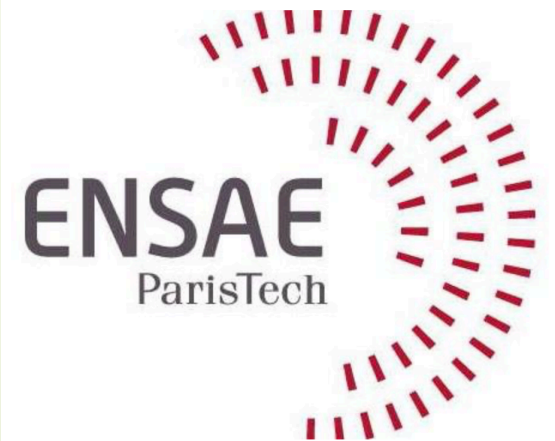
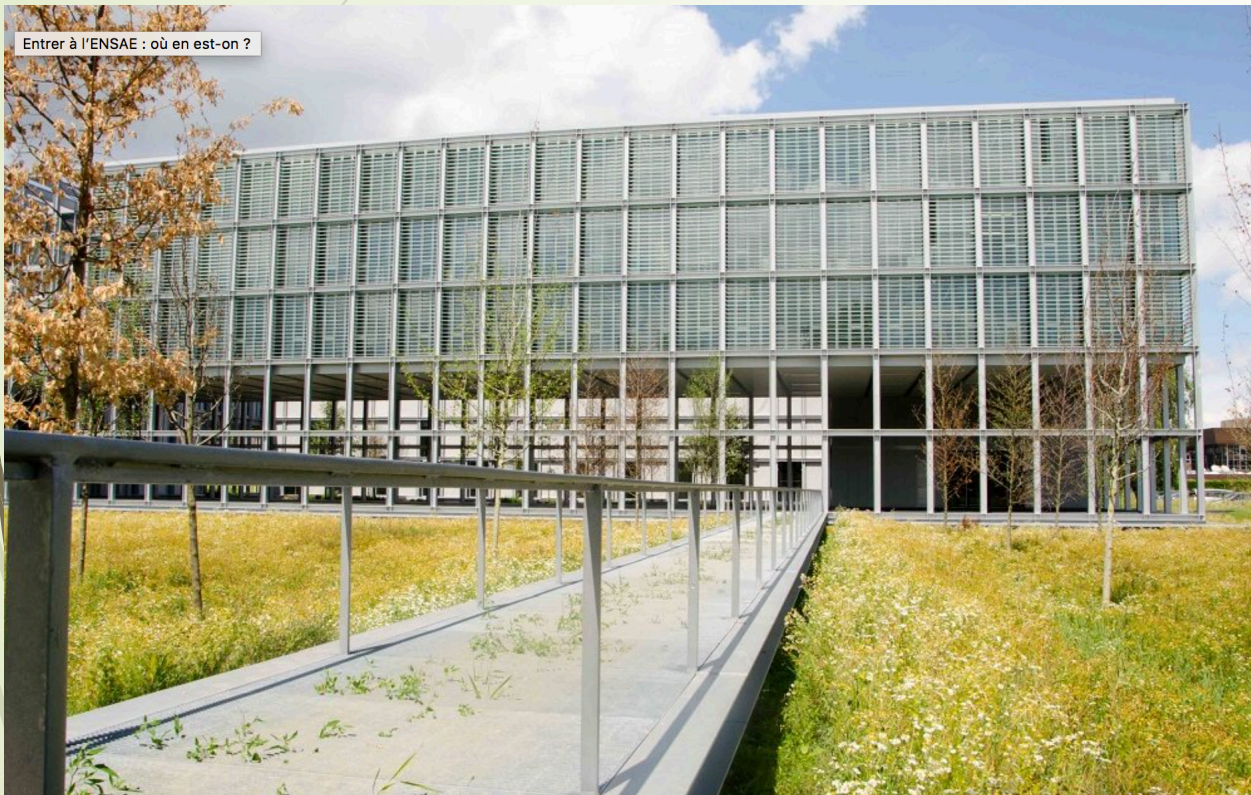


8 DataScience Courses at ENSAE

Guillaume Lecué



ENSAE – Ecole Nationale de la Statistique et de l'Administration Economique





Accéder à l'école, y suivre des enseignements et passer les examens

- inscription en ligne <https://inscription.ensae.fr/>
- Portail ENSAE / messagerie / pamplémousse (informations : salles de cours, dates d'examens, EdT, etc.)
- Configuration de votre badge étudiant au bureau de Christophe LAGARDE (3081 au 3e étage)



First semester starting in October:


- 1) **Statistical learning theory** by Arnak Dalalyan
- 2) **Estimation non paramétrique** by Cristina BUTUCEA
- 3) **Statistiques en grandes dimensions** by Alexander Tsybakov
- 4) **Modèles à chaînes de Markov cachées
et méthodes de Monte Carlo séquentielles** by Nicolas Chopin
- 5) **Enchères et matching: apprentissage et approximations** by Vianney Perchet

Second semester starting last week of January:

- 1) **Optimal Transport: Theory, Computations, Statistics, and ML Applications** by Marco Cuturi
- 2) **Online learning and aggregation** by Alexander Tsybakov
- 3) **A mathematical introduction to Compressed Sensing** by Guillaume Lécué



Courses organisation

- ▶ Each course is about 20 - 30 hours long (course + TD + TP)
 - ▶ Numerus clausus: 15 for Tsybakov's courses
 - ▶ Courses are mainly theoretical but closely related to applications
 - ▶ Prerequisites: background in probability theory, mathematical analysis, convex optimization
 - ▶ Exams or projects
- 

Statistical Learning Theory

Lecturer: [Arnak Dalalyan](#)

Examen écrit



Course contents (20h de cours / TD)

➤ Basic notions

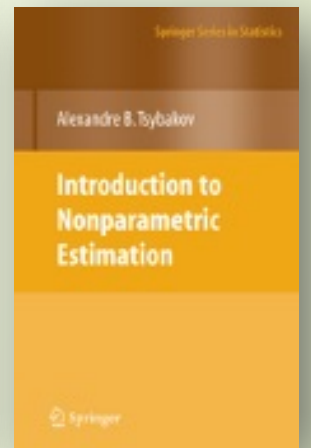
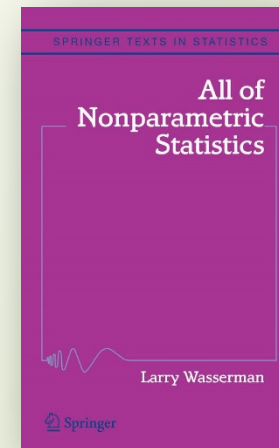
- Three main problems of statistical learning: regression, classification and density estimation.
- Bayes predictor and links between the three main problems.
- Empirical risk minimization

➤ Density Estimation

- piecewise linear estimation
- bias-variance tradeoff
- minimax risk over the Holder classes

➤ Adaptive estimation

- bandwidth selection by minimizing an unbiased risk estimator
- Lepski's method
- thresholding in nonparametric regression





Estimation non paramétrique

Lecturer: [Cristina BUTUCEA](#)

Examen écrit + petit projet

Course contents (24h total: 15h de cours + 9 heures de TD)

Goal: Design and study optimal nonparametric methods in

- Classical models: density, regression
- Applications: mixture of populations, confidentiality preserving data, large covariance matrices
- **Estimation**
 - Kernel methods
 - Projection estimators (wavelet and Fourier basis)
 - Local polynomial estimators, Splines
- **Hypothesis Testing**
 - Separation rates
 - Aggregation of test procedures
- **Uncertainty quantification:** build confidence sets for functions



Statistiques en grandes dimensions

Lecturer: [Alexandre Tsybakov](#)

Examen écrit

Course contents (14 heures de cours et 8 heures de TD)

- **Modèle de suites Gaussiennes**
- **Sparsité et procédures de seuillage**
- **Régression linéaire en grande dimension. Méthodes BIC, LASSO, Dantzig, square root lasso**
- **Inégalité d'oracle et sélection des variables**
- **Estimation de matrices de grande dimension de faible rang**
- **Inférence sur les réseaux et modèle stochastique à blocs**

Modèles à chaînes de Markov cachées et méthodes de Monte Carlo séquentielles

Lecturer: [Nicolas Chopin](#)

Examen : projet

Course contents

- Modèles à chaîne de Markov cachée : modèle supposant un processus markovien X_t observé imparfaitement et avec bruit.
- Nombreuses applications en épidémiologie (X_t =nombre d'infectés), écologie (X_t = nombre d'individus), robotique/navigation/pistage (X_t =position du robot ou du véhicule), finance (X_t =volatilité de l'actif sous-jacent), etc.
- Filtrage (Apprentissage séquentiel) de tels modèles requiert le développement de méthodes de Monte Carlo spécifiques, permettant un traitement séquentiel rapide des données.



Auction and Matchings: Learning and Approximations

Lecturer: Vianney Perchet

Validation: Examen

8 lectures Courses/TD (18 hours)

At the junction of Mathematics, Economics and Computer Science

- * Auctions Theory (strategies, equilibrium, revenue maximization)
- * Stable Matchings (algorithms, variants)
- * Learning (sample complexity, learning against agents, learning while earning)
- * Approximation (Prophet inequalities, Secretary problems, etc)
- * Online algorithms (Competitive ratio)
- * Real life examples and applications (Parcoursup, online advertisement, new/future possibilities)

Optimal Transport: Theory, Computations, Statistics, and ML Applications

Lecturer: [Marco Cuturi](#)

Examen : Mémoire avec implémentation (python)

organisation : 8 lectures + 4 practical sessions for a total of 18 hours

3 lectures on theory

- Monge and Kantorovich Problems, duality in OT, 2-Wasserstein geometry and the Brenier theorem.
- Closed forms: Applications to transport between Gaussians, Transport in 1D,
- Caffarelli contraction theorem, regularity theory (Figalli).

3 lectures on computations and statistics

- Algorithmic overview: network flow solvers in the discrete world, Benamou-Brenier formula in the PDE world.
- Statistical results and the curse of dimensionality
- Regularized approaches to compute optimal transport.

2 lectures on applications

- Handling measures with the Wasserstein geometry: computation of barycenters, clusters
- Automatic differentiation with the Sinkhorn algorithm. Wasserstein regression
- Wasserstein GANs
- Applications to Biology (cell pathways) and NLP (alignment of multilingual corpora)

4 practical sessions.

- = 1D transport, transport between Gaussians, network flow solver type algorithms
- = Sinkhorn algorithm, color transfer, retrieval, biology.
- = Sorting using the Sinkhorn algorithm, backpropagation of the Sinkhorn algorithm
- = Wasserstein GANs



Online learning and aggregation

Lecturer: [Alexandre Tsybakov](#)

Examen : Ecrit

Course contents (21 heures de cours / TD)

- Apprentissage séquentiel
- Algorithme de gradient en ligne
- Agrégation (à poids exponentiels) en ligne
- Liens avec l'apprentissage statistique
- Introduction aux problèmes de bandits



Compressed sensing

Lecturer: [Guillaume Lécué](#)

Examen : oral + notebook python

Course contents (21 heures de cours / TD / TP)

- Complexité algorithmique
- relaxation convexe
- matrices aléatoires
- Algorithmes
- grandes matrices de faible rang
- détection de communautés dans les graphes