$

- L'erreur quadratique moyenne est :

$$EQM(\hat{p}_{\text{EMV}}) = \text{Var}(\hat{p}_{\text{EMV}}) + [b(\hat{p}_{\text{EMV}})]^2 = \frac{p(1-p)}{nm}.$$

- Comme $\lim_{n\to\infty} EQM(\hat{p}_{\text{EMV}}) = 0$, $\hat{p}_{\text{EMV}}$ est consistant.

Prob 3 : Estimation par MEV pour la Loi Bin (Solution)

- On observe les fréquences suivantes :

x	$f_X(x)$
0	8
1	18
2	12
3	7
4	5

- L'EMV de p est donné par :

$$\hat{p} = \frac{\sum x_i f_X(x_i)}{nm}.$$

- En remplaçant par les valeurs observées :

$$\hat{p} = \frac{(0 \times 8) + (1 \times 18) + (2 \times 12) + (3 \times 7) + (4 \times 5)}{(8 + 18 + 12 + 7 + 5) \times 4}.$$

- En effectuant le calcul, nous obtenons une estimation numérique de $\hat{p} = \frac{92}{200} = 0.46$.

Prob 3 : Estimation par MEV pour la Loi Bin (Solution)

- La fonction de répartition d'une loi binomiale est :

$$F(x) = P(X \leq x) = \sum_{k=0}^{x} \binom{m}{k} p^k (1-p)^{m-k}.$$

- Ainsi,

$$P(X \leq 2) = \sum_{k=0}^{2} \binom{m}{k} \hat{p}^k (1-\hat{p})^{m-k}.$$

- En utilisant l'estimation de $\hat{p}$, nous obtenons une valeur numérique.

Prob 3 : Estimation par MEV pour la Loi Bin (Solution)

- Nous souhaitons estimer $P(X \leq 2)$, qui s'écrit :

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2).$$

- La fonction de masse de probabilité de la loi binomiale donne :

$$P(X = x) = \binom{m}{x} \hat{p}^x (1 - \hat{p})^{m-x}.$$

- Calculons chaque terme avec $\hat{p} = 0.46$ et $m = 4$:

$$P(X = 0) = \binom{4}{0} (0.46)^0 (1 - 0.46)^4 = (0.54)^4 = 0.0859.$$

$$P(X = 1) = \binom{4}{1} (0.46)^1 (1 - 0.46)^3 = 4 \times (0.46) \times (0.54)^3 = 0.2919.$$

$$P(X = 2) = \binom{4}{2} (0.46)^2 (1 - 0.46)^2 = 6 \times (0.46)^2 \times (0.54)^2 = 0.3504.$$

- En sommant ces probabilités :

$$P(X \leq 2) = 0.0859 + 0.2919 + 0.3504 = 0.7282.$$

Table des Matières

Test t en Régression Linéaire

Problème 4 : Test t en Régression Linéaire

- Une université souhaite étudier l'impact du nombre d'heures de révision (X) sur les notes obtenues à un examen (Y).
- On modélise cette relation par une régression linéaire simple :

$$Y = \beta_0 + \beta_1 X + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2).$$

- L'objectif est de tester si X a un effect sur Y ou non. On considère alors l'hypothèse suivante :

$$H_0 : \beta_1 = 0 \quad \text{contre} \quad H_1 : \beta_1 \neq 0.$$

- Cela permet de vérifier si le nombre d'heures de révision a un effet significatif sur la note finale.

Problème 4 : Test t en Régression Linéaire (Solution)

- L'estimateur des moindres carrés pour β_1 est donné par :

$$\hat{\beta}_1 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}.$$

- Son écart-type est estimé par :

$$s_{\hat{\beta}_1} = \frac{s}{\sqrt{\sum(X_i - \bar{X})^2}}, \quad \text{où} \quad s^2 = \frac{1}{n-2}\sum(Y_i - \hat{Y}_i)^2.$$

- La statistique de test suit une loi de Student à $n-2$ degrés de liberté :

$$t = \frac{\hat{\beta}_1}{s_{\hat{\beta}_1}}.$$

- Si $|t| > t_{\alpha/2,n-2}$, on rejette H_0.

Problème 4 : Test t en Régression Linéaire (Solution)

- On collecte les données suivantes :

X (Heures de révision)	Y (Note à l'examen)
2	50
3	55
5	65
7	70
8	75
10	85

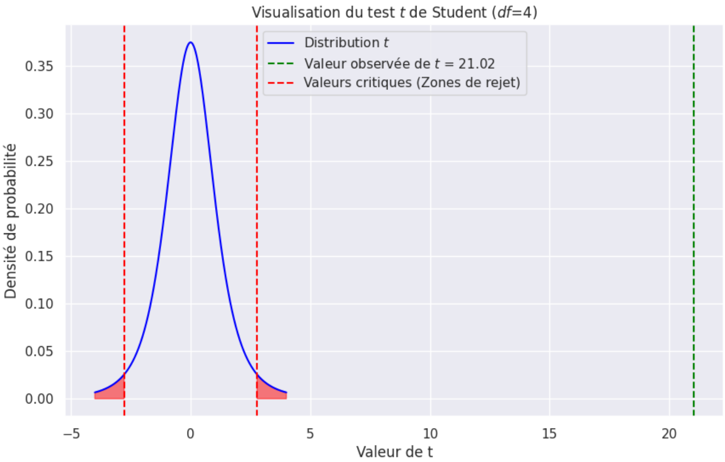
- On utilise Python pour estimer β_1 et effectuer le test t.

Problème 4 : Test t en Régression Linéaire (Solution)

```
import numpy as np
import scipy.stats as stats
import statsmodels.api as sm
X = np.array([2, 3, 5, 7, 8, 10]) # Données
Y = np.array([50, 55, 65, 70, 75, 85]) # Données
X = sm.add_constant(X) # Ajout d'une colonne de 1 pour l'ordonnée à l'origine
model = sm.OLS(Y, X).fit() # Régression linéaire
print(model.summary()) # Affichage des résultats
beta1_hat = model.params[1] # Extraction de beta1_hat
se_beta1 = model.bse[1] # Extraction de se_beta1 pour le test t
t_stat = model.tvalues[1] # Extraction de t_stat pour le test t
p_value = model.pvalues[1] # Extraction de la p_value pour le test t
alpha = 0.05 # Fixer le niveau de signification à 0.05
t_crit = stats.t.ppf(1 - alpha/2, df=len(Y) - 2) # calcul de la statitique critique
print(f"Estimateur de beta1 : {beta1_hat:.4f}")
print(f"Écart-type : {se_beta1:.4f}")
print(f"Statistique t : {t_stat:.4f}")
print(f"p-value : {p_value:.4f}")
print(f"Valeur critique pour alpha={alpha}: {t_crit:.4f}")
if abs(t_stat) > t_crit:
    print("Rejet de H0 : beta1 est significatif")
else:
    print("On ne rejette pas H0 : beta1 n'est pas significatif")
```

```
Estimateur de beta1 : 4.1993; Écart-type : 0.1997
Statistique t : 21.0246; p-value : 0.0000
Valeur critique pour alpha=0.05: 2.7764; Rejet de H0 : beta1 est significatif
```

Problème 4 : Test t en Régression Linéaire (Solution)



Problème 4 : Test t en Régression Linéaire (Solution)

- Si la p-value est inférieure à $\alpha = 0.05$, on rejette H_0, ce qui signifie que X a un effet significatif sur Y.
- Si la p-value est plus grande, alors on ne peut pas conclure que X influence Y.
- La valeur de $\hat{\beta}_1$ nous donne une estimation de l'effet d'une heure de révision supplémentaire sur la note obtenue.

Problème 4 : Test t en Régression Linéaire (Solution)

- Le test t en régression linéaire permet de tester si une variable explicative a un effet significatif sur la variable dépendante.
- Il repose sur l'estimation du coefficient β_1 et son écart-type.
- Si $|t|$ est supérieur à la valeur critique de Student, on rejette H_0 et on conclut que la variable X a un impact significatif.

Table des Matières

Problème 5 : Visualisation des Lois des Grands Nombres et du Théorème Central Limite

Problème 5 - Visualisation des LGN & TCL : Objectifs

- Explorer les trois théorèmes fondamentaux des probabilités :
  1. La **loi faible des grands nombres** : comment la moyenne empirique fluctue autour de l'espérance.
  2. La **loi forte des grands nombres** : convergence des trajectoires individuelles.
  3. Le **théorème central limite** : distribution asymptotique des moyennes.
- Manipuler un jeu de données simulé et ajuster les paramètres pour comprendre leur impact.
- Expliquer ce que vous observez à partir des graphiques générés.

Problème 5 - Visualisation des LGN & TCL

- **Rappel: Différence entre la Loi Faible et la Loi Forte des Grands Nombres**
  - La Loi Faible des Grands Nombres concerne une forte probabilité : $\overline{X}_n$ sera proche de μ pour une grande taille d'échantillon, mais pas nécessairement pour toutes les séquences d'observations.
    - **Formulation mathématique :**

$$P(|\overline{X}_n - \mu| \geq \epsilon) \to 0 \quad \text{quand} \quad n \to \infty, \quad \forall \epsilon > 0.$$

    - **Exemple :** Si nous lançons une pièce et calculons la proportion de 'face', la probabilité que cette proportion s'écarte significativement de 0.5 devient très faible lorsque le nombre de lancers augmente.
  - La Loi Forte des Grands Nombres concerne une convergence presque sûre : elle garantit que $\overline{X}_n$ converge vers μ pour (presque) toutes les séquences possibles des variables aléatoires.
    - **Formulation mathématique :**

$$P\left(\lim_{n\to\infty} \overline{X}_n = \mu\right) = 1.$$

    - **Exemple :** Si nous lançons une pièce et calculons la proportion de 'face', cette proportion atteindra finalement 0.5 et y restera avec probabilité 1.

Problème 5 - Visualisation des LGN & TCL

- **Idée Principale du Théorème Central Limite (TCL) :** Peu importe la distribution initiale des X_i, la moyenne ou la somme des X_i suit une distribution normale lorsque n est suffisamment grand.
- **Formule pour la Moyenne :**

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1) \quad \text{quand} \quad n \to \infty.$$

- **Exemple Simple :** Si nous lançons un dé n fois et calculons la moyenne des résultats obtenus, la distribution de cette moyenne deviendra progressivement normale à mesure que n augmente, même si les résultats individuels du dé suivent une distribution uniforme.

Problème 5 : Visualisation des LGN & TCL

- Nous allons considérer une distribution uniforme $U(0, 1)$.
- Cela permet de voir comment la convergence des moyennes s'applique à un cas plus général.
- Vous devrez exécuter du code et analyser les figures obtenues.

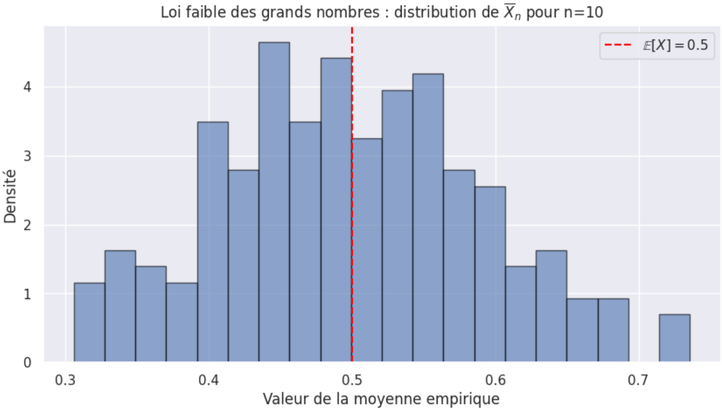
Problème 5 : Loi Faible des Grands Nombres

- On génère n réalisations d'une variable aléatoire uniforme $U(0,1)$.
- L'estimateur qui est considéré est la moyenne empirique $\overline{X}$. Elle est calculée pour chaque taille d'échantillon n.
- **Expérience :** Modifiez n dans le code ci-dessous et observez son effet sur la dispersion de la moyenne.

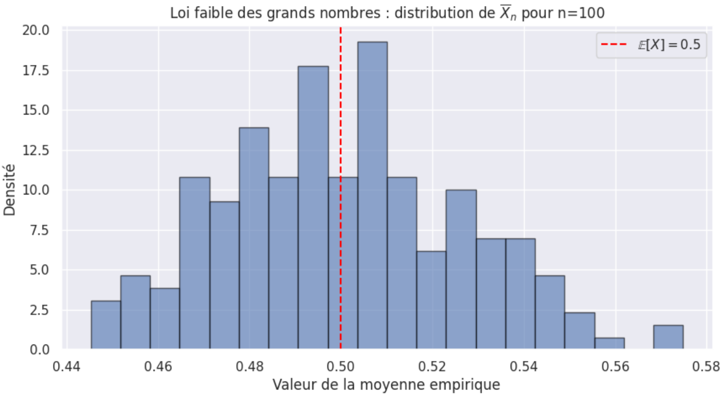
Problème 5 : Loi Faible des Grands Nombres

```python
import numpy as np
import matplotlib.pyplot as plt

np.random.seed(42)

n_values = [10, 100, 1000, 5000]  # Différentes tailles d'échantillon
num_trials = 200  # Nombre de répétitions

for n in n_values:
    X = np.random.uniform(0, 1, (num_trials, n))  # Variables uniformes
    means = np.mean(X, axis=1)  # Moyennes empiriques

    plt.figure(figsize=(10, 5))
    plt.hist(means, bins=20, density=True, alpha=0.6, color='b', edgecolor='black')
    plt.axvline(0.5, color='red', linestyle='dashed', label=r'$\mathbb{E}[X] = 0.5$')
    plt.title(f"Loi faible des grands nombres : distribution de $\overline{{X}}_n$
              f"pour n={n}")
    plt.xlabel("Valeur de la moyenne empirique")
    plt.ylabel("Densité")
    plt.legend()
    plt.show()
```

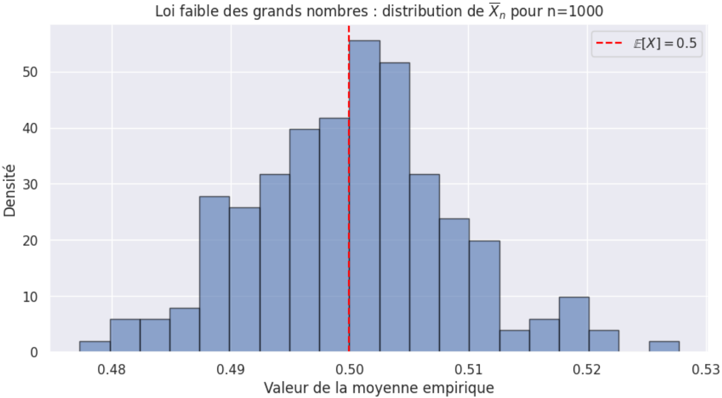
Problème 5 : Lois Faible des Grands nombre (Solution)



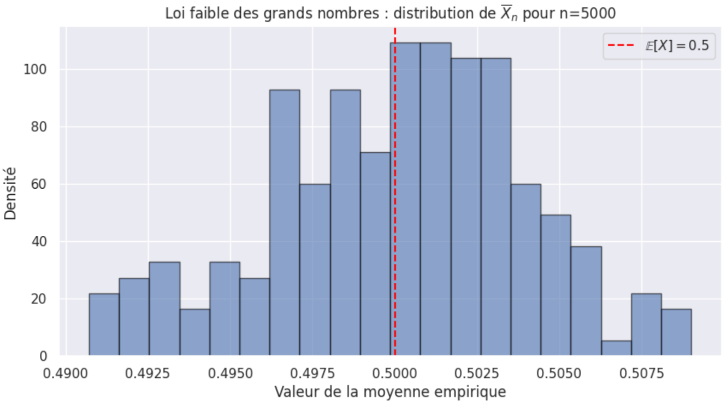
Problème 5 : Lois Faible des Grands nombre (Solution)



Problème 5 : Lois Faible des Grands nombre (Solution)



Problème 5 : Lois Faible des Grands nombre (Solution)



- On observe que la dispersion diminue avec l'augmentation de la taille de l'échantillion où la distribution des moyennes emipirques see concentre de plus en plus autoure de $\mathbb{E}[X]$

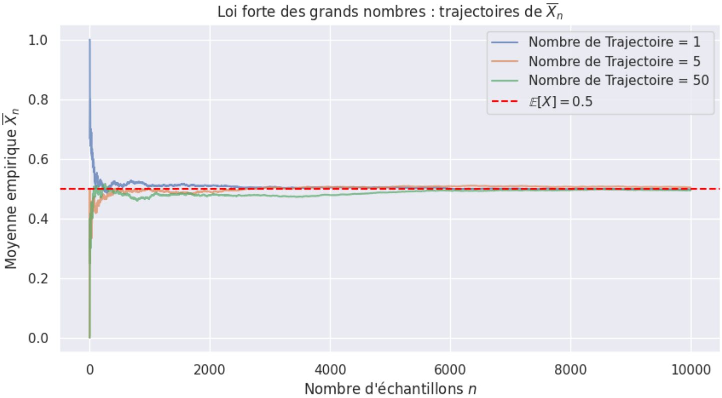
Problème 5 : Lois Forte des Grands nombre Nombres

- Nous allons tracer les trajectoires individuelles de la moyenne empirique.
- **Expérience :** Faites varier le nombre de trajectoires num_{trials} pour voir si toutes convergent.
- Que remarquez-vous ?

Loi Forte des Grands Nombres

```
n = 10000  # Taille de l'échantillon
num_trials = [1, 5, 50]  # Nombre de trajectoires à afficher  # Nombre de trajectoires

X = np.random.uniform(0, 1, (num_trials, n))  # Variables uniformes
means = np.cumsum(X, axis=1) / np.arange(1, n + 1)  # Moyennes cumulatives

plt.figure(figsize=(10, 5))
for i in range(num_trials):
    plt.plot(means[i, :], alpha=0.7, label=f'Trajectoire {i+1}')

plt.axhline(0.5, color='red', linestyle='dashed', label=r'$\mathbb{E}[X] = 0.5$')
plt.title("Loi forte des grands nombres : trajectoires de $\overline{{X}}_n$")
plt.xlabel("Nombre d'échantillons $n$")
plt.ylabel("Moyenne empirique $\overline{X}_n$")
plt.legend()
plt.show()
```

Problème 5 : Lois Forte des Grands nombre (Solution)



- Que ce soit pour une expérience à 1 seul essai, ou pour des expériences à plusieurs essais (5 et 50 essais dans le graphe), $\overline{X}$ converge vers 1 pour une taille d'échantillon n élevée.

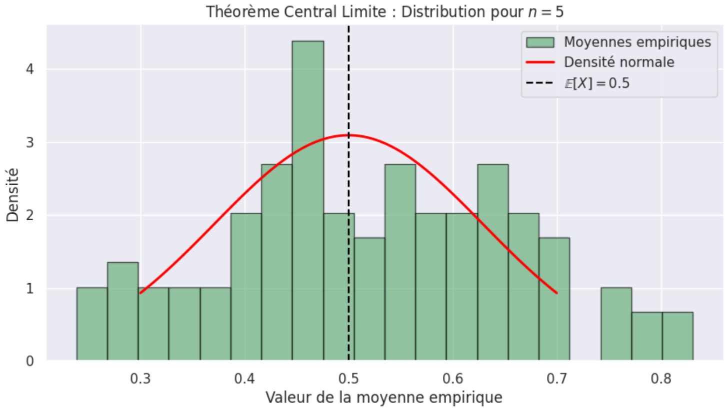
Problème 5 : Théorème Central Limite

- On observe la distribution des moyennes empiriques pour différentes tailles d'échantillon.
- **Expérience :** Changez n $(5, 10, 25, 50, 500)$ et analysez la forme de l'histogramme.
- De quelle distribution $\overline{X}$ s'approche lorsqu'on augmente n?

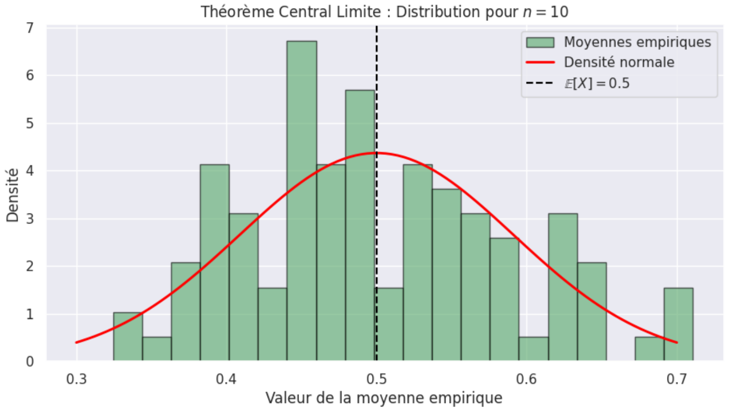
Problème 5 : Théorème Central Limite

```
import scipy.stats as stats

num_trials = 100
n_values = [5, 10, 25, 50, 500]

for n in n_values:
    X_tcl = np.random.uniform(0, 1, (num_trials, n))
    sample_means = np.mean(X_tcl, axis=1)

    plt.figure(figsize=(10, 5))
    plt.hist(sample_means, bins=20, density=True, alpha=0.6, color='g', edgecolor='black',

    # Superposition d'une loi normale
    mu, sigma = 0.5, np.sqrt(1/12 / n)
    x_vals = np.linspace(0.3, 0.7, 100)
    plt.plot(x_vals, stats.norm.pdf(x_vals, mu, sigma), color="red", lw=2, label="Densité r

    plt.axvline(0.5, color='black', linestyle='dashed', label=r'$\mathbb{E}[X] = 0.5$')
    plt.title(f"Théorème Central Limite : Distribution pour $n={n}$")
    plt.xlabel("Valeur de la moyenne empirique")
    plt.ylabel("Densité")
    plt.legend()
    plt.show()
```

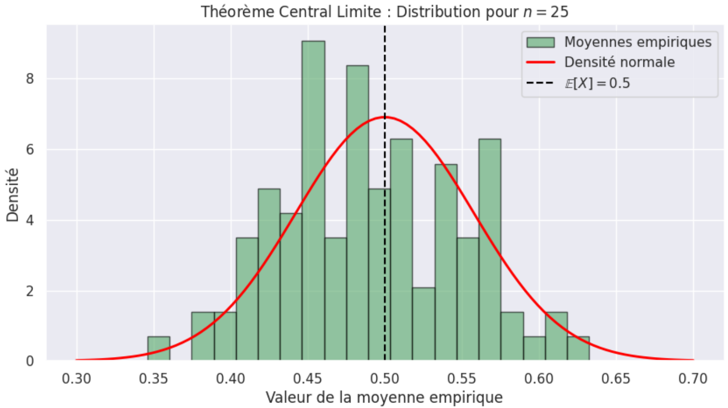
Problème 5 : Théorème Central Limite (Solution)



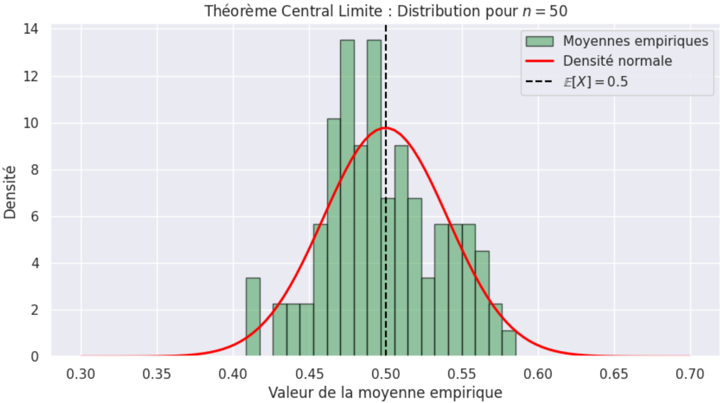
Problème 5 : Théorème Central Limite (Solution)



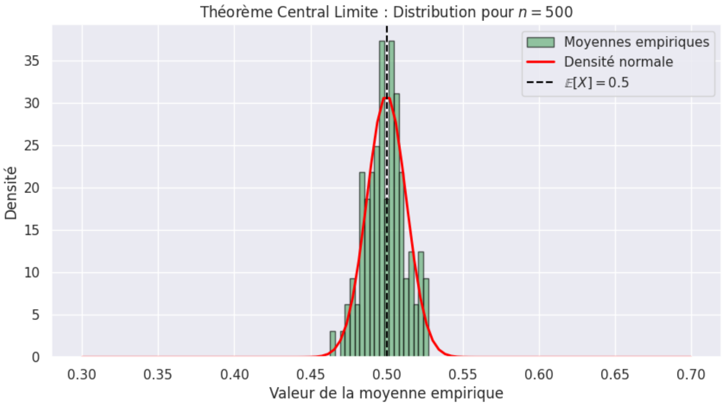
Problème 5 : Théorème Central Limite (Solution)



Problème 5 : Théorème Central Limite (Solution)



Problème 5 : Théorème Central Limite (Solution)



- La distribution de $\overline{X}$ se rapproche de plus en plus d'une distribution normale.

Problème 5 : LGN & TCL

- **Loi faible des grands nombres** : Plus n est grand, plus la moyenne empirique se concentre autour de $\mathbb{E}[X]$.
- **Loi forte des grands nombres** : Chaque trajectoire individuelle converge vers $\mathbb{E}[X]$.
- **Théorème central limite** : Lorsque n augmente, la distribution des moyennes suit une loi normale.