

Programme du M1 Math&AS (2025-2026)

Les UEs dotées d'une astérisque (*) sont mutualisées avec les étudiant.e.s du master ISADS.

Les UEs dotées de deux astérisques (**) sont mutualisées avec les étudiant.e.s du master AMS.

L'UE dotée de trois astérisques (***) est mutualisée avec les étudiant.e.s du master 1 Informatique.

L'UE dotée de quatre astérisques (****) est partiellement mutualisée entre le M1 et le M2 Math&AS.

Optimisation numérique (**)

De très nombreux problèmes en industrie, en physique et en économie consistent en la minimisation (ou la maximisation) d'une fonction objective. Ce cours vise à présenter un grand nombre de méthodes numériques qui ont été développées pour résoudre de tels problèmes. Ces méthodes peuvent être locales ou globales, déterministes ou stochastiques. De nombreux exemples seront implémentés sur machine afin d'illustrer l'emploi en pratique de ces méthodes.

Probabilités (**)

Le module est consacré principalement à l'étude des chaînes de Markov à espace d'états discret, avec des applications aux marches aléatoires et à des processus à valeurs dans un espace d'états discret. Nous approfondirons également les notions d'espérance conditionnelle et de loi conditionnelle.

Algorithmique randomisée et recherche opérationnelle (***)

Ce cours porte sur les méthodes de résolution des problèmes d'optimisation difficiles pour lesquels il n'existe pas de méthode de résolution polynomiale connue. Il comprend :

- * Programmation linéaire, méthode du simplexe, dualité.
- * Heuristiques et méta-heuristiques (Recherche Tabou, Recuit Simulé, Algorithme Génétique).
- * Méthodes de résolution exacte : Branch and Bound et Branch and Cut.

Par ailleurs, le cours aborde également l'algorithmique randomisée à travers l'étude des méthodes de Monte-Carlo et de Las-Vegas.

Traitement des données et machine learning (*)

Cette UE a pour objectif d'initier les étudiants aux méthodes fondamentales du traitement des données et du machine learning, en les aidant à identifier les algorithmes les plus adaptés aux différentes tâches d'apprentissage, sur la base de leurs propriétés théoriques et de résultats expérimentaux.

Le cours comprend :

- * L'analyse des données et la synthèse de l'information contenue dans un jeu de données.
- * L'étude des méthodes d'analyse factorielle (ACP, AFD et ACM).

Ce programme explore également les principales méthodes d'apprentissage supervisé en intelligence artificielle, en mettant l'accent sur les algorithmes classiques de classification et une introduction aux réseaux de neurones. Plus précisément, nous aborderons les sujets suivants.

Introduction à l'Apprentissage Supervisé : définition et principe de l'apprentissage supervisé; différences entre classification et régression; évaluation des modèles : métriques de performance.

Régression logistique : modélisation de la probabilité d'appartenance à une classe; fonction sigmoïde et interprétation des coefficients; application à la classification binaire et multiclassées.

Machines à Vecteurs de Support (SVM) : principe des hyperplans séparateurs; utilisation des noyaux pour la classification non linéaire; réglage des paramètres (C, gamma) et impact sur la performance.

Introduction aux Réseaux de Neurones : architecture d'un réseau de neurones artificiel; fonction d'activation et propagation du gradient.

Enfin, une partie de cette UE est spécifiquement dédiée aux étudiants du M1 Math&AS et porte sur l'étude théorique de certaines séries chronologiques.

Inférence statistique et théorie de l'apprentissage ()**

Ce cours fournit les outils nécessaires à la compréhension des travaux avancés en statistique et en apprentissage automatique. Il couvre :

- * Les tests statistiques, intervalles de confiance et propriétés des estimateurs classiques.
- * La statistique asymptotique et la théorie de la décision.
- * La théorie de l'apprentissage statistique : formulation des problèmes, minimisation du risque empirique, inégalités de concentration et théorie des processus empiriques.

Méthodes de régression (*)

Modèle linéaire général et cas particulier de la régression simple ou multiple, traitement par la méthode des moindres carrés. Propriétés des estimateurs avec ou sans normalité. Décomposition de la variance. Ratios de Student et tests basés sur les matrices de contraste. Intervalles de prédiction. Test de Fisher général pour modèles emboîtés. Approche par vraisemblance et lien entre le LRT et le test de Fisher. Diagnostics (résidus studentisés, QQ-plots, autres indicateurs usuels). Introduction à la pénalisation de la dimension. Une attention particulière sera

portée à la gestion des régresseurs catégoriels (interprétation des coefficients et tests associés).

Logiciels statistiques et Python (*)

Cette UE inclut trois langages de programmation :

R : Programmation en R, manipulation d'objets, boucles, fonctions et applications statistiques.

SAS : Concepts généraux de programmation en SAS, manipulation de données, macros et jointures.

Python : L'UE met particulièrement l'accent sur Python, en développant ses fondamentaux et son application à l'analyse et au traitement des données. Plus précisément, le programme de Python est structuré comme suit.

Programmation Orientée Objet (POO) en Python : classes et objets; héritage et polymorphisme; encapsulation et abstraction; applications pratiques et bonnes pratiques de conception.

Introduction aux bibliothèques de base : NumPy; Math et SciPy; Scikit-Learn.

Manipulation de données avec Pandas.

Anglais

Ce cours aide les étudiants à se préparer aux exigences du monde professionnel. Contenu : Job Interview, Debating, CV et lettre de motivation, rédaction d'essais, compréhension orale et préparation au TOEIC.

Séminaire invité entreprise et PPEI (**)**

Cette UE comprend

Préparation professionnelle : CV, lettres de motivation et entraînements aux entretiens.

Rencontres avec des experts en IA et Data Science.

Ces interventions visent à présenter les différents aspects du métier de data scientist et à mettre les étudiants en contact avec des entreprises et partenaires susceptibles de proposer des opportunités de stage.

Programme du M2 Math&AS (2025-2026)

Le Master 2 Math&AS s'adresse aux étudiants titulaires d'un Master 1 Math&AS ou d'un Master 1 de Mathématiques avec des bases solides en statistique théorique, probabilités, Python et R.

Méthodes de régression avancées (*)

Cette UE couvre les techniques avancées de régression, des modèles linéaires classiques aux approches pénalisées et non linéaires, avec un accent particulier sur les méthodes statistiques et l'évaluation des modèles. Plus précisément, le cours développe les thèmes suivants.

Régression linéaire multiple et comparaison de modèles : rappel sur la régression linéaire multiple; tests entre modèles emboîtés (test de Fisher, test de vraisemblance)

Régression linéaire et hétéroscédasticité : impact de l'hétéroscédasticité sur l'estimation des coefficients; méthodes de correction et estimation robuste; introduction aux modèles linéaires généralisés (GLM).

Régularisation et régressions pénalisées : principe de la pénalisation en régression; régression LASSO : sélection de variables et parcimonie; régression ridge : contrôle du sur-ajustement; régression ElasticNet : compromis entre LASSO et ridge.

Régression PLS et méthodes en grande dimension.

Régressions non linéaires et bases de splines.

Apprentissage non supervisé (*)

Ce programme introduit les bases de l'intelligence artificielle, les concepts fondamentaux de l'apprentissage non supervisé et les techniques de clustering. Il couvre également la réduction de dimension et les modèles de mélanges de lois. Plus précisément, le cours développe les thèmes suivants.

Introduction à l'intelligence artificielle et à l'apprentissage : définition et principes de l'intelligence artificielle; concepts d'apprentissage supervisé et non supervisé; applications et enjeux actuels de l'IA.

Clustering et algorithmes non supervisés : présentation des méthodes de clustering; algorithme K-means et ses extensions (K-medoids, K-means++...); classification hiérarchique : CAH et CDH.

Modèles de mélanges de lois : estimation par l'algorithme EM; applications en segmentation et détection de groupes latents.

Statistique non paramétrique (*)

Le module est une introduction à la statistique non paramétrique. Il est partiellement mutualisé et comprend une première partie commune aux deux masters M2 ISADS et Math&AS, ainsi qu'une partie d'approfondissement théorique destinée aux étudiants de M2 Math&AS.

Dans la première partie, les bases de la statistique non paramétrique seront introduites. Les sujets traités incluent :

- * la construction des estimateurs à noyau et des estimateurs par projection de la fonction de densité,
- * les modèles de régression,
- * les estimateurs dans les problèmes inverses,
- * les régions de confiance non paramétriques,
- * les tests non paramétriques.

Cette partie théorique sera enrichie par des séances de travaux dirigés et des mises en œuvre des méthodes sur des données réelles, permettant ainsi d'acquérir une expérience pratique.

La deuxième partie, destinée aux étudiants de M2 Math&AS, se concentre sur un approfondissement théorique des méthodes. Elle traite notamment des estimateurs adaptatifs et des bornes inférieures pour démontrer l'optimalité au sens minimax des estimateurs vus en cours dans des modèles à densité.

Statistique bayésienne (*)

Cette UE est divisée en deux parties. La première, commune aux deux masters M2 ISADS et M2 Math&AS, est une introduction à la statistique bayésienne paramétrique. Nous y étudierons le modèle bayésien paramétrique à travers plusieurs concepts fondamentaux : calcul d'une loi a posteriori, prédicteur et intervalles de prédiction, estimateur de Bayes, ainsi que la distinction entre intervalles de confiance et intervalles de crédibilité.

La seconde partie est spécifiquement destinée aux étudiants du M2 Math&AS et vise à les faire travailler sur divers aspects de la statistique bayésienne non paramétrique à travers plusieurs projets. Ces projets, dont les thèmes peuvent varier d'une année à l'autre, permettent d'explorer des sujets tels que l'Uncertainty Quantification, les vitesses de contraction de la loi a posteriori en statistique bayésienne non paramétrique, les lois a priori gaussiennes et leurs applications, le processus de Dirichlet et ses mélanges, ainsi que l'adaptation bayésienne.

Techniques de prévision (*)

Après une introduction sur les séries chronologiques, plusieurs méthodes pour effectuer des prévisions seront étudiées ainsi que les notions nécessaires :

- * Méthode de la moyenne mobile pour estimer les différentes composantes d'une série chronologique.

- * Méthodes de lissage exponentiel.
- * Processus stationnaires : cas général, test pour un bruit blanc.
- * Processus ARMA : généralités, identification.
- * Processus ARIMA et SARIMA.

Les différentes méthodes seront mises en pratique sur des exemples à l'aide du logiciel R.

Statistique computationnelle

Cette UE est une introduction aux méthodes de Monte-Carlo.

Le premier chapitre porte sur la génération de variables aléatoires avec en particulier les méthodes d'inversion et de rejet.

Le deuxième chapitre est consacré au calcul approché d'une intégrale par la méthode de Monte-Carlo, dont des méthodes de réduction de la variance.

Le troisième chapitre porte sur les méthodes de Monte-Carlo par Chaîne de Markov classiques : algorithme de Metropolis-Hastings et échantillonneur de Gibbs. Application au modèle de mélange.

Les différents algorithmes sont mis en œuvre sur des exemples à l'aide du logiciel R.

Deep Learning

Cette UE vise à offrir une vue d'ensemble complète sur le deep learning, en partant des éléments fondamentaux (y compris les bases mathématiques des différentes composantes) et en allant jusqu'à l'implémentation des algorithmes à l'origine des révolutions technologiques de ces dernières années (intelligence artificielle générative, grands modèles de langage, traitement des images).

Plus précisément, les sujets abordés s'appuieront sur la bibliothèque PyTorch et incluront :

- * Les bases des réseaux de neurones : couches linéaires, fonctions d'activation, régularisation, optimisation par descente de gradient.
- * Le traitement des grandes quantités de données et l'optimisation avec un accélérateur graphique (GPU).
- * Le traitement des images : réseaux convolutifs, normalisation, connexions résiduelles.
- * Le traitement du langage : tokenisation, couches récurrentes, mécanisme d'attention, modèle Transformer.

L'UE se termine par un projet permettant d'appliquer les notions acquises dans un cadre réaliste et stimulant. Par exemple, il pourra s'agir de l'implémentation d'un modèle de langage similaire à GPT ou BERT, d'un modèle de génération d'images inspiré de DALL-E ou Stable Diffusion, d'un système de question-réponse sur des documents fournis, d'un outil de reconnaissance d'images médicales, ou encore d'un agent d'apprentissage par renforcement capable de jouer à des jeux vidéo.

Projet Data Challenge

Dans cette UE, les étudiants travailleront en binômes ou trinômes sur des défis proposés par un intervenant extérieur, un data scientist expérimenté issu du monde de l'entreprise. Ces challenges, variés dans leur nature, mobiliseront différentes compétences en machine learning, intelligence artificielle et analyse de données, en fonction du problème étudié.

L'objectif est double : permettre aux étudiants de mettre en pratique les concepts vus dans leurs autres cours et renforcer leurs compétences en programmation et en analyse de données à travers des problématiques concrètes, proches des enjeux rencontrés en entreprise.

Ouverture à la recherche

Cette UE est consacrée au développement de la théorie des processus stochastiques. Dans un premier temps, l'accent est mis sur le mouvement brownien, les processus gaussiens et le modèle de Galton-Watson. Ensuite, l'étude se poursuit avec des processus dont les trajectoires ne sont plus nécessairement continues, comme les processus de Lévy. Les propriétés théoriques de ces processus sont analysées, et leurs nombreuses applications sont illustrées dans divers domaines, allant de la finance et l'actuariat aux neurosciences, ainsi qu'à l'approximation des algorithmes de descente de gradient stochastique.

Machine learning et quantification d'incertitude

Cette UE est séparée en deux parties indépendantes. La première est consacrée aux arbres de classification et forêts aléatoires. La deuxième est consacrée aux méthodes d'explicabilité en machine learning en général, comme par exemple la méthode de la valeur de Shapley. Les aspects théoriques et algorithmiques des méthodes seront abordés, ainsi que leur implémentation pratique. Des applications à des jeux de données réels seront proposés.