



Programme  
des enseignements de 3<sup>e</sup> année  
Filière Data Science & Génie Statistique  
ANNÉE SCOLAIRE 2022 / 2023



École nationale  
de la statistique  
et de l'analyse  
de l'information

# FILIÈRE DATA SCIENCE & GÉNIE STATISTIQUE

ANNÉE SCOLAIRE 2022/2023

*DATA SCIENCE & ADVANCED STATISTICAL ENGINEERING SPECIALIZATION*

*2022/2023 ACADEMIC YEAR*

# Table des matières

Présentation de la filière .....	4
Descriptifs des enseignements communs.....	9
UE : APPRENTISSAGE AUTOMATIQUE (MACHINE LEARNING) .....	10
APPRENTISSAGE AUTOMATIQUE .....	11
APPRENTISSAGE PROFOND .....	12
TRAITEMENT AUTOMATIQUE DU LANGAGE ET FOUILLE DU WEB .....	13
UE : COMPLÉMENTS D'APPRENTISSAGE AUTOMATIQUE (ADVANCED MACHINE LEARNING).....	14
Régularisation.....	16
APPRENTISSAGE À GRANDE ÉCHELLE .....	17
Apprentissage PROFOND AVANCÉ .....	18
APPRENTISSAGE PAR RENFORCEMENT.....	19
UE : PROJETS.....	20
PROJET METHODOLOGIQUE .....	21
PROJET DE FIN D'ETUDES.....	22
DATA CHALLENGE.....	23
ANGLAIS.....	24
Descriptifs des enseignements de la filière .....	25
UE SPECIFIQUES FILIERE GS.....	26
MODELISATION NON LINEAIRE .....	27
THEORIE DES VALEURS EXTREMES .....	28
STATISTIQUE BAYESIENNE.....	30
MODELES MARKOVIENS .....	32
ANALYSE DE DONNEES FONCTIONNELLES .....	33
FILTRAGE.....	35
STATISTIQUE DES PROCESSUS.....	36
SERIES TEMPORELLES AVANCEES .....	38
PLAN D'EXPERIENCES.....	39
FIABILITE .....	40
UE : PROJET PROFESSIONNEL ET STAGES .....	41
DROIT DU TRAVAIL .....	42

## Présentation de la filière

---

La formation d'ingénieur de l'Ensaï inclut 6 filières de spécialisation. Toutes ces filières forment aux métiers de la Data Science, avec une maîtrise des outils permettant l'extraction, l'analyse et la fouille de données et une capacité à choisir les modalités de traitements des données massives (Big Data) et des techniques d'apprentissage automatique (machine learning). Selon les spécialisations, ces compétences sont spécifiques à un domaine ou transversales. L'ensemble des filières continue à former aux compétences transversales (soft skills) et à la valorisation des travaux menés dans un contexte professionnel et international. Lors des cours et du projet méthodologique en anglais, les élèves travaillent toutes les compétences linguistiques et communicationnelles et approfondissent leurs connaissances liées au monde de l'entreprise et de la recherche. La séquence de Tronc Commun mêlant enseignements scientifiques, projets et anglais conclut la formation à l'autonomie et la capacité à mettre en œuvre des analyses de données en situation complexe. Un stage de fin d'études est à réaliser à l'issue de la scolarité, qui permet de mettre en œuvre dans un cadre professionnel une démarche scientifique autour d'une problématique en lien avec les enseignements de la filière.

Même si l'un des débouchés revendiqués de la filière est l'industrie, la caractéristique de cette filière est sa transversalité. A l'issue de cette spécialisation, les étudiants sont capables de s'adapter à des problématiques provenant de différents secteurs d'activités comme l'industrie, le secteur bancaire, l'environnement, les services. C'est aussi la voie à privilégier pour ceux qui se destinent à la recherche et au développement dans le domaine des statistiques.

### Philosophie pédagogique

L'objectif est de former des ingénieurs experts en statistique et modélisation stochastique et capables de s'adapter à tous les domaines où les statistiques sont nécessaires. Dans cette optique, les cours dispensés ont pour objectifs de renforcer et d'élargir les compétences en modélisation statistique des étudiants et de leur fournir des compétences pluridisciplinaires et transversales. A l'issue de leur formation, les étudiants de cette filière auront les compétences spécifiques suivantes :

- Savoir introduire de nouvelles méthodes statistiques et expertiser les pratiques existantes ;
- Capacité à mettre en œuvre des méthodes d'analyse d'image ainsi que les techniques de filtrage ;
- Capacité à mobiliser et d'utiliser à bon escient des outils d'aide à la décision dans le domaine de la gestion des risques (industriels (fiabilité-qualité), environnementaux et financiers) ;
- Capacité à construire des modèles complexes ou novateurs de prévisions ;
- Capacité à s'adapter à des problématiques provenant de différents secteurs d'activités comme l'industrie, le secteur bancaire, l'environnement, les services, en maîtrisant les techniques statistiques requises (statistique bayésienne, statistique des processus, chaînes de Markov, *machine learning*...) et les modalités de traitement des données massives ;
- Capacité à innover et à mener des activités de recherche en statistique fondamentale et appliquée.

### Thématiques abordées

Pour répondre aux exigences de la filière, en plus du tronc commun à tous les élèves de troisième année, l'enseignement est divisé en quatre unités d'enseignement (UE). L'accent est placé sur la maîtrise des outils probabilistes et statistiques pour la modélisation de processus aléatoire dans le domaine de l'industrie. Les thématiques principales abordées incluent la statistique des processus, les systèmes dynamiques aléatoires, le filtrage particulier, la qualité et la fiabilité, le traitement de l'image et l'analyse de survie. La filière intègre également une UE sur l'apprentissage automatique permettant d'acquérir de l'expérience dans l'utilisation et le développement des techniques d'apprentissage automatique (Machine Learning, Deep Learning) et de maîtriser les traitements de données volumineuses et hétérogènes.

Transversalement à ces unités d'enseignement, les applications en informatique (R, Python, Matlab) sont omniprésentes à travers les divers projets qui jalonnent la filière.

Des séminaires professionnels présentent la richesse des métiers offerts en ingénierie statistique et sont en même temps l'occasion d'une présentation par les praticiens des outils ou modèles statistiques utilisés dans l'entreprise.

La pratique de la langue anglaise est renforcée, à travers la rédaction et présentation d'un projet méthodologique correspondant à une initiation à la veille scientifique sur une durée de cinq mois, et à travers certains enseignements pouvant être dispensés en anglais.

### Option Formation Par la Recherche

L'Ensaï offre la possibilité aux élèves de 3ème année qui le souhaitent de se préparer en vue d'une carrière de chercheur en entreprise au sein des services Recherche et Développement ou dans le secteur académique. Dans le cadre de l'option formation par la recherche (Ofpr), ces élèves bénéficient au cours de leur dernière année d'études à l'Ensaï d'un aménagement de leur scolarité qui leur permet de suivre le master 2 de mathématiques fondamentales parcours Aléa à l'Université Rennes 1. La filière comprend ainsi des cours fondamentaux et théoriques en statistique sur lesquels vont s'appuyer le parcours recherche.

A l'issue de ce parcours, ils pourront poursuivre cette formation par une thèse académique ou de type Cifre (Convention Industrielle de Formation par la Recherche).

Chaque année, sur l'effectif total de la filière, environ un tiers poursuit ses études par une thèse Cifre ou académique. Les thèses académiques sont en général encadrées dans des laboratoires de recherche de l'Inserm, de l'Inria, du Crest, du CNRS, de l'Inra ou dépendant d'autres Grandes Écoles. Les thématiques principales de recherche, couvrent de nombreux domaines comme l'industrie, les sciences de l'ingénieur, la fiabilité, l'énergie et l'environnement, la biologie, la santé. A l'issue de la thèse, les types d'emploi possibles sont : chercheurs et enseignants chercheurs dans le domaine des mathématiques appliquées et ingénieurs de recherche (industrie, sociétés de service, ...). Parmi les entreprises signataires de thèses Cifre ou organismes de recherche, on peut citer Renault SA, EUROCOPTER, Thales, Safran, Airbus, SNCF, Orange, EDF, AXA, BNP, ONERA, INRA, INRIA, CEA...

### Débouchés

La filière permet aux étudiants d'accéder aux métiers de la recherche et de l'ingénierie de l'aléatoire dans les domaines de l'industrie, dans les sociétés de conseils et les organismes nationaux d'études et de recherche. Les secteurs concernés sont l'aéronautique, l'agroalimentaire, l'environnement et l'énergie, l'assurance et finance, le biomédical. L'équipe pédagogique collabore, par le biais de contrats de recherche ou de suivi pédagogique, avec EDF, RTE, Air liquide, SNCF, Météo France, Safran, Thalès, AIRBUS, IFREMER, Orange, CEA, INRIA, INRA. L'introduction des séminaires consacrés aux études de cas réels, la pratique des principaux logiciels statistiques, les nombreux projets associés à certains enseignements et les stages en milieu professionnel contribuent à former des ingénieurs opérationnels en milieu professionnel. Les débouchés principaux sont : data-scientist junior, ingénieur R&D, ingénieur sûreté et fonctionnement, ingénieur qualité, consultant statisticien au sein de bureaux d'études de grandes entreprises et d'organismes publics, ...

	Volume horaire				Crédits
	Cours	Ateliers	Projets	Total	
<b>UE1 Machine learning</b>					
Machine learning	18	12		30	2,5
Apprentissage profond	6	6		12	1
Webmining et traitement du langage	9	12		21	1,5
<b>Total</b>	<b>33</b>	<b>30</b>		<b>63</b>	<b>5</b>
<b>UE1 Bis Compléments de Machine learning</b>					
Apprentissage par renforcement	15			15	1
Apprentissage profond avancé	12			12	1
Apprentissage à grande échelle	6	6		12	1
Régularisation	9	6		15	1
<b>Total</b>	<b>42</b>	<b>12</b>		<b>54</b>	<b>4</b>
<b>UE2 Modélisation</b>					
Modélisation non linéaire	15	6		21	2
Théorie des valeurs extrêmes	16,5	6		22,5	1
Statistique bayésienne	9	6		15	1
<b>Total</b>	<b>40,5</b>	<b>18</b>		<b>58,5</b>	<b>4</b>
<b>UE3 Statistique et signal</b>					
Modèles markoviens en analyse d'images ou Functional data analysis (SD)	30	15		45	3
Filtrage linéaire et non linéaire	9	3		12	1
<b>Total</b>	<b>39</b>	<b>18</b>		<b>57</b>	<b>4</b>
<b>UE4 Processus et prévision</b>					
Statistique des processus	21	6		27	2
Séries temporelles avancées	3	21		24	2
<b>Total</b>	<b>24</b>	<b>27</b>		<b>51</b>	<b>4</b>
<b>UE5 Qualité/Fiabilité</b>					
Plans d'expériences	18			18	1
Fiabilité	21	6		27	2
<b>Total</b>	<b>39</b>	<b>6</b>		<b>45</b>	<b>3</b>
<b>UE Projets</b>					
Projet méthodologique		9	27	36	2,5
Projet de fin d'étude		9	27	36	2,5
Data Challenge		12		12	0
Anglais		30		30	1
<b>Total</b>		<b>60</b>	<b>54</b>	<b>114</b>	<b>6</b>
<b>UE Projet professionnel et stages</b>					
Stage de fin d'études					25
Stage d'application					5
Séminaires et projet professionnels (dont droit du travail)	30			30	0
<b>Total</b>	<b>30</b>			<b>30</b>	<b>30</b>
Sport		30		30	0
<b>TOTAL</b>	<b>247,5</b>	<b>201</b>	<b>54</b>	<b>502,5</b>	<b>60</b>

# Présentation des UEs

## UE Apprentissage Automatique

### *Machine Learning*

ECTS : 5 – Volume horaire : 63h

L'apprentissage automatique (machine-learning) est un paradigme essentiellement différent des approches statistiques exploratoires (statistiques au sens classique) ou explicatives (économétrie). Il vise un objectif de prédiction dans la continuité des méthodes d'apprentissage statistique supervisé introduites lors des premières années de la formation d'ingénieur. Largement utilisé dans l'ensemble des professions statistiques à l'heure actuelle (les métiers de la *data science*), l'apprentissage statistique est incontournable dans la formation de l'ingénieur statisticien et trouve de nombreuses applications: prédiction des cours basés à partir d'articles de presse en finance, détection de maladie par imagerie médicale en santé, recommandation de produits en marketing, compression d'images ou encore modèles de traitement du langage, toutes ces applications reposent sur les mêmes bases.

L'UE se compose de 3 matières : apprentissage statistique (*machine-learning*), apprentissage profond (*deep-learning*) et traitement automatique de la langue et fouille du web (*natural language processing and webmining*). L'ensemble de ces matières permettent de mettre en œuvre les techniques classiques, en développant un esprit critique sur leurs limites (sur-apprentissage, grande dimension, représentativité de l'échantillon) et en utilisant des données non structurées (texte, image...). L'UE est accompagnée de séminaires professionnels sur des sujets émergents dans le champs de l'apprentissage statistique.

## UE Compléments d'Apprentissage Automatique

### *Advanced Machine Learning*

ECTS : 4 – Volume horaire : 54h

L'apprentissage automatique (machine-learning) possède de nombreux cas d'application nécessitant une théorie spécifique. C'est le cas par exemple des applications aux images ou à langue (réseaux de neurones convolutifs, récurrents ou avec mécanisme d'attention) ou aux systèmes autonomes (systèmes de recommandation, apprentissage par renforcement). Des problèmes comme le sur-apprentissage possèdent des solutions provenant de la théorie de la compression du signal (régularisation). Enfin, l'apprentissage sur des données suffisamment massives engendre un arbitrage entre qualité et temps de calcul (apprentissage à grande échelle).

L'UE se compose de 4 matières : régularisation, apprentissage à grande échelle, apprentissage profond avancé et apprentissage par renforcement.

## UE Spécifiques DSGS

ECTS : 15 – Volume horaire : 211,5h

Cette filière et les UE associées renforcent les connaissances en modélisation statistique, en abordant les thèmes de la qualité et la fiabilité, du traitement de l'image et du signal, ainsi que la prévision et ses applications. Les élèves sont ainsi capables de s'adapter à des problématiques provenant de différents secteurs d'activité comme l'industrie, le secteur bancaire, l'environnement, les services...

La filière DSGS inclut 4 UE spécifiques : modélisation, statistique et signal, processus et prévision, qualité/fiabilité.

## UE Projets

### *Projects*

ECTS : 6 – Volume horaire : 114h

Les projets sont l'occasion pour les étudiants de se mettre dans une démarche de recherche et développement : diagnostique d'un problème nouveau, lecture de la littérature scientifique sur le sujet, approfondissement d'article, résolution d'un problème en respectant un compromis entre les règles de l'art d'une part, et les contraintes humaines, financières et techniques de l'autre. Ils permettent par ailleurs aux élèves de mesurer l'utilité de toutes les notions acquises au cours des trois années de formation.

Un Data Challenge complète ces cas d'études concrets, à travers la réalisation d'un projet sur un temps court et des contraintes spécifiques. Le cours d'anglais vise soit à préparer les étudiants à passer le TOEIC (niveau B2) ou soit à consolider/entretenir leur niveau.

## UE Projet Professionnels et Stages

### *Professional Project and Internships*

ECTS : 30

Cette UE a pour but de contribuer à la bonne insertion professionnelle des étudiants. Les séminaires professionnels aideront les étudiants à s'orienter dans les choix futurs de leur carrière, notamment le stage. Le cours de droit du travail permet aux étudiants de prendre connaissance le contexte juridique de l'entreprise.

En rejoignant les équipes statistiques d'une entreprise, d'une administration ou d'un centre de recherche, l'élève ingénieur est invité à mettre en application les connaissances acquises au cours de sa scolarité. Le stage de fin d'études doit permettre à l'étudiant d'atteindre au moins deux objectifs :

- Un objectif technique : une mission est confiée à l'élève stagiaire et il doit, à l'aide de ses connaissances théoriques et ses aptitudes, tenter de répondre à cette commande en utilisant au mieux votre environnement proche.
- Un objectif professionnel : l'élève stagiaire évolue dans un environnement professionnel et il doit mettre à profit cette période de stage pour parfaire sa connaissance du monde du travail, accroître ses capacités relationnelles et adopter une démarche d'insertion.



# Descriptifs des enseignements communs

---

UE 1 – Machine Learning

# UE : APPRENTISSAGE AUTOMATIQUE (MACHINE LEARNING)

Correspondant de l'UE : Sébastien Da Veiga  
 Nombre d'ECTS : 5

Volume horaire de travail élève : env. 125h  
 (enseignements + travail personnel)  
 Nombre d'heures d'enseignement : 63h

## Finalité de l'UE :

L'apprentissage automatique (machine-learning) est un paradigme essentiellement différent des approches statistiques exploratoires (statistiques au sens classique) ou explicatives (économétrie). Il vise un objectif de prédiction dans la continuité des méthodes d'apprentissage statistique supervisé introduites lors des premières années de la formation d'ingénieur. Largement utilisé dans l'ensemble des professions statistiques à l'heure actuelle (les métiers de la *data science*), l'apprentissage statistique est incontournable dans la formation de l'ingénieur statisticien et trouve de nombreuses applications: prédiction des cours basés à partir d'articles de presse en finance, détection de maladie par imagerie médicale en santé, recommandation de produits en marketing, compression d'images ou encore modèles de traitement du langage, toutes ces applications reposent sur les mêmes bases.

## Structuration de l'UE :

L'UE se compose de 3 matières : apprentissage statistique (*machine-learning*), apprentissage profond (*deep-learning*) et traitement automatique de la langue et fouille du web (*natural language processing and webmining*). L'ensemble de ces matières permettent de mettre en œuvre les techniques classiques, en développant un esprit critique sur leurs limites (sur-apprentissage, grande dimension, représentativité de l'échantillon) et en utilisant des données non structurées (texte, image...). L'UE est accompagnée de séminaires professionnels sur des sujets émergents dans le champ de l'apprentissage statistique.

## Compétences ou acquis d'apprentissage à l'issue de l'UE :

Cette UE permet de maîtriser des méthodes et des outils de l'ingénieur (identification, modélisation et résolution de problèmes même non familiers et incomplètement définis, l'utilisation des approches numériques et des outils informatiques, l'analyse et la conception de systèmes) en développant l'aptitude à étudier et résoudre des problèmes complexes, à concevoir et mettre en œuvre des projets de collecte et d'analyse d'informations et à concevoir et mettre en œuvre des algorithmes prédictifs de *machine-learning*.

## Les pré-requis de l'UE :

Modélisation statistique de 2<sup>ème</sup> année, méthodes d'optimisation et d'algorithmique, aisance en R et Python.

UE1 - Machine Learning

## APPRENTISSAGE AUTOMATIQUE

### *Machine Learning*

<i>Enseignant</i>	: François PORTIER (Ensaï) et Sébastien DA VEIGA (Ensaï)
<i>Nombre d'ECTS</i>	: 2,5
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: Env. 60h
<i>Répartition des enseignements</i>	: 18h de cours et 12h d'ateliers
<i>Langue d'enseignement</i>	: Anglais (cours) et Français (ateliers)
<i>Logiciels</i>	: R
<i>Documents pédagogiques</i>	: supports de cours, bibliographie et fiches de TP
<i>Pré-requis</i>	: régression, régression logistique, convergence de lois de probabilité, théorème central-limite, algèbre linéaire ; classification par arbre ; programmation avec R ; programmation orientée objet ; optimisation : maximisation d'une fonction, algorithme de Newton

### Modalités d'évaluation :

- 1 compte-rendu de TP
- examen final

### Acquis d'apprentissage (objectifs) :

Comprendre les différents modèles de l'état de l'art (modèle linéaire: régression et classification, pénalisation, méthodes locales, SVM, forêt) ; Connaître les cas d'usage de ces modèles ; Savoir comparer empiriquement différents modèles pour une tâche donnée ; Savoir implémenter les méthodes étudiées en Python ou R

### Principales notions abordées :

Supervised learning; Regression; Classification; Empirical risk minimization; Model evaluation; Cross validation; Functional approximation; Model complexity; Large scale optimization; Stochastic gradient descent; Regularization; RIDGE and LASSO; Support Vector Machine; Kernel trick; Ensemble methods; Aggregation and Boosting; Random forest.

### Références bibliographiques :

- T. HASTIE, J. FRIEDMAN and R. TIBSHIRANI. *The elements of Statistical Learning: Data Mining, Inference and Prediction* (2nd ed.), 2009
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013), *An introduction to statistical learning*, New York: springer.
- Statistical learning with sparsity: the lasso and generalizations, T Hastie, R Tibshirani, M Wainwright – 2019  
Link: <https://web.stanford.edu/~hastie/StatLearnSparsity/>

UE1 - Machine Learning

## APPRENTISSAGE PROFOND

### *Deep Learning*

<i>Enseignant</i>	: Benjamin GIRAULT (Ensay)
<i>Nombre d'ECTS</i>	: 1
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 25h
<i>Répartition des enseignements</i>	: 6h de cours et 6h d'ateliers.
<i>Langue d'enseignement</i>	: Anglais
<i>Logiciels</i>	: Python (tensorflow)
<i>Documents pédagogiques</i>	: supports de cours, bibliographie et fiches de TP
<i>Pré-requis</i>	: R, Python, modélisation statistique, apprentissage statistique, optimisation de fonctions

### Modalités d'évaluation :

1 Quizz et 1 compte-rendu de TP

### Acquis d'apprentissage (objectifs) :

- identifier une tâche particulièrement adaptée pour un réseau de neurones et/ou un réseau de neurones profond
- identifier et implémenter un réseau de neurone approprié pour un tâche d'apprentissage supervisé donnée (architecture, fonction de coût, méthode d'optimisation)
- utiliser et spécialiser un réseau de neurones pré-entraîné

### Principales notions abordées :

Les réseaux de neurones profonds sont au coeur d'avancées rapides en traitement d'image et de la langue depuis les années 2010. Ce cours présente ces modèles, leur fonctionnement, ainsi que comment les utiliser.

- Principe des réseaux de neurones
- propriétés des réseaux de neurones simples
- descente de gradient
- réseaux de neurones profonds
- architectures particulières : réseaux à convolution ; réseaux récurrents.

### Références bibliographiques :

- Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. 2016

UE1 - Machine Learning

# TRAITEMENT AUTOMATIQUE DU LANGAGE ET FOUILLE DU WEB

*Natural language processing and webmining*

<i>Enseignant</i>	: Guillaume Gravier (Irisa)
<i>Nombre d'ECTS</i>	: 1,5
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 38h
<i>Répartition des enseignements</i>	: Cours : 9h – Atelier : 12h
<i>Langue d'enseignement</i>	: Anglais
<i>Logiciels</i>	: Python
<i>Documents pédagogiques</i>	: Support de cours, Supports de TP
<i>Pré-requis</i>	: Programmation avec Python, Apprentissage statistique

## Modalités d'évaluation :

Projet

## Acquis d'apprentissage (objectifs) :

- collecter des données, extraire de l'information et apparier des sources textuelles
- choisir une méthode de traitement automatique de la langue pour une tâche classique (classification, analyse de sentiment, détection > d'entités...)
- se repérer parmi le foisonnement des modèles d'étude de la langue

## Principales notions abordées :

1. What's natural language and its processing
2. The representation of words
3. The representation and classification of documents
4. Language modeling and contextual word embedding
5. Sentence-level tagging (token level tasks)
6. Sequence to sequence models and transformers
7. Overview of standard NLP tasks today

## Références bibliographiques :

- Daniel Jurafsky, James H. Martin. *Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition*, 2nd edition, Prentice-Hall, 2009. Draft of the 3rd edition partly available at <https://web.stanford.edu/~jurafsky/slp3>.
- Yoav Goldberg. *Neural Network Methods for Natural Language Processing*. 2017. An earlier draft is freely available online at <http://u.cs.biu.ac.il/~yogo/nlp.pdf>.
- Kevin Gimpel's lectures (Toyota Technological Institute at Chicago and UChicago) on Natural Language Processing (<https://ttic.uchicago.edu/~kgimpel/teaching/31190-s18/index.html>) and on Advanced Natural Language Processing (<https://ttic.uchicago.edu/~kgimpel/teaching/31210-s19/index.html>).

UE 1bis – Compléments de Machine Learning

# UE : COMPLÉMENTS D'APPRENTISSAGE AUTOMATIQUE (ADVANCED MACHINE LEARNING)

*Correspondant de l'UE*
*Filières concernées*

: Rémi Pépin

Ingénierie des données

Marketing

Génie statistique

*Nombre d'ECTS*

: 4-6 (cf. ci-dessous)

*Volume horaire de travail élève*
*(enseignements + travail personnel)*

: 100-150h (cf. ci-dessous)

*Nombre d'heures d'enseignement*

: 54h-78h (cf. ci-dessous)

## Finalité de l'UE :

L'apprentissage automatique (machine-learning) possède de nombreux cas d'application nécessitant une théorie spécifique. C'est le cas par exemple des applications aux images ou à langue (réseaux de neurones convolutifs, récurrents ou avec mécanisme d'attention) ou aux systèmes autonomes (systèmes de recommandation, apprentissage par renforcement). Des problèmes comme le sur-apprentissage possèdent des solutions provenant de la théorie de la compression du signal (régularisation). Enfin, l'apprentissage sur des données suffisamment massives engendre un arbitrage entre qualité et temps de calcul (apprentissage à grande échelle).

## Structuration de l'UE :

L'UE se compose de 6 matières, et seule la filière ingénierie des données la suit dans son intégralité. Le tableau ci-dessous résume les ECTS et horaires par filière :

	Heures de cours	Heures de travail	ECTS	Ingénierie des données	Marketing	Génie statistique
Régularisation	15	30	1	X	X	X
Apprentissage à grande échelle	12	25	1	X		X
Apprentissage profond avancé	12	25	1	X	X	X
Traitement automatique de la langue avancé	12	18	1	X	X	
Systèmes de recommandation	12	25	1	X	X	
Apprentissage par renforcement	15	25	1	X	X	X
<b>Total</b>				<b>72 h cours 148 h perso 6 ECTS</b>	<b>60 h cours 123 h perso 5 ECTS</b>	<b>54h cours 105h perso 4 ECTS</b>

## Compétences ou acquis d'apprentissage à l'issue de l'UE :

Cette UE permet de consolider les acquis d'apprentissage de l'UE 1 Apprentissage statistique et de développer des compétences pour résoudre des problèmes spécifiques à des sous-disciplines d'intérêt (traitement de l'image, traitement du signal, traitement de la langue, systèmes autonomes, données massives...).

## Les pré-requis de l'UE :

UE 1 Apprentissage statistique, méthodes d'optimisation et d'algorithmique, panorama du big data, aisance en Python.

UE 1bis - Compléments de Machine Learning

## REGULARISATION

### *Regularization*

<i>Enseignant</i>	: Cédric HERZET (Inria) & Clément ELVIRA (Inria)
<i>Filières concernées</i>	Ingénierie des données Marketing Génie statistique
<i>Nombre d'ECTS</i>	: 1
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 30h
<i>Répartition des enseignements</i>	: Cours : 9h – Atelier : 6h
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: Python
<i>Documents pédagogiques</i>	: Supports de cours, Supports de TP, Bibliographie
<i>Pré-requis</i>	: Optimisation, Python, Algèbre

### Modalités d'évaluation :

Un examen sur table de 2 heures avec questions de cours et résolution de problèmes, 1 compte-rendu de TP

### Acquis d'apprentissage (objectifs) :

- identifier les cas d'apprentissage statistique et les problèmes inverses où la régularisation est utile : comprendre quels sont les motivations et les objectifs de la pénalisation
- connaître les modèles de régularisation les plus courants et savoir quelles caractéristiques de reconstruction ils favorisent choisir une régularisation parmi les méthodes les plus courantes et estimer un modèle régularisé par une méthode de descente de gradient
- connaître et comprendre les différents types de problèmes d'optimisation et les algorithmes qui permettent de les résoudre numériquement

### Principales notions abordées :

- Ingrédients principaux des problèmes inverses et d'apprentissage statistique + exemples pratiques
- Motivations et objectifs de la régularisation
- Types de régularisation et fonctions de régularisation couramment rencontrées
- Caractérisation des problèmes pénalisés: existence de solution, unicité, conditions d'optimalité.
- Méthodes numériques de résolution de problèmes d'optimisation
- Conditions théoriques de reconstruction correcte

### Références bibliographiques :

- Hastie, Trevor, Robert Tibshirani, and Martin Wainwright. 2015. *Statistical Learning with Sparsity: The Lasso and Generalizations*. CRC Press.
- C. Bishop. *Pattern recognition and machine learning*. Springer-Verlag New York, 2006.
- S. Foucart and H. Rauhut. *A mathematical introduction to compressive sensing*. Applied and Numerical Harmonic Analysis. Birkhäuser, 2013.
- D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, USA, 2003.



UE 1bis - Compléments de Machine Learning

## APPRENTISSAGE À GRANDE ÉCHELLE

### *Large-scale machine learning*

<i>Enseignant</i>	: Cédric HERZET (ENSAI)
<i>Filières concernées</i>	: Génie statistique Ingénierie des données
<i>Nombre d'ECTS</i>	: 1
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 25h
<i>Répartition des enseignements</i>	: - Cours : 6h - Atelier : 6h
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: Python
<i>Documents pédagogiques</i>	: support de cours, fiches de TP
<i>Pré-requis</i>	: panorama du big data, machine-learning, apprentissage statistique, optimisation

### Modalités d'évaluation :

2 quizz en contrôle continu, 1 TP noté

### Acquis d'apprentissage (objectifs) :

- Le passage à des bases de données à grande échelle modifie certains usages en apprentissage statistiques. Nous en verrons quelques exemples, avec les justifications théoriques sous-jacentes.
  - objectif 1 : être en mesure de choisir des approches appropriées pour un problème « de grande taille » donné, ce qui nécessite de...
  - Objectif 2 : connaître le comportement en termes de taille de stockage et/ou de temps de calcul et/ou de qualité d'estimation des approches présentées.

### Principales notions abordées :

- la descente de gradient stochastique
- données et modèles parcimonieux (matrices de rang faible, attributs de Fourier aléatoires, distillation, ...)
- le calcul distribué

### Références bibliographiques :

- Introduction to High Performance Computing for Scientists and Engineers, Georg Hager, Gerhard Wellein, CRC Press, 2010
- Introduction to High Performance Scientific Computing, Victor Eijkhout, Edmond Chow, Robert van de Geijn, 2014
- Bekkerman, Ron, Mikhail Bilenko, and John Langford. n.d. *Scaling up Machine Learning: Parallel and Distributed Approaches*.
- The MIT Press. n.d. "Large-Scale Kernel Machines." Accessed September 1, 2020. <https://mitpress.mit.edu/books/large-scale-kernel-machines>.

UE 1bis - Complément de Machine Learning

## APPRENTISSAGE PROFOND AVANCÉ

### *Advanced Deep Learning*

<i>Enseignant</i>	: Benjamin GIRAULT (Ensaï)
<i>Nombre d'ECTS</i>	: 1
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 25h
<i>Répartition des enseignements</i>	: 3h de cours et 9h d'ateliers.
<i>Langue d'enseignement</i>	: Anglais
<i>Logiciels</i>	: Python (tensorflow)
<i>Documents pédagogiques</i>	: supports de cours, bibliographie et fiches de TP
<i>Pré-requis</i>	: R, Python, algèbre linéaire, optimisation de fonctions, apprentissage profond

### Modalités d'évaluation :

1 Quizz et 1 compte-rendu de TP

### Acquis d'apprentissage (objectifs) :

- Les réseaux de neurones profonds s'adaptent à de nombreuses tâches. Ce cours donne quelques exemples de ces adaptations.
  - objectif : adapter l'architecture d'un réseau de neurones pour des applications non-triviales (multi-vue, image vers texte, texte vers image, compression, génération d'adversaires, ...)
  - identifier une tâche particulièrement adaptée pour un réseau de neurones et/ou un réseau de neurones profond
  -

### Principales notions abordées :

- Quelques exemples d'architectures non-triviales
- Auto-encodeurs
- Réseaux adversariaux
- Réseaux de neurones bayésiens

### Références bibliographiques :

- Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. 2016

UE 1bis - Complément de Machine Learning

## APPRENTISSAGE PAR RENFORCEMENT

### *Re-inforcement learning*

<i>Enseignants</i>	: Frédéric Logé (Sonio) Rémi Besson (Sonio)
<i>Filières concernées</i>	: Marketing Ingénierie des données Génie statistique
<i>Nombre d'ECTS</i>	: 1
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 25h
<i>Répartition des enseignements</i>	: 12h de cours / ateliers
<i>Langue d'enseignement</i>	: Anglais
<i>Logiciels</i>	: Python
<i>Documents pédagogiques</i>	: supports de cours, bibliographie et fiches de TP
<i>Pré-requis</i>	: Apprentissage profond et Chaînes de Markov facilitent la compréhension

### Modalités d'évaluation :

1 mini-projet

### Acquis d'apprentissage (objectifs) :

- Objectif 1 : savoir modéliser une situation comme un problème d'apprentissage par renforcement (choix d'une fonction de récompense, d'un espace d'état, d'un espace d'actions...)
- Objectif 2 : être capable d'apprendre une politique à partir de données et de l'évaluer
- Objectif 3 : comprendre les limites de la modélisation par apprentissage par renforcement

### Principales notions abordées :

- modèles de bandits
- dilemme exploration-exploitation
- algorithmes de bandits
- regret
- processus de décision markovien (MDP)
- équation de Bellman
- fonction état-valeur
- algorithmes pour MDP

### Références bibliographiques :

- Tor Lattimore and Csaba Szepesvári, Bandit Algorithms, 2020 (<https://torlattimore.com/downloads/book/book.pdf>)
- Sutton and Barto, Introduction to Reinforcement Learning, 2020 (<http://incompleteideas.net/book/RLbook2020.pdf>)
- David Silver course on RL: <https://www.davidsilver.uk/teaching/>
- GYM, OpenAI, <https://gym.openai.com/>

UE Projets

## UE : PROJETS

<i>Correspondant de l'UE</i>	: Sébastien DA VEIGA
<i>Nombre d'ECTS</i>	: 6
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: Entre 120h et 150h
<i>Nombre d'heures d'enseignement</i>	: Suivis réguliers avec les encadrants

### Finalité de l'UE :

Les projets sont l'occasion pour les étudiants de mettre en œuvre leurs connaissances acquises à l'ENSAI sur des cas d'études concrets. Ils visent à mettre en œuvre les outils et connaissances acquises en statistique, en informatique et en économie, dans une démarche de résolution de problèmes concrets type ingénieur.

Les projets se déclinent en deux versions: le projet méthodologique, en langue anglaise, vise à approfondir une thématique centrée autour d'un ou plusieurs articles scientifiques ; le projet de fin d'études (dans toutes les filières sauf biostatistique), plus appliqué, nécessairement sur des données issues d'une collecte, vise à proposer une solution pratique à une problématique générale proposée par une entreprise ou un laboratoire de recherche. À eux deux, ces projets couvrent toute l'étendue d'une démarche de développement: diagnostique d'un problème nouveau, lecture de la littérature scientifique sur le sujet, résolution d'un problème en respectant un compromis entre les règles de l'art d'une part, et les contraintes humaines, financières et techniques de l'autre. Ils permettent par ailleurs aux élèves de mesurer l'utilité de toutes les notions acquises au cours des trois années de formation.

Selon les filières, la réalisation d'un Data Challenge complète ces cas d'études concrets, à travers la réalisation d'un projet sur un temps court et des contraintes spécifiques.

### Structuration de l'UE :

Projet méthodologique: approfondissement d'une démarche de revue de littérature en langue anglaise ; constitue la partie théorique de recherche d'information dans une démarche de recherche et développement. Les élèves désireux de développer un projet de recherche académique peuvent privilégier un projet plus conséquent.

Projet de fin d'étude: approfondissement d'une démarche pratique, sachant composer avec des contraintes opposées, entre rigueur scientifique et nécessités pratiques ; constitue la partie implémentation dans une démarche de recherche et développement.

Data Challenge (optionnel, selon les filières) : rassembler sur une période très courte différentes équipes de profils variés afin de collaborer sur un projet.

### Compétences ou acquis d'apprentissage à l'issue de l'UE :

Ces projets concluent la formation d'ingénieur de l'Ensaï, et mobilisent un ensemble de compétences de l'ingénieur : capacité à trouver l'information pertinente, à faire une veille scientifique, à prendre en compte les enjeux de l'entreprise, à travailler dans un contexte international, tout en mobilisant des compétences techniques pour résoudre des problèmes complexes, et mener une démarche scientifique.

### Les pré-requis de l'UE :

Méthodes de travail des projets de 1<sup>ère</sup> et 2<sup>ème</sup> année.

UE Projets

## PROJET METHODOLOGIQUE

### *Methodological project*

<i>Enseignant</i>	: Divers intervenants
<i>Filières concernées</i>	: Toutes filières, avec des modalités différentes
<i>Nombre d'ECTS</i>	: 1,5 ou 2,5 selon modalités (sauf filière Biostatistique) 3 (filière Biostatistique)
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: Entre 40h et 75 h selon modalités
<i>Répartition des enseignements</i>	: 9h d'ateliers, et suivis réguliers
<i>Langue d'enseignement</i>	: Anglais
<i>Logiciels</i>	: /
<i>Documents pédagogiques</i>	: /
<i>Pré-requis</i>	: Connaissances générales en statistiques (1 <sup>e</sup> et 2 <sup>e</sup> années)

### Modalités d'évaluation :

Le projet méthodologique consiste en la production d'un article de synthèse sur un sujet de recherche à choisir parmi un catalogue. L'évaluation tient compte de l'article rédigé et de la réalisation d'une soutenance.

### Acquis d'apprentissage (objectifs) :

Les objectifs du projet méthodologique, et donc les compétences qui sont renforcées grâce à celui-ci, sont multiples:

- familiarisation avec la forme des productions académiques (articles notamment), en lecture comme en écriture
- capacité à faire une revue de littérature mélangeant ouvrages scientifiques et professionnels
- prise de conscience des enjeux autour de la reproductibilité des résultats de recherche
- communication sur des sujets techniques

Pour les étudiants désirant s'initier à la recherche, et pour l'ensemble des étudiants de la filière Biostatistique, le projet méthodologique prendra la forme de l'étude détaillée, la reproduction et/ou la mise en perspective d'un ou de plusieurs articles, au lieu d'une revue de littérature.

À cela s'ajoute les objectifs spécifiques à la production d'un travail technique en langue anglaise: mise en œuvre d'un projet complexe en langue anglaise, communication écrite et orale, acquisition d'un vocabulaire spécialisé, maîtrise de différents niveaux de langues en terme de style (oral vs. écrit) et de technicité (vulgarisation vs. spécialisation), mise en place de stratégies pour faire face à des difficultés linguistiques.

### Principales notions abordées :

Travail de recherche en groupe suivi par un chercheur (env. 5 séances) et un professeur d'anglais (4 séances). Des séances de travail personnel sont réservées dans l'emploi du temps.

UE Projets

## PROJET DE FIN D'ETUDES

### *Methodological project*

<i>Enseignant</i>	: Divers intervenants
<i>Filières concernées</i>	Toutes filières sauf Biostatistique
<i>Nombre d'ECTS</i>	: 2,5 ou 3,5 selon modalités
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: Entre 60h et 90 h
<i>Répartition des enseignements</i>	: 9h d'ateliers et suivis réguliers
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: /
<i>Documents pédagogiques</i>	: /
<i>Pré-requis</i>	: Connaissance de statistiques des trois années de l'ENSAI

### Modalités d'évaluation :

Le projet de fin d'études consiste en la production d'une étude statistique de niveau professionnel dans le monde de l'entreprise ou de la recherche, parmi un catalogue de sujet mis à disposition des élèves. Le projet est évalué à travers un rapport et une soutenance.

Les élèves désireux de s'initier à la recherche passant plus de temps sur le projet méthodologique, les attendus du projet de fin d'études sont ajustés pour tenir compte du moindre temps consacré que leurs camarades.

### Acquis d'apprentissage (objectifs) :

Les objectifs du projet de fin d'études, et donc les compétences qui sont renforcées grâce à celui-ci, sont multiples:

- mise en situation professionnelle
- capacité à définir une stratégie d'étude en réponse à une demande client
- mobilisation des compétences techniques (statistiques, économiques, informatiques)
- compromis entre rigueur scientifique et contraintes pratiques (limitations financières, logicielles, cognitives, temporelles...)
- travail de groupe
- gestion d'un projet sur le temps long
- communication (écrite, orale) sur des sujets techniques

### Principales notions abordées :

Travail autonome en groupe suivi par un professionnel de l'entreprise ou de la recherche (env. 5 séances).

UE Projets

## DATA CHALLENGE

### *Data Challenge*

<i>Enseignant</i>	: Divers intervenants industriels (correspondante : Salima El Kolei)
<i>Nombre d'ECTS</i>	: Pas d'attribution d'ECTS
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 2 journées
<i>Répartition des enseignements</i>	: 12h d'ateliers
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: /
<i>Documents pédagogiques</i>	: /
<i>Pré-requis</i>	: Méthodes de travail des projets, Compétences statistiques et informatiques de 3ème année

### Modalités d'évaluation :

Les élèves participent au data challenge proposé à l'Ensaï ouvert également aux élèves de deuxième année. Il n'y a pas d'évaluation.

### Acquis d'apprentissage (objectifs) :

Le data challenge permet de rassembler sur une période très courte différentes équipes de profils variés afin de collaborer sur un projet. Cette expérience se rapproche des conditions réelles dans laquelle évoluent les datascientists au sein des entreprises. Il permet, à partir des mécanismes du jeu, de dynamiser et d'articuler la pédagogie autour d'un besoin concret d'entreprise et d'un événement qui s'achève par une évaluation objective. De nombreux challenges sont proposés autour de la Data ou présentant des problématiques Data importantes.

L'objectif de ce cours est de valoriser les compétences transversales acquises dans ce contexte opérationnel. Les compétences qui sont renforcées grâce à celui-ci sont multiples:

- Comprendre les problèmes à résoudre.
- Travailler en mode projet avec des contraintes.
- S'intégrer et s'adapter dans un contexte pluridisciplinaire. Selon les challenges, les compétences seront mobilisées à géométrie variable.
- S'adapter à la réalité de la Data d'entreprise (données non structurées, manquantes, volumétrie...)
- Communication orale des résultats (pitch...)

### Principales notions abordées :

Travail en groupe sur un temps court.

UE Projets

## ANGLAIS

### English

<i>Enseignant</i>	: Divers intervenants (correspondant : Todd Donahue)
<i>Nombre d'ECTS</i>	: 1
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 40h
<i>Répartition des enseignements</i>	: 15h de cours, 15h d'aide au projet de fin d'études
<i>Langue d'enseignement</i>	: Anglais
<i>Logiciels</i>	: Sans objet
<i>Documents pédagogiques</i>	: Sous Moodle
<i>Pré-requis</i>	: Aucun

### Modalités d'évaluation :

L'examen final prend la forme d'une simulation d'entretien d'embauche. Cet examen oral durera environ 25 minutes, sera noté, et permettra d'évaluer le niveau d'expression orale sur l'échelle CECRL (Cadre européen commun de référence pour les langues). Le CV et la lettre faite pour cet exercice seront évalués et feront partie de la note finale. L'anglais est également évalué à travers le rapport écrit et la soutenance orale du projet de fin d'études. Le niveau acquis apparaîtra sur le Supplément au diplôme. L'objectif de la CTI (Commission des Titres d'Ingénieur) pour tous les élèves ingénieurs est d'atteindre le niveau B2.

### Acquis d'apprentissage (objectifs) :

- maîtriser une ou plusieurs langues étrangères
- savoir candidater et réussir un recrutement en langue anglaise
- contextualiser et prendre en compte les enjeux et les besoins de la société
- se connaître, s'auto-évaluer, gérer ses compétences, opérer ses choix professionnels
- s'intégrer et évoluer dans un groupe pour mener à bien un projet dans un contexte international et/ou pluriculturel
- savoir identifier les informations pertinentes, à les évaluer et à les exploiter

### Principales notions abordées :

Pour les élèves qui n'ont pas eu un score d'au moins 785 au TOEIC : pendant les 5 premières séances, la plupart des cours seront basés sur la préparation à cet examen. Les ressources informatiques de l'Ecole doivent aussi être mises à profit (pages Moodle, TOEIC Mastery), ainsi que les méthodes disponibles à la bibliothèque. Pour les autres élèves, les cours seront organisés par groupe de niveau et conçus afin de les préparer à affronter le monde professionnel sur le plan international. Les thèmes suivants seront traités : « Leading meetings », « Interviews », « Presentations », « Taking decisions », et « Negotiating deals », et « Cultural and Political Current Events ». Ensuite, les 5 dernières séances seront consacrées au travail de rédaction/correction des rapports faits en anglais dans chaque filière ainsi qu'à la préparation des soutenances orales. Chaque responsable de filière indiquera aux élèves, en début d'année, le projet concerné et les modalités de notation. Les élèves recevront des consignes détaillées avant de démarrer ces cinq séances, afin d'arriver à la première séance avec une première version ou extrait de leur rapport en anglais prêt pour correction et relecture. Pour tout complément d'information, chaque élève peut consulter le Programme des enseignements : Langues étrangères, distribué au début de l'année académique.

Références bibliographiques : Définies par chaque intervenant.



# Descriptifs des enseignements de la filière

---

UE Spécifiques filière GS

## UE SPECIFIQUES FILIERE GS

<i>Correspondant de l'UE</i>	: Sébastien Da Veiga
<i>Nombre d'ECTS</i>	: 16
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: De 25 à 30h par ECTS
<i>Nombre d'heures d'enseignement</i>	: 221h

### Finalité des UE :

Cette filière et les UE associées renforcent les connaissances en modélisation statistique, en abordant les thèmes de la qualité et la fiabilité, du traitement de l'image et du signal, ainsi que la prévision et ses applications, notamment dans le domaine de l'environnement. Les élèves sont ainsi capables de s'adapter à des problématiques provenant de différents secteurs d'activité comme l'industrie, le secteur bancaire, l'environnement, les services...

### Structuration de l'UE :

La filière GS inclut 4 UE spécifiques : modélisation, traitement de l'image, processus et prévision, qualité/fiabilité.

### Compétences ou acquis d'apprentissage à l'issue de l'UE :

- Savoir introduire de nouvelles méthodes statistiques et expertiser les pratiques existantes ;
- Capacité à mettre en œuvre des méthodes d'analyse d'image ainsi que les techniques de filtrage;
- Capacité à mobiliser et d'utiliser à bon escient des outils d'aide à la décision dans le domaine de la gestion des risques (industriels (fiabilité-qualité), environnementaux et financiers) ;
- Capacité à construire des modèles complexes ou novateurs de prévisions ;
- Capacité à s'adapter à des problématiques provenant de différents secteurs d'activités comme l'industrie, le secteur bancaire, l'environnement, les services, en maîtrisant les techniques statistiques requises (statistique bayésienne, statistique des processus, chaînes de Markov, *machine learning*...) et les modalités de traitement des données massives ;
- Capacité à innover et à mener des activités de recherche en statistique fondamentale et appliquée.

UE 2 – Modélisation

## MODELISATION NON LINEAIRE

### *Non Linear modelisation*

<i>Enseignant</i>	: Marian Hristache (ENSAI)
<i>Nombre d'ECTS</i>	: 2
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 33h
<i>Répartition des enseignements</i>	: 15h de cours, 6h de TP
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: R/Python
<i>Documents pédagogiques</i>	: Slides
<i>Pré-requis</i>	: algèbre linéaire, une connaissance de la programmation R ou Python, régression linéaire multiple

### Modalités d'évaluation :

TP noté

### Acquis d'apprentissage (objectifs) :

- Connaître les modèles de régression paramétrique non-linéaires et non paramétrique ;
- savoir utiliser les procédures d'estimation standard pour ce type de modélisation ;
- être en mesure de déployer la méthodologie étudiée sur des données réelles.

### Principales notions abordées :

1. Améliorations possibles du modèle linéaire
2. Extensions du modèle linéaire (polynôme, spline, ondelettes)
3. Modèle additif généralisé
4. Modèle paramétrique non linéaire

### Références bibliographiques :

- ANTONIADIS, J. BERRUYER, R. CARMONA. *Régression non linéaire et applications*. Economica, 1992
- R.L. EUBANK. *Nonparametric Regression and Spline Smoothing* : 2nd Edition. Dekker 1999
- D. RUPPERT, M. P. WAND, R. J. CARROLL. *Semiparametric Regression*. Cambridge University Press, 2003
- J. O. RAWLINGS, S. G. PANTULA, D. A. DICKEY. *Applied Regression Analysis. A Research Tool (2<sup>nd</sup> ed.)*. Springer, 1998.
- T. P. RYAN. *Modern Regression Methods (2<sup>nd</sup> ed.)*. Wiley, 2008
- T. HASTIE, R. TIBSHIRANI, Generalized additive models, *Statist. Sci* 1, 297-318, 1986.
- T. HASTIE, R. TIBSHIRANI, J. FRIEDMAN, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2<sup>nd</sup> ed.)*, Springer, 2009
- R. TIBSHIRANI, Regression shrinkage and selection via the lasso. *J. Royal. Statist. Soc B.*, Vol. 58, No. 1, pages 267-288, 1996.

UE 2 – Modélisation

## THEORIE DES VALEURS EXTREMES

### *Extreme Value Theory*

<i>Enseignant</i>	: Gilles Stupfler (ENSAI) – cours. Yasser ABBAS
<i>Nombre d'ECTS</i>	: 1
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 25h
<i>Répartition des enseignements</i>	: 16h30 de cours & 6h de TP
<i>Langue d'enseignement</i>	: Anglais
<i>Logiciels</i>	: R/Python
<i>Documents pédagogiques</i>	: Notes de cours, exercices intégrés, correction des exercices, code R
<i>Pré-requis</i>	: Intégration, Probabilités et Statistique de 1A

### Modalités d'évaluation :

Examen écrit

### Acquis d'apprentissage (objectifs) :

- Comprendre la différence entre la statistique classique et celle des valeurs extrêmes ;
- Maîtriser la gamme des modèles statistiques usuels de la théorie des valeurs extrêmes ;
- Savoir utiliser les procédures d'estimation standard des paramètres extrêmes d'une distribution.

### Principales notions abordées :

Modèles de valeurs extrêmes : théorème de Fisher-Tippett-Gnedenko, famille de lois GEV, théorème de Pickands-Balkema-de Haan, famille de lois GP. Domaines d'attraction de Fréchet, Weibull et Gumbel. Exemples classiques.

Estimation dans les modèles à queue lourde : définition, méthode d'extrapolation de Weissman, estimation de l'indice de queue par quasi-vraisemblance (estimateur de Hill), biais, choix du seuil, estimation de quantile extrême. Cas pratique sur données actuarielles (avec R).

Estimation hors des modèles à queue lourde : extrapolation via l'approximation GP, estimation des paramètres (estimateur de Pickands, estimateur des moments, pseudo-vraisemblance), estimation de quantile extrême. Cas pratique sur données environnementales (avec R).

### Références bibliographiques :

- Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*, John Wiley & Sons, Chichester.
- Billingsley, P. (1995). *Probability and measure (third edition)*, John Wiley and Sons.
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*, Springer-Verlag, London.
- Dekkers, A.L.M., Einmahl, J.H.J., de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution, *Annals of Statistics* 17(4): 1833-1855.
- Diebolt, J., Guillou, A., Rached, I. (2007). Approximation of the distribution of excesses through a generalized probability-weighted moments method, *Journal of Statistical Planning and Inference* 137(3): 841-857.
- Drees, H., Ferreira, A., de Haan, L. (2004). On maximum likelihood estimation of the extreme value index, *Annals of Applied Probability* 14(3): 1179-1201.
- Embrechts, P., Klüppelberg, C., Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*, Springer-Verlag, Berlin.

- de Haan, L., Ferreira, A. (2006). *Extreme Value Theory: An Introduction*, Springer, New York. Erratum : <https://personal.eur.nl/ldehaan/EVTbook.correction.pdf>
- Hall, P. (1982). On estimating the endpoint of a distribution, *Annals of Statistics* 10(2): 556-568.
- Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution, *Annals of Statistics* 3(5): 1163-1174.
- Hosking, J.R.M., Wallis, J.R. (1987). Parameter and quantile estimation for the generalized Pareto distribution, *Technometrics* 29(3): 339-349.
- Pickands, J. (1975). Statistical inference using extreme order statistics, *Annals of Statistics* 3(1): 119-131.
- Resnick, S.I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, Springer, New York.
- Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations, *Journal of the American Statistical Association* 73(364): 812-815.
- Zhou, C. (2009). Existence and consistency of the maximum likelihood estimator for the extreme value index, *Journal of Multivariate Analysis* 100: 794-815.

UE 2 – Modélisation

# STATISTIQUE BAYESIENNE

## Bayesian Inference

<i>Enseignant</i>	: Sophie ANCELET (IRSN) Eric PARENT (AgroParisTech/INRA)
<i>Nombre d'ECTS</i>	: 1
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 33h
<i>Répartition des enseignements</i>	: - Cours : 9h - Atelier : 6h
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: R (packages rjags, rstan, coda, ggcmc), JAGS, STAN
<i>Documents pédagogiques</i>	: Slides + Documents R Markdown pour corrections TP
<i>Pré-requis</i>	: Probabilité, statistique inférentielle, SAS, R (1A) Régression, GLM, Chaîne de Markov, Calcul bayésien (2A)

La statistique bayésienne se distingue par l'utilisation d'une loi de probabilité sur l'ensemble des objets (ex : paramètres, modèles, etc...) sur lesquels l'inférence est menée. Etant donné un modèle probabiliste paramétrique (cas exclusivement considéré dans ce cours), les paramètres inconnus du modèle sont traités comme des variables aléatoires : une loi de probabilité dite loi a priori leur est assignée afin de décrire un état de connaissances (i.e., incertitude épistémique) avant la collecte de données. La formule de Bayes permet ensuite une mise à jour de cet état de connaissances via l'assimilation de données observées afin d'obtenir une loi de probabilité dite loi a posteriori sur les paramètres inconnus. En pratique, la loi a posteriori n'a généralement pas une forme analytique connue. Différentes approximations de la loi a posteriori seront présentées, notamment les Algorithmes Monte-Carlo par Chaînes de Markov (MCMC) (formes classiques et plus avancées) dans la continuité du cours de calcul bayésien de 2A. Une initiation aux logiciels gratuits JAGS et STAN (packages R « rjags » et « rstan ») sera proposée lors des ateliers. Le cours traitera enfin du problème de la sélection et de la validation prédictive de modèles sous le paradigme bayésien, et sera illustrée par des exemples empreints aux sciences de la vie au sein desquelles l'emploi des méthodes statistiques bayésiennes progresse considérablement.

*Enseignement destiné aux élèves de la filière « Data Science, Modélisation économique & Santé » et « Data Science en Santé & Biostatistique ».*

### Principales notions abordées :

Formule de Bayes / Loi a priori / Loi a posteriori / Loi prédictive a posteriori / Incertitude épistémique  
 Estimateurs bayésiens / Intervalles de crédibilité  
 Lois a priori conjuguées / Lois a priori de Jeffreys  
 Modèles à variables latentes et représentation hiérarchique / Graphe acyclique orienté  
 Approximations déterministes de la loi a posteriori  
 Algorithmes MCMC (Principe, Gibbs/Metropolis-Hastings/Dynamiques Hamiltoniennes, diagnostics de convergence)  
 Validation prédictive bayésienne (posterior predictive check, cross validation)  
 Sélection bayésienne de modèles (Facteur de Bayes, Deviance Information Criterion, Widely Applicable Information Criterion)

### Acquis d'apprentissage (objectifs) :

- Expliquer le principe général de la démarche statistique bayésienne
- Faire un choix pertinent de lois a priori
- Formuler l'écriture probabiliste d'un modèle à variables latentes sous forme hiérarchique et donner sa représentation sous forme de graphe acyclique orienté
- Calcul de lois a posteriori (*uniquement pour SSV*)

- Mener l'inférence bayésienne de modèles classiques (e.g., régression linéaire, GLM) et de modèles à variables latentes (e.g., modèles mixtes) à l'aide des packages R « rjags » et « rstan »
- Mener une étude de convergence d'un algorithme MCMC
- Comparer différents modèles à l'aide de critères de sélection bayésiens
- Valider un modèle – du point de vue prédictif - sous le paradigme bayésien

## Modalités d'évaluation :

Devoir maison

## Références bibliographiques :

Collectif BIOBAYES: ALBERT I., ANCELET S., DAVID O., DENIS J.B., MAKOWSKI D., PARENT E., SOUBEYRAND S. Méthodes statistiques bayésiennes. Bases théoriques et applications en alimentation, environnement et génétique. *Ellipses*, 2015

GELMAN, A., CARLIN, J. B., STERN, H. S. and RUBIN, D. B. Bayesian data analysis. CRC Press, 2013

UE 3 - Statistique et signal

## MODELES MARKOVIENS

*Markovian models in image analysis*

*Enseignant* : Charles Kervrann (Inria)

*Nombre d'ECTS* : 3

*Volume horaire de travail élève* : 69h  
(enseignements + travail personnel)

*Répartition des enseignements* : 30h de cours, 15h de TP

*Langue d'enseignement* : Français

*Logiciels* : Matlab

*Documents pédagogiques* : Slides

*Pré-requis* : chaînes de Markov, statistique exploratoire multivariée

### Modalités d'évaluation :

Examen écrit et TP noté

### Acquis d'apprentissage (objectifs) :

- Maîtriser les principaux concepts et outils de la modélisation markovienne/gibbsienne, en se limitant au cas de graphes de dépendance finis
- Savoir utiliser leur emploi sur divers problèmes archétypes (modélisation et classification de texture, segmentation au sens des niveaux de gris, restauration d'image avec préservation de discontinuités, etc.).
- Mise en pratique sur des cas concrets

### Principales notions abordées :

Aux frontières de la physique statistique, de la théorie des probabilités et du traitement du signal, les champs de Markov et les distributions de Gibbs fournissent un cadre probabiliste très bien adapté à la description et à la résolution des problèmes d'analyse d'images où un nombre considérable de variables, observées ou non, interagissent localement. Les propriétés markoviennes résultantes permettent de définir des techniques d'échantillonnage, d'estimation bayésienne des variables non-observées, et d'estimation des paramètres mis en jeu. Dans ce cours sera abordé les points suivants : champs markoviens gaussiens sur grilles, champ de Markov sur graphe, problèmes inverses et estimation bayésienne.

### Références bibliographiques :

- J. BESAG, Spatial interaction and the statistical analysis of lattice systems, *J. Royal Statist. Soc.*, 36:192-236, 1974
- S. GEMAN, D. GEMAN, *Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images*, *IEEE Trans. Pattern Anal. Machine Intell.*, 6(6):721-741, 1984
- R. KINDERMANN, J.L. SNELL, *Markov random fields and their applications*, Amer. Math. Soc., Providence, 1980
- R. CHELLAPPA, A.K. JAIN (Eds.), *Markov random fields: theory and applications*, Academic Press, Boston, 1993
- B. CHALMOND, *Éléments de modélisation pour l'analyse d'images (Mathématiques et Applications (33))* (French Edition), Springer-SMAI 2000.



UE 3 – Statistique et signal

## ANALYSE DE DONNEES FONCTIONNELLES

### *Functional Data Analysis*

<i>Professors</i>	: Valentin PATILEA (Ensaï) & Eftychia SOLEA (Ensaï)
<i>ECTS Credits</i>	: 3
<i>Estimated personal workload</i> (beyond lecture and tutorial time)	: 1.5 h personal workload per 1h lecture
<i>Lectures and Tutorials</i>	: 30 hrs (ENSAI) including 3 hrs of independent work
<i>Teaching language</i>	: English
<i>Software</i>	: R
<i>Course materials</i>	: Lecture notes and textbook (see below)
<i>Prerequisites</i>	: Statistical inference and methods, Multivariate statistical analysis

### Learning Objectives :

This course aims to provide an introduction to functional data analysis. The fundamental statistical tools for modeling and analyzing such data will be explored. This course introduces ideas and methodology in functional data analysis (FDA) as well as the use of software. Students will learn the idea of different methods and the related theory, and also the numerical and estimation routines to perform functional data analysis. Students will also have an opportunity to learn how to apply FDA to a wide array of application areas. The course will demonstrate applications where FDA techniques have clear advantage over classical multivariate techniques. Some recent development in FDA will also be discussed.

### Main Subjects covered :

Chapter 1. Introduction.

Chapter 2. Representing functional data and exploratory data analysis. Including: basic expansions, FPCA, derivatives, penalised smoothing, registration, fda package.

Chapter 3. Elements of Hilbert space theory and random functions.

Chapter 4. Estimation and inference from a random sample. Including, estimation of functional principal component analysis (FPCA), testing hypothesis about the mean.

Chapter 5. Functional Linear regression models. Including: Functional linear regression models with scalar or functional response variable (function-on-scalar, scalar-on-function and function-on-function models).

Chapter 6. Functional Generalised Linear Models.

Chapter 7. Analysis of functional time series and the ftsa package.

Chapter 8. Further problems.

### Evaluation criteria :

The final grade will be determined by three criteria: Homework (20 %), Project (35%), Final exam (45%)

### References :

- RAMSAY, J.O. and SILVERMAN, B. W. Functional Data Analysis. Springer. 2005.
- RAMSAY, J.O., HOOKER, G. and GRAVES, S. Functional Data Analysis in R and Matlab. Springer. 2009.
- SHI, J. Q. and CHOI, T. Gaussian Process Regression Analysis for Functional Data. Chapman & Hall/CRC Press. 2011.
- HORMANN, S. and KIDZINSKI, L., HALLIN, M. Dynamic Functional Principal Components. JRSSB, Vol. 77, No. 2, pp. 319-348. arXiv 1210.7192v5. 2015.
- SHANG, H. L. ftsa: An R package for analysing functional time series. The R journal, 64-72. 2013.
- HORVATH, L. and KOKOSZKA, P. Inference for Functional Data with Applications. Springer Series in Statistics, Volume XIV. 2012.

7. KOKOSZKA, P and REIMHER, M. Introduction to Functional Data Analysis. Chapman & Hall/CRC, Texts in Statistical Science. 2017.

UE 3 - Statistique et signal

## FILTRAGE

### Filtering

<i>Enseignant</i>	: Salima El Kolei (Ensaï)
<i>Nombre d'ECTS</i>	: 1
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 18h
<i>Répartition des enseignements</i>	: 9h de cours, 3h de TP
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: R
<i>Documents pédagogiques</i>	: Notes de cours
<i>Pré-requis</i>	: chaînes de Markov

### Modalités d'évaluation :

TP noté

### Acquis d'apprentissage (objectifs) :

- Maîtriser les fondements théoriques du filtre de Kalman et ses extensions.
- Comprendre les algorithmes de filtrage linéaire et non linéaires qui en découlent
- Mobiliser ses acquis afin d'illustrer ses utilisations à travers différentes applications en ingénierie.

### Principales notions abordées :

1 : Estimation bayésienne : importance et prise en compte d'une information a priori, estimateur du minimum de variance et loi conditionnelle, conditionnement dans les vecteurs aléatoires gaussiens.

2 : Systèmes linéaires gaussiens : filtre et lisseur de Kalman.

3 : Systèmes non-linéaires : filtres sous-optimaux obtenus par linéarisation (filtre de Kalman étendu) ou par approximation gaussienne.

4 : Modèles d'état généraux : filtre bayésien et approximation Monte Carlo (filtre particulaire).

5 : Illustration de la problématique et mise en œuvre des algorithmes.

### Références bibliographiques :

- A. H. JAZWINSKI, *Stochastic processes and filtering theory*, Academic Press, 1970.
- B. D. O. ANDERSON, J. B. MOORE, *Optimal filtering*, Prentice-Hall, 1979.
- O. CAPPÉ, É. MOULINES, T. RYDEN, *Inference in hidden Markov models*, Springer, 2005.
- J. DURBIN, S. J. KOOPMAN, *Time series analysis by state space methods*, 2nd edition, Oxford University Press, 2012.
- R. DOUC, É. MOULINES, D. S. STOFFER, *Nonlinear time series: theory, methods and applications with R examples*, CRC Press, 2014.
- R. H. SHUMWAY, D. S. STOFFER, *Time series analysis and its applications, with R examples*, 4th edition, Springer, 2017.

UE 4 - Processus et prévision

## STATISTIQUE DES PROCESSUS

### *Statistic of Stochastic Processes*

<i>Enseignant</i>	: Lionel Truquet (ENSAI)
<i>Nombre d'ECTS</i>	: 2
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 46h
<i>Répartition des enseignements</i>	: 21h de cours et 6h d'atelier
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: R (coda, bps, spline, HiddenMarkov, depmixS4)
<i>Documents pédagogiques</i>	: Polycopié
<i>Pré-requis</i>	: Probabilité, statistique inférentielle, SAS, R (1A) Chaîne de Markov, Calcul Bayésien, Martingales, Méthodes de rééchantillonnage (2A)

Un processus stochastique est un ensemble de variables aléatoires indexées par le temps ou l'espace. La modélisation stochastique est utilisée pour la modélisation de phénomènes temporels ou spatiaux dans de nombreux domaines comme la biologie, la fiabilité ou la finance. Ce cours a pour objectif d'étudier les propriétés théoriques de différents types de processus stochastiques, notamment les chaînes de Markov, les processus de Poisson et les processus de Markov à sauts. Ce cours sera également illustré par des des exemples de modélisation et d'inférence statistique.

*Enseignement destiné aux élèves de la filière « Data Science, Modélisation économique & Santé » et « Data Science en Santé & Biostatistique ».*

### Principales notions abordées :

Processus Stochastiques - Inférence paramétrique  
 Processus de Renouvellement  
 Processus de Poisson - Points de rupture  
 Chaîne de Markov - Etats absorbants - Modèle de Markov Caché  
 Processus de Markov à sauts – Modèles phylogénétiques  
 Files d'attente (*uniquement pour « Génie Statistique »*)  
 Aperçu des chaînes de Markov inhomogène (*uniquement pour « Génie Statistique »*)

### Acquis d'apprentissage (objectifs) :

- Définir les processus stochastiques suivants chaîne de Markov, processus de Poisson, processus de Markov à sauts)
- Effectuer des calculs de base pour les processus de Poisson et les chaînes de Markov, y compris les modèles de files d'attente simples.
- Décrire des modèles statistiques faisant intervenir des processus stochastiques
- Développer les méthodes d'inférence fréquentistes ou bayésiennes dérivées des modèles de processus stochastiques.

### Modalités d'évaluation :

Examen écrit

## Références bibliographiques :

- ASMUSSEN, S. Applied probability and queues. Second edition. Springer 2003.
- BOSQ, D. Statistique mathématique et statistique des processus. Lavoisier, 2012.
- DELMAS, J-F., JOURDAIN, B. Modèles aléatoires. Applications aux sciences de l'ingénieur et du vivant. Springer 2006.
- FOATA, D., FUCHS, A. Processus Stochastiques (2<sup>e</sup> éd.). Dunod 2004.
- PARDOUX, E. Processus de Markov et applications. Algorithmes, génome et finance. Dunod 2007.

UE 4 - Processus et prévision

## SERIES TEMPORELLES AVANCEES

### *Advanced Times Series*

<i>Enseignant</i>	: Vincent Lefieux (RTE, UPMC)
<i>Nombre d'ECTS</i>	: 2
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 36h
<i>Répartition des enseignements</i>	: 3h de cours, 21h d'ateliers
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: R/ SAS
<i>Documents pédagogiques</i>	: Slides/Manuscrit
<i>Pré-requis</i>	: cours de séries temporelles en 2A

### Modalités d'évaluation :

Projet

### Acquis d'apprentissage (objectifs) :

- Maîtriser les techniques d'analyse des séries temporelles multivariées couramment utilisées dans les applications
- Savoir utiliser les procédures d'estimation dans le cadre non stationnaire (présence de racines unités et cointégration)
- Mettre en œuvre les méthodes d'estimation et de test présentées dans le cours sur données réelles

### Principales notions abordées :

1 : Processus stationnaire (rappels et extensions) : Estimation et tests processus ARMA, SARIMA, GARCH, EGARCH...

2 : Processus non stationnaire : estimation et tests processus à racine unité et estimation des relations de cointégration.

### Références bibliographiques :

- BANERJEE, A., DOLADO, J.J., GALBRAITH, J.W., & HENDRY, D.F., 1993, *Co-integration, Error Correction, and the Econometric Analysis of non-stationary data*, Oxford University Press, New-York.
- BROCKWELL, P.J., & DAVIS, R.A., 1998, *Time Series : Theory and Methods*, Springer Series in Statistics, Springer Verlag, Second Edition.
- BROCKWELL, P.J., & DAVIS, R.A., 2002, *Introduction to Time Series and Forecasting (2<sup>nd</sup> ed.)*, Springer-Verlag, New-York.
- BOURBONNAIS, R., 2011, *Econométrie*, 8<sup>ème</sup> ed, Dunod.
- FRANCO, C., & ZAKOIAN, J., 2009, *Modèles GARCH*, Economica.
- GOURIEROUX, G., 1992, *Modèles ARCH et applications financières*, Economica.
- GOURIEROUX, G., & MONFORT, A., 1995, *Séries temporelles et modèles dynamiques*, 2<sup>ème</sup> ed., Economica.
- HAMILTON, J.D., 1994, *Time Series Analysis*, Princeton Univ. Press.
- MADDALA, G.S., & KIM, I.M., 1998, *Unit Roots, Cointegration and Structural Change*, Cambridge University Press.
- MIGNON, V., LARDIC, S., 2002, *Econométrie des séries temporelles macroéconomiques et financières*, Economica.
- PFAFF, B., 2008, *Analysis of Integrated and Cointegrated Time Series with R*, 2<sup>nd</sup> ed., Springer.
- TSAY, R.S., 2010, *Analysis of Financial Time Series*, 3<sup>rd</sup> edition, Wiley.

UE 5 – Qualité et Fiabilité

## PLAN D'EXPERIENCES

### *Experimental design*

<i>Enseignant</i>	: Walter Tinsson (Université de PAU)
<i>Nombre d'ECTS</i>	: 1
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 27h
<i>Répartition des enseignements</i>	: 18h de cours,
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: Sans objet
<i>Documents pédagogiques</i>	: Slides/Manuscrit
<i>Pré-requis</i>	: Calcul matriciel, Optimisation, Bases de la régression et d'analyse de variance...

### Modalités d'évaluation :

Examen écrit

### Acquis d'apprentissage (objectifs) :

- Comprendre les principes fondateurs des stratégies d'expérimentation
- Apprendre à choisir, construire un dispositif expérimental
- Acquérir les outils d'analyse des plans d'expériences (utilisation du logiciel R)

### Principales notions abordées :

Principes fondateurs et présentation des grandes familles de plans  
 Les Outils d'Analyse : modèle linéaire, analyse de la variance  
 Plans factoriels complets et fractionnaires, Optimalité  
 Expériences Accélérées

### Références bibliographiques :

- AZAIS, J.-M., BARDET, J.-M. Le modèle linéaire par l'exemple (2<sup>e</sup> éd.). Dunod, 2012.
- DROESBEKE, J.-J., FINE, J., SAPORTA, G. (Eds Scientifiques). Plans d'expériences : Applications à l'entreprise. Technip, 1997.

UE 5 – Qualité et Fiabilité

## FIABILITE

### *Reliability*

<i>Enseignant</i>	: Nikolaos Limnios (UTC Compiègne)
<i>Nombre d'ECTS</i>	: 2
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 39h
<i>Répartition des enseignements</i>	: 21h de cours, 6h de TD
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: R
<i>Documents pédagogiques</i>	: photocopié du cours
<i>Pré-requis</i>	: statistique des processus, chaîne de Markov

### Modalités d'évaluation :

Examen écrit

### Acquis d'apprentissage (objectifs) :

- Maîtriser la gamme des modèles probabilistes usuels de fiabilité
- Savoir utiliser les procédures d'estimation statistique, en passant du cas plus simple de temps de défaillances indépendants et identiquement distribués au cas plus complexe de temps de défaillances dépendants.
- Concevoir et améliorer la politique de maintenance

### Principales notions abordées :

Dans ce cours il sera introduit dans un premier temps les processus de renouvellement et leur estimation puis sera abordé les processus de Markov dans le cadre de la fiabilité à temps continu. Enfin, des cas plus généraux de processus semi-markoviens seront présentés en détail et illustrés à travers des exemples en sûreté de fonctionnement.

1 : Introduction à la fiabilité

2 : Fiabilité en temps discret : chaînes de Markov

3 : Fiabilité en temps continu : Processus de Markov

4 : Fiabilité en temps continu : Processus Semi-Markoviens

### Références bibliographiques :

Cf. photocopié du cours.



UE Stages

## UE : PROJET PROFESSIONNEL ET STAGES

<i>Correspondant de l'UE</i>	: Patrick Gandubert
<i>Nombre d'ECTS</i>	: 30
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: Travail en entreprise
<i>Nombre d'heures d'enseignement</i>	: 30h (séminaires)

### Finalité de l'UE :

Cette UE correspond à des temps pédagogiques en lien direct avec les entreprises. Les séminaires professionnels ont pour objectif de présenter aux étudiants diverses problématiques auxquelles ils seront confrontés dans leur environnement professionnel. Il permet d'apporter des compléments par rapport à certains cours, et fait le lien entre les enseignements et les applications pratiques qui en découlent. Le projet professionnel permet de préparer les étudiants à leur entrée dans la vie professionnelle et aux stages, il est réalisé sur la 2ème et 3ème année de formation. Des simulations d'entretien de recrutement sont organisées en 3e année. Elles sont assurées par des recruteurs d'entreprises et d'organisations partenaires de l'Ensaï. Les stages (application en 2ème année, fin d'études en 3ème année) permettent aux élèves de mettre en pratique les enseignements de mathématiques appliquées, d'informatique et d'économie dans un cadre professionnel. Le stage de fin d'études, d'une durée de 20 semaines minimum, vise à appliquer les enseignements de 3ème année et à acquérir de l'expérience pour assurer la transition vers l'emploi. Il constitue une étape essentielle de mise en situation professionnelle pour le futur ingénieur qui dispose à ce stade de l'ensemble des bagages techniques de la formation.

### Structuration de l'UE :

Le stage de fin d'études constitue la majeure partie de l'évaluation de cette UE (25 ECTS). L'Ensaï exige une forte adéquation entre le contenu du stage et la filière de spécialisation de 3e année. Il fait l'objet d'une procédure de validation par le responsable de filière et par le département des relations avec les entreprises. L'évaluation tient compte de la capacité d'intégration de l'étudiant dans l'entreprise, ses capacités d'initiative et de satisfaction au regard des objectifs du stage, et de la qualité du rapport et de la soutenance réalisée devant un jury composé d'un président, d'un vice-président, tous les deux issus du monde de l'entreprise et d'un permanent de l'école. Le stage d'application de 2ème année est pris en compte dans cette UE (5 ECTS). Les séminaires professionnels ne sont pas évalués.

### Compétences ou acquis d'apprentissage à l'issue de l'UE :

Le stage de fin d'études (et l'UE) comprend un objectif technique - il s'agit de répondre à la commande, à la problématique inscrite dans le thème du stage à l'aide des connaissances acquises - et un objectif professionnel - il s'agit de parfaire la connaissance du monde du travail, de développer des capacités relationnelles et d'adopter une démarche d'insertion dans le monde professionnel.

### Les pré-requis de l'UE :

Aucun

UE Projet Professionnel

## DROIT DU TRAVAIL

### *Work Law*

<i>Enseignant</i>	: Charlotte GRUNDMAN, Avocat au Barreau de Paris
<i>Nombre d'ECTS</i>	: 0
<i>Volume horaire de travail élève (enseignements + travail personnel)</i>	: 15h
<i>Répartition des enseignements</i>	: Cours : 3h – Atelier : 6h
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: Sans objet
<i>Documents pédagogiques</i>	: Distribués pendant le cours
<i>Pré-requis</i>	: Aucun

### Modalités d'évaluation :

Exposé d'un cas pratique réalisé lors des TD.

### Acquis d'apprentissage (objectifs) :

La matière étant extrêmement vaste et complexe, il est ici proposé aux étudiants une approche didactique et vivante du sujet, l'objectif de l'enseignement étant de permettre aux étudiants qui travailleront dans un futur proche en entreprise d'avoir compris certaines notions pratiques essentielles en droit du travail.

### Principales notions abordées :

Hormis le cours d'amphi, il sera systématiquement proposé aux étudiants, après l'étude d'une notion, un exercice visant à mettre en pratique la notion abordée. Le cours commun (3 heures) traite des notions suivantes : Comprendre d'où l'on vient pour savoir où on va (introduction historique au droit du travail, les sources du droit du travail, ordre public absolu et ordre public social), les instances de contrôle du droit du travail, formation et exécution du contrat de travail, la rupture du contrat à durée indéterminée. Pour les TD, la première heure de cours sera consacrée à l'étude d'un chapitre (la modification du contrat de travail, le recrutement, les droits fondamentaux du salarié). Cet exposé sera suivi d'une mise en situation pratique, où les étudiants devront par groupe répondre à un cas pratique. Un rapporteur sera désigné par groupe, et la notation se fera à cette occasion.