



Programme des enseignements

3^e année

Filière Génie Statistique

ANNÉE SCOLAIRE 2019 / 2020



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

FILIÈRE GÉNIE STATISTIQUE

ANNÉE SCOLAIRE 2019/2020

ADVANCED STATISTICAL ENGINEERING SPECIALIZATION

2019/2020 ACADEMIC YEAR

Table des matières

Présentation de la filière Génie statistique.....	4
Philosophie pédagogique	4
Thématiques abordées.....	4
Option Formation Par la Recherche	4
Descriptifs des enseignements communs.....	7
UE0 Tronc commun	8
Droit du Travail.....	8
Anglais	11
Sport.....	13
UE1 – Machine Learning.....	14
Machine Learning.....	14
Machine Learning – Réseaux de neurones avancés.....	16
Régression pénalisée et sélection de modèles.....	17
Apprentissage statistique à grande échelle	18
Webmining et traitement du langage	20
Descriptifs des enseignements de la filière.....	21
UE2 - Modélisation	22
Modélisation non linéaire	22
Théorie des valeurs extrêmes	23
Modèles additifs semi-paramétriques	25
Statistique Bayésienne	26
UE3 - Traitement de l'image.....	28
Modèles Markoviens en analyse d'images	28
Projet d'analyse d'images	29
Filtrage linéaire et non linéaire	30
UE4 – Processus et prévision.....	31
Statistique des processus	31
Géostatistique	32
Prévision semi-paramétrique	33
Séries temporelles avancées.....	34
UE5 – Qualité/Fiabilité	35
Plans d'expériences.....	35
Introduction à la sûreté de fonctionnement.....	36
Fiabilité.....	37
UE – Projet de fin d'études.....	38
Projet méthodologique	38
Projet de fin d'études.....	39
Data challenge.....	40
UE - Séminaires professionnels	41
Séminaires professionnels.....	41

Présentation de la filière Génie statistique

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 formation, 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 en statistiques polyvalents, 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 statistiques des étudiants.

Thématiques abordées

Les thèmes abordés incluent la qualité et la fiabilité, le traitement de l'image et du signal, ainsi que la prévision et ses applications, notamment dans le domaine de l'environnement.

Pour répondre aux exigences de la filière, en plus du tronc commun à tous les élèves de troisième année, l'enseignement est divisé en six unités d'enseignement.

En plus, tout au long de l'année, les étudiants ont à gérer en parallèle de leur scolarité deux projets. Le projet méthodologique est une initiation à la veille scientifique avec un enseignant-chercheur. Le projet de fin d'étude représente l'équivalent d'un travail d'ingénieurs sur un à deux mois.

Transversalement à ces unités d'enseignement, les applications en informatique (R, Matlab, SAS) sont omniprésentes. 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 langue anglaise n'est pas négligée puisque des enseignements donnent lieu à l'écriture de deux mémoires en anglais et deux enseignements sont dispensés dans cette langue.

Option Formation Par la Recherche

L'Ensaï offre la possibilité aux élèves de 3^{ème} 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 recherche en statistique à l'Université Rennes 1.

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. En ce qui concerne les entreprises signataires de thèses Cifre ou organismes de recherche, on peut citer Renault SA, EUROCOPTER, SNCF, EDF, GLAXOSMITHKLINE, AXA, BNP, INNOPSYS, ONERA, INRA, INRIA, CEA...

Débouchés :

Consultant Statisticien, Data Scientist, Ingénieur maîtrise des risques, chargé d'études statistiques, Ingénieur recherche et développement

Ils sont partenaires ou interviennent dans la filière :

AIR LIQUIDE, EDF, CEA, RTE, SNCF, THALES, SNECMA, ESTIA, METEO France...

Volume horaire						
	Cours	Ateliers	Projets	Total	Crédits	Enseignant-e-s
UE0 Tronc commun						
Droit du Travail	3	6		9	0.5	Charlotte GRUNDMAN
Anglais	30			30	1.5	
Sport		30		30	0	
Total	33	36	0	69	2	
UE1 Machine learning						
Machine learning	18	21		39	3	Hong-Phuong DANG, Romaric GAUDEL, Fabien NAVARRO, Brigitte GELEIN
Machine-learning – Réseaux de neurones avancés	3	9		12	0	Romaric GAUDEL
Régression pénalisée et sélection de modèles	9	6		15	1	
Apprentissage statistique à grande échelle	9	9		18	1.5	Arthur KATOSSKY, Rémi PÉPIN
Webmining et traitement du langage	9	12		21	1.5	Arthur KATOSSKY
Total	48	57	0	105	7	
UE2 Modélisation						
Modélisation non linéaire	12	6		18	1	Fabien NAVARRO
Théorie des valeurs extrêmes	16.5			16.5	1	Marian HRISTACHE
Théorie des valeurs extrêmes – Compléments		6		6	0.5	Erwann MENARD-COMMAULT
Modèle additifs semi-paramétriques	6	6		12	1	Margaux BREGERE
Statistique bayésienne	9	6		15	1	Sophie ANCELET, Eric PARENT
Total	43,5	24	0	67,5	4.5	
UE3 Traitement de l'image						
Modèles markoviens en analyse d'images	30			30	2	Charles KERVRANN
Projet d'analyse d'images		15		15	1	
Filtrage linéaire et non linéaire	9	3		12	1	François LEGLAND
Total	39	18	0	57	4	
UE4 Processus et prévision						
Statistique des processus	12			12	1	Myriam VIMOND
Statistique des processus — Compléments (GS)	9	6		15	1	Myriam VIMOND
Géostatistique	15			15	1	Thomas ROMARY
Prévision semi-paramétrique	12			12	1	Vincent LEFIEUX
Séries temporelles avancées	3	21		24	1	Vincent LEFIEUX
Total	51	27	0	78	5	
UE5 Qualité/Fiabilité						
Plans d'expériences	18			18	1.5	Walter TINSSON
Introduction à la sûreté de fonctionnement	6			6	0	Perrine BROY
Fiabilité	21	6		27	2	Nikolas LIMNIOS
Total	45	6	0	51	3.5	
Projet de fin d'étude						
Projet méthodologique		3	9	12	1	
Projet de fin d'étude		9	27	36	3	
Data Challenge		12		12	0	
Total	0	24	36	60	4	
Séminaire professionnel						
Séminaire professionnel	30			30	0	
TOTAL	289,5	192	36	517,5	30	

UE Stage 3A	Crédits 25
--------------------	-----------------------

Descriptifs des enseignements communs

UE0 Tronc commun

Droit du Travail

Work Law

Cours : 3h • Atelier : 6h

Enseignant : Charlotte GRUNDMAN, Avocat au Barreau de Paris.

Correspondant : Ronan LE SAOUT

Objectif pédagogique :

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.

A cette fin, et 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.

Afin de satisfaire le plus possible à cet objectif, il est ainsi proposé l'organisation suivante des cours :

Cours commun (3 heures) :

Chapitre 1 : Comprendre d'où l'on vient pour savoir où on va :

- Introduction historique au droit du travail
- Les sources du droit du travail
 - sources imposées,
 - sources négociées
- Ordre public absolu et ordre public social

Chapitre 2 : les instances de contrôle du droit du travail

- L'inspecteur du travail
- Les multiples juges du droit du travail
- Point sur la procédure prud'homale

Chapitre 3 : Formation et exécution du contrat de travail

- la qualification du contrat de travail : « faux artisans, faux auto-entrepreneurs et vrai salarié ».
- le contrat à durée indéterminée, norme juridique et sociale
- la période d'essai après la loi du 25 juin 2008 : définition, durée et rupture
- les principales clauses du contrat de travail :
 - la clause de mobilité
 - la clause de non-concurrence

Chapitre 4 : la rupture du contrat à durée indéterminée

- le licenciement pour motif personnel
- le licenciement pour motif économique
- la démission du salarié
- les autres modes de rupture

Les TD :

La première heure de cours sera consacrée à l'étude d'un chapitre. 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.

Chapitre 1 : La modification du contrat de travail

Modification du contrat de travail et changement des conditions de travail

- la durée du travail (focus sur le forfait-jour)
- le lieu de travail
- la rémunération

Chapitre 2 : Le recrutement

Chapitre 3 : les droits fondamentaux du salarié

- Le fait religieux en entreprise
- Vie personnelle et technologies de l'information et de la communication (TIC)
- La mise en place de moyens de contrôle via les TIC en entreprise
- Harcèlements
- Maladie et maternité du salarié

Langue d'enseignement

Français

UE0 – Tronc commun

Anglais

English

Cours : 30h (dont 15h d'aide au projet)

Enseignant : Divers intervenants

Correspondant : Todd DONAHUE

Objectif pédagogique

Les élèves qui n'ont pas passé ou qui n'ont pas réussi le TOEIC l'année dernière auront progressé dans les compétences requises – c'est à dire, la compréhension orale, la reconnaissance des erreurs, les pièges grammaticaux, et la compréhension écrite. Les autres auront acquis les compétences nécessaires pour affronter le monde professionnel. Ils auront vu les tournures qui aident à diriger et à participer à des réunions, à prendre des décisions, et à négocier. Ils se seront entraînés à faire des présentations. Ils auront rédigé un projet en anglais et préparé la soutenance de ce projet.

Contenu de la matière

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.

Pré-requis

Aucun

Contrôle des connaissances

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*. Le CV et la lettre faite pour cet exercice seront évalués et feront partie de la note finale. Le niveau acquis apparaîtra sur le Supplément au diplôme. L'objectif de la CTI[†] pour tous les élèves ingénieurs est d'atteindre le niveau B2.

* le Cadre européen commun de référence pour les langues.

† la Commission des Titres d'Ingénieur.

Références bibliographiques

- Arbogast, B., *30 Days to the TOEIC Test*, Canada: Peterson's, 2002.

- Schramper-Azar, B., *Understanding and Using English Grammar*, New York: Longman, 1999.
- Buckwalter, Elvis, et.al, *Boostez votre score au TOEIC-spécial étudiants*, Paris: Eyrolles, 2009.
- Gear, Jolene, *Cambridge Grammar and Vocabulary for the TOEIC Test*, Cambridge: Cambridge University Press, 2010.
- Lecomte, Stéphane, et. al, *La Grammaire au TOEIC et au TOEFL : Mode d'emploi*, Paris: Ophrys, 2008.
- Loughed, Lin, *Tests complets pour le nouveau TOEIC (4^{ème} ed.)*, Paris: Pearson Education France, 2008.
- MBA Center, *New TOEIC Study Book*, Paris: MBA Center Publications, 2007.

Langue d'enseignement

Anglais

Pour tout complément d'information, chaque élève peut consulter le Programme des enseignements : Langues étrangères, distribué au début de l'année académique.

UE0 – Tronc commun

Sport

Sport

Atelier : 30h

Enseignant : Divers intervenants

Correspondant : Julien LEPAGE

Cours facultatif

Objectif de la matière

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

Contenu de la matière

9 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.

Prise en compte dans la scolarité

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 Ensaï. 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.

UE1 – Machine Learning

Machine Learning

Machine Learning

Cours : 18h • Atelier : 21h

Enseignants : Hong-Phuong DANG (Ensaï), Romaric GAUDEL (Ensaï), Fabien NAVARRO (Ensaï) et Brigitte GELEIN (Ensaï)

Correspondant : Arthur KATOSSKY (Ensaï)

Objectif pédagogique :

Ce cours présente les principes de l'apprentissage automatique (Machine Learning) ainsi que les modèles les plus utilisés.

Contenu de la matière

- Principes de l'apprentissage automatique
 - Apprentissage supervisé vs. non-supervisé ; échantillon d'entraînement et de validation, overfitting, erreur de généralisation ; fonction de coût (loss function) et minimisation d'une erreur ; évaluation des méthodes non-supervisées ; méthodes vues en 2A en tant que méthodes d'apprentissage
- Réseaux de neurones
 - Principe des réseaux de neurones ; propriétés des réseaux de neurones simples ; descente de gradient ; réseaux de neurones profonds ; architectures particulières (ex: réseaux de convolution ; réseaux récurrents ; ...) ; réduction de la dimension à l'aide de réseaux de neurones (auto-encodeurs ; word2vec ; ...).
- Méthodes d'agrégation
 - Quelques rappels et approfondissements (CART, multiregression trees), Bagging, random forests, Boosting, XGBoost, Stacking (agrégation de modèles de types différents par construction d'un modèle « superviseur » qui combine au mieux les prédictions des modèles primaires.)
- Support Vector Machines
 - Classification par hyper-plan séparateur ; classifieur de marge maximale ; données non linéairement séparable et méthodes à noyau ; SVM multi-classe ; liens avec d'autres modèles (logistique, réseaux de neurones) ; descente de gradient

Compétences

- 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

Pré-requis

R, Python, algèbre linéaire, optimisation de fonctions

Contrôle des connaissances

Des TP notés + un examen final

Références bibliographiques

- Andrew 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

Langue d'enseignement

Français

UE1 – Machine Learning

Machine Learning – Réseaux de neurones avancés

Machine Learning

Cours : 3h • Atelier : 9h

Enseignant : Romaric GAUDEL (Ensaï)

Correspondant : Romaric GAUDEL (Ensaï)

Objectif pédagogique

Ce cours s'intéresse à des réseaux de neurones aux architectures plus complexes.

Contenu de la matière

Connaissances

- Réseaux de neurones pour séries temporelles
- Modèles Génératifs (auto-encodeurs (variationnels), réseaux antagonistes génératifs)

Compétences

- Mise en application des modèles étudiés

Pré-requis

Python, principes de l'apprentissage automatique, réseaux de neurones

Contrôle des connaissances

TP notés

Références bibliographiques

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

Langue d'enseignement

Français

UE1 – Machine Learning

Régression pénalisée et sélection de modèles

Penalized problems and model selection

Cours : 9h • Atelier : 6h

Enseignants : Cédric HERZET (INRIA) & Clément ELVIRA (INRIA)

Correspondant : Arthur KATOSKY (Ensaï)

Objectif pédagogique :

De nombreuses tâches d'apprentissage et de traitement du signal visent à retrouver un ensemble de grandeurs inconnues (état d'un système, modèle génératif, etc) à partir de données.

Malheureusement, dans de nombreuses situations, les données disponibles s'avèrent insuffisantes pour lever l'ambiguïté sur les quantités à inférer ou les estimer avec une précision suffisante.

Une manière classique de contourner ce problème consiste à introduire une information « a priori » sur la solution recherchée.

Plus particulièrement, dans ce cours nous montrons comment lever l'ambiguïté inhérente à certains problèmes en « pénalisant » les solutions ne présentant pas certaines caractéristiques d'intérêt.

L'objectif de ce cours est d'identifier et manipuler les méthodes de pénalisation les plus courantes.

Contenu de la matière

- Identifier la pénalisation la plus adaptée à une tâche
- Résoudre un problème d'optimisation comportant un terme de régularisation
- Régler les paramètres du modèle

Pré-requis

- Algèbre linéaire
- Optimisation convexe
- Programmation en Python

Contrôle des connaissances

TP notés + examen final

Références bibliographiques

- 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.

Langue d'enseignement

Français

UE1 – Machine Learning

Apprentissage statistique à grande échelle

Large-scale Machine-Learning

Cours : 9h • Atelier : 9h

Enseignant : Arthur KATOSSKY (Ensaï) et Rémi PEPIN (Ensaï)

Correspondant : Arthur KATOSSKY (Ensaï)

Objectif de la matière

Au cours de la dernière décennie, nous avons assisté à l'émergence d'applications numériques nécessitant de faire face à de gigantesques quantités de données, générées de plus en plus rapidement. Ces applications (surveillance de réseaux, biologie et médecine, applications financières, réseaux sociaux, etc.) nécessitent un besoin grandissant de techniques capables d'analyser et de traiter ces grandes masses d'information, avec précision et efficacité. La statistique rejoint ici les sciences du numérique, et plus précisément l'informatique répartie, pour proposer de nouvelles approches, relatives au Big Data. Les techniques et les modèles doivent prendre en compte le volume pléthorique de ces données, mais également leur génération rapide en continu (vélocité) ainsi que la diversité de leur format (variété) et la qualité de l'information (véracité), appelés communément les 4V du Big Data.

Contenu de la matière

- Les différents « v »
- Principes, avantages et inconvénients d'un système réparti
- connaître les stratégies de tolérance aux fautes (duplication des données, exécution avec erreurs)

Compétences

- Identifier l'architecture adaptée à une tâche (exécution séquentielle et/ou parallèle, exécution en mémoire et/ou en flux, exécution locale et/ou distante).
- Lancer des calculs sur une architecture Big Data (notamment, appliquer les paradigme Map-Reduce).
- exécuter des calculs volumineux – et en particulier des calculs statistiques – sur des prestataires de calcul (IaaS ou PaaS comme Amazon Web Service, Google Cloud Platform ou autre)

Pré-requis

Algorithmique.

Contrôle des connaissances

À déterminer.

Références bibliographiques

- Analyses des Big Data : quels usages, quels défis ? Note d'analyse du Commissariat général à la stratégie et la prospective
- Pirmin Lemberger, Marc Batty, Médéric Morel, Jean-Luc Raffaëlli. Big Data et machine learning - Manuel du data scientist, Dunod, 2015.
- Rudi Bruchez. Les bases de données NoSQL et le BigData : Comprendre et mettre en œuvre, Eyrolles (2015)

Langue d'enseignement

Français.

UE1 – Machine Learning

Webmining et traitement du langage

Webmining & NLP

Cours : 9h • Atelier : 12h

Enseignant : Arthur KATOSSKY (Ensaï)

Correspondant : Arthur KATOSSKY (Ensaï)

Objectif pédagogique :

Le cours de *webmining & natural language processing* répond à plusieurs objectifs :

- pratiquer la collecte de données, l'extraction d'information et l'appariement de sources
- équiper les élèves avec des outils théoriques pour l'étude des données textuelles
- faire comprendre les grandes approches qui structurent le foisonnement de modèles de la langue
- présenter des exemples concrets d'applications dans les différents domaines d'application des élèves
- donner la capacité de réaliser des tâches classiques en étude de texte: classification, analyse de sentiment, détection d'entités, etc.

Contenu de la matière

- Introduction au traitement automatique du langage (*natural language processing*)
- Grandes catégories de modèles : *bag-of-words* et *tf-idf* ; réseaux de neurones (LSTM, GRU, etc.) ; plongements de mots (word2vec, GloVe, fasttext, Elmo, BERT, etc.) ; modèles probabilistes (HMM, CRF, LDA, etc.)
- Applications : classification, analyse de sentiment, détection d'entités, etc.
- Traitement de données textuelles et extraction d'information
- Collecte de données sur le web et utilisation d'une API

Pré-requis

Apprentissage statistique (réseaux de neurones) ; apprentissage statistique à grande échelle ; statistique bayésienne ; chaînes de Markov

Contrôle des connaissances

Projet

Références bibliographiques

Communiquée ultérieurement

Langue d'enseignement

Français.

Descriptifs des enseignements de la filière

UE2 - Modélisation

Modélisation non linéaire

Nonlinear Modeling

Cours : 12h • Atelier : 6h

Enseignant : Fabien NAVARRO (Ensay)

Correspondant : Salima EL KOLEI

Enseignement destiné aux élèves des filières « Statistique pour les sciences de la vie » et « Génie statistique »

Objectif pédagogique

Ce cours a pour but d'introduire les modèles de régression paramétrique non-linéaires et une approche non paramétriques alternative à la méthode des polynômes locaux, pour l'estimation d'une régression.

Contenu de la matière

1. Extension du modèle linéaire
 - 1.1. base de fonctions
 - 1.2. les splines
 - 1.3. les ondelettes
2. Modèle paramétrique non linéaire
 - 2.1. M-estimation
 - 2.2. intervalles de confiance et tests
3. Splines pénalisées et modèles mixtes

Pré-requis

Les cours de régression de deuxième année

Contrôle des connaissances

A déterminer

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 (2nd ed.)*. Springer, 1998.
- T. P. RYAN. *Modern Regression Methods (2nd ed.)*. Wiley, 2008

Langue d'enseignement

Français

UE2 - Modélisation

Théorie des valeurs extrêmes

Extreme Value Theory

Cours : 16h30 • Atelier : 6h

Enseignant : Marian HRISTACHE (Ensaï) et Erwann MENARD-COMMAULT

Correspondant Salima EL KOLEI

Enseignement destiné aux élèves des filières "Génie Statistique" et "Gestion des Risques et Ingénierie Financière"

Objectif pédagogique

A l'issue de cet enseignement, les élèves devront connaître les trois types de lois de valeurs extrêmes, comprendre comment on les obtient comme lois limites du maximum (minimum) d'un échantillon, savoir caractériser leur domaines d'attraction et estimer leurs paramètres.

Contenu de la matière

1. Comportement asymptotique des sommes des variables i.i.d.
 - 1.1. Lois des grands nombres
 - 1.2. Lois stables, Fonctions à variation lente à l'infini,
 - 1.3. Domaine d'attraction d'une loi stable
2. Comportement asymptotique des observations extrêmes
 - 2.1. Limite en probabilité, Convergence en loi
 - 2.2. Domaines d'attraction pour les lois max-stables, Conditions de Von Mises
 - 2.3. Comportement asymptotique des k-statistiques extrêmes
3. Estimation dans les modèles de valeurs extrêmes
 - 3.1. Modèles de valeurs extrêmes et lois de Pareto généralisées
 - 3.2. Méthode du maximum de vraisemblance et des moments pondérés
4. Estimation de l'indice de queue : Estimateur de Pickands, de Hill et de Dekker-Einmahl-De Haan

Pré-requis

Probabilités et Statistique de première année.

Contrôle des connaissances

Examen écrit

Références bibliographiques

- J. BEIRLANT, Y. GOEGEBEUR, J. SEGERS, J. TEUGELS, *Statistics of Extremes*, Wiley, 2004
- J. BEIRLANT, J. TEUGELS, P. VYNCKIER, *Practical Analysis of Extreme Values*, Leuven University Press, 1996
- E. CASTILLO, *Extreme Value Theory in Engineering*, Academic press, 1988
- P. EMBRECHTS, C. KLUPPELBERG, T. MIKOSCH, *Modelling Extremal Events for Insurance and Finance*, Springer, 1997
- J. GALAMBOS, *Advanced Probability Theory* (2nd ed.), Dekker, New York, 1995
- S. KOTZ, S. NADARAJAH, *Extreme Value Distributions. Theory and Applications*, Imperial College Press, 2000
- R-D. REISS, M. THOMAS, *Statistical Analysis of Extreme Values* (3rd ed.), Birkhäuser, 2007
- S.I. RESNICK, *Extreme Values, Regular Variation and Point Processes*, Springer, 1987

Langue d'enseignement

Français

UE2 - Modélisation

Modèles additifs semi-paramétriques

Semiparametric additive models

Cours : 6h • Atelier : 6h

Enseignant : Margaux BREGERE (EDF)

Correspondant Salima EL KOLEI

Enseignement destiné aux élèves de la filière "Génie Statistique"

Objectif pédagogique

Dans la continuité du cours de modélisation non linéaire, ce cours a pour but de présenter les méthodes de régression pénalisées, les différentes bases de spline existantes, et l'utilisation de ces outils dans des modèles semi-paramétriques additifs. Ces méthodes seront mises en pratique sur des données, en particulier des données de consommation électrique.

Contenu de la matière

- 1) Méthode de régression pénalisée : régression ridge et lasso
- 2) Les splines; présentation de l'outil splines, les différentes bases qui existent
- 3) Modèle additifs: estimation, validation
 - algorithmes d'estimation et résultats: du backfitting à mgcv
 - le future: lasso, grouped lasso
 - démarche de modélisation en pratique
- 4) Exemples d'applications: électricité

Contrôle des connaissances

Compte rendu de TP

Références bibliographiques

- P. CRAVEN, G. WAHBA, Smoothing noisy data with spline functions: estimated the correct degree of smoothing by the method of general cross validation," *Numerische Mathematik*, 31, 377-403, 1979.
- B. EFRON, L. JOHNSTONE, T. HASTIE, R. TIBSHIRANI, Least angle regression, *Ann. Stat.* 2003.
- 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
- A. PIERROT, Y. GOUDE, Short-Term Electricity Load Forecasting With Generalized Additive Models, *Proceedings of ISAP power*, 2011.
- R. TIBSHIRANI, Regression shrinkage and selection via the lasso. *J. Royal. Statist. Soc B.*, Vol. 58, No. 1, pages 267-288, 1996.
- S. WOOD, Y. GOUDE Y, S. SHAW : Generalized additive models for large datasets. preprint, 2011.
- S. WOOD, *Generalized Additive Models : An Introduction with R*. CRC/Chapman &Hall, 2006.
- S. WOOD, mgcv:GAMs and Generalized Ridge Regression for R. *R News* 1(2), *R News*, 1(2), 20-25, 2001.
- H. ZOU, The adaptive lasso and its oracle properties, *Journal of the American Statistical Association*, 101, 1418-1429, 2006.

Langue d'enseignement

Français

UE2 -Modélisation

Statistique Bayésienne

Bayesian Statistics

Cours : 9h • Atelier : 6h

Enseignant : Sophie ANCELET et Eric PARENT

Correspondant : Salima EL KOLEI

Enseignement destiné aux élèves de la filière "Génie Statistique", « Ingénierie statistique des territoires et de la santé » et « Statistique pour les sciences de la vie »

Objectif pédagogique

A l'issue de cet enseignement, les élèves devront maîtriser les connaissances de base pour l'analyse de données par approche statistique bayésienne. Les problèmes traités seront empreints aux sciences de la vie où l'emploi des méthodes bayésiennes progresse considérablement. Cependant, le caractère général de ces méthodes, utiles dans bien d'autres domaines d'application, sera clairement établi. À l'issue de ce module, l'étudiant doit être 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.

Contenu de la matière

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 traités sous le logiciel WINBUGS ou JAGS en salle informatique.

Pré-requis

Cours de statistique bayésienne en deuxième année

Contrôle des connaissances

Projet court

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.)

Langue d'enseignement

Français

UE3 - Traitement de l'image

Modèles Markoviens en analyse d'images

Markovian Models for Image Analysis

Cours : 30h

Enseignant : Charles KERVRANN (INRIA)

Correspondant : Salima EL KOLEI

Enseignement destiné aux élèves de la filière "Génie Statistique"

Objectif pédagogique

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. Le but de ce cours est d'introduire simplement les principaux concepts et outils de la modélisation markovienne/gibbsienne, en se limitant au cas de graphes de dépendance finis, et d'illustrer 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.). Ce cours est complété par un projet en MATLAB.

Contenu de la matière

1. Un avant-goût gaussien : champs markoviens gaussiens sur grilles
2. Champ de Markov sur graphe
 - 2.1. Distribution de Gibbs et graphe d'indépendance associé
 - 2.2. Lecture markovienne du graphe, Échantillonnage
 - 2.3. Estimation des paramètres sur données complètes
 - 2.4. Application à la classification de texture
3. Problèmes inverses et estimation bayésienne
 - 3.1. Problèmes inverses «mal posés», Exemples : restauration, segmentation
 - 3.2. Estimation bayésienne optimale des variables cachées
 - 3.3. Estimation des paramètres avec «données incomplètes»

Pré-requis

Les cours de chaînes de Markov et statistique exploratoire multivariée de deuxième année

Contrôle des connaissances

Examen écrit

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

Langue d'enseignement

Français

UE3 - Traitement de l'image

Projet d'analyse d'images

Image Analysis Project

Atelier : 15h

Enseignant : Yunjiao LU (INRIA)

Correspondant : Salima EL KOLEI

Enseignement destiné aux élèves de la filière "Génie Statistique"

Objectif pédagogique

Ce projet permettra aux étudiants d'aborder les applications utilisant des modèles statistiques d'images : l'analyse en composantes principales, le tirage d'échantillons, l'estimation des paramètres, la résolution d'un problème de maximum de vraisemblance. Il leur permettra de se familiariser avec des algorithmes d'optimisation stochastiques (recuit simulé) sur un problème pratique de détection de visages/symboles dans une image via un apprentissage statistique des variations photométriques observées sur une base de données-image.

Contenu de la matière

Au cours de ce projet, les étudiants devront compléter en langage MATLAB un programme de détection de visages dans des images quelconques et évaluer ces détections en fonction des paramètres d'optimisation/apprentissage.

Pré-requis

Le cours de Modèles Markoviens en analyse d'images de 3^{ème} année

Contrôle des connaissances

L'évaluation portera sur les comptes rendus de travaux pratiques.

Langue d'enseignement

Français

UE3 – Traitement de l'image

Filtrage linéaire et non linéaire

Linear and Nonlinear Filtering

Cours : 9h • Atelier : 3h

Enseignant : François LEGLAND (IRISA)

Correspondant : Salima EL KOLEI

Enseignement destiné aux élèves de la filière "Génie Statistique"

Objectif pédagogique

En toute généralité, le filtrage consiste à estimer de façon récursive un état caché au vu d'observations. Dans le cas particulier des systèmes linéaires gaussiens, il existe une solution explicite, appelée filtre de Kalman. Cet algorithme a été développé dans les années soixante en réponse à des problèmes en recherche aérospatiale, et trouve de nos jours de nombreuses applications en aéronautique, robotique, vision, télécommunications, économie, biologie, etc. Dans le cas de modèles d'état plus généraux, des méthodes de simulation Monte Carlo très efficaces sont apparues dans les années quatre-vingt-dix, sous le nom de filtres particulaires. L'objectif du cours est de présenter les fondements théoriques du filtre de Kalman et ses différentes extensions, et d'illustrer son utilisation par différentes applications en ingénierie.

Contenu de la matière

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

Systèmes linéaires gaussiens : filtre et lisseur de Kalman.

Systèmes non-linéaires : linéarisation (filtre de Kalman étendu) ou approximation gaussienne (filtre de Kalman *unscented*).

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

Illustration de la problématique et mise en œuvre des algorithmes.

Pré-requis

Le cours de chaînes de Markov de deuxième année

Contrôle des connaissances

TP noté

Références bibliographiques

- 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.
- A. GELB, ed., *Applied optimal estimation*, The MIT Press, 1974.
- A. H. JAZWINSKI, *Stochastic processes and filtering theory*, Academic Press, 1970.

Langue d'enseignement

Français

UE4 – Processus et prévision

Statistique des processus

Statistics of Stochastic Processes

Cours : 21h • Atelier : 6h

Enseignant : Myriam VIMOND (Ensaï)

Correspondant : Salima EL KOLEI

Enseignement destiné aux élèves des filières « Génie statistique » et « Statistique pour les sciences de la vie »

Objectif pédagogique

Il s'agit de présenter les modélisations des principaux phénomènes aléatoires dépendants du temps rencontrés dans l'industrie et en sciences de la vie (hors séries temporelles et traitement du signal). On se limitera aux phénomènes dont l'espace des états est discret. Dans chaque cas la présentation portera autant sur les outils probabilistes que sur l'inférence statistique dans ces modèles.

Contenu de la matière

1. Compléments sur les chaînes de Markov.
2. Processus de branchement.
3. Processus de Poisson.
4. Processus de Markov à sauts.
5. Modèles de files d'attente.

Pré-requis

Les cours de Probabilités et Statistique de première année et le cours de chaînes de Markov

Contrôle des connaissances

Examen écrit.

Références bibliographiques

- ASMUSSEN, S. Applied probability and queues. Second edition. Springer 2003.
- 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^e éd.). Dunod 2004.
- KIMMEL, M., AXELROD, D. Branching processes in biology. Springer 2002.
- PARDOUX, E. Processus de Markov et applications. Algorithmes, génome et finance. Dunod 2007.

Langue d'enseignement

Français

UE4 – Processus et prévisions

Géostatistique

Geostatistics

Cours : 15h

Enseignant : Thomas ROMARY (ENSMP)

Correspondant : Salima EL KOLEI

Enseignement destiné aux élèves de la filière "Génie Statistique"

Objectif pédagogique

Donner une introduction aux méthodes statistiques spatiales

Contenu de la matière

1. Cadre mathématiques
 - 1.1. Fonctions aléatoires d'ordre 2
 - 1.2. Fonctions aléatoires intrinsèques
 - 1.3. Régularité des fonctions aléatoires
2. Analyse structural
 - 2.1. Variographie
 - 2.2. Modèles usuels
3. Estimation
 - 3.1. Krigeage simple et ordinaire
 - 3.2. Krigeage universel
 - 3.3. Multivariable: cokrigeage
4. Simulation
5. Approche bayésienne
 - 5.1. Krigeage bayésien
 - 5.2. Problèmes inverses

Contrôle des connaissances

Examen écrit

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

Langue d'enseignement

Français

UE4 - Processus et prévision

Prévision semi-paramétrique

Semiparametric Forecasting

Cours : 12h

Enseignant : Vincent LEFIEUX (RTE)

Correspondant : Salima EL KOLEI

Enseignement destiné aux élèves de la filière "Génie Statistique"

Objectif pédagogique

Le succès des modèles paramétriques dans le cadre des séries temporelles n'est plus à démontrer. Les modèles ARMA et ARCH, entre autres, illustrent cette popularité. Néanmoins des difficultés certaines apparaissent si les données sous-jacentes ne répondent pas à la structure prédéterminée des modèles paramétriques.

Si les modèles non-paramétriques lèvent bel et bien cet à priori dans la forme du modèle, le « fléau de la dimension » les rend souvent caduques en pratique. Il existe des alternatives comme les modèles semi-paramétriques, qui permettent de réduire la dimension du problème traité, en ne considérant que des combinaisons « intelligentes » de variables explicatives (passé du processus et éventuellement variables exogènes).

Ce cours a pour but de décrire quelques-uns des résultats récents sur les modèles semi-paramétriques appliqués aux séries temporelles.

Contenu de la matière

Après une présentation synthétique des méthodes existantes, les étudiants seront amenés à analyser de manière plus fine un modèle (à partir d'un article). Une synthèse devra être rédigée et les étudiants présenteront leurs résultats par oral.

Pré-requis

Cours de séries temporelles 2^e année

Contrôle des connaissances

Présentation orale.

Références bibliographiques

- Cai Z., Fan J., Yao Q., Functional-Coefficient Regression Models for Nonlinear Time Series, Journal of the American Statistical Association, 2000, Vol. 95, No. 451
- Carbon M., Delecroix M., Non-parametric vs parametric forecasting in time series: a computational point of view, Applied Stochastic Models and Data Analysis, 1993, Vol. 9
- Li K. C., Sliced Inverse Regression for dimension reduction, Journal of the American Statistical Association, 1991, Vol. 86, No. 414
- Xia Y., An H. Z., Projection Pursuit autoregression in time series, Journal of Time Series Analysis, 1999, Vol. 20, No. 6
- Xia Y., Tong H., Li W. K., An adaptive estimation of dimension reduction space, Journal of the Royal Statistical Society, Ser. B, 2002, Vol. 64, No. 3

Langue d'enseignement

Français

UE4 - Processus et prévision

Séries temporelles avancées

Advanced Time Series

Cours : 3h • Atelier : 21h

Enseignant : Vincent LEFIEUX (RTE)

Correspondant : Salima EL KOLEI

Enseignement destiné aux élèves de la filière "Génie Statistique" et de la filière "Gestion des risques et ingénierie financière"

Objectif pédagogique

Ce cours a pour but de décrire quelques résultats « avancés » sur les modèles paramétriques de séries temporelles, en particulier la cointégration.

Contenu de la matière

Modèles VAR et cointégration

Modèles ARCH et GARCH

Pré-requis

Cours de séries temporelles 2^e année

Contrôle des connaissances

TP noté

Références bibliographiques

- Hamilton J. D., Time Series Analysis, Princeton University Press, 1994,
- Lütkepohl H., New Introduction to Multiple Time Series Analysis, Springer, 2005

Langue d'enseignement

Français

UE5 – Qualité/Fiabilité

Plans d'expériences

Experiment Design

Cours : 18h

Enseignant : Walter TINSSON (Université de PAU)

Correspondant : Lionel TRUQUET

Enseignement destiné aux élèves des filières « Statistique pour les sciences de la vie » et « Génie statistique »

Objectif pédagogique

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)

Contenu de la matière

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

Pré-requis

Cours de régression et d'Anova

Contrôle des connaissances

Examen pratique en salle informatique

Références bibliographiques

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

Langue d'enseignement

Français

UE5 – Qualité/Fiabilité

Introduction à la sûreté de fonctionnement

Risk Management and Operation Safety

Cours : 6h

Enseignant : Perrine BROY

Correspondant : Salima EL KOLEI

Enseignement destiné aux élèves de la filière "Génie Statistique"

Objectif pédagogique

Le cours a pour objectif de fournir des notions de sûreté de fonctionnement, et des méthodes pour manager des projets dans lesquels la sûreté de fonctionnement est imposée. Ces notions sont illustrées par des exemples d'application au cours d'exercices.

A la sortie du cours, les élèves doivent pouvoir être familiers avec les notions de sûreté de fonctionnement, suivre le cours de fiabilité et réaliser des projets simples de sûreté de fonctionnement.

Contenu de la matière

1. Histoire de la sûreté de fonctionnement
2. Diagrammes de fiabilité
3. Arbres de défaillances
4. Arbres d'événements
5. Réseaux bayésiens
6. Exemples d'applications à EDF

Pré-requis

Références bibliographiques

- VILLEMEUR, Sûreté de fonctionnement des systèmes, Ed. Eyrolles, 1988
- A. LANNOY, Maîtrise des risques et sûreté de fonctionnement, repères historiques et méthodologiques, Lavoisier.
- K. MURPHY, Dynamic Bayesian Networks: Representation, Inference and Learning, PhD thesis

Langue d'enseignement

Français

UE5 - Fiabilité

Fiabilité

Reliability

Cours : 21h • Atelier : 6h

Enseignant : Nikolas LIMNIOS

Correspondant : Salima EL KOLEI

Enseignement destiné aux élèves de la filière "Génie Statistique"

Objectif pédagogique

L'objectif de ce cours est de présenter aux étudiants les problèmes de base en fiabilité des systèmes, en particulier la modélisation probabiliste et l'estimation statistique, en passant du cas plus simple de temps de défaillances i.i.d. au cas plus complexe de temps de défaillances dépendants. Les processus stochastiques de base pour la modélisation et l'estimation qui seront présentés sont les processus de renouvellement, les processus de Markov et les processus semi-Markoviens qui généralisent les cas précédents. Une introduction plus détaillée sera donnée en ce qui concerne les processus semi-Markoviens.

Contenu de la matière

1. Introduction à la fiabilité (~2h)
 - 1.1. Définitions et exemples
 - 1.2. Processus de renouvellement
 - 1.3. Estimation
2. Fiabilité en temps discret – Chaîne de Markov (~5h)
 - 2.1. Introduction
 - 2.2. Chaînes de Markov
 - 2.3. Fiabilité
 - 2.4. Estimation
3. Fiabilité en temps continu - Processus de Markov (~3h)
 - 3.1. Introduction
 - 3.2. Processus de Markov
 - 3.3. Fiabilité
 - 3.4. Estimation
4. Fiabilité en temps continu – Processus semi-Markoviens (~5h)
 - 4.1. Introduction
 - 4.2. Processus semi-Markoviens
 - 4.3. Fiabilité
 - 4.4. Estimation

Pré-requis

Le cours des processus stochastiques, en particulier de chaînes et processus de Markov.

Contrôle des connaissances

Examen écrit + mini projet

Références bibliographiques

Cf. Poly du cours

Langue d'enseignement

Français

UE – Projet de fin d'études

Projet méthodologique

Methodological project

Atelier : 3h • Projet : 9h

Enseignants : Divers intervenants

Correspondant : Arthur KATOSSKY (Ensay)

Enseignement destiné à l'ensemble des élèves des six filières

Objectif pédagogique

Le projet méthodologique consiste en la production d'un article de synthèse sur un sujet de recherche à choisir parmi un catalogue. Ses objectifs sont multiples:

- familiarisation avec la forme des productions académiques
- mise en œuvre d'une démarche scientifique rigoureuse
- prise de conscience des enjeux autour de la reproductibilité des résultats de recherche
- travail en binôme
- communication sur des sujets techniques

À cela s'ajoute les objectifs spécifiques à la production d'un travail scientifique en langue anglaise (expression écrite et orale, vocabulaire spécialisé, vulgarisation, etc.).

Contenu de la matière

Travail de recherche en groupe suivi par un chercheur (env. 5 séances) et un professeur d'anglais (4 séances).

Pré-requis

Contrôle des connaissances/ Evaluation

Projet

Références bibliographiques

Selon les projets

Langue d'enseignement

Anglais

UE – Projet de fin d'études

Projet de fin d'études

End of study project

Atelier : 9h • Projet : 27h

Enseignant : Divers intervenants

Correspondant : Arthur KATOSSKY (Ensaï)

Enseignement destiné à l'ensemble des élèves des six filières

Objectif pédagogique

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. Ses objectifs 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

Contenu de la matière

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

Pré-requis

Références bibliographiques

Selon les projets

Contrôle des connaissances

Évaluation: projet avec soutenance

Langue d'enseignement

Français

UE – Projet de fin d'études

Data challenge

Data Challenge

Atelier : 12h

Enseignant : Divers intervenants industriels

Correspondant : Salima EL KOLEI

Enseignement destiné à l'ensemble des élèves des six filières

Objectif du cours électif

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 et d'évaluer les compétences transversales acquises dans ce contexte opérationnel.

Contenu de la matière

Les élèves devront participer au data challenge proposé à l'Ensaï ouvert également aux élèves de deuxième année.

Compétences acquises

- 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...)

Prérequis

- Compétences en statistiques et informatiques de 1A, 2A et 3A.
- Compétences transversales mobilisées dans les projets 1A, 2A et 3A.

UE - Séminaires professionnels

Séminaires professionnels

Professional Seminar

Cours : 30h

Enseignant : Divers intervenants

Correspondant : Salima EL KOLEI

Enseignement destiné aux élèves de la filière "Génie Statistique"

Objectif pédagogique

Le séminaire professionnel a 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. Il est assuré en général par des professionnels.

Contrôle des connaissances

Sans évaluation

Langue d'enseignement

Français