



PROGRAMME DES ENSEIGNEMENTS
2013-2014

TROISIÈME ANNÉE
DATA SCIENCE

Quels savoirs pour quels métiers ?

Compétences acquises

- Aider à la décision grâce à des approches quantitatives adaptées au problème posé en utilisant des sources statistiques existantes ou en construisant des protocoles expérimentaux.
- Mobiliser des méthodes avancées, ainsi que les outils informatiques associés, en statistique, économétrie, analyse des données massives ("Big Data").
- Analyser les résultats en s'appuyant sur des connaissances théoriques approfondies en statistique, économie, sociologie ou marketing.
- Mener des activités de recherche, tant académiques que dans l'entreprise, que ce soit en économie ou sociologie appliquées, en marketing quantitatif ou en statistique fondamentale.

Profils et parcours. La voie Data Science permet d'acquérir un profil hautement qualifié en statistique et en économétrie appliquées pour lequel les débouchés sur le marché du travail sont extrêmement variés, du Data Scientist à l'enseignant-chercheur, en passant par le conseil, l'expertise statistique publique ou industrielle. Suivant les parcours, ces compétences pratiques s'accompagneront de connaissances théoriques approfondies en sciences sociales (économie, sociologie), marketing, mathématiques appliquées (probabilité, statistique) et éventuellement de compétences basiques en biologie.

Le "Data Scientist" ainsi formé possède une expertise scientifique de très haut niveau qui lui permet d'aider à la prise de décision dans de nombreux domaines : évaluation de politiques publiques, évaluation des politiques commerciales des entreprises, finance, biostatistique, imagerie, statistique d'enquête, ou d'effectuer de la recherche fondamentale. Ce profil polyvalent peut mener aussi bien à des carrières d'expert qu'à des postes décisionnels ou d'encadrement en entreprise.

Métiers. On assiste depuis plusieurs années à une explosion du volume de données disponibles, dans des domaines très divers (par exemple la génétique, les neurosciences, la climatologie, mais aussi la finance, le marketing et les sciences humaines et sociales). Après une période où les questions portaient principalement sur le stockage et la conservation de ces données, ce sont maintenant les problématiques liées à leur exploitation statistique et à leur analyse qui apparaissent comme porteuses d'enjeux d'importance. Les emplois générés nécessitent tout à la fois des compétences techniques et une compréhension stratégique des enjeux sous-jacents.

Le caractère transversal des méthodes quantitatives présentées dans cette voie permet aux étudiants d'accéder à une large palette de métiers, aussi bien dans le public que dans le privé. Cette voie développe, entre autres, les compétences attendues pour les postes de "Chief Data Officer" qui émergent dans le contexte lié aux "Big Data".

Trois modules, ou domaines d'applications, sont proposés :

Le module **Statistique et apprentissage** (DS1) vise à concevoir les modèles statistiques, à trouver ou construire les bases de données (plans d'expérience, pré-traitements, etc.) et à organiser les tests ou l'apprentissage statistique pour soutenir la prise de décision en rationalité limitée. Il offre des cours dans différents domaines : les enseignements de biostatistique conduisent aux métiers de statisticiens exercés en laboratoire pharmaceutique, dans l'industrie agro-alimentaire ou au sein d'organismes publics tels que l'INSERM ; les enseignements liés aux enquêtes statistiques conduisent aux métiers de méthodologues au sein des instituts de sondage, dans les services statistiques et d'études des grandes entreprises et des administrations, et au sein des sociétés de conseil ; les enseignements orientés vers l'apprentissage et les statistiques en grande dimension conduisent aux métiers d'experts statistiques dans l'industrie ou les grandes entreprises utilisant de grandes bases de données telles que Google ou Amazon. Ces enseignements sont soutenus par des cours théoriques fondamentaux et mènent aussi à la recherche en statistique.

Le module **Economie et sociologie quantitatives** (DS2) vise à étudier, modéliser et tester les comportements des agents pour éclairer les décideurs publics ou privés dans leurs choix et l'appréhension quantifiée de leurs impacts. Il s'agit plus précisément, d'une part, d'être capable d'élaborer un processus de construction de données économiques et sociales et, d'autre part, de savoir utiliser et analyser ces données grâce aux théories économiques et sociologiques récentes. Ce module mène à des postes d'études dans des institutions chargées d'évaluer l'efficacité des politiques publiques ou d'étudier les comportements des agents économiques : INSEE, ministère des finances, mais aussi ministère de l'éducation nationale, ministère du travail ou de la santé, organismes nationaux tels que les caisses de sécurité sociale, l'UNEDIC, l'OFCE, la Banque de France, le Crédoc, internationaux tels que l'OCDE, la Banque mondiale, les institutions européennes ou le FMI, ou encore les instituts de sondage et des laboratoires de recherche (Institut des politiques publiques par exemple). Couplé à un master, ce module mène en effet aussi à la recherche en économie appliquée ou en sociologie quantitative.

Le module **Marketing quantitatif et aide à la décision** (DS3) vise à mettre les données massives ("Big Data") issues de la révolution digitale et les données classiques au service de la compréhension des attentes et des comportements des consommateurs pour optimiser les offres et la relation client. Ce module mène aux carrières d'experts statistiques dans les grandes entreprises, les sociétés de conseil et les instituts de sondage. Les diplômés de l'ENSAE qui ont choisi ce profil ont eu des carrières particulièrement dynamiques grâce au caractère exclusif de la formation de l'école dans ce domaine. Les entreprises tournées vers la clientèle grand public (Amazon, Google, CDiscount, Orange, Canal+ etc.) sont des débouchés privilégiés pour valoriser la maîtrise du quantitatif acquise dans ce module. Il offre également de nombreux débouchés dans les sociétés spécialisées dans les enquêtes d'opinion et de mesure d'audience. La capacité des diplômés à extraire de l'information pertinente à partir de données et à évoluer par la suite dans l'entreprise est ici un avantage décisif.

Organisation des cours de 3ème année : voies et modules de spécialisation

L'organisation des cours de 3ème année répond à une logique de métiers. Chaque voie de spécialisation délimite un grand domaine de débouchés (décrits ci-dessus) nécessitant des compétences homogènes.

Chaque voie de spécialisation est constituée d'un tronc commun et de plusieurs modules de spécialisation.

Les cours de tronc commun présentent des méthodes ou des applications d'une portée très générale dans le domaine d'intérêt de la voie. Ils donnent un panorama complet de l'état de l'art dans ce domaine. Certains sont fondamentaux (et obligatoires), d'autres sont des cours d'ouvertures pouvant être choisis librement par les élèves.

Les modules de spécialisation sont des séquences de cours cohérentes en vue d'exercer les métiers types qui constituent les débouchés de la voie. Pour autant, ces cours ne sont pas nécessairement appliqués, car il peut être nécessaire d'approfondir un domaine théorique spécifique avant de pouvoir en présenter les champs d'application. Chaque élève choisit un module, dont il suit tous les cours obligatoires. Il suit également des enseignements optionnels du module (ou les cours d'un autre module s'il souhaite se constituer un profil polyvalent).

Au total, chaque élève doit valider 30 crédits ECTS par semestre. Parmi l'ensemble des cours validés par l'élève, au moins 10 doivent appartenir à leur voie de spécialisation (soit typiquement 2/3 des crédits annuels). Le volant de cours pouvant être choisi "hors voie" permet aux élèves de se constituer des profils personnalisés panachant, dans certaines limites, les cours de différentes voies de spécialisation.

Les élèves effectuent en outre un "groupe de travail", projet de recherche théorique ou appliquée, mené par groupes de 2 à 4 élèves tout au long de l'année et donnant lieu à la rédaction d'un rapport au format d'un article scientifique.

Enfin, les élèves peuvent valider les cours de "Formation par la recherche" organisés par le CREST au titre des cours optionnels de chaque voie. Le descriptif de ces cours est disponible sur le site internet du CREST.

Tronc commun de la voie Data Science

Les cours suivants sont obligatoires :

- **Apprentissage et Datamining**
- **Méthodes statistiques de l'économétrie**
- **Sondages**

Les cours optionnels de tronc commun sont les suivants :

- **Analyse des données approfondie**
- **Développements récents en économétrie semi et non paramétrique**
- **Éléments logiciels pour le traitement des données massives, C++**
- **Histoire de la statistique**
- **Macroéconométrie (séries temporelles avancées)**
- **Outils pour l'analyse des données massives**
- **Séminaire de sondages**
- **Statistical Analysis of Network Data with Applications in Marketing**
- **Statistique et économétrie spatiale**
- **Structural Econometrics**

Module Marketing quantitatif et aide à la décision

Cours obligatoires :

- **Bases de données et web**
- **Économétrie avancée des variables qualitatives**
- **Evaluation of public policies**
- **Marketing**
- **Économétrie du marketing**
- **Méthodes de scoring**

Cours optionnels :

- **Bootstrap et rééchantillonnage**
- **Computational statistics**
- **Démographie**
- **Duration Models**
- **Economics of Innovation and Intellectual Property**
- **Économie de l'énergie**
- **Empirical Corporate Finance**
- **L'enquête statistique en sciences sociales**
- **Gestion humaine dans l'entreprise**
- **Macroéconométrie appliquée**
- **Modèles statistiques dynamiques à variables cachées**

- **Sociologie de la conso. et des modes de vie**
- **Statistique bayésienne**
- **Topics in Empirical Industrial Organization**

Module Statistique et Apprentissage

Cours obligatoires :

- **Bootstrap et rééchantillonnage**
- **Computational Statistics**
- **Estimation non paramétrique**
- **Modèles GARCH et à volatilité stochastique**
- **Statistique bayésienne**

Cours optionnels :

- **Apprentissage statistique en génomique**
- **Bases de données et web**
- **Copules et applications**
- **Duration Models**
- **Économétrie avancée des variables qualitatives**
- **Evaluation of Public Policies**
- **Introduction aux méthodes statistiques pour la biologie moléculaire**
- **Méthodes de scoring**
- **Méthodes statistiques en finance ? modèles à chaînes de Markov**
- **Modèles statistiques dynamiques à variables cachées**
- **Statistique asymptotique et robuste**
- **Statistique pour l'épidémiologie**

Module Economie et sociologie quantitatives

Cours obligatoires :

- **Duration Models**
- **Econométrie avancée des variables qualitatives**
- **Evaluation of public policies**
- **Séminaire de sociologie quantitative**

Cours optionnels :

- **Advanced Empirical Organization**
- **Bootstrap et rééchantillonnage**
- **Computational Statistics**

- **Démographie**
- **Development Economics**
- **Economics of Education : structural approches**
- **Economics of Innovation and Intellectual Property**
- **Economie de l'éducation : politiques publiques**
- **Economie de l'énergie**
- **Économie de l'environnement**
- **Économie de la santé**
- **Empirical Corporate Finance**
- **L'enquête statistique en sciences sociales**
- **Labor Economics and Employment Policies**
- **Macroéconométrie appliquée**
- **Modèles d'acteur en sociologie**
- **Modèles statistiques dynamiques à variables cachées**
- **Sociologie de la conso. et des modes de vie**
- **Sociologie de l'immigration**
- **Statistique bayésienne**
- **Topics in Empirical Industrial Organization**

Ces différents cours sont présentés en détail dans la suite de ce fascicule. Le tableau ci-dessous indique pour chaque cours :

- le nom de l'enseignant
- le semestre où le cours est programmé
- le volume horaire
- le nombre de crédits ECTS
- les voies et modules de l'ENSAE où il est ouvert/obligatoire (S désigne par exemple la voie statistique, S1 le premier module de cette voie ; une lettre en gras indique que le cours est obligatoire dans la voie ou le module correspondant)
- le numéro de page du catalogue où le cours est décrit

code	intitulé du cours	voie	cours+tp	crédit	examen	intervenant	correspondant	page
PREMIER SEMESTRE								
SE309	Advanced Econometrics of Qualitative Data	3A, 3A - DS1, 3A - DS2, 3A - DS3, 3A - AMFEL, CESS, APE	14+0	3	mém.	Kamionka	Zakri	25
AE301	Advanced Industrial Organization	3A, 3A - DS2, 3A - AMFEL, 3A - AMPE2, 3A - PPE2, CESS, APE	24+0	3	oral	Chambolle, Lin-nemer, Vergé	Sauvagnat	9
SE306	Apprentissage et datamining	3A, 3A - DS1, 3A - DS2, 3A - DS3, CESS, TSI	14+10	3	écrit+CC	Dalalyan	Brunel	23
TSI304	Bayesian Data Analysis	3A, 3A - DS1, 3A - DS2, 3A - DS3	24+12	3		Gelman	Brunel	40
OMI303	C++	3A, 3A - DS1, 3A - FQ1, 3A - FQ2, CESS	12+12	3	mém.	Durut	Cazorla	19
SE331	Computational statistics	3A, 3A - DS1, 3A - DS2, 3A - DS3, CESS	12+8	3	mém.	Robert	Brunel	37
AE324	Development Economics	3A, 3A - DS2, 3A - PPE2, CESS, APE	24+0	3	écrit	Dumas, Lambert	Mohkam	12
SE332	Développements récents en économétrie semi et non paramétrique	3A, 3A - DS1, 3A - DS2, 3A - DS3, APE	20+0	3	mém+sout.	D'Hautfoeuille, Gautier	Zakri	38
SE307	Duration Models	3A, 3A - DS1, 3A - DS2, 3A - DS3, 3A - PPE1, 3A - PPE2, 3A - GRR, 3A - A, CESS, APE, SOCSTAT, Institut des Actuariers	14+0	3	écrit	Lopez	Milhaud	24
AE326	Economics of Education : empirical evidence and public policies	3A, 3A - DS2, CESS, APE	24+0	3	écrit+CC	Behaghel, Gurgand	Zakri	13
AE306	Economie de l'environnement	3A, 3A - DS2, 3A - AMFEL, CESS	14+0	3	mém.	Bourgeon	Sauvagnat	11
OMI309	Eléments logiciels pour le traitement des données massives	3A, 3A - DS1, 3A - DS2, 3A - DS3, 3A - FQ1, 3A - FQ2, CESS	0+16	3	mém.	Durut	Cazorla	20
SE317	Evaluation of Public Policies	3A, 3A - DS1, 3A - DS2, 3A - DS3, 3A - AMFEL, 3A - PPE2, CESS, APE, EPP	24+0	3	écrit	Crépon, Givord	Zakri	32
SE318	Introduction aux méthodes statistiques en biologie moléculaire	3A, 3A - DS1, CESS, TSI	16+4	3	mém.	Matias, Neuvial	Brunel	34
SE312	Macroéconométrie (Séries Temporelles Avancées)	3A, 3A - DS1, 3A - DS2, 3A - DS3, 3A - PPE1, 3A - PPE2, 3A - GRR, 3A - FQ1, 3A - FQ2, CESS, TSI	18+4	3	écrit	Doz	Zakri	28
DES311	Marketing	3A, 3A - DS3, 3A - AMFEL, CESS	20+0	3	CC	Balagué	Zakri	14
SE321	Modèles à espace d'état et filtrage particulière	3A, 3A - DS1, 3A - FQ2, CESS, TSI	18+0	3	mém.	Chopin	Brunel	36
DES314	Modèles d'acteur en sociologie	3A, 3A - DS2, CESS, SOCSTAT	16+0	3	écrit	Lemel	Zakri	15
SE310	Modèles statistiques dynamiques à variables cachées	3A, 3A - DS1, 3A - DS2, 3A - DS3, 3A - PPE1, 3A - PPE2, 3A - FQ2, CESS, MASEF, TSI	14+0	3	écrit	Zakoian	Brunel	27
DES317	Sociologie de la consommation et des modes de vie	3A, 3A - DS2, 3A - DS3, CESS, SOCSTAT	16+0	3	écrit	de Saint Pol	Zakri	17
SE314	Sondages	3A, 3A - DS1, 3A - DS2, 3A - DS3, CESS	18+8	3	écrit	Sautory	Zakri	29
SE334	Statistical Analysis of Network Data with applications in Marketing	3A, 3A - DS1, 3A - DS2, 3A - DS3	16+0	3	mém.	Kolaczyk	Brunel	39
SE315	Statistique asymptotique et robuste	3A, 3A - DS1, CESS	18+0	3	mém.	Bertail	Brunel	31
SE301	Statistique bayésienne	3A, 3A - DS1, 3A - DS2, 3A - DS3, CESS	18+0	3	mém.	Rousseau	Brunel	22

code	intitulé du cours	voie	cours+tp	crédit	examen	intervenant	correspondant	page
DEUXIEME SEMESTRE								
SE305	Analyse des données approfondie	3A, 3A - DS1, 3A - DS2, 3A - DS3, CESS	20+0	2,5	écrit	Bel, Blumenthal, Robin Vert	Zakri	75
SE335	Apprentissage statistique en génomique	3A, 3A - DS1	16+0	2,5	mém.		Brunel	82
OMI308	Base de données et Web	3A, 3A - DS1, 3A - DS3, CESS	0+20	2,5	mém.		Cazorla	66
SE303	Bootstrap et rééchantillonnage	3A, 3A - DS1, 3A - DS2, 3A - DS3	16+0	2,5	mém.	Bertail	Brunel	72
FA331	Copules et applications	3A, 3A - DS1, 3A - GRR, 3A - FQ2, 3A - A, MASEF, TSI, MA - P7	16+0	2,5	écrit	Fermanian	Fermanian	64
DES319	Démographie	3A, 3A - DS2, 3A - DS3, 3A - A, CESS	14+0	2,5	CC	Bonnet, Deses-quelles, Pailhe, Solaz	Zakri	59
DES318	Econométrie du marketing	3A, 3A - DS3, 3A - AMFE1, 3A - AMFE2, CESS	16+0	2,5	CC	Belagué, Wilner	Zakri	57
AE340	Economics of Education : structural approaches	3A, 3A - DS2, EPP	24+0	2,5	écrit+tut.	Belzil	Zakri	46
AE344	Economics of Innovation and Intellectual Property	3A, 3A - DS2, 3A - DS3, 3A - AMFE1, 3A - AMFE2	14+0	2,5	mém.	Patnam	Sauvagnat	49
AE342	Economie de l'énergie	3A, 3A - DS2, 3A - DS3, 3A - AMFE1, 3A - AMFE2, 3A - PPE1, 3A - PPE2	16+0	2,5	écrit	Joudon, Ollivier, Trochet	Sauvagnat	48
AE312	Empirical Corporate Finance	3A, 3A - DS2, 3A - DS3, 3A - AMFE1, 3A - AMFE2	24+0	2,5	mém.	Patnam	Sauvagnat	42
SE329	Estimation non paramétrique	3A, 3A - DS1, CESS, MA - P7	16+8	2,5	écrit	Tsybakov	Brunel	81
DES305	Gestion humaine dans l'entreprise	3A, 3A - DS3, 3A - AMFE2, CESS	10+2	2,5	mém.	Grandcoing	Sauvagnat	52
AE308	Health Economics	3A, 3A - DS2, 3A - AMFE1, 3A - A, CESS, EPP	20+0	2,5	mém.	Bardey, Rochaix	Sauvagnat	41
SE2S4	Histoire de la statistique	3A, 3A - DS1, 3A - DS2, 3A - DS3, SOCSTAT	18+0	2,5	mém.	Davezies, Didier	Zakri	70
AE318	Labor economics and Employment Policies	3A, 3A - DS2, 3A - AMFE1, 3A - PPE1, 3A - PPE2, APE, EPP	24+0	2,5	écrit	Cahuc, Kramarz, Roux	Mohkam	44
AE319	Macroéconométrie appliquée	3A, 3A - DS2, 3A - DS3, 3A - PPE1, 3A - PPE2, CESS	20+0	2,5	mém.	Belly, Dubois, Le Bihan	Mohkam	45
SE304	Méthodes statistiques de l'économétrie	3A, 3A - DS1, 3A - DS2, 3A - DS3, CESS	16+0	2,5	mém+sout.	Monfort	Brunel	73
SE311	Modèles GARCH et à volatilité stochastique	3A, 3A - DS1, 3A - PPE1, 3A - FQ2, CESS, TSI, MA - P7	14+0	2,5	écrit	Francoq	Francoq	76
OMI310	Outils pour l'analyse des données massives	3A, 3A - DS1, 3A - DS2, 3A - DS3	18+6	2,5			Cazorla	68
OMI301	Projet informatique	3A, 3A - DS1, 3A - DS2, 3A - DS3, CESS	0+10	5	mém.	diff. enseignants	Cazorla	65
FA306	Scoring	3A, 3A - DS1, 3A - DS3, 3A - GRR, 3A - A, CESS, MASEF	12+0	2,5	mém.	Georges, Jacomy	Zakri	63
DES316	Séminaire de sociologie quantitative	3A, 3A - DS2, CESS, SOCSTAT	20+0	2,5	CC	Petev	Zakri	55
SE336	Séminaire de sondages	3A, 3A - DS1, 3A - DS2, 3A - DS3	20+0	2,5	CC		Zakri	83
DES322	Sociologie de l'immigration et des inégalités ethnoraciales	3A, 3A - DS2, SOCSTAT	14+0	2,5	mém.	Safi	Zakri	61
DES313	Sociologie et enquête statistique	3A, 3A - DS2, 3A - DS3, CESS, SOCSTAT	16+0	2,5	mém.	Amossé, Vallet	Zakri	53
SE337	Statistique et économétrie spatiale	3A, 3A - DS1, 3A - DS2, 3A - DS3	14+6	2,5	mém.	Bouayad Agha, Vedrine	Zakri	77
SE313	Statistique pour l'épidémiologie	3A, 3A - DS1, CESS, TSI	10+4	2,5	mém.	Clemençon	Brunel	79
SE328	Structural Econometrics	3A, 3A - DS1, 3A - DS2, 3A - DS3, CESS, APE	18+2	2,5	CC	Belzil	Zakri	80
AE345	Topics in Empirical Industrial Organization	3A, 3A - DS2, 3A - DS3, 3A - AMFE1, 3A - AMFE2, APE	18+0	2,5	mém.	Février	Sauvagnat	50

Advanced Industrial Organization (AE301)

Professeur : Claire Chambolle - INRA
Laurent Linnemer - ENSAE-CREST
Thibaud Vergé - ENSAE-CREST

premier semestre
Cours : 24 heures
TP : 0 heures

Objectifs

The course is in two parts. A first part is devoted to market interactions under asymmetric or incomplete information. A second part is devoted to the economics of retailing.

Plan

Information and Strategic Behavior (six two-hour sessions)

The first two sessions are devoted to static models of competition. We study how competition is affected by imperfect information in the product market. First, consumers might be imperfectly informed about firms. Next, a firm might be imperfectly informed about the other firms.

1. Varian (1980) ; Bagwell-Ramey (1994) ; Bagwell-Lee (2010) ;
2. MacMin (1980) ; Hansen (1988) ; Spulber (1995).

Sessions 3 and 4 are devoted to models of entry. We examine how entry/exit decisions are affected by imperfect information, one firm (usually the incumbent) being more informed about market conditions than the other (usually the entrant or a smaller rival).

3. Reputation : Selten (1978) ; Kreps-Wilson (1982).
4. Limit-pricing : Milgrom-Roberts (1982)

The last two sessions are devoted signals of product quality. We study, through the example of quality, how asymmetric information affects the interaction between firms and consumers.

5. Milgrom-Roberts (1986)
6. Bagwell-Riordan (1991)

Economics of retailing (six two-hour sessions)

Retailing is not only an important sector of our everyday life ; it is also one that raises a number of interesting and complex economic issues. We focus on some of these issues from a theoretical, empirical and policy point-of-view. A first session presents a broad overview of the retailing sector in France as well as in neighboring countries and introduces some of the issues that we will focus on later : How does competition between retailers take place ? What are the consequences of relationships between retailers and their suppliers ?

Allain M.L et C. Chambolle, (2003), 'Economie de la distribution ', Repères, éditions La découverte.

The second session on "competitive strategies in the retailing industry" analyses the various types of competition between retailers : retailers compete in prices but we should keep in mind that retailers are inherently multiproduct firms. They also compete on line of products and formats. Finally, promotions and advertising play in important role in the interactions between retailers.

Johnson J.P et D. Myatt (2003), Champsaur P. et J.-C. Rochet (1989),

The next three sessions focus on various aspects of the vertical relationships between the manufacturers

and the retailers. We first look at the standard textbook vertical restraints models and see how vertical contracts (resale price maintenance, exclusive territories, exclusive and/or selective distribution) allow firms to solve vertical coordination problems (double marginalization, service free-riding, etc.) but may also lead to the softening of inter-brand competition.

Motta, M. , (2004), Rey, P. and J. Tirole, (1986), Rey, P. and J. Stiglitz, (1995).

The fourth session will then look at the advantages and disadvantages of a common type of vertical agreement, namely franchising. Franchising has several benefits for both the franchisor and the franchisee as it may eliminate liquidity constraints or solve asymmetric information problems, but it may also be anticompetitive as it eliminates, or at least soften intra-brand competition.

Mathewson G.F and R. A. Winter (1985)..

The fifth session will look at the increasing shift in negotiation power that occurred in the recent years. In this session, we focus among other on the role of retailer's private label and category management.

Bonnet, C. et Dubois, P., (2010), Dhar, S. K., Raju, J. S., Subramanian, U. and Y. Wang (2009), Rochet, J.C, (1988)

In the sixth and final session, we focus on the effects of retail regulations on market structure, looking in particular at the French grocery retailing industry. A first issue will be the impact of entry regulations (Loi Royer and Loi Raffarin) on the industry structure ; the second issue will be the impact of the ban on below-cost pricing (especially Loi Galland) on retail prices.

Allain, Chambolle et Vergé (2008), Bertrand, M. and F. Kramarz (2002).

Références

Economie de l'environnement (AE306)

Professeur : Jean-Marc Bourgeon - INRA

premier semestre

Cours : 14 heures

TP : 0 heures

Objectifs

La vocation de cet enseignement est de montrer comment l'analyse microéconomique peut être utilisée dans un domaine d'application important : l'environnement. Il fera le "pont " entre la théorie microéconomique et des applications concrètes. Les illustrations seront aussi bien françaises qu'étrangères.

Plan

1. **Introduction et rappel sur les externalités**
2. **Les instruments économiques** - Les taxes ; les marchés de droits : illustrations américaines et européenne.
3. **L'évaluation des avantages** - L'évaluation contingente ; les autres méthodes.
4. **Economie des ressources naturelles** - Règle de Hotelling et croissance avec ressources épuisables ; ressources renouvelables.
5. **Application** : Effet de serre.

Références

- BARDE J.-Ph. (1992), *Economie et politique de l'environnement*, Presses universitaires de France, Paris [73 BAR 01 A].
- BAUMOL W.J., OATES W.E. (1988), *The Theory of Environmental Policy*, Second edition, Cambridge University Press [73 BAU 00 A].
- DESAIGUES B., POINT P. (1993), *Economie du patrimoine naturel. La valorisation des bénéfices de protection de l'environnement*, Economica [73 DES 00 A].
- HENRY C. (1989), *Microeconomics for Public Policy. Helping the Invisible Hand*, Clarendon Press, Oxford [63 HEN 01 A].
- NEHER P.A. (1990), *Natural Resources Economics. Conservation and Exploitation*, Cambridge University Press [73 NEH 00 A].

Development Economics (AE324)

Professeur : Christelle Dumas - Université de Cergy-Pontoise
Sylvie Lambert - INRA - PSE

premier semestre
Cours : 24 heures
TP : 0 heures

Objectifs

Ce cours analyse les aspects théoriques du développement économique et les principes de modélisation applicables aux pays en développement.

Plan

1. **Introduction** - Pauvreté. Inégalité.
2. **Croissance et convergence**
3. **Géographie et institutions**
4. **Modèle dual** - Complémentarités entre secteurs
5. **Ouverture et croissance**
6. **Politiques** - Dette, aide et stabilisation.
7. **Nutrition et santé**
8. **Ménages** - Allocation des ressources au sein du ménage. Modèles de ménages ruraux.
9. **Education**
10. **Migrations et transferts**
11. **Risque** - Marchés manquants et arrangements contractuels. Assurance et crédit. Accès à la terre. Relations contractuelles jointes.

Références

Bardhan, P. et C. Udry (1999), *Development Microeconomics*, Oxford university press.

Economics of Education : empirical evidence and public policies (AE326)

Professeur : Luc Behaghel - Paris School of Economics-INRA
Marc Gurgand - CNRS-PSE

premier semestre
Cours : 24 heures
TP : 0 heures

Objectifs

This course provides an introduction to the economic analysis (both theoretical and empirical) of the investment in and provision of education.

The theoretical background that explains individual and public investment in education is reviewed and linked to empirical evidence. One important application is the analysis of the returns to education both at the micro and macro level. The course will also analyze the production and provision of education. Examples are the importance of inputs such as teachers and class size, the role of incentives and the analysis of peer effects.

Each course will consist of a general introduction by the instructors, followed by presentation of papers by students.

Plan

Références

Marketing (DES311)

Professeur : Christine Balagué - Telecom Paris Sud

premier semestre

Cours : 20 heures

TP : 0 heures

Objectifs

- Introduire les concepts fondamentaux en marketing comme mode de création et de gestion de valeur
- Apprendre aux étudiants à maîtriser ces concepts en les appliquant à des cas réels récents d'entreprises
- Initier les étudiants aux nouvelles tendances du marketing sur les nouvelles technologies

Les séances sont constituées d'un cours théorique sur les principaux concepts, suivi de la présentation d'un diagnostic marketing mené par un groupe d'étudiants sur un cas réel d'entreprise. Le déroulement est calqué sur la démarche adoptée par les responsables marketing.

Plan

1. Introduction : historique, définitions, concepts, outils, métiers
2. La création de valeur, l'analyse stratégique et le plan d'action
3. Etudes et recherches en marketing
4. Segmentation, ciblage, positionnement
5. Stratégie produit et marque
6. Innovation et lancement d'une nouvelle offre
7. Politique de prix
8. Stratégie de communication
9. E-marketing, réseaux sociaux

Références

- Marketing Management (2010), Philip Kotler, Kevin Keller, Delphine Manceau, Bernard Dubois, 13ème édition, Ed. Pearson
- L'essentiel du marketing (2010) Eric Vernet, 3ème édition, Ed. Eyrolles
- Marketing, Fondements et méthodes des recherches en Marketing (2009) Bernard Pras, Elyette Roux, Pierre Desmet, 4ème édition, Ed. Dunod
- Facebook, Twitter et les autres, comment intégrer les réseaux sociaux dans une stratégie d'entreprise (2010) Christine Balagué, Editions Pearson
- Le marketing en ligne (2011) Christine Balagué, Editions Pearson

Modèles d'acteur en sociologie (DES314)

Professeur : Yannick Lemel - INSEE-CREST

premier semestre

Cours : 16 heures

TP : 0 heures

Objectifs

Les modèles d'acteurs - représentations d'individus agissant sous un système de contraintes afin d'atteindre les buts qu'ils se sont donnés - ont toujours été utilisés de manière plus ou moins explicite par les sociologues.

La nouveauté de ces vingt dernières années est moins dans l'utilisation de modèles de ce type que dans l'effort consenti pour les préciser et les formaliser. L'objectif fondamental poursuivi est de permettre l'analyse du fonctionnement de systèmes sociaux en s'appuyant sur une modélisation adaptée des interrelations qui lient les agents et organisent leurs actions. Le cours ne prétend pas traiter du problème dans toute sa généralité mais introduira les éléments de base retenus par les sociologues pour ce faire. Seront distingués et traités : les types d'objectifs poursuivis par les agents, les règles au sein desquelles ceux-ci inscrivent (de manière plus ou moins consciente) leur action, les modalités d'agrégations des comportements individuels et, enfin, les conséquences pour les individus de leur insertion dans des réseaux structurés d'interrelation. Les différentes notions et concepts seront illustrés à l'aide d'exemples empruntés aux différents champs sociologiques utilisant des modèles d'acteurs : sociologie de la famille, du travail, l'analyse des réseaux sociaux, etc.

Plan

1. **Ressemblances et différences entre approches d'économistes et de sociologues.** Grounded theory et démarche d'abstraction descendante. Architecture générale d'un modèle d'acteur dans une perspective sociologique. Niveaux d'analyse, micro, méso et macro. Place de l'école dite du 'rational choice'. Comportements prosociaux et altruisme : du rôle de l'intérêt personnel
2. **Systèmes normatifs et règles de comportement** - Normes et rôles. Les sanctions. L'intériorisation minimale des règles comme a priori de l'approche sociologique. Valeurs et systèmes de valeurs. Systèmes culturels. Les approches des systèmes normatifs suivant le degré d'intériorisation qu'ils présupposent. Problèmes d'observation.
3. **Les objectifs poursuivis par les agents** - Notion de 'bien social'. Pouvoir, prestige et richesse. Eléments de définition. Raisons de les distinguer. Problèmes de mesure et d'identification des buts.
4. **L'individualisme structurel.** Les réseaux d'appartenance comme structurant l'action. La notion d'embeddedness. Structure sociale et rationalité individuelle. Maintenance et création des systèmes normatifs. L'organisation formelle comme système d'action. Exit, voice and loyalty. Notion de 'corporate actor'.
5. **L'approche des comportements individuels** - Phénomènes émergents et effets pervers. La théorie des jeux appliquée aux comportements des foules. Notion de corporate actor. L'organisation formelle comme système d'action. Exit, voice and loyalty. Marché et 'corporate actors'.

Références

- DEGENNE A. et LEMEL Y., Sociologie des Comportements Intentionnels, Economica. Paris - 2006. [91 DEG 00 B]
- BOUDON R., Sociology that Really Matters , in European Sociological Review, (vol 18,3), 2002 [Cf JSTOR]].
- BOUDON R., L'inégalité des chances : la mobilité sociale dans les société industrielles, A. Colin. Paris - 1973 [91 BOU 00 G].
- COLEMAN J., Foundation of Social Theory, Harvard U.P. Cambridge - 1990 [91 COL 00 B].
- GRANOVETTER M., Economic Action and Social Structure : The Problem of Embeddedness, in American Journal of Sociology (vol.91, no3), 1985. [Cf JSTOR]
- HEDSTROM P. and BEARMAN P. (eds.), The Oxford Handbook of Analytical Sociology, Oxford University Press, Oxford, 2009
- KELLERHALS et alii, Microsociologie de la famille, PUF. Paris - 1984.[91 KEL 00 B]
- OLSON M., La logique de l'action collective, PUF. Paris - 1975 [91 OLS 00 A].

Sociologie de la consommation et des modes de vie (DES317)

Professeur : Thibaut de Saint Pol - INSEE-CREST

premier semestre

Cours : 16 heures

TP : 0 heures

Objectifs

Consommer, ce n'est pas seulement répondre à des besoins naturels, mais s'inscrire dans un ensemble de codes, de pratiques et de représentations essentielles dans la construction de l'identité des individus et des groupes.

Ce cours a pour but de montrer comment l'étude de la consommation, en se plaçant au croisement d'autres champs de la sociologie comme la famille, le travail ou la culture, apporte un éclairage décisif pour la compréhension de nos sociétés.

Plan

1. **Introduction** - Approcher la consommation en sociologie. De l'étude des budgets ouvriers à l'analyse des comportements selon la classe sociale.
2. **La diversité des modes de vies et l'évolution des pratiques** - Intérêts et limites des enquêtes Budget de famille. Effets d'âge et effets de générations.
3. **Sociologie des biens de consommation** - Biens et besoins. Le luxe, la valeur et les goûts. Gas-pillage et consommation ostentatoire.
4. **Le corps est-il 'notre plus bel objet de consommation' ?** - Le poids des apparences dans les interactions sociales. Le corps comme capital. L'obésité comme symptôme de la surconsommation.
5. **Inégalités de consommation. L'exemple de la santé** - La santé, un bien comme les autres ? L'accès aux soins et à la prévention. Le rôle de l'état et la mise en œuvre de politiques publiques.
6. **Comprendre la décision d'achat** - La rationalité du consommateur et le rôle des marques. Lieux de vente et lieux d'achats. Logiques et analyses critiques de la publicité.
7. **Sociologie de la mode** - Aspiration à l'imitation et quête de distinction. Les phénomènes de diffusion.
8. **Consommation de masse et cohésion sociale** - La consommation comme lien social. Désenchantement, exclus et critiques de la société de consommation.

Références

Baudrillard J., La société de consommation, Paris, Denoël, 1970. [91 BAU 02 D]

Bourdieu P., La Distinction, Paris, éditions de Minuit, 1979. [91 BOU 01 A]

Douglas M., Isherwood B., The world of goods : towards an anthropology of consumption, New York : Basic books, 1979.

Halbwachs M., La classe ouvrière et les niveaux de vie, Paris, Alcan, 1912.

Mauss M., 'Essai sur le don, forme et raison de l'échange dans les sociétés archaïques', 1923, rééd. in Sociologie et Anthropologie, Paris, PUF, 'Quadrige', 2001.

Muniz A., O'Guinn T., 'Brand community', Journal of Consumer Research, March 2001.

de Saint Pol T., Le Corps désirable. Hommes et femmes face à leur poids, Paris : Presses Universitaires

de France, coll. 'Le lien social', 2010. [91 SAI 03 A 1]

Simmel G., ' La mode ', Tragédie de la culture et autres essais, Paris, Rivages, 1988.

Veblen T., La théorie de la classe de loisirs, Gallimard 1970, (1899). [91 VEB 00 A]

C++ (OMI303)

Professeur : Matthieu Durut - Lokad

premier semestre

Cours : 12 heures

TP : 12 heures

Objectifs

Le langage C++, langage dit ' orienté objet ' et développé dans les années 80 par Bjarne Stroustrup, est actuellement un des standards de l'industrie. Ses performances et sa syntaxe, proches de celles du langage C, couplées à son côté objet, qui permet de structurer facilement le code, lui ont permis de s'imposer rapidement, entre autres dans l'industrie financière.

L'objectif de ce cours est d'initier au C++ les élèves de 3^e année n'ayant pas pris l'option C++ en 2^e année. Il suppose la connaissance d'au moins un langage de programmation (C, Pascal, Java, Python, etc.). L'évaluation sera faite au travers de la réalisation d'un projet.

Les TD aborderont des sujets utiles dans le monde professionnel (génération de nombres aléatoires, méthode de MonteCarlo, interfaçage C++/Excel).

Les élèves ayant suivi le cours de 2^e année de C++ en sont dispensés pour valider le module.

Plan

1. **Rappels de C.**
2. **Introduction à l'objet** - Classes. Polymorphisme. Construction/destruction. Exemple traité : écriture d'un objet ' nombre complexe '.
3. **Allocation dynamique** - Particularités de l'allocation dynamique en C++. Constructeur de copie. Exemple traité : écriture d'un objet ' matrices '.
4. **Structuration objet d'un code** - Héritage. Classes abstraites. Applications à une problématique concrète de mathématiques pour la finance. Exemple traité : génération de nombres aléatoires et méthode de MonteCarlo ; écriture d'un pricer.
5. **Interface avec Excel** - Dynamic Link Library (DLL). Ecriture et compilation d'une DLL callable depuis Excel ou VBA. Exemple traité : appel du pricer écrit dans les TDs précédents depuis Excel.

Références

ECKEL B. : *Thinking in C++*, disponible gratuitement sur www.mindview.net

STROUSTRUP B. (1996) : *Le langage C++*, Prentice Education.

Eléments logiciels pour le traitement des données massives (OMI309)

Professeur : Matthieu Durut - Lokad

premier semestre

Cours : 0 heures

TP : 16 heures

Objectifs

Ce cours, destiné aux étudiants déjà familiers avec les concepts de base d'un langage orienté objet ou fonctionnel, traite des aspects techniques informatiques que revêtent l'exploration et l'utilisation statistiques de données massives, des thèmes actuellement désignés par le terme 'Big Data'.

Le traitement de ces jeux de données très volumineux soulève de nombreux enjeux informatiques : lorsque les performances sont importantes, de nombreuses abstractions logicielles volent en éclat et une compréhension plus fine de nombreux phénomènes devient alors indispensable. Le cours présente certains de ces aspects : impact de la connaissance du hardware sur la performance, enjeux de la mémoire, qu'elle soit locale ou répartie sur de nombreuses machines, programmation répartie et concurrente, concepts de base d'algorithmie, afin de distinguer et d'estimer ce qui est lent de ce qui est rapide, ce qui est petit de ce qui est grand, ce qui nécessite d'être réparti de ce qui peut être réalisé sur une simple machine.

Plan

1 Hardware

- RAM DDR et SDR.
- Latence des CPUs. Loop unrolling.
- Architectures SMP et NUMA.
- Caches, Cache hit et Cache miss, Cohérence.
- Disques durs HDD et SSD.
- Quelques ordres de grandeurs.

2 Gestion de la mémoire

- Concepts de pile et tas. Call stack. Context switching. Trashing des caches.
- RAII.
- Smart pointers.
- Garbage-collection.

3 Programmation concurrente

- Mutual Exclusion.
- Verrous, Sections critiques, concepts de Starvation Freedom et Deadlock Freedom.
- Verrou de Peterson.
- Race conditions.
- Quelques structures de données thread-safe.
- Variables volatiles, réordonnancement des instructions, Memory Fence, et modèle mémoire C++/C#, Java.

4 Programmation répartie

- GPU.
- Grid-computing. Map-Reduce.
- Cloud Computing.
- Stacks Hadoop et Azure.

- Systèmes de stockage. TSV, SQL, No-SQL par les Key-value store.
- CAP theorem et strong consistency. Impact sur l'architecture et la responsabilité dans les Key-value stores.

5 Algorithmie

- Complexité espace/temps.
- Tris (Heapsort, Quicksort).
- Table de hachage et structures de données classiques.
- Binary Search Tree et Red Black Tree.
- Quelques algorithmes sur les chaînes de caractères.

6 Outils de développement

- Versionning de source.
- Intégration continue.
- Tests unitaires et TDD.
- Probing.
- Gestion des tickets.

7 Eléments d'architecture logicielle

- La POO revisitée.
- Langages fonctionnels.
- DDD et CQRS.

Le cours sera évalué par un projet informatique orienté mathématiques financières ou statistiques, au choix des étudiants. L'objectif du projet sera de donner une dimension concrète à des concepts déjà traités plus abstraitement dans des cours de statistiques ou de finance, en utilisant une partie des outils introduits durant les cours.

Références

ECKEL B. : Thinking in C++, disponible gratuitement sur <http://www.mindview.net>.

Meyers S. : Effective C++

Alexandrescu A. : Modern C++ Design : Generic Programming and Design Patterns Applied

Drepper U. : What every programmer should know about memory

Shavit N. et Herlihy M. : The Art of Multiprocessor Programming

Statistique bayésienne (SE301)

Professeur : Judith Rousseau - ENSAE - CREST

premier semestre

Cours : 18 heures

TP : 0 heures

Objectifs

Ce cours présente les motivations de l'analyse statistique bayésienne, à la fois par rapport à la théorie de la décision et ses différentes notions d'optimalité (minimaxité, admissibilité, invariance) et en termes d'utilisation de l'information a priori disponible. Il considère ensuite les méthodes de modélisation a priori et de calcul des estimateurs de Bayes en estimation ponctuelle et pour les tests d'hypothèses. Les divers concepts seront illustrés dans le cadre des modèles linéaires généralisés, afin de montrer l'applicabilité et la pertinence de l'approche bayésienne. Le cours cherche à approfondir un point particulier par séance, les bases devant être acquises par une lecture préalable du chapitre correspondant.

Plan

1. **Théorie de la décision.** Définitions, modèles et motivations. Minimaxité, règles maximins et lois les moins favorables. Admissibilité et classes complètes. Invariance et meilleurs estimateurs équivariants.
2. **Modélisation des informations a priori.** Représentation probabiliste des informations. Choix de lois conjuguées. Extension aux mélanges de lois conjuguées. Cadre non-informatif et lois de référence. Sensibilité des réponses à la loi a priori.
3. **Inférence bayésienne.** Estimation ponctuelle. Le cas particulier des modèles de régression. Tests d'hypothèses et comparaison avec l'approche de Neyman-Pearson. Méthodes de calcul des estimateurs de Bayes.

Références

Berger J.O. (1985), Statistical Decision Theory and Bayesian Analysis. Springer Verlag, New-York [21 BER 03 A].

Bernardo J.M. et A.F.M. Smith (1994) Bayesian Theory. Wiley, New-York [21 BER 02 A].

Robert C. (1992) L'Analyse Statistique Bayésienne. Economica, Paris [21 ROB 01 B].

Robert C. (1994) The Bayesian Choice. Springer Verlag, New-York [21 ROB 01 C].

Robert C. (1996) Méthodes de Monte-Carlo par chaînes de Markov, Economica, Paris [21 ROB 01 D]

Apprentissage et datamining (SE306)

Professeur : Arnak Dalalyan - ENSAE - CREST

premier semestre

Cours : 14 heures

TP : 10 heures

Objectifs

Ce cours a pour objectif de présenter les bases de l'apprentissage statistique en se focalisant essentiellement sur deux problématiques de l'apprentissage supervisé : classification binaire et régression. Les algorithmes d'apprentissage les plus couramment utilisés seront présentés et leurs propriétés statistiques seront discutées. Ces algorithmes seront appliqués aux données réelles lors des séances de TP.

Plan

- Problématique et exemples
- Minimisation du risque empirique
- Algorithme des plus proches voisins
- Arbres de décision
- Forêts aléatoires
- Machines à vecteurs de support
- Boosting
- Algorithme EM pour l'estimation d'un mélange de gaussiennes

Références

1. Luc Devroye, Laszlo Györfi and Gabor Lugosi. A Probabilistic Theory of Pattern Recognition. Springer 1996.
2. Trevor Hastie, Robert Tibshirani and Jerome Friedman. The Elements of Statistical Learning : Data Mining, Inference, and Prediction. Springer 2009.
3. Williams, Graham. Data Mining with Rattle and R : The Art of Excavating Data for Knowledge Discovery Springer 2011.
4. Luis Torgo. Data Mining with R : Learning with Case Studies. Chapman & Hall/CRC Data Mining and Knowledge Discovery Series 2010.

Duration Models (SE307)

Professeur : Olivier Lopez - ENSAE ParisTech - Université Paris VI

premier semestre

Cours : 14 heures

TP : 0 heures

Objectifs

In this course, we introduce the main statistical tools used to model and infer duration phenomena commonly encountered in actuarial science. Elaborating lifetables, designed to reflect the mortality of an insured population, is the main example that we consider, although additional examples coming from both life and non-life insurance are also proposed. In the last part of this course, we consider prospective models used to analyze and forecast the evolution of mortality through time, which is a key issue in evaluating the longevity risk.

Experience obtained : at the end of this course, the students will know how to

- construct models adapted to describing duration phenomenon
- perform statistical estimation in these models, taking their specificities into account
- design lifetables reflecting the mortality of an insured portfolio
- use prospective models to explain and forecast the evolution of mortality

Plan

1. **General concepts in survival analysis.** Characterizations of the distribution of lifetimes. Incomplete observations : censoring and truncation.
2. **Non parametric estimation.** Kaplan-Meier estimator of the survival function. Nelson-Aalen estimator of the cumulative hazard rate. Estimation of death rates.
3. **Smoothing techniques.** Moving averages, splines. Whittaker-Henderson smoothing. Bayesian smoothing.
4. **Parametric modelling.** Maximum likelihood estimation for censored and truncated lifetimes. Estimation from death rates. Classical distributions used to model lifetimes. Goodness-of-fit procedures.
5. **Prospective models.** Lee-Carter model. Cairns-Blake-Dowd model. Extensions.

Références

Delwarde, A., Denuit, M. (2006) *Construction de tables de mortalité périodiques et prospectives*, Paris, Ed. Economica.

Kalbfleisch, J.D., Prentice, R.L. (2002) *The statistical analysis of failure time data, Second Ed.* New-York, Wiley.

Planchet, F., Thérond, P. (2011) *Modélisation statistique des phénomènes de durée, applications actuarielles*, Paris, Ed. Economica.

Advanced Econometrics of Qualitative Data (SE309)

Professeur : Thierry Kamionka - ENSAE-CREST

premier semestre

Cours : 14 heures

TP : 0 heures

Objectifs

L'objectif de ce cours est de présenter les modèles non linéaires (logit, probit, tobit) sur données de panel. Souvent, sur de telles données, la variable modélisée est qualitative ou censurée. C'est le cas, par exemple, lorsque le salaire ne peut être observé que s'il dépasse un seuil ou que l'on cherche à modéliser le résultat d'une succession de choix (mode de transport, bien consommé, décision d'activité, technologie de production, filière d'éducation, actif financier). Deux approches sont alors communément utilisées (effets fixes et effets aléatoires). Ces deux modes de traitement de l'hétérogénéité individuelle sur données de panel comportent des avantages et inconvénients respectifs. Dans le cadre de ce cours, sont présentées différentes méthodes d'estimation dont certaines sont basées sur des simulations.

Plan

1. Rappels sur les modèles à variables qualitatives ou censurées. Modèles binaires. Modèles polytomiques.
2. Modèles à effets fixes sur données de panel. Vraisemblance conditionnelle. Extension et généralisation
3. Modèles à effets aléatoires sur données de panel. Modèles univariés. Application au modèle probit.
4. Modèles dynamiques. Présentation. Différentes techniques d'estimation par simulation
5. Les modèles bivariés. Modélisation en coupe. Modélisation en panel.
6. Tests de spécification. Test du Score. Application.

Références

- Arellano M. et R. Carrasco (2003), “*Binary Choice Panel Data Models With Predetermined Variables*”, *Journal of Econometrics*, 115, pp. 125-157.
- Chamberlain G. (1985), *Panel Data*, in *Handbook of Econometrics*, vol. 2, édité par Z. Griliches and M.D. Intriligator. North-Holland . [28 GRI 00 A]
- Gouriéroux C. et A. Monfort (1997), *Simulation-Based Econometric Methods*, Oxford University Press. [28 GOU 00 E]
- Heckman J.J. (1981), “*The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete time-Discrete Data Stochastic Process*”.
- Honoré B. et E. Kyriazidou (2000), “*Estimation of Tobit-Type Models With Individual Specific Effects*”, *Econometric Reviews*, Vol. 19, 3, pp. 341-66.
- Honoré B. et E. Kyriazidou (2000), “*Panel Data Discrete Choice Models With Lagged Dependent Variables*”, *Econometrica*, Vol. 68, 4, pp. 839-74.
- Hyslop D. (1999), “*State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women*”, *Econometrica*, vol. 65, 6, 1255-94.
- Magnac T. (2004), “*Panel Binary Variables and Sufficiency : Generalizing Conditional Logit*”, *Econometrica*, Vol. 72, 6, 1859-76.

- Wooldridge J.M. (2005), “*Simple Solution To Initial Conditions In Dynamic, Nonlinear Panel Data Models With Unobserved Heterogeneity*”, Journal of Applied Econometrics, vol. 20, 1, pp. 39-54.

Modèles statistiques dynamiques à variables cachées (SE310)

Professeur : Jean-Michel Zakoïan - ENSAE-CREST

premier semestre

Cours : 14 heures

TP : 0 heures

Objectifs

Les modèles dynamiques faisant intervenir des variables inobservables, ou facteurs, forment une classe très riche et très importante dans les domaines économiques et financiers. Parmi les modèles appartenant à cette classe on peut citer : les modèles d'espace d'états linéaires (modèles à facteurs linéaires, modèles à composantes inobservables...), modèles dynamiques avec données manquantes ou à changement de régime, modèles à changement de temps, modèles ARCH à facteurs, modèles à volatilité stochastique. Les facteurs, peuvent avoir une interprétation économique ou être des outils purement statistiques. La difficulté fondamentale rencontrée est que, sauf pour quelques modèles simples, les méthodes usuelles d'estimation et de lissage (i.e. d'approximation des facteurs) conduisent à des problèmes numériques redoutables. L'objectif du cours est d'étudier les principales méthodes permettant de résoudre ces problèmes : les méthodes récursives, pour les modèles espace d'états linéaires et les modèles à changements de régimes, et les méthodes de moments et méthodes simulées pour tous les autres types de modèles.

Plan

1. **Introduction** - Exemples de modèles dynamiques à variables cachées.
2. **Les modèles à changements de régime markoviens** - Définition, exemples. Stationnarité. Filtrage, lissage et prévision avec l'algorithme d'Hamilton. Propriétés asymptotiques de l'estimateur du maximum de vraisemblance dans le cas AR à changement de régime.
3. **Les modèles espace-états linéaires** - Filtrage et lissage de Kalman (filtre de covariance et d'information). Prévision, estimation et tests. Application au modèle à volatilité stochastique.
4. **La méthode des moments généralisée** - Principe de la méthode. Utilisation des représentations ARMA. Exemples d'applications.
5. **Les techniques de simulation de variables aléatoires** - Méthode d'inversion. Méthode d'acceptation-rejet. Méthode de Monte-Carlo par chaînes de Markov (MCMC). Algorithme de Gibbs. Algorithme d'Hastings-Metropolis. Algorithmes hybrides.
6. **Méthodes des moments simulés et extensions** - Méthode des moments simulés. Inférence indirecte. Méthode du pseudo-maximum de vraisemblance simulé.
7. **Méthodes du maximum de vraisemblance simulé, approche bayésienne** - Importance sampling. Algorithmes MCMC dans le cadre classique et dans le cadre bayésien.

Références

GOURIEROUX C. and MONFORT A. (1996) : *Simulation based econometric methods*, Oxford Univ. Press.
HALL A. (2005) : *Generalized Method of Moments*, (Advanced Texts in Econometrics), Oxford Univ. Press.
HAMILTON J. (1994) : *Time series analysis*, Princeton University Press.
ROBERT C. (1996) : *Méthodes de Monte-Carlo par chaînes de Markov*, Economica.

Macroéconométrie (Séries Temporelles Avancées) (SE312)

Professeur : Catherine Doz - Université Paris I

premier semestre

Cours : 18 heures

TP : 4 heures

Objectifs

L'objectif principal du cours est l'étude des modèles VAR non stationnaires cointégrés. Les modèles VAR permettent de modéliser simultanément les comportements dynamiques d'un petit nombre de séries, et sont utilisés notamment pour faire de la prévision, pour évaluer l'impact des chocs sur les différentes variables ou pour tester des hypothèses structurelles. Le cours commence par des rappels et compléments sur les modèles VAR stationnaires, puis présente les techniques spécifiques au cadre non stationnaire lorsque des relations de cointégration existent entre certaines variables.

Plan

- Processus VAR stationnaires : rappels (processus vectoriels stationnaires, VAR stationnaires et représentation canonique, processus des innovations, estimation, tests, utilisation en prévision). Tests de causalité, fonctions impulsion-réponse, modèles VAR structurels.
- Processus vectoriels intégrés, notion de cointégration, approche en 2 étapes de Engle-Granger, forme ECM d'un VAR cointégré, représentation de Beveridge-Nelson et notion de tendances communes, procédure d'estimation de Johansen (EMV), tests portant sur le nombre de relations de cointégration et la forme du terme constant, tests de restrictions linéaires sur la matrice de cointégration et les paramètres d'ajustement, fonctions impulsion-réponse, modèles structurels.

Références

- BROCKWELL P.J. et DAVIS R.A. (1990). Time Series. Theory and Methods. Springer-Verlag.
- GOURIEROUX C. et MONFORT A. (1995). Séries temporelles et modèles dynamiques, 2ème ed., Economica.
- HAMILTON J.D. (1994). Time Series Analysis, Princeton Univ. Press.
- JOHANSEN S. (1995). Likelihood-based inference in cointegrated Vector Auto-Regressive models, Oxford Univ Press.
- JUSELIUS K. (2006). The cointegrated VAR model, Oxford Univ Press.
- LÜTKEPOHL H. (2005). New Introduction to Multiple Time Series Analysis, Springer Verlag.

Sondages (SE314)

Professeur : Olivier Sautory - INSEE

premier semestre

Cours : 18 heures

TP : 8 heures

Objectifs

L'objectif de ce cours est d'introduire les principaux concepts de la théorie des sondages en soulignant la nature particulière de l'aléa et l'importance des calculs de précision, et de présenter les plans de sondage les plus classiques. Le rôle de l'information auxiliaire sera mis en exergue, que ce soit lors de la phase d'échantillonnage ou lors de l'estimation. Ce cours sera illustré par des exemples d'enquêtes, et les procédures de sondages de SAS seront brièvement présentées.

Principaux acquis de la formation : à l'issue du cours, l'étudiant saura

- Définir un plan de sondage (méthode et algorithme) adapté au problème posé ;
- Calculer la précision des estimateurs (moyenne, total, proportion) issus d'une enquête par sondage ;
- Utiliser des variables auxiliaires pour améliorer la qualité des estimateurs ;
- Repondérer les données pour corriger de la non-réponse totale et mettre en place des stratégies d'imputation pour corriger de la non-réponse partielle.

Plan

1. **Généralités sur les enquêtes par sondage** - Les bases de sondage. La notion d'estimation et de précision. Les différents types d'erreur : erreur d'échantillonnage, erreur de mesure, non-réponse.
2. **Sondage aléatoire simple** - Estimation d'une moyenne, d'un total, d'une proportion. Calcul et estimation de la précision. Détermination de la taille de l'échantillon. Estimation d'un ratio. Estimation sur domaine.
3. **Sondage à probabilités inégales** - Estimation d'un total, d'une moyenne, précision. Choix des probabilités de tirage, cas du tirage à probabilités proportionnelles à la taille.
4. **Stratification** - Estimation, précision. Allocation de l'échantillon entre les strates : allocation optimale, allocation proportionnelle, etc. Constitution des strates : choix des variables de strates, choix du nombre de strates, etc.
5. **Sondage à plusieurs degrés** - Le sondage en grappes : estimation d'un total, précision. Cas d'un sondage aléatoire simple de grappes. Comparaison avec le sondage aléatoire simple. La taille des grappes. L'effet de grappes. Le sondage à deux degrés : estimation d'un total, précision. Cas d'un sondage aléatoire simple à chaque degré. Comparaison avec le sondage aléatoire simple. Les sondages auto-pondérés.
6. **Les sondages empiriques** - La méthode des quotas (principes, 'biais', 'précision'). La méthode des itinéraires. La méthode des unités-types. Le volontariat.
7. **Estimation par le ratio, post-stratification** - Définitions, propriétés, comparaison avec le sondage aléatoire simple.
8. **Estimation par régression, calage** - Estimation par différence. Estimation par régression : définition, propriété. Aperçu sur les méthodes de calage.

9. **Correction de la non-réponse** - Aperçu sur les méthodes de traitement de la non-réponse totale (méthodes de repondération) et les méthodes de traitement de la non-réponse partielle (méthodes d'imputation).

Références

- Ardilly P. (1994) : Les techniques de sondage, Technip, Paris [29 ARD 00 A]
Cochran W.G. (1977) : Sampling techniques (3ème édition), J. Wiley, New-York [29 COC 00 A]
Deroo M., Dussaix A.M. (1980) : Pratique et analyse des enquêtes par sondage, PUF [29 DER 00 A]
Desabie J. (1966) : Théorie et pratique des sondages, Dunod, Paris [29 DES 00 A]
Grosbras J.-M. (1987) : Méthodes statistiques des sondages, Economica, Paris [29 GRO 00 A]
Raj (1968) : Sampling theory, Mc Graw-Hill, New-York [29 RAJ 00 A] Sarndal C.E., Swenson B., Wretman J. (1997) : Model assisted survey sampling, Springer [29 SAR 00 A]
Tillé Y. (2001) : Théorie des sondages, Dunod [29 TIL 00 A]

Statistique asymptotique et robuste (SE315)

Professeur : Patrice Bertail - Université Paris Ouest Nanterre - CREST premier semestre
Cours : 18 heures
TP : 0 heures

Objectifs

L'objectif de ce cours est de familiariser les étudiants à la théorie asymptotique à la LeCam et de leur permettre une étude systématique des modèles semi-paramétriques. Un modèle semi-paramétrique est une famille de loi indexée par un paramètre de dimension finie et un paramètre de nuisance en général de dimension infinie. De tels modèles apparaissent dans pratiquement tous les types d'applications connues, lorsqu'on renonce à faire des hypothèses paramétriques (par exemple de normalité). Après avoir introduit les idées essentielles, nous donnerons quelques éléments de base sur les processus empiriques. Les résultats sur ces processus permettent notamment d'obtenir la convergence et la normalité asymptotique d'une large classe d'estimateurs. Nous aborderons pour cela quelques notions de différentiabilité des paramètres (vus comme des fonctions sur des espaces de probabilités) permettant de définir la notion de gradient ou de fonction d'influence (voir le cours de statistique robuste) et de généraliser la delta méthode. Nous montrons alors comment il est possible de généraliser la notion de borne de Cramer-Rao pour des modèles semi-paramétriques via la notion d'espace tangent (l'information localement apportée respectivement par le paramètre de nuisance et le paramètre d'intérêt). Ces techniques permettent de construire des estimateurs efficaces pour de larges classes de modèles.

Plan

1. **Introduction.** Introduction et rappel. Les problèmes liés à l'estimateur du maximum de vraisemblance.
Modèles semi-paramétriques et bornes de Cramer-Rao.
2. **Processus empirique.** Outils mathématiques : entropie à crochet, entropie intégrale. Théorèmes de Glivenko Cantelli. Théorèmes de Donsker.
3. **Notions de différentiabilité et méthode delta fonctionnelle.** Fonctionnelles de Von-Mises. Gateaux-Fréchet-différentiabilité. Hadamard différentiabilité par rapport à un cône tangent. Bornes d'efficacité semi-paramétriques. Construction d'estimateurs efficaces.
4. **Statistique asymptotique.** Modèle régulier au sens de LeCam Estimateur du maximum de vraisemblance/ Estimateurs en une étape discrétisés. Contiguïté. Normalité asymptotique locale. Théorème de convolution et efficacité asymptotique.

Références

Bickel P.J., Klaassen, Ritov et J. Wellner (1993), Efficient and adaptive estimation in semi-parametric models, Hopkins. [pas en bibli]
Le cam L. (1986), Asymptotic methods in Statistical Decision Theory, Springer. [21 LE 00 C]
Van der vaart A.(1998), Asymptotic Statistics, Cambridge. [21 VAA 00 A]

Evaluation of Public Policies (SE317)

Professeur : Bruno Crépon - ENSAE-CREST
Pauline Givord - INSEE-CREST

premier semestre
Cours : 24 heures
TP : 0 heures

Objectifs

L'objectif du cours est de présenter l'éventail des méthodes récentes développées pour l'évaluation empirique des politiques publiques. On s'intéressera à l'identification d'un effet causal, et l'accent est en particulier mis sur les problèmes que posent l'hétérogénéité de l'effet du traitement et la présence d'effet de sélection. Une attention particulière sera donnée aux aspects pratiques d'une évaluation (accès aux données pertinentes, hypothèses identifiantes, interprétation des résultats). Le cours présentera les différentes méthodes, en s'appuyant en particulier sur la lecture critique d'articles récents. Il se concentre plus sur la façon d'utiliser les méthodes économétriques que sur les propriétés statistiques de ces méthodes, supposées maîtrisées. On présentera plus les intuitions des méthodes statistiques que sur les démonstrations.

Plan

1. Introduction : paramètres d'intérêt, contexte d'observation et causalité
2. Différences de Différences
3. Experimentation
4. Matching
5. Sélection sur inobservables
6. Variables instrumentales
7. Regression Discontinuities
8. Quantiles
9. Bornes
10. Modèles de durée
11. Externalités et effets de bouclage

Références

Angrist J. et G. Imbens (1994), *Identification and estimation of local average treatment effects*, *Econometrica* vol. n62(2), pp 467-475.
Angrist J., G. Imbens G. et D. Rubin (1996), *Identification of Causal Effects Using Instrumental Variables*, *Journal of American Statistical Association*, Vol. 91 (Jun.), n434, pp 444-455.
Angrist J. et A. Krueger (2001), *Instrumental Variables and the Search for Identification : From Supply and Demand to Natural Experiments*, *Journal of Economic Perspectives*, Vol. 15(4), pp. 69-85.
Angrist J. et A. Krueger (1999), *Empirical Strategies in Labor Economics*, *Handbook of Labor Economics*, Vol. 3, Ed. Ashenfelter and Card. [61 ASH 00 A :]
Bertrand M., E. Duflo et S. Mullainathan (2004), *How much should we trust Differences-in-Differences Estimates ?*, *Quarterly Journal of Economic*, vol. 119 (Feb.), pp. 249-275.

Card D. (2001), *Estimating the Return to Schooling : Progress of some Persistent Econometric Problems*, *Econometrica*, Vol. 69, n5, pp. 1127-1160.

Duflo E. *Empirical Methods*..

Hahn J. ; P. Todd et W. Van der Klaauw (2001), *Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design*, *Econometrica* Vol. 69 (1), pp. 201-209.

Rosenzweig M. et K. Wolpin (2000), *Natural "Natural Experiments" in Economics*, *Journal of Economic Literature*, Vol. XXXVIII, pp. 827-874.

Introduction aux méthodes statistiques en biologie moléculaire (SE318)

Professeur : Catherine Matias - CNRS
Pierre Neuvial - Institut Curie

premier semestre
Cours : 16 heures
TP : 4 heures

Objectifs

Ce cours a pour but de présenter les méthodes statistiques développées dans le cadre des données biologiques issues de technologies haut-débit. La caractéristique principale de cet ensemble de données réside dans la très grande dimension de l'espace des variables, nécessitant à la fois de la modélisation statistique mais aussi le développement d'algorithmes performants.

Le cours abordera un certain nombre de techniques statistiques actuelles : les chaînes de Markov cachées, la théorie des valeurs extrêmes, les tests multiples et la notion de false discovery rate (FDR), les méthodes de segmentation, le 'Lasso' pour la régression de signaux 'sparses', l'analyse de réseaux, etc. Il permettra aussi de se familiariser avec les différents types de données disponibles à l'heure actuelle : séquences, puces cDNA, puces CGH (comparative genomic hybridization), réseaux de régulation géniques, réseaux métaboliques, SNPs (single nucleotide polymorphisms).

Plan

1. (C. Matias) Analyse de séquences biologiques.
 - Introduction à la biologie moléculaire : concepts de base (organisation de la cellule, transcription, traduction), phénomènes d'épissage, signaux épigénétiques, données de séquences.
 - Détection de gènes à partir de chaînes de Markov cachées (HMM).
2. (C. Matias) Génomique comparative : alignement de séquences, algorithme de Needleman et Wunsch, loi du score maximal, modèles pair-Markov cachés, reconstruction phylogénétique.
3. (P. Neuvial) Transcriptomique I : données de puces, détection de gènes différenciellement exprimés, tests multiples.
4. (P. Neuvial) Transcriptomique II.
5. (C. Matias) Réseaux de co-expression de gènes : inférence à partir de données transcriptome, modèles graphiques Gaussiens, inférence en grande dimension, introduction à la régression Lasso.
6. (Pierre N.) Puces CGH (comparative genomic hybridization) et puces SNPs : segmentation de signaux (HMMs et/ou détection de ruptures) + techniques lasso.
7. (C. Matias) Analyse de réseaux biologiques : modèle d'Erdos-Rényi, familles exponentielles (ERGM : Exponential Random Graph model), modèles de mélange (SBM : stochastic block model). Classification des noeuds (détection de communautés, groupes SBM). Applications aux réseaux sociaux et aux réseaux biologiques.
8. (C. Matias) Variants génétiques : allèles, SNP (single nucleotide polymorphism), données castémoïn et données familiales. Tests univariés usuels et tests d'enrichissement.

Références

DURBIN R. ET AL., Biological Sequence Analysis, Cambridge University Press. Cambridge, New-York - 1998 [25 DUR 00 A]
KOLACZYK E.D. Statistical Analysis of Network Data : Methods and Models (Springer Series in Statistics) LAURITZEN S., Graphical Models, Clarendon Press. Oxford - 1996 [28 LAU 00 A]

SPEED T., Statistical analysis of gene expression microarray data, Chapman & Hall. Londres - 2003
E 00 A

Modèles à espace d'état et filtrage particulière (SE321)

Professeur : Nicolas Chopin - ENSAE-CREST

premier semestre

Cours : 18 heures

TP : 0 heures

Objectifs

Les modèles dits à chaîne de Markov cachée (ou à espace d'état), sont des modèles de séries temporelles faisant intervenir un 'signal' (un processus (X_t) markovien décrivant l'état d'un système) observé de façon imparfaite et bruitée sous forme de données, par ex.

$Y_t = f(X_t) + \epsilon_t$. Ces modèles sont très utilisés dans de nombreuses disciplines :

Finance : volatilité stochastique (X_t est la volatilité non-observée)

Ingénierie : pistage de cible (X_t est la position d'un mobile dont on essaie de retrouver la trajectoire ; reconnaissance de la parole (X_t est un phonème)

Biostatistique : Ecologie (X_t =taille d'une population) ; Epidémiologie (X_t =nombre d'infectés).

Le but de ce cours est de présenter les méthodes modernes d'analyse séquentielle de tels modèles, basés sur des algorithmes particuliers (Monte Carlo séquentiel). On traitera notamment les problèmes du filtrage, du lissage, de prédiction, et d'estimation des paramètres.

A la fin du cours, nous évoquerons aussi rapidement l'extension de tels algorithmes à des problèmes non-séquentiels, notamment en Statistique Bayésienne.

Pré-requis :

cours 2A simulation et Monte Carlo, ou cours similaire Les cours de 3A de 'Statistique Computationnelle' et de 'Statistique Bayésienne' sont conseillés mais non obligatoires.

Acquis de la formation :

A la fin du cours, l'étudiant sera en mesure :

d'énoncer les propriétés principales des modèles HMM

de mettre en oeuvre un filtre particulier pour filtrer et lisser un modèle HMM donné

d'estimer les paramètres d'un tel modèle à partir de différentes méthodes

Plan

1. Introduction : définition des HMM (Hidden Markov models), propriétés principales, notion de filtrage, lissage, et prédiction, formules forward-backward.
2. HMM discrets, algorithme de Baum-Petrie
3. HMM linéaire Gaussian, algorithme de Kalman
4. Algorithmes SMC pour le filtrage d'un modèle HMM
5. Estimation dans les modèles HMM
6. Introduction aux applications non-séquentielles des algorithmes SMC

Références

Bibliographie :

Del Moral (2004). Feynman-Kac formulae, Springer.

Cappé, Moulines and Ryden (2010) , Inference in Hidden Markov Models (Springer Series in Statistics)

Computational statistics (SE331)

Professeur : Christian P Robert - Paris-Dauphine - CREST

premier semestre

Cours : 12 heures

TP : 8 heures

Objectifs

Ce cours vise à sensibiliser les étudiants aux méthodes de simulation en présentant les principes et en les adossant sur une programmation en R des méthodes de Monte Carlo et de Monte Carlo par chaînes de Markov (MCMC). Les illustrations sont principalement fondées sur des exemples bayésiens mais les principes dépassent ce cadre pour traiter les modèles complexes envisagés dans les disciplines utilisatrices de la statistique, comme l'économétrie, la finance, la génétique, l'écologie ou l'épidémiologie.

Plan

Simulation de lois de probabilité

Contrôle de convergence

Intégration par la méthode de Monte Carlo

Optimisation stochastique

Méthodes de Monte Carlo par chaînes de Markov

Echantillonneur de Gibbs

Contrôle de convergence pour les algorithmes MCMC

Références

Robert, C. & Casella, G. (2009) *Introducing Monte Carlo Methods with R*, Springer

Développements récents en économétrie semi et non paramétrique (SE332)

Professeur : Xavier D'Haultfoeuille - ENSAE-CREST
Eric Gautier - ENSAE - CREST

premier semestre
Cours : 20 heures
TP : 0 heures

Objectifs

This course is an introduction to recent development in semi and nonparametric models in econometrics. Economic models rarely impose parametric restrictions on error terms. Theoretical restrictions rather take the form of exclusion restrictions, monotonicity... We investigate the consequence of relaxing such parametric assumptions, both in terms of identification and estimation. The course presents methods which are to become classical in applied econometrics, such as quantile regressions or single index models, but also general modern ideas such as control functions in IV models or partial identification.

Plan

- Quantile regression

Motivation, Estimation, Quantile IV.

Quantile restrictions in nonlinear models : binary model, censored regressions.

- Estimation of semiparametric models

Estimation without nuisance parameters

Semiparametric two steps estimators

Sieve estimator

- Instrumental variables in semi or nonparametric models

A benchmark : the linear model.

The estimating equation approach.

The control function approach.

- Semi and nonparametric random coefficients models

Examples : random coefficients in linear, discrete choice, panel data or treatment effects models.

- Partially identified models

Examples of partially identified models : missing data, incomplete models.

Inference on parameters or the identification region.

Références

Chernozhukov V., H. Hong and E. Tamer (2007), Estimation and Confidence Regions for Parameter Sets in Econometric Models, *Econometrica*, 75, 1243-1284.

Horowitz, J. L. (1998), *Semiparametric Methods in Econometrics*, Springer-Verlag.

Imbens, G. and Newey, W. K. (2009), Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity", *Econometrica*, forthcoming.

Koenker, R. (2005) *Quantile Regression*, Econometric Society Monograph Series, Cambridge University Press.

Manski, C. (2003), *Partial Identification of Probability Distributions*, Springer.

Wooldridge, J. W. (2002), *Econometric Analysis of Cross Section and Panel Data*.

Statistical Analysis of Network Data with applications in Marketing (SE334)

Professeur : Eric Kolaczyk - Boston University

premier semestre

Cours : 16 heures

TP : 0 heures

Objectifs

Researchers from a wide range of disciplines, including marketing, are more and more involved with the collection and statistical analysis of network-indexed data, and statistical methods and models are being developed in this area at a furious pace. The goal of this course is to present an overview of the foundations common to the statistical analysis of network data across the disciplines, from a statistical perspective, with an eye towards topics with particular relevance to marketing.

Plan

Network mapping

Network characterization

Community detection

Network sampling

Predicting network structure and attributes

Dynamics of network processes

Références

Kolaczyk, E.D. (2009). Statistical Analysis of Network Data : Methods and Models. Springer, New York.
Readings from the literature on networks and marketing.

Bayesian Data Analysis (TSI304)

Professeur : Andrew Gelman -

premier semestre

Cours : 24 heures

TP : 12 heures

Objectifs

We will cover statistical modeling, inference, and model checking and evaluation from a Bayesian perspective.

When the course is over, students should be able to :

- Understand and use the Bayesian concepts of combination of information, inference, and model checking
- Formulate basic applied statistical models and compute with them.

Plan

Topics include the empirical construction of probabilities, hierarchical models, Markov chain simulation, distributional approximations, regression models, and nonlinear and nonparametric models. We consider examples from social and health sciences, and computation in R and Stan.

Références

- Bayesian Data Analysis, third edition, by Gelman, Carlin, Stern, Dunson, Vehtari, Rubin (to appear)
- Bayesian Data Analysis, second edition, by Gelman, Carlin, Stern, & Rubin, CRC Press (2003)
- Bayesian Core, by Marin and Robert, Springer (2007)

Health Economics (AE308)

Professeur : David Bardey - Universidad del Rosario
Lise Rochaix - Haute Autorité de la Santé

deuxième semestre
Cours : 20 heures
TP : 0 heures

Objectifs

Le cours procède à une analyse micro-économique des systèmes de santé. Il met en regard des travaux empiriques (étude sur données individuelles des comportements d'offre et de demande de soins) et théoriques (applications des modèles de régulation au secteur de la santé). Il examine les réformes engagées récemment en France et à l'étranger et cherche à déterminer si leurs effets sont conformes aux prédictions des modèles théoriques de régulation appliqués au secteur de la santé.

Plan

1. **Assurance, gestion actuarielle et gestion des risques longs**- Assurance maladie et concurrence : mécanisme de marche et neutralité actuarielle ; Tarification actuarielle et gestion des risques longs (risque de reclassification, efficacité ex ante/ex post).
2. **Assurance maladie, antisélection et sélection des risques**-Les réactions des assureurs selon le mode de régulation (approches théoriques et résultats empiriques) ; Sélection des risques (pratiques de sélection, conséquences en terme d'efficacité, régulation).
3. **Assurance maladie et risque moral**-Risque moral ex ante ; Risque moral ex post (aspects théoriques et résultats empiriques).
4. **Assurance maladie et intégration verticale**-Présentation des différentes formes d'intégration verticale ; Analyses théoriques.
5. **Les différents modes de régulation des producteurs de soins**- Les schémas de rémunération : paiements prospectifs et rétrospectif ; La demande induite (aspects théoriques et résultats empiriques) ; Organisation des soins primaires, qualité des soins et prise en charge des pathologies chroniques.
6. **Régulation des dépenses hospitalières**- Tarification à la pathologie : présentation du système français ("tarification à l'activité") ; Analyse théorique de la tarification à la pathologie (équité entre établissements, sélection des patients et prise de décision médicale).
7. **Régulation des dépenses pharmaceutiques**- La régulation du médicament en France ; Innovation, médicaments génériques.

Références

Arrow Kenneth J., "Uncertainty and the welfare economics of medical care", *American Economic Review*, vol. LIII, n.5, dec. 1963 (cf. Jstor)
Culyer A.J. et J.P. Newhouse, *Handbook of Health Economics*, North-Holland, [53 CUL 00 A [Usuel]]
Newhouse Joseph P., *Free For All ? : Lessons from the Rand Health Insurance Experiment*, Harvard University Press, (1996) [73 NEW 00 A]
Phelps Charles E., *Health Economics*, Harper et Collins, New-York, (1992) [63 PHE 01 A]

Empirical Corporate Finance (AE312)

Professeur : Manasa Patnam - ENSAE-CREST

deuxième semestre

Cours : 24 heures

TP : 0 heures

Objectifs

The objective of this course is to present an introduction to the empirical analysis of modern corporate finance. Each part of the courses first discusses and reviews the rationale and ideas behind the theories and then focuses on testing the theory. The course will present a broad range of papers that cover the basic methodologies used in empirical corporate finance. The goal of the course is to provide a solid foundation so that the frontiers of research in corporate finance can be examined.

Plan

1 CAPITAL STRUCTURE

Topics covered in this section include, Modigliani-Miller & firm financial structure, debt instruments and credit market, equity instruments and IPO's, payout policy. The empirical part of this section, discusses, first, the general determinants of firm capital structure (supplementary readings) and subsequently evidence that link firms' leverage policy to structural characteristics of product markets (Eckbo, Chapter 13). Solutions to endogeneity problems that commonly arise in this empirical investigation are also discussed.

2 FINANCIAL DISTRESS : BANKRUPTCY, RE-ORGANIZATION AND LIQUIDITY

Topics covered in this section include, default and bankruptcy in a perfect market, costs of bankruptcy and financial distress, financial distress and firm value, leverage, agency costs, asymmetric information. Empirical work on the costs and benefits of bankruptcy procedure is reviewed, with a discussion on optimal bankruptcy reorganization procedure - how different international bankruptcy law procedures affect firm liquidation.

3 CORPORATE GOVERNANCE & AGENCY COSTS

Topics covered in this section include, ownership structure, managerial incentives, takeovers and buyouts, shareholder value, monitoring by board of directors, regulation. The empirical part of this section is divided in two parts. The first part surveys literature on firm organizational structure, examining the relevance of group/family ownership, shareholder protection and managerial 'entrenchment', and firm networks in affecting various firm policies. The second part discusses differences in incentive structure (compensation, stock options etc.) provided to managers and executives and the factors (risk aversion, productivity, information asymmetry etc.) that affect it.

4 CORPORATE INVESTMENT AND FIRM VALUE

This section reviews the econometric issues surrounding the most popular technique in empirical corporate finance, event studies. Event studies aim to isolate the price impact of information content on corporate actions and are relevant for understanding corporate policy decisions. The section discusses measures of firm value (Tobins q) and how firm related 'events' affect this measure. The section also introduces the growing literature on behavioral corporate finance. The idea is to incorporate behavioral biases of CEO's and managers in the analysis of corporate policy decisions, in particular financing and investment patterns.

Références

The main textbooks used for this course are :

1. Tirole, J., 2005, The Theory of Corporate Finance, Princeton University Press. (Tirole)
2. Berk, J. and DeMarzo, P., 2007,. Corporate Finance, 1st edition, Pearson-Addison Wesley : Boston. (Berk & DeMarzo)
3. Eckbo, B.E., 2008, Handbook of Empirical Corporate Finance, Volume 2, North Holland : Amsterdam. (Eckbo)

Labor economics and Employment Policies (AE318)

Professeur :	Pierre Cahuc - ENSAE-CREST et Ecole Polytechnique	deuxième semestre
	Francis Kramarz - ENSAE-CREST	Cours : 24 heures
	Sébastien Roux - ENSAE-CREST	TP : 0 heures

Objectifs

Contemporary analyses of the labour market emphasise the importance of the movement of jobs and labour. At the heart of these analyses are models that formalise job-seeking behaviours and the processes for matching labour with newly created jobs. The objective of this course is to present these models and show how they can provide a better empirical and theoretical understanding of the determinants of employment, inequalities and the consequences of public intervention in the labour market.

Plan

1. **Labor supply** - The consequences of negative income taxes.
2. **Labor demand** - The consequences of low wage subsidies. The sources and consequences of hiring and firing costs.
3. **Job search** - Estimation of duration models.
4. **Job search in equilibrium : matching models** - Job creations and job destructions. The theory and empirics of wages.
5. **Equilibrium search with on-the-job-search and wage determination** - The estimation of equilibrium search models.
6. **Efficiency wages** - Evidence about efficiency wage.
7. **Collective bargaining and wage differentials** - Inter-industry wages differential. Offshoring and employment.
8. **Minimum wage** - Minimum wages in France and in the US.
9. **Discrimination** - Discrimination in the US.

Références

Ashenfelter, O., D. et Card (éd.), *Handbook of Labor Economics*, volumes 3A, 3B, 3C, Amsterdam : Elsevier Science, North Holland.

Cahuc, P., et A. Zylberberg (2004), *Labor Economics*, MIT Press.

Pissarides, C. (2000), *Equilibrium Unemployment Theory*, 2ème éd, MIT Press.

Macroéconométrie appliquée (AE319)

Professeur : Pierre-Olivier Beffy - EXANE
Eric Dubois - INSEE
Hervé Le Bihan - Banque de France

deuxième semestre
Cours : 20 heures
TP : 0 heures

Objectifs

L'objectif de ce cours est d'une part, d'introduire un ensemble de techniques statistiques et de modélisations utilisées dans la construction et l'estimation de modèles macroéconomiques, et d'autre, part de décrire les conditions de leur usage, à la fois dans la recherche macroéconomique empirique et dans un contexte opérationnel de prévision. Les aspects techniques d'économétrie, en général abordés dans d'autres cours de deuxième ou troisième année, feront peu l'objet de démonstrations.

Après une introduction portant sur l'historique et les différents objectifs de la modélisation macroéconométrique (tests de théories économiques, prévisions, simulations de politique économique), les différents types d'approches et de méthodes seront présentés. Elles seront illustrées par des exemples dans le domaine de la politique économique et de la prévision.

Plan

1. **Introduction** - Décompositions tendance/cycle
2. **Modèles statistiques descriptifs : Modèles VAR, VAR structurels, VECM, approche 'General to Specific'** - Techniques d'estimation. Tests de spécification. Prévision. Tests de causalité, fonctions de réponse.
3. **Modèles macroéconométriques structurels 'traditionnels'** - Cadre comptable, organisation des équations et des blocs d'équations, spécifications usuelles des équations importantes. Méthodes d'analyse de résolution et propriétés des modèles. Utilisation en variante et prévision. Illustrations.
4. **Modèles à anticipations rationnelles et modèle DSGE** - Méthodes de résolution. Techniques d'estimation (GMM, ML, inférence indirecte, approche bayésienne). Propriétés. Exemples d'utilisation (nouvelle courbe de Phillips, règles et chocs de politique monétaire).

Références

De Jong et C. Dave (2007) Structural macroeconometrics, Princeton University Press
Canova F. (2005) Methods for Applied Macroeconomic Research, Princeton University Press
Hamilton J. (1994) Time Series Analysis, Princeton University Press

Economics of Education : structural approaches (AE340)

Professeur : Christian Belzil - ENSAE-CREST, Ecole Polytechnique deuxième semestre
Cours : 24 heures
TP : 0 heures

Objectifs

The main objective of this course is to introduce the students to a modern analysis of dynamic models aimed at explaining the human capital accumulation behavior. As several empirical papers will be discussed, students will need to be introduced to Dynamic Programming techniques as well as econometric techniques required to implement these models on actual data. There seems to be no textbook which covers all the material. As a consequence, the course will focus on articles.

Plan

1. Introduction to Dynamic Programming
Bellman Equation and the Optimality principle
continuous control variables, Euler equations.
finite vs infinite horizon
Discrete choices
2. Introduction to human capital
Becker, Gary Human capital : A Theoretical and Empirical Analysis with Special Reference to Education, Columbia University Press
The Ben Porath Model
3. An introduction to Mincerian wage regression Models
Mincer, Jacob "Schooling, Experience and Earnings" Columbia University 1974.
Heckman, James, Petra Todd and Lance Lochner, 2002, "Earnings Functions, Rates of Return and Treatment effects : The Mincer Equation and Beyond" NBER Working Paper 11544 and IZA Working Paper 1700
4. The estimation of Structural and Reduced-Form Dynamic Discrete Choice Models
Eckstein and Wolpin (1989) The estimation of Dynamic Stochastic Discrete Choice Models, Journal of Human Resources
Rust, John (1994) "Structural Estimation of Markov Decision Processes" in R. Engle and D. McFadden, Handbook of Econometrics, North-holland : Amsterdam, 3081-4143.
5. Endogenous Schooling Models
Belzil, Christian and Hansen, Jorgen (2002) "Unobserved Ability and the Return to Schooling" Econometrica, vol 70, No 6.
Belzil, Christian and Hansen, Jorgen (2003) "Structural Estimates of the Intergenerational Education Correlation", forthcoming in Journal of Applied Econometrics.
Cameron, Stephen and Heckman, James (1998) "Life Cycle Schooling and Dynamic Selection Bias : Models and Evidence for Five Cohorts of American Males" Journal of Political Economy, 106 (2), 262-333.
Keane, Michael P. and Wolpin, Kenneth (1997) "The Career Decisions of Young Men" Journal of Political Economy, 105 (3), 473-522.

6. Measuring Returns to Schooling

Belzil, Christian (2007) "The return to Schooling in Structural Dynamic Models : European economic review

Card, David (2000) "The Causal Effect of Education on Earnings" Handbook of Labor Economics, edited by David Card and Orley Ashenfelter, North-Holland Publishers.

7. Human Capital and Growth Theory

8. Education, Human Capital and Risk

9. Other topics (School quality, Thye effects of Class Size)

Références

Economie de l'énergie (AE342)

Professeur : Laurent Joudon - EDF
 Timothée Ollivier - EDF
 Jean-Michel Trochet - EDF

deuxième semestre
Cours : 16 heures
TP : 0 heures

Objectifs

Ce cours présente les principaux fondements économiques de l'exploitation des ressources naturelles et des problèmes d'environnement tels que la pollution. Les concepts d'externalité, de bien public (absence de rivalité et d'exclusion) et de défaillance des marchés justifient l'intervention de l'Etat. Les outils de régulation environnementale sont analysés grâce à la notion d'efficacité économique et à l'approche coût-bénéfice. La théorie économique du risque et du comportement en situation d'incertitude permet d'analyser la pertinence du principe de précaution.

Plan

I- Analyse économique de l'environnement
Efficiences des marchés et environnement
Analyse coût-bénéfice
Méthodes d'évaluation de l'environnement
II- Analyse des politiques environnementales
Politiques décentralisées : droits de propriété, responsabilité légale. Le théorème de Coase.
Politiques réglementaires versus instruments économiques de régulation.
III- Gestion des ressources naturelles
Exploitation optimale des ressources non renouvelables
Gestion des ressources renouvelables
IV- Risques environnementaux
Risque et précaution
Valeur d'option versus diversification intertemporelle du risque
Analyse coût-bénéfice en situation d'incertitude

Références

Livres :
R Perman, Y Ma, J McGilvray & MS Common, (2003), Natural resource and environmental economics, (third ed), Pearson Education, Harlow.
Hanley, N., Shogren, J. F., & B. White (2007). Environmental economics in theory and practice. Oxford university press, New York.

Economics of Innovation and Intellectual Property (AE344)

Professeur : Manasa Patnam - ENSAE-CREST

deuxième semestre

Cours : 14 heures

TP : 0 heures

Objectifs

This course provides a helicopter tour of the economics of innovation and intellectual property. The course covers in particular recent events in the area of intellectual property, such as the "patent wars", litigation involving patent trolls, standard essential patents, and the sharing of patent in form of patent pools or commons.

Plan

Topic 1 : Introduction
Topic 2 : Science and university research
Topic 3 : Returns to R&D
Topic 4 : Technology spillovers
Topic 5 : Localization of knowledge
Topic 6 : Patents
Topic 7 : Trademarks and Branding
Topic 8 : Copyright
Topic 9 : The choice between formal and informal intellectual property
Topic 10 : Technology Transfer
Topic 11 : Strategic patenting
Topic 12 : Patent thickets
Topic 13 : Markets for Technology
Topic 14 : Patent Litigation and Enforcement

Références

Greenhalgh C. and M. Rogers (2010) : Innovation, Intellectual Property, and Economic Growth, Princeton University Press.

Scotchmer Suzanne (2006) : Innovation and Incentives, MIT Press.

Topics in Empirical Industrial Organization (AE345)

Professeur : Philippe Février - ENSAE-CREST

deuxième semestre

Cours : 18 heures

TP : 0 heures

Objectifs

The course is dealing with an empirical analysis of IO. The knowledge of IO and Applied Econometrics is preferred but is not required. The primary focus of the course will be on the use of econometric analysis and data both for descriptive and measurement purposes, and to test the predictions of economic theories. Papers will be discussed in detail with an emphasis on data, sources of identification, and estimation techniques.

Plan

1. Overview

Reiss, Peter C., and Frank A. Wolak. 2004. 'Structural Econometric Modeling : Rationales and Examples from Industrial Organization,' Engle, R., and D. McFadden (eds.), Handbook of Econometrics, Volume V. Amsterdam : North-Holland Company., pp. 1-42.

Bresnahan, Timothy. 1989. "Empirical Studies of Industries with Market Power," in Schmalensee and Willig (eds.), Handbook of Industrial Organization, Volume II : 1011-1058. Amsterdam : North-Holland.

2. Estimation Techniques in Empirical IO

Parametric estimation (OLS, IV, GMM)

Non-parametric estimation

3. Demand estimation

Pakes, A. 2003. 'A Reconsideration of Hedonic Price Indexes with an Application to PCs,' American Economic Review 93 (5) : 1578-1596.

Berry, S. 1994. 'Estimating Discrete Choice Models of Product Differentiation,' Rand Journal of Economics 25(2) 242-262.

Berry, Steve, Levinsohn, James, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium,"

Econometrica 63 (July) : 841-990.

Nevo, Aviv. 2001. "Measuring Market Power in the Ready-to-Eat Cereal Industry," Econometrica 69 (2) : 307-322.

4. Auctions

Jean-Jacques Laffont, Herve Ossard, and Quang Vuong. Econometrics of first-price auctions. Econometrica, 63(4) :953-980, July 1995

Emmanuel Guerre, Isabelle Perrigne, and Quang Vuong. Optimal nonparametric estimation of first-price auctions. Econometrica, 68(3) :525- 574, May 2000

Phil Haile and Elie Tamer. Inference with an incomplete model of english auctions. Journal of Political Economy, 111 :1-51, 2003

5. Price discrimination

P. Leslie, "Price Discrimination in Broadway Theater," RJE 35(3), Autumn 2004, pp. 520-541.

G. Crawford, "The Discriminatory Incentives to Bundle in the Cable Television Industry," QME

6(1),

March 2008, pp. 41-78.

Shepard, Andrea. 1991. "Price Discrimination and Retail Configuration," *Journal of Political Economy* 99 (February) : 30-53.

6. **Contracts**

Chiappori, P. and B. Salanie (2000) "Testing for Asymmetric Information in Insurance Markets," *Journal of Political Economy*, 108, 56-78

Amy Finkelstein and James Poterba. Adverse selection in insurance markets : Policyholder evidence from the uk annuity market. *Journal of Political Economy*, 112(1) :183-208, February 2004

7. **Dynamic demand**

Aguirregabiria, V. (1999) "The Dynamics of Markups and Inventories in Retailing Firms," *Review of Economic Studies*, 66, 275-308

Aguirregabiria, V. and Mira, P. Swapping the Nested Fixed Point Algorithm : A Class of Estimators for Discrete Markov Decision Models' *Econometrica*, 2002, 70, pp. 1519-1543.

Aguirregabiria, V. and Mira, P. Sequential Estimation of Dynamic Discrete Games.' *Econometrica*, 2007, 75, pp. 1-53.

Hotz, J. and Miller, R. Conditional Choice Probabilities and the Estimation of Dynamic Models.' *Review of Economic Studies*, 1993, 60, pp. 497-529.

Rust, John. 1987. "Optimal Replacement of GMC Bus Engines : An Empirical Model of Harold Zurcher," *Econometrica* 55(5) : 999-1033.

8. **Entry and Exit**

Berry, S. (1992) "Estimation of a Model of Entry in the Airline Industry," *Econometrica*, 60, 889-917

Bresnahan, T. and P. Reiss (1990) "Entry in Monopoly Markets," *Review of Economic Studies*, 57, 531-553

Tamer, Elie (2003), "Incomplete simultaneous discrete response models with multiple equilibria", *Review of Economic Studies*, 70, 147-165.

9. **Vertical relations**

Berto Villas-Boas, Sofia (2007), "Vertical Relationships Between Manufacturers and Retailers : Inference With Limited Data", *Review of Economic Studies*, 74(2), 625-652.

Bonnet, Céline and Pierre Dubois (2004), "Inference on Vertical Contracts between Manufacturers and Retailers Allowing for Non Linear Pricing and Resale Price Maintenance".

Bonnet, Céline and Pierre Dubois (2007), "Non Linear Contracting and Endogenous Buyer Power between Manufacturers and Retailers : Identification and Estimation on Differentiated Products".

Références

Gestion humaine dans l'entreprise (DES305)

Professeur : Catherine Grandcoing - CNAM

deuxième semestre

Cours : 10 heures

TP : 2 heures

Objectifs

Dans un contexte économique et social agité, ce cours a pour vocation de travailler sur les stratégies de vie professionnelle. Plusieurs thèmes seront abordés :

- Définir son projet professionnel
- Arbitrages et conséquences du 'work life balance'
- Optimiser son auto-présentation
- Savoir décrypter et gérer les codes de l'entreprise

Deux angles complémentaires soutiendront les réflexions menées :

- le nécessaire et subtil équilibre du triangle des trois quotients (intellectuel, émotionnel et corporel)
- la carrière professionnelle gérée sous forme de projet à long terme : objectifs professionnels et privés, étapes de la vie, re-programmations professionnelles

A partir de situations professionnelles concrètes (entretien de recrutement, point d'évaluation de carrière, choix de mobilité) abordées sous forme de jeux de rôles, travail de réflexions collectives sur la gestion de carrière, l'insertion professionnelle et l'optimisation de ses propres ressources humaines.

Le déroulement des séances sera fondé sur une participation active des étudiants permettant l'élaboration commune d'enseignements pratiques.

Un projet réalisé en groupe d'élèves complétera l'enseignement : le thème de ce travail sera choisi lors des séances de cours et s'appuiera sur des interviews d'anciens, des références bibliographiques et une réflexion personnelle.

Plan

Références

Sociologie et enquête statistique (DES313)

Professeur : Thomas Amossé - Centre d'études de l'emploi
Louis-André Vallet - ENSAE-CREST - MK3

deuxième semestre
Cours : 16 heures
TP : 0 heures

Objectifs

Le cours est centré sur l'enquête statistique en sciences sociales, à partir de l'étude de cas récents. Chaque séance comprendra deux parties pour lesquelles assiduité et participation des étudiants sont requises. Pour des raisons d'organisation, ce cours accueille au maximum 35 étudiants.

La première partie prendra la forme d'un séminaire. Elle sera bâtie sur l'analyse d'une enquête menée dans la statistique publique française ou dans un programme international. Pour chaque enquête, un sous-groupe (d'au plus 5 étudiants) rassemblera, avec l'aide des enseignants, les principales pièces du dossier. Le cas échéant, il prendra contact avec le responsable de l'enquête (qui pourra être invité par les enseignants). Dans un exposé d'au plus 45 minutes, les étudiants présenteront la problématique de l'enquête, les traits originaux de sa méthodologie, ses résultats publiés les plus marquants. La perspective adoptée sera critique et constructive. L'exposé, bâti autour d'une présentation Powerpoint, fera l'objet d'une note écrite de synthèse (5 pages). Il sera éventuellement accompagné d'une note écrite d'analyse de publications (5 pages). Une discussion avec l'ensemble des participants du séminaire s'ensuivra.

Dans la seconde partie du cours, les enseignants traiteront une question de méthodologie générale en rapport avec l'enquête statistique en sciences sociales. L'exposé, d'environ une demi-heure, s'appuiera sur la présentation d'un article remis aux étudiants et sera suivi d'une discussion finale.

Plan

- **Enquêtes analysées en 2006 et 2007** : Les enquêtes FQP « Formation et Qualification Professionnelle » (INSEE), L'enquête « Passage à la retraite des immigrés » (CNAV), Les enquêtes « Sans domicile » (INED et INSEE), Les enquêtes « Emploi » (INSEE), L'enquête « Familles et employeurs » (INED), Les enquêtes « Conditions de travail » (DARES), L'Echantillon Démographique Permanent (INSEE), L'enquête SHARE (Survey of Health, Ageing and Retirement in Europe).
- **Thèmes des exposés de seconde partie** : Construction et mise en oeuvre d'une enquête dans la statistique publique en France ; Enquête par questionnaire et enquête par entretien ; Les différents modes de collecte et leurs effets ; La formulation des questions d'enquête et ses effets ; L'exploration de champ et la démarche prosopographique ; Analyse factorielle et analyse de régression en sociologie ; Le raisonnement empirique : donner un sens aux données ; Causalité, statistique et sociologie.

Références

BEAUD S. et WEBER F. (2003). *Guide de l'enquête de terrain*, Paris, La Découverte. [91 BEA 00 B]
COMBESSIE J.-C. (2003). *La méthode en sociologie*, Paris, La Découverte.[pas en bibli]
GHIGLIONE R., MATALON B. (1998). *Les enquêtes sociologiques : théories et pratiques*, Paris, Colin. [91 GHI 00 A]
GOLDTHORPE J.H. (2000). *On Sociology : Numbers, Narratives, and the Integration of Research and Theory*, Oxford University Press. [91 GOL 00 A]

- GROVES R.M., COUPER M.P. (1998) *Nonresponse in Household Interview Surveys*, London, New York, Wiley. [29 GRO 02 A]
- MAXIM P.S. (1999) *Quantitative Research Methods in the Social Sciences*, Oxford University Press [91 MAX 00 A]
- QUIVY R. et VAN CAMPENHOUDT L. (1995). *Manuel de recherche en sciences sociales*, Paris, Dunod.[91 QUI 00 A]
- SCHUMAN H. et PRESSER S. (1996). *Questions and Answers in Attitude Surveys : Experiments on Question Form, Wording, and Context*, Thousand Oaks, London, New Delhi, Sage.[29 SCH 00 A]
- SINGLETON R.A. et STRAITS B.C. (2005). *Approaches to Social Research*, New York, Oxford University Press.[91 SIN 01 A]

Séminaire de sociologie quantitative (DES316)

Professeur : Ivaylo Petev - Stanford University, Department of Sociology, et C deuxièmement semestre

Cours : 20 heures

TP : 0 heures

Objectifs

Ce séminaire est destiné aux étudiants qui envisagent de pratiquer la sociologie quantitative, à ceux qui souhaitent utiliser ses méthodes dans d'autres disciplines comme le marketing ou la gestion et à ceux qui veulent simplement être capables de pratiquer une lecture critique de travaux de sociologie quantitative. Il est fondé sur la lecture et la discussion d'articles de sociologie empirique mettant en œuvre des méthodes statistiques, très simples ou plus sophistiquées.

Dans certains cas, l'usage des statistiques est mis au service d'une démarche inductive, d'exploration des données. Dans d'autres cas, il contribue à une démarche hypothético-déductive, où on teste un modèle sociologique. On aura l'occasion de se demander si la distinction entre ces deux types d'usage est aussi tranchée qu'on pourrait le croire a priori. On examinera ses liens avec la distinction entre statistique descriptive et économétrie.

On examinera systématiquement les conditions du succès (ou de l'échec) de la méthode quantitative : qualité des données et travail sur celles-ci, adéquation de la méthode aux données et à l'objet.

Enfin, on s'interrogera sur les limites de la quantification et l'articulation des méthodes quantitatives et qualitatives.

Chaque séance comprendra la présentation critique d'un ou plusieurs articles par un étudiant, suivie d'une discussion générale. Cette discussion suppose que chacun ait lu les articles étudiés.

Plan

1. **Introduction.** Comment lire les papiers. Répartition du travail entre les étudiants.
2. **L'usage exploratoire des méthodes descriptives.** De simples tableaux et graphiques peuvent-ils nous aider à rompre avec des idées préconçues et à explorer la société ? Quel est l'apport de l'analyse de données ?
3. **Des méthodes descriptives avec modèle.** Des modèles sous-jacents aux tableaux croisés ? Peut-on rendre visible et prévisible le lien entre positions sociales, dispositions mentales et actions ?
4. **L'économétrie.** La mise en œuvre de l'économétrie suppose-t-elle une hypothèse préalable ? Modèles ad hoc ou modèles à portée générale ? Le raffinement économétrique apporte-t-il des idées neuves ?
5. **Les frontières de la quantification.** Que faire de la critique des statistiques ? Peut-on ouvrir de nouveaux territoires à la sociologie quantitative ? Quel lien entre la statistique et les autres méthodes d'observation ? La statistique textuelle peut-elle réconcilier « quali » et « quanti » ? Quelle est la place de la sociologie quantitative dans la sociologie en France et à l'étranger ?

Références

*NB : des indications concernant la lecture des *livres figurant ici seront données lors de la première séance.*

- *BAUDELLOT C., ESTABLET R., 2006, *Suicide, l'envers de notre monde*, Seuil [pas en bibli]
- DESROSIERES A., 1978 « Marché matrimonial et structure des classes sociales », *Actes de la recherche en sciences sociales*, no 20
- SAPIRO G., 1996, « La raison littéraire », *Actes de la recherche en sciences sociales*, no 111-112
- VALLET L-A., 1999, « Quarante années de mobilité sociale en France. L'évolution de la fluidité sociale à la lumière de modèles récents », *Revue française de sociologie*, no 1
- BOURDIEU P., 1973, « L'opinion publique n'existe pas », *Les Temps Modernes*, no 318
- PETTITT B. et WESTERN B., 2004, « Mass Imprisonment and the Life Course : Race and Class Inequality in U.S. Incarceration », *American Sociological Review*, no 2
- GODECHOT O. et MARIOT N., 2004, « Structure relationnelle des jurys de thèse », *Revue française de sociologie*, no 2
- BRESSOUX P. KRAMARZ F. et PROST C., 2007, « Teachers' Training, Class Size, and Students' Outcomes : Learning from Administrative Forecasting Mistakes » *Crest Working Paper*
- GOLLAC M., 1997 « Des chiffres insensés ? Pourquoi et comment on donne un sens aux données », *Revue française de sociologie*, no 1
- BESSIERE C., 1997, « Suivre des enquêteurs Insee », *Genèses*, no 29

Econométrie du marketing (DES318)

Professeur : Christine Balagué - Telecom Paris Sud
Lionel Wilner - INSEE - CREST

deuxième semestre
Cours : 16 heures
TP : 0 heures

Objectifs

L'objectif de ce cours est de présenter quelques méthodes économétriques utilisées dans la recherche appliquée en marketing. Il contient 2 parties présentant 2 approches complémentaires, l'une centrée sur l'économie industrielle appliquée au marketing, l'autre sur le marketing opérationnel. Chaque séance combinera aspects théoriques et applications empiriques issues de travaux récents consacrés à la recherche des stratégies de prix optimales. Les exemples couverts incluront : choix de produits, segmentation de clientèle, estimation de la demande, estimation des search costs à partir de la dispersion des prix.

Plan

Partie Marketing opérationnelle

- Comportement du consommateur : modèles stochastiques
- Modèle de prévision de ventes d'un nouveau produit
- Equations structurelles

Partie Economie Industrielle appliquée au Marketing

- Econométrie du choix des consommateurs : décision d'achat, choix entre marques
- Estimation de la demande : modèles à coefficients aléatoires Berry-Levinsohn-Pakes (BLP), applications (céréales, minivan, télévision satellite) et implémentation.
- Estimation des search costs à partir de la dispersion des prix

Références

Partie Economie Industrielle appliquée au Marketing

1. Franses, P.H. and Montgomery, A.L. (2002), *Econometric Models in Marketing*, *Advances in Econometrics*.
- Franses, P.H. and Paap, R. (2001), *Quantitative Models in Marketing Research*, Cambridge University Press.
2. Berry, S. (1994), 'Estimating Discrete Choice Models of Product Differentiation', *Rand Journal of Economics*, Vol. 25(2), pp.242-262.
- Berry, S., Levinsohn, J. and Pakes, A. (1995), 'Automobile Prices in Market Equilibrium', *Econometrica*, Vol. 63(4), pp.841-990.
- Nevo, A. (2001), 'Measuring Market Power in the Ready-to-Eat Cereal Industry', *Econometrica*, Vol. 69(2), pp.307-322.
- Petrin, A. (2002), 'Quantifying the Benefits of New Products : The Case of the Minivan', *Journal of Political Economy*, Vol. 110 (4), pp.705-729.
- Dubé, J.-P., Fox, J. and Su, C.-L., (2011), 'Improving the Numerical Performance of Discrete Choice Random Coefficients Demand Estimation', *Econometrica*, to appear.
- Judd, K.L. and Su, C.-L., (2011), 'Constrained Optimization Approaches to Estimation of Structural Models', *Econometrica*, to appear.

3. Hong, H. and Shum, M. (2006), 'Using price distributions to estimate search costs', *RAND Journal of Economics*, Vol. 37, pp.257-275.
- Moraga-Gonzalez, J.L. and Wildenbeest, M.R. (2008), 'Maximum likelihood estimation of search costs', *European Economic Review*, Vol. 52, pp.820-848.

Démographie (DES319)

Professeur : Carole Bonnet - INED
Aline Desesquelles -
Ariane Pailhe - INED
Anne Solaz - INED

deuxième semestre
Cours : 14 heures
TP : 0 heures

Objectifs

Le débat public contemporain réserve une large place aux questions démographiques, qu'il s'agisse des évolutions de la fécondité et de la mortalité, du vieillissement des populations, de l'intensification des migrations internationales ou encore des changements des structures familiales. La démographie occupe de fait une place centrale dans les politiques publiques, à l'échelle locale comme à l'échelle internationale.

L'objectif de ce cours est de fournir les connaissances théoriques et techniques nécessaires à la compréhension de ces grands enjeux contemporains. Au fil de ces 7 séances thématiques, les principaux concepts et outils de l'analyse démographique seront présentés et mis en relation avec les questions démographiques d'actualité et leurs conséquences économiques et sociales.

Principaux acquis de la formation

A l'issue du cours, l'étudiant saura :

Décrire les différentes sources, les problèmes de mesure des données démographiques et les enjeux démographiques actuels ;

Calculer et interpréter les principaux indicateurs démographiques, maîtriser les différentes approches de l'analyse démographique (transversale et longitudinale) ;

Décrire les principes de la dynamique des populations ;

Synthétiser et avoir un regard critique sur un article scientifique de démographie.

Plan

Séance 1 : Introduction, principes démographiques de base

Séance 2 : Analyse de la mortalité

Séance 3 : Analyse de la conjugalité et de la fécondité

Séance 4 : Dynamique des populations

Séance 5 : Migrations internationales

Séance 6 : Projections démographiques

Séance 7 : Retraites

Références

BLANCHET Didier, LEGROS Florence, Eds (2008) Démographie : tendances, incertitudes, implications, Revue Economique, vol 59, n° 5

CASELLI Graziella, VALLIN Jacques, WUNSCH Guillaume (2001-2006), Démographie : analyse et synthèse. INED -Presses Universitaires de France, 8 volumes.

MESLE France, TOULEMON Laurent, VERON Jacques (dir) 2011, Dictionnaire de démographie et des sciences de la population, Armand Colin, Paris 528p.

LERIDON Henri, TOULEMON Laurent (1997), Démographie. Approche statistique et dynamique des populations. Economica, 440 p.

PRESTON Samuel, HEUVELINE Patrick, GUILLOT Michel (2001), Demography. Measuring and Modeling Population Processes. Blackwell, 291 p.

Sociologie de l'immigration et des inégalités ethno-raciales (DES322)

Professeur : Mirna Safi - CREST et Sciences Po Paris

deuxième semestre

Cours : 14 heures

TP : 0 heures

Objectifs

L'objectif de cet enseignement est de transmettre aux étudiants une bonne connaissance de la recherche sociologique sur l'immigration et ses plus récentes avancées à la fois théoriques empiriques. En plus du traitement des problématiques classiques qui lui sont liées (migration et processus d'intégration), la thématique migratoire sera abordée en tant que phénomène producteur d'inégalités dans une société d'accueil, ces inégalités étant liées à l'origine ou la couleur de la peau.

Longtemps traitée comme un sujet mineur en France, l'immigration se trouve aux fondements de la sociologie américaine et notamment les travaux de l'école de Chicago. Cet enseignement proposera ainsi un passage en revue des grandes étapes de la théorisation sociologique de l'intégration des immigrés avec les théories de multidimensionnalité, de multiculturalisme et de segmentation. On s'intéressera également aux nouveaux développements théoriques sur les migrations internationales, le transnationalisme et les relations inter-ethniques. Dans un second temps, l'enseignement accordera une importance particulière à l'étude des formes structurelles d'inégalités qui peuvent caractériser certains groupes minoritaires, notamment sur les marchés du travail et du logement, à l'école, dans la ville et dans la sphère civique et politique. Cela permettra de traiter la question plus générale des discriminations liées à l'origine ou la couleur de la peau, leurs mesures et les politiques publiques qui permettent de les combattre.

A côté de ces grandes perspectives théoriques, une attention particulière sera prêtée à l'analyse empirique des problématiques migratoires, notamment avec des outils quantitatifs. On s'attardera ainsi sur la définition et la catégorisation statistique des populations immigrées et issues de l'immigration en passant en revue les débats méthodologiques et politiques autour de ces questions. Seront également présentées les grandes enquêtes permettant de traiter l'immigration en France et à l'étranger ainsi que les principales variables permettant une analyse fine du phénomène migratoire et des inégalités qui peuvent en découler. Enfin, l'accent sera mis sur l'apport de l'analyse comparative ; il s'agit de comparer le parcours de plusieurs groupes immigrés dans une société d'accueil mais aussi de comparer l'intégration des immigrés dans différents pays d'accueil (essentiellement France, Etats-Unis, Grande-Bretagne et Allemagne).

Plan

Références

- Alba, R.D. & Nee, V. (2005), *Remaking the American mainstream*, Harvard University Press
- Gordon, M. (1964), *Assimilation in American life : the role of race, religion, and national origins*, Oxford University Press
- Massey, D.S. & Denton, N.A. (1993), *American Apartheid : Segregation and the Making of the Underclass*, Harvard University Press, Cambridge,MA.
- Meurs, D., A. Pailhé et P. Simon (2005), 'Persistance des inégalités entre générations liées à l'immigration : l'accès à l'emploi des immigrés et de leurs descendants en France'. *Population*, vol. 61, n° 5-6, p. 763-802..
- Portes, A. & Zhou, M. (1993), "The New Second Generation : segmented assimilation and its variants", *The Annals of the American Academy of Political and Social Sciences* 530, 74-96.

- Portes, A. (1997), "Immigration Theory for a New Century", *International Migration Review* 31, 799–825.
- Simon, P. (1998), 'Nationalité et origine dans la statistique française'. *Population*, vol. 53, n° 3, p. 541-567.
- Tribalat, M., B. Riandey et P. Simon. (1996), *De l'immigration à l'assimilation. Enquête sur les populations d'origine étrangère en France*, La découverte/INED.
- Waters, M. et K. Eschbach. 1995. 'Immigration and ethnic and racial inequality in the United States'. *Annual Review of Sociology*, vol. 21, n° 1, p. 419-446.
- Waters, M.C. & Jimenez, T.R. (2005), 'Assessing Immigrant Assimilation : New Empirical and Theoretical Challenges', *Annual Review of Sociology* 31, 105–125.
- Zhou, M. (1997), "Segmented Assimilation : Issues, Controversies, and Recent Research on the New Second Generation", *International Migration Review*, 31(4), 975-1008.

Scoring (FA306)

Professeur : Pierre Georges - Crédit Foncier
Damien Jacomy - EDF

deuxième semestre
Cours : 12 heures
TP : 0 heures

Objectifs

La maîtrise du risque de défaillance des emprunteurs est un enjeu majeur pour un établissement de crédit.

L'objectif du prêteur est donc d'identifier, parmi les demandeurs de crédit, les individus les plus risqués (qui ont une forte probabilité de ne pas rembourser leur crédit). Ainsi, la plupart des établissements de crédit utilisent des scores pour estimer le risque du demandeur, que ce soit un particulier pour un crédit auto, un crédit immobilier, ou une entreprise. Les variables utilisées par le score sont les informations caractéristiques du client (âge, revenu, CSP, situation maritale, etc.) et celles du bien financé (prix de la voiture, apport personnel, nombre de pièces du logement, etc.).

Disposant ainsi d'un outil de prévision, le prêteur peut alors piloter son risque par la politique d'octroi basée sur le score : refuser les individus dont la probabilité de défaillance est supérieure à une barre donnée.

L'objectif de ce cours est de présenter la modélisation du risque de défaillance, les méthodes d'estimation, la performance de ces scores et leurs utilisations pratiques par les établissements de crédit.

Plan

1. **Présentation de la problématique** - Modélisation par deux approches : LOGIT, PROBIT ou analyse discriminante.
2. **Méthodes d'estimation** - Recherche des variables explicatives. Variables qualitatives, quantitatives, croisements, estimation par maximum de vraisemblance.
3. **Sélection endogène et réintégration des refusés.**
4. **Mesure de performance.**
5. **Politique d'octroi, suivi de la qualité d'un score, application à la tarification différenciée.**
6. **Scores polytomiques et scores de comportement.**

Références

BARDOS M., *Analyse discriminante : application au risque et au scoring financier*, Dunod. Paris - 2001. [23 BAR 02 A]

CELEUX G. et NAKACHE J. P., *Analyse discriminante sur variables qualitatives*, Polytechnica. Paris - 1994 [21 CEL 00 A]

GOURIEROUX C., *Econométrie des variables qualitatives*, Economica. Paris - 1989 [28 GOU 00 A]

GOURIEROUX C. and JASIAK J., *The econometrics of individual risk : credit, insurance and marketing*, Princeton University Press. 2007.

Copules et applications (FA331)

Professeur : Jean-David Fermanian - ENSAE-CREST

deuxième semestre

Cours : 16 heures

TP : 0 heures

Objectifs

Ce cours étudiera la dépendance entre variables aléatoires via le concept de copules. Ces outils probabilistes, définis dans les années cinquante, ont été redécouverts il y a quelques années, et ont donné lieu depuis à une abondante production scientifique. Les praticiens eux-mêmes ne peuvent ignorer les copules, car la modélisation multivariée est désormais au cœur des problématiques financières. Mais du fait de leur généralité, les copules constituent un outil susceptible d'applications dans de nombreux autres domaines, notamment biologie, médecine, fiabilité etc. Nous introduirons certains concepts et mesures de dépendance, les principales familles de copules et certaines propriétés probabilistes associées. Nous nous intéresserons à l'inférence statistique des copules. Enfin, nous étudierons les applications des copules pour modéliser les risques joints, qu'ils soient des temps de défauts ou des rendements d'actifs.

Plan

- Généralités sur les copules- Définitions de base : densité, copules de survie, théorème de Sklar. Bornes de Fréchet. Familles de copules classiques (Gaussienne, Student, Archimédiennes, Marshall-Olkin, etc). Méthodes de génération de nouvelles copules. Modèles de copules à facteurs. Mesures de dépendances : tail indicators, tau de Kendall etc.
- Statistique des copules- Estimation paramétrique et semi-paramétrique. Estimation non-paramétrique, copule empirique. Choix de la juste copule. Simulation de copules. Problèmes divers : test d'indépendance, bootstrap etc.
- Applications en gestion des risques-Dépendance entre facteurs de risques. Dépendance entre temps de défauts et valeurs d'actifs. Risques granulaires, risques opérationnels
- Applications en valorisation-Valorisation des CDO. Options exotiques sur plusieurs sous-jacents. Le problème de la 'martingalité'.

Références

- Cherubini, U. et Luciano, E. (2004). *Copula Methods in Finance*, Wiley.
- Fermanian, J-D., Radulovic, D. & Wegkamp, M. (2004). The weak convergence of empirical and smoothed empirical copula processes. *Bernoulli*, 847-860.
- Fermanian, J-D. (2005). Goodness-of-fit tests for copulas. *J. Multivariate Anal.*, 95, 119-152.
- Genest, C., Ghoudi, K. & Rivest, L.P. (1993). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions, *Biometrika*, 82, 543-552.
- Joe, H. (2001). *Multivariate models and dependence concepts*, Chapman & Hall.
- Luciano E. et Marena, M. (2002). *Copulae as a new tool in financial modelling*. Working paper.
- Nelsen, R. (2006). *An introduction to copulas*. Springer.
- Shi, J. & Louis, T. (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, 51, 1384-1399.

Projet informatique (OMI301)

Professeur : diff. enseignants -

deuxième semestre

Cours : 0 heures

TP : 10 heures

Objectifs

Il est proposé à chaque élève de troisième année de réaliser un projet de programmation individuel dans le langage de son choix. Ce projet pourra ou non se rapporter à la spécialisation choisie par l'étudiant, et est donc ouvert à l'ensemble de la promotion. Ce projet est néanmoins ambitieux et n'est réalisable qu'à la condition de maîtriser le langage de programmation envisagé. La voie normale d'acquisition de cette maîtrise est bien entendu de suivre ou d'avoir suivi un ou plusieurs enseignements de programmation, mais ceci n'est pas imposé.

Conditions de réalisation - Le projet est individuel. Néanmoins, un projet très ambitieux requérant au plus deux personnes pourra être validé si la tâche de chacun est précisée individuellement. Les projets nécessitant un effectif supérieur devraient faire l'objet d'une proposition de Groupe de Travail. Enfin, un étudiant ne peut réaliser qu'un unique projet informatique dans l'année.

Définition du sujet - Chaque sujet sera dactylographié, contiendra une description précise du projet et détaillera l'environnement de développement retenu (langage, accès base de données éventuels). Un recueil de sujets sera distribué début décembre, mais les étudiants peuvent également proposer un sujet personnel. Le sujet devra être original : en aucun cas, il ne peut y avoir double emploi avec un autre cours du cursus. Le Groupe de Travail ne constitue pas un cas particulier : le projet devra être spécifiquement conçu pour ce cours et demander un travail supplémentaire à l'étudiant. Les sujets seront affectés début février.

Validation du sujet - Tous les sujets proposés par les encadrants ou les élèves seront validés par les enseignants en informatique. Les conditions essentielles sont l'originalité du sujet et sa difficulté, ainsi que la motivation de l'étudiant. Si nécessaire, la connaissance du langage retenu intervient aussi.

Choix du langage - Tous les langages enseignés à l'ENSAE sont possibles, à savoir Python, C++, Java, y compris les langages statistiques (SAS, R, etc.). Néanmoins, d'autres langages de programmation restent possibles, sachant qu'une maîtrise minimale sera alors un pré-requis puisqu'ils ne pourront faire l'objet d'aucun cours. Il est également possible d'inclure une partie gestion de base de données dans la proposition. Nous disposons à l'école d'Oracle, SQLyog et d'Access.

Liens avec les cours de programmation - Les enseignements de langage font l'objet d'un contrôle des connaissances séparé, dans lequel la réalisation du présent projet informatique n'intervient pas.

Suivi - Chaque projet sera suivi par la personne l'ayant proposé. Pour les sujets proposés par les étudiants, si une telle personne n'est pas prévue, les enseignants en informatique en désigneront une. L'encadrant défini avec l'élève le planning des suivis. Le volume du suivi est limité à une dizaine d'heures.

Remise du travail et notation - Tous les projets donneront lieu à la rédaction d'un rapport de présentation et à une soutenance orale devant un jury présidé par une personnalité extérieure à la formation, assistée de la personne chargée du suivi. L'application informatique sera mise à la disposition du jury, qui est tenu aux règles classiques de confidentialité de l'école. Pour permettre au jury d'évaluer dans des délais raisonnables le travail des étudiants, le rapport écrit et l'application informatique devront être remis pour fin mai. Les modalités de cette remise seront précisées en temps utile à chacun. La notation est entièrement confiée au jury qui évaluera les différents aspects du travail : qualité de la réalisation informatique, respect du 'cahier des charges' initial, qualité du rapport écrit et de la soutenance orale.

Base de données et Web (OMI308)

Professeur :

deuxième semestre

Cours : 0 heures

TP : 20 heures

Objectifs

Le cours de base de données et Web a un triple objectif.

En premier lieu, la partie base de données du cours vise à transférer un noyau de connaissances élémentaires permettant une utilisation d'un système de gestion de bases de données à un niveau "utilisateur spécialisé" voire à un niveau "administrateur de bases de données". Le cours comprend donc une présentation générale des fonctionnalités qui doivent être satisfaites par un système de gestion de bases de données, une présentation détaillée du modèle relationnel qui est le modèle sous-jacent aux systèmes de gestion de bases de données actuellement commercialisés et deux aperçus des orientations nouvelles des systèmes de gestion de bases de données à savoir les bases de données déductives et les bases de données orientées objets.

Ensuite, la seconde partie du cours présente la programmation Web. L'ambition est de donner les outils nécessaires à la construction de sites Internet dynamiques sophistiqués, c'est à dire permettant de gérer le contenu des pages en fonction de l'identité d'un internaute (cas d'un relevé de compte), du contexte de la navigation (cas d'un panier d'achat). La démarche suivie sera celle de la construction d'un site, depuis sa conception jusqu'à sa mise en ligne. Comme nous le verrons, la programmation web est ' composite ' : les concepts et les langages seront introduits à mesure qu'ils seront utiles.

Des travaux dirigés permettent de comprendre et appliquer les notions vues dans chacune des parties.

Enfin, l'ensemble des éléments du cours est mis en oeuvre lors d'un projet qui permet de valider l'enseignement.

Plan

1. **INTRODUCTION** : Historique : Systèmes de gestion de fichier - Systèmes de gestion de bases de données ; Fonctionnalités d'un SGBD.
2. **QUELQUES MODELES DE DONNEES** : Modèle Entité - Association, modèle Réseau, modèle Hiérarchique
3. **LE MODELE RELATIONNEL** : Modèle de Données, langages de Manipulation, l'algèbre relationnelle, le calcul, langages relationnels commerciaux ; SQL, QUEL, QBE ; Extension des langages ; Plongement dans un langage de programmation, couplage avec Prolog et langages de règles
4. **CONCEPTION DE SCHEMA** : Motivation : mises à jour et anomalies ; dépendances fonctionnelles ; relations non sous première forme normale ; formes normales ; décomposition ; sans perte d'information ; sans perte de dépendances
5. **CONCURRENCE D'ACCES** : Base cohérente, transactions, maintien de la cohérence lors d'accès concurrents
6. **ORIENTATIONS** : Bases de Connaissance, bases de données orientées objets.
7. **Introduction au Web et à HTML**
8. **Présentation de PHP.**
9. **Réalisation d'un cahier des charges**

10. **Modélisation d'un site**

N.B. : PHP est distribué librement et gratuitement sous licence GNU GPL.

Références

Programmation Web avec PHP, L.Lacroix, N. Leprince, C. Boggero, C. Lauer, Eyrolles [85 LAC 00 A]
JavaScript The Definitive Guide, David Flanagan, O'Reilly [Cf OSI]

Outils pour l'analyse des données massives (OMI310)

Professeur :

deuxième semestre

Cours : 18 heures

TP : 6 heures

Objectifs

Objectif

Ce cours introductif vise à présenter quelques concepts de base, mathématiques et informatiques, qui servent à traiter les ' Big Data '. Au cœur de ces concepts, on insistera sur l'algorithmique distribuée en général, et sur l'optimisation distribuée en particulier. Des séances sur machines permettront de mettre ces concepts en œuvre sur des données réelles. On illustrera ces outils sur des problèmes récents dans le domaine du commerce électronique.

Principaux acquis de la formation

A l'issue de ce cours, l'étudiant saura utiliser un réseau de machines virtuelles pour y distribuer des données et y lancer des calculs de type ' Map/Reduce '. Il saura par exemple tirer parti du cadre ' Map/Reduce ' pour programmer un algorithme descente de gradient distribuée et l'appliquer à l'estimation des paramètres de différents modèles de régression. Sur un plan plus conceptuel, il saura montrer des résultats de convergence pour l'algorithme de descente de gradient distribué et saura également comprendre le rôle du graphe de machines dans la vitesse de convergence grâce à des outils de théorie des graphes.

Plan

Partie I : Rappels

Rappels d'optimisation

Rappels d'apprentissage

Notions de théorie des graphes

Partie II : Notions d'algorithmique distribuée

MPI/MapReduce

Cas d'utilisation typiques

Partie III : Optimisation Distribuée

Enoncés

Algorithmes

Convergence

Partie IV : Expériences sur données réelles

Références

D.P. Berstekas, J.N. Tsitsiklis, Parallel and Distributed Computation : Numerical Methods, Prentice-Hall, 1989.

S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends in Machine Learning, 3 :1-122, 2011

J. Dean and S. Ghemawat. Mapreduce : simplified data processing on large clusters. Communications of the ACM, 51 :107-113, 2008

J.C. Duchi, A. Agarwal, and M.J. Wainwright. Dual averaging for distributed optimization :

Convergence analysis and network scaling. IEEE Transactions on Automatic Control 57(3) :592-606, 2012

Histoire de la statistique (SE2S4)

Professeur : Laurent Davezies - ENSAE-CREST
Emmanuel Didier - CNRS

deuxième semestre
Cours : 18 heures
TP : 0 heures

Objectifs

Le cours sera l'occasion de prendre le contre-pied des approches théoriques de la statistique qui insistent sur le cohérence logique intrinsèque. Nous mettrons en évidence les contingences historiques de leur invention et les contraintes irréductibles de leur pratique. Nous chercherons à décrire diverses formes d'articulation entre les évolutions sociales et celles des outils de quantification et de leurs usages. Comment expliquer que de nouvelles techniques ou institutions apparaissent ? Quels acteurs s'en emparent ? Dans quels contextes politiques et économiques ? Peut-on malgré ces influences sociales parler d'une autonomie des formalismes mathématiques ?

Plan

Séance 1.

Présentation du cours.

Exposé introductif : comment s'articulent changement social et changement de la statistique ?

Choix des dossiers par les élèves.

Séance 2.

La Grande Dépression aux USA : de nouveaux outils statistiques pour observer la société ?

Séance 3.

Intervention d'Alain Desrosières.

Présentation de l'histoire des statistiques.

Séance 4.

Les Lumières et l'Invention des Moindres-Carrés Ordinaires (MCO)

Séance 5.

Intervention de J.J. Droesbeke.

Histoire des statistiques inférentielles.

Séance 6.

Crise de l'Etat providence et décompte de l'exclusion sociale.

Visionnage du film Parole d'enquêtrices, par l'INSEE en deux parties qui porte sur l'enquête Sans Domicile de 2001. Avec des populations particulièrement difficiles, comment échantillonner ? Comment questionner ?

Séance 7.

Histoire récente des statistiques de police en France

Comment une même série statistique, l'état 4001, peut avoir été inventé pendant les années 1970 dans le cadre de l'Etat providence et être réutilisée par le New Public Management.

Séance 8.

Intervention de Michel Armatte.

Crise écologique et modèles de prévision

Séance 9

Exposé des élèves.

Références

ARMATTE M. (1995) Histoire du modèle linéaire. Formes et usages en économie et économétrie jusqu'en 1945. Thèse de doctorat - EHESS - Paris. [pas en bibliothèque]

DESROSIERES A. (1993) La politique des grands nombres. Histoire de la raison Statistique, La Découverte, Paris [20 DES 00 A] et [20 DES 00 B].

DIDIER E. (2010) En quoi consiste l'Amérique. La Découverte, Paris.

HALD A. (1998) A History of Mathematical Statistics from 1750 to 1930, John Wiley, New York [21 HAL 01 A]

INSEE (1987) Pour une histoire de la statistique, vol 1 et 2. [30 INS 00 C]

MORGAN M. (1990) The History of Econometric Ideas, Cambridge University Press [62 MOR 00 A]

STIGLER S. (1986) The History of Statistics. The measurement of uncertainty before 1900, Harvard University Press [30 STI 00 A]

Bootstrap et rééchantillonnage (SE303)

Professeur : Patrice Bertail - Université Paris Ouest Nanterre - CREST deuxième semestre
Cours : 16 heures
TP : 0 heures

Objectifs

L'objectif de ce cours est présenter les techniques de base des méthodes de rééchantillonnage (bootstrap, jackknife etc). Ces méthodes ont connu un développement récent important ces dernières années comme alternatives aux méthodes asymptotiques usuelles. Ces dernières sont basées sur des applications du théorème Central-Limite. Les méthodes de bootstrap proposent, dans des problèmes d'application divers, une alternative à la détermination de distributions d'échantillonnage de statistiques particulières. Ceci permet de résoudre, même dans des situations complexes (modèles paramétriques et non-paramétriques), les problèmes standards d'inférence (biais d'un estimateur, intervalles de confiance, tests d'hypothèses, prédiction). Ces méthodes seront présentées de façon théorique, puis appliquées aux modèles classiques de la statistique et de l'économétrie (modèle linéaire, séries chronologiques etc). Ces méthodes utilisent de façon intensive la puissance de calcul des ordinateurs. On évoquera aussi dans ce cours, d'autres méthodes d'estimation, plus spécifiques, basées sur des techniques de simulation.

Plan

1. Introduction aux principes de base du bootstrap
2. Méthode de Monte-Carlo
3. Estimation du biais d'un estimateur
4. Intervalles de confiance (différentes méthodes)
5. Tests d'hypothèses (calcul de p-valeurs etc)
6. Propriétés théoriques du bootstrap
7. Bootstrap dans les modèles de régression (y compris la prédiction)
8. Le bootstrap itéré
9. Divers thèmes d'application : (séries chronologiques, bootstrap lissé, jackknife etc)
10. Autres méthodes d'estimation par simulation (moments ou maximum de vraisemblance simulés)

Références

Simar, L. (2001), An Invitation to the Bootstrap : Panacea for Statistical Inference ? (une copie de ce handout sera disponible avant le premier cours)
Davison, A.C., Hinkley, D.V. (1997), Bootstrap Methods and their Applications, Cambridge UP. [21 DAV 03 A]
Efron, B. et Tibshirani, R.J. (1993), An Introduction to the Bootstrap, Chapman and Hall, NY. [21 EFR 00 B]
Gouriéroux, C. et A. Monfort (1994), Simulation Based Econometric Methods, CORE Lecture Series, CORE, Université Catholique de Louvain, Louvain-la-Neuve, Belgium [28 GOU 00 E]
Hall, P. (1992), The Bootstrap and Edgeworth Expansion, Springer-Verlag [21 HAL 00 A]

Méthodes statistiques de l'économétrie (SE304)

Professeur : Alain Monfort - INSEE-CREST

deuxième semestre

Cours : 16 heures

TP : 0 heures

Objectifs

Ce cours a pour objectif de présenter les méthodes économétriques récentes sous une forme unifiée. La théorie statistique centrale sera celle des estimateurs extrémaux ainsi que celle des méthodes de tests et régions de confiance qui en découlent. Cette théorie générale sera appliquée à la méthode des moindres carrés non linéaires et des régressions quantiles, aux méthodes du maximum de vraisemblance et du pseudo maximum de vraisemblance, à la méthode des moments généralisée, à la méthode des moindres carrés asymptotiques (par exemple les méthodes de Berkson et du chi-deux minimum) et aux méthodes fondées sur des simulations, en particulier l'inférence indirecte. Des applications à de nombreux modèles sont considérées.

Plan

MODELISATION ECONOMETRIQUE

Modèles paramétriques et semi-paramétriques

Modèles d'échantillonnage, conditionnels statiques et dynamiques

Problèmes statistiques

Construction d'un modèle économétrique

INFORMATION

Information de Kullback

Information de Fisher

ESTIMATEURS EXTREMAUX

Définitions

Convergence

Normalité asymptotique

Tests

Régions de confiance

M-estimateurs

M-estimateurs quasi-généralisés

Modèles dynamiques

BORNES D'EFFICACITE ASYMPTOTIQUES

Borne paramétrique

Borne semi-paramétrique

MOINDRES CARRES NON LINEAIRES, REGRESSIONS QUANTILES

Définitions

Exemples : modèles index, splines, réseaux de neurones

Propriétés asymptotiques

Estimateur robuste de la matrice de variance-covariance asymptotique

Cas de l'homoscédasticité conditionnelle

Méthodes des moindres écarts absolus

- Regressions quantiles
- Applications.
- METHODES DU PSEUDO MAXIMUM DE VRAISEMBLANCE (PMV)
- Méthodes du PMV du 1er ordre (PMV1)
- Famille exponentielle linéaire
- Méthode du PMV1 Quasi généralisée
- Optimalité semi-paramétrique
- Méthode du PMV du 2è ordre (PMV2)
- Famille exponentielle quadratique
- Méthode du PMV du 4ème ordre (PMV4)
- Applications
- METHODE DES MOMENTS GENERALISEE (MG)
- Méthode des moments simple
- Définition et propriétés de la méthode des MG
- Métrique optimale
- Cas de la définition du paramètre par une espérance conditionnelle
- Optimalité semi-paramétrique
- Instruments optimaux
- Méthodes des doubles moindres carrés linéaires et non linéaires
- Applications
- METHODES DES MOINDRES CARRES ASYMPTOTIQUES (MCA)
- Paramètre d'intérêt, paramètre auxiliaire
- Définition et propriétés des MCA
- Méthode optimale
- Méthode de Berkson
- Méthode du chi-deux minimum
- Test d'hypothèses mixtes
- Applications
- METHODES FONDEES SUR DES SIMULATIONS POUR LES MODELES DYNAMIQUES
- Deux types de modèles dynamiques
- Le problème de base
- Modèles espace-état
- Méthodes fondées sur des simulations
- Exemples
- INFERENCE INDIRECTE
- Principes généraux
- Un exercice
- Propriétés asymptotiques
- Méthodes équivalentes
- Applications
- Méthodes des moments simulés

Références

C. Gouriéroux et A. Monfort, ' Statistique et modèles économétriques ', 2 volumes, Economica (1995).
 C. Gouriéroux et A. Monfort, ' Simulation Based Econometric Method, Oxford University Press (1996)

Analyse des données approfondie (SE305)

Professeur : Liliame Bel - AgroParisTech
David Blumenthal - AgroParisTech
Stéphane Robin - INRA

deuxième semestre
Cours : 20 heures
TP : 0 heures

Objectifs

L'objectif de ce cours est de présenter des développements récents de méthodes statistiques qui complètent et étendent les méthodes d'analyse des données vues en première année. Les méthodes d'analyse factorielle et de clustering sont des méthodes descriptives très utilisées en pratique dans l'industrie et les services. Cependant, elles ne permettent pas de faire d'inférence statistique, c'est pourquoi on assiste à un développement de méthodes fondées sur des modèles statistiques. Ce cours présente des extensions des méthodes d'analyse des données usuelles, les modèles pour l'analyse des données fonctionnelles et les modèles de mélange qui permettent de réaliser des classifications non supervisées. De nombreux exemples (analyse sensorielle, données de pollution, réseaux sociaux) illustrent les méthodes présentées.

Plan

1. Traitement de données pour l'analyse sensorielle (6h)
 - a. Analyse des tableaux à 3 dimensions
 - b. Analyse Procrustéenne Généralisée
 - c. Modèles de cartographies des préférences
2. Analyse de données fonctionnelles (6h)
 - a. Représentation, décomposition, lissage et analyse en composantes principales de données fonctionnelles.
 - b. Modèles linéaires fonctionnels : ANOVA, régression fonctionnelle.
 - c. Classification et discrimination de courbes.
3. Modèles de mélange (8h)
 - a. Mélange de lois gaussiennes, algorithme EM
 - b. Choix du nombre de groupes
 - c. Classification non supervisée des sommets d'un graphe. Méthodes variationnelles. Application à l'analyse des réseaux sociaux.

Références

Mardia K.V., Kent J.T. Bibby J.M. Multivariate Analysis, Academic Press, 1979
Bartholomew D.J. Knott, M Latent variable models and factor analysis, Arnold, 1999
Ramsay J.O., Silverman B.W. Applied Functional data Analysis, Second Edition , Springer-Verlag, 2002
Ramsay J.O., Silverman B.W. Functional data Analysis, Springer-Verlag, 2005
Geoffrey McLachlan , David Peel, Finite Mixture Models Wiley Series in Probability and Statistics, 2000

Modèles GARCH et à volatilité stochastique (SE311)

Professeur : Christian Francq - ENSAE-CREST

deuxième semestre

Cours : 14 heures

TP : 0 heures

Objectifs

Ce cours est une introduction à la modélisation des séries financières. Les méthodes d'analyse ont connu une évolution rapide dans les vingt-cinq dernières années. On distingue deux grandes classes de modèles : les modèles GARCH dans lesquels la volatilité est une fonction des observations passées et les modèles à volatilité stochastique faisant intervenir des variables latentes dans la spécification de la volatilité. L'analyse de ces classes de modèles repose sur des concepts probabilistes et des méthodes d'inférence statistique souvent sophistiqués. L'objectif est de présenter les modèles classiques, leurs motivations empiriques, les principales propriétés des solutions et de donner une vue d'ensemble sur les techniques d'inférence statistique adaptées.

Plan

1. **Introduction** - Rappels sur les modèles classiques de séries temporelles. Propriétés des séries financières et insuffisance des modèles ARMA.
2. **Processus conditionnellement hétéroscédastiques** - Processus GARCH : définitions, représentations. Conditions de stationnarité stricte et au second-ordre, propriétés des distributions marginales, prévisions. Asymétries : modèles GARCH exponentiel et GARCH à seuil. Agrégation et modèles GARCH faibles. Inférence statistique des modèles ARMA-GARCH. Identification des ordres. Estimation par MCO et quasi-maximum de vraisemblance : propriétés asymptotiques. Tests d'hétéroscédasticité conditionnelle.
3. **Modèles à volatilité stochastique** - Motivations et lien avec le temps continu. Modèle canonique : stationnarité et représentations ARMA. Méthodes d'estimation : filtre de Kalman, GMM, méthode utilisant les représentations ARMA, méthodes simulées. Modèles à changement de régime markovien.

Références

BOLLERSLEV T.P., ENGLE R.F. and NELSON D.B. (1994) : *ARCH models*, in : R.F Engle and D.L.McFadden, eds., *Handbook of Econometrics*, Vol. 4 Chap 49, 2959-3038.

GHYSELS E., HARVEY A.C. and RENAULT E. (1996) : *Stochastic volatility*, In C.R Rao and G.S. Maddala (eds.), *Statistical Methods in Finance* (Amsterdam : North-Holland), 119-191.

GOURIEROUX C. (1992) : *Modèles ARCH et applications financières*, Economica.

HAMILTON J. (1994) : *Time-Series Analysis*, Princeton University Press.

TSAY R.S. (2002) : *Analysis of financial time series*. Wiley.

STRAUMANN D. (2005) : *Estimation in conditionally heteroscedastic time series models*. Lecture Notes in Statistics, Springer Berlin Heidelberg.

Statistique et économétrie spatiale (SE337)

Professeur : Salima Bouayad Agha -
Lionel Vedrine -

deuxième semestre
Cours : 14 heures
TP : 6 heures

Objectifs

L'utilisation de données spatiales connaît un développement important dans de nombreux domaines : sciences de la terre, environnement et climatologie, épidémiologie, économétrie, analyse d'image, parmi d'autres. Comme pour les séries temporelles, la statistique spatiale se différencie de la statistique classique par le fait que les observations sont dépendantes. Son originalité tient au fait que dans l'espace, les interactions peuvent être multidirectionnelles.

Après avoir passé en revue les différents types de données spatiales, le cours présente les outils permettant de mesurer le degré de signification statistique des configurations et des relations spatiales de données géoréférencées. Le cours s'attache ensuite plus spécifiquement à l'étude des données économiques. Il s'agit d'étendre les méthodes de l'économétrie standard en considérant les principaux problèmes rencontrés dans l'utilisation de ces données (hétérogénéité des observations, interaction spatiale).

De nombreux exemples utilisant R ou STATA illustrent les sujets abordés. Le cours sera complété de 3 TP, l'un sur la cartographie et les méthodes exploratoires des données, les 2 autres sur l'économétrie.

Principaux acquis de la formation

A l'issue du cours, l'étudiant saura :

- Détecter dans les données ou dans les résultats de l'estimation par les MCO la présence d'autocorrélation et/ou d'hétérogénéité spatiale ;
- Enoncer les conséquences sur l'estimateur MCO de la présence d'autocorrélation et/ou d'hétérogénéité spatiale ;
- Mettre en œuvre les méthodes d'estimation et d'inférence adaptée, tester ses résultats et choisir la spécification spatiale la plus adaptée ;
- Interpréter les paramètres associés aux variables spatiales et calculer les multiplicateurs spatiaux ;
- Conduire une analyse économétrique spatiale complète avec des données réelles sous STATA et sous R (spécification, choix de la méthode d'estimation, tests, etc.) et interpréter les résultats.

Plan

PARTIE 1 : Statistique Spatiale

La boîte à outils d'analyse des données spatiales : Matrices de contiguïté, Matrices de pondération spatiale

Analyse exploratoire des données spatiales : Outils de représentation, Tests et indices locaux

PARTIE 2 : Econométrie spatiale sur données en coupe

Auto-corrélation et hétérogénéité spatiale

PARTIE 3 : Introduction à l'économétrie spatiale sur données de panel

Références

- Droesbeke, J.J., Lejeune, M. et Saporta, G. (2006), Analyse statistique des données spatiales, Technip
- Anselin L. (1988), Spatial Econometrics : Methods and Models, Kluwer Academic Publishers
- Anselin, L., Florax, R.J.G.M. et Rey, S.J. (2004), Advances in Spatial Econometrics, Springer
- LeSage, J. et Pace, K.R. (2009), Introduction to Spatial Econometrics, CRC Press
- Anselin, L, Le Gallo, J., et Jayet, J. (2007), Spatial Panel Econometrics, In L. Matyas and P. Sevestre (Eds.), The Econometrics of Panel Data, Fundamentals and Recent Developments in Theory and Practice (3rd Edition). Dordrecht, Kluwer

Statistique pour l'épidémiologie (SE313)

Professeur : Stephan Clemençon - TELECOM ParisTech

deuxième semestre

Cours : 10 heures

TP : 4 heures

Objectifs

Ce cours présente un panorama succinct des modèles stochastiques utilisés en épidémiologie pour décrire l'évolution de maladies infectieuses transmissibles au sein d'une population et des méthodes statistiques adaptées au type de données rencontrées dans ce champ d'application. Il vise en particulier à permettre aux étudiants d'appréhender des modèles markoviens de population stratifiée simples, l'asymptotique 'grande population' prévalant dans les applications statistiques et les outils probabilistes tels que le couplage pour l'analyse de ces modèles.

Plan

Séance 1 - Introduction générale du cours. La modélisation déterministe. Le modèle de Reed-Frost.

Séance 2 -Le modèle SIR standard. Construction et extensions.

Séance 3 - Complément sur les modèles markoviens et le couplage. Applications à l'étude des modèles épidémiologiques

Séance 4 - Asymptotique grande population pour les modèles SIR.

Séance 5 - Modèles épidémiologiques markoviens densité dépendants à sauts

Séance 6 - Estimation statistique .

Séance 7 - Estimation statistique (suite)

Références

MODE C., Stochastic Processes in Epidemiology - HIV/AIDS and Other Infectious Diseases. (2000) World Scientific Publishing Company, Singapore.[17 MOD 00 B]

Andersson H. & Britton, T. Stochastic Epidemic Models and Their Statistical Analysis. (2000) Lecture Notes in Statistics, n°151, Springer. [17 AND 01 A]

Structural Econometrics (SE328)

Professeur : Christian Belzil - ENSAE-CREST, Ecole Polytechnique

deuxième semestre

Cours : 18 heures

TP : 2 heures

Objectifs

The lectures will deal with the following issues :

Plan

1. Discrete Stochastic Dynamic Programming Models (DSDPM). Introduction and definitions, the Bellman equation, introduction to recursive methods.
2. An example of DSDPM. Solving a partial equilibrium search problem, maximum likelihood estimation.
3. General Framework. State Space, Law of Motions, Classification of solution methods, Dimension of integrals, Curse of dimensionality.
4. Direct Solution Methods. Schooling Decisions as an optimal stopping problem, Extreme value-dynamic Logit Model (Rust's method).
5. Computationally Intensive Methods. Simulation Methods, Interpolation Methods.
6. Identification and Estimation. A theorem by Hotz and Miller, the degree of underidentification, restrictions on preferences.
7. Alternative Solution Methods. Estimation by Conditional Choice Probabilities, OLS estimation, Non rational expectations (Expectation Parametrization Method).
8. Applications. Examples of computer programs for various methods.

Références

Bellman, R. (1957), Dynamic Programming, Princeton University Press, Princeton.

Eckstein, Z. et K. Wolpin (1989), 'The Specification and Estimation of Dynamic Stochastic Discrete Choice Models', Journal of Human Resources, 24, 562-598.

Keane, M. P. et K. Wolpin (1997), 'The Career Decisions of Young Men', Journal of Political Economy, 105, 473-522.

Rust, J. (1994), 'Structural Estimation of Markov Decision Processes' in Handbook of Econometrics, Engle, R. et D. McFadden (eds.), Elsevier Science, North-Holland Publishers, Amsterdam, 3081-4143.

Estimation non paramétrique (SE329)

Professeur : Alexandre Tsybakov - ENSAE-CREST

deuxième semestre

Cours : 16 heures

TP : 8 heures

Objectifs

L'objet de ce cours est de présenter quelques méthodes classiques de l'estimation non-paramétrique et d'estimation statistique en grande dimension. Les thèmes suivants seront abordés :

Plan

1. Estimateurs à noyaux et par projection d'une densité de probabilité. Adaptation par validation croisée.
2. Estimation non-paramétrique d'une fonction de régression. Estimateurs par polynômes locaux, par projection et par la méthode de splines. Vitesses de convergence, inégalités d'oracle et adaptation.
3. Régularisation et méthodes pénalisées. Procédures de seuillage, BIC et Lasso. Estimation statistique en grande dimension, sélection de variables et sparsité. Compressed sensing.

Références

- A.B.Tsybakov : Introduction to Nonparametric Estimation. Springer, New York, 2009.
- A.B.Tsybakov : Apprentissage statistique et estimation non-paramétrique. Polycopié de l'Ecole Polytechnique, 2010.
- L. Wasserman : All of Nonparametric Statistics. Springer, New York, 2006.
- L. Devroye : A Course in Density Estimation. Birkhauser, Boston, 1987.
- A.Nemirovski : Topics in non-parametric statistics. Ecole d'Eté de Probabilités de Saint-Flour XXVIII - 1998. Lecture Notes in Mathematics, v. 1738. Springer, 2000.

Apprentissage statistique en génomique (SE335)

Professeur : Jean-Philippe Vert - Mines ParisTech

deuxième semestre

Cours : 16 heures

TP : 0 heures

Objectifs

Les technologies modernes de séquençage de génome ou de puces à ADN sont en train de révolutionner la biologie et la recherche biomédicale. En fournissant d'immenses quantités de données sur les propriétés moléculaires des organismes vivants, elles ouvrent la voie à l'analyse quantitative de systèmes biologiques. La génomique repose ainsi, de plus en plus, sur l'utilisation de méthodes statistiques pour extraire de l'information pertinente de ces données, en particulier pour inférer automatiquement des modèles prédictifs sur les systèmes biologiques. L'objectif de ce cours est de présenter quelques méthodes d'apprentissage statistique à travers plusieurs applications en analyse de données génomiques.

Plan

Après une présentation générale de la théorie de l'apprentissage statistique, notamment de la classification supervisée, nous explorerons différentes approches 'modernes' pour l'inférence linéaire et non-linéaire en grande dimension et la sélection de variables, notamment (i) les méthodes à noyaux, en particulier les machines à vecteur de support (SVM), (ii) les méthodes parcimonieuses par pénalisation, notamment le Lasso, et (iii) les méthodes d'aggrégation, en particulier les forêts aléatoires. Ces méthodes seront illustrées sur deux problèmes importants en génomique : la classification supervisée de données génomiques pour le diagnostic et le pronostic en oncologie, d'une part, et la reconstruction de réseaux de régulation génétique, d'autre part. Ce cours s'appuiera notamment sur le cours 'Machine learning for computational systems biology' que j'ai enseigné au Malaysia Genome Institute en novembre 2010.

Références

T Hastie, R Tibshirani et J Friedman (2001) The elements of statistical learning : data mining, inference, and prediction. Springer. E Barillot, L Calzone, P Hupé, JP Vert, A Zynoviev (2011) Computational systems biology of cancer. CRC Press (to appear).

Séminaire de sondages (SE336)

Professeur :

deuxième semestre

Cours : 20 heures

TP : 0 heures

Objectifs

Ce séminaire est destiné aux étudiants qui envisagent de produire ou d'utiliser des données issues d'enquêtes, en tant que responsable d'enquêtes ou en tant que chargé d'études (tant dans le domaine économique, sociologique que marketing). Il vise à apporter un approfondissement et un regard critique sur les méthodes de conception et de production de données (échantillonnage, correction de la non-réponse, redressement, estimation amp ;hellip ;) ainsi que sur leur utilisation par des méthodes d'analyse statistique appropriées. Il constitue la suite du cours Sondages I, avec une orientation forte vers les aspects pratiques des enquêtes par sondage.

Chaque séance comprendra une présentation par un expert du domaine étudié, abondamment illustrée d'exemples d'application.

Principaux acquis de la formation

amp ;Agrave ; l'issue du cours, l'étudiant saura :

Choisir parmi les méthodes d'échantillonnage la méthode la plus efficace compte tenu des objectifs de l'enquête et de l'information disponible avant l'enquête ;

Mettre en œuvre les méthodes classiques d'imputation pour traiter la non-réponse partielle, et de repondération pour traiter la non-réponse totale ;

Mettre en œuvre les méthodes de calage classique et de calage généralisé ;

Estimer la précision des estimateurs de paramètres simples ou complexes ;

Utiliser des méthodes d'analyse des données d'enquête qui tiennent compte du plan de sondage.

Plan

1 et 2. Méthodes d'échantillonnage : rappels sur les méthodes classiques d'échantillonnage direct ; sondage équilibré et la macro Cube ; échantillonnage indirect et partage des poids

3. Traitement de la non-réponse totale : causes de la non-réponse, mécanismes de réponse (ignorable, non ignorable) ; analyse des facteurs influençant la non-réponse ; modèles de réponse et repondération

4. Traitement de la non-réponse partielle : imputation par prédiction, imputation aléatoire.

5. Méthodes de calage : théorie du calage sur marges, la macro Calmar, calage et correction de la non-réponse, le calage généralisé

6. Estimation d'un paramètre complexe : estimateur par substitution, estimation de variance (technique de linéarisation, méthodes de Bootstrap)

7. Analyse des données d'enquêtes : la prise en compte du plan de sondage dans les statistiques, les tests classiques, les modèles économétriques ; le traitement des valeurs influentes

8. Méthodologie de collecte : le paradigme de l'erreur d'enquête totale, indicateurs de qualité de la collecte, enquêtes multimodes

9 et 10. Compléments divers : méthodes empiriques (quotas), estimation sur petits domaines, coordination d'échantillons, enquêtes répétées dans le temps

Références

- Ardilly P. (2006), Les techniques de sondage, Technip, 2ème édition
- Haziza D. (2009), Imputation and inference in the presence of missing data, Handbook of Statistics, volume 29, Sample Surveys.
- Lavallée P. (2002), Sondage indirect, méthode généralisée du partage des poids, éditions de l'Université de Bruxelles
- Samp ;auml ;rndal C.-E., Swensson B., Wretman J.(1992), Model Assisted Survey Sampling, Springer-Verlag