

Deuxième Journée de Rencontre Dijon - Besançon en Probabilités et Statistique

Université de Bourgogne, IMB, Dijon, 28 janvier 2011

Salle René Baire, Bâtiment Mirande, 4^{ème} étage

9h30-10h00 : Landy Rabehasaina (Univ. Franche-Comté)

Problèmes de premier passage en dimension $1\$$ et $2\$$: applications à la Fiabilité et la Réassurance.

10h10-10h40 : Elise Mostacci (IMB et Plateforme Protéomique Dijon) :

Utilisation de la PLS dans un modèle de survie en grande dimension

Résumé : L'identification de nouveaux biomarqueurs diagnostiques ou pronostiques est un des objectifs majeurs de la recherche clinique. Les données de protéomique clinique sont constituées d'une matrice de covariables X , de taille $n \times p$, dont chaque ligne correspond à l'enregistrement de l'intensité des protéines à l'instant t pour un patient et d'une variable réponse Y dont les valeurs peuvent être des classes (type de cancer, sain/malade) ou des délais de survenue d'un événement (décès, rechute). La grande dimension des données issues de la spectrométrie de masse où le nombre de variables p excède très largement le nombre de patients n rend impossible l'utilisation des outils statistiques classiques. Pour résoudre le problème de la dimension, il est nécessaire de réduire la dimensionnalité des données tout en conservant l'information pertinente. Deux possibilités s'offrent à nous : la sélection de variables ou la construction de nouvelles variables sous forme de composantes qui résument l'information de l'ensemble des données à partir des variables d'origine. Bien que faciles à interpréter, les méthodes de sélection de variables, univariées, ne tiennent pas compte des interactions entre variables. La construction de nouvelles composantes semble être une bonne alternative, plus particulièrement l'utilisation de la PLS qui permet de résumer l'information contenue dans les covariables en tenant compte de la réponse d'intérêt. Les composantes PLS restent toutefois difficiles à interpréter; c'est pourquoi certaines méthodes proposent de conjuguer sélection de variables et réduction de dimension en introduisant des pénalités sur les variables afin de supprimer des composantes celles qui ne contribuent pas à la prédiction. On parle alors de PLS creuse. L'objectif de notre travail est l'estimation du temps de survie dans le cas $n \ll p$ par le biais d'un modèle combinant PLS creuse et modèle de survie. Dans cet exposé, nous vous présenterons un

état de l'art des méthodes utilisant la PLS creuse, d'abord dans le cadre d'une classification binaire des patients puis de l'estimation de la survie.

10h50-11h15 : Pause café

11h15-11h45 : Peggy Cénac-Guesdon (IMB Dijon)

Arbres de contexte, chaînes de Markov de longueur variable et sources dynamiques

Résumé Les séquences aléatoires de lettres, que l'on peut penser être A,C,G,T pour des séquences biologiques ou 0; 1 pour des bits, peuvent être vues comme des chaînes au sens probabiliste ou comme des sources au sens de la théorie de l'information. Nous avons souhaité éclaircir ce lien pour les Chaînes de Markov à Longueur Variable (VLMC). C'est pourquoi, après avoir établi un cadre probabiliste pour les VLMC, nous montrons que toute VLMC est une source dynamique, dont nous construisons explicitement la transformation. Sur deux exemples (le "peigne" et le "bambou fleuri"), nous trouvons une condition nécessaire et suffisante d'existence et d'unicité de mesure stationnaire pour la VLMC. Ces deux exemples sont détaillés de telle sorte que sont fournies les séries de Dirichlet associées ainsi que les fonctions génératrices d'occurrence d'un motif.

11h55-12h25 : Francial Giscard Libégué (Univ. Franche-Comté)

Introduction à la méthode des noyaux associés continus

Résumé Depuis les travaux de Rosenblatt (1956) et Parzen (1962), la méthode du noyau a été utilisée pour diverses fonctions sur toute la droite réelle. Dans le cas de l'estimation d'une densité de probabilité, cette méthode présente des insuffisances et crée des biais de bordure lorsque le support de la densité à estimer est borné au moins d'un côté. Plusieurs solutions ont été proposées dans la littérature aux cas par cas mais celles-ci restent encore incomplètes dans la pratique. Dans cet exposé, nous présentons la méthode des noyaux associés continus qui englobe celle de Rosenblatt-Parzen, ainsi que tous les autres noyaux particuliers. Nous définissons d'abord le noyau associé continu, nous montrons comment le construire à partir d'une loi de probabilité paramétrée. Ensuite, nous l'appliquons pour estimer une densité de probabilité. Nous étudions les différentes propriétés liées à cet estimateur puis nous donnons le principe général de réduction de biais de bordure. Enfin, nos résultats seront illustrés par des simulations et des perspectives seront évoquées.

12h35-14h00 : Pause déjeuner

14h00-14h30 : Camelia Goga (IMB Dijon)

Applications des estimateurs de type ridge et en composantes principales en sondages

14h40 - 15h10 : Bruno Saussereau (Univ. Franche-Comté)

Lois de conservations stochastiques avec bruit fractionnaire: existence, unicité et mesure invariante.

Résumé : On s'intéresse à une perturbation stochastique de type mouvement brownien fractionnaire d'une équation hyperbolique du premier ordre (type équation de Burgers sans viscosité). L'existence et l'unicité de la solution est étudiée via une formule de Lax-Oleinik. La méthode des caractéristiques généralisées associées au principe variationnel nous permettra de construire une unique mesure invariante pour le problème considéré.

15h20-15h50 : Alexis Flesch (Univ. Franche-Comté)

Bandes de confiance par vraisemblance empirique pour l'inférence de fonctions moyennes dans le cadre d'événements récurrents, en concurrence, avec événement terminal et censure à droite.

16h00-16h20 Pause café

16h20-16h50 : Davit Varron (Univ. Franche-Comté)

Comportements asymptotiques de la Lebesgue-répartition sur un segment des accroissements fonctionnels du processus empirique uniforme