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

# Programme des enseignements de 3e année Filière GS

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

Campus de Ker Lann, 51 rue Blaise
Tél : 33 (0)2 99 05 32 32 / scolarite@

www.ensai.fr

# FILIÈRE GÉNIE STATISTIQUE

## ANNÉE SCOLAIRE 2020/2021

## *ADVANCED STATISTICAL ENGINEERING SPECIALIZATION*

### *2020/2021 ACADEMIC YEAR*

Campus de Ker Lann, 51 rue Blaise Pascal - BP37203 - 35172 BRUZ CEDEX
Tél : 33 (0)2 99 05 32 32 / scolarite@ensai.fr

www.ensai.fr

# Présentation de la filière

La formation d'ingénieur de l'Ensai 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 maitrise 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 particulaire, 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, SAS) 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'Ensai 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'Ensai 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 sureté et fonctionnement, ingénieur qualité, consultant statisticien au sein de bureaux d'études de grandes entreprises et d'organismes publics, ...

| | Volume horaire | | | | |
|---|---|---|---|---|---|
| | **Cours** | **Ateliers** | **Projets** | **Total** | **Crédits** |
| **UE0 Tronc commun** | | | | | |
| Droit du Travail | 3 | 6 | | 9 | 0,5 |
| Anglais | 30 | | | 30 | 1,5 |
| Sport | | 30 | | 30 | 0 |
| **Total** | **33** | **36** | | **69** | **2** |
| **UE1 Machine learning** | | | | | |
| Machine learning | 21 | 30 | | 51 | 3 |
| Régression pénalisée et sélection de modèles | 9 | 6 | | 15 | 1 |
| Apprentissage statistique à grande échelle | 6 | 9 | | 15 | 1,5 |
| Webmining et traitement du langage | 9 | 12 | | 21 | 1,5 |
| **Total** | **45** | **57** | | **102** | **7** |
| **UE2 Modélisation** | | | | | |
| Modélisation non linéaire | 15 | 6 | | 21 | 2 |
| Théorie des valeurs extrêmes | 12,5 | 4 | | 16,5 | 1 |
| Statistique bayésienne | 9 | 6 | | 15 | 1 |
| **Total** | **36,5** | **16** | | **52,5** | **4** |
| **UE3 Traitement de l'image** | | | | | |
| Modèles markoviens en analyse d'images | 30 | 15 | | 45 | 3 |
| Filtrage linéaire et non linéaire | 9 | 3 | | 12 | 1 |
| **Total** | **39** | **18** | | **57** | **4** |
| **UE4 Processus et prévision** | | | | | |
| Statistique des processus | 21 | 6 | | 27 | 2 |
| Géostatistique | 15 | | | 15 | 1 |
| Séries temporelles avancées | 3 | 21 | | 24 | 2 |
| **Total** | **39** | **27** | | **66** | **5** |
| **UE5 Qualité/Fiabilité** | | | | | |
| Plans d'expériences | 18 | | | 18 | 1 |
| Fiabilité | 21 | 6 | | 27 | 2 |
| **Total** | **39** | **6** | | **45** | **3** |
| **UE Projet de fin d'étude** | | | | | |
| Projet méthodologique | | 9 | 27 | 36 | 2,5 |
| Projet de fin d'étude | | 9 | 27 | 36 | 2,5 |
| Data Challenge | | 12 | | 12 | 0 |
| **Total** | | **30** | **54** | **84** | **5** |
| **UE Projet professionnel et stages** | | | | | |
| Stage de fin d'études | | | | | 25 |
| Stage d'application | | | | | 5 |
| Séminaires et projet professionnels | 30 | | | 30 | 0 |
| **Total** | **30** | | | **30** | **30** |
| **TOTAL** | **261,5** | **190** | **54** | **505,5** | **60** |

Campus de Ker Lann, 51 rue Blaise Pascal - BP37203 - 35172 BRUZ CEDEX
Tél : 33 (0)2 99 05 32 32 / scolarite@ensai.fr

www.ensai.fr

# Descriptifs des enseignements communs

Campus de Ker Lann, 51 rue Blaise Pascal - BP37203 - 35172 BRUZ CEDEX
Tél : 33 (0)2 99 05 32 32 / scolarite@ensai.fr

www.ensai.fr

UE 0

# UE : COURS D'OUVERTURE

Correspondant de l'UE : Ronan Le Saout
*Nombre d'ECTS*
*Volume horaire de travail élève*
*(enseignements + travail personnel)*
Nombre d'heures d'enseignement : 2
: Entre 50h et 60h

: 39h

## Finalité de l'UE :

À la fin de cette UE, notamment grâce à l'enseignement de l'anglais, les élèves seront capables de mettre en œuvre les compétences linguistiques et culturelles qui facilitent la suivie des cours scientifiques dispensés en anglais ou d'autres langues, le travail dans un environnement professionnel international et la compréhension des normes culturelles dans les pays étrangers. À travers le cours de droit du travail, les élèves acquerront les connaissances dans une discipline autre que la statistique, l'économie et l'informatique nécessaires pour mieux appréhender le contexte juridique de l'entreprise. Cette UE vise également le développement des compétences transversales (*soft skills)* qui aideront les élèves à réussir les projets académiques de leur formation, à intégrer le marché du travail et à devenir les citoyens éclairés.

## Structuration de l'UE :

L'UE 0 de la 3ème année se compose de deux matières obligatoires, l'anglais et le droit du travail, ainsi que le sport de manière optionnelle.

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

En anglais les élèves travaillent toutes les compétences linguistiques pour atteindre le niveau B2 du CECR et progresser vers un niveau C1. Lors des cours d'anglais et de l'aide au projet en anglais, les élèves développement également les connaissances liées au monde de l'entreprise et de la recherche ainsi que les compétences transversales *(soft skills).* Le cours du droit du travail permet aux étudiants d'identifier et comprendre certaines notions pratiques essentielles en gestion des ressources humaines en entreprise.

## Les pré-requis de l'UE :

Aucun

Campus de Ker Lann, 51 rue Blaise Pascal - BP37203 - 35172 BRUZ CEDEX
Tél : 33 (0)2 99 05 32 32 / scolarite@ensai.fr

www.ensai.fr

UE 0 Cours d'ouverture

# ANGLAIS

*English*

Enseignant : Divers intervenants (correspondant : Todd Donahue)
*Nombre d'ECTS*
*Volume horaire de travail élève*
*(enseignements + travail personnel)*
*Répartition des enseignements*
Langue d'enseignement : 1
: 40h

: 15h de cours, 15h d'aide au projet de fin d'études
: Anglais
Logiciels : Sans objet
Documents pédagogiques : Sous Moodle
Pré-requis : 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 « Negociating 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.

UE Cours d'ouverture

# DROIT DU TRAVAIL

*Work Law*

Enseignant : Charlotte GRUNDMAN, Avocat au Barreau de Paris
*Nombre d'ECTS*
*Volume horaire de travail élève*
*(enseignements + travail personnel)*

*Répartition des enseignements*
Langue d'enseignement : 1
: 15h

: Cours : 3h • Atelier : 6h
: Français
Logiciels : Sans objet
Documents pédagogiques : Distribués pendant le cours
Pré-requis : 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.

UE Cours d'ouverture

# SPORT

*Sport*

Enseignant : Divers intervenants (correspondant : Jullien Lepage)
*Nombre d'ECTS*
*Volume horaire de travail élève*
(enseignements + travail personnel) : 0
: 30

## Modalités d'évaluation :

La participation à une activité sportive peut donner lieu à l'attribution d'un bonus ajouté sur la moyenne du semestre concerné. Le niveau de ce bonus est précisé dans une circulaire d'application en début d'année académique. Il varie selon l'assiduité aux séances, l'engagement et la participation aux compétitions tout au long de l'année. Pour être définitive, la liste des élèves bénéficiant de ces bonus doit être validée par le directeur des études.

Un bonus peut être exceptionnellement attribué en dehors des activités sportives réalisées dans le cadre Ensai. Pour y prétendre, les élèves concernés doivent remplir les 3 conditions suivantes:
- pratiquer régulièrement une activité sportive et participer aux compétitions liées ;
- posséder un niveau national (voir très bon niveau régional suivant le sport en question) ;
- déposer une demande argumentée auprès de la direction des études et du service sport en début d'année scolaire, afin de faire valider le programme d'entraînement, des compétitions et les modalités de diffusion des performances. Pour certains ayant des contraintes sportives, des aménagements horaires pourront d'ailleurs être ainsi envisagés si besoin.

## Acquis d'apprentissage (objectifs) :

L'objectif est d'amener les élèves à maintenir un esprit sportif, sortir du strict cadre académique et développer leurs capacités physiques.

## Principales notions abordées :

Neuf activités sportives sont proposées par l'école : Badminton, Basket, Football, Hand-ball, Tennis de table, Tennis débutant, Volley-ball, Cross-training, Course à pied/préparation physique/coaching sportif. Outre les entraînements, les élèves inscrits peuvent être amenés à participer à des compétitions.

UE 1

# *UE : APPRENTISSAGE AUTOMATIQUE (MACHINE LEARNING)*

Correspondant de l'UE : Arthur Katossky
*Nombre d'ECTS*

*Volume horaire de travail élève*
*(enseignements + travail personnel)*
Nombre d'heures d'enseignement : 6 pour les filières ISTS, GDR et SSV, 7 pour GS et MQRM, 8 pour SID
: De 25h à 30h par ECTS

: 78h pour les filières ISTS, GDR et SSV, 102h pour GS et MQRM, 120h pour SID

## 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 automatique 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 4 matières : apprentissage automatique (Machine-learning), régression pénalisées et régularisation, apprentissage statistique à grande échelle, 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…). Selon les filières de spécialisation, des séminaires complémentaires (systèmes de recommandation…) sont introduits.

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

Cette UE permet de maitriser 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, s'intégrant dans une architecture informatique de données volumineuses (big data).

## Les pré-requis de l'UE :

Modélisation statistique de 2ème année, méthodes d'optimisation et d'algorithmique, panorama du big data, aisance en R et Python.

Campus de Ker Lann, 51 rue Blaise Pascal - BP37203 - 35172 BRUZ CEDEX
Tél : 33 (0)2 99 05 32 32 / scolarite@ensai.fr

www.ensai.fr

UE Machine Learning

# APPRENTISSAGE AUTOMATIQUE

*Machine Learning*

Enseignant : Hong-Phuong DANG (Ensai), Romaric GAUDEL (Ensai), Fabien NAVARRO (Ensai) et Brigitte GELEIN (Ensai)
*Nombre d'ECTS*
*Volume horaire de travail élève*
*(enseignements + travail personnel)*

*Répartition des enseignements*

Langue d'enseignement : 2 (ISTS, GDR, SSV), 3 (GS, MQRM) ou 4 (SID)
: De 25h à 30h par ECTS

: Pour les 27h en filières ISTS, GDR et SSV, il y a 15h de cours et 12h d'ateliers. Pour les 24h complémentaires en GS et MQRM, il y a 6h de cours et 12h d'ateliers. Pour les 36h complémentaires en SID, il y a 12h de cours et 18h d'ateliers.
: Français
Logiciels : R et Python
Documents pédagogiques : supports de cours, bibliographie et fiches de TP
Pré-requis : R, Python, modélisation statistique, algèbre linéaire, optimisation de fonctions

## Modalités d'évaluation :

Contrôle continu à la discrétion des intervenants, un QCM, compte-rendus de TP (1 en ISTS, GDR et SSV, 4 en GS et MQRM, 4 en SID)

## Acquis d'apprentissage (objectifs) :

- Identifier comment résoudre une tâche par apprentissage automatique
- Choisir un modèle a priori adapté à une tâche
- Utiliser un modèle de l'état de l'art (SVM, réseau de neurones, forêt, …)
- Comparer empiriquement différents modèles pour une tâche donnée

## Principales notions abordées :

Un rappel des principes de l'apprentissage statistique et automatique sera effectué. L'ensemble des filières aborderont les réseaux de neurones (y compris deep learning), les méthodes d'agrégation (forêts aléatoires, bagging, boosting, stacking) et les séparateurs à vaste marge (Support Vector Machines). Les réseaux de neurones avancés seront abordés en GS, MQRM et SID, les systèmes de recommandation en MQRM et SID.

## Références bibliographiques :

Campus de Ker Lann, 51 rue Blaise Pascal - BP37203 - 35172 BRUZ CEDEX
Tél : 33 (0)2 99 05 32 32 / scolarite@ensai.fr

www.ensai.fr

- Endrew Ng. Machine Learning Yearning. Disponible gratuitement au lien https://www.deeplearning.ai/machine-learning-yearning/.
- Rémi Gilleron. Apprentissage machine - Clé de l'intelligence artificielle - Une introduction pour non-spécialistes. Ellipses.
- Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. 2016

UE Machine Learning

# REGRESSION PENALISEE ET SELECTION DE MODELES

*Penalized problems and model selection*

Enseignant : Cédric HERZET (INRIA) & Clément ELVIRA (INRIA)
*Nombre d'ECTS*
*Volume horaire de travail élève*
*(enseignements + travail personnel)*

*Répartition des enseignements*
Langue d'enseignement : 1
: 30h

: Cours : 9h • Atelier : 6h
: Français
Logiciels : Python
Documents pédagogiques : Supports de cours, Supports de TP, Bibliographie
Pré-requis : 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. Birkhaüser, 2013.
- D. P. Bertsekas. Nonlinear Programming. Athena Scientific, USA, 2003.

UE Machine Learning

# APPRENTISSAGE A GRANDE ECHELLE

*Large scale machine-learning*

Enseignant : Romaric Gaudel (ENSAI)
*Nombre d'ECTS*
*Volume horaire de travail élève*
*(enseignements + travail personnel)*

*Répartition des enseignements*
Langue d'enseignement : 1,5
: 25h

: Cours : 6h • Atelier : 9h
: Français
Logiciels : Python
Documents pédagogiques : support de cours, fiches de TP
Pré-requis : panorama du big data, machine-learning, 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 : être en mesure de choisir des approches appropriées pour un problème donné. Nécessite de
- objectif : 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
- 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 Machine Learning

# TRAITEMENT AUTOMATIQUE DU LANGAGE ET FOUILLE DU WEB

*Webmining et NLP*

Enseignant : Guillaume Gravier (Irisa)
*Nombre d'ECTS*
*Volume horaire de travail élève*
*(enseignements + travail personnel)*

*Répartition des enseignements*
Langue d'enseignement : 1,5 sauf pour filière informatique (SID) : 2,5
: 38h sauf pour filière informatique (SID) : 56h

: Cours : 9h • Atelier : 12h (6h complémentaires en SID)
: Français
Logiciels : Python
Documents pédagogiques : Support de cours, Supports de TP
Pré-requis : Python, Machine Learning

## Modalités d'évaluation :

Projet (de taille plus importante en SID)

## 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/nnlp.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 Projets

# UE : PROJETS

Correspondant de l'UE : Arthur Katossky
*Nombre d'ECTS*
*Volume horaire de travail élève*
*(enseignements + travail personnel)*
Nombre d'heures d'enseignement : 5
: Entre 120h et 150h

: 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 connaissance acquises en statistique, en informatique et en économie, dans une démarche de résolution de problèmes concrets type ingénieur.
Les projets de déclinent en deux versions: le projet académique, en langue anglaise, vise à approfondir une thématique centrée autour d'un ou plusieurs articles scientifiques ; le projet de fin d'études, 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 rigoureuse, à la pointe de la recherche scientifique, en langue anglaise ; constitue la partie théorique de recherche d'information dans une démarche de recherche et développement.
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'Ensai, 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 1ère et 2ème année.

Campus de Ker Lann, 51 rue Blaise Pascal - BP37203 - 35172 BRUZ CEDEX
Tél : 33 (0)2 99 05 32 32 / scolarite@ensai.fr

www.ensai.fr

Campus de Ker Lann, 51 rue Blaise Pascal - BP37203 - 35172 BRUZ CEDEX
Tél : 33 (0)2 99 05 32 32 / scolarite@ensai.fr

www.ensai.fr

UE Projets

# PROJET MÉTHODOLOGIQUE

*Methodological project*

Enseignant : Divers intervenants
*Nombre d'ECTS*
*Volume horaire de travail élève*
*(enseignements + travail personnel)*

*Répartition des enseignements*
Langue d'enseignement : 2,5
: Entre 60h et 75 h

: 9h d'ateliers, et suivis réguliers
: Anglais
Logiciels : /
Documents pédagogiques : /
Pré-requis : /

## 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
- mise en œuvre d'une démarche scientifique rigoureuse
- prise de conscience des enjeux autour de la reproductibilité des résultats de recherche
- communication sur des sujets techniques

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).

UE Projets

# PROJET DE FIN D'ETUDES

*Methodological project*

Enseignant : Divers intervenants
*Nombre d'ECTS*
*Volume horaire de travail élève*
*(enseignements + travail personnel)*

*Répartition des enseignements*
Langue d'enseignement : 2,5
: Entre 60h et 75 h

: 9h d'ateliers, et suivis réguliers
: Français
Logiciels : /
Documents pédagogiques : /
Pré-requis : /

## 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.

## 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*

Enseignant : Divers intervenants industriels (correspondante : Salima El Kolei)
*Nombre d'ECTS*
*Volume horaire de travail élève*
*(enseignements + travail personnel)*

*Répartition des enseignements*
Langue d'enseignement : Pas d'attribution d'ECTS
: 2 journées

: 12h d'ateliers
: Français
Logiciels : /
Documents pédagogiques : /
Pré-requis : Méthodes de travail des projets, Compétences statistiques et informatiques de 3ème année

## Modalités d'évaluation :

Les élèves participent au data challenge proposé à l'Ensai 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 0

# UE : PROJET PROFESSIONNEL ET STAGES

Correspondant de l'UE : Patrick Gandubert
*Nombre d'ECTS*
*Volume horaire de travail élève*
*(enseignements + travail personnel)*
Nombre d'heures d'enseignement : 30
: Travail en entreprise

: 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'Ensai. 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'Ensai 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

Campus de Ker Lann, 51 rue Blaise Pascal - BP37203 - 35172 BRUZ CEDEX
Tél : 33 (0)2 99 05 32 32 / scolarite@ensai.fr

www.ensai.fr

Campus de Ker Lann, 51 rue Blaise Pascal - BP37203 - 35172 BRUZ CEDEX
Tél : 33 (0)2 99 05 32 32 / scolarite@ensai.fr

www.ensai.fr

# Descriptifs des enseignements de la filière

Campus de Ker Lann, 51 rue Blaise Pascal - BP37203 - 35172 BRUZ CEDEX
Tél : 33 (0)2 99 05 32 32 / scolarite@ensai.fr

www.ensai.fr

UE Spécifiques filière GS

# UE SPÉCIFIQUES FILIÈRE GS

Correspondant de l'UE : Salima El Kolei
*Nombre d'ECTS*
*Volume horaire de travail élève*
*(enseignements + travail personnel)*
Nombre d'heures d'enseignement : 16
: De 25 à 30h par ECTS

: 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.

Campus de Ker Lann, 51 rue Blaise Pascal - BP37203 - 35172 BRUZ CEDEX
Tél : 33 (0)2 99 05 32 32 / scolarite@ensai.fr

www.ensai.fr

UE 2– Modélisation

# MODÉLISATION NON LINEAIRE

*Non Linear modelisation*

Enseignant : Fabien Navarro (ENSAI)
*Nombre d'ECTS*
*Volume horaire de travail élève*
*(enseignements + travail personnel)*

*Répartition des enseignements*

Langue d'enseignement : 2
: 33

: 15h de cours, 6h de TP
dont 6h en autonomie

: Français
Logiciels : R/Python
Documents pédagogiques : Slides
Pré-requis : algèbre linéaire ; une connaissance de la programmation R ou Python ; régression linéaire multiple

## Modalités d'évaluation :

TP noté

## Acquis d'apprentissage (objectifs) :

- Connaitre 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 :

Campus de Ker Lann, 51 rue Blaise Pascal - BP37203 - 35172 BRUZ CEDEX
Tél : 33 (0)2 99 05 32 32 / scolarite@ensai.fr

www.ensai.fr

- 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 (2nd ed.)*. Springer, 1998.
- T. P. RYAN. *Modern Regression Methods (2nd 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 (2nd 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

# THÉORIE DES VALEURS EXTREMES

*Extreme Value Theory*

Enseignant : Gilles Stupfler (ENSAI)
*Nombre d'ECTS*
*Volume horaire de travail élève*
*(enseignements + travail personnel)*

*Répartition des enseignements*

Langue d'enseignement : 1
: 25

: 16h30 de cours
dont 4h30 en autonomie

: Français
Logiciels : R/Python
Documents pédagogiques : Notes de cours, exercices intégrés, correction des exercices, code R
Pré-requis : 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.

UE 2– Modélisation

# STATISTIQUE BAYESIENNE

*Bayesian Statistics*

Enseignant : Sophie Ancelet (IRSN)
*Nombre d'ECTS*
*Volume horaire de travail élève*
*(enseignements + travail personnel)*

*Répartition des enseignements*

Langue d'enseignement : 1
: 23

: 9h de cours, 6h TP

: Français
Logiciels : R & WINBUGS, OPENBUGS & JAGS
Documents pédagogiques : Cliquez ici pour taper du texte.
Pré-requis : Cours de calcul bayésien 2A

## Modalités d'évaluation :

Examen écrit

## Acquis d'apprentissage (objectifs) :

- Maîtriser les connaissances de base pour l'analyse de données par approche statistique bayésienne.
- Avoir une connaissance des modèles hiérarchiques, réseau bayésien et graphes d'indépendance.
- Etre capable de mettre en œuvre les méthodes enseignées pour mener des inférences bayésiennes de données, notamment à l'aide des logiciels WINBUGS, OPENBUGS et JAGS.

## Principales notions abordées :

Un rappel de cours est fait concernant les principes de la modélisation statistique bayésienne. L'accent sera mis sur l'analyse bayésienne par les méthodes de Monte Carlo par Chaînes de Markov (MCMC). Aux travers d'exemples, seront abordés les notions de graphe d'indépendance conditionnelle, réseau bayésien, convergence des Chaînes de Markov, inférence, prédiction, validation et comparaison de modèles dans un cadre bayésien. Les exemples seront traîtés sous le logiciel WINBUGS ou JAGS en salle informatique.

## Références bibliographiques :

- Collectif BIOBAYES: Albert I., **Ancelet S.**, David O., Denis J.B., Makowski D., Parent E., Soubeyrand S. **(2015)** Méthodes statistiques bayésiennes. Bases théoriques et applications en alimentation, environnement et génétique. *ELLIPSES, ISBN : 978234000501*
- Carlin, B. P. and Louis, T.A. (2009). Bayesian Methods for Data Analysis. Chapman & HALL/CRC, third edition, (535 pp.)
- Gelman, A., Carlin, J. B., Stern, H. S. and. Rubin, D. B (2004). Bayesian data analysis. Texts in Statistical Science. Chapman & HALL/CRC, second edition, (668 pp.)
- Robert, C. P. (2001). The Bayesian choice. Springer, (second edition) (604 pp.)
- Lunn, D.J., Thomas, A., Best, N. and Spiegelhalter, D. (2000). WinBUGS -- a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing, 10: 325-337.
- Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1996). Markov chain Monte Carlo in practice. Chapman and Hall, (486 pp.)


UE 3 : Traitement de l'image

# MODELES MARKOVIENS EN ANALYSE D'IMAGE

*Markovian models in image analysis*

Enseignant : Charles Kervrann (INRIA)
*Nombre d'ECTS*
*Volume horaire de travail élève*
*(enseignements + travail personnel)*

*Répartition des enseignements*

Langue d'enseignement : 3
: 69

: 30h de cours, 15h de TP

: 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) :

- Maitriser les principaux concepts et outils de la modélisation markovienne/gibbsienne, en se limitant au cas de graphes de dépendance finis

- **S**avoir 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.).
- **M**ise 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 bayesienne 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 : Traitement de l'image

# FILTRAGE

*Filtering*

Enseignant : François Le Gland (IRISA, INRIA)
*Nombre d'ECTS*
*Volume horaire de travail élève*
*(enseignements + travail personnel)*

*Répartition des enseignements*

Langue d'enseignement : 1
: 18h

: 9h de cours, 3h de TP

: Français
Logiciels : R
Documents pédagogiques : Notes de cours
Pré-requis : 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

*Process Statistics*

Enseignant : Myriam Vimond (ENSAI)
*Nombre d'ECTS*
*Volume horaire de travail élève*
*(enseignements + travail personnel)*

*Répartition des enseignements*

Langue d'enseignement : 2
: 39h

: 25h30 de cours, 1h30 de TD
dont 3h en autonomie

: Français
Logiciels : R (packages coda, HiddenMarkov, depmisS4, Matrix, bps, splines)
Documents pédagogiques : Notes de cours sous format papier
Pré-requis : Probabilité, Statistique inférentielle, Chaîne de Markov, Calcul Bayésien

## Modalités d'évaluation :

Examen écrit

## Acquis d'apprentissage (objectifs) :

- Décrire les modèles de processus stochastiques couramment utilisés (chaîne de Markov, processus de Poisson, processus de Markov à sauts)
- Être capable de faire des calculs de base pour les processus de Poisson et les processus de Markov, y compris les modèles de files d'attente simples.
- Maîtriser les méthodes d'inférence fréquentistes ou bayésiennes dérivées des modèles de processus stochastiques.
- Mettre en œuvre un modèle stochastique du cours (processus de Poisson, chaine de Markov ou modèle de markov caché) à des fins de modélisation statistique

## Principales notions abordées :

Les processus stochastiques sont utilisés pour la modélisation de phénomènes temporels ou spatiaux dans de nombreux domaines comme la biologie, la fiabilité ou la finance. L'objectif de ce cours est de présenter l'étude de trois processus stochastiques à des fins de modélisation statistique.
1. Processus de Poisson homogène et inhomogène.
2. Chaîne de Markov, Modèle de Markov Caché.
3. Processus de Markov à sauts, Files d'attente.
Une liste d'exercices non corrigés est fournie pour permettre aux étudiants de s'exercer en supplément des cours.

## Références bibliographiques :

ASMUSSEN, S. Applied probability and queues. Second edition. Springer 2003.
BOSQ, D. Statistique mathématique et statistique des processus. Lavoisier, 2012.
DACUNHA-CASTELLE, D., DUFLO, M. Probabilités et statistiques II, Problèmes à temps mobile. Masson, 1993.
DELMAS, J-F., JOURDAIN, B. Modèles aléatoires. Applications aux sciences de l'ingénieur et du vivant. Springer 2006.
FOATA, D., FUCHS,A. Processus Stochastiques (2e éd.). Dunod 2004.
FUCHS, C., Inference for Diffusion Processes, With Application in Life Sciences. Springer 2013.
KIMMEL, M., AXELROD, D. Branching processes in biology. Springer 2002.

PARDOUX, E. Processus de Markov et applications. Algorithmes, génome et finance. Dunod 2007.

UE 4– Processus et prévision

# GEOSTATISTIQUE

*Geostatistics*

Enseignant : Thomas Romary (MINES PARIS TECH)
*Nombre d'ECTS*
*Volume horaire de travail élève*
*(enseignements + travail personnel)*

*Répartition des enseignements*

Langue d'enseignement : 1
: 23h

: 15h de cours,

: Français
Logiciels : R
Documents pédagogiques : notes de cours sous format papier
Pré-requis : probabilité, statistique inférentielle, calcul bayésien

## Modalités d'évaluation :

TP noté

## Acquis d'apprentissage (objectifs) :

- Mobiliser ses connaissances de probabilité et de statistiques inférentielles pour la modélisation de phénomènes spatiaux ;
- Mettre en œuvre une analyse des données modélisées par un processus spatial ;
- Savoir utiliser les méthodes de statistiques spatiaux

## Principales notions abordées :

La géostatistique constitue une branche importante de la statistique spatiale. Les exemples les plus simples concernent les problèmes tels que l'interpolation des précipitations, les applications géologiques et minières et de façon plus complexe les applications démographiques. Dans ce cours sera abordé l'introduction à l'analyse structurale (variogramme) ; l'interpolation des données par les méthodes de krigeage et cokrigeage et enfin la simulation et estimation de tels processus sera présenté.

## Références bibliographiques :

- CHILÈS J.P. and DELFINER P. (2012) Geostatistics. Modeling spatial uncertainty (2nd ed.). Wiley.
- DIGGLE P.J. AND RIBEIRO P.J. (2007) Model-based Geostatistics. Springer.
- LANTUÉJOUL C. (2002) Geostatistical simulation ; models and algorithms. Springer.
- MATHERON G. (1970) La théorie des variables régionalisées et ses applications. Les cahiers du centre de morphologie mathématiques de Fontainebleau. Ecole des Mines de Paris (fascicule 5).
- STEIN M. L. (1999) Interpolation of Spatial Data: Some Theory for Kriging. Springer

UE 4– Processus et prévision

# SERIES TEMPORELLES AVANCEES

*Advanced Times Series*

Enseignant : Vincent Lefieux (RTE, UPMC)
*Nombre d'ECTS*
*Volume horaire de travail élève*
*(enseignements + travail personnel)*

*Répartition des enseignements*

Langue d'enseignement : 2
: 36h

: 3h de cours, 21h ateliers

: Français
Logiciels : R/ SAS
Documents pédagogiques :
Pré-requis : cours de séries temporelles en 2A

## Modalités d'évaluation :

Projet

## Acquis d'apprentissage (objectifs) :

- Maitriser 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 (2nd ed.)*, Springer-Verlag, New-York.
- BOURBONNAIS,R., 2011, *Econométrie*, 8ème ed, Dunod.
- FRANCQ, 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*, 2ème 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*, 2nd ed., Springer.
- TSAY, R.S, 2010, *Analysis of Financial Time Series*, 3rd edition, Wiley.


UE 5– Qualité et Fiabilité

# FIABILITE

*Reliability*

Enseignant : Nikolaos Limnios (UTC Compiègne)
*Nombre d'ECTS*
*Volume horaire de travail élève*
*(enseignements + travail personnel)*

*Répartition des enseignements*

Langue d'enseignement : 2
: 39h

: 21h de cours, 6h TD

: Français
Logiciels : R
Documents pédagogiques : polycopié du cours
Pré-requis : statistique des processus, chaine 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 sureté de fonctionnement.

1 : Introduction à la fiabilité
2 : Fiabilité en temps discret : chaines de Markov
3 : Fiabilité en temps continu : Processus de Markov
4 : Fiabilité en temps continu : Processus Semi-Markoviens

## Références bibliographiques :

Cf. polycopié du cours.

UE 5– Qualité et Fiabilité

# PLAN D'EXPERIENCES

*Experimental design*

Enseignant : Walter Tinsson (Université de PAU)
*Nombre d'ECTS*
*Volume horaire de travail élève*
*(enseignements + travail personnel)*

*Répartition des enseignements*

Langue d'enseignement : 1
: 27h

: 18h de cours,

: Français
Logiciels : Cliquez ici pour taper du texte.
Documents pédagogiques :
Pré-requis : 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 (2e éd.). Dunod, 2012.
- DROESBEKE, J.-J., FINE, J., SAPORTA, G. (Eds Scientifiques). Plans d'expériences : Applications à l'entreprise. Technip, 1997.

Campus de Ker Lann, 51 rue Blaise Pascal - BP37203 - 35172 BRUZ CEDEX
Tél : 33 (0)2 99 05 32 32 / scolarite@ensai.fr

www.ensai.fr