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 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 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 4

- **Fonction :** Identique à la question 3.
- **Objectif :** Trouver $\nabla_{\mathbf{x}}^2 h(\mathbf{x})$
- **Solution :** $\nabla_{\mathbf{x}}^2 h(\mathbf{x}) = \mathbf{d}\mathbf{d}^\top e^{-\mathbf{d}^\top \mathbf{x}} + 2\mathbf{B}$
- **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 5

- **Fonction :** $j(\mathbf{x}) = \|\mathbf{C}\mathbf{x} + \mathbf{d}\|^2$
- **Objectif :** Trouver $\nabla_{\mathbf{x}} j(\mathbf{x})$
- **Solution :** $\nabla_{\mathbf{x}} j(\mathbf{x}) = 2\mathbf{C}^\top(\mathbf{C}\mathbf{x} + \mathbf{d})$
- **Justification :**
  - La dérivée est obtenue en utilisant la règle de la chaîne pour la fonction norme au carré appliquée à l'expression linéaire affine en $\mathbf{x}$.

Problème 1 : Question 6

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

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 $P(X \leq 2)$.

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 :

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

  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 $P(X \leq 2)$.

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.

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

```
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 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 : 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

```python
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()
```