

Seminaire Stats & Machine Learning

2024-2025

Séminaire Stats & Machine Learning

2025-2026

(romain.perier@universite-paris-saclay.fr, bertrand.even@universite-paris-saclay.fr)

Les *jeudis*, de 13h à 14h en salle 0A1, Institut de Maths d'Orsay [\[accès\]](#).

Lorsque la présentation aura lieu en distanciel, ce sera via le lien suivant :

<https://bbb.imo.universite-paris-saclay.fr/rooms/3ik-vmc-adh-n8p/join>

Les abstracts ainsi que les slides et/ou les offres de stage/thèse des orateurs se trouvent en bas du programme.

<u>Date</u>	<u>Invité</u>	<u>Titre</u>
02/10	Mathieu Even	Causal Machine Learning: Introduction and some Research Directions
09/10	Gilles Stoltz	Trouver un bon stage de M2 et surtout une thèse
16/10	PAS DE SÉMINAIRE	
23/10	Angelina Roche	ACP : des données multivariées aux processus ponctuels
30/10	PAS DE SÉMINAIRE	
06/11	Luca Ganassali	Causalité : de la découverte de structure à l'inférence des effets
13/11	Mohamed Ndaoud	Robust High-Dimensional Linear Regression: Convex and Non-Convex Estimators for Adversarial and Heavy-Tailed Settings
20/11	Marianne Clausel	Challenges in causal machine learning : time series analysis and fairness
27/11	Pierfrancesco Urbani	Introduction to dynamical mean field theory: results and methods
04/12	Pierre Jacob	
11/12	Tabea Rebafka	
18/12	PAS DE SÉMINAIRE	
08/01	Thanh Mai Pham Ngoc	
15/01	Perrine Lacroix	

Abstracts

02/10:

Causal Machine Learning: Introduction and some Research Directions

Mathieu Even, [INRIA](#)

Email: mathieu.even@inria.fr

Abstract:

Causal Machine Learning is at the intersection between Causal Inference, that seeks at quantifying the effect of an intervention (such as a treatment) on an outcome (such as for instance survival), and Machine Learning, that seeks at predicting as best as possible some information based on prior knowledge and data.

In this presentation, I will make an introduction to causal inference and causal ML, by starting from simple examples and presenting classical methodologies to quantify the effect of a treatment, from a very high point of view.

I will then present three research projects in this direction, that involve causal inference, privacy, LLM auditing and human-AI interactions.

09/10:

Trouver un bon stage de M2 et surtout une thèse

Gilles Stoltz, [CNRS](#) / [Laboratoire de mathématiques d'Orsay](#)

Email: gilles.stoltz@universite-paris-saclay.fr

Résumé:

Je voudrais vous présenter deux sujets de thèse (en bandits stochastiques, et en inférence causale), mais aussi et surtout, vous expliquer au passage ma vision sur les enjeux de l'année de M2 : i) trouver un sujet mais surtout un financement de thèse, avec des encadrants disponibles et ii) considérer les cours du M2 comme autant d'énoncés et d'opportunités de thèses, ce qui doit induire une attitude en cours adéquate, se dégageant du souci scolaire pour viser l'action de recherche. Je partagerai en particulier mes sentiments très positifs, mais aussi mes points d'attention, concernant les thèses CIFRE.

23/10:

ACP : des données multivariées aux processus ponctuels

Angelina Roche, MAP5

Email: angelina.roche@u-paris.fr

Résumé:

L'ACP est une méthode classique de réduction de la dimension en statistique multivariée. Les bases de l'ACP pour données multivariées seront rappelées dans un premier temps. Dans un second temps, je présenterai son extension à des données qui se présentent sous formes de courbes (les données dites fonctionnelles) puis à des données qui se présentent sous la forme de processus ponctuels.

06/11:

Causalité : de la découverte de structure à l'inférence des effets

Luca Ganassali, Laboratoire de mathématiques d'Orsay

Email: luca.ganassali@universite-paris-saclay.fr

Résumé :

Ce séminaire explorera les deux grands cadres mathématiques de la causalité, les graphes causaux et les potential outcomes, et les questions statistiques fondamentales qu'ils soulèvent. Comment découvrir la structure causale sous-jacente à partir de données ? Comment inférer des effets causaux ? Nous discuterons de sujets choisis : identifiabilité des graphes dans le cadre gaussien, puis estimation des effets causaux via le modèle des potential outcomes, et ses liens avec le transport optimal. Ces sujets recèlent des questions ouvertes que nous discuterons.

13/11:

Robust High-Dimensional Linear Regression: Convex and Non-Convex Estimators for Adversarial and Heavy-Tailed Settings

[Mohamed Ndaoud](#), [ESSEC Business School](#) / [CREST](#)

Email: b00791786@essec.edu

Abstract :

We study high-dimensional linear regression under adversarial contamination and heavy-tailed noise. We establish minimax lower bounds that capture the effect of both corruption and noise tails, and we propose two complementary estimators. The first is a convex, tuning-free variant of the Square-Root Slope that is adaptive, computationally efficient, and minimax optimal up to logarithmic factors. The second is a non-convex M-estimator based on alpha divergence that removes outlier-induced bias and achieves minimax-optimal deviation bounds, improved robustness to large outliers, and asymptotic normality. Together, these approaches provide practical and theoretically grounded methods for robust high-dimensional regression in adversarial and heavy-tailed settings.

20/11:

Challenges in causal machine learning : time series analysis and fairness

[Marianne Clausel](#), [Université de Lorraine](#)

Email: marianne.clausel@univ-lorraine.fr

Abstract :

This talk will focus on two challenges in causal ML. The first one deals with causal discovery for complex data as time series, or group of variables.

The second one addresses the question of data augmentation and related topics as relationships between fairness and causality.

27/11:

Introduction to dynamical mean field theory: results and methods

[Pierfrancesco Urbani](#), [CNRS](#) / [CEA](#)

Email: pierfrancesco.urbani@ipht.fr

Abstract:

I will discuss dynamical mean field theory that is a tool for studying dynamical systems in high dimension.

I will first review the applications of this technique to problems in machine learning and then I will describe the technique in a nutshell in two settings: high-dimensional inference (the spiked matrix-tensor problem) and dynamical systems appearing in simple recurrent neural networks.

04/12:

Pierre Jacob, [ESSEC Business School](#)

Email: pierre.jacob.work@gmail.com

Unbiased Markov Chain Monte Carlo: what, why, and how

Abstract: This presentation presents methods, called unbiased Markov chain Monte Carlo, to remove the initialization or burn-in bias from Markov chain Monte Carlo estimates, with consequences on parallel computing, convergence diagnostics and performance assessment. These methods rely on couplings of Markov chains generated by the MCMC algorithms, where the pairs of chains need to coalesce after a random time. Such constructions are not trivial but are often feasible by combining common random numbers or synchronous couplings, reflection couplings, and maximal or gamma-couplings. The current theoretical understanding of these couplings in the context of MCMC is very limited, thus giving rise to many research directions. The presentation does not assume prior knowledge about MCMC, and presents on-going research efforts on this topic, which is at the interface of Bayesian computation and applied probability. The presentation follows roughly the document available at <https://arxiv.org/abs/2406.06851>, co-written with Yves F. Atchadé (Boston University).

11/12:

Tabea Rebafka, [AgroParisTech](#)

Email: tabea.rebafka1@agroparistech.fr

08/01:

Thanh Mai Pham Ngoc, [Université Sorbonne Paris Nord](#)

Email: phamngoc@math.univ-paris13.fr

15/01:

Perrine Lacroix, [Laboratoire de mathématiques Jean Leray](#)

Email: perrine.lacroix@univ-nantes.fr