

Notations

Introduction

L'objectif de ce document est de formaliser un ensemble de notations cohérentes pour l'ensemble du cours.

Nous conseillons NBViewer pour visualiser ce Jupyter Notebook.

Vecteur aléatoire en entrée

Soit $p \in \mathbb{N}$ le nombre de composantes du vecteur d'entrée.

Soit $\mathcal{X} \subset \mathbb{R}^p$ le support de \mathbf{X} .

Soit $\mathbf{X} = (X_1, \dots, X_p)^T \in \mathcal{X}$ le vecteur aléatoire d'entrée.

On note $i \in \{1, \dots, p\}$ l'indice d'une composante du vecteur d'entrée \mathbf{X} . Pour tout $i = 1, \dots, p$, on a donc $X_i \in \mathbb{R}$.

Densité de probabilité du vecteur d'entrée

On note f la densité de probabilité du vecteur aléatoire \mathbf{X} . Dans le contexte de l'inférence (c'est à dire de l'étape B), lorsqu'on considère un modèle probabiliste paramétrique, on peut chercher à estimer les paramètres du modèle s'ajustant au mieux à un échantillon donné de \mathbf{X} .

Soit $m \in \mathbb{N}$ le nombre de paramètres de la densité de probabilité f . Soit $\Theta \subset \mathbb{R}^m$ le sous-ensemble des paramètres. Alors $f(\mathbf{X}, \boldsymbol{\theta})$ est la densité de probabilité dont le vecteur de paramètres est $\boldsymbol{\theta} \in \Theta$.

Variable aléatoire en sortie

On note $i \in \{1, \dots, p\}$ l'indice d'une composante du vecteur d'entrée \mathbf{X} . Pour tout $i = 1, \dots, p$, on a donc $X_i \in \mathbb{R}$.

On note f la densité de probabilité du vecteur aléatoire \mathbf{X} .

Soit $g : \mathcal{X} \rightarrow \mathbb{R}$ une fonction. On considère la variable aléatoire :

$$Y = g(\mathbf{X}).$$

Nous allons estimer l'espérance de Y :

$$\mathbb{E}(Y) = \int_{\mathcal{X}} g(\mathbf{X}) f(\mathbf{X}) d\mathbf{X}.$$

De plus, nous allons estimer la variance de Y :

$$\mathbb{V}(Y) = \mathbb{E}[(Y - \mathbb{E}(Y))^2].$$

Pour un seuil $s \in \mathbb{R}$ fixé, on peut souhaiter la probabilité de dépasser le seuil s :

$$\mathbb{P}(Y > s) = \int_{\mathcal{X}} g(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}$$

où $\mathbf{1}_{g(\mathbf{x}) > s}$ est la fonction indicatrice définie par :

$$\mathbf{1}_{g(\mathbf{x}) > s}(\mathbf{x}) = \begin{cases} 1, & \text{si } g(\mathbf{x}) > s, \\ 0, & \text{sinon.} \end{cases}$$

Echantillon Monte-Carlo simple

Soit $n \in \mathbb{N}$ un entier représentant la taille de l'échantillon.

Soit $\{\mathbf{X}^{(j)}\}_{j=1, \dots, n}$ un échantillon i.i.d. du vecteur aléatoire \mathbf{X} . Par conséquent, la réalisation $x_i^{(j)}$ est la i -ème composante de la j -ème réalisation, pour $i = 1, \dots, p$ et $j = 1, \dots, n$. Le plan d'expériences A associé est donc :

$$A = \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_p^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \cdots & x_p^{(2)} \\ \vdots & \vdots & & \vdots \\ x_1^{(n)} & x_2^{(n)} & \cdots & x_p^{(n)} \end{pmatrix}$$

où chaque ligne représente une réalisation du vecteur aléatoire \mathbf{X} et chaque colonne représente une composante.

Soit

$$y^{(j)} = g(\mathbf{x}^{(j)})$$

pour $j = 1, \dots, n$. L'estimateur Monte-Carlo de la moyenne empirique est :

$$\bar{y} = \frac{1}{n} \sum_{j=1}^n y^{(j)}.$$

Modèle paramétrique

Dans certains cas, on suppose que la fonction g est un modèle paramétrique dont on cherche à déterminer les paramètres. Il peut s'agir d'un contexte de calage de modèle (dans l'étape B') ou d'estimation des hyper-paramètres d'un méta-modèle (dans l'étape C).

Soit $q \in \mathbb{N}$ le nombre de paramètres de la fonction g . Soit $\mathcal{B} \subset \mathbb{R}^q$ le sous-ensemble des paramètres. Pour tout $\beta \in \mathcal{B}$, on considère la fonction $y = g(\mathbf{X}, \beta)$.

Répliques

Pour observer la variabilité d'un estimateur, on considère parfois des répliques d'une simulation de Monte-Carlo. Dans ce cas, le nombre de répliques est noté $r \in \mathbb{N}$ et chaque réplique porte l'indice $k = 1, \dots, r$.