

---

# Programme M2 IM



## Option Modélisation Stochastique et Statistique

Année 2021-2022

---



Cette option offre des compétences poussées en modélisation stochastique (calcul stochastique et méthodes numériques probabilistes) et en statistiques (machine learning, CART, sélection d'estimateurs ou de modèles ...) et consolide les connaissances informatiques acquises en M1 (C++, R, Scilab) par l'assimilation de nouveaux outils (Hadoop, Spark, SPSS ...). Il vise à former des ingénieurs dans la plupart des domaines nécessitant une analyse stochastique ou statistique (Big Data) : finance, marketing, environnement, biologie, industrie médicale et pharmaceutique, transports, énergie, assurance ...

---

## UE Mathématiques Appliquées

*Stochastic calculus and applications* : This course is devoted to the introduction of the basic concepts of continuous time stochastic processes which are used in many fields : physics, finance, biology, medicine, filtering theory, decision theory. It will consist of a presentation of Brownian motion, Itô integral, stochastic differential equations and Girsanov theorem. Several applications will be given.

## UE Mathématiques pour la Modélisation

*Probabilistic numerical methods* : Probabilistic numerical methods are widely used in machine learning algorithms as well as in mathematical finance for pricing financial derivatives and computing strategies. The course will present the basic methods used for simulating random variables and implementing the Monte-Carlo methods. Simulation in Scilab of stochastic processes used in mathematical finance, such as Brownian motion and solutions to stochastic differential equations, will be discussed as well.

*Advanced statistics* : This course focuses on three pillars of modern statistical inference: parameter estimation, hypothesis testing, and model selection. Its aim is to provide a good understanding of the current methods via a thorough treatment of the existing theoretical guarantees. A particular emphasis will be placed on the asymptotic setting.

*Modeling Studies* : In this course the student will apply the theoretical knowledge obtained during the courses of stochastic calculus, probabilistic computational methods and advanced statistics in some applied contexts related to the quantitative risk management. We will focus on some theoretical elements concernign Gaussian mixture models, spherical and elliptical models, interdependencies of risks, copula models (Archimedean and extreme values copulas), dependence measures between risks, the aggregation problem, Value at Risk, Multivariate extremal risks. The student will develop a project, during this course, starting from recent research publications on quantitative risk management theory. An important code (R or Python) contribution in the final project will be required. The project will be present at the end of the course, and it represents the evaluation criterion.

### References

- McNEIL A., FREY R., EMBRECHTS P. (2005) : Quantitative Risk Management, Priceton Series in Finance

- 
- EMBRECHTS, P. et al. (2010) : Modelling Extremal Events: for Insurance and Finance, Springer,
  - R. B. NELSEN : An introduction to Copulas, volume 139 of Lecture Notes in Statistics. Springer-Verlag, New York, 1999.

## UE Numérique et Informatique

**Statistical learning methods** : This course will focus on machine learning methods and Bayesian statistics approach. It will cover the following topics: Linear regression, linear models for classification, Kernelized SVM's, Neural networks, MCMC inference, PCA, clustering (K-means, hierarchical, DBSCAN). It will thus cover the main chapters of Murphy's book *Machine Learning: a probabilistic approach*. Several applications will be realized in R or Python.

**Statistical computational methods** : To apply theoretical models to the real world it is useful to know some softwares able to implement statistical methods. R and Python are open source softwares easy to use. The first part of this course will be devoted to the decision trees theory, in particular the CART and Random Forests algorithms developed by Breiman and his coauthors. These statistical tools define nonlinear models in the regression and classification context and are widely used in companies because of the easy interpretability of the outputs but also because of their potential coupling with variables selection. The R package VSURF will be used. The second part of the course will review various statistical methods (Linear Regression, PCA, Clustering...), apply them on real data sets and interpret the results. At the end the student will have a concrete view of the work of an applied statistician and will be autonomous in the implementation of his theoretical knowledge with SAS.

**Technologies of Big Data** : The aim of this course is to learn how to install and how to use an architecture for Big Data. This course will present the different softwares used for Big Data treatment : Apache Pig, Hive, HBase, Spark ...

## UE Métiers 1&2

**Analyse de la conjoncture économique**: La publication d'un taux de chômage moins bon que prévu, ou la révision à la baisse de prévisions de croissance peuvent suffire à faire plonger la bourse. Chaque jour plusieurs indicateurs conjoncturels sont publiés. La question de la fiabilité de ces indicateurs, comme celle de l'information qu'ils apportent sont vitales pour poser un diagnostic pertinent de la situation économique présente et à

---

venir. Le problème est qu' à la différence des variables financières qui sont connues en temps réel, la plupart des agrégats ne sont connus qu'avec plusieurs mois de retards. C'est pour palier cette lacune que la conjoncture propose des instruments spécifiques permettant de prévoir le passé proche et le présent. Alors, si comme le disait Pierre Dac, « la prévision est un art difficile, surtout lorsqu'elle concerne l'avenir ? », qu'en est-il des prévisions du passé et du présent proposées par les conjoncturistes ? Sont-ils les oracles des temps modernes ? A quoi servent les enquêtes de conjoncture ? Est-il possible de prévoir les retournements de l'activité ? Peut-on faire confiance aux modèles macro économétriques ? Les variables financières sont-elles vraiment impossibles à prévoir ? Pourquoi a-t-on sous-estimé l'ampleur de la crise de 2008 ? Ce cours répond notamment à ces questions en familiarisant les étudiants avec les pratiques de l'analyse conjoncturelle. Les aspects pratiques de l'analyse conjoncturelle et l'utilisation d'outils statistiques élaborés seront privilégiés. Le contenu pourra évoluer en fonction des événements économiques.

**Big Data & Analytics:** This course offers an introduction to data science as well as various software tools. It provides a comprehensive presentation of neural networks: deep, convolutional, recurrent, adversarial and generative. It also provides an introduction to the tools routinely used by data analysis practitioners. An important part of the course is devoted to practical case studies on computers, using Jupiter notebooks. More specifically, we will study the categorization of images, semantic segmentation of images and speech recognition. Part of the evaluation will take the form of a Kaggle challenge.

Syllabus:

Introduction: methodological framework, introduction to Kaggle and Python Notebooks.

Convolutional Neural Networks (CNNs).

Recurrent Neural Networks.

Adversarial and generative networks.

IBM Data Science tools.

**Building Recommender Systems:** A recommender system (RS) can help to influence your customers' behaviour directly but entertainingly. In this course, we will build RS's using different approaches: content-based, collaborative filtering, context-aware, or a hybrid one. We will learn about the theory behind diverse mathematical models of an RS task: matrix and tensor decompositions, associative rules, neighbourhood methods, learning to rank, and metric learning. For the practical part, we will employ classical machine learning (such as `scikit-learn`), deep learning (e.g. `pytorch`), and big data frameworks (e.g. `pyspark`). During the lectures, we will talk not only about theorems but also about applications of RS's making the clients of companies and non-profit organisations happier. No prior knowledge of the subject is necessary. Python programming experience is mandatory. Statistical learning fundamentals will be nice to have.

**Artificial Intelligence Seminars:** Le traitement automatique du langage naturel est un des axes forts du développement de l'intelligence artificielle. La théorie sera agrémentée de

---

cas pratiques et la finalité du cours est de maîtriser différentes techniques (lemmatisation, tokenisation, vectorisation...) et des outils (spacy, nltk...) pour le traitement et la compréhension de la donnée textuelle. Les utilisations sont nombreuses comme par exemple l'élaboration d'un chatbots ou d'un assistant vocal. Ce cours sera composé de la mise en place des fondamentaux pour ensuite pouvoir pratiquer lors d'un projet. Le prérequis sera de connaître la programmation en python.

## UE Management

Le module doit permettre de développer des compétences en matière de raisonnement juridique et d'analyser des situations pratiques en matière de management. Le programme intègre notamment les aspects suivants :

### 1. Développement durable/Responsabilité Sociétale des Entreprises

Comprendre comment les objectifs socio-économiques sont conciliables avec les préoccupations de qualité de vie, de protection environnementale et d'équité entre générations

### 2. Ethique

Eclairer le rôle de l'ingénieur face à l'avenir technologique. Développer une attitude critique, prospective vis-à-vis des liens entre connaissance et monde social. Cerner les mécanismes et les enjeux du concept de responsabilité dans le monde technologique actuel. Reconnaître les dimensions éthiques et sociales de l'ingénierie

### 3. Innovation et Entreprise- Hackaton

Favoriser la création d'entreprise. Susciter des vocations d'entrepreneur.

### 4. Culture juridique

Initier les étudiants à la matière juridique. Connaître les modalités juridiques pour la protection des logiciels et celle des bases de données. Faciliter leur insertion professionnelle (contrat de travail)

Pour les alternants, ce module (3 ECTS) sera réalisé sur une semaine au second semestre pendant la période de stage. Pour les non-alternants, il aura lieu au premier semestre.

---

## **UE Projet de Fin d'Etude**

Pour les alternants, ce projet de fin d'étude consistera en une présentation à mi-parcours de leurs stage en entreprise (9 ECTS). Pour les non-alternants il consistera en la restitution d'un projet appliqué sous la direction d'un tuteur (6 ECTS).

## **UE Projet Innovation-Recherche multidisciplinaire & Networking**

Cette UE ne concerne que les alternants (5 ECTS). Elle a lieu pendant deux semaines complètes à l'Université au second semestre. Elle est constituée de projets multidisciplinaires, d'une immersion dans le milieu de la recherche afin d'en découvrir l'intérêt pour un futur ingénieur et vise à fournir, par un travail de retour d'expérience, le développement d'un réseau professionnel.

## **UE Stage**

Cette UE correspond au stage en entreprise qui se déroulera de début mars à fin août pour les non-alternants (30 ECTS) et en un stage d'un an dans l'entreprise pour les alternants (22 ECTS).

| FORMATION INITIALE  |      |
|---|------|
| UE  | ECTS |
| MATHÉMATIQUES APPLIQUEES  | 3    |
| MATHEMATIQUES POUR LA MODELISATION<br>NUMERIQUE ET INFORMATIQUE | 6    |
| MATIÈRES 1&2  | 6    |
| PROJET DE FIN D'ETUDE<br>MANAGEMENT                             | 6    |
| STAGE   | 30   |

| FORMATION EN ALTERNANCE   |      |
|---|------|
| UE  | ECTS |
| MATHÉMATIQUES APPLIQUEES  | 3    |
| MATHEMATIQUES POUR LA MODELISATION<br>NUMERIQUE ET INFORMATIQUE | 6    |
| MATIÈRES 1&2  | 6    |
| PROJET DE FIN D'ETUDE   | 9    |
| PROJET INNOVATION-RECHERCHE<br>MULTIDISCIPLINAIRE & NETWORKING  | 5    |
| MANAGEMENT  | 3    |
| STAGE   | 22   |