Chapitre 3 - Estimation

Marouane Il Idrissi
il_idrissi.marouane@uqam.ca

STT 1000 - Automne 2025
Département de Mathématiques, Université du Québec à Montréal

UQÀM

Estimation

Dans le **Chapitre 1**, on a défini des **variables aléatoires** pour modéliser des phénomènes

Dans le **Chapitre 2**, on a **analysé des échantillons de données observées**

Dans **ce chapitre**, on va **combiner ce que l'on a vu jusqu'à présent**, en supposant que **les données observées sont des réalisations de variables aléatoires**

Cette supposition va nous permettre d'**estimer** la valeur des **paramètres de la loi de ces v.a.s** à partir **d'un échantillon observé**

Par exemple:
Supposons que **notre échantillon observé** provient de **réalisations d'une v.a.** de loi $\mathcal{N}(\mu, \sigma^2)$ et que je ne connaisse ni la valeur de μ ni celle de σ^2.

☞ **Est-il possible de calculer "une bonne estimation de μ et σ^2"?**
☞ **Comment savoir si cette estimation est bonne?**
☞ **Sur quelle base puis-je comparer deux estimations de μ?** ☞ **Comment construire des estimation?**

Notes

Plan du chapitre

1. Estimation ponctuelle

2. Propriétés des estimateurs

3. Illustrations

4. Méthodes d'estimation

Sommaire

Notes

Estimation - Échantillon i.i.d.

Un **échantillon** est **une collection** $X_1, \ldots, X_n$ de **variables aléatoires** ★
Un **échantillon observé (Chap2)** est **une réalisation de cette collection de variables aléatoire**

☞ **Dans la suite, on notera n la taille de l'échantillon**

Dans tout ce cours, nous allons supposer que:

1. Les variables aléatoires $X_1, \ldots, X_n$ sont **(mutuellement) indépendantes** ;
2. **Toutes les variables qui composent l'échantillon ont la même loi, avec les mêmes paramètres**.

Si ces **deux conditions sont respectées**, on parle d'**échantillon indépendant et identiquement distribué (i.i.d.)** ★

Estimation - Loi et paramètre de la population

☞ **Dans la suite de ce cours, on va toujours supposer que la taille de la population est infinie**

Chaque **v.a.** qui compose **l'échantillon** représente **un individu de la population**

La **loi de ces v.a.s** est appelée la **loi de la population** ★
Ou **loi-mère**, ou encore **loi populationnelle**

Dans ce cours, on ne va s'intéresser **qu'à des lois populationnelles à paramètres**
μ et σ pour une normale, $[a, b]$ pour une uniforme, λ pour une exponentielle

Les **paramètres de la loi de la population** s'appellent les **paramètres de la population** ★

Problématique de l'estimation (paramétrique):
Les **paramètres de la population sont inconnus et on cherche à les estimer**

Notes

Estimation - Estimateur et estimation

Pour **estimer un paramètre**, on va faire appel à un **estimateur de ce paramètre**

L'**estimateur d'un paramètre**, c'est **une fonction de l'échantillon** $X_1, \ldots, X_n$ ★

☞ **Un estimateur d'un paramètre est une variable aléatoire**

Un **échantillon observé** $x_1, \ldots, x_n$ est la **réalisation d'un échantillon** ★

x_1 est une réalisation de X_1, x_2 de X_2...

L'**estimation d'un paramètre**, c'est **une fonction de l'échantillon observé** $x_1, \ldots, x_n$ ★

☞ **Une estimation d'un paramètre est un nombre**

Exemple: **Qui peut être un estimateur, qui peut être une estimation?** ✍

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Estimation - Notations

Pour **formaliser le problème de l'estimation de paramètre**, on va d'abord introduire **quelques notations**:

- **Échantillon**: $\mathbf{X} = (X_1, \ldots, X_n)$ ★
- **Échantillon observé**: $\mathbf{x} = (x_1, \ldots, x_n)$ ★
- **Paramètre (générique)**: θ ★
- **Ensemble de définition du paramètre**: Θ ★
- **Densité de la loi de la population**: $f(s; \theta)$
- **Estimateur (générique) d'un paramètre**: $\widehat{\theta}(\mathbf{X})$ ★
- **Estimation d'un paramètre**: $\widehat{\theta}(\mathbf{x})$ ★

Donnez θ et Θ pour les lois de la population suivantes:

☞ Loi de Bernoulli $\mathcal{B}(p)$ ✍

☞ Loi exponentielle $\mathcal{E}(\lambda)$ ✍

☞ Loi normale $\mathcal{N}(\mu, \sigma^2)$ ✍

Notes

Estimation - Pour fixer les idées

Soit un **échantillon i.i.d** $X_1, \ldots, X_5$ de **loi populationnelle** $\mathcal{B}(p)$
On cherche à **estimer le paramètre de la population** $\theta = p \in \Theta = [0, 1]$

On propose un **estimateur pour** p

$$\widehat{\theta}(\mathbf{X}) = \frac{1}{5}(X_1 + X_2 + X_3 + X_4 + X_5)$$

Si on **observe un échantillon** $\mathbf{x} = (1, 0, 1, 1, 0)$ alors

$$\widehat{\theta}(\mathbf{x}) = \frac{1}{5}(1 + 0 + 1 + 1 + 0) = 3/5$$

qui est **une estimation de** p

☞ **Est-ce que $\widehat{\theta}$ est un bon estimateur de p?**

Est-ce qu'on peut **faire confiance a l'estimation**?

Sommaire

Notes

Propriétés des estimateurs - Biais

Le **biais d'un estimateur** $\widehat{\theta}(\mathbf{X})$ permet de **jauger si cet estimateur** **vise juste en moyenne**

Intuition "avec les mains":

1. Supposons que l'on ait accès a plusieurs échantillons observés d'un même échantillon
2. On calcule **une estimation par échantillon observé**
3. **La moyenne de ces estimations doit être "proche" de la vraie valeur de** θ

Définition *(Biais d'un estimateur)*. ★

Soit $\mathbf{X} = (X_1, \ldots, X_n)$ un échantillon i.i.d. et $\widehat{\theta}(\mathbf{X})$ un estimateur d'un paramètre de population $\theta \in \Theta$. Le **biais de $\widehat{\theta}$ pour θ** est défini par:

$$\mathrm{biais}_\theta\left(\widehat{\theta}(\mathbf{X})\right) = \mathbb{E}\left[\widehat{\theta}(\mathbf{X})\right] - \theta.$$

Si $\mathrm{biais}_\theta\left(\widehat{\theta}(\mathbf{X})\right) = 0$, on dit que $\widehat{\theta}(\mathbf{X})$ est **un estimateur sans biais** pour θ.

Supposons que $\mathbb{E}\left[X_1\right] = \mu$

☞ Quel est le biais de $\widehat{\mu}_1(\mathbf{X}) = \frac{3}{n}\sum_i^n X_i$ pour μ? ✍

☞ Quel est le biais de $\widehat{\mu}_2(\mathbf{X}) = \frac{1}{n}\sum_i^n X_i$ pour μ? ✍

Biais - Estimateur de la variance

Supposons maintenant que $\mathbb{E}\left[X_1\right] = \mu$ et $\mathbb{V}\left(X_1\right) = \sigma^2$

☞ Montrons que, pour $i = 1, \ldots, n$ ✍

$$\mathbb{E}\left[\left(X_i - \frac{1}{n}\sum_{i=1}^n X_i\right)^2\right] = \sigma^2\frac{n-1}{n}$$

1. Calculez le biais de ✍
$$\widehat{\sigma}_1^2(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^n \left(X_i - \overline{X}_n\right)^2$$
2. Calculez le biais de ✍
$$\widehat{\sigma}_2^2(\mathbf{X}) = \frac{1}{n-1}\sum_{i=1}^n \left(X_i - \overline{X}_n\right)^2 = S^2$$

☞ **Maintenant on sait d'où vient le "$n-1$"!**

Notes

Erreur quadratique moyenne - Comparer deux estimateurs

Comment faire pour comparer **deux estimateurs sans biais** pour le **même paramètre de la population**?

☞ Comparer leur **erreur quadratique moyenne!**

Définition *(Erreur quadratique moyenne (EQM) d'un estimateur).* ★

On appelle **erreur quadratique moyenne d'un estimateur** $\widehat{\theta}(\mathbf{X})$ du paramètre θ la quantité

$$\mathrm{EQM}_\theta\left(\widehat{\theta}(\mathbf{X})\right) = \mathbb{E}\left[\left(\widehat{\theta}(\mathbf{X}) - \theta\right)^2\right]$$

On dira que $\widehat{\theta}_1(\mathbf{X})$ est un **meilleur estimateur que** $\widehat{\theta}_2(\mathbf{X})$ **pour** θ **au sens de l'EQM** si ★

$$\mathrm{EQM}_\theta\left(\widehat{\theta}_1(\mathbf{X})\right) \leq \mathrm{EQM}_\theta\left(\widehat{\theta}_2(\mathbf{X})\right)$$

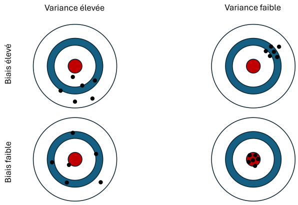
Résultat *(Compromis biais-variance).* ★✍

$$\mathrm{EQM}_\theta\left(\widehat{\theta}(\mathbf{X})\right) = \mathbb{V}\left(\widehat{\theta}(\mathbf{X})\right) + \mathrm{biais}_\theta\left(\widehat{\theta}(\mathbf{X})\right)^2$$

Erreur quadratique moyenne - Compromis biais-variance

Ainsi, ce qui importe pour l'EQM est:

- Le **biais de l'estimateur** (si on vise juste)
- La **variance de l'estimateur** (si on vise dispersé)



Notes

Erreur quadratique moyenne - Efficacité relative

Définition *(Efficacité relative entre deux estimateurs).* ★

Soient $\widehat{\theta}_1(\mathbf{X})$ et $\widehat{\theta}_2(\mathbf{X})$ deux estimateurs de θ. **L'efficacité relative** entre $\widehat{\theta}_1(\mathbf{X})$ et $\widehat{\theta}_1(\mathbf{X})$ est donnée par le ratio

$$\mathrm{e}_\theta\left(\widehat{\theta}_1(\mathbf{X}), \widehat{\theta}_2(\mathbf{X})\right) = \frac{\mathrm{EQM}_\theta\left(\widehat{\theta}_2(\mathbf{X})\right)}{\mathrm{EQM}_\theta\left(\widehat{\theta}_1(\mathbf{X})\right)}$$

- ☞ Si $\mathrm{e}_\theta\left(\widehat{\theta}_1(\mathbf{X}), \widehat{\theta}_2(\mathbf{X})\right) < 1$, $\widehat{\theta}_1(\mathbf{X})$ est **moins efficace pour θ** que $\widehat{\theta}_2(\mathbf{X})$
- ☞ Si $\mathrm{e}_\theta\left(\widehat{\theta}_1(\mathbf{X}), \widehat{\theta}_2(\mathbf{X})\right) = 1$, $\widehat{\theta}_1(\mathbf{X})$ et $\widehat{\theta}_1(\mathbf{X})$ ont **la même efficacité pour θ**
- ☞ Si $\mathrm{e}_\theta\left(\widehat{\theta}_1(\mathbf{X}), \widehat{\theta}_2(\mathbf{X})\right) > 1$, $\widehat{\theta}_1(\mathbf{X})$ est **plus efficace pour θ** que $\widehat{\theta}_1(\mathbf{X})$

Résultat *(Efficacité relative d'estimateurs sans biais).* ★✍

Si $\widehat{\theta}_1(\mathbf{X})$ et $\widehat{\theta}_2(\mathbf{X})$ sont sans biais pour θ, alors

$$\mathrm{e}_\theta\left(\widehat{\theta}_1(\mathbf{X}), \widehat{\theta}_2(\mathbf{X})\right) = \frac{\mathbb{V}\left(\widehat{\theta}_2(\mathbf{X})\right)}{\mathbb{V}\left(\widehat{\theta}_1(\mathbf{X})\right)}$$

Efficacité relative - Exemple

Soit $X_1, \ldots, X_n$ un échantillon i.i.d. de taille n pair, tel que $\mathbb{E}[X_1] = \mu$ et $\mathbb{V}(X_1) = \sigma^2$

On propose les deux estimateurs suivants pour μ

$$\widehat{\mu}_1(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{et} \qquad \widehat{\mu}_2(\mathbf{X}) = \frac{1}{n/2}\sum_{i=1}^{n/2} X_i$$

1. Quel est l'EQM de $\widehat{\mu}_1(\mathbf{X})$? ✍
2. Quel est l'EQM de $\widehat{\mu}_2(\mathbf{X})$? ✍
3. Quelle est l'efficacité relative entre $\widehat{\mu}_1(\mathbf{X})$ et $\widehat{\mu}_2(\mathbf{X})$? ✍
4. Est-ce un résultat intuitif? Pourquoi? ✍

Notes

Sommaire

Échantillon de loi de Bernoulli

Soit un **échantillon i.i.d** $X_1, \ldots, X_5$ de **loi populationnelle** $\mathcal{B}(p)$
On cherche à **estimer le paramètre de la population** $p \in [0, 1]$

On propose **trois estimateurs**:

$$\widehat{p}_1(\mathbf{X}) = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \widehat{p}_2(\mathbf{X}) = 1 - \overline{X}_n = 1 - \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \widehat{p}_3(\mathbf{X}) = \frac{1}{2}(X_1 + X_n)$$

1. Quel est le biais de chaque estimateur? ✍
2. Quelle est la variance de chaque estimateur? ✍
3. Quel est l'EQM de chaque estimateur? ✍
4. Quelle sont les efficacités relatives de chaque estimateur par rapport à $\widehat{p}_1(\mathbf{X})$? ✍
5. Pour quelle valeur du paramètre de la population p est-ce que $\widehat{p}_1$ et $\widehat{p}_2$ ont la même efficacité? ✍
6. Quel est le meilleur estimateur de p? ✍

Notes

Échantillon Gaussien - Estimateur de la moyenne de la population

Soit un **échantillon i.i.d** $X_1, \ldots, X_5$ de **loi de la population** $\mathcal{N}(\mu, \sigma^2)$
Pour commencer, on cherche à **estimer le paramètre de la population** $\mu \in \mathbb{R}$

On propose **deux estimateurs**:

$$\widehat{\mu}_1(\mathbf{X}) = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \widehat{\mu}_2(\mathbf{X}) = \frac{1}{2}(X_1 + X_n)$$

1. Quel est le biais de chaque estimateur? ✍
2. Quelle est la variance de chaque estimateur? ✍
3. Quel est l'EQM de chaque estimateur? ✍
4. Quelle est l'efficacité relative entre $\widehat{\mu}_1(\mathbf{X})$ et $\widehat{\mu}_2(\mathbf{X})$? ✍
5. Quel est le meilleur estimateur de μ? ✍

Deux estimateurs notables

Focus sur **deux estimateurs notables** qu'il est important de connaître
On les a déjà croisés leur version pour **échantillon observé** au chapitre précédent

1. **L'estimateur de la moyenne empirique** ★

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

☞ Sans biais pour estimer **l'espérance de la loi de la population**

2. **L'estimateur de la variance empirique** ★

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n} (X_1 - \overline{X}_n)$$

☞ Sans biais pour estimer **la variance de la loi de la population**

☞ Ces estimateurs sont même **les meilleurs estimateurs sans biais au sens de l'EQM!**
Plus de détails au cours de STT2000!

Notes

Sommaire

Méthodes d’estimation

Jusqu’à présent, **on a étudié les propriétés d’estimateurs qui vous ont été donnés**

Mais **existe-t-il des méthodes systématiques pour construire des estimateurs?**

Il en existe plusieurs, dont **deux principales que nous allons étudier dans ce cours**:

1. **La méthode des moments**
2. **La méthode du maximum de vraisemblance**

☞ **Ces deux méthodes offrent une manière générique de construire des estimateurs quelle que soit la loi de la population**

Mais elles ne produisent pas forcément les mêmes estimateurs

Notes

Méthodes des moments - Notations et définitions

Définition *(Moments d'une variable aléatoire).* ★

Soit X une variable aléatoire à densité f_θ pour $\theta \in \Theta$. On appelle le **moment d'ordre k**, pour $k \in \mathcal{N}$, la quantité $\mathbb{E}\left[X^k\right]$.

Définition *(Moments empiriques (ou échantillonaux)).* ★

Soit $X_1, \ldots, X_n$ un échantillon i.i.d. de loi populationnelle a densité f_θ pour $\theta \in \Theta$. On appelle le **moment empirique de l'échantillon à l'ordre k**, pour $k \in \mathcal{N}$, la quantité:

$$\frac{1}{n}\sum_{i=1}^{n} X_i^k$$

☞ Pour $k = 1$, le moment est **l'espérance**, et le moment empirique est **la moyenne empirique**

Méthodes des moments - Recette

Supposons que l'ont ait accès à un échantillon i.i.d $X_1, \ldots, X_n$
Étapes pour **construire un estimateur par la méthode des moments**: ★

1. **Compter combien de paramètres on souhaite estimer** (la dimension de Θ)
   Supposons que l'on ait m paramètres à estimer
2. Calculer (ou reconnaître) tous les moments d'ordre 1 à m

   Poser les m équations à m inconnues

$$\begin{cases} \mathbb{E}[X] = \frac{1}{n}\sum_{i=1}^{n} X_i \\ \vdots \\ \mathbb{E}[X^m] = \frac{1}{n}\sum_{i=1}^{n} X_i^m \end{cases}$$

   On va "matcher" les m premiers moments aux moments empiriques
3. Trouver le θ qui résout ces équations
   ☞ **C'est l'estimateur par la méthode des moments!**

Généralement, on note l'estimateur obtenu par la méthode des moments $\widehat{\theta}_{MM}(\mathbf{X})$

Notes

Méthodes des moments - Échantillon de Bernoulli

Soit un **échantillon i.i.d** $X_1, \ldots, X_5$ de **loi populationnelle** $\mathcal{B}(p)$
On cherche à **estimer par la méthode des moments le paramètre de la population** $p \in [0, 1]$

1. Combien de paramètres a-t-on a estimer? ✍
2. Quelles sont les valeurs des moments nécessaires? ✍
3. Quel est le système d'équation à résoudre? ✍
4. En déduire un estimateur par la méthode des moments? ✍
5. Est-il biaisé? Quelle est sa variance? Quel est son EQM? ✍

Méthodes des moments - Échantillon Gaussien

Soit un **échantillon i.i.d** $X_1, \ldots, X_5$ de **loi populationnelle** $\mathcal{N}(\mu, \sigma^2)$
On cherche à **estimer par la méthode des moments les paramètres de la population** $(\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)$

1. Combien de paramètres a-t-on a estimer? ✍
2. Quelles sont les valeurs des moments nécessaires? ✍
3. Quel est le système d'équation à résoudre? ✍
4. En déduire un estimateur par la méthode des moments? ✍
5. Est-il biaisé? Quelle est sa variance? Quel est son EQM? ✍

Notes

Méthodes du maximum de vraisemblance - Principes et notations

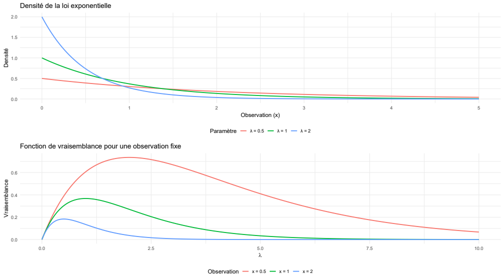
La **vraisemblance**, c'est ce que mesure la **densité de probabilité**
Rappel: La densité de la population s'écrit $f(s;\theta)$

Si on fixe un θ, et qu'on regarde $f(s;\theta)$ comme fonction de s:
☞ Indique la vraisemblance d'observer la valeur s

Si on fixe une observation x_1, et qu'on regarde $f(x_1;\theta)$ comme une valeur de θ:
☞ Indique la vraisemblance que le choix θ ait pu produit x_1 comme observation

C'est cette deuxième interprétation qui va nous intéresser!

Méthodes du maximum de vraisemblance - Loi exponentielle



Notes

Méthodes du maximum de vraisemblance - Vraisemblance de l'échantillon

Définition *(Vraisemblance d'un échantillon).* ★

Soit $\mathbf{X} = (X_1, \ldots, X_n)$ un échantillon i.i.d. de loi populationnelle admettant comme densité $f(s;\theta)$. On appelle **fonction de vraisemblance de l'échantillon** la fonction

$$L(\theta;\mathbf{X}) = \prod_{i=1}^{n} f(x_i;\theta)$$

☞ **C'est la loi joint du vecteur aléatoire X, vu comme une fonction de θ!**

Intuition: Trouver le θ qui aura la plus grande vraisemblance d'avoir produit cet échantillon!

C'est exactement la définition de **l'estimateur du maximum de vraisemblance** (EMV):

$$\widehat{\theta}_{MV}(\mathbf{X}) = \operatorname*{argmax}_{\theta\in\Theta} L(\theta;\mathbf{X})$$

Méthodes du maximum de vraisemblance - En pratique

Maximiser un produit de densités, c'est compliqué...
Maximiser une somme de densités, c'est plus facile!

On préfère travailler avec la **log-vraisemblance de l'échantillon** ★

$$\ell(\theta;\mathbf{X}) = \log\left(L(\theta;\mathbf{X})\right) = \sum_{i=1}^{n} \log\left(f(x_i;\theta)\right)$$

☞ Comme le log est une fonction **croissante**, optimiser L ou ℓ **donne le même résultat!**

Recette: ★

1. Calculer la vraisemblance puis la log-vraisemblance de l'échantillon ;
2. Calculer la dérivée **par rapport à** θ, puis l'annuler ;
3. Vérifier que la **dérivée seconde de la log-vraisemblance est négative** pour s'assurer la solution trouvée est un maximum.

Rappel: Dérivée seconde négative indique que la fonction est concave ✍

Notes

Méthodes du maximum de vraisemblance - Échantillon Bernoulli

Soit un **échantillon i.i.d** $X_1, \dots, X_5$ de **loi populationnelle** $\mathcal{B}(p)$
On cherche à **estimer par la méthode des moments le paramètre de la population** $p \in [0, 1]$

1. Quelle est la fonction de masse de probabilité $f(t; p)$ de la loi populationnelle? ✍
2. Quelle est la vraisemblance de l'échantillon? Sa log-vraisemblance? ✍
3. Quelle est la dérivée de la log-vraisemblance? Annulez-là pour trouver l'estimateur du maximum de vraisemblance. ✍
4. Quelle est dérivée seconde de la log-vraisemblance? ✍
5. L'EMV est-il biaisé? Quelle est sa variance? Quel est son EQM? ✍

Méthodes du maximum de vraisemblance - Échantillon Gaussien

Soit un **échantillon i.i.d** $X_1, \dots, X_5$ de **loi populationnelle** $\mathcal{N}(\mu, \sigma^2)$
On cherche à **estimer par la méthode du maximum de vraisemblance les paramètres de la population** $(\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)$

1. Quelle est la fonction de densité $f(t; \mu, \sigma^2)$ de la loi populationnelle? ✍
2. Quelle est la vraisemblance de l'échantillon? Sa log-vraisemblance? ✍
3. Quel est le gradient de la log-vraisemblance? Annulez-le pour trouver l'estimateur du maximum de vraisemblance. ✍
4. Quelle est la matrice Hessienne de la log-vraisemblance? ✍
5. L'EMV est-il biaisé? Quelle est sa variance? Quel est son EQM? ✍

Notes