



# Programme des enseignements

3<sup>e</sup> année

**Filière Génie Statistique**

ANNEE SCOLAIRE 2018 / 2019

## **FILIÈRE GÉNIE STATISTIQUE**

**ANNÉE SCOLAIRE 2018/2019**

***ADVANCED STATISTICAL ENGINEERING SPECIALIZATION***

***2018/2019 ACADEMIC YEAR***



# Table des matières

<b>Présentation de la filière Génie statistique.....</b>	<b>4</b>
Philosophie pédagogique .....	4
Thématiques abordées.....	4
Option Formation Par la Recherche .....	4
<b>Descriptifs des enseignements du tronc commun.....</b>	<b>7</b>
UE Tronc commun .....	8
Droit du Travail.....	8
Anglais .....	10
Sport .....	12
<b>Descriptifs des enseignements de la filière.....</b>	<b>13</b>
UE - Modélisation .....	14
Modélisation non linéaire .....	14
Théorie des valeurs extrêmes .....	15
Modèles avancés d'ingénierie financière.....	16
Modèles additifs semi-paramétriques .....	17
UE - Traitement de l'image.....	18
Modèles Markoviens en analyse d'images .....	18
Projet d'analyse d'images .....	19
Filtrage linéaire et non linéaire .....	20
UE – Processus et prévision.....	21
Statistique des processus .....	21
Géostatistique .....	22
Prévision semi-paramétrique .....	23
Séries temporelles avancées .....	24
UE – Qualité/Fiabilité .....	25
Plans d'expériences.....	25
Introduction à la sûreté de fonctionnement.....	26
Fiabilité .....	27
UE - Outils d'aide à la décision .....	28
Méthodes modernes d'apprentissage .....	28
Réseaux de neurones .....	29
Statistique Bayésienne .....	30
UE – Informatique .....	32
Python .....	32
WinBUGS .....	33
UE – Projets .....	34
Projet de fin d'études.....	34
Projet de recherche méthodologique .....	35
UE - Séminaires professionnels .....	36
Séminaires professionnels.....	36

# 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.

## Thématisques 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 3ème année qui le souhaitent de se préparer en vue d'une carrière de chercheur en entreprise au sein des services Recherche et Développement ou dans le secteur académique. Dans le cadre de l'option formation par la recherche (Ofpr), ces élèves bénéficient au cours de leur dernière année d'études à l'Ensaï d'un aménagement de leur scolarité qui leur permet de suivre le master 2 de 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...

**Enseignements de 3<sup>ème</sup> année**  
**Filière "Génie Statistique"**

	Volume horaire				Crédits
	Cours	Atelier	Projet	Total	
<b>Tronc commun</b>					
Droit du travail	3	6		9	1,5
Anglais	30			30	2,5
Sport		30		30	
<b>Total</b>	<b>33</b>	<b>36</b>		<b>69</b>	<b>4</b>
<b>Modélisation</b>					
Modélisation non linéaire <sup>1</sup>	12	6		18	1
Théorie des valeurs extrêmes <sup>2</sup>	16,5	6		22,5	1
Modèles d'ingénierie financière <sup>2,5</sup>	15			15	1
Modèle additifs semi-paramétriques	6	6		12	1
<b>Total</b>	<b>49,5</b>	<b>18</b>		<b>67,5</b>	<b>4</b>
<b>Traitement de l'image</b>					
Modèles markoviens en analyse d'images	30			30	2
Projet d'analyse d'images		15		15	1
Filtrage linéaire et non linéaire <sup>5</sup>	9	3		12	1
<b>Total</b>	<b>39</b>	<b>18</b>		<b>57</b>	<b>4</b>
<b>Processus et prévision</b>					
Statistique des processus <sup>1</sup>	21	6		27	2
Géostatistique	15			15	1
Prévision semi-paramétrique	12			12	1
Séries temporelles avancées	3	21		24	1
<b>Total</b>	<b>51</b>	<b>27</b>		<b>78</b>	<b>5</b>
<b>Qualité/Fiabilité</b>					
Plans d'expériences	18			18	2
Introduction sûreté de fonctionnement	6			6	
Fiabilité	21	6		27	3
<b>Total</b>	<b>45</b>	<b>6</b>		<b>51</b>	<b>5</b>
<b>Outils d'aide à la décision</b>					
Méthodes modernes d'apprentissage <sup>5</sup>	30			30	1,75
Réseaux de neurones <sup>3</sup>	6	12		18	0,25
Statistique bayésienne <sup>4,5,1</sup>	9	6		15	1
<b>Total</b>	<b>45</b>	<b>18</b>		<b>63</b>	<b>3</b>
<b>Informatique</b>					
Python	9	6		15	1
Introduction WinBUGS	3	3		6	
<b>Total</b>	<b>12</b>	<b>9</b>		<b>21</b>	<b>1</b>
<b>Projets</b>					
Projet de fin d'études		12	60	72	2
Projet recherche méthodologique		3	9	12	2
<b>Total</b>		<b>15</b>	<b>69</b>	<b>84</b>	<b>4</b>
<b>Séminaire professionnel</b>					
		21		21	
		<b>21</b>		<b>21</b>	
	<b>274,5</b>	<b>168</b>	<b>69</b>	<b>511,5</b>	<b>30</b>

1 Enseignement commun à la filière Statistique pour les sciences de la vie.

2 Enseignement commun à la filière Gestion des risques et Ingénierie financière

3 Enseignement commun à la filière Statistique et Ingénierie des Données

4 Enseignement commun à la filière Ingénierie statistique des territoires et de la santé

5 Enseignement optionnel pour le Master de Mathématiques, parcours statistique, de l'Université de Rennes 1

UE Stage	Crédits
	<b>30</b>

## **Descriptifs des enseignements du tronc commun**

UE 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 comprises 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

UE – 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<sup>†</sup> 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.

- Schrampfer-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.
- Lougheed, Lin, *Tests complets pour le nouveau TOEIC (4<sup>ème</sup> 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.**

UE – 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
- 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.

## **Descriptifs des enseignements de la filière**

UE - Modélisation

## Modélisation non linéaire

### ***Nonlinear Modeling***

Cours : 12h • Atelier : 6h

Enseignant : Fabien NAVARRO (Ensaï)  
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* (2<sup>nd</sup> ed.). Springer, 1998.
- T. P. RYAN. *Modern Regression Methods* (2<sup>nd</sup> ed.). Wiley, 2008

### **Langue d'enseignement**

Français

UE - 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* (2<sup>nd</sup> 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* (3<sup>rd</sup> ed.), Birkhäuser, 2007
- S.I. RESNICK, *Extreme Values, Regular Variation and Point Processes*, Springer, 1987

### **Langue d'enseignement**

Français

UE – Modélisation

## Modèles avancés d'ingénierie financière

### *Advanced Financial Engineering Models*

Cours : 15h

Enseignant : Valentin PATILEA (Ensaï)  
Correspondant : Valentin PATILEA (Ensaï)

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

#### **Objectif pédagogique**

Etre en mesure d'appliquer les techniques d'estimation non paramétrique (estimateurs à noyau, splines, polynômes locaux) à des problèmes courants en finance, comme par exemple l'évaluation des options et le scoring. Plus précisément, pour ces deux applications possibles, les objectifs sont, d'une part, être en mesure d'estimer une densité risque-neutre et les prix des options européennes à l'aide des estimateurs non paramétriques. D'autre part, être capable de créer un score à l'aide des modèles semi-paramétriques à indice(s). Les techniques d'estimation non paramétrique seront également utilisées pour la validation des modèles de régression usuels.

#### **Contenu de la matière**

1. Rappels d'estimation non paramétrique.
2. Bref rappels des notions de finance (arbitrage, densité risque-neutre,...)
3. Estimation de la densité risque-neutre et des prix des options.
4. Exemples d'applications.
5. Rappels sur les modèles à indice(s).
6. Construction d'un score semi-paramétrique.
7. Test d'adéquations non paramétriques pour les modèles de régression.
8. Exemples d'applications.

#### **Pré-requis**

Les connaissances en estimation non-paramétrique faciliteront la compréhension du cours.

#### **Contrôle des connaissances**

Examen écrit et/ou mini projet

#### **Références bibliographiques**

- W. HÄRDLE, M. MÜLLER, S. SPERLICH, A. WERWATZ, *Nonparametric and Semiparametric Models*, Springer, 2004..
- Y. AÏT-SAHALIA, J. DUARTE, Nonparametric option pricing under shape restrictions (Journal of Econometrics 2003, vol. 116, 9-47)..
- YATCHEW, A. & HÄRDLE, W., 2006, *Nonparametric state price density estimation using constrained least squares and the bootstrap*, Journal of Econometrics, vol. 133, 579-599.

#### **Langue d'enseignement**

Français

UE - 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 pénalisé : 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 (2<sup>nd</sup> 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

UE - 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 bayesienne 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 bayesienne
  - 3.1. Problèmes inverses «mal posés», Exemples : restauration, segmentation
  - 3.2. Estimation bayesienne 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

UE - Traitement de l'image

## Projet d'analyse d'images

### *Image Analysis Project*

Atelier : 15h

Enseignant : (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<sup>ème</sup> année

### **Contrôle des connaissances**

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

### **Langue d'enseignement**

Français

UE – 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 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

#### 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

UE – 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<sup>e</sup> 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

UE – 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 bayesien
  - 5.2. Problèmes inverses

### **Contrôle des connaissances**

Examen écrit

### **Références bibliographiques**

- CHILÈS J.P. and DELFINER P. (2012) Geostatistics. Modeling spatial uncertainty (2<sup>nd</sup> 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

UE - 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<sup>e</sup> 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 adaptative estimation of dimension reduction space, Journal of the Royal Statistical Society, Ser. B, 2002, Vol. 64, No. 3

### Langue d'enseignement

Français

UE - 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<sup>e</sup> 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

UE – 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 », et « Ingénierie statistique des territoires et de la santé »*

### **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<sup>e</sup> 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

UE – Qualité/Fiabilité

## Introduction à la sûreté de fonctionnement

### ***Risk Management and Operation Safety***

Cours : 6h

Enseignant : Tiphaine VOUILLOT (Ligeron)  
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

UE - Fiabilité

## Fiabilité

### ***Reliability***

Cours : 21h • Atelier : 6h

Enseignant : Sophie MERCIER

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 - Outils d'aide à la décision

## Méthodes modernes d'apprentissage

### *Modern Statistical Learning*

Cours : 30h

Enseignant : Badih GHATTAS (Université Aix-Marseille II)

Correspondant : Salima EL KOLEI

*Enseignement destiné aux élèves de la filière "Génie Statistique"*

### Objectif pédagogique

Ce cours présente des méthodes modernes de prévision basées essentiellement sur la segmentation de l'espace des variables explicatives. Les techniques, relativement modernes, utilisées ici peuvent être vues comme une généralisation directe ou non d'outils statistiques plus anciens.

Plusieurs exemples d'application serviront à l'illustration de l'utilisation de ces techniques dans des domaines variés : environnement, médecine, bioinformatique, industrie,...

### Contenu de la matière

1. Introduction à l'apprentissage statistique
2. CART et ses variantes
3. Agrégation d'arbres par bootstrap
4. Les réseaux bayésiens
5. Séparation linéaire et SVM

### Pré-requis

### Contrôle des connaissances

Étude d'un dossier / TP noté

### Références bibliographiques

### Langue d'enseignement

Français

UE - Outils d'aide à la décision

## Réseaux de neurones

### **Neural Networks**

Cours : 6h • Atelier : 12h

Enseignant : Romaric GAUDEL (ENSAI) et Maxime PERALTA

Correspondant : Romaric GAUDEL (ENSAI)

*Enseignement destiné aux élèves des filières « Génie statistique » et « Statistique et Ingénierie des Données » et « Marketing quantitatif et revenue management »*

### **Objectif pédagogique**

Les réseaux de neurones artificiels sont des outils inspirés par la structure du cerveau, créés à l'origine pour imiter certaines fonctions présentes dans le monde animal. Ces outils ont connu un fort développement depuis, et ont étendu considérablement leurs champs d'applications. Récemment, ils ont fait un retour en force en tant que modèle générique pour l'Apprentissage Statistique. Les réseaux de neurones sont utilisés en Apprentissage Statistique pour faire de la classification, pour prédire le comportement d'un système, pour compresser l'information, pour percevoir, pour mémoriser, pour anticiper...

Dans ce cours, nous aurons une introduction aux réseaux de neurones, à leurs divers variantes et aux applications différentes. Des travaux pratiques amèneront à utiliser et apprendre des réseaux de neurones à l'aide de TensorFlow.

### **Contenu de la matière**

- Neurones artificiels et réseaux de neurones artificiels.
- Réseaux particuliers : réseaux convolutionnels, réseaux récurrents, auto-encodeurs, machines de Boltzmann, word2vec ...
- Adaptation de la descente de gradient pour accélérer l'apprentissage

### **Compétences attendues**

- Représenter un réseau de neurones
- Choisir une structure de réseau de neurone adaptée à une tâche particulière
- Utiliser un réseau de neurone fourni
- Apprendre un réseau de neurone

### **Contrôle des connaissances**

Sera indiqué par l'enseignant.

### **Références bibliographiques**

Ian Goodfellow, Yoshua Bengio, Aaron Courville. Deep Learning. MIT Press, 2016.

### **Prérequis**

Bases de la programmation (pour les travaux pratiques)

### **Langue d'enseignement**

Français

UE -Outils d'aide à la décision

## 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

UE – Informatique

## Python

Cours : 9h • Atelier : 6h

Enseignant : Fabien NAVARRO (ENSAI)

Correspondant : Salima EL KOLEI

*Enseignement destiné aux élèves de la filière « Génie Statistique »*

### Objectif pédagogique

Ce cours est une initiation au langage Python pour le calcul scientifique et la programmation statistique. L'objectif principal est d'acquérir les bases nécessaires du langage Python pour la mise en œuvre des méthodes statistiques usuelles.

### Contenu de la matière

1. Introduction à la programmation Python
2. Importation, manipulation et visualisation
3. Mise en œuvre des techniques statistiques

### Contrôle des connaissances

Tp noté et/ou mini projet

### Références bibliographiques

W. MCKINNEY. Python for data analysis: Data wrangling with Pandas, NumPy, and IPython. O'Reilly Media, Inc., 2012.

### Langue d'enseignement

Français

UE – Informatique

## WinBUGS

Cours : 3h • Atelier : 3h

Enseignant : Eric PARENT (Agrocampus Paris)

Correspondant : Salima EL KOLEI

*Enseignement destiné aux élèves de la filière « Génie Statistique »*

### Objectif pédagogique

WinBUGS fait partie du projet BUGS (Bayesian inference Using Gibbs Sampler) qui vise à rendre simple la pratique des méthodes MCMC aux statisticiens. Il a été développé par l'unité MRC Biostatistics de l'université de Cambridge. WinBUGS peut être utilisé de différentes manières :

- Via une interface « clique-bouton » qui permet de contrôler l'analyse,
- En utilisant des modèles définis par des interfaces graphiques, appelés DoodleBUGS,
- Via d'autres logiciels tels que R (en particulier via le package R2WinBUGS)

Au terme de ce cours d'initiation au logiciel WinBUGS, le participant sera capable d'implémenter différents types de modèles, de comprendre les différents algorithmes MCMC et d'utiliser WinBUGS (de manière indépendante ou au travers du logiciel R). Toutes les analyses seront illustrées sur des exemples concrets (biostatistique, psychologie, etc) au moyen de WinBUGS et chaque partie de la formation sera combinée à des séances pratiques en vue de renforcer les concepts introduits.

### Contenu de la matière

- Moyenne : estimation d'une moyenne, test sur une moyenne, comparaison de deux moyennes pour des données normales.
- Régression logistique.
- Modèle linéaire, modèle non linéaire.
- Modèle linéaire mixte, modèle non linéaire mixte.

### Pré-requis

Cours de statistique computationnelle en 2A et statistique bayésienne de 3A.

### Langue d'enseignement

Français

UE – Projets

## Projet de fin d'études

### *Final Year Project*

Atelier : 12h • Projet : 60h

Enseignant : Divers tuteurs  
Correspondant : Salima EL KOLEI

*Enseignement destiné aux élèves de la filière "Génie Statistique"*

### Objectif pédagogique

Travail de recherche méthodologique et de mise en œuvre des méthodes retenues sur des données fournies par l'animateur

### Contrôle des connaissances

L'évaluation prend en compte :

- La qualité de la note d'étape à rendre pour la fin janvier
- la qualité du rapport (pédagogie, rigueur scientifique, réponse à la problématique du client)
- la qualité de la présentation orale
- l'implication des étudiants et la force de proposition durant le projet

### Références bibliographiques

- BACCINI, P. BESSE, S. CANU, S. DÉJEAN, B. LAURENT, C. MARTEAU, P. MARTIN, H. MILHEM, WikiStat - Pour rédiger un rapport de statistique, Université de Toulouse, 2012, [http://www.math.univ-toulouse.fr/~besse/Wikistat/pdf/st-m-redac\\_rapport-stat.pdf](http://www.math.univ-toulouse.fr/~besse/Wikistat/pdf/st-m-redac_rapport-stat.pdf)
- I. MEURANT, Unité de Psychologie Sociale - La rédaction d'un mémoire et d'un article scientifique, Université Libre de Bruxelles, 2008, <http://www.ulb.ac.be/psycho/psysoc/ressources/redaction.htm>
- E. SAUDRAIS, Le petit typographe rationnel - Typographie scientifique, 2005

### Langue d'enseignement

Français

UE - Projets

## Projet de recherche méthodologique

### *Methodological Research Project*

Atelier : 3h • Projet : 9h

Enseignant : Divers tuteurs  
Correspondant : Salima EL KOLEI

*Enseignement destiné aux élèves de la filière "Génie Statistique"*

#### **Objectif pédagogique**

Ce projet a pour but d'approfondir des notions vues en cours ou d'en découvrir d'autres, d'effectuer un travail de modélisation. A la fin du projet, un rapport de vulgarisation scientifique (10-20 pages) doit être produit en anglais. Le travail est réalisé en binôme.

#### **Contrôle des connaissances**

L'évaluation prend en compte la qualité du rapport (pédagogie, rigueur scientifique, anglais), de la présentation orale, ainsi que l'implication des étudiants dans l'avancement du travail et le respect des différents délais demandés.

#### **Références bibliographiques**

- B. MALMFORS ; P. GARNSWORTHY ; M. GROSSMAN, Writing and presenting scientific papers, Nottingham University Press, 2004
- A. WALLWORK, English for Writing Research Papers, Springer, 2011
- A. WALLWORK, English for Presentations at International Conferences, Springer, 2011
- F. VIVALDI, Mathematical Writing, an undergraduate course, University of London, 2011
- J. MARRON, Effective Writing in mathematical statistics, Statistica Neerlandica, 53 – 1,1999
- N. J. HIGHAM, Handbook of writing for the mathematical sciences, SIAM. Society for\_Industrial and Applied Mathematics, 1998

#### **Langue d'enseignement**

Français

UE - Séminaires professionnels

## Séminaires professionnels

### *Professional Seminar*

Atelier : 21h

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