Statistique & Apprentissage

Paul-Henry Cournède

paul-henry.cournede@centralesupelec.fr

2018-2019

Organisation du cours

2 Parties

- Partie I : Éléments de Statistique Mathématique
  - 5 Amphis, 5 TDs, dont 2 amphis pour commencer, 2 TDs pour finir.
  - 1 Contrôle sans document le 20/05 : exercices mathématiques
- Partie II : Apprentissage statistique

  (= Statistical Learning ≈ Machine Learning)
  - 4 Amphis, 4 TDs
  - 1 Contrôle sans document le 07/06 : QCM
  - 4 séances de TPs sous Python

Supports

- Notes de cours et Enoncés de TDs
  - disponibles cette semaine à la reprographie pour la partie I
  - disponibles un peu plus tard pour la partie II
- Corrigés des exercices disponibles sur EDUNAO après chaque TD

TDs

- Avancés + Normaux

Mille excuses par avance...

- (Généralement) pas de réponse aux questions par mail.

Définition

“ Statistique : Branche des mathématiques ayant pour objet l'analyse et l'interprétation de données quantifiables.”

Dictionnaire TLF

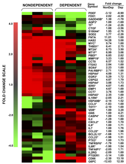
Définition

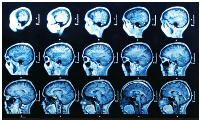
" Statistique : Branche des mathématiques ayant pour objet l'analyse et l'interprétation de données quantifiables."

Dictionnaire TLF

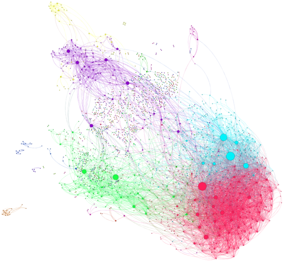
Data everywhere !

Médecine



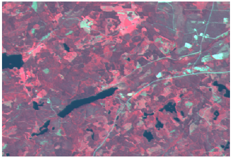


Réseaux sociaux



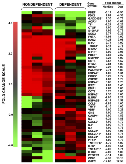
Finance

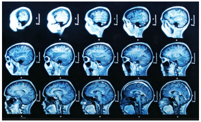
Environnement, Agriculture



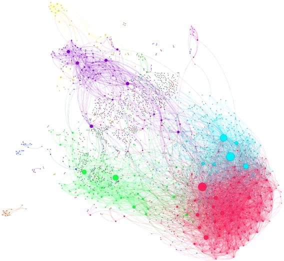
Data everywhere !

Médecine



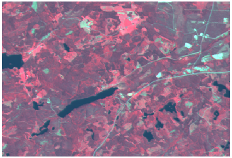


Réseaux sociaux



Finance

Environnement, Agriculture



Disponibilité des données, algorithmes et méthodes statistiques de traitement des données, moyens de calcul $\Longrightarrow$ Révolution de l'Intelligence Artificielle

Objectifs du Cours de Statistique

Définition

Une donnée est la réalisation d'une variable aléatoire.

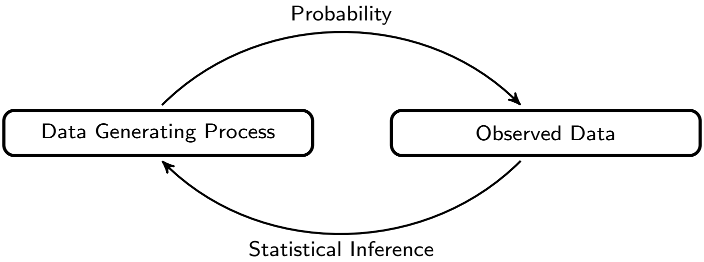
Objectifs du Cours de Statistique

Définition

Une donnée est la réalisation d'une variable aléatoire.

La variable aléatoire représente un phénomène inconnu que l'on souhaite étudier.

Pour cela la Statistique se place **dans l'espace des observations**, et à partir de données, nous réalisons une **inférence sur la loi de probabilité inconnue** de la variable aléatoire sous-jacente.



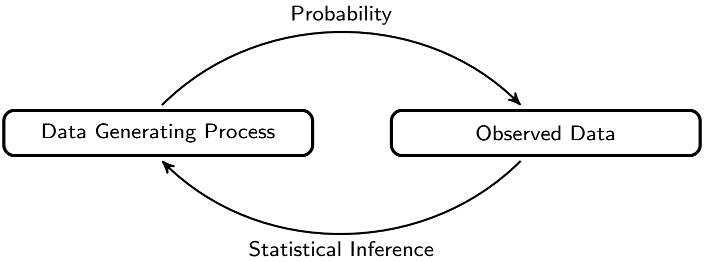
Objectifs du Cours de Statistique

Définition

Une donnée est la réalisation d'une variable aléatoire.

La variable aléatoire représente un phénomène inconnu que l'on souhaite étudier.

Pour cela la Statistique se place **dans l'espace des observations**, et à partir de données, nous réalisons une **inférence sur la loi de probabilité inconnue** de la variable aléatoire sous-jacente.



Avec deux objectifs :

- analyser un phénomène passé (**statistique descriptive**)
- prédire un phénomène futur à partir de l'analyse du passé (**statistique prédictive ou inférentielle**)

Ce que nous allons voir ...

1. Modèles statistiques paramétriques $\quad X \sim P_\theta, \theta \in \Theta$

Ce que nous allons voir ...

1. Modèles statistiques paramétriques $\quad X \sim P_\theta, \theta \in \Theta$
2. Estimation Paramétrique

soient :$x_1, x_2, \ldots, x_N,$ observations de $X,$ on cherche $\hat{\theta}$ qui approche θ

Ce que nous allons voir ...

1. Modèles statistiques paramétriques $X \sim P_\theta, \theta \in \Theta$
2. Estimation Paramétrique

   soient :$x_1, x_2, \ldots, x_N,$ observations de $X,$ on cherche $\hat{\theta}$ qui approche θ
3. Caractérisation de l'incertitude d'estimation par des intervalles de confiance

   On cherche θ_1, θ_2 tels que $\theta \in [\theta_1, \theta_2]$ à 95%

Ce que nous allons voir ...

1. Modèles statistiques paramétriques $X \sim P_\theta, \theta \in \Theta$
2. Estimation Paramétrique

   soient :$x_1, x_2, \ldots, x_N,$ observations de $X,$ on cherche $\hat{\theta}$ qui approche θ
3. Caractérisation de l'incertitude d'estimation par des intervalles de confiance

   On cherche θ_1, θ_2 tels que $\theta \in [\theta_1, \theta_2]$ à 95%
4. Tests d'Hypothèses pour la validation des modèles

   $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ $\Longrightarrow$ Test paramétrique
   ou
   $H_0 : X \sim P_\theta$ versus $H_1 : X \sim Q_\lambda$ $\Longrightarrow$ Test d'ajustement

Ce que nous allons voir ...

1. Modèles statistiques paramétriques $X \sim P_\theta, \theta \in \Theta$
2. Estimation Paramétrique

   soient :$x_1, x_2, \ldots, x_N,$ observations de $X,$ on cherche $\hat{\theta}$ qui approche θ
3. Caractérisation de l'incertitude d'estimation par des intervalles de confiance

   On cherche θ_1, θ_2 tels que $\theta \in [\theta_1, \theta_2]$ à 95%
4. Tests d'Hypothèses pour la validation des modèles

   $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ $\Longrightarrow$ Test paramétrique
   ou
   $H_0 : X \sim P_\theta$ versus $H_1 : X \sim Q_\lambda$ $\Longrightarrow$ Test d'ajustement
5. Cas des modèles conditionnels : l'apprentissage supervisé

   $Y \sim P_\theta(Y|X)$, ou Y est la variable à expliquer, X les variables explicatives.
   $Y \in \mathbb{R}$ $\Longrightarrow$ Régression ; $Y \in \{0; 1; \ldots; K\}$ $\Longrightarrow$ Classification

Ce que nous allons voir ...

1. Modèles statistiques paramétriques $\qquad X \sim P_\theta, \theta \in \Theta$
2. Estimation Paramétrique

   soient :$x_1, x_2, \ldots, x_N,$ observations de $X,$ on cherche $\hat{\theta}$ qui approche θ
3. Caractérisation de l'incertitude d'estimation par des intervalles de confiance

   On cherche θ_1, θ_2 tels que $\theta \in [\theta_1, \theta_2]$ à 95%
4. Tests d'Hypothèses pour la validation des modèles

   $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ $\Longrightarrow$ Test paramétrique
   ou
   $H_0 : X \sim P_\theta$ versus $H_1 : X \sim Q_\lambda$ $\Longrightarrow$ Test d'ajustement
5. Cas des modèles conditionnels : l'apprentissage supervisé

   $Y \sim P_\theta(Y|X)$, ou Y est la variable à expliquer, X les variables explicatives.
   $Y \in \mathbb{R}$ $\Longrightarrow$ Régression ; $Y \in \{0; 1; \ldots; K\}$ $\Longrightarrow$ Classification
6. Réduction de dimension pour les données de grande dimension

   $x \in \mathbb{R}^p$, on cherche $y \in \mathbb{R}^m$, $m \ll p$, et ψ tels que $\|x - \psi(y)\| < \epsilon$

Ce que nous allons voir ...

1. Modèles statistiques paramétriques $X \sim P_\theta, \theta \in \Theta$
2. Estimation Paramétrique

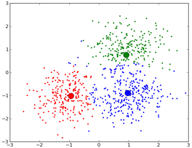
   soient :$x_1, x_2, \ldots, x_N$, observations de X, on cherche $\hat{\theta}$ qui approche θ
3. Caractérisation de l'incertitude d'estimation par des intervalles de confiance

   On cherche θ_1, θ_2 tels que $\theta \in [\theta_1, \theta_2]$ à 95%
4. Tests d'Hypothèses pour la validation des modèles

   $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ $\Longrightarrow$ Test paramétrique
   ou
   $H_0 : X \sim P_\theta$ versus $H_1 : X \sim Q_\lambda$ $\Longrightarrow$ Test d'ajustement
5. Cas des modèles conditionnels : l'apprentissage supervisé

   $Y \sim P_\theta(Y|X)$, ou Y est la variable à expliquer, X les variables explicatives.
   $Y \in \mathbb{R}$ $\Longrightarrow$ Régression ; $Y \in \{0; 1; \ldots; K\}$ $\Longrightarrow$ Classification
6. Réduction de dimension pour les données de grande dimension

   $x \in \mathbb{R}^p$, on cherche $y \in \mathbb{R}^m$, $m \ll p$, et ψ tels que $\|x - \psi(y)\| < \epsilon$
7. Identification de sous-populations de variables aléatoires : Clustering



I - Modélisation Statistique

I.1 - Cadre Probabiliste

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, soit $(\mathcal{X}, \mathcal{A})$ un espace mesurable

Définition

Une variable aléatoire (v.a.) X est une application mesurable de Ω dans $\mathcal{X}$.
Soit $\omega \in \Omega$. La valeur $X(\omega)$ est une réalisation de la v.a. aussi appelée observation ou donnée.

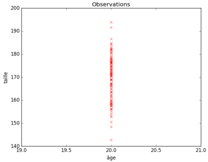
I - Modélisation Statistique

I.1 - Cadre Probabiliste

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, soit $(\mathcal{X}, \mathcal{A})$ un espace mesurable

Définition

Une variable aléatoire (v.a.) X est une application mesurable de Ω dans $\mathcal{X}$.
Soit $\omega \in \Omega$. La valeur $X(\omega)$ est une réalisation de la v.a. aussi appelée observation ou donnée.



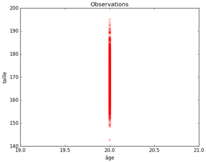
I - Modélisation Statistique

I.1 - Cadre Probabiliste

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, soit $(\mathcal{X}, \mathcal{A})$ un espace mesurable

Définition

Une variable aléatoire (v.a.) X est une application mesurable de Ω dans $\mathcal{X}$.
Soit $\omega \in \Omega$. La valeur $X(\omega)$ est une réalisation de la v.a. aussi appelée observation ou donnée.



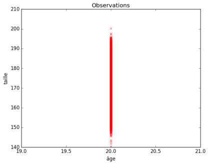
I - Modélisation Statistique

I.1 - Cadre Probabiliste

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, soit $(\mathcal{X}, \mathcal{A})$ un espace mesurable

Définition

Une variable aléatoire (v.a.) X est une application mesurable de Ω dans $\mathcal{X}$.
Soit $\omega \in \Omega$. La valeur $X(\omega)$ est une réalisation de la v.a. aussi appelée observation ou donnée.



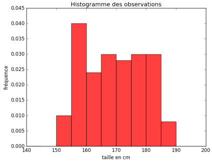
I - Modélisation Statistique

I.1 - Cadre Probabiliste

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, soit $(\mathcal{X}, \mathcal{A})$ un espace mesurable

Définition

Une variable aléatoire (v.a.) X est une application mesurable de Ω dans $\mathcal{X}$.
Soit $\omega \in \Omega$. La valeur $X(\omega)$ est une réalisation de la v.a. aussi appelée observation ou donnée.



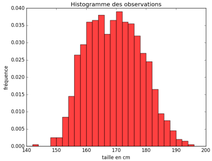
I - Modélisation Statistique

I.1 - Cadre Probabiliste

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, soit $(\mathcal{X}, \mathcal{A})$ un espace mesurable

Définition

Une variable aléatoire (v.a.) X est une application mesurable de Ω dans $\mathcal{X}$.
Soit $\omega \in \Omega$. La valeur $X(\omega)$ est une réalisation de la v.a. aussi appelée observation ou donnée.



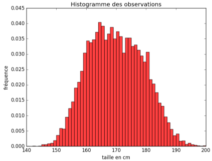
I - Modélisation Statistique

I.1 - Cadre Probabiliste

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, soit $(\mathcal{X}, \mathcal{A})$ un espace mesurable

Définition

Une variable aléatoire (v.a.) X est une application mesurable de Ω dans $\mathcal{X}$.
Soit $\omega \in \Omega$. La valeur $X(\omega)$ est une réalisation de la v.a. aussi appelée observation ou donnée.



I - Modélisation Statistique

I.1 - Cadre Probabiliste

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, soit $(\mathcal{X}, \mathcal{A})$ un espace mesurable

Définition

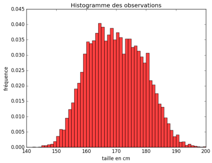
Une variable aléatoire (v.a.) X est une application mesurable de Ω dans $\mathcal{X}$.
Soit $\omega \in \Omega$. La valeur $X(\omega)$ est une réalisation de la v.a. aussi appelée observation ou donnée.

Définition

Soit : X v.a. de $(\Omega, \mathcal{F}, \mathbb{P})$ dans $(\mathcal{X}, \mathcal{A})$. L'application mesurable :

$$P_X : \begin{array}{l} \mathcal{A} \to [0;1] \\ A \mapsto \mathbb{P}\left(X^{-1}(A)\right) \end{array}$$

est une mesure de probabilité sur $\mathcal{A}$, appelée loi de probabilité de la variable aléatoire X.
On note : $X \sim P_X$.



I - Modélisation Statistique

I.1 - Cadre Probabiliste

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, soit $(\mathcal{X}, \mathcal{A})$ un espace mesurable

Définition

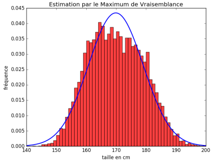
Une variable aléatoire (v.a.) X est une application mesurable de Ω dans $\mathcal{X}$.
Soit $\omega \in \Omega$. La valeur $X(\omega)$ est une réalisation de la v.a. aussi appelée observation ou donnée.

Définition

Soit : X v.a. de $(\Omega, \mathcal{F}, \mathbb{P})$ dans $(\mathcal{X}, \mathcal{A})$. L'application mesurable :

$$P_X : \quad \begin{array}{l} \mathcal{A} \to [0;1] \\ A \mapsto \mathbb{P}\left(X^{-1}(A)\right) \end{array}$$

est une mesure de probabilité sur $\mathcal{A}$, appelée loi de probabilité de la variable aléatoire X.
On note : $X \sim P_X$.



I - Modélisation Statistique

I.1 - Cadre Probabiliste

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, soit $(\mathcal{X}, \mathcal{A})$ un espace mesurable

Définition

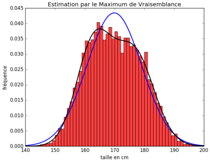
Une variable aléatoire (v.a.) X est une application mesurable de Ω dans $\mathcal{X}$.
Soit $\omega \in \Omega$. La valeur $X(\omega)$ est une réalisation de la v.a. aussi appelée observation ou donnée.

Définition

Soit : X v.a. de $(\Omega, \mathcal{F}, \mathbb{P})$ dans $(\mathcal{X}, \mathcal{A})$. L'application mesurable :

$$P_X : \quad \begin{array}{l} \mathcal{A} \to [0;1] \\ A \mapsto \mathbb{P}\left(X^{-1}(A)\right) \end{array}$$

est une mesure de probabilité sur $\mathcal{A}$, appelée loi de probabilité de la variable aléatoire X.
On note : $X \sim P_X$.



I - Modélisation Statistique

I.1 - Cadre Probabiliste

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, soit $(\mathcal{X}, \mathcal{A})$ un espace mesurable

Définition

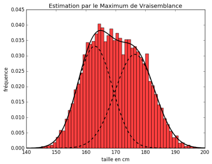
Une variable aléatoire (v.a.) X est une application mesurable de Ω dans $\mathcal{X}$.
Soit $\omega \in \Omega$. La valeur $X(\omega)$ est une réalisation de la v.a. aussi appelée observation ou donnée.

Définition

Soit : X v.a. de $(\Omega, \mathcal{F}, \mathbb{P})$ dans $(\mathcal{X}, \mathcal{A})$. L'application mesurable :

$$P_X : \quad \begin{array}{l} \mathcal{A} \to [0;1] \\ A \mapsto \mathbb{P}\left(X^{-1}(A)\right) \end{array}$$

est une mesure de probabilité sur $\mathcal{A}$, appelée loi de probabilité de la variable aléatoire X.
On note : $X \sim P_X$.



I - Modélisation Statistique

I.1 - Cadre Probabiliste

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, soit $(\mathcal{X}, \mathcal{A})$ un espace mesurable

Définition

Une variable aléatoire (v.a.) X est une application mesurable de Ω dans $\mathcal{X}$.
Soit $\omega \in \Omega$. La valeur $X(\omega)$ est une réalisation de la v.a. aussi appelée observation ou donnée.

Définition

Soit : X v.a. de $(\Omega, \mathcal{F}, \mathbb{P})$ dans $(\mathcal{X}, \mathcal{A})$. L'application mesurable :

$$P_X : \begin{array}{l} \mathcal{A} \to [0;1] \\ A \mapsto \mathbb{P}\left(X^{-1}(A)\right) \end{array}$$

est une mesure de probabilité sur $\mathcal{A}$, appelée loi de probabilité de la variable aléatoire X.
On note : $X \sim P_X$.

Définition

Une mesure σ-finie P est dite dominée par une mesure σ-finie μ si tous les ensembles négligeables pour P le sont aussi pour μ.

I - Modélisation Statistique

I.1 - Cadre Probabiliste

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, soit $(\mathcal{X}, \mathcal{A})$ un espace mesurable

Définition

Une variable aléatoire (v.a.) X est une application mesurable de Ω dans $\mathcal{X}$.
Soit $\omega \in \Omega$. La valeur $X(\omega)$ est une réalisation de la v.a. aussi appelée observation ou donnée.

Définition

Soit : X v.a. de $(\Omega, \mathcal{F}, \mathbb{P})$ dans $(\mathcal{X}, \mathcal{A})$. L'application mesurable :

$$P_X : \quad \begin{array}{l} \mathcal{A} \to [0;1] \\ A \mapsto \mathbb{P}\left(X^{-1}(A)\right) \end{array}$$

est une mesure de probabilité sur $\mathcal{A}$, appelée loi de probabilité de la variable aléatoire X.
On note : $X \sim P_X$.

Définition

Une mesure σ-finie P est dite dominée par une mesure σ-finie μ si tous les ensembles négligeables pour P le sont aussi pour μ.

Théorème

Soit P mesure de probabilité. P admet une densité par rapport à μ ssi P est dominée par μ. La densité est alors la dérivée de Radon-Nikodym $\frac{dP}{d\mu}$.

I - Modélisation Statistique

I.1 - Cadre Probabiliste

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, soit $(\mathcal{X}, \mathcal{A})$ un espace mesurable

Définition

Une variable aléatoire (v.a.) X est une application mesurable de Ω dans $\mathcal{X}$.
Soit $\omega \in \Omega$. La valeur $X(\omega)$ est une réalisation de la v.a. aussi appelée observation ou donnée.

Définition

Soit : X v.a. de $(\Omega, \mathcal{F}, \mathbb{P})$ dans $(\mathcal{X}, \mathcal{A})$. L'application :

$$P_X : \begin{array}{l} \mathcal{A} \to [0;1] \\ A \mapsto \mathbb{P}\left(X^{-1}(A)\right) \end{array}$$

est une mesure de probabilité sur $\mathcal{A}$, appelée loi de probabilité de la variable aléatoire X.
On note : $X \sim P_X$.

Définition

Une mesure σ-finie P est dite dominée par une mesure σ-finie μ si tous les ensembles négligeables pour P le sont aussi pour μ.

Théorème

Soit P mesure de probabilité. P admet une densité par rapport à μ ssi P est dominée par μ. La densité est alors la dérivée de Radon-Nikodym $\frac{dP}{d\mu}$.

I - Modélisation Statistique

I.1 - Cadre Probabiliste

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, soit $(\mathcal{X}, \mathcal{A})$ un espace mesurable

Définition

Une variable aléatoire (v.a.) X est une application mesurable de Ω dans $\mathcal{X}$.
Soit $\omega \in \Omega$. La valeur $X(\omega)$ est une réalisation de la v.a. aussi appelée observation ou donnée.

Définition

Soit : X v.a. de $(\Omega, \mathcal{F}, \mathbb{P})$ dans $(\mathcal{X}, \mathcal{A})$. L'application :

$$P_X : \begin{array}{l} \mathcal{A} \to [0;1] \\ A \mapsto \mathbb{P}\left(X^{-1}(A)\right) \end{array}$$

est une mesure de probabilité sur $\mathcal{A}$, appelée loi de probabilité de la variable aléatoire X.
On note : $X \sim P_X$.

Définition

Une mesure σ-finie P est dite dominée par une mesure σ-finie μ si tous les ensembles négligeables pour P le sont aussi pour μ.

Théorème

Soit P mesure de probabilité. P admet une densité par rapport à μ ssi P est dominée par μ. La densité est alors la dérivée de Radon-Nikodym $\frac{dP}{d\mu}$.

Dans ce cours, $\mathcal{X} \subset \mathbb{R}^d$:
Variables aléatoires continues $\Longrightarrow$ mesure de référence $\mu =$ mesure de Lebesgue.

$$A \in \mathcal{A} \subset \mathcal{B}(\mathbb{R}^d) : P_X(A) = \int_A p(x)\lambda_d(dx)$$

I - Modélisation Statistique

I.1 - Cadre Probabiliste

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, soit $(\mathcal{X}, \mathcal{A})$ un espace mesurable

Définition

Une variable aléatoire (v.a.) X est une application mesurable de Ω dans $\mathcal{X}$.
Soit $\omega \in \Omega$. La valeur $X(\omega)$ est une réalisation de la v.a. aussi appelée observation ou donnée.

Définition

Soit : X v.a. de $(\Omega, \mathcal{F}, \mathbb{P})$ dans $(\mathcal{X}, \mathcal{A})$. L'application :

$$P_X : \begin{array}{l} \mathcal{A} \to [0;1] \\ A \mapsto \mathbb{P}\left(X^{-1}(A)\right) \end{array}$$

est une mesure de probabilité sur $\mathcal{A}$, appelée loi de probabilité de la variable aléatoire X.
On note : $X \sim P_X$.

Définition

Une mesure σ-finie P est dite dominée par une mesure σ-finie μ si tous les ensembles négligeables pour P le sont aussi pour μ.

Théorème

Soit P mesure de probabilité. P admet une densité par rapport à μ ssi P est dominée par μ. La densité est alors la dérivée de Radon-Nikodym $\frac{dP}{d\mu}$.

Dans ce cours, $\mathcal{X} \subset \mathbb{R}^d$:
Variables aléatoires continues $\Longrightarrow$ mesure de référence $\mu =$ mesure de Lebesgue.

$$A \in \mathcal{A} \subset \mathcal{B}(\mathbb{R}^d) : P_X(A) = \int_A p(x)\lambda_d(dx)$$

Variables aléatoires discrètes $\Longrightarrow$ mesure de référence $\mu =$ mesure de comptage.

$$A \in \mathcal{A} = \mathcal{P}(\mathcal{X}) : P_X(A) = \int_A p(x)K(dx) = \sum_{\alpha \in A} \mathbb{P}(X = \alpha) = \sum_{\alpha \in A} P_X(\alpha)$$

I - Modélisation Statistique

I.1 - Cadre Probabiliste

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, soit $(\mathcal{X}, \mathcal{A})$ un espace mesurable

Définition

Une variable aléatoire (v.a.) X est une application mesurable de Ω dans $\mathcal{X}$.
Soit $\omega \in \Omega$. La valeur $X(\omega)$ est une réalisation de la v.a. aussi appelée observation ou donnée.

Définition

Soit : X v.a. de $(\Omega, \mathcal{F}, \mathbb{P})$ dans $(\mathcal{X}, \mathcal{A})$. L'application :

$$P_X : \begin{array}{l} \mathcal{A} \to [0;1] \\ A \mapsto \mathbb{P}\left(X^{-1}(A)\right) \end{array}$$

est une mesure de probabilité sur $\mathcal{A}$, appelée loi de probabilité de la variable aléatoire X.
On note : $X \sim P_X$.

Définition

Une mesure σ-finie P est dite dominée par une mesure σ-finie μ si tous les ensembles négligeables pour P le sont aussi pour μ.

Théorème

Soit P mesure de probabilité. P admet une densité par rapport à μ ssi P est dominée par μ. La densité est alors la dérivée de Radon-Nikodym $\frac{dP}{d\mu}$.

Dans ce cours, $\mathcal{X} \subset \mathbb{R}^d$:
Variables aléatoires continues $\Longrightarrow$ mesure de référence $\mu =$ mesure de Lebesgue.

$$A \in \mathcal{A} \subset \mathcal{B}(\mathbb{R}^d) : P_X(A) = \int_A p(x)\lambda_d(dx)$$

Variables aléatoires discrètes $\Longrightarrow$ mesure de référence $\mu =$ mesure de comptage.

$$A \in \mathcal{A} = \mathcal{P}(\mathcal{X}) : P_X(A) = \int_A p(x)K(dx) = \sum_{\alpha \in A} \mathbb{P}(X = \alpha) = \sum_{\alpha \in A} P_X(\alpha)$$

Toutes les lois des v.a. considérées admettront une densité par rapport à l'une ou l'autre de ces mesures de référence.

I - Modélisation Statistique

I.1 - Cadre Probabiliste

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, soit $(\mathcal{X}, \mathcal{A})$ un espace mesurable

Définition

Une variable aléatoire (v.a.) X est une application mesurable de Ω dans $\mathcal{X}$.
Soit $\omega \in \Omega$. La valeur $X(\omega)$ est une réalisation de la v.a. aussi appelée observation ou donnée.

Définition

Soit : X v.a. de $(\Omega, \mathcal{F}, \mathbb{P})$ dans $(\mathcal{X}, \mathcal{A})$. L'application :

$$P_X : \begin{array}{l} \mathcal{A} \to [0;1] \\ A \mapsto \mathbb{P}\left(X^{-1}(A)\right) \end{array}$$

est une mesure de probabilité sur $\mathcal{A}$, appelée loi de probabilité de la variable aléatoire X.
On note : $X \sim P_X$.

Définition

Une mesure σ-finie P est dite dominée par une mesure σ-finie μ si tous les ensembles négligeables pour P le sont aussi pour μ.

Théorème

Soit P mesure de probabilité. P admet une densité par rapport à μ ssi P est dominée par μ. La densité est alors la dérivée de Radon-Nikodym $\frac{dP}{d\mu}$.

Dans ce cours, $\mathcal{X} \subset \mathbb{R}^d$:
Variables aléatoires continues $\Longrightarrow$ mesure de référence $\mu =$ mesure de Lebesgue.

$$A \in \mathcal{A} \subset \mathcal{B}(\mathbb{R}^d) : P_X(A) = \int_A p(x)\lambda_d(dx)$$

Variables aléatoires discrètes $\Longrightarrow$ mesure de référence $\mu =$ mesure de comptage.

$$A \in \mathcal{A} = \mathcal{P}(\mathcal{X}) : P_X(A) = \int_A p(x)K(dx) = \sum_{\alpha \in A} \mathbb{P}(X = \alpha) = \sum_{\alpha \in A} P_X(\alpha)$$

Toutes les lois des v.a. considérées admettront une densité par rapport à l'une ou l'autre de ces mesures de référence.

Définition : Échantillon aléatoire

Soit : X v.a. de $(\Omega, \mathcal{F}, \mathbb{P})$ dans $(\mathcal{X}, \mathcal{A})$, de loi P_X.
On appelle échantillon aléatoire de taille N la v.a. :

$$E : \begin{array}{l} (\Omega^N, \mathcal{F}^{\otimes N}) \to (\mathcal{X}^N, \mathcal{A}^{\otimes N}) \\ (\omega_1, \ldots, \omega_N) \to (X(w_1), \ldots, X(w_N)) \end{array}$$

Si E est de loi $\mathbb{P}_X^{\otimes N}$, il est alors équivalent de considérer $X_1, \ldots, X_N$, N variables aléatoires indépendantes de même loi P_X.

Le vecteur aléatoire $(X_1, \ldots, X_N)$ est alors dit échantillon aléatoire indépendant identiquement distribué (i.i.d.).

Le vecteur d'observations associées $(x_1, \ldots, x_N)$ est dit échantillon d'observations.

Dans ce cours, $\mathcal{X} \subset \mathbb{R}^d$:
Variables aléatoires continues $\Longrightarrow$ mesure de référence $\mu =$ mesure de Lebesgue.

$$A \in \mathcal{A} \subset \mathcal{B}(\mathbb{R}^d) : P_X(A) = \int_A p(x)\lambda_d(dx)$$

Variables aléatoires discrètes à valeurs dans $\Longrightarrow$ mesure de référence $\mu =$ mesure de comptage.

$$A \in \mathcal{A} = \mathcal{P}(\mathcal{X}) : P_X(A) = \int_A p(x)K(dx) = \sum_{\alpha \in A} \mathbb{P}(X = \alpha) = \sum_{\alpha \in A} P_X(\alpha)$$

Toutes les lois des v.a. considérées admettront une densité par rapport à l'une ou l'autre de ces mesures de référence.

Définition : Échantillon aléatoire

Soit : X v.a. de $(\Omega, \mathcal{F}, \mathbb{P})$ dans $(\mathcal{X}, \mathcal{A})$, de loi P_X.
On appelle échantillon aléatoire de taille N la v.a. :

$$E : \begin{array}{l} (\Omega^N, \mathcal{F}^{\otimes N}) \rightarrow (\mathcal{X}^N, \mathcal{A}^{\otimes N}) \\ (\omega_1, \ldots, \omega_N) \rightarrow (X(w_1), \ldots, X(w_N)) \end{array}$$

Si E est de loi $\mathbb{P}_X^{\otimes N}$, il est alors équivalent de considérer $X_1, \ldots, X_N$, N variables aléatoires indépendantes de même loi P_X.

Le vecteur aléatoire $(X_1, \ldots, X_N)$ est alors dit échantillon aléatoire indépendant identiquement distribué (i.i.d.).

Le vecteur d'observations associées $(x_1, \ldots, x_N)$ est dit échantillon d'observations.

Dans ce cours, $\mathcal{X} \subset \mathbb{R}^d$:
Variables aléatoires continues $\Longrightarrow$ mesure de référence $\mu =$ mesure de Lebesgue.

$$A \in \mathcal{A} \subset \mathcal{B}(\mathbb{R}^d) : P_X(A) = \int_A p(x)\lambda_d(dx)$$

Variables aléatoires discrètes à valeurs dans $\Longrightarrow$ mesure de référence $\mu =$ mesure de comptage.

$$A \in \mathcal{A} = \mathcal{P}(\mathcal{X}) : P_X(A) = \int_A p(x)K(dx) = \sum_{\alpha \in A} \mathbb{P}(X = \alpha) = \sum_{\alpha \in A} P_X(\alpha)$$

Toutes les lois des v.a. considérées admettront une densité par rapport à l'une ou l'autre de ces mesures de référence.

Définition : Échantillon aléatoire

Soit : X v.a. de $(\Omega, \mathcal{F}, \mathbb{P})$ dans $(\mathcal{X}, \mathcal{A})$, de loi P_X.
On appelle échantillon aléatoire de taille N la v.a. :

$$E : \begin{array}{l} (\Omega^N, \mathcal{F}^{\otimes N}) \to (\mathcal{X}^N, \mathcal{A}^{\otimes N}) \\ (\omega_1, \ldots, \omega_N) \to (X(w_1), \ldots, X(w_N)) \end{array}$$

Si E est de loi $\mathbb{P}_X^{\otimes N}$, il est alors équivalent de considérer $X_1, \ldots, X_N$, N variables aléatoires indépendantes de même loi P_X.

Le vecteur aléatoire $(X_1, \ldots, X_N)$ est alors dit échantillon aléatoire indépendant identiquement distribué (i.i.d.).

Le vecteur d'observations associées $(x_1, \ldots, x_N)$ est dit échantillon d'observations.

I.2 - La Méthode Empirique

Définition

Soit $(X_1, \ldots, X_N)$, un échantillon aléatoire dans $(\mathcal{X}^N, \mathcal{A}^{\otimes N})$, soit $(\mathcal{Z}, \mathcal{C})$ un espace mesurable.
Soit T une application mesurable de $(\mathcal{X}^N, \mathcal{A}^{\otimes N})$ dans $(\mathcal{Z}, \mathcal{C})$.
Alors $T_N = T \circ (X_1, \ldots, X_N) = T(X_1, \ldots, X_N)$ est une variable aléatoire de $(\Omega^N, \mathcal{F}^{\otimes N})$ dans $(\mathcal{Z}, \mathcal{C})$ appelée statistique de l'échantillon.

Dans ce cours, $\mathcal{X} \subset \mathbb{R}^d$:
Variables aléatoires continues $\Longrightarrow$ mesure de référence $\mu =$ mesure de Lebesgue.

$$A \in \mathcal{A} \subset \mathcal{B}(\mathbb{R}^d) : P_X(A) = \int_A p(x)\lambda_d(dx)$$

Variables aléatoires discrètes à valeurs dans $\Longrightarrow$ mesure de référence $\mu =$ mesure de comptage.

$$A \in \mathcal{A} = \mathcal{P}(\mathcal{X}) : P_X(A) = \int_A p(x)K(dx) \;= \sum_{\alpha \in A} \mathbb{P}(X = \alpha) = \sum_{\alpha \in A} P_X(\alpha)$$

Toutes les lois des v.a. considérées admettront une densité par rapport à l'une ou l'autre de ces mesures de référence.

Définition : Échantillon aléatoire

Soit : X v.a. de $(\Omega, \mathcal{F}, \mathbb{P})$ dans $(\mathcal{X}, \mathcal{A})$, de loi P_X.
On appelle échantillon aléatoire de taille N la v.a. :

$$E : \begin{array}{l} (\Omega^N, \mathcal{F}^{\otimes N}) \to (\mathcal{X}^N, \mathcal{A}^{\otimes N}) \\ (\omega_1, \ldots, \omega_N) \to (X(w_1), \ldots, X(w_N)) \end{array}$$

Si E est de loi $\mathbb{P}_X^{\otimes N}$, il est alors équivalent de considérer $X_1, \ldots, X_N$, N variables aléatoires indépendantes de même loi P_X.

Le vecteur aléatoire $(X_1, \ldots, X_N)$ est alors dit échantillon aléatoire indépendant identiquement distribué (i.i.d.).

Le vecteur d'observations associées $(x_1, \ldots, x_N)$ est dit échantillon d'observations.

I.2 - La Méthode Empirique

Définition

Soit $(X_1, \ldots, X_N)$, un échantillon aléatoire dans $(\mathcal{X}^N, \mathcal{A}^{\otimes N})$, soit $(\mathcal{Z}, \mathcal{C})$ un espace mesurable.
Soit T une application mesurable de $(\mathcal{X}^N, \mathcal{A}^{\otimes N})$ dans $(\mathcal{Z}, \mathcal{C})$.
Alors $T_N = T \circ (X_1, \ldots, X_N) = T(X_1, \ldots, X_N)$ est une variable aléatoire de $(\Omega^N, \mathcal{F}^{\otimes N})$ dans $(\mathcal{Z}, \mathcal{C})$ appelée statistique de l'échantillon.

Exemples

- La moyenne empirique :

$$T(X_1, \ldots, X_N) = \frac{1}{N} \sum_{i=1}^{N} X_i = \bar{X}$$

Dans ce cours, $\mathcal{X} \subset \mathbb{R}^d$:
Variables aléatoires continues $\Longrightarrow$ mesure de référence $\mu =$ mesure de Lebesgue.

$$A \in \mathcal{A} \subset \mathcal{B}(\mathbb{R}^d) : P_X(A) = \int_A p(x)\lambda_d(dx)$$

Variables aléatoires discrètes à valeurs dans $\Longrightarrow$ mesure de référence $\mu =$ mesure de comptage.

$$A \in \mathcal{A} = \mathcal{P}(\mathcal{X}) : P_X(A) = \int_A p(x)K(dx) = \sum_{\alpha \in A} \mathbb{P}(X = \alpha) = \sum_{\alpha \in A} P_X(\alpha)$$

Toutes les lois des v.a. considérées admettront une densité par rapport à l'une ou l'autre de ces mesures de référence.

Définition : Échantillon aléatoire

Soit : X v.a. de $(\Omega, \mathcal{F}, \mathbb{P})$ dans $(\mathcal{X}, \mathcal{A})$, de loi P_X.
On appelle échantillon aléatoire de taille N la v.a. :

$$E : \begin{array}{l} (\Omega^N, \mathcal{F}^{\otimes N}) \to (\mathcal{X}^N, \mathcal{A}^{\otimes N}) \\ (\omega_1, \dots, \omega_N) \to (X(w_1), \dots, X(w_N)) \end{array}$$

Si E est de loi $\mathbb{P}_X^{\otimes N}$, il est alors équivalent de considérer $X_1, \dots, X_N$, N variables aléatoires indépendantes de même loi P_X.

Le vecteur aléatoire $(X_1, \dots, X_N)$ est alors dit échantillon aléatoire indépendant identiquement distribué (i.i.d.).

Le vecteur d'observations associées $(x_1, \dots, x_N)$ est dit échantillon d'observations.

I.2 - La Méthode Empirique

Définition

Soit $(X_1, \dots, X_N)$, un échantillon aléatoire dans $(\mathcal{X}^N, \mathcal{A}^{\otimes N})$, soit $(\mathcal{Z}, \mathcal{C})$ un espace mesurable.
Soit T une application mesurable de $(\mathcal{X}^N, \mathcal{A}^{\otimes N})$ dans $(\mathcal{Z}, \mathcal{C})$.
Alors $T_N = T \circ (X_1, \dots, X_N) = T(X_1, \dots, X_N)$ est une variable aléatoire de $(\Omega^N, \mathcal{F}^{\otimes N})$ dans $(\mathcal{Z}, \mathcal{C})$ appelée statistique de l'échantillon.

Exemples

- La moyenne empirique :
$$T(X_1, \dots, X_N) = \frac{1}{N}\sum_{i=1}^N X_i = \bar{X}$$
- Le maximum :
$T(X_1, \dots, X_N) = \max(X_1, \dots, X_N)$

Dans ce cours, $\mathcal{X} \subset \mathbb{R}^d$:
Variables aléatoires continues $\Longrightarrow$ mesure de référence $\mu =$ mesure de Lebesgue.

$$A \in \mathcal{A} \subset \mathcal{B}(\mathbb{R}^d) : P_X(A) = \int_A p(x)\lambda_d(dx)$$

Variables aléatoires discrètes à valeurs dans $\Longrightarrow$ mesure de référence $\mu =$ mesure de comptage.

$$A \in \mathcal{A} = \mathcal{P}(\mathcal{X}) : P_X(A) = \int_A p(x)K(dx) = \sum_{\alpha \in A} \mathbb{P}(X = \alpha) = \sum_{\alpha \in A} P_X(\alpha)$$

Toutes les lois des v.a. considérées admettront une densité par rapport à l'une ou l'autre de ces mesures de référence.

Définition : Échantillon aléatoire

Soit : X v.a. de $(\Omega, \mathcal{F}, \mathbb{P})$ dans $(\mathcal{X}, \mathcal{A})$, de loi P_X.
On appelle échantillon aléatoire de taille N la v.a. :

$$E : \begin{array}{l} (\Omega^N, \mathcal{F}^{\otimes N}) \to (\mathcal{X}^N, \mathcal{A}^{\otimes N}) \\ (\omega_1, \ldots, \omega_N) \to (X(w_1), \ldots, X(w_N)) \end{array}$$

Si E est de loi $\mathbb{P}_X^{\otimes N}$, il est alors équivalent de considérer $X_1, \ldots, X_N$, N variables aléatoires indépendantes de même loi P_X.

Le vecteur aléatoire $(X_1, \ldots, X_N)$ est alors dit échantillon aléatoire indépendant identiquement distribué (i.i.d.).

Le vecteur d'observations associées $(x_1, \ldots, x_N)$ est dit échantillon d'observations.

I.2 - La Méthode Empirique

Définition

Soit $(X_1, \ldots, X_N)$, un échantillon aléatoire dans $(\mathcal{X}^N, \mathcal{A}^{\otimes N})$, soit $(\mathcal{Z}, \mathcal{C})$ un espace mesurable.
Soit T une application mesurable de $(\mathcal{X}^N, \mathcal{A}^{\otimes N})$ dans $(\mathcal{Z}, \mathcal{C})$.
Alors $T_N = T \circ (X_1, \ldots, X_N) = T(X_1, \ldots, X_N)$ est une variable aléatoire de $(\Omega^N, \mathcal{F}^{\otimes N})$ dans $(\mathcal{Z}, \mathcal{C})$ appelée statistique de l'échantillon.

Exemples

- La moyenne empirique :
$$T(X_1, \ldots, X_N) = \frac{1}{N}\sum_{i=1}^{N} X_i = \bar{X}$$
- Le maximum :
$T(X_1, \ldots, X_N) = \max(X_1, \ldots, X_N)$

Définition

Soit $(X_1, \ldots, X_N)$, un échantillon aléatoire i.i.d. de loi P_X.
La mesure empirique est une statistique de l'échantillon à valeurs dans l'espaces des mesures de probabilité sur $(\mathcal{X}, \mathcal{A})$ définie par :

$$\hat{P}(X_1, \ldots, X_N) = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_i}$$

où δ_{X_i} est la mesure de Dirac en X_i : $\forall A \in \mathcal{A}$,

$$\delta_{X_i}(A) = \begin{cases} 1 \text{ si } X_i \in A \\ 0 \text{ sinon .} \end{cases}$$

Dans ce cours, $\mathcal{X} \subset \mathbb{R}^d$:
Variables aléatoires continues $\Longrightarrow$ mesure de référence $\mu =$ mesure de Lebesgue.

$$A \in \mathcal{A} \subset \mathcal{B}(\mathbb{R}^d) : P_X(A) = \int_A p(x)\lambda_d(dx)$$

Variables aléatoires discrètes à valeurs dans $\Longrightarrow$ mesure de référence $\mu =$ mesure de comptage.

$$A \in \mathcal{A} = \mathcal{P}(\mathcal{X}) : P_X(A) = \int_A p(x)K(dx) = \sum_{\alpha \in A} \mathbb{P}(X = \alpha) = \sum_{\alpha \in A} P_X(\alpha)$$

Toutes les lois des v.a. considérées admettront une densité par rapport à l'une ou l'autre de ces mesures de référence.

Définition : Échantillon aléatoire

Soit : X v.a. de $(\Omega, \mathcal{F}, \mathbb{P})$ dans $(\mathcal{X}, \mathcal{A})$, de loi P_X.
On appelle échantillon aléatoire de taille N la v.a. :

$$E : \begin{array}{l} (\Omega^N, \mathcal{F}^{\otimes N}) \to (\mathcal{X}^N, \mathcal{A}^{\otimes N}) \\ (\omega_1, \ldots, \omega_N) \to (X(w_1), \ldots, X(w_N)) \end{array}$$

Si E est de loi $\mathbb{P}_X^{\otimes N}$, il est alors équivalent de considérer $X_1, \ldots, X_N$, N variables aléatoires indépendantes de même loi P_X.

Le vecteur aléatoire $(X_1, \ldots, X_N)$ est alors dit échantillon aléatoire indépendant identiquement distribué (i.i.d.).

Le vecteur d'observations associées $(x_1, \ldots, x_N)$ est dit échantillon d'observations.

I.2 - La Méthode Empirique

Définition

Soit $(X_1, \ldots, X_N)$, un échantillon aléatoire dans $(\mathcal{X}^N, \mathcal{A}^{\otimes N})$, soit $(\mathcal{Z}, \mathcal{C})$ un espace mesurable.
Soit T une application mesurable de $(\mathcal{X}^N, \mathcal{A}^{\otimes N})$ dans $(\mathcal{Z}, \mathcal{C})$.
Alors $T_N = T \circ (X_1, \ldots, X_N) = T(X_1, \ldots, X_N)$ est une variable aléatoire de $(\Omega^N, \mathcal{F}^{\otimes N})$ dans $(\mathcal{Z}, \mathcal{C})$ appelée statistique de l'échantillon.

Exemples

- La moyenne empirique :
$$T(X_1, \ldots, X_N) = \frac{1}{N}\sum_{i=1}^N X_i = \bar{X}$$
- Le maximum :
$T(X_1, \ldots, X_N) = \max(X_1, \ldots, X_N)$

Définition

Soit $(X_1, \ldots, X_N)$, un échantillon aléatoire i.i.d. de loi P_X.
La mesure empirique est une statistique de l'échantillon à valeurs dans l'espaces des mesures de probabilité sur $(\mathcal{X}, \mathcal{A})$ définie par :

$$\hat{P}(X_1, \ldots, X_N) = \frac{1}{N}\sum_{i=1}^N \delta_{X_i}$$

où δ_{X_i} est la mesure de Dirac en X_i : $\forall A \in \mathcal{A}$,

$$\delta_{X_i}(A) = \begin{cases} 1 \text{ si } X_i \in A \\ 0 \text{ sinon } . \end{cases}$$

Soit $(x_1, \ldots, x_N)$ un échantillon d'observations, $\hat{P}(x_1, \ldots, x_N) = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}$ sera la mesure empirique associée à l'échantillon d'observations.

I.2 - La Méthode Empirique

Définition

Soit $(X_1, \ldots, X_N)$, un échantillon aléatoire dans $(\mathcal{X}^N, \mathcal{A}^{\otimes N})$, soit $(\mathcal{Z}, \mathcal{C})$ un espace mesurable.
Soit T une application mesurable de $(\mathcal{X}^N, \mathcal{A}^{\otimes N})$ dans $(\mathcal{Z}, \mathcal{C})$.
Alors $T_N = T \circ (X_1, \ldots, X_N) = T(X_1, \ldots, X_N)$ est une variable aléatoire de $(\Omega^N, \mathcal{F}^{\otimes N})$ dans $(\mathcal{Z}, \mathcal{C})$ appelée statistique de l'échantillon.

Exemples

- La moyenne empirique :
$$T(X_1, \ldots, X_N) = \frac{1}{N} \sum_{i=1}^{N} X_i = \bar{X}$$
- Le maximum :
$T(X_1, \ldots, X_N) = \max(X_1, \ldots, X_N)$

Définition

Soit $(X_1, \ldots, X_N)$, un échantillon aléatoire i.i.d. de loi P_X.
La mesure empirique est une statistique de l'échantillon à valeurs dans l'espaces des mesures de probabilité sur $(\mathcal{X}, \mathcal{A})$ définie par :

$$\hat{P}(X_1, \ldots, X_N) = \frac{1}{N} \sum_{i=1}^{N} \delta_{X_i}$$

où δ_{X_i} est la mesure de Dirac en X_i : $\forall A \in \mathcal{A}$,
$$\delta_{X_i}(A) = \begin{cases} 1 \text{ si } X_i \in A \\ 0 \text{ sinon .} \end{cases}$$
Soit $(x_1, \ldots, x_N)$ un échantillon d'observations,
$\hat{P}(x_1, \ldots, x_N) = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i}$ sera la mesure empirique associée à l'échantillon d'observations.

I.2 - La Méthode Empirique

Définition

Soit $(X_1, \ldots, X_N)$, un échantillon aléatoire dans $(\mathcal{X}^N, \mathcal{A}^{\otimes N})$, soit $(\mathcal{Z}, \mathcal{C})$ un espace mesurable.
Soit T une application mesurable de $(\mathcal{X}^N, \mathcal{A}^{\otimes N})$ dans $(\mathcal{Z}, \mathcal{C})$.
Alors $T_N = T \circ (X_1, \ldots, X_N) = T(X_1, \ldots, X_N)$ est une variable aléatoire de $(\Omega^N, \mathcal{F}^{\otimes N})$ dans $(\mathcal{Z}, \mathcal{C})$ appelée statistique de l'échantillon.

Exemples

- La moyenne empirique :

$$T(X_1, \ldots, X_N) = \frac{1}{N} \sum_{i=1}^{N} X_i = \bar{X}$$

- Le maximum :
$T(X_1, \ldots, X_N) = \max(X_1, \ldots, X_N)$

Définition

Soit $(X_1, \ldots, X_N)$, un échantillon aléatoire i.i.d. de loi P_X.
La mesure empirique est une statistique de l'échantillon à valeurs dans l'espaces des mesures de probabilité sur $(\mathcal{X}, \mathcal{A})$ définie par :

$$\hat{P}(X_1, \ldots, X_N) = \frac{1}{N} \sum_{i=1}^{N} \delta_{X_i}$$

où δ_{X_i} est la mesure de Dirac en X_i : $\forall A \in \mathcal{A}$,

$$\delta_{X_i}(A) = \begin{cases} 1 \text{ si } X_i \in A \\ 0 \text{ sinon .} \end{cases}$$

Soit $(x_1, \ldots, x_N)$ un échantillon d'observations,
$\hat{P}(x_1, \ldots, x_N) = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i}$ sera la mesure empirique associée à l'échantillon d'observations.

Méthode empirique

$X \sim P_X$, P_X inconnue.

La méthode empirique vise à approcher les caractéristiques de P_X à partir d'un échantillon d'observations $(x_1, \ldots, x_N)$ pour des v.a. i.i.d. de loi P_X.

Si la caractéristique s'exprime par une fonctionnelle G sur l'ensemble des mesures de probabilité sur $(\mathcal{X}, \mathcal{A})$, alors :
la caractéristique empirique approche la vraie caractéristique

$$G(\hat{P}(x_1, \ldots, x_N)) \approx G(P_X)$$

où la caractéristique empirique est la fonctionnelle G appliquée à la mesure empirique.

I.2 - La Méthode Empirique

Définition

Soit $(X_1, \dots, X_N)$, un échantillon aléatoire dans $(\mathcal{X}^N, \mathcal{A}^{\otimes N})$, soit $(\mathcal{Z}, \mathcal{C})$ un espace mesurable.
Soit T une application mesurable de $(\mathcal{X}^N, \mathcal{A}^{\otimes N})$ dans $(\mathcal{Z}, \mathcal{C})$.
Alors $T_N = T \circ (X_1, \dots, X_N) = T(X_1, \dots, X_N)$ est une variable aléatoire de $(\Omega^N, \mathcal{F}^{\otimes N})$ dans $(\mathcal{Z}, \mathcal{C})$ appelée statistique de l'échantillon.

Exemples

- La moyenne empirique :

$$T(X_1, \dots, X_N) = \frac{1}{N}\sum_{i=1}^{N} X_i = \bar{X}$$

- Le maximum :
  $T(X_1, \dots, X_N) = \max(X_1, \dots, X_N)$

Définition

Soit $(X_1, \dots, X_N)$, un échantillon aléatoire i.i.d. de loi P_X.
La mesure empirique est une statistique de l'échantillon à valeurs dans l'espaces des mesures de probabilité sur $(\mathcal{X}, \mathcal{A})$ définie par :

$$\hat{P}(X_1, \dots, X_N) = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_i}$$

où δ_{X_i} est la mesure de Dirac en X_i : $\forall A \in \mathcal{A}$,

$$\delta_{X_i}(A) = \begin{cases} 1 \text{ si } X_i \in A \\ 0 \text{ sinon .} \end{cases}$$

Soit $(x_1, \dots, x_N)$ un échantillon d'observations, $\hat{P}(x_1, \dots, x_N) = \frac{1}{N}\sum_{i=1}^{N} \delta_{x_i}$ sera la mesure empirique associée à l'échantillon d'observations.

Méthode empirique

$X \sim P_X$, P_X inconnue.

La méthode empirique vise à approcher les caractéristiques de P_X à partir d'un échantillon d'observations $(x_1, \dots, x_N)$ pour des v.a. i.i.d. de loi P_X.

Si la caractéristique s'exprime par une fonctionnelle G sur l'ensemble des mesures de probabilité sur $(\mathcal{X}, \mathcal{A})$, alors :
la caractéristique empirique approche la vraie caractéristique

$$G(\hat{P}(x_1, \dots, x_N)) \approx G(P_X)$$

où la caractéristique empirique est la fonctionnelle G appliquée à la mesure empirique.

Exemples

- La moyenne empirique
  Soit X, v.a. dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, intégrable, de loi P.

$$G : P \mapsto \mathbb{E}(X) = \int_{\mathbb{R}} x\, P(dx)$$
$$= \int_{\mathbb{R}} x\, p(x)\, dx \text{ (si } P \text{ admet une densité } p)$$

En appliquant G à la mesure empirique :

$$\begin{aligned} G(\hat{P}(x_1, \dots, x_N)) &= \int_{\mathbb{R}} x\, \hat{P}(x_1, \dots, x_N)\,(dx) \\ &= \int_{\mathbb{R}} x \left(\frac{1}{N}\sum_{i=1}^{N}\delta_{x_i}\right)(dx) \\ &= \frac{1}{N}\sum_{i=1}^{N}\int_{\mathbb{R}} x\, \delta_{x_i}(dx) \\ &= \frac{1}{N}\sum_{i=1}^{N} x_i \end{aligned}$$

Méthode empirique

$X \sim P_X$, P_X inconnue.

La méthode empirique vise à approcher les caractéristiques de P_X à partir d'un échantillon d'observations $(x_1, \ldots, x_N)$ pour des v.a. i.i.d. de loi P_X.

Si la caractéristique s'exprime par une fonctionnelle G sur l'ensemble des mesures de probabilité sur $(\mathcal{X}, \mathcal{A})$, alors :
la caractéristique empirique approche la vraie caractéristique

$$G(\hat{P}(x_1, \ldots, x_N)) \approx G(P_X)$$

où la caractéristique empirique est la fonctionnelle G appliquée à la mesure empirique.

Exemples

- La moyenne empirique
  Soit X, v.a. dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, intégrable, de loi P.

$$\begin{aligned} G : P \mapsto \mathbb{E}(X) &= \int_{\mathbb{R}} x\, P(dx) \\ &= \int_{\mathbb{R}} x\, p(x)\, dx \text{ (si } P \text{ admet une densité } p) \end{aligned}$$

  En appliquant G à la mesure empirique :

$$\begin{aligned} G(\hat{P}(x_1, \ldots, x_N)) &= \int_{\mathbb{R}} x\, \hat{P}(x_1, \ldots, x_N)\,(dx) \\ &= \int_{\mathbb{R}} x \left(\frac{1}{N} \sum_{i=1}^{N} \delta_{x_i} \right) (dx) \\ &= \frac{1}{N} \sum_{i=1}^{N} \int_{\mathbb{R}} x\, \delta_{x_i}(dx) \\ &= \frac{1}{N} \sum_{i=1}^{N} x_i \end{aligned}$$

Méthode empirique

$X \sim P_X$, P_X inconnue.

La méthode empirique vise à approcher les caractéristiques de P_X à partir d'un échantillon d'observations $(x_1, \ldots, x_N)$ pour des v.a. i.i.d. de loi P_X.

Si la caractéristique s'exprime par une fonctionnelle G sur l'ensemble des mesures de probabilité sur $(\mathcal{X}, \mathcal{A})$, alors :
la caractéristique empirique approche la vraie caractéristique

$$G(\hat{P}(x_1, \ldots, x_N)) \approx G(P_X)$$

où la caractéristique empirique est la fonctionnelle G appliquée à la mesure empirique.

Exemples

- La moyenne empirique
  Soit X, v.a. dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, intégrable, de loi P.
  $$\begin{aligned} G : P \mapsto \mathbb{E}(X) &= \int_{\mathbb{R}} x\, P(dx) \\ &= \int_{\mathbb{R}} x\, p(x)\, dx \text{ (si } P \text{ admet une densité } p) \end{aligned}$$
  En appliquant G à la mesure empirique :
  $$\begin{aligned} G(\hat{P}(x_1, \ldots, x_N)) &= \int_{\mathbb{R}} x\, \hat{P}(x_1, \ldots, x_N)\,(dx) \\ &= \int_{\mathbb{R}} x \left(\frac{1}{N} \sum_{i=1}^{N} \delta_{x_i} \right) (dx) \\ &= \frac{1}{N} \sum_{i=1}^{N} \int_{\mathbb{R}} x\, \delta_{x_i}(dx) \\ &= \frac{1}{N} \sum_{i=1}^{N} x_i \end{aligned}$$

- $\mathbb{E}(h(X))$
  Idem : si $h : (\mathcal{X}, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, P-intégrable.
  $$G : P \mapsto \mathbb{E}(h(X)) = \int_{\mathbb{R}} h(x)\, P(dx)$$
  En appliquant G à la mesure empirique :
  $$\begin{aligned} G(\hat{P}(x_1, \ldots, x_N)) &= \int_{\mathbb{R}} h(x) \left(\frac{1}{N} \sum_{i=1}^{N} \delta_{x_i} \right) (dx) \\ &= \frac{1}{N} \sum_{i=1}^{N} h(x_i) \end{aligned}$$

Méthode empirique

$X \sim P_X$, P_X inconnue.

La méthode empirique vise à approcher les caractéristiques de P_X à partir d'un échantillon d'observations $(x_1, \ldots, x_N)$ pour des v.a. i.i.d. de loi P_X.

Si la caractéristique s'exprime par une fonctionnelle G sur l'ensemble des mesures de probabilité sur $(\mathcal{X}, \mathcal{A})$, alors :
la caractéristique empirique approche la vraie caractéristique

$$G(\hat{P}(x_1, \ldots, x_N)) \approx G(P_X)$$

où la caractéristique empirique est la fonctionnelle G appliquée à la mesure empirique.

Exemples

- La moyenne empirique
  Soit X, v.a. dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, intégrable, de loi P.

$$\begin{aligned} G : P \mapsto \mathbb{E}(X) &= \int_{\mathbb{R}} x\, P(dx) \\ &= \int_{\mathbb{R}} x\, p(x)\, dx \text{ (si } P \text{ admet une densité } p) \end{aligned}$$

  En appliquant G à la mesure empirique :

$$\begin{aligned} G(\hat{P}(x_1, \ldots, x_N)) &= \int_{\mathbb{R}} x\, \hat{P}(x_1, \ldots, x_N)\,(dx) \\ &= \int_{\mathbb{R}} x \left(\frac{1}{N} \sum_{i=1}^{N} \delta_{x_i} \right)(dx) \\ &= \frac{1}{N} \sum_{i=1}^{N} \int_{\mathbb{R}} x\, \delta_{x_i}(dx) \\ &= \frac{1}{N} \sum_{i=1}^{N} x_i \end{aligned}$$

- $\mathbb{E}(h(X))$
  Idem : si $h : (\mathcal{X}, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, P-intégrable.

$$G : P \mapsto \mathbb{E}(h(X)) = \int_{\mathbb{R}} h(x)\, P(dx)$$

  En appliquant G à la mesure empirique :

$$\begin{aligned} G(\hat{P}(x_1, \ldots, x_N)) &= \int_{\mathbb{R}} h(x) \left(\frac{1}{N} \sum_{i=1}^{N} \delta_{x_i} \right)(dx) \\ &= \frac{1}{N} \sum_{i=1}^{N} h(x_i) \end{aligned}$$

- Moments et variance empiriques
  En prenant $h(X) = X^k$, nous obtenons les moments empiriques

$$\hat{m}_k(x_1, \ldots, x_N) = \frac{1}{N} \sum_{i=1}^{N} x_i^k$$

  On en déduit la variance empirique :
  $G(P) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$, et donc

$$G(\hat{P}(x_1, \ldots, x_N)) = \frac{1}{N} \sum_{i=1}^{N} x_i^2 - \left(\frac{1}{N} \sum_{i=1}^{N} x_i \right)^2$$

  On note S^2 la statistique associée.

Méthode empirique

$X \sim P_X$, P_X inconnue.

La méthode empirique vise à approcher les caractéristiques de P_X à partir d'un échantillon d'observations $(x_1, \ldots, x_N)$ pour des v.a. i.i.d. de loi P_X.

Si la caractéristique s'exprime par une fonctionnelle G sur l'ensemble des mesures de probabilité sur $(\mathcal{X}, \mathcal{A})$, alors : la caractéristique empirique approche la vraie caractéristique

$$G(\hat{P}(x_1, \ldots, x_N)) \approx G(P_X)$$

où la caractéristique empirique est la fonctionnelle G appliquée à la mesure empirique.

Exemples

- La moyenne empirique
  Soit X, v.a. dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, intégrable, de loi P.
  $$\begin{aligned} G : P \mapsto \mathbb{E}(X) &= \int_{\mathbb{R}} x\, P(dx) \\ &= \int_{\mathbb{R}} x\, p(x)\, dx \text{ (si } P \text{ admet une densité } p) \end{aligned}$$
  En appliquant G à la mesure empirique :
  $$\begin{aligned} G(\hat{P}(x_1, \ldots, x_N)) &= \int_{\mathbb{R}} x\, \hat{P}(x_1, \ldots, x_N)\, (dx) \\ &= \int_{\mathbb{R}} x \left(\frac{1}{N} \sum_{i=1}^N \delta_{x_i} \right) (dx) \\ &= \frac{1}{N} \sum_{i=1}^N \int_{\mathbb{R}} x\, \delta_{x_i}(dx) \\ &= \frac{1}{N} \sum_{i=1}^N x_i \end{aligned}$$

- $\mathbb{E}(h(X))$
  Idem : si $h : (\mathcal{X}, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, P-intégrable.
  $$G : P \mapsto \mathbb{E}(h(X)) = \int_{\mathbb{R}} h(x)\, P(dx)$$
  En appliquant G à la mesure empirique :
  $$\begin{aligned} G(\hat{P}(x_1, \ldots, x_N)) &= \int_{\mathbb{R}} h(x) \left(\frac{1}{N} \sum_{i=1}^N \delta_{x_i} \right) (dx) \\ &= \frac{1}{N} \sum_{i=1}^N h(x_i) \end{aligned}$$

- Moments et variance empiriques
  En prenant $h(X) = X^k$, nous obtenons les moments empiriques
  $$\hat{m}_k(x_1, \ldots, x_N) = \frac{1}{N} \sum_{i=1}^N x_i^k$$
  On en déduit la variance empirique :
  $G(P) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$, et donc
  $$G(\hat{P}(x_1, \ldots, x_N)) = \frac{1}{N} \sum_{i=1}^N x_i^2 - \left(\frac{1}{N} \sum_{i=1}^N x_i \right)^2$$
  On note S^2 la statistique associée.

- Fonction de répartition empirique
  $[G(P)](x) = \int_{-\infty}^x P(dx)$, soit :
  $$\begin{aligned} G(\hat{P}(x_1, \ldots, x_N))(x) &= \int_{-\infty}^x \left(\frac{1}{N} \sum_{i=1}^N \delta_{x_i} \right) (dx) \\ &= \frac{1}{N} \sum_{i=1}^N \int_{-\infty}^x \delta_{x_i}(dx) \\ &= \frac{1}{N} \sum_{i=1}^N \mathbb{I}_{]-\infty;x]}(x_i) \end{aligned}$$

- $\mathbb{E}(h(X))$

  Idem : si $h : (\mathcal{X}, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, P-intégrable.

  $$G : P \mapsto \mathbb{E}(h(X)) = \int_{\mathbb{R}} h(x)\, P(dx)$$

  En appliquant G à la mesure empirique :

  $$\begin{aligned} G(\hat{P}(x_1, \ldots, x_N)) &= \int_{\mathbb{R}} h(x) \left(\frac{1}{N} \sum_{i=1}^{N} \delta_{x_i} \right) (dx) \\ &= \frac{1}{N} \sum_{i=1}^{N} h(x_i) \end{aligned}$$

- Moments et variance empiriques

  En prenant $h(X) = X^k$, nous obtenons les moments empiriques

  $$\hat{m}_k(x_1, \ldots, x_N) = \frac{1}{N} \sum_{i=1}^{N} x_i^k$$

  On en déduit la variance empirique :

  $$G(P) = \mathbb{E}(X^2) - \mathbb{E}(X)^2, \text{ et donc}$$

  $$G(\hat{P}(x_1, \ldots, x_N)) = \frac{1}{N} \sum_{i=1}^{N} x_i^2 - \left(\frac{1}{N} \sum_{i=1}^{N} x_i \right)^2$$

  On note S^2 la statistique associée.

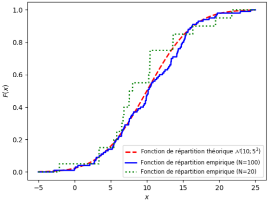
- Fonction de répartition empirique

  $[G(P)](x) = \int_{-\infty}^{x} P(dx)$, soit :

  $$\begin{aligned} G(\hat{P}(x_1, \ldots, x_N))(x) &= \int_{-\infty}^{x} \left(\frac{1}{N} \sum_{i=1}^{N} \delta_{x_i} \right) (dx) \\ &= \frac{1}{N} \sum_{i=1}^{N} \int_{-\infty}^{x} \delta_{x_i}(dx) \\ &= \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}_{]-\infty;x]}(x_i) \end{aligned}$$

- $\mathbb{E}(h(X))$
  Idem : si $h : (\mathcal{X}, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, P-intégrable.
  $$G : P \mapsto \mathbb{E}(h(X)) = \int_{\mathbb{R}} h(x)\, P(dx)$$
  En appliquant G à la mesure empirique :
  $$\begin{aligned} G(\hat{P}(x_1, \dots, x_N)) &= \int_{\mathbb{R}} h(x) \left(\frac{1}{N} \sum_{i=1}^{N} \delta_{x_i} \right)(dx) \\ &= \frac{1}{N} \sum_{i=1}^{N} h(x_i) \end{aligned}$$
- Moments et variance empiriques
  En prenant $h(X) = X^k$, nous obtenons les moments empiriques
  $$\hat{m}_k(x_1, \dots, x_N) = \frac{1}{N} \sum_{i=1}^{N} x_i^k$$
  On en déduit la variance empirique :
  $G(P) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$, et donc
  $$G(\hat{P}(x_1, \dots, x_N)) = \frac{1}{N} \sum_{i=1}^{N} x_i^2 - \left(\frac{1}{N} \sum_{i=1}^{N} x_i \right)^2$$
  On note S^2 la statistique associée.
- Fonction de répartition empirique
  $[G(P)](x) = \int_{-\infty}^{x} P(dx)$, soit :
  $$\begin{aligned} G(\hat{P}(x_1, \dots, x_N))(x) &= \int_{-\infty}^{x} \left(\frac{1}{N} \sum_{i=1}^{N} \delta_{x_i} \right)(dx) \\ &= \frac{1}{N} \sum_{i=1}^{N} \int_{-\infty}^{x} \delta_{x_i}(dx) \\ &= \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}_{]-\infty;x]}(x_i) \end{aligned}$$

Ex. fonctions de répartitions empiriques pour $\mathcal{N}(10; 5)$



- $\mathbb{E}(h(X))$

  Idem : si $h : (\mathcal{X}, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, P-intégrable.

  $$G : P \mapsto \mathbb{E}(h(X)) = \int_{\mathbb{R}} h(x)\, P(dx)$$

  En appliquant G à la mesure empirique :

  $$\begin{aligned} G(\hat{P}(x_1, \ldots, x_N)) &= \int_{\mathbb{R}} h(x) \left(\frac{1}{N} \sum_{i=1}^{N} \delta_{x_i} \right) (dx) \\ &= \frac{1}{N} \sum_{i=1}^{N} h(x_i) \end{aligned}$$

- Moments et variance empiriques

  En prenant $h(X) = X^k$, nous obtenons les moments empiriques

  $$\hat{m}_k(x_1, \ldots, x_N) = \frac{1}{N} \sum_{i=1}^{N} x_i^k$$

  On en déduit la variance empirique :

  $G(P) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$, et donc

  $$G(\hat{P}(x_1, \ldots, x_N)) = \frac{1}{N} \sum_{i=1}^{N} x_i^2 - \left(\frac{1}{N} \sum_{i=1}^{N} x_i \right)^2$$

  On note S^2 la statistique associée.

- Fonction de répartition empirique

  $[G(P)](x) = \int_{-\infty}^{x} P(dx)$, soit :

  $$\begin{aligned} G(\hat{P}(x_1, \ldots, x_N))(x) &= \int_{-\infty}^{x} \left(\frac{1}{N} \sum_{i=1}^{N} \delta_{x_i} \right) (dx) \\ &= \frac{1}{N} \sum_{i=1}^{N} \int_{-\infty}^{x} \delta_{x_i}(dx) \\ &= \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}_{]-\infty;x]}(x_i) \end{aligned}$$

- $\mathbb{E}(h(X))$
  Idem : si $h : (\mathcal{X}, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, P-intégrable.
  $$G : P \mapsto \mathbb{E}(h(X)) = \int_{\mathbb{R}} h(x)\, P(dx)$$
  En appliquant G à la mesure empirique :
  $$\begin{aligned} G(\hat{P}(x_1, \ldots, x_N)) &= \int_{\mathbb{R}} h(x) \left(\frac{1}{N} \sum_{i=1}^{N} \delta_{x_i} \right) (dx) \\ &= \frac{1}{N} \sum_{i=1}^{N} h(x_i) \end{aligned}$$
- Moments et variance empiriques
  En prenant $h(X) = X^k$, nous obtenons les moments empiriques
  $$\hat{m}_k(x_1, \ldots, x_N) = \frac{1}{N} \sum_{i=1}^{N} x_i^k$$
  On en déduit la variance empirique :
  $G(P) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$, et donc
  $$G(\hat{P}(x_1, \ldots, x_N)) = \frac{1}{N} \sum_{i=1}^{N} x_i^2 - \left(\frac{1}{N} \sum_{i=1}^{N} x_i \right)^2$$
  On note S^2 la statistique associée.
- Fonction de répartition empirique
  $[G(P)](x) = \int_{-\infty}^{x} P(dx)$, soit :
  $$\begin{aligned} G(\hat{P}(x_1, \ldots, x_N))(x) &= \int_{-\infty}^{x} \left(\frac{1}{N} \sum_{i=1}^{N} \delta_{x_i} \right) (dx) \\ &= \frac{1}{N} \sum_{i=1}^{N} \int_{-\infty}^{x} \delta_{x_i}(dx) \\ &= \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}_{]-\infty;x]}(x_i) \end{aligned}$$

I.3 - Modèles Statistiques Paramétriques

Définition

Un modèle statistique sur $(\mathcal{X}, \mathcal{A})$ est une famille de lois de probabilités sur $(\mathcal{X}, \mathcal{A})$.

Si cette famille peut être paramétrée par $\theta \in \Theta \subset \mathbb{R}^p$, on parle de modèle statistique paramétrique. On note :

$$\mathcal{M}_\Theta = \{P_\theta, \theta \in \Theta\}$$

Si la famille est dominée par une mesure σ-finie, alors les lois admettent des densités. On note :

$$\mathcal{M}_\Theta = \{p_\theta, \theta \in \Theta\}$$

- $\mathbb{E}(h(X))$
  Idem : si $h : (\mathcal{X}, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, P-intégrable.
  $$G : P \mapsto \mathbb{E}(h(X)) = \int_{\mathbb{R}} h(x)\, P(dx)$$
  En appliquant G à la mesure empirique :
  $$\begin{aligned} G(\hat{P}(x_1, \ldots, x_N)) &= \int_{\mathbb{R}} h(x) \left(\frac{1}{N} \sum_{i=1}^{N} \delta_{x_i} \right) (dx) \\ &= \frac{1}{N} \sum_{i=1}^{N} h(x_i) \end{aligned}$$
- Moments et variance empiriques
  En prenant $h(X) = X^k$, nous obtenons les moments empiriques
  $$\hat{m}_k(x_1, \ldots, x_N) = \frac{1}{N} \sum_{i=1}^{N} x_i^k$$
  On en déduit la variance empirique :
  $G(P) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$, et donc
  $$G(\hat{P}(x_1, \ldots, x_N)) = \frac{1}{N} \sum_{i=1}^{N} x_i^2 - \left(\frac{1}{N} \sum_{i=1}^{N} x_i \right)^2$$
  On note S^2 la statistique associée.
- Fonction de répartition empirique
  $[G(P)](x) = \int_{-\infty}^{x} P(dx)$, soit :
  $$\begin{aligned} G(\hat{P}(x_1, \ldots, x_N))(x) &= \int_{-\infty}^{x} \left(\frac{1}{N} \sum_{i=1}^{N} \delta_{x_i} \right) (dx) \\ &= \frac{1}{N} \sum_{i=1}^{N} \int_{-\infty}^{x} \delta_{x_i}(dx) \\ &= \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}_{]-\infty;x]}(x_i) \end{aligned}$$

I.3 - Modèles Statistiques Paramétriques

Définition

Un modèle statistique sur $(\mathcal{X}, \mathcal{A})$ est une famille de lois de probabilités sur $(\mathcal{X}, \mathcal{A})$.

Si cette famille peut être paramétrée par $\theta \in \Theta \subset \mathbb{R}^p$, on parle de modèle statistique paramétrique. On note :

$$\mathcal{M}_\Theta = \{P_\theta, \theta \in \Theta\}$$

Si la famille est dominée par une mesure σ-finie, alors les lois admettent des densités. On note :

$$\mathcal{M}_\Theta = \{p_\theta, \theta \in \Theta\}$$

Exemples

- Famille des lois de Poisson
  $\mathcal{M}_\Theta = \{P_\lambda, \lambda \in \mathbb{R}_+\}$, $P_\lambda(X = k) = \dfrac{\lambda^k}{k!} e^{-\lambda}$.

- $\mathbb{E}(h(X))$
  Idem : si $h : (\mathcal{X}, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, P-intégrable.
  $G : P \mapsto \mathbb{E}(h(X)) = \int_{\mathbb{R}} h(x)\, P(dx)$
  En appliquant G à la mesure empirique :
  $$G(\hat{P}(x_1, \ldots, x_N)) = \int_{\mathbb{R}} h(x) \left(\frac{1}{N} \sum_{i=1}^{N} \delta_{x_i} \right) (dx) = \frac{1}{N} \sum_{i=1}^{N} h(x_i)$$
- Moments et variance empiriques
  En prenant $h(X) = X^k$, nous obtenons les moments empiriques
  $$\hat{m}_k(x_1, \ldots, x_N) = \frac{1}{N} \sum_{i=1}^{N} x_i^k$$
  On en déduit la variance empirique :
  $G(P) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$, et donc
  $$G(\hat{P}(x_1, \ldots, x_N)) = \frac{1}{N} \sum_{i=1}^{N} x_i^2 - \left(\frac{1}{N} \sum_{i=1}^{N} x_i \right)^2$$
  On note S^2 la statistique associée.
- Fonction de répartition empirique
  $[G(P)](x) = \int_{-\infty}^{x} P(dx)$, soit :
  $$G(\hat{P}(x_1, \ldots, x_N))(x) = \int_{-\infty}^{x} \left(\frac{1}{N} \sum_{i=1}^{N} \delta_{x_i} \right) (dx) = \frac{1}{N} \sum_{i=1}^{N} \int_{-\infty}^{x} \delta_{x_i}(dx) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}_{]-\infty;x]}(x_i)$$

I.3 - Modèles Statistiques Paramétriques

Définition

Un modèle statistique sur $(\mathcal{X}, \mathcal{A})$ est une famille de lois de probabilités sur $(\mathcal{X}, \mathcal{A})$.

Si cette famille peut être paramétrée par $\theta \in \Theta \subset \mathbb{R}^p$, on parle de modèle statistique paramétrique. On note :

$$\mathcal{M}_\Theta = \{P_\theta, \theta \in \Theta\}$$

Si la famille est dominée par une mesure σ-finie, alors les lois admettent des densités. On note :

$$\mathcal{M}_\Theta = \{p_\theta, \theta \in \Theta\}$$

Exemples

- Famille des lois de Poisson
  $\mathcal{M}_\Theta = \{P_\lambda, \lambda \in \mathbb{R}_+\}$, $P_\lambda(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$.
- Famille des lois Gaussiennes
  $\mathcal{M}_\Theta = \left\{p_{(\mu,\sigma^2)}, \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_+^*\right\}$,
  $P_{(\mu,\sigma^2)}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$.

- $\mathbb{E}(h(X))$

  Idem : si $h : (\mathcal{X}, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, P-intégrable.

  $$G : P \mapsto \mathbb{E}(h(X)) = \int_{\mathbb{R}} h(x)\, P(dx)$$

  En appliquant G à la mesure empirique :

  $$\begin{aligned} G(\hat{P}(x_1, \ldots, x_N)) &= \int_{\mathbb{R}} h(x) \left(\frac{1}{N} \sum_{i=1}^{N} \delta_{x_i} \right)(dx) \\ &= \frac{1}{N} \sum_{i=1}^{N} h(x_i) \end{aligned}$$

- Moments et variance empiriques

  En prenant $h(X) = X^k$, nous obtenons les moments empiriques

  $$\hat{m}_k(x_1, \ldots, x_N) = \frac{1}{N} \sum_{i=1}^{N} x_i^k$$

  On en déduit la variance empirique :

  $$G(P) = \mathbb{E}(X^2) - \mathbb{E}(X)^2, \text{ et donc}$$

  $$G(\hat{P}(x_1, \ldots, x_N)) = \frac{1}{N} \sum_{i=1}^{N} x_i^2 - \left(\frac{1}{N} \sum_{i=1}^{N} x_i \right)^2$$

  On note S^2 la statistique associée.

- Fonction de répartition empirique

  $[G(P)](x) = \int_{-\infty}^{x} P(dx)$, soit :

  $$\begin{aligned} G(\hat{P}(x_1, \ldots, x_N))(x) &= \int_{-\infty}^{x} \left(\frac{1}{N} \sum_{i=1}^{N} \delta_{x_i} \right)(dx) \\ &= \frac{1}{N} \sum_{i=1}^{N} \int_{-\infty}^{x} \delta_{x_i}(dx) \\ &= \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}_{]-\infty;x]}(x_i) \end{aligned}$$

I.3 - Modèles Statistiques Paramétriques

Définition

Un modèle statistique sur $(\mathcal{X}, \mathcal{A})$ est une famille de lois de probabilités sur $(\mathcal{X}, \mathcal{A})$.

Si cette famille peut être paramétrée par $\theta \in \Theta \subset \mathbb{R}^p$, on parle de modèle statistique paramétrique. On note :

$$\mathcal{M}_\Theta = \{P_\theta, \theta \in \Theta\}$$

Si la famille est dominée par une mesure σ-finie, alors les lois admettent des densités. On note :

$$\mathcal{M}_\Theta = \{p_\theta, \theta \in \Theta\}$$

Exemples

- Famille des lois de Poisson

  $$\mathcal{M}_\Theta = \{P_\lambda, \lambda \in \mathbb{R}_+\},\ P_\lambda(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

- Famille des lois Gaussiennes

  $$\mathcal{M}_\Theta = \left\{ p_{(\mu,\sigma^2)}, \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_+^* \right\},$$

  $$p_{(\mu,\sigma^2)}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

- Modèles de régression

  $X \in (\mathcal{X}, \mathcal{A})$: variables explicatives, $Y \in (\mathcal{Y}, \mathcal{B})$: variables à expliquer.

  Un modèle de régression est une famille de lois de densités conditionnelles

  $$\mathcal{M}_\Theta = \{p_\theta(y|x), \theta \in \Theta\}$$

  Par exemple, régression linéaire : Y v.a. dans $\mathbb{R}$, X v.a. dans $\mathbb{R}^p$, $\xi \sim \mathcal{N}(0; \sigma^2)$.

  $$Y = \beta^t X + \xi,$$

  $$p_{(\beta,\sigma^2)}(y|x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\beta^t x}{\sigma}\right)^2}$$

Définition

Un modèle statistique sur $(\mathcal{X}, \mathcal{A})$ est une famille de lois de probabilités sur $(\mathcal{X}, \mathcal{A})$.

Si cette famille peut être paramétrée par $\theta \in \Theta \subset \mathbb{R}^p$, on parle de modèle statistique paramétrique. On note :

$$\mathcal{M}_\Theta = \{P_\theta , \theta \in \Theta\}$$

Si la famille est dominée par une mesure σ-finie, alors les lois admettent des densités. On note :

$$\mathcal{M}_\Theta = \{p_\theta , \theta \in \Theta\}$$

Exemples

- Famille des lois de Poisson
  $$\mathcal{M}_\Theta = \{P_\lambda , \lambda \in \mathbb{R}_+\},\ P_\lambda(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$
- Famille des lois Gaussiennes
  $$\mathcal{M}_\Theta = \left\{p_{(\mu,\sigma^2)}, \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_+^*\right\},$$
  $$P_{(\mu,\sigma^2)}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$
- Modèles de régression

  $X \in (\mathcal{X}, \mathcal{A})$: variables explicatives, $Y \in (\mathcal{Y}, \mathcal{B})$: variables à expliquer.

  Un modèle de régression est une famille de lois de densités conditionnelles
  $$\mathcal{M}_\Theta = \{p_\theta(y|x), \theta \in \Theta\}$$

  Par exemple, régression linéaire : Y v.a. dans $\mathbb{R}$, X v.a. dans $\mathbb{R}^p$, $\xi \sim \mathcal{N}(0; \sigma^2)$.
  $$Y = \beta^t X + \xi,$$
  $$p_{(\beta,\sigma^2)}(y|x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\beta^t x}{\sigma}\right)^2}$$

I.3 - Modèles Statistiques Paramétriques

Définition

Un modèle statistique sur $(\mathcal{X}, \mathcal{A})$ est une famille de lois de probabilités sur $(\mathcal{X}, \mathcal{A})$.

Si cette famille peut être paramétrée par $\theta \in \Theta \subset \mathbb{R}^p$, on parle de modèle statistique paramétrique. On note :

$$\mathcal{M}_\Theta = \{P_\theta, \theta \in \Theta\}$$

Si la famille est dominée par une mesure σ-finie, alors les lois admettent des densités. On note :

$$\mathcal{M}_\Theta = \{p_\theta, \theta \in \Theta\}$$

Exemples

- Famille des lois de Poisson
  $$\mathcal{M}_\Theta = \{P_\lambda, \lambda \in \mathbb{R}_+\},\ P_\lambda(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$
- Famille des lois Gaussiennes
  $$\mathcal{M}_\Theta = \left\{p_{(\mu,\sigma^2)}, \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_+^*\right\},$$
  $$P_{(\mu,\sigma^2)}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$
- Modèles de régression

  $X \in (\mathcal{X}, \mathcal{A})$: variables explicatives, $Y \in (\mathcal{Y}, \mathcal{B})$: variables à expliquer.

  Un modèle de régression est une famille de lois de densités conditionnelles
  $$\mathcal{M}_\Theta = \{p_\theta(y|x), \theta \in \Theta\}$$

  Par exemple, régression linéaire : Y v.a. dans $\mathbb{R}$, X v.a. dans $\mathbb{R}^p$, $\xi \sim \mathcal{N}(0; \sigma^2)$.
  $$Y = \beta^t X + \xi,$$
  $$p_{(\beta,\sigma^2)}(y|x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\beta^t x}{\sigma}\right)^2}$$

II - Estimation Paramétrique

Soit $X \sim P^*$, P^* inconnue. On suppose que P^* appartient à un modèle paramétrique $P^* \in \mathcal{M}_\Theta \{P_\theta, \theta \in \Theta\}$, et donc $\exists \theta^* \in \Theta$ tel que $P^* = P_{\theta^*}$.

I.3 - Modèles Statistiques Paramétriques

Définition

Un modèle statistique sur $(\mathcal{X}, \mathcal{A})$ est une famille de lois de probabilités sur $(\mathcal{X}, \mathcal{A})$.

Si cette famille peut être paramétrée par $\theta \in \Theta \subset \mathbb{R}^p$, on parle de modèle statistique paramétrique. On note :

$$\mathcal{M}_\Theta = \{P_\theta \, , \theta \in \Theta\}$$

Si la famille est dominée par une mesure σ-finie, alors les lois admettent des densités. On note :

$$\mathcal{M}_\Theta = \{p_\theta \, , \theta \in \Theta\}$$

Exemples

- Famille des lois de Poisson
$$\mathcal{M}_\Theta = \{P_\lambda \, , \lambda \in \mathbb{R}_+\}, \; P_\lambda(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$
- Famille des lois Gaussiennes
$$\mathcal{M}_\Theta = \left\{p_{(\mu,\sigma^2)}, \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_+^*\right\},$$
$$P_{(\mu,\sigma^2)}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$
- Modèles de régression

  $X \in (\mathcal{X}, \mathcal{A})$: variables explicatives, $Y \in (\mathcal{Y}, \mathcal{B})$: variables à expliquer.

  Un modèle de régression est une famille de lois de densités conditionnelles
$$\mathcal{M}_\Theta = \{p_\theta(y|x), \theta \in \Theta\}$$
  Par exemple, régression linéaire : Y v.a. dans $\mathbb{R}$, X v.a. dans $\mathbb{R}^p$, $\xi \sim \mathcal{N}(0; \sigma^2)$.
$$Y = \beta^t X + \xi,$$
$$p_{(\beta,\sigma^2)}(y|x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\beta^t x}{\sigma}\right)^2}$$

II - Estimation Paramétrique

Soit $X \sim P^*$, P^* inconnue. On suppose que P^* appartient à un modèle paramétrique $P^* \in \mathcal{M}_\Theta \{P_\theta \, , \theta \in \Theta\}$, et donc $\exists \theta^* \in \Theta$ tel que $P^* = P_{\theta^*}$.

$\Longrightarrow$ Le problème d'inférence sur P^* est alors ramené au problème d'estimation du paramètre θ^*.

I.3 - Modèles Statistiques Paramétriques

Définition

Un modèle statistique sur $(\mathcal{X}, \mathcal{A})$ est une famille de lois de probabilités sur $(\mathcal{X}, \mathcal{A})$.

Si cette famille peut être paramétrée par $\theta \in \Theta \subset \mathbb{R}^p$, on parle de modèle statistique paramétrique. On note :

$$\mathcal{M}_\Theta = \{P_\theta \,, \theta \in \Theta\}$$

Si la famille est dominée par une mesure σ-finie, alors les lois admettent des densités. On note :

$$\mathcal{M}_\Theta = \{p_\theta \,, \theta \in \Theta\}$$

Exemples

- Famille des lois de Poisson
$$\mathcal{M}_\Theta = \{P_\lambda \,, \lambda \in \mathbb{R}_+\},\ P_\lambda(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$
- Famille des lois Gaussiennes
$$\mathcal{M}_\Theta = \left\{p_{(\mu,\sigma^2)}, \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_+^*\right\},$$
$$P_{(\mu,\sigma^2)}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$
- Modèles de régression

  $X \in (\mathcal{X}, \mathcal{A})$: variables explicatives, $Y \in (\mathcal{Y}, \mathcal{B})$: variables à expliquer.

  Un modèle de régression est une famille de lois de densités conditionnelles
$$\mathcal{M}_\Theta = \{p_\theta(y|x), \theta \in \Theta\}$$

  Par exemple, régression linéaire : Y v.a. dans $\mathbb{R}$, X v.a. dans $\mathbb{R}^p$, $\xi \sim \mathcal{N}(0; \sigma^2)$.
$$Y = \beta^t X + \xi,$$
$$p_{(\beta,\sigma^2)}(y|x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\beta^t x}{\sigma}\right)^2}$$

II - Estimation Paramétrique

Soit $X \sim P^*$, P^* inconnue. On suppose que P^* appartient à un modèle paramétrique $P^* \in \mathcal{M}_\Theta \{P_\theta \,, \theta \in \Theta\}$, et donc $\exists \theta^* \in \Theta$ tel que $P^* = P_{\theta^*}$.

$\Longrightarrow$ Le problème d'inférence sur P^* est alors ramené au problème d'estimation du paramètre θ^*.

Définition

Soit $(X_1, \ldots, X_N)$ un échantillon i.i.d. pour une loi P_{θ^*}. Un estimateur de θ^* est une statistique $T(X_1, \ldots, X_N)$ à valeurs dans Θ visant à estimer θ^*.

I.3 - Modèles Statistiques Paramétriques

Définition

Un modèle statistique sur $(\mathcal{X}, \mathcal{A})$ est une famille de lois de probabilités sur $(\mathcal{X}, \mathcal{A})$.

Si cette famille peut être paramétrée par $\theta \in \Theta \subset \mathbb{R}^p$, on parle de modèle statistique paramétrique. On note :

$$\mathcal{M}_\Theta = \{P_\theta , \theta \in \Theta\}$$

Si la famille est dominée par une mesure σ-finie, alors les lois admettent des densités. On note :

$$\mathcal{M}_\Theta = \{p_\theta , \theta \in \Theta\}$$

Exemples

- Famille des lois de Poisson
$$\mathcal{M}_\Theta = \{P_\lambda , \lambda \in \mathbb{R}_+\},\ P_\lambda(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$
- Famille des lois Gaussiennes
$$\mathcal{M}_\Theta = \left\{p_{(\mu,\sigma^2)}, \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_+^*\right\},$$
$$P_{(\mu,\sigma^2)}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$
- Modèles de régression

  $X \in (\mathcal{X}, \mathcal{A})$: variables explicatives, $Y \in (\mathcal{Y}, \mathcal{B})$: variables à expliquer.

  Un modèle de régression est une famille de lois de densités conditionnelles
$$\mathcal{M}_\Theta = \{p_\theta(y|x), \theta \in \Theta\}$$
  Par exemple, régression linéaire : Y v.a. dans $\mathbb{R}$, X v.a. dans $\mathbb{R}^p$, $\xi \sim \mathcal{N}(0; \sigma^2)$.
$$Y = \beta^t X + \xi,$$
$$p_{(\beta,\sigma^2)}(y|x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\beta^t x}{\sigma}\right)^2}$$

II - Estimation Paramétrique

Soit $X \sim P^*$, P^* inconnue. On suppose que P^* appartient à un modèle paramétrique $P^* \in \mathcal{M}_\Theta \{P_\theta , \theta \in \Theta\}$, et donc $\exists \theta^* \in \Theta$ tel que $P^* = P_{\theta^*}$.

$\Longrightarrow$ Le problème d'inférence sur P^* est alors ramené au problème d'estimation du paramètre θ^*.

Définition

Soit $(X_1, \ldots, X_N)$ un échantillon i.i.d. pour une loi P_{θ^*}. Un estimateur de θ^* est une statistique $T(X_1, \ldots, X_N)$ à valeurs dans Θ visant à estimer θ^*.

II.1 - Méthodes d'estimation ponctuelle

II.1.a - Méthode des moments

I.3 - Modèles Statistiques Paramétriques

Définition

Un modèle statistique sur $(\mathcal{X}, \mathcal{A})$ est une famille de lois de probabilités sur $(\mathcal{X}, \mathcal{A})$.

Si cette famille peut être paramétrée par $\theta \in \Theta \subset \mathbb{R}^p$, on parle de modèle statistique paramétrique. On note :

$$\mathcal{M}_\Theta = \{P_\theta, \theta \in \Theta\}$$

Si la famille est dominée par une mesure σ-finie, alors les lois admettent des densités. On note :

$$\mathcal{M}_\Theta = \{p_\theta, \theta \in \Theta\}$$

Exemples

- Famille des lois de Poisson
  $\mathcal{M}_\Theta = \{P_\lambda, \lambda \in \mathbb{R}_+\}$, $P_\lambda(X = k) = \dfrac{\lambda^k}{k!} e^{-\lambda}$.
- Famille des lois Gaussiennes
  $$\mathcal{M}_\Theta = \left\{p_{(\mu,\sigma^2)}, \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_+^*\right\},$$
  $$P_{(\mu,\sigma^2)}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$
- Modèles de régression

  $X \in (\mathcal{X}, \mathcal{A})$: variables explicatives, $Y \in (\mathcal{Y}, \mathcal{B})$: variables à expliquer.

  Un modèle de régression est une famille de lois de densités conditionnelles
  $$\mathcal{M}_\Theta = \{p_\theta(y|x), \theta \in \Theta\}$$

  Par exemple, régression linéaire : Y v.a. dans $\mathbb{R}$, X v.a. dans $\mathbb{R}^p$, $\xi \sim \mathcal{N}(0; \sigma^2)$.
  $$Y = \beta^t X + \xi,$$
  $$p_{(\beta,\sigma^2)}(y|x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\beta^t x}{\sigma}\right)^2}$$

II - Estimation Paramétrique

Soit $X \sim P^*$, P^* inconnue. On suppose que P^* appartient à un modèle paramétrique $P^* \in \mathcal{M}_\Theta \{P_\theta, \theta \in \Theta\}$, et donc $\exists \theta^* \in \Theta$ tel que $P^* = P_{\theta^*}$.

$\Longrightarrow$ Le problème d'inférence sur P^* est alors ramené au problème d'estimation du paramètre θ^*.

Définition

Soit $(X_1, \ldots, X_N)$ un échantillon i.i.d. pour une loi P_{θ^*}. Un estimateur de θ^* est une statistique $T(X_1, \ldots, X_N)$ à valeurs dans Θ visant à estimer θ^*.

II.1 - Méthodes d'estimation ponctuelle

II.1.a - Méthode des moments

$X \sim P_{\theta^*}$, $P_{\theta^*} \in \mathcal{M}_\Theta = \{P_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$

On considère la fonctionnelle G sur l'espaces des mesures de probabilités sur $(\mathcal{X}, \mathcal{A})$, $G : P \mapsto (\mathbb{E}_P(X), \ldots, \mathbb{E}_P(X^p))$.

Alors, pour tout $\theta \in \Theta$, $G(P_\theta) := g(\theta)$.

I.3 - Modèles Statistiques Paramétriques

Définition

Un modèle statistique sur $(\mathcal{X}, \mathcal{A})$ est une famille de lois de probabilités sur $(\mathcal{X}, \mathcal{A})$.

Si cette famille peut être paramétrée par $\theta \in \Theta \subset \mathbb{R}^p$, on parle de modèle statistique paramétrique. On note :

$$\mathcal{M}_\Theta = \{P_\theta, \theta \in \Theta\}$$

Si la famille est dominée par une mesure σ-finie, alors les lois admettent des densités. On note :

$$\mathcal{M}_\Theta = \{p_\theta, \theta \in \Theta\}$$

Exemples

- Famille des lois de Poisson
$$\mathcal{M}_\Theta = \{P_\lambda, \lambda \in \mathbb{R}_+\},\ P_\lambda(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$
- Famille des lois Gaussiennes
$$\mathcal{M}_\Theta = \left\{p_{(\mu,\sigma^2)}, \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_+^*\right\},$$
$$P_{(\mu,\sigma^2)}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$
- Modèles de régression

  $X \in (\mathcal{X}, \mathcal{A})$: variables explicatives, $Y \in (\mathcal{Y}, \mathcal{B})$: variables à expliquer.

  Un modèle de régression est une famille de lois de densités conditionnelles
$$\mathcal{M}_\Theta = \{p_\theta(y|x), \theta \in \Theta\}$$

  Par exemple, régression linéaire : Y v.a. dans $\mathbb{R}$, X v.a. dans $\mathbb{R}^p$, $\xi \sim \mathcal{N}(0; \sigma^2)$.
$$Y = \beta^t X + \xi,$$
$$p_{(\beta,\sigma^2)}(y|x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\beta^t x}{\sigma}\right)^2}$$

II - Estimation Paramétrique

Soit $X \sim P^*$, P^* inconnue. On suppose que P^* appartient à un modèle paramétrique $P^* \in \mathcal{M}_\Theta \{P_\theta, \theta \in \Theta\}$, et donc $\exists \theta^* \in \Theta$ tel que $P^* = P_{\theta^*}$.

$\Longrightarrow$ Le problème d'inférence sur P^* est alors ramené au problème d'estimation du paramètre θ^*.

Définition

Soit $(X_1, \ldots, X_N)$ un échantillon i.i.d. pour une loi P_{θ^*}. Un estimateur de θ^* est une statistique $T(X_1, \ldots, X_N)$ à valeurs dans Θ visant à estimer θ^*.

II.1 - Méthodes d'estimation ponctuelle

II.1.a - Méthode des moments

$X \sim P_{\theta^*}$, $P_{\theta^*} \in \mathcal{M}_\Theta = \{P_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$

On considère la fonctionnelle G sur l'espaces des mesures de probabilités sur $(\mathcal{X}, \mathcal{A})$, $G : P \mapsto (\mathbb{E}_P(X), \ldots, \mathbb{E}_P(X^p))$.

Alors, pour tout $\theta \in \Theta$, $G(P_\theta) := g(\theta)$.

Soit $(x_1, \ldots, x_N)$ un échantillon d'observations i.i.d. pour X, on calcule G en la mesure empirique

$\hat{g} = G\left(\hat{P}(x_1, \ldots, x_N)\right) \Rightarrow p$ premiers moments empiriques

et on cherche $\hat{\theta}$ tel que $g(\hat{\theta}) = \hat{g}$
(ou tel que $\hat{\theta} = \arg\min_\theta \|g(\theta) - \hat{g}\|^2$)

I.3 - Modèles Statistiques Paramétriques

Définition

Un modèle statistique sur $(\mathcal{X}, \mathcal{A})$ est une famille de lois de probabilités sur $(\mathcal{X}, \mathcal{A})$.

Si cette famille peut être paramétrée par $\theta \in \Theta \subset \mathbb{R}^p$, on parle de modèle statistique paramétrique. On note :

$$\mathcal{M}_\Theta = \{P_\theta, \theta \in \Theta\}$$

Si la famille est dominée par une mesure σ-finie, alors les lois admettent des densités. On note :

$$\mathcal{M}_\Theta = \{p_\theta, \theta \in \Theta\}$$

Exemples

- Famille des lois de Poisson
  $\mathcal{M}_\Theta = \{P_\lambda, \lambda \in \mathbb{R}_+\}$, $P_\lambda(X = k) = \dfrac{\lambda^k}{k!} e^{-\lambda}$.
- Famille des lois Gaussiennes
  $$\mathcal{M}_\Theta = \left\{p_{(\mu,\sigma^2)}, \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_+^*\right\},$$
  $$P_{(\mu,\sigma^2)}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$
- Modèles de régression

  $X \in (\mathcal{X}, \mathcal{A})$: variables explicatives, $Y \in (\mathcal{Y}, \mathcal{B})$: variables à expliquer.

  Un modèle de régression est une famille de lois de densités conditionnelles
  $$\mathcal{M}_\Theta = \{p_\theta(y|x), \theta \in \Theta\}$$
  Par exemple, régression linéaire : Y v.a. dans $\mathbb{R}$, X v.a. dans $\mathbb{R}^p$, $\xi \sim \mathcal{N}(0; \sigma^2)$.
  $$Y = \beta^t X + \xi,$$
  $$p_{(\beta,\sigma^2)}(y|x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\beta^t x}{\sigma}\right)^2}$$

II - Estimation Paramétrique

Soit $X \sim P^*$, P^* inconnue. On suppose que P^* appartient à un modèle paramétrique $P^* \in \mathcal{M}_\Theta \{P_\theta, \theta \in \Theta\}$, et donc $\exists \theta^* \in \Theta$ tel que $P^* = P_{\theta^*}$.

$\Longrightarrow$ Le problème d'inférence sur P^* est alors ramené au problème d'estimation du paramètre θ^*.

Définition

Soit $(X_1, \ldots, X_N)$ un échantillon i.i.d. pour une loi P_{θ^*}. Un estimateur de θ^* est une statistique $T(X_1, \ldots, X_N)$ à valeurs dans Θ visant à estimer θ^*.

II.1 - Méthodes d'estimation ponctuelle

II.1.a - Méthode des moments

$X \sim P_{\theta^*}$, $P_{\theta^*} \in \mathcal{M}_\Theta = \{P_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$

On considère la fonctionnelle G sur l'espaces des mesures de probabilités sur $(\mathcal{X}, \mathcal{A})$, $G : P \mapsto (\mathbb{E}_P(X), \ldots, \mathbb{E}_P(X^p))$.

Alors, pour tout $\theta \in \Theta$, $G(P_\theta) := g(\theta)$.

Soit $(x_1, \ldots, x_N)$ un échantillon d'observations i.i.d. pour X, on calcule G en la mesure empirique

$\hat{g} = G\left(\hat{P}(x_1, \ldots, x_N)\right) \Rightarrow p$ premiers moments empiriques

et on cherche $\hat{\theta}$ tel que $g(\hat{\theta}) = \hat{g}$
(ou tel que $\hat{\theta} = \arg\min_\theta \|g(\theta) - \hat{g}\|^2$)

$\Longrightarrow$ $\hat{\theta}$ est un estimateur de θ^*.

II - Estimation Paramétrique

Soit $X \sim P^*$, P^* inconnue. On suppose que P^* appartient à un modèle paramétrique $P^* \in \mathcal{M}_\Theta \{P_\theta, \theta \in \Theta\}$, et donc $\exists \theta^* \in \Theta$ tel que $P^* = P_{\theta^*}$.

$\Longrightarrow$ Le problème d'inférence sur P^* est alors ramené au problème d'estimation du paramètre θ^*.

Définition

Soit $(X_1, \dots, X_N)$ un échantillon i.i.d. pour une loi P_{θ^*}. Un estimateur de θ^* est une statistique $T(X_1, \dots, X_N)$ à valeurs dans Θ visant à estimer θ^*.

II.1 - Méthodes d'estimation ponctuelle

II.1.a - Méthode des moments

$X \sim P_{\theta^*}$, $P_{\theta^*} \in \mathcal{M}_\Theta = \{P_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$

On considère la fonctionnelle G sur l'espaces des mesures de probabilités sur $(\mathcal{X}, \mathcal{A})$, $G : P \mapsto (\mathbb{E}_P(X), \dots, \mathbb{E}_P(X^p))$.

Alors, pour tout $\theta \in \Theta$, $G(P_\theta) := g(\theta)$.

Soit $(x_1, \dots, x_N)$ un échantillon d'observations i.i.d. pour X, on calcule G en la mesure empirique

$\hat{g} = G\left(\hat{P}(x_1, \dots, x_N)\right) \Rightarrow p$ premiers moments empiriques

et on cherche $\hat{\theta}$ tel que $g(\hat{\theta}) = \hat{g}$
(ou tel que $\hat{\theta} = \arg\min_\theta \|g(\theta) - \hat{g}\|^2$)

$\Longrightarrow$ $\hat{\theta}$ est un estimateur de θ^*.

II - Estimation Paramétrique

Soit $X \sim P^*$, P^* inconnue. On suppose que P^* appartient à un modèle paramétrique $P^* \in \mathcal{M}_\Theta \{P_\theta, \theta \in \Theta\}$, et donc $\exists \theta^* \in \Theta$ tel que $P^* = P_{\theta^*}$.

$\Longrightarrow$ Le problème d'inférence sur P^* est alors ramené au problème d'estimation du paramètre θ^*.

Définition

Soit $(X_1, \ldots, X_N)$ un échantillon i.i.d. pour une loi P_{θ^*}. Un estimateur de θ^* est une statistique $T(X_1, \ldots, X_N)$ à valeurs dans Θ visant à estimer θ^*.

II.1 - Méthodes d'estimation ponctuelle

II.1.a - Méthode des moments

$X \sim P_{\theta^*}$, $P_{\theta^*} \in \mathcal{M}_\Theta = \{P_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$

On considère la fonctionnelle G sur l'espaces des mesures de probabilités sur $(\mathcal{X}, \mathcal{A})$, $G : P \mapsto (\mathbb{E}_P(X), \ldots, \mathbb{E}_P(X^p))$.

Alors, pour tout $\theta \in \Theta$, $G(P_\theta) := g(\theta)$.

Soit $(x_1, \ldots, x_N)$ un échantillon d'observations i.i.d. pour X, on calcule G en la mesure empirique

$\hat{g} = G\left(\hat{P}(x_1, \ldots, x_N)\right) \Rightarrow p$ premiers moments empiriques

et on cherche $\hat{\theta}$ tel que $g(\hat{\theta}) = \hat{g}$
(ou tel que $\hat{\theta} = \arg\min_\theta \|g(\theta) - \hat{g}\|^2$)

$\Longrightarrow$ $\hat{\theta}$ est un estimateur de θ^*.

II.1.b - Maximum de vraisemblance

Définition

Soit $\mathcal{M}_\Theta$, modèle paramétrique dominé, représenté par une famille de densités : $\mathcal{M}_\Theta = \{p_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$

Soit $X \sim p_{\theta^*}$, $p_{\theta^*} \in \mathcal{M}_\Theta$. Soit une observation $x \in \mathcal{X}$ pour la variable aléatoire X.

La fonction : $\theta \in \Theta$, $\theta \mapsto p_\theta(x)$ est appelée vraisemblance du paramètre θ en x, on la notera $\mathcal{L}(\theta; x)$.

Pour un échantillon d'observations $(x_1, x_2, \ldots, x_N)$ pour des variables aléatoires i.i.d, la vraisemblance s'écrit :

$$\mathcal{L}(\theta; x_1, x_2, \ldots, x_N) = \prod_{i=1}^N p_\theta(x_i) ,$$

II - Estimation Paramétrique

Soit $X \sim P^*$, P^* inconnue. On suppose que P^* appartient à un modèle paramétrique $P^* \in \mathcal{M}_\Theta \{P_\theta, \theta \in \Theta\}$, et donc $\exists \theta^* \in \Theta$ tel que $P^* = P_{\theta^*}$.

$\Longrightarrow$ Le problème d'inférence sur P^* est alors ramené au problème d'estimation du paramètre θ^*.

Définition

Soit $(X_1, \dots, X_N)$ un échantillon i.i.d. pour une loi P_{θ^*}. Un estimateur de θ^* est une statistique $T(X_1, \dots, X_N)$ à valeurs dans Θ visant à estimer θ^*.

II.1 - Méthodes d'estimation ponctuelle

II.1.a - Méthode des moments

$X \sim P_{\theta^*}$, $P_{\theta^*} \in \mathcal{M}_\Theta = \{P_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$

On considère la fonctionnelle G sur l'espaces des mesures de probabilités sur $(\mathcal{X}, \mathcal{A})$, $G : P \mapsto (\mathbb{E}_P(X), \dots, \mathbb{E}_P(X^p))$.

Alors, pour tout $\theta \in \Theta$, $G(P_\theta) := g(\theta)$.

Soit $(x_1, \dots, x_N)$ un échantillon d'observations i.i.d. pour X, on calcule G en la mesure empirique

$\hat{g} = G\left(\hat{P}(x_1, \dots, x_N)\right) \Rightarrow p$ premiers moments empiriques

et on cherche $\hat{\theta}$ tel que $g(\hat{\theta}) = \hat{g}$
(ou tel que $\hat{\theta} = \arg\min_\theta \|g(\theta) - \hat{g}\|^2$)

$\Longrightarrow$ $\hat{\theta}$ est un estimateur de θ^*.

II.1.b - Maximum de vraisemblance

Définition

Soit $\mathcal{M}_\Theta$, modèle paramétrique dominé, représenté par une famille de densités : $\mathcal{M}_\Theta = \{p_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$

Soit $X \sim p_{\theta^*}$, $p_{\theta^*} \in \mathcal{M}_\Theta$. Soit une observation $x \in \mathcal{X}$ pour la variable aléatoire X.

La fonction : $\theta \in \Theta, \theta \mapsto p_\theta(x)$ est appelée vraisemblance du paramètre θ en x, on la notera $\mathcal{L}(\theta; x)$.

Pour un échantillon d'observations $(x_1, x_2, \dots, x_N)$ pour des variables aléatoires i.i.d, la vraisemblance s'écrit :

$$\mathcal{L}(\theta; x_1, x_2, \dots, x_N) = \prod_{i=1}^N p_\theta(x_i) \ ,$$

ou

$$\mathcal{L}(\theta; x_1, x_2, \dots, x_N) = \prod_{i=1}^N p(x_i|\theta) \ .$$

II - Estimation Paramétrique

Soit $X \sim P^*$, P^* inconnue. On suppose que P^* appartient à un modèle paramétrique $P^* \in \mathcal{M}_\Theta \{P_\theta, \theta \in \Theta\}$, et donc $\exists \theta^* \in \Theta$ tel que $P^* = P_{\theta^*}$.

$\Longrightarrow$ Le problème d'inférence sur P^* est alors ramené au problème d'estimation du paramètre θ^*.

Définition

Soit $(X_1, \dots, X_N)$ un échantillon i.i.d. pour une loi P_{θ^*}. Un estimateur de θ^* est une statistique $T(X_1, \dots, X_N)$ à valeurs dans Θ visant à estimer θ^*.

II.1 - Méthodes d'estimation ponctuelle

II.1.a - Méthode des moments

$X \sim P_{\theta^*}$, $P_{\theta^*} \in \mathcal{M}_\Theta = \{P_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$

On considère la fonctionnelle G sur l'espaces des mesures de probabilités sur $(\mathcal{X}, \mathcal{A})$, $G : P \mapsto (\mathbb{E}_P(X), \dots, \mathbb{E}_P(X^p))$.

Alors, pour tout $\theta \in \Theta$, $G(P_\theta) := g(\theta)$.

Soit $(x_1, \dots, x_N)$ un échantillon d'observations i.i.d. pour X, on calcule G en la mesure empirique

$\hat{g} = G\left(\hat{P}(x_1, \dots, x_N)\right) \Rightarrow p$ premiers moments empiriques

et on cherche $\hat{\theta}$ tel que $g(\hat{\theta}) = \hat{g}$
(ou tel que $\hat{\theta} = \arg\min_\theta \|g(\theta) - \hat{g}\|^2$)

$\Longrightarrow$ $\hat{\theta}$ est un estimateur de θ^*.

II.1.b - Maximum de vraisemblance

Définition

Soit $\mathcal{M}_\Theta$, modèle paramétrique dominé, représenté par une famille de densités : $\mathcal{M}_\Theta = \{p_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$

Soit $X \sim p_{\theta^*}$, $p_{\theta^*} \in \mathcal{M}_\Theta$. Soit une observation $x \in \mathcal{X}$ pour la variable aléatoire X.

La fonction : $\theta \in \Theta$, $\theta \mapsto p_\theta(x)$ est appelée vraisemblance du paramètre θ en x, on la notera $\mathcal{L}(\theta; x)$.

Pour un échantillon d'observations $(x_1, x_2, \dots, x_N)$ pour des variables aléatoires i.i.d, la vraisemblance s'écrit :

$$\mathcal{L}(\theta; x_1, x_2, \dots, x_N) = \prod_{i=1}^N p_\theta(x_i) ,$$

ou

$$\mathcal{L}(\theta; x_1, x_2, \dots, x_N) = \prod_{i=1}^N p(x_i|\theta) .$$

La vraisemblance du paramètre θ correspond ainsi à sa plausibilité pour un certain jeu de données observées.

II - Estimation Paramétrique

Soit $X \sim P^*$, P^* inconnue. On suppose que P^* appartient à un modèle paramétrique $P^* \in \mathcal{M}_\Theta \{P_\theta, \theta \in \Theta\}$, et donc $\exists \theta^* \in \Theta$ tel que $P^* = P_{\theta^*}$.

$\Longrightarrow$ Le problème d'inférence sur P^* est alors ramené au problème d'estimation du paramètre θ^*.

Définition

Soit $(X_1, \ldots, X_N)$ un échantillon i.i.d. pour une loi P_{θ^*}. Un estimateur de θ^* est une statistique $T(X_1, \ldots, X_N)$ à valeurs dans Θ visant à estimer θ^*.

II.1 - Méthodes d'estimation ponctuelle

II.1.a - Méthode des moments

$X \sim P_{\theta^*}$, $P_{\theta^*} \in \mathcal{M}_\Theta = \{P_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$

On considère la fonctionnelle G sur l'espaces des mesures de probabilités sur $(\mathcal{X}, \mathcal{A})$, $G : P \mapsto (\mathbb{E}_P(X), \ldots, \mathbb{E}_P(X^p))$.

Alors, pour tout $\theta \in \Theta$, $G(P_\theta) := g(\theta)$.

Soit $(x_1, \ldots, x_N)$ un échantillon d'observations i.i.d. pour X, on calcule G en la mesure empirique

$\hat{g} = G\left(\hat{P}(x_1, \ldots, x_N)\right) \Rightarrow p$ premiers moments empiriques

et on cherche $\hat{\theta}$ tel que $g(\hat{\theta}) = \hat{g}$
(ou tel que $\hat{\theta} = \arg\min_\theta \|g(\theta) - \hat{g}\|^2$)

$\Longrightarrow$ $\hat{\theta}$ est un estimateur de θ^*.

II.1.b - Maximum de vraisemblance

Définition

Soit $\mathcal{M}_\Theta$, modèle paramétrique dominé, représenté par une famille de densités : $\mathcal{M}_\Theta = \{p_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$

Soit $X \sim p_{\theta^*}$, $p_{\theta^*} \in \mathcal{M}_\Theta$. Soit une observation $x \in \mathcal{X}$ pour la variable aléatoire X.

La fonction : $\theta \in \Theta, \theta \mapsto p_\theta(x)$ est appelée vraisemblance du paramètre θ en x, on la notera $\mathcal{L}(\theta; x)$.

Pour un échantillon d'observations $(x_1, x_2, \ldots, x_N)$ pour des variables aléatoires i.i.d, la vraisemblance s'écrit :

$$\mathcal{L}(\theta; x_1, x_2, \ldots, x_N) = \prod_{i=1}^N p_\theta(x_i) \ ,$$

ou

$$\mathcal{L}(\theta; x_1, x_2, \ldots, x_N) = \prod_{i=1}^N p(x_i|\theta) \ .$$

La vraisemblance du paramètre θ correspond ainsi à sa plausibilité pour un certain jeu de données observées.

Définition

Pour une observation ou un échantillon d'observations x, si $\hat{\theta}(x)$ maximise la vraisemblance $\mathcal{L}(\theta; x)$ sur Θ
$\hat{\theta}(x)$ sera appelée estimation du maximum de vraisemblance et l'estimateur associé $\hat{\theta}(X)$, estimateur du maximum de vraisemblance (EMV).

II - Estimation Paramétrique

Soit $X \sim P^*$, P^* inconnue. On suppose que P^* appartient à un modèle paramétrique $P^* \in \mathcal{M}_\Theta \{P_\theta, \theta \in \Theta\}$, et donc $\exists \theta^* \in \Theta$ tel que $P^* = P_{\theta^*}$.

$\Longrightarrow$ Le problème d'inférence sur P^* est alors ramené au problème d'estimation du paramètre θ^*.

Définition

Soit $(X_1, \ldots, X_N)$ un échantillon i.i.d. pour une loi P_{θ^*}. Un estimateur de θ^* est une statistique $T(X_1, \ldots, X_N)$ à valeurs dans Θ visant à estimer θ^*.

II.1 - Méthodes d'estimation ponctuelle

II.1.a - Méthode des moments

$X \sim P_{\theta^*}$, $P_{\theta^*} \in \mathcal{M}_\Theta = \{P_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$

On considère la fonctionnelle G sur l'espaces des mesures de probabilités sur $(\mathcal{X}, \mathcal{A})$, $G : P \mapsto (\mathbb{E}_P(X), \ldots, \mathbb{E}_P(X^p))$.

Alors, pour tout $\theta \in \Theta$, $G(P_\theta) := g(\theta)$.

Soit $(x_1, \ldots, x_N)$ un échantillon d'observations i.i.d. pour X, on calcule G en la mesure empirique

$\hat{g} = G\left(\hat{P}(x_1, \ldots, x_N)\right) \Rightarrow p$ premiers moments empiriques

et on cherche $\hat{\theta}$ tel que $g(\hat{\theta}) = \hat{g}$
(ou tel que $\hat{\theta} = \arg\min_\theta \|g(\theta) - \hat{g}\|^2$)

$\Longrightarrow$ $\hat{\theta}$ est un estimateur de θ^*.

II.1.b - Maximum de vraisemblance

Définition

Soit $\mathcal{M}_\Theta$, modèle paramétrique dominé, représenté par une famille de densités : $\mathcal{M}_\Theta = \{p_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$

Soit $X \sim p_{\theta^*}$, $p_{\theta^*} \in \mathcal{M}_\Theta$. Soit une observation $x \in \mathcal{X}$ pour la variable aléatoire X.

La fonction : $\theta \in \Theta, \theta \mapsto p_\theta(x)$ est appelée vraisemblance du paramètre θ en x, on la notera $\mathcal{L}(\theta; x)$.

Pour un échantillon d'observations $(x_1, x_2, \ldots, x_N)$ pour des variables aléatoires i.i.d, la vraisemblance s'écrit :

$$\mathcal{L}(\theta; x_1, x_2, \ldots, x_N) = \prod_{i=1}^N p_\theta(x_i) ,$$

ou

$$\mathcal{L}(\theta; x_1, x_2, \ldots, x_N) = \prod_{i=1}^N p(x_i|\theta) .$$

La vraisemblance du paramètre θ correspond ainsi à sa plausibilité pour un certain jeu de données observées.

Définition

Pour une observation ou un échantillon d'observations x, si $\hat{\theta}(x)$ maximise la vraisemblance $\mathcal{L}(\theta; x)$ sur Θ
$\hat{\theta}(x)$ sera appelée estimation du maximum de vraisemblance et l'estimateur associé $\hat{\theta}(X)$, estimateur du maximum de vraisemblance (EMV).

On note :

$$\hat{\theta}(x) = \arg\max_{\theta \in \Theta} \mathcal{L}(\theta; x).$$

II - Estimation Paramétrique

Soit $X \sim P^*$, P^* inconnue. On suppose que P^* appartient à un modèle paramétrique $P^* \in \mathcal{M}_\Theta \{P_\theta, \theta \in \Theta\}$, et donc $\exists \theta^* \in \Theta$ tel que $P^* = P_{\theta^*}$.

$\Longrightarrow$ Le problème d'inférence sur P^* est alors ramené au problème d'estimation du paramètre θ^*.

Définition

Soit $(X_1, \ldots, X_N)$ un échantillon i.i.d. pour une loi P_{θ^*}. Un estimateur de θ^* est une statistique $T(X_1, \ldots, X_N)$ à valeurs dans Θ visant à estimer θ^*.

II.1 - Méthodes d'estimation ponctuelle

II.1.a - Méthode des moments

$X \sim P_{\theta^*}$, $P_{\theta^*} \in \mathcal{M}_\Theta = \{P_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$

On considère la fonctionnelle G sur l'espaces des mesures de probabilités sur $(\mathcal{X}, \mathcal{A})$, $G : P \mapsto (\mathbb{E}_P(X), \ldots, \mathbb{E}_P(X^p))$.

Alors, pour tout $\theta \in \Theta$, $G(P_\theta) := g(\theta)$.

Soit $(x_1, \ldots, x_N)$ un échantillon d'observations i.i.d. pour X, on calcule G en la mesure empirique

$\hat{g} = G\left(\hat{P}(x_1, \ldots, x_N)\right) \Rightarrow p$ premiers moments empiriques

et on cherche $\hat{\theta}$ tel que $g(\hat{\theta}) = \hat{g}$
(ou tel que $\hat{\theta} = \arg\min_\theta \|g(\theta) - \hat{g}\|^2$)

$\Longrightarrow$ $\hat{\theta}$ est un estimateur de θ^*.

II.1.b - Maximum de vraisemblance

Définition

Soit $\mathcal{M}_\Theta$, modèle paramétrique dominé, représenté par une famille de densités : $\mathcal{M}_\Theta = \{p_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$

Soit $X \sim p_{\theta^*}$, $p_{\theta^*} \in \mathcal{M}_\Theta$. Soit une observation $x \in \mathcal{X}$ pour la variable aléatoire X.

La fonction : $\theta \in \Theta, \theta \mapsto p_\theta(x)$ est appelée vraisemblance du paramètre θ en x, on la notera $\mathcal{L}(\theta; x)$.

Pour un échantillon d'observations $(x_1, x_2, \ldots, x_N)$ pour des variables aléatoires i.i.d, la vraisemblance s'écrit :

$$\mathcal{L}(\theta; x_1, x_2, \ldots, x_N) = \prod_{i=1}^N p_\theta(x_i) ,$$

ou

$$\mathcal{L}(\theta; x_1, x_2, \ldots, x_N) = \prod_{i=1}^N p(x_i|\theta) .$$

La vraisemblance du paramètre θ correspond ainsi à sa plausibilité pour un certain jeu de données observées.

Définition

Pour une observation ou un échantillon d'observations x, si $\hat{\theta}(x)$ maximise la vraisemblance $\mathcal{L}(\theta; x)$ sur Θ
$\hat{\theta}(x)$ sera appelée estimation du maximum de vraisemblance et l'estimateur associé $\hat{\theta}(X)$, estimateur du maximum de vraisemblance (EMV).

On note :

$$\hat{\theta}(x) = \arg\max_{\theta \in \Theta} \mathcal{L}(\theta; x).$$

Remarques

- Souvent on maximise plutôt $\ell(\theta; x) = \ln(\mathcal{L}(\theta; x))$

II - Estimation Paramétrique

Soit $X \sim P^*$, P^* inconnue. On suppose que P^* appartient à un modèle paramétrique $P^* \in \mathcal{M}_\Theta \{P_\theta, \theta \in \Theta\}$, et donc $\exists \theta^* \in \Theta$ tel que $P^* = P_{\theta^*}$.

$\Longrightarrow$ Le problème d'inférence sur P^* est alors ramené au problème d'estimation du paramètre θ^*.

Définition

Soit $(X_1, \ldots, X_N)$ un échantillon i.i.d. pour une loi P_{θ^*}. Un estimateur de θ^* est une statistique $T(X_1, \ldots, X_N)$ à valeurs dans Θ visant à estimer θ^*.

II.1 - Méthodes d'estimation ponctuelle

II.1.a - Méthode des moments

$X \sim P_{\theta^*}$, $P_{\theta^*} \in \mathcal{M}_\Theta = \{P_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$

On considère la fonctionnelle G sur l'espaces des mesures de probabilités sur $(\mathcal{X}, \mathcal{A})$, $G : P \mapsto (\mathbb{E}_P(X), \ldots, \mathbb{E}_P(X^p))$.

Alors, pour tout $\theta \in \Theta$, $G(P_\theta) := g(\theta)$.

Soit $(x_1, \ldots, x_N)$ un échantillon d'observations i.i.d. pour X, on calcule G en la mesure empirique

$\hat{g} = G\left(\hat{P}(x_1, \ldots, x_N)\right) \Rightarrow p$ premiers moments empiriques

et on cherche $\hat{\theta}$ tel que $g(\hat{\theta}) = \hat{g}$
(ou tel que $\hat{\theta} = \arg\min_\theta \|g(\theta) - \hat{g}\|^2$)

$\Longrightarrow$ $\hat{\theta}$ est un estimateur de θ^*.

II.1.b - Maximum de vraisemblance

Définition

Soit $\mathcal{M}_\Theta$, modèle paramétrique dominé, représenté par une famille de densités : $\mathcal{M}_\Theta = \{p_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$

Soit $X \sim p_{\theta^*}$, $p_{\theta^*} \in \mathcal{M}_\Theta$. Soit une observation $x \in \mathcal{X}$ pour la variable aléatoire X.

La fonction : $\theta \in \Theta, \theta \mapsto p_\theta(x)$ est appelée vraisemblance du paramètre θ en x, on la notera $\mathcal{L}(\theta; x)$.

Pour un échantillon d'observations $(x_1, x_2, \ldots, x_N)$ pour des variables aléatoires i.i.d, la vraisemblance s'écrit :

$$\mathcal{L}(\theta; x_1, x_2, \ldots, x_N) = \prod_{i=1}^N p_\theta(x_i) \ ,$$

ou

$$\mathcal{L}(\theta; x_1, x_2, \ldots, x_N) = \prod_{i=1}^N p(x_i|\theta) \ .$$

La vraisemblance du paramètre θ correspond ainsi à sa plausibilité pour un certain jeu de données observées.

Définition

Pour une observation ou un échantillon d'observations x, si $\hat{\theta}(x)$ maximise la vraisemblance $\mathcal{L}(\theta; x)$ sur Θ
$\hat{\theta}(x)$ sera appelée estimation du maximum de vraisemblance et l'estimateur associé $\hat{\theta}(X)$, estimateur du maximum de vraisemblance (EMV).

On note :

$$\hat{\theta}(x) = \arg\max_{\theta \in \Theta} \mathcal{L}(\theta; x).$$

Remarques

- Souvent on maximise plutôt $\ell(\theta; x) = \ln(\mathcal{L}(\theta; x))$
- Existence et unicité pas forcément assurées

II - Estimation Paramétrique

Soit $X \sim P^*$, P^* inconnue. On suppose que P^* appartient à un modèle paramétrique $P^* \in \mathcal{M}_\Theta \{P_\theta, \theta \in \Theta\}$, et donc $\exists \theta^* \in \Theta$ tel que $P^* = P_{\theta^*}$.

$\Longrightarrow$ Le problème d'inférence sur P^* est alors ramené au problème d'estimation du paramètre θ^*.

Définition

Soit $(X_1, \ldots, X_N)$ un échantillon i.i.d. pour une loi P_{θ^*}. Un estimateur de θ^* est une statistique $T(X_1, \ldots, X_N)$ à valeurs dans Θ visant à estimer θ^*.

II.1 - Méthodes d'estimation ponctuelle

II.1.a - Méthode des moments

$X \sim P_{\theta^*}$, $P_{\theta^*} \in \mathcal{M}_\Theta = \{P_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$

On considère la fonctionnelle G sur l'espaces des mesures de probabilités sur $(\mathcal{X}, \mathcal{A})$, $G : P \mapsto (\mathbb{E}_P(X), \ldots, \mathbb{E}_P(X^p))$.

Alors, pour tout $\theta \in \Theta$, $G(P_\theta) := g(\theta)$.

Soit $(x_1, \ldots, x_N)$ un échantillon d'observations i.i.d. pour X, on calcule G en la mesure empirique

$\hat{g} = G\left(\hat{P}(x_1, \ldots, x_N)\right) \Rightarrow p$ premiers moments empiriques

et on cherche $\hat{\theta}$ tel que $g(\hat{\theta}) = \hat{g}$
(ou tel que $\hat{\theta} = \arg\min_\theta \|g(\theta) - \hat{g}\|^2$)

$\Longrightarrow$ $\hat{\theta}$ est un estimateur de θ^*.

II.1.b - Maximum de vraisemblance

Définition

Soit $\mathcal{M}_\Theta$, modèle paramétrique dominé, représenté par une famille de densités : $\mathcal{M}_\Theta = \{p_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$

Soit $X \sim p_{\theta^*}$, $p_{\theta^*} \in \mathcal{M}_\Theta$. Soit une observation $x \in \mathcal{X}$ pour la variable aléatoire X.

La fonction : $\theta \in \Theta$, $\theta \mapsto p_\theta(x)$ est appelée vraisemblance du paramètre θ en x, on la notera $\mathcal{L}(\theta; x)$.

Pour un échantillon d'observations $(x_1, x_2, \ldots, x_N)$ pour des variables aléatoires i.i.d, la vraisemblance s'écrit :

$$\mathcal{L}(\theta; x_1, x_2, \ldots, x_N) = \prod_{i=1}^N p_\theta(x_i) \ ,$$

ou

$$\mathcal{L}(\theta; x_1, x_2, \ldots, x_N) = \prod_{i=1}^N p(x_i|\theta) \ .$$

La vraisemblance du paramètre θ correspond ainsi à sa plausibilité pour un certain jeu de données observées.

Définition

Pour une observation ou un échantillon d'observations x, si $\hat{\theta}(x)$ maximise la vraisemblance $\mathcal{L}(\theta; x)$ sur Θ
$\hat{\theta}(x)$ sera appelée estimation du maximum de vraisemblance et l'estimateur associé $\hat{\theta}(X)$, estimateur du maximum de vraisemblance (EMV).

On note :

$$\hat{\theta}(x) = \arg\max_{\theta \in \Theta} \mathcal{L}(\theta; x).$$

Remarques

- Souvent on maximise plutôt $\ell(\theta; x) = \ln(\mathcal{L}(\theta; x))$
- Existence et unicité pas forcément assurées
- Si $\hat{\theta}$ est un point intérieur à Θ, CN d'optimalité : $\nabla\mathcal{L}(\theta) = 0$ (CNS si $\mathcal{L}(\theta)$ concave et Θ convexe).