
Séance d'affichage

OLU AWOSOGA, University of Lethbridge

Étude pilote randomisée contrôlée multi-sites pour la thérapie de couple basée sur la congruence pour les problèmes de jeu

Cette étude a été menée de 2009 à 2011 en Ontario et en Alberta, au Canada, en vue de comparer le statut de couples en traitement avec des couples contrôles sans traitement ou suivant le traitement habituel. Les couples en traitement ont suivi une thérapie de couple de 12 semaines basée sur la congruence alors que les couples contrôles ont été vus lors de trois brèves rencontres réparties sur 12 semaines. La moyenne initiale du score de jeu DSM-IV était de 8,7/10. Le taux de rétention des couples traités au suivi de 2 mois était de 89 %, comparativement à 78 % pour les couples non traités. Il est intéressant de noter un effet positif inattendu sur les couples du groupe contrôle en raison de la participation à une étude. L'échantillon (N=30; 15 couples) était constitué à 66 % de couples dont l'homme a un problème de jeu et à 34 % de femmes ayant le même problème. La thérapie de couple basée sur la congruence s'est avérée bien acceptée.

ENTAO CHEN, Acadia University

Modèle multinomial hiérarchique basé sur la loi bêta pour l'analyse de la structure d'âges d'animaux sauvages

Une loi bêta modifiée servant à estimer la forme de la survie des animaux est emboîtée dans un modèle multinomial hiérarchique développé pour estimer la distribution sous-jacente des âges annuels par usage exclusif du dénombrement annuel par âge des récoltes. L'évaluation de la performance et de la robustesse du modèle est effectuée à l'aide d'études de Monte-Carlo et montre la précision et le taux de couverture selon différents scénarios de simulation. Une analyse plus approfondie de la performance est obtenue par une application réelle à deux jeux de données d'âge à la récolte.

HAO (NELSON) CHEN, University of British Columbia

Méthodes basées sur les arbres pour l'émulation d'un modèle informatique complexe

Plusieurs phénomènes complexes sont difficiles à investiguer à l'aide d'expériences physiques contrôlées. Les modèles informatiques deviennent alors des solutions de rechange importantes pour comprendre de tels phénomènes. Un processus gaussien est souvent utilisé en tant que substitut statistique pour la relation entre les entrées et sorties d'un modèle informatique. Cependant, les processus gaussiens impliquent une forte hypothèse de stationnarité des sorties. Cet inconvénient peut toutefois être corrigé avec des méthodes à base d'arbres qui découpent l'espace de sortie et ajustent des substituts statistiques différents dans chaque sous-région. Cette affiche présente une revue de plusieurs méthodes basées sur les arbres. Une comparaison entre différentes méthodes basées sur les arbres est ensuite entreprise par simulation et la méthode optimale est identifiée.

JINGJIA CHU, University of Western Ontario

Modélisation du risque commun entre les actions : un modèle multivarié de séries chronologiques avec une structure GARCH additive

Les modèles GARCH DCC (Engle et Sheppard, 2001) ont été bien étudiés pour décrire les matrices de covariance conditionnelle et de corrélation alors que le risque commun dans les séries ne peut pas être saisi intuitivement par les modèles GARCH multivariés existants. Une nouvelle classe de modèles de séries chronologiques multivariés avec une structure additive GARCH est proposée. La covariance conditionnelle dynamique entre les séries est agrégée par un terme de risque commun qui a été la clé pour caractériser la corrélation conditionnelle. Le modèle que nous avons proposé ici peut être mis en œuvre pour s'appliquer à la gestion de portefeuilles institutionnels afin de déterminer les poids sur les actions et les titres à revenu fixe.

HYUKJUN GWEON, University of Waterloo

Codification automatisée des professions par l'apprentissage automatique

La codification des professions fait référence à la catégorisation d'une réponse narrative d'un répondant à une enquête dans l'une des centaines de codes des professions. À la place d'une codification manuelle, nous proposons une codification automatisée par l'apprentissage automatique. La codification automatisée constitue un défi, car les réponses consistent habituellement seulement de quelques mots alors que des centaines de catégories sont possibles. Néanmoins, les approches d'apprentissage automatique peuvent être utiles pour une fraction d'enregistrements. Nous incluons des résultats de simulation en utilisant des données sur les professions provenant d'une agence allemande.

FANG HE, University of Western Ontario

Conception de sous-scène dans le paquet rgl du progiciel R

Il y a plusieurs paquets dans le progiciel R (R Core Team, 2014) qui traitent des graphiques en 3D, comme rgl, ggplot2, etc. L'avantage de rgl est le contrôle de la souris pour faire pivoter ou zoomer la scène. Il y avait encore des limites dans rgl avant la version 0.94. Nos changements ont été inspirés du paquet grid de graphiques en 2D (Murrell, 2006). Nous avons mis en place des « sous-scènes » qui ont des propriétés similaires à celles des « fenêtres d'affichage qui forment une « zone rectangulaire qui fournit un contexte pour dessiner » (Murrell, 2006). Ces nouvelles caractéristiques pourraient devenir la base de l'édition interactive dans le futur.

JAMES KIBERD, Dalhousie University

Le bon, le mauvais et l'invocable : qualité des études et signification statistique, comme les prédicteurs des taux de citation d'études

La fréquence à laquelle un article académique est cité indique son taux de contribution aux ouvrages publiés. Les études de haute qualité et celles qui communiquent des résultats importants tendent à être citées plus fréquemment dans les ouvrages publiés par rapport aux études de moins bonne qualité et celles qui présentent des résultats non significatifs. Nous analyserons les contributions relatives de la qualité des études, la signification statistique rapportée, et le consensus concernant le domaine (basé sur des méta-analyses publiées) sur la fréquence des citations pour les études mentionnées dans un échantillon de méta-analyses de Cochrane. L'incidence de ces facteurs sur la fréquence des citations peut révéler la façon probablement biaisée dont les auteurs évaluent les ouvrages publiés.

KAREN A KOPCIUK, Cancer Epidemiology and Prevention Research, AHS/University of Calgary

Imputation basée sur la distribution pour des données métabolomiques censurées à gauche

Les instruments de spectrométrie de masse utilisés pour mesurer les caractéristiques métabolomiques possèdent des limites de détection qui peuvent se traduire par un nombre important de données manquantes. L'imputation d'une valeur constante entraîne un biais et une réduction de la variabilité dans leurs distributions alors que les laisser tomber pourrait fausser les relations avec d'autres métabolites. Puisque l'estimation dans les méthodes basées sur la projection, telles que les composantes principales et les modèles de régression des moindres carrés partiels, est basée sur la variation, cela peut entraîner des résultats trompeurs. Une méthode d'imputation basée sur la distribution est proposée pour récupérer les valeurs des données manquantes et comparée à des valeurs imputées constantes à l'aide des données simulées. La performance est évaluée en utilisant l'aire sous la courbe ROC.

TAKUMA KUROSAWA, Tokyo University of Science

Normalité asymptotique des estimateurs du maximum de vraisemblance partielle pour des séries chronologiques binaires sur des modèles de transition

Dans cette étude, nous considérons un modèle de transition pour traiter des séries chronologiques binaires dont les observations binaires dépendent d'observations passées et des covariables. Pour estimer les paramètres du modèle, nous pouvons utiliser la méthode d'estimation du maximum de vraisemblance partielle. Nous montrons que les estimateurs sont asymptotiquement normaux sous certaines conditions. En outre, nous étudions ce comportement asymptotique des estimateurs avec taille finie de l'échantillon à l'aide d'études de simulation.

WAN-CHEN LEE, Health Canada

Stratégies statistiques pour évaluer l'exposition aux à des mélanges chimiques pendant la grossesse

Des scientifiques ont découvert que l'exposition précoce, soit dans l'utérus ou pendant les premiers stades de développement de la petite enfance, à des produits chimiques industriels peut altérer le développement. Par exemple, en affectant le cerveau pour modifier le niveau d'attention, la capacité d'apprentissage, ou le comportement; ou encore par d'autres impacts tels que l'altération de l'endroit où les cellules adipeuses s'accumulent ou en modifiant le développement d'organes, menant à une prédisposition au cancer dans le futur. On développe des critères statistiques pour évaluer l'exposition à des mélanges chimiques pendant la grossesse, qui sont ensuite utilisés dans l'évaluation du risque de produits chimiques comme le mercure, le plomb, les phtalates et les polluants organiques persistants.

MICHAEL LI, McMaster University

Modélisation en haute résolution des données animales par télémétrie : modèles à chaîne de Markov cachée et extensions

La classification des données de séries chronologiques en groupes discrets peut améliorer la prévision en plus de fournir un aperçu de la nature des états sous-jacents, non observables du système. Cependant, l'hétérogénéité temporelle et l'autocorrélation (persistance) en occupation de groupe peuvent masquer ces signaux. Nous utilisons des modèles à chaîne de Markov d'état latent et cachée (CMC), deux techniques de classification standard, pour modéliser à haute résolution les données de mouvement des panthères de la Floride sur une base horaire. Cela permet une hétérogénéité temporelle des probabilités de transition, une extension du modèle simple, mais rarement explorée, résout les problèmes de modélisation CMC précédents et précise les comportements des panthères.

JASON LOEPPKY, University of British Columbia

Paramétrage du processus gaussien pour la modélisation d'une fonction de boîte noire

Dans cette affiche, nous discutons deux paramétrisations alternatives du modèle de processus gaussien qui sont souvent utilisées dans la littérature. Nous commençons par discuter de l'invariance des deux modèles. Nous discutons d'abord le manque d'invariance et puis discutons l'absence d'une invariance numérique entre ces deux paramétrages. Nous discutons ensuite l'interprétation de paramétrages et montrons leurs conséquences sur la modélisation du code informatique. De plus, nous montrons à l'aide d'une série d'exemples qu'une paramétrisation est habituellement bien meilleure en terme de qualité de la prévision pour un ensemble de données test.

BIN LUO, University of Western Ontario

Outils statistiques spatiaux pour la détection d'orage de mutations pangénomique sous un système d'échantillonnage de sondes de microréseaux

En génétique, l'étude des « pluies » de mutations peut aider à mieux comprendre les mécanismes mutagènes. Une méthode économique est d'utiliser un tableau spécifique de génotypage de l'organisme qui est conçu pour détecter les mutations à des sites définis sur les sondes pour l'ensemble du génome. Des mutations dans les sites sans sondes ne sont pas observées. Pour établir des outils statistiques formels pour la détection de mutations du génome entier, plusieurs statistiques de test sont proposées et sont basées sur les propriétés spatiales de tableau de la sonde. La performance au niveau de la puissance des tests statistiques est évaluée sous les processus de classement de Neyman-Scott à l'aide de simulations de Monte-Carlo. Les statistiques présentant de bonnes performances sont recommandées comme outils de dépistage pour les généticiens.

ADAM MUIR, Dalhousie University

Est-ce qu'un article de revue vient de dévorer votre citation? L'impact des revues sur la fréquence de citation des études primaires

Les articles de revue occupent une place dans le haut de la hiérarchie de la preuve et sont parmi les contributions qui ont le plus d'influence à la littérature. Il a été établi que les articles de revue sont fréquemment cités. Par contre, il n'est pas clair si

ces revues créent des citations dans un domaine donné en attirant l'attention sur les publications scientifiques qu'elles citent. Dans cette enquête, nous examinons les publications scientifiques citées par les articles de revues Cochrane et explorons les taux auxquels les publications scientifiques sont citées avant et après la publication de l'article de revue.

DREW NEISH, University of Guelph

Partitionnement des données à l'aide de modèles de régression de mélanges Dirichlet multinomiaux

Les analyses de composition du microbiome humain sont possibles à l'aide de méthodes de séquençage avancées, où le résultat est une abondance de taxons bactériens différents dans chaque échantillon de microbiome. Précédemment, modèle de mélanges de Dirichlet multinomial était utilisé pour modéliser de telles données métagénomiques de microbes, où chaque composante du mélange représente des méta-communautés distinctes qui présentent des compositions de biotes semblables. Par contre, l'identification de l'association de covariables environnementales ou biologiques avec l'abondance dans différentes méta-communautés demeure un problème sérieux. Ici, un modèle de régression de mélanges Dirichlet multinomiaux est proposé et présenté. Ces modèles permettent une recherche probabiliste de la relation entre l'abondance bactérienne et les covariables environnementales ou biologiques au sein de chaque méta-communauté présumée.

AURÉLIEN NICOSIA, Laval University

Un modèle général de marche aléatoire directionnelle : application au mouvement animalier

Dans cette affiche, nous proposons un modèle de marche aléatoire directionnelle général pour décrire le mouvement d'un animal qui prend en compte les caractéristiques de l'environnement. Un processus circulaire-linéaire modélise la direction et la distance entre deux localisations consécutives de l'animal. Une structure de processus cachée permet des situations où l'animal manifeste des comportements de mouvement différents. L'originalité principale de l'approche proposée est que beaucoup de cibles environnementales peuvent être simultanément incluses dans le modèle directionnel. L'inférence du modèle utilise l'algorithme EM. Nous illustrons son utilisation en modélisant le mouvement d'un animal dans la forêt boréale du Canada.

ANDREW PORTER, University of Guelph

Estimation de la variance de l'erreur expérimentale pour des plans factoriels fractionnaires

En analysant les plans factoriels fractionnaires il peut être incertain quelles interactions d'ordre supérieur sont inactives et ainsi elles peuvent être groupées afin d'estimer la variance de l'erreur expérimentale. Il est encore moins clair comment grouper les moyennes des carrés qui sont contaminées par un effet actif peut avoir un effet sur le biais ou une erreur de Type I. Nous présentons des résultats de simulation sur la performance de quatre méthodes pour définir la variance de l'erreur expérimentale lorsqu'une des moyennes des carrés est active. En remplaçant la plus grande contribution à l'erreur quadratique moyenne nous espérons que la statistique d'ordre maximal de la loi du khi carré démontrera une baisse satisfaisante du biais.

SHI QIU, University of Saskatchewan

Modèle de comparaison validé croisé et détection de région divergente en employant iIS et iWAIC pour la cartographie d'une maladie

Deux problèmes statistiques se posent dans l'utilisation de modèles hiérarchiques bayésiens pour la cartographie des maladies. Le premier consiste à comparer la qualité d'ajustement des différents modèles qui peuvent être utilisés pour tester différentes hypothèses. Le deuxième problème est d'identifier des valeurs aberrantes/régions divergentes avec risque résiduel inhabituellement élevé ou faible de la maladie. L'évaluation du modèle selon la méthode de validation croisée «leave-one-out» (LOOCV) a été utilisée pour ces deux problèmes. Cependant, la LOOCV réelle prend beaucoup de temps. Cet article présente deux méthodes, à savoir iIS et iWAIC, pour approcher LOOCV avec seulement des échantillons de chaînes de Markov simulés d'une distribution a posteriori basée sur un jeu de données complet.

RHONDA ROSYCHUK, University of Alberta

Étude des statistiques de scanographie spatiale basées sur des données de durée avant événement

La statistique de scanographie spatiale est utilisée pour déterminer les groupements géographiques d'un nombre plus élevé de cas de maladie. Elle permet également de détecter les zones géographiques avec un temps plus long jusqu'à l'événement au moyen de lois appropriées. D'autres auteurs ont proposé des lois exponentielle et de Weibull pour ces durées avant événement. Nous proposons la loi log-Weibull comme autre loi pour les statistiques de scanographie spatiale et nous comparons les trois lois par simulation pour analyser la censure de type I. Ces méthodes sont aussi illustrées avec des données de temps de consultations de spécialistes (cardiologues ou médecins internes) par des malades sortants qui se sont présentés au service des urgences pour une fibrillation atriale et un flutter auriculaire en Alberta de 1999 à 2011.

ASANAO SHIMOKAWA, Tokyo University of Science

Création de modèles arborescents de prédiction basés sur des covariables par intervalle

Dans cette étude, nous considérons l'arbre de survie construit à l'aide de l'algorithme CART (arbres de classification et de régression). Dans le modèle que nous proposons, les covariables peuvent servir de variables par intervalle symboliques. Ce modèle permet d'inclure un concept dans plusieurs nœuds terminaux dans un modèle arborescent. Ensuite, un nouveau concept sera prédit avec tous les nœuds terminaux en fonction de la fréquence observée du concept inclus dans chaque nœud terminal. Pour montrer l'utilité de ce modèle, nous présentons les résultats d'application de l'approche proposée en fonction des covariables obtenues par imagerie par résonance magnétique (IRM) sur des patients présentant des métastases cérébrales provenant du cancer du sein.

ANJALI SILVA, University of Guelph

Analyse comparative des techniques de classification des données de séquençage ARN

Le séquençage de l'ARN (ou RNA-seq) est une approche profonde de la profilation de transcriptomes basée sur le séquençage. Ce séquençage fournit des comptages de transcriptions, offrant ainsi une méthode pour quantifier le niveau d'expression des gènes. Or malgré la vaste disponibilité de données RNA-seq, l'interprétation de ces données dans leur contexte biologique demeure un défi. À l'aide d'algorithmes de classification, il est possible d'effectuer une étude systématique des relations entre gènes pour identifier des gènes qui partagent des profils d'expression similaires. Nous présentons une étude comparative de trois techniques de classification avec des données RNA-seq obtenues d'une étude d'expression génétique sur la réponse du maïs à la restriction de l'azote. Nous analysons les classes de gènes identifiées par chaque méthode afin de déterminer leur importance biologique.

GABRIELLE SIMONEAU, McGill University

Comparaison empirique de méthodes de méta-analyse des données individuelles sur la précision du diagnostic

La méta-analyse des données individuelles (DI) présente de nombreux avantages. Dans le contexte des études sur la précision du diagnostic, la sensibilité et la spécificité communes sont généralement présentées pour un seuil donné et la méta-analyse est effectuée via une approche bivariable. Avec les DI, il est possible d'obtenir la sensibilité et la spécificité communes pour n'importe quel seuil. Une façon d'analyser ces données est d'appliquer le modèle BREM à chaque seuil. Une autre idée consiste à analyser les résultats pour tous les seuils simultanément, ce qui permet de tenir compte de la corrélation au sein de l'étude. Notre objectif est de comparer deux approches multivariées au modèle BREM lorsque les DI sont disponibles, et ce de manière empirique.

YUBIN SUNG, University of Guelph

Méthode en deux étapes pour l'analyse d'associations génétiques avec des caractères longitudinaux multiples d'échantillons de sujets apparentés

Nous proposons une procédure en deux étapes pour identifier les effets pléiotropiques sur des caractères longitudinaux multiples dans un ensemble de données familial. Dans un premier temps, nous analysons chaque caractère longitudinal via un modèle à trois niveaux à effets mixtes. Les effets aléatoires au niveau du sujet et de la famille permettent de mesurer les effets génétiques spécifiques au sujet et les corrélations intraclasse entre sujets au sein de la famille, respectivement. Dans un second temps, nous effectuons un test d'association simultané entre un polymorphisme du nucléotide simple (ou SNP) et tous les effets spécifiques au sujet pour de multiples caractères longitudinaux. Nous réalisons ces opérations à l'aide d'une méthode de scoring de la quasi-vraisemblance dans laquelle la structure des corrélations entre sujets apparentés est ajustée.

FODE TOUNKARA, Université Laval

Copule archimédienne pour des données de Bernoulli corrélées

Cette présentation montre que la famille des copules archimédiennes fournit une large classe de modèles pour la surdispersion dans une expérience de Bernoulli. Ces modèles contiennent des paramètres pour la probabilité marginale de succès et un paramètre pour la dépendance. Nous présentons deux applications. Premièrement, nous construisons un intervalle de confiance pour la corrélation intra-grappe. La deuxième application concerne l'estimation de la taille d'une population fermée à partir d'une expérience de capture-recapture. Nous utilisons le critère AIC pour la sélection d'un modèle.

DOUGLAS WHITAKER, University of Florida

Compréhension estudiantine des histogrammes et diagrammes à bâtons : résultats des évaluations LOCUS

Les histogrammes et les diagrammes à bâtons sont des outils de visualisation de données standard, et la maîtrise de leur utilisation est requise pour comprendre plusieurs sujets statistiques importants. Ces formes de visualisation de données font souvent partie du curriculum mathématique au primaire et au secondaire en plus d'être enseignées dans les cours d'introduction à la statistique à l'université. L'inclusion d'histogrammes et de diagrammes à bâtons dans les documents standard n'implique pas nécessairement que les étudiants comprennent ces concepts au niveau attendu. En se servant des données des évaluations LOCUS (NSF DRL-1118168) administrées à plus de 3500 élèves de 12 à 18 ans aux États-Unis, cette affiche illustre le niveau de compréhension actuel des élèves ainsi que des fausses idées communes à propos des histogrammes et des diagrammes à bâtons.

DOUGLAS WHITAKER, University of Florida

Transition et collaboration : l'identité d'un enseignant d'un cours de perfectionnement de placement en statistique

La statistique est de plus en plus visible et présente dans les programmes d'études primaires et secondaires. Cependant, bien que cette matière soit largement reconnue comme indépendante des mathématiques au niveau tertiaire et au-delà, au primaire et au secondaire l'instruction en statistique se fait généralement dans le cadre des cours de mathématiques. Or, en l'absence d'une préparation et d'un développement professionnel appropriés, il peut en résulter un traitement insuffisant de la matière. Les enseignants de mathématiques chargés d'enseigner un contenu statistique auront peut-être à se forger une nouvelle identité d'enseignant de statistique pour être efficaces. Cette affiche examine l'identité d'un enseignant de mathématiques en exercice dans le secondaire participant à l'enseignement en collaboration d'un cours de perfectionnement de placement en statistique.

KEVIN WILSON, Dalhousie University

Hors de vue, oublié? Effets de la citation par des articles de synthèse sur la fréquence de citation dans la littérature épidémiologique

L'influence d'un texte académique est couramment évaluée selon la fréquence à laquelle il est cité dans la littérature. Bien entendu, les revues systématiques figurent parmi les contributions les plus fréquemment citées et occupent une place privilégiée dans la hiérarchie des preuves. Dans cette étude, nous évaluons l'effet des articles de synthèse (généraux, systématiques et de méta-analyse) sur la fréquence des citations des articles primaires dans une cohorte d'études épidémiologiques publiées en 2005. À l'aide d'une régression de Poisson avec des équations d'estimation générales, nous cherchons à savoir si les articles de synthèse affectent la fréquence de citation de la littérature primaire par les chercheurs.

LINGYUN YE, Canadian Centre for Vaccinology
Validité de la méthode cas-témoins avec tests négatifs

Nous étudions le devis cas-témoins avec tests négatifs (CTTN), une extension du devis cas-témoins traditionnel, dans laquelle la population à l'étude se compose de sujets accédant aux services de santé pour une maladie respiratoire aiguë. Des tests en laboratoire sont réalisés pour confirmer l'issue de la maladie. En raison de sa simplicité, la méthode CTTN est devenue la référence absolue pour estimer l'efficacité des vaccins (EV) contre la grippe et le rotavirus. En modélisant les séries de cas et de cas-témoins comme des processus de Poisson indépendants, nous montrons que le devis CTTN offre une estimation convergente de l'EV. Nous discutons aussi de la justification, de l'interprétation et de plusieurs problèmes de méthodologie que pose le devis CTTN.

BOYKO ZLATEV, University of Alberta
Étude de sélections de poèmes par des méthodes statistiques

Nous étudions un échantillon de sélections de poèmes de P.B. Shelley en lui appliquant diverses méthodes statistiques. Les sélections, publiées par diverses maisons d'édition entre 1840 et 2014, sont tirées de différentes éditions et anthologies. Nous appliquons une mise à l'échelle multidimensionnelle à la fois aux sélections et aux poèmes, les distances étant calculées en transformant les coefficients de corrélation appropriés. Nous effectuons alors une analyse de groupement et de classification sur les données à l'échelle. Nous tirons de cette étude des conclusions sur l'évolution et sur certains aspects non évidents de l'accueil connu par les poèmes de Shelley.