

Stéphane MALLAT

CHAIRE SCIENCES DES DONNÉES

Apprentissage et génération par échantillonnage aléatoire

17 janvier > 13 mars 2024



Cours et séminaires

Les cours auront lieu le mercredi, de 9 h 30 à 11 h,
ils seront suivis des séminaires de 11 h 15 à 12 h 30.
Amphithéâtre Marguerite de Navarre.

17 janvier 2024

COURS **Apprentissage et génération par échantillonnage aléatoire (1)**

SÉMINAIRE **Challenges de données (1)** - Stéphane Mallat

24 janvier 2024

COURS **Apprentissage et génération par échantillonnage aléatoire (2)**

SÉMINAIRE **Challenges de données (2)** - Stéphane Mallat

31 janvier 2024

COURS **Apprentissage et génération par échantillonnage aléatoire (3)**

SÉMINAIRE **Présentation de gagnants de challenges 2023**

- Stéphane Mallat

7 février 2024

COURS **Apprentissage et génération par échantillonnage aléatoire (4)**

SÉMINAIRE **Modèles génératifs pour la simulation de systèmes
à l'équilibre** - Marylou Gabriel, professeur à l'École Polytechnique

14 février 2024

COURS **Apprentissage et génération par échantillonnage aléatoire (5)**

SÉMINAIRE **Modèles causaux : les modèles génératifs en support
à des interventions** - Michèle Sebag, directrice de recherche au CNRS

28 février 2024

COURS **Apprentissage et génération par échantillonnage aléatoire (6)**

SÉMINAIRE **Échantillonnage par distributions auxiliaires :
de la téléportation à l'échantillonnage d'importance par chaîne
de Markov** - Randal Douc, professeur de Telecom Sudparis

6 mars 2024

COURS **Apprentissage et génération par échantillonnage aléatoire (7)**

SÉMINAIRE **Calculs bayésiens approximatifs (ABC) : d'un substitut élémentaire
à une nouvelle forme d'inférence** - Christian Robert, professeur
à l'Université Paris-Dauphine

13 mars 2024

COURS **Apprentissage et génération par échantillonnage aléatoire (8)**

SÉMINAIRE **Une vision alternative des modèles de diffusion par débruitage**
- Francis Bach, directeur de recherche à l'INRIA

La modélisation de données de grandes dimensions est essentiellement probabiliste. L'apprentissage de modèles, l'inférence et la génération de nouvelles données nécessitent d'échantillonner ces distributions de probabilités. Des résultats impressionnants sont obtenus avec des réseaux de neurones, pour la génération d'images, de sons, de textes ou de champs physiques. On suivra un chemin qui va des bases mathématiques jusqu'aux frontières algorithmiques de la génération aléatoire.

Le cours introduit le cadre mathématique de l'apprentissage et de l'inférence statistique par méthode de Monte Carlo, ainsi que les algorithmes de générations aléatoires. On considère les modèles de champs de Markov, qui explicitent l'indépendance conditionnelle de variables, et sont caractérisés par des énergies de Gibbs. Le cours présente les algorithmes d'échantillonnage basés sur des chaînes de Markov, et notamment l'algorithme de Metropolis-Hasting et l'échantillonnage de Gibbs. On aborde ensuite l'échantillonnage par l'équation de Langevin, issue de l'équation de Fokker-Planck. Le cours se termine par une présentation des algorithmes de génération par score-diffusion, qui permettent d'échantillonner des distributions de probabilités complexes, en estimant le score avec des réseaux de neurones.

Image : Bouquet virtuel sphérique composé avec l'aide d'un système d'IA génératives (Dall-e via ChatGPT de Bing nov 2023), avec courbes hélicoïdales et croissance semi-aléatoire, Lamiot