

Modélisation probabiliste et statistiques

Cours de première année (EM)
1^{er} semestre

Code : GST11 / Crédit ECTS = à préciser
Nombre d'heures total 45, dont : présentiel 21, travail personnel 21

Les prépas HES-S et Cachan sont d'office dispensés de cette UV.
Aucune autre dispense n'est acquise.
Les cas particuliers sont à soumettre au coordonnateur

Responsable : Denis Berthier, INT/DSI,
e-mail : Denis.Berthier@int-evry.fr

Enseignants et contacts :

Denis Berthier Directeur d'Etudes, coordonnateur, poste 4122, pièce C105-01, (Denis.Berthier@int-evry.fr),
Annabelle Ioannic-Chardin, Maître de Conférences (Annabelle.Ioannic-Chardin@int-evry.fr)
Luc Leclere, Chargé d'Enseignement et de Recherche, poste 4469, Luc.Leclere@int-evry.fr
Frederic Lehmann, Maître de Conférences, Frederic.Lehmann@int-evry.fr
Michel Marot, Chargé d'Enseignement et de Recherche, Michel.Marot@int-evry.fr.

Introduction du cours

La modélisation est la base de toute activité scientifique (qu'il s'agisse de sciences « dures » ou « molles », pour employer des termes aussi courants que contestables). Modéliser, c'est :

- avoir certains objectifs en relation avec une situation, laquelle peut être seulement partiellement ou imparfaitement connue,
- appliquer à cette situation un certain point de vue, tenant compte des objectifs spécifiques,
- extraire de cette situation un certain nombre de caractéristiques jugées pertinentes, tout en négligeant des caractéristiques jugées non pertinentes par rapport au problème qu'on se pose, au point de vue adopté et/ou aux actions envisagées,
- établir une relation entre ces caractéristiques et les propriétés formelles d'un modèle (modèle mathématique, en général, dans le cadre d'une activité scientifique).

De ce fait, un modèle n'est jamais « la réalité », même si cela est trop souvent oublié. À une situation initiale partiellement informulée et/ou indéterminée, le modèle superpose une « grille de lecture » ou « grille d'interprétation », un vocabulaire, des concepts, un langage, des méthodes. Les différentes disciplines scientifiques dites « de base » (maths, probas, théories physiques, etc) fournissent des cadres et des outils de modélisation pour des disciplines moins formalisées ou plus appliquées.

Comme telles, les probabilités et statistiques sont à la fois des disciplines scientifiques (c'est l'aspect développé dans le polycopié) et des outils de modélisation (c'est l'aspect qui sera développé dans les TD).

Vos enseignants, ou le coordonnateur, sont accessibles en dehors des cours, au cas où un problème particulier se poserait. Dans ce cas, n'hésitez pas à les contacter aussitôt, de façon à le corriger au plus vite. Vous pouvez aussi nous joindre par email.

Description du cours

Finalité de l'enseignement :

L'usage des statistiques peut être dévoyé pour apporter de pseudo preuves d'assertions publicitaires, idéologiques ou autres. (Sur l'utilisation des statistiques, et plus généralement sur « l'usage du nombre dans les débats de société: justice, sociologie, médias, statistiques », on pourra consulter le site <http://www.unil.ch.pennombre>). Maîtriser les statistiques ne se réduit pas à l'application de formules apprises par cœur (ou retrouvées dans le polycopié). Il faut comprendre les conditions d'utilisation de telle ou telle méthode, et savoir apprécier, dans chaque cas, si elles sont plus ou moins satisfaites - bref, développer un esprit critique.

Les considérations de l'introduction ont ainsi leur versant pragmatique ; l'un des buts de ce cours est en effet de vous faire affronter des situations comme elles se présentent dans la vie réelle : vous allez vous trouver devant des énoncés de problèmes (délibérément) vagues ou incomplets et vous devrez faire les hypothèses complémentaires plausibles qui vous permettront de vous raccrocher aux méthodes rigoureuses apprises dans le cours.

Il faut apprendre à lire et analyser un énoncé, éliminer les informations parasites, préciser des informations vagues, lever les ambiguïtés, formuler des hypothèses, synthétiser des informations dispersées, non directement utilisables, pour en faire des informations utiles. De cette manière, on arrive à une formulation mathématique de l'énoncé initial.

A la fin, il faut exprimer / interpréter la solution trouvée dans les termes de l'énoncé. Le travail de rédaction est important et ne doit pas être laissé au lecteur. Il est aussi un indicateur de votre degré de compréhension.

Tout cela bien entendu en appliquant le principe d'économie, ou principe du rasoir d'Occam : on ne doit jamais rajouter que le strict nécessaire, sans envisager des hypothèses abracadabrantes.

Objectifs d'apprentissage :

Il s'agit d'acquérir les connaissances de base en probabilités et statistiques, nécessaires à la compréhension d'autres cours (économétrie, marketing, comptabilité, etc.).

Au terme de ce cours, vous devrez comprendre et savoir utiliser pour modéliser dans les contextes décrits en introduction et savoir exploiter pour calculs, les notions et méthodes suivantes :

A) en probabilités

- probabilité,
- variable aléatoire, loi de probabilité, fonction de répartition, espérance, écart-type, probabilité conditionnelle, indépendance,
- loi uniforme, loi de Bernoulli, loi binomiale, loi de Poisson
- loi normale, théorème central limite et théorèmes d'approximations, utilisation des tables de la loi normale,

B) en statistiques

- population, individu, variable, échantillon, distribution statistique
- corrélation
- régression linéaire, méthode des moindres carrés,
- théorie des sondages, estimation et échantillonnage de la moyenne et d'une proportion,
- test du khi (si le temps le permet)

Attentes du cours et approche pédagogique

Le cours se présente sous la forme de 19,5 heures de TD et 1,5 heure de TP. Il est découpé en séances de 3 heures. Toutes sont obligatoires et des contrôles de présence systématiques sont effectués.

Avant chaque séance, vous devez avoir étudié dans le polycopié le chapitre concerné et avoir essayé de faire les exercices associés (voir, plus loin, le calendrier pour la correspondance entre séances et chapitres du cours).

Au cours de la séance, votre enseignant peut être amené à faire quelques rappels de cours, ou à donner des explications sur des points mal compris. En aucun cas, la durée de ces parties de cours ne doit dépasser le quart de la séance. Votre enseignant peut vous envoyer au tableau pour vous faire traiter l'un des exercices que vous aviez à préparer pour la séance en cours. Ce système, basé sur un travail personnel de préparation, n'a d'intérêt que si vous suivez le rythme prévu au planning, c-à-d que vous avez effectivement étudié avant la séance le(s) chapitre(s) prévu(s), et préparé les exercices associés.

Le polycopié constitue la référence en matière de connaissances « théoriques » qu'il faudra avoir acquises. Néanmoins, cela ne signifie pas qu'il suffit de connaître le polycopié pour savoir faire les exercices que vous

aurez à traiter pendant les contrôles. Le prologue du polycopié définit la philosophie générale de cette UV, les exercices inclus dans ce polycopié constituent des exemples illustrant cette philosophie.

Les exercices faits en TD ne sont pas « au programme ». Néanmoins, ils sont des compléments indispensables à une véritable compréhension du cours. Quelqu'un qui saurait le polycopié sur le bout des doigts ne serait pas forcément en mesure de faire les exercices. En effet, le polycopié présente le côté « théorique » de la discipline, avec quelques exemples, tandis que les exercices permettent de l'utiliser « en situation » (cf le prologue). Ils fournissent le mode d'emploi du polycopié devant un problème réel.

Malgré son niveau élémentaire, ce cours est en général considéré comme difficile ; il demande un travail sérieux et continu, même pour certains qui ont l'impression trompeuse d'avoir déjà tout vu et de tout savoir.

Evaluation

Le contrôle comporte deux volets, comptant chacun (sauf cas particuliers liés à des absences justifiées) pour 50% de la note finale 1 :

- le contrôle continu, constitué de deux épreuves écrites d'environ 45 mn,
- le contrôle final 1, constitué d'un examen d'une heure et demi en fin de semestre, portant sur l'ensemble du programme.

En cas d'échec (note finale 1 inférieure à 10), un contrôle final 2 est organisé. La note finale 2 est calculée en tenant compte du contrôle continu, de la même manière que la note finale 1.

Le polycopié, et seul le polycopié, est autorisé pendant les contrôles (final et continus). Aucun échange de documents, d'information, de calculatrice, etc, n'est autorisé pendant les contrôles. Tout manquement est sanctionné par l'exclusion immédiate de la salle et par la note 0. Vous devez donc avoir avec vous le polycopié (utile pour les tables statistiques) et une calculatrice.

Le contrôle continu se fera en partie sous forme d'interrogations écrites planifiées, selon les règles suivantes :

- 1- afin de conserver l'homogénéité de la notation sur toute la promotion, les sujets d'interrogation sont identiques pour tous les groupes, les interrogations doivent donc être simultanées, elles se passeront cette année en salle d'examen. Chaque interrogation est corrigée par un correcteur unique,
- 2- toutes les interrogations ont le même poids dans la note de contrôle continu,
- 3- une interrogation se situe au début de la séance prévue à cet effet, démarre à l'heure précise programmée pour le début de la séance et dure précisément la durée indiquée sur l'énoncé, variable selon la difficulté des exercices (environ 45 mn). Si un élève arrive en retard à l'interrogation, il dispose en tout et pour tout du temps restant au moment de son arrivée, sans délai supplémentaire, et est noté comme s'il avait disposé du même temps que les autres - sauf raison valable de retard, dûment justifiée par écrit, auquel cas il est considéré comme légitimement absent à cette interrogation,
- 4- une absence à une interrogation, non justifiée *via le service de scolarité*, selon les mêmes règles que pour un contrôle final, entraîne la note 0 pour cette interrogation,
- 5- une absence dûment justifiée à une interrogation (*via le service de scolarité*) entraîne la dispense de cette interrogation, la moyenne des interrogations étant alors prise sur l'ensemble des autres interrogations, et le poids du contrôle continu dans la note finale modifié en conséquence,
- 6- il n'y a en aucun cas rattrapage d'une interrogation,
- 7- une interrogation peut porter sur l'ensemble des cours antérieurs, à l'exception d'un cours qui aurait eu lieu la veille ou le jour même,
- 8- à la fin de la durée prévue pour le contrôle, les exercices seront immédiatement corrigés au tableau. Il est indispensable de ne pas quitter la salle avant cette correction, qui permet de voir à chaud ce qui doit être amélioré dans votre travail ; en outre, cela agacerait profondément le coordinateur, qui pourrait être tenté de reposer le même exercice au contrôle final,

9- afin d'assurer un réel encadrement pédagogique personnalisé, *tout changement de groupe est strictement interdit*, sauf accord préalable du coordinateur en début d'année.

La note de contrôle continu est la moyenne de ces interrogations écrites, modulée par une *note de participation* comprise entre -3 et +3. Les points négatifs sont acquis par les absences non justifiées (1/2 point par absence) ; les points positifs sont acquis par la participation *active* aux séances.

Plan du cours

Correspondance cours-chapitres et contrôles continus

Probabilités :

CM 1 : Présentation + chapitres 1 et 2 : Probabilités - Variables aléatoires
CI 1 et CI 2 : chapitres 1 et 2 : Probabilités - Variables aléatoires
CI 3 et CI 4 : chapitre 3, première partie : Lois discrètes usuelles
CI 5 et CI 6 : chapitre 3, deuxième partie : Lois normales, TCL
CI 7 : chapitre 3, deuxième partie (suite) : Lois normales, TCL
CM 2 : Contrôle continu 1 et corrigé

Statistiques :

CI 8 et TP 1 : chapitres 7 et 8 : Corrélation, régression
CI 9 et CI 10 : chapitre 9 : Estimation et échantillonnage
CM 3 : Contrôle continu 2 et corrigé

Contrôle final (1h30)

	TRAVAIL PERSONNEL REQUIS AVANT LA SEANCE	THEME, CONTENU ET SUPPORTS DE LA SEANCE
SEANCE 1 CM1	Etude des chapitres 1 et 2 Préparation des exercices	SEANCE DE CADRAGE SUR OBJECTIFS METHODE OBLIGATIONS CONSIGNES
SEANCE 2 CI 1 + CI 2	Etude des chapitres 1 et 2 Préparation des exercices	Probabilités, variables aléatoires
SEANCE 3 CI 3 + CI 4	Etude du chapitre 3 , 1 ^{ère} partie Préparation des exercices	Lois discrètes
SEANCE 4 CI 5 + CI 6	Etude du chapitre 3, 2 ^{ème} partie Préparation des exercices	Lois continues et théorème central limite
SEANCE 5 CI7	Etude du chapitre 3, 2 ^{ème} partie Préparation des exercices	Théorème central limite
SEANCE 6 CM2 = cc1	Révision de la partie probabilités	Contrôle continu 1 et corrigé
SEANCE 7 CI 8 + TP 1	Lecture des chapitres 5 et 6 Etude du chapitre 7 Préparation des exercices	Corrélation et régression
SEANCE 8 CI 9 + CI 10	Etude du chapitre 9 Préparation des exercices	Estimation et échantillonnage
SEANCE 9 CM 3 = cc2	Révision de la partie statistiques	Contrôle continu 2 et corrigé

Bibliographie

Référence de base :

Polycopié de Denis Berthier : « Modélisation probabiliste et statistiques ».

Références complémentaires :

Bénichou P.: *Statistique descriptive et probabilités*, Polycopié INT.

Evrard Y., Pras B. & Roux E.: *MARKET: études et recherches en marketing: Fondements, Méthodes*, Éditions Nathan, Paris, 1993.

Malinvaud E. : *Méthodes statistiques de l'économétrie*, Dunod 1978.

Saporta G.: *Probabilités, analyse des données et statistique*, Éditions Technip, Paris, 1990.

Pour des études plus poussées :

Delmas J.P.: *Introduction aux probabilités*, Polycopié INT, Édition 1994.

Neveu J.: *Bases mathématiques du calcul des probabilités*, Éditions Masson, Paris, 1964.

Pearl J.: *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann, San mateo, California, 1988.