

Journées de Biostatistique 2023

jeudi 16 novembre 2023 - vendredi 17 novembre 2023

Institut de Mathématiques de Toulouse

Recueil des résumés

Contents

Robustesse aux données manquantes : comparaison entre modèle mixte et analyse en composantes principales fonctionnelles	1
Analyse des trajectoires de soin, clustering et survie auprès des patients souffrant d'insuffisance cardiaque	1
Données RNA-seq semi-synthétiques : pourquoi normaliser alors qu'on peut générer des faux positifs ?	2
ShiBa logiciel d'inférence bayésienne sans codage	2
Analyse des pseudo-observations par la méthode des moments généralisée comme alternative à l'analyse de survie Bayésienne	3
Instrumental Variables: state of the art and practical recommandations	4
Comparison between different randomization methods for treatment allocation with a continuous factor as stratification criterion	4
Statistical properties of minimal sufficient balance randomization approach in a stratification context, a simulation study	5
Puissance conditionnelle lors de l'analyse intermédiaire en essai clinique	6
High-dimension Mechanistic Model Building using LASSO Approaches : Application to Ebola Vaccination	6
Modèle conjoint pour données répétées et évènements temporelles avec un âge pathologique latent : s'affranchir de l'âge de début de maladie	7
Statistique bayésienne distribuée, effet centre et power prior	8
Integration of External Data in the Design and Analysis of Clinical Trials for Rare Diseases	8
Tipping point for survival analyses	9
Test bayésien à deux intervalles	9
Modélisation des expériences de préférence de place conditionnée	10
Comparaison de deux méthodes d'estimation de l'incidence de l'infection à VIH	11
Détection des patients transplantés rénaux à risque de perdre leur greffon par long short-term memory	12

Inférence variationnelle structurée pour les modèles de Markov caché	13
Risk ratio, odds ratio, risk difference... Which causal measure is easier to generalize? . .	13
The importance of jointly analyzing treatment administration and toxicity associated with targeted therapies: a case study of regorafenib in soft tissue sarcoma patients	13
Individualized treatment rule characterization via Gaussian process surrogate	13

Données longitudinales / 3**Robustesse aux données manquantes : comparaison entre modèle mixte et analyse en composantes principales fonctionnelles****Auteurs:** Corentin Ségalas¹; Cécile Proust-Lima¹; Robin Genuer¹¹ *Université de Bordeaux / Bordeaux Population Health***Auteur correspondant** corentin.segalas@u-bordeaux.fr

En recherche médicale, la collecte de données longitudinales est très fréquente. L'analyse de ces données permet de décrire l'évolution temporelle d'un processus biologique tout en prenant en compte l'importante variabilité individuelle. Par exemple, les cohortes qui étudient l'évolution cognitive du sujet âgé intègrent les résultats des patients à différents tests psychométriques collectés sur des intervalles de temps réguliers. L'étude rétrospective de ces trajectoires permet de mieux décrire l'évolution cognitive du sujet âgé et constitue un outil précieux dans la compréhension de l'histoire naturelle de la démence.

Le modèle statistique le plus couramment utilisé dans l'étude de données longitudinales dans ce contexte est le modèle mixte. Il permet de modéliser avec une grande flexibilité des trajectoires variées, hétérogènes d'un sujet à l'autre avec souvent peu de mesures par sujet et la présence de données manquantes. Ce modèle est connu pour être, dans certains cas, robuste à la présence de données manquantes, fréquentes dans les données longitudinales. Néanmoins, il s'agit d'un modèle paramétrique et des hypothèses souvent fortes sont nécessaires pour estimer le modèle.

Si l'on considère ces données longitudinales comme des données fonctionnelles, c'est-à-dire comme des réalisations aléatoires d'une fonction inconnue sous-jacente, il devient envisageable d'appliquer d'autres outils statistiques. En particulier, l'analyse en composantes principales fonctionnelles (ACPF) est un outil qui permet de décrire des trajectoires temporelles, sans hypothèses paramétriques. Néanmoins, si cette méthode a été éprouvée sur des données denses et régulières, son utilisation sur des données longitudinales avec, en particulier, la présence de données manquantes reste à explorer.

Nous avons réalisé une étude empirique qui montre que l'ACPF se comporte bien malgré la présence de données manquantes et peut se comparer avec l'approche par modèle mixte. Une application de l'ACPF sur des données issues de la cohorte 3C qui étudie l'évolution du sujet âgé permet d'illustrer son intérêt.

Apprentissage statistique et grande dimension / 4**Analyse des trajectoires de soin, clustering et survie auprès des patients souffrant d'insuffisance cardiaque****Auteurs:** Tristan Amadei¹; Tristan Kirscher¹; Antoine Klein¹**Co-auteurs:** Juliette Murriss²; Anne-Isabelle Tropeano³; Sandrine Katsahian⁴¹ *ENSAE Paris – IP Paris*² *HeKA, Inserm, Inria Paris*³ *Hôpital Européen Georges Pompidou, AP-HP, Inserm*⁴ *Hôpital Européen Georges Pompidou, AP-HP, HeKA, Inserm, Inria Paris***Auteurs correspondants:** antoine.klein@ensae.fr, tristan.amadei@ensae.fr, tristan.kirscher@ensae.fr

L'insuffisance cardiaque (IC) entraîne environ 200 000 hospitalisations par an en France et est associée à une surmortalité élevée. Alors que les patients IC sont de plus en plus âgés, la compréhension

des causes de décès en milieu hospitalier est une problématique de santé publique majeure. Dans ce travail, nous proposons une méthodologie innovante visant à identifier les trajectoires de soin fréquentes et à étudier leur impact sur la survie globale.

Les données utilisées sont extraites de l'Echantillon Généraliste des Bénéficiaires sur les patients IC avec une première hospitalisation entre 2010 et 2016. Dans un premier temps, nous procédons à l'exploration de motifs séquentiels pour caractériser les séquences de soin à partir des diagnostics d'hospitalisation (Groupes Homogènes de Malades, GHM). La deuxième étape consiste à utiliser un algorithme de clustering afin d'apporter un caractère interprétable. Pour ce faire, une nouvelle métrique a été développée pour évaluer la distance entre deux GHM. Enfin, une analyse de survie est réalisée en utilisant des approches semi-paramétriques (modèle de Cox à hasards proportionnels, pénalisé et splines) et des méthodes d'ensemble.

10 051 patients ont été inclus, représentant 85 594 hospitalisations. Un total de 6 618 séquences a été identifié et cinq clusters ont été construits. Les multiples hospitalisations pour IC étaient les motifs les plus fréquents, souvent combinées au décès, et en particulier pour deux clusters incluant des patients plus âgés et avec moins d'hospitalisations. Les seconds motifs fréquents identifiés étaient les complications pulmonaires et cardiaques. Les principaux facteurs associés au décès étaient l'âge, le sexe et la durée de séjour. Les associations trajectoires et décès n'étaient en revanche pas significatives.

Dans l'ensemble, l'approche proposée présente un intérêt méthodologique pour l'analyse des trajectoires de soin, clustering et survie chez les patients souffrant d'insuffisance cardiaque et peut être aisément transposée à d'autres problématiques cliniques.

Apprentissage statistique et grande dimension / 5

Données RNA-seq semi-synthétiques : pourquoi normaliser alors qu'on peut générer des faux positifs ?

Auteurs: Boris Hejblum¹; Kalidou Ba^{None}; Rodolphe Thiébaud^{None}; Denis Agniel^{None}

¹ Inserm U1219

Auteur correspondant boris.hejblum@inserm.fr

En revisitant les analyses de données RNA-seq semi-synthétiques simulées par Li et al. (*Genome Biology* 2022), nous soulignons l'importance de la prise en compte de la taille de librairie et de l'utilisation d'une normalisation adéquate avant d'envisager toute manipulation des observations supposant leur échangeabilité. Après avoir proposé une nouvelle stratégie de simulation corrigée, nous montrons que seules certaines méthodes d'analyse différentielle souffrent réellement d'une inflation exagérée du taux de faux positifs dans les études de grandes tailles avec une forte variabilité biologique (telles que les études chez l'humain) — en particulier edgeR et DESeq2. De plus, nous montrons que le comportement du test des rangs de Wilcoxon n'est en rien supérieur à d'autres tests spécialisés pour l'analyse différentielle de données RNA-seq tels que dearseq. Ces derniers sont de surcroît souvent plus versatiles et peuvent prendre en compte des designs expérimentaux relativement complexes, tandis que le test des rangs ne peut s'appliquer que pour la comparaison non-ajustée entre deux groupes.

Statistique bayésienne / 6

ShiBa logiciel d'inférence bayésienne sans codage

Auteur: Thibaut Fabacher¹

Co-auteurs: Erik-A. Sauleau¹; François Lefebvre²; François Severac²; Julien Godet³; Nicolas Meyer²

¹ Université de Strasbourg, Laboratoire ICube UMR7357

² Hôpitaux Universitaires de Strasbourg

³ *Université de Strasbourg*

Auteur correspondant ea.sauleau@unistra.fr

La diffusion de l'inférence bayésienne dans la communauté scientifique, y compris parmi les spécialistes de biométrie reste limitée. Cette timidité s'explique en partie par l'absence d'outil logiciel omnibus et ne nécessitant pas de codage. Notre logiciel ShiBa, sous licence CC-by-nc 4.0, vient combler ces manques.

ShiBa est un développement utilisant Shiny pour R (Chang W et al (2023). shiny: Web Application Framework for R. <https://shiny.posit.co/>, <https://github.com/rstudio/shiny>) permettant une interface homme-machine conviviale et une prise en main facile. Le logiciel est un outil d'analyse statistique n'utilisant que l'inférence bayésienne (d'où le nom ShiBa associant Shiny et Bayes), et repose essentiellement sur STAN (<https://mc-stan.org/support/>). L'utilisation des Hamiltonian Monte Carlo permet une convergence plus rapide, appréciée par l'utilisateur, tout en gardant de bonnes propriétés de convergence. Outre la possibilité de réaliser des inférences simples (estimation de moyennes, de la fréquence des modalités d'une variable catégorielle), ShiBa permet l'estimation des paramètres de modèles linéaires généralisés : régression linéaire, logistique, Poisson et Beta. Les distributions a priori sont par défaut celles de STAN mais l'utilisateur a la possibilité de changer les hyperparamètres. Par exemple dans la régression linéaire de y sur x, STAN propose un prior normal sur le paramètre de moyenne nulle et d'écart-type 2,5 fois le rapport des écarts-type entre y et x. Mais à l'aide d'une fenêtre pop-up, l'utilisateur peut les changer et voit graphiquement la représentation du prior.

Les résultats de l'inférence sont présentés graphiquement et sous forme tabulaire (médiane et intervalle empirique, par défaut à 95%) avec un indicateur de bonne convergence basé sur l'absence d'autocorrélation résiduelle et sur un Rhat inférieur à 1,10 (pour chaque paramètre). A des fins de sélection de modèle, le WAIC est présenté.

L'aide à l'utilisation se fait par info-bulles pour chaque paramètre ou option ainsi que par un Wiki. Il reste à tester!

Données longitudinales / 7

Analyse des pseudo-observations par la méthode des moments généralisée comme alternative à l'analyse de survie Bayésienne

Auteur: Léa Orsini¹

Co-auteurs: Caroline Brard ²; David Dejaridin ³; Emmanuel Lesaffre ⁴; Gwénaél Le Teuff ¹

¹ *CESP, INSERM U1018, Université Paris-Saclay, Villejuif, France*

² *Clinical Development Operations, Ipsen Innovationn, Les Ulis, France*

³ *Product Development, Data Sciences, F. Hoffmann-La Roche AG, Basel, Switzerland*

⁴ *I-Biostat, KU-Leuven, Leuven, Belgium*

Auteur correspondant lea.orsini@gustaveroussy.fr

En Bayésien, la formulation d'un modèle de survie à risques proportionnels nécessite généralement la modélisation de la fonction de risque de base. Elle peut être paramétrique et suppose alors des hypothèses fortes ou non paramétrique conduisant à une implémentation complexe. En fréquentiste, les pseudo-observations définies par Andersen sont devenues une alternative à l'analyse de survie par le modèle de Cox mais elles sont surtout avantageuses pour des modélisations plus complexes telles que les modèles multi-états ou les événements récurrents. L'avantage des pseudo-observations est de s'affranchir de la complexité des données censurées en les transformant en données longitudinales, ensuite analysées par les équations d'estimations généralisées (GEE). L'objectif est de proposer une nouvelle alternative à l'analyse de survie Bayésienne reposant sur l'analyse des pseudo-observations. Nous proposons d'utiliser la méthode des moments généralisée (GMM) qui repose sur la définition d'une fonction quadratique de moments. Dans le cadre fréquentiste, certains auteurs ont montré que l'approche GMM donne des estimateurs plus efficaces que l'approche GEE lorsque la matrice de travail est mal spécifiée. Contrairement au GEE, une version Bayésienne a également été proposée, basée sur une pseudo-vraisemblance. Nous avons donc étendu les approches GMM (fréquentiste et Bayésien) aux spécificités de l'analyse des pseudo-observations et comparé leurs

performances, par une étude de simulation d'essais randomisés, à celles des modèles de Cox, GEE et Bayésien exponentiel par morceaux. La version fréquentiste donne des performances similaires au GEE. Le GMM Bayésien surestime légèrement l'effet traitement pour des petits échantillons. Pour illustration, trois analyses post-hoc ont été réalisées sur des essais cliniques, de différentes tailles, incluant des patients atteints du Sarcome d'Ewing. Les modèles GMM ont donné des estimations proches du modèle de Cox. L'analyse Bayésienne des pseudo-observations ouvre de nouvelles perspectives pour l'analyse de survie Bayésienne ne nécessitant pas la spécification de la fonction de risque de base.

Inférence causale et essais cliniques / 8

Instrumental Variables: state of the art and practical recommendations

Auteur: Remi Khellaf¹

¹ *INRIA*

Auteur correspondant remi.khellaf@inria.fr

In this work, we provide a comprehensive theoretical and empirical exploration of the integration of instrumental variables (IV) in causal analysis. Specifically, we focus on the estimation of the Average Treatment Effect (ATE) when confronted with the challenge of unmeasured confounding variables.

We begin by introducing the conceptual foundations and methodological underpinnings of the IV estimator, highlighting the critical assumptions, potential violations, and strategies for mitigating such violations. Comparative simulations involving well-known ATE estimators, including Inverse Propensity Score Weighting, the G-Formula, and IV estimation, are presented, demonstrating their performance across a diverse range of scenarios.

Then, acknowledging that practitioners often rely on the sometimes unrealistic linearity of outcome assumption in ATE estimation, we detail a more flexible nonparametric approach that facilitates the computation of the Local Average Treatment Effect (LATE). This method requires an additional assumption, monotonicity, ensuring a monotonous relationship between treatment and the instrumental variable, and integrates it within the framework of Principal Stratification. Empirical and analytical results are showcased, emphasizing the efficacy of this methodology while advocating for the need for caution in LATE estimation.

Our findings reveal challenges in ATE estimation using IV in scenarios with limited sample sizes and the inherent complexity of interpreting results in nonparametric approaches, where the target population for LATE estimation may remain unidentified solely based on available data.

In conclusion, this presentation aims at gaining a comprehensive understanding of when and how to judiciously incorporate instrumental variables into causal analysis, leading to more accurate and insightful conclusions.

Inférence causale et essais cliniques / 9

Comparison between different randomization methods for treatment allocation with a continuous factor as stratification criterion

Auteurs: Aymeric Molinard¹; François Montestruc²; Thomas Blondel³

¹ *INSA Rennes*

² *eXYSTAT*³ *Ceva Santé Animale***Auteur correspondant** aymericmolinar@yahoo.fr

Randomization is a key step in clinical trials to ensure a valid estimation of treatment effect. Most popular randomization method is stratification with blocks. This method can cause serious imbalances which makes this method unworkable in case of small sample size trials or incorporation of several prognostic factors. Minimization can overcome these problems, by accounting for many factors at the design stage, even when the number of patients is low.

The origin of this work is a future study with Ceva Santé Animale, about testing a new medicine on dogs. Stratification had to be applied to allocate 150 individuals to three treatment arms, depending on four qualitative factors and a continuous variable. Continuous criteria are often imbalanced at baseline without any clinically relevant classes to categorize the factor. A new randomization method accounting for both continuous and categorical factors with many strata was developed. This method is derived from classical Pocock and Simon's minimization and accounting for continuous factors directly without transformation into categorical variables while maintaining randomness.

The main objective of this work was to compare different randomization methods, including stratification with blocks, classical minimization and the new randomization method. Simulations were run on randomly generated sets of individuals, to estimate the impact of the randomization method, the number of factors used and the number of individuals on the imbalance between treatment arms.

The method proposed, i.e. non-deterministic neighborhood-based minimization, allows to consider continuous covariate in a way which is at least as efficient as categorized-based usual methods. Minimization is preferable in terms of balance when the number of patients decreases. In the case of this study, the final decision was to use the new randomization method, which creates similar imbalance as stratification with blocks and classical minimization for qualitative factors, but a smaller one for the quantitative variable.

Inférence causale et essais cliniques / 10

Statistical properties of minimal sufficient balance randomization approach in a stratification context, a simulation study

Auteur: Hugo CANNAFARINA¹**Co-auteurs:** Armelle GUENEGOU ARNOUX²; Sandrine KATSAHIAN²¹ *ENSAI*² *Assistance Publique-Hôpitaux de Paris URC APHP.Centre - INSERM CIC1418-EC - INSERM-INRIA HeKA***Auteur correspondant** hugo.cannafarina@eleve.ensai.fr

Introduction – Randomization is a crucial step in clinical trials and ensures balance across treatment groups. Several approaches exist (e.g. stratified permuted blocks or covariate adaptive minimization). Some of them were introduced recently such as Zhao et al Minimum Sufficient Balance (MSB) in 2015. The aim of this work is to assess the performance of MSB to grasp a better understanding of its strengths and limitations.

Methods – (i) Stratified permuted 4-blocks, (ii) Pocock & Simon's minimization (with 2 and 4 classes) and (iii) MSB were applied and compared one to another in simulated stratified and unstratified settings. The simulated scheme was based on Sepsicool-1 trial data. Data augmentation was performed using NORTA method in order to generate correlated datasets with any specified distribution. Scenarios of 5,000 datasets each were considered with variations in terms of sample sizes and correlation structures. Evaluation criteria included imbalance tests, adjusted and unadjusted statistical power, and RMSE of observed and estimated treatment effect.

Results – MSB outperformed at minimizing covariates imbalance with stratified Students-t-test and

stratified Wilcoxon signed-rank test, as well as non-stratified Students-t-test (for both augmented and non-augmented scenarios). Minimization was better than all the others for non-stratified Wilcoxon signed-rank test. Since, in practice, covariate adaptive randomization procedures are rarely stratified, as the important covariates are included in the randomization, MSB remains a relevant candidate. However, no randomization procedure was clearly better than the other at reaching the true treatment effect.

Conclusion – This study showed that MSB is a valuable randomization approach in adaptive design to control for group imbalance, and to maintain a high probability in showing treatment effect.

Keywords - Randomized clinical trial, adaptive design, data simulation

Inférence causale et essais cliniques / 11

Puissance conditionnelle lors de l'analyse intermédiaire en essai clinique

Auteur: Juliette Murriss¹

Co-auteurs: Maud MEGRET²; Sandrine KATSAHIAN³

¹ HeKA, Inserm, Inria Paris

² ENSAI

³ Assistance Publique-Hôpitaux de Paris URC APHP.Centre - INSERM CIC1418-EC - INSERM-INRIA HeKA

Auteur correspondant maud.megret@eleve.ensai.fr

Introduction – L'analyse intermédiaire (AI) lors d'un essai clinique peut permettre d'évaluer le critère de jugement principal avant le recrutement ou la fin de suivi de l'ensemble des patients. Cela mène à l'arrêt précoce ou à la poursuite de l'essai. Cette étape inclut notamment l'évaluation de la puissance conditionnelle (PC, probabilité d'obtenir un résultat significatif à la fin de l'étude, conditionnellement aux hypothèses initiales et aux données collectées à l'AI). Notre objectif est de fournir les éléments nécessaires et suffisants au calcul de cette PC, ainsi que ses propriétés.

Méthodes – Trois zones sont définies pour étudier la PC : 1) une zone défavorable à la poursuite de l'essai, 2) une zone prometteuse à la poursuite après réévaluation du nombre de patients, et 3) une zone favorable à la poursuite dans les conditions initiales. A partir des données d'un essai clinique randomisé contrôlé en réanimation, l'évolution de la PC est explorée selon le nombre total de patients inclus et la différence observée de critère de jugement (binaire) entre les deux bras de traitement au moment de l'AI.

Résultats – La PC est positivement associée au nombre de patients total inclus à l'AI. La marge de différence entre randomisation et critère de jugement principal au moment de l'AI puis finale est plus importante. Une PC calculée lors de l'AI considérée dans la zone prometteuse a de plus grande chance de conduire à une puissance théorique finale élevée. Néanmoins, nous n'avons pas pu mettre en évidence d'association directe entre PC et significativité de la différence de critère de jugement en fin d'étude. Enfin, la PC peut être calculée au global ou par sous-groupe d'intérêt.

Conclusion – L'exploration des propriétés de la PC a mené à l'élaboration de fiches de bonnes pratiques pour des statisticiens.

Mots-clés - Essai clinique, analyse intermédiaire, puissance conditionnelle

Apprentissage statistique et grande dimension / 12

High-dimension Mechanistic Model Building using LASSO Approaches : Application to Ebola Vaccination

Auteurs: Auriane Gabaut¹; Mélanie Prague¹

¹ Université de Bordeaux, Inria, Inserm, Bordeaux Population Health Research Center, SISTM Team ; Vaccine Research Institute, Créteil, France

Auteur correspondant aurianegabaut@gmail.com

Constructing non-linear mixed-effects models (NLMEM) deepens our comprehension of biological processes. Specifically, NLMEM facilitates the incorporation of inter-individual variability by parameterizing models at the individual level within a population framework. To do so, parameters combine fixed effects, capturing population-level relationships with covariates, and random effects, accounting for individual deviations. Estimation in NLMEM is achievable through maximum likelihood methods, such as the Stochastic Approximation Expectation-Maximization (SAEM-Dempster, 1977; Kuhn & Lavielle, 2005) algorithm. However, this approach is computationally intensive, and selecting covariates that define individual-level parameters cannot be done by comparing all possible models.

For the optimized construction of NLMEM, traditional methodologies rely on modified stepwise approaches (SCM-Jonsson, 1998; COSSAC-Ayral, 2021). Alternatively, the Stochastic Approximation for Model Building Algorithm (SAMBA-Prague & Lavielle, 2022) builds the covariate model on the posterior realization of the parameters. Within a low-dimensional context, SAMBA efficiently and more rapidly constructs models by minimizing an information criterion. However, we aim to extend it for high-dimensional settings—such as those involving transcriptomic data. Initially, SAMBA employs a stepwise AIC algorithm for covariate selection. Our proposal integrates a multivariate LASSO approach, offering a more nuanced treatment of parameter correlations. This methodology incorporates a whitening step (Perrot-Dockès, 2018) and a stability selection process (Meinshausen & Bühlmann, 2010).

We validated our approach through simulations imitating the dynamics of the humoral immune response to an Ebola vaccine (Pasin, 2019). These simulations were replicated 100 times, each involving 100 individuals and 200 covariates. Remarkably, the False Discovery Rate for the proposed method was reduced by a factor of 10, while maintaining a similar False Negative Rate. This indicates enhanced control over the False Positive Rate. We applied our method using data from the Prevac/Prevac-UP trial (Prevac-UP Team, 2022), which compares two licensed vaccines for Ebola in Africa.

Données longitudinales / 13

Modèle conjoint pour données répétées et événements temporelles avec un âge pathologique latent : s'affranchir de l'âge de début de maladie

Auteurs: Juliette ORTHOLAND¹; Stanley Durrleman²; Sophie Tezenas du Montcel³

¹ ICM, INRIA

² INRIA, ICM

³ APHP, ICM

Auteur correspondant juliette.ortholand@icm-institute.org

Introduction Les maladies chroniques ont souvent des progressions hétérogènes, ce qui représente un défi dans la mise au point de thérapies efficaces. Le développement de modèles de progression a permis de mieux comprendre cette hétérogénéité. Les modèles conjoints de mesures répétées et d'événements temporels ont ainsi prouvé leur efficacité, mais nécessitent un âge de début de maladie. Or, pour ces pathologies les processus biologiques sous-jacents peuvent commencer avant l'apparition des premiers symptômes, comme c'est le cas pour la Sclérose Latérale Amyotrophique (SLA). L'âge de début de la maladie est donc mal défini.

Objectif Dans ce travail, nous proposons un modèle conjoint à effets mixtes non linéaires avec un âge pathologique latent, afin de s'affranchir de la nécessité d'un âge de début de maladie.

Méthode Nous avons appliqué notre modèle au contexte de la SLA. Nous avons validé notre modèle en effectuant une analyse de sensibilité sur des données simulées. Nous avons ensuite comparé notre

modèle à un modèle conjoint de l'état de l'art et à des modèles de référence en survie et en données répétées, à la fois sur des données simulées et réelles.

Résultats Nous avons montré qu'avec la baisse de la qualité des données (moins de points, plus de bruit), notre modèle tend à surestimer la médiane de la survie et à sous-estimer la croissance absolue des données répétées. Sur les données réelles de la SLA, notre modèle est significativement plus performant que le modèle conjoint de l'état de l'art, pour le biais absolu (4,083(4,244) contre 5,42(5,687)[p-value=1,5e-212]) et pour l'AUC moyenne (0,672(0,072) contre 0,624(0,094)[p-value=2,9e-03]), qui mesure l'ordonnement des événements.

Conclusion Nous avons montré que dans le contexte d'un début de maladie mal défini, notre approche est plus adaptée que l'état de l'art. Nous avons également extrait des conseils d'application pour les futurs utilisateurs.

Statistique bayésienne / 14

Statistique bayésienne distribuée, effet centre et power prior

Auteur: Erik-A. Sauleau¹

Co-auteur: Bineta Faye²

¹ Université de Strasbourg, Laboratoire ICube UMR7357

² Hôpitaux Universitaires de Strasbourg

Auteur correspondant ea.sauleau@unistra.fr

L'exploitation des données de santé suppose le respect de contraintes réglementaires fortes notamment lors de l'analyse des données de plusieurs établissements (ou centres). Porter les analyses statistiques, et les algorithmes d'apprentissage, vers les données, et non plus agréger les données, permet de résoudre ces problèmes réglementaires.

L'inférence bayésienne se prête à ce cas des statistiques, dites distribuées, par la mise en commun des distributions *a posteriori* des différents centres (et non de leurs données). Une difficulté vient quand on cherche un effet centre ou que les données d'un centre sont biaisées. On peut assimiler ce biais à l'hétérogénéité d'un centre par rapport aux autres et utiliser la technique du *power prior*. Cette technique, efficace notamment dans les GLM en conjonction avec le *g-prior* pour les coefficients du prédicteur linéaire, permet de mesurer l'hétérogénéité entre des centres « historiques » et le centre « courant ». Elle pondère la vraisemblance de chaque centre historique par un coefficient α_0 sur lequel un prior Beta est placé.

Si le centre biaisé est parmi les centres historiques le α_0 correspondant sera faible, différent des autres. Le problème est que le biais n'est pas visible si le centre biaisé est le centre courant. On propose alors une « cartographie » de l'hétérogénéité en utilisant tour à tour chaque centre comme courant et les autres comme historiques. Le produit entre α_0 et la taille de chaque centre permet de repérer le centre biaisé, lorsqu'il est en position « historique » par un produit moindre que celui des autres centres. Lorsque le centre biaisé est en position « courante », tous les produits sont du même ordre de grandeur. Le centre biaisé doit ensuite être supprimé ou laissé dans les centres « historiques » pour compenser le biais sur les estimations des paramètres du prédicteur linéaire.

Statistique bayésienne / 15

Integration of External Data in the Design and Analysis of Clinical Trials for Rare Diseases

Auteur: Emilie Jounot¹

Co-auteur: Lucie Truffaut-Chalet¹

¹ Servier

Auteur correspondant emilie.jounot@servier.com

Background: In the rare neurological disease field, it is challenging to demonstrate treatment effect in the clinical trials for new treatments. Some neurological diseases have very few patients available for recruitment. As a result, a sufficiently powered randomized clinical trial (RCT) is not feasible. The importance of incorporating historical data into the planning and analyses of clinical trials is on the rise, particularly for extending the control group of a new trial. However, the use of external data also introduces the risk of potential bias, as the historical control population may be rather different from the RCT.

Method : There are several approaches to eliminate or reduce the bias due to population difference. The methods used during the internship are either frequentist (propensity score matching (PSM) and inverse probability of treatment weighting (IPTW)), or a combination of frequentist and Bayesian methods (PSM or IPTW combined with power prior or commensurate prior).

Purpose: The aim is to compare the operational characteristics of various methods allowing to borrow information from historical data, in the context of a new trial with a 2:1 randomization ratio. The analyses are conducted using simulated data from a new clinical trial combined with an external cohort acting as historical data. In order to take into account discrepancies between historical and current data, different scenarios of natural change in the progression of the disease (called drift) and imbalance in patient group characteristics are implemented.

Results: Methods of borrowing historical data can increase power without creating too much bias or inflation of the type I error. Whether it is really beneficial depends on the hypotheses and the settings.

Données longitudinales / 16

Tipping point for survival analyses

Auteur: Clement Daniel¹

Co-auteur: estelle lambert

¹ Servier

Auteur correspondant clement.daniel@servier.com

Background: Traditionally, tipping point analysis explores the influence of missingness on the overall conclusion of the treatment difference by shifting the imputed missing values in the treatment group towards the reference group until the result becomes non-significant. Over the past years, tremendous efforts have been made to develop statistically rigorous methods for tipping point analysis in clinical trials for continuous and for binary or categorical endpoints. However, less attention has been paid to studies with time-to-event outcomes.

Methods: The objective of this research project is to investigate how methods that are widely used for time-to-event outcomes can be extended in a clinically meaningful and interpretable way to test the censor-at-random assumption. Several approaches for conducting such analyses based on multiple imputation using parametric, semi-parametric, and non-parametric imputation models are investigated and their operating characteristics via simulation are evaluated.

Results: It appeared that the piecewise exponential multiple imputation (PEMI) has a smaller bias compared with other methods when estimating treatment effect while preserving the Type I error rates, but this might be explained by the fact the data were simulated following a piecewise exponential hazard. The range of methods that we studied allows the analyst to match a method with the analysis planned for a particular clinical trial, whether this analysis be parametric, semi-parametric, or non-parametric.

Statistique bayésienne / 17

Test bayésien à deux intervalles

Auteurs: Nicolas MEYER¹; Erik-André SAULEAU²

¹ *iCUBE UMR 7357, Strasbourg. Hôpitaux Universitaires de Strasbourg*

² *Université de Strasbourg, Laboratoire iCUBE UMR 7357.*

Auteur correspondant nmeyer@unistra.fr

La théorie bayésienne permet de réaliser des tests mais avec deux formulations selon que l'on teste la présence d'une variable dans un modèle ou la valeur d'un paramètre déjà inclus dans un modèle. Le Null Hypothesis Bayesian Test spécifie une hypothèse nulle ponctuelle, problématique car la probabilité qu'un paramètre vaille une valeur précise est mathématiquement nulle, et une hypothèse alternative. Différentes alternatives ont été proposées (Region Of Practical Equivalence (ROPE) de Kruschke ou le Hill-and-Chimney prior (HCP) de Tendeiro). Mais la ROPE néglige de formuler une hypothèse alternative tandis que le HCP constitue une loi a priori dont la forme parfois très irrégulière s'interprète mal et toujours focalisée sur le rejet de l'hypothèse nulle.

Le test à deux intervalles (2IT) est une formulation bayésienne alternative du test sur la valeur d'un paramètre. Le test formule une Hypothèse d'Absence d'effet (HA) et d'une Hypothèse de Présence d'effet (HP) sur l'espace de la loi a priori du paramètre. Les intervalles peuvent être joints ou disjoints, de longueurs identiques ou non. Le 2IT permet d'intégrer dans un même concept les tests de supériorité, de non-infériorité et d'équivalence. Des simulations montrent ses excellentes propriétés statistiques. Les probabilités $\Pr(HA|HA)$ et $\Pr(HP|HP)$ de conclure correctement à HA ou à HP croissent rapidement à mesure que les effectifs augmentent. Les valeurs de $\Pr(HA|HP)$ et $\Pr(HP|HA)$ sont toujours inférieures à 0,05 et tendent rapidement vers 0.

Le 2IT pallie toutes les limites des tests actuels classiques ou bayésiens. Il permet d'accepter formellement l'une des deux hypothèses testées, sans recourir à une hypothèse nulle ponctuelle, en ne testant que des effets plausibles. Les rôles de la taille d'effet et de la taille de l'échantillon sur la probabilité de conclure une hypothèse sont clairement individualisés. Il donne en outre un cadre formel à la notion de sérendipité.

Inférence causale et essais cliniques / 18

Modélisation des expériences de préférence de place conditionnée

Auteurs: David Tseitline¹; Emmanuel Curis²; Nadia Benturquia³

¹ *ENSAI*

² *UR 7537 BioSTM*

³ *UMR-S 1144, faculté de pharmacie, université Paris Cité —INSERM*

Auteur correspondant david.tseitline@eleve.ensai.fr

Parmi les protocoles expérimentaux permettant d'évaluer le pouvoir addictogène d'une substance et le risque de rechute après un sevrage, la préférence de place conditionnée joue un rôle important. Son principe repose sur un dispositif à au moins deux compartiments : dans l'un, l'animal cobaye (le plus souvent, une souris ou un rat) a reçu le produit étudié, alors que dans l'autre il reçoit un placebo, délivré dans les mêmes conditions (« phase de conditionnement »). L'idée est alors de quantifier la préférence que développe l'animal pour le compartiment dans lequel il reçoit la substance (compartiment « conditionnant ») ; cette préférence est évaluée par le temps qu'il y passe le jour du test après la phase de conditionnement. Le plan expérimental comporte aussi, le plus souvent, une première partie pendant laquelle les deux compartiments sont explorés sans substance ni placebo, afin d'évaluer la préférence naturelle de l'animal pour l'un ou l'autre des compartiments (« pré-test »).

Si ce principe est simple, la mise en pratique est délicate et de nombreuses variantes existent : nombre de compartiments (deux ou trois, avec un compartiment « neutre »), structure des compartiments et des repères que peut avoir l'animal pour les distinguer, temps total de suivi de l'animal ; façon de choisir quel compartiment contiendra la substance ; façon de tenir compte de la préférence naturelle de l'animal. Outre ces choix expérimentaux, il existe aussi une très grande variété de méthodes d'

analyse des résultats, dont une bonne part porte sur la façon de prendre en compte le fait **1**) que les temps passés dans chaque compartiment sont liés, puisque leur somme vaut nécessairement le temps total de suivi, imposé et **2**) qu'il est nécessaire de comparer les temps après le conditionnement (test) aux temps lors du pré-test. Plusieurs formules sont utilisées dans la littérature pour cela : différence des temps ; différence des temps rapportés au temps passé dans le compartiment neutre (lorsqu'il existe)··· et le débat fait toujours rage, avec de nouvelles propositions [1]. Cette variété de méthodes d'analyses, la plupart sur des bases empiriques, rend délicate la comparaison des résultats de la littérature.

Afin d'y remédier, nous proposons un modèle de simulation de données de préférence de place conditionné. Cette simulation génère les temps passés dans chaque compartiment, en respectant la contrainte du temps total. Pour cela, le modèle est exprimé en fraction de temps total passé dans chaque compartiment et un modèle inspiré des modèles de régression logistique multinomiale est proposé : si t_i est le temps passé dans le i -ème compartiment, et si le premier compartiment sert de référence, alors $\ln \frac{y_i}{y_1} = f_i(\theta)$, où θ représente l'ensemble des paramètres du modèle (décrivant, par exemple, le sexe de l'animal, sa préférence spontanée pour le compartiment i , l'effet de la substance étudiée···). En pratique, le modèle proposé est linéaire : $f_i(\cdot) = X_i + \cdot$, avec ε une variable aléatoire centrée, gaussienne. Le cas échéant, certains termes du vecteur θ peuvent être aléatoires (conduisant à un modèle à effets mixtes).

Ce modèle permet de reproduire raisonnablement correctement les données expérimentales, dans le cas d'une expérience réalisée sur quelques centaines de rats, avec de la cocaïne [2]. Il permet aussi d'étudier l'influence du choix de la méthode d'analyse et du plan expérimental sur les conclusions tirées de l'expérience. Nous présenterons les principaux résultats, en termes de risque de première espèce (de voir à tort un effet addictogène de la substance, par exemple) mais aussi en termes de puissance et de biais.

Données longitudinales / 19

Comparaison de deux méthodes d'estimation de l'incidence de l'infection à VIH

Auteur: Julie MUZZOLON¹

Co-auteurs: Amber Kunkel²; Françoise Cazein²; Florence Lot²; Ahmadou Alioum³

¹ Santé Publique France, Université Lyon 1

² Santé Publique France

³ ISPED, INSERM

Auteur correspondant muzzolonjulie@yahoo.fr

Comparaison de deux méthodes d'estimation de l'incidence de l'infection à VIH

Julie Muzzolon^{1, 3}, Amber Kunkel¹, Françoise Cazein¹, Florence Lot¹, Ahmadou Alioum²

1. Unité VIH-hépatites B/C-IST, Direction des maladies infectieuses, Santé publique France

2. ISPED/Inserm U1219, Bordeaux Population Health, Université de Bordeaux

3. Master Mathématiques, Modélisation et Science des données, Université Lyon 1

Contexte

Il est nécessaire d'estimer l'incidence annuelle de l'infection à VIH pour comprendre la progression de l'épidémie. Cependant, ces estimations ne sont pas réalisées de façon régulière en France.

Deux grandes familles de méthodes d'estimation existent : les modèles de rétro-calcul basés sur les délais entre l'infection et le diagnostic et les méthodes basées sur les marqueurs d'infection récente. Notre objectif est de déterminer quelle approche donne les meilleurs résultats sur les données françaises.

Méthodes

Nous avons construit une base de données simulées qui contient la valeur du marqueur d'infection

récente TM au diagnostic et des trajectoires de mesures de CD4, de stades cliniques depuis l'infection jusqu'au diagnostic VIH chez les hommes ayant des relations sexuelles avec des hommes (HSH). Nous avons simulé la période Covid en considérant une baisse des probabilités d'être diagnostiqué en 2020 ($\times 1/2$) et 2021 ($\times 2/3$).

Avec cette base de données simulées, nous avons comparé deux modèles : un modèle de rétro-calcul basé sur les mesures de CD4 au diagnostic (modèle « ECDC ») et un modèle d'extrapolation stratifiée basé sur le marqueur d'infection récente TM (méthode « Le Vu »).

Résultats

Base sans période Covid

Les deux modèles retrouvent la tendance de l'incidence simulée de 2008 à 2022, en revanche l'incidence est sur-estimée par la méthode Le Vu, en particulier chez les primo-testeurs. Le modèle de rétro-calcul ECDC estime mieux l'incidence, mais les intervalles de confiance sont très larges pour les années récentes.

Base avec période Covid

La baisse du nombre de diagnostics entraîne une baisse de l'incidence estimée sur 2020-2021 pour la méthode Le Vu. Le modèle ECDC donne quant à lui des tendances différentes selon les paramètres choisis car tous les taux de diagnostic sont ré-estimés, y compris pour les années antérieures, mais cet effet est limité.

Conclusion

Concernant l'incidence chez les HSH entre 2008 et 2022, le modèle ECDC donne de bonnes estimations, mais la tendance est moins fiable sur les années récentes. Le modèle Le Vu fournit une tendance similaire à la tendance simulée, mais l'incidence est surestimée. Dans les deux cas, il faut interpréter avec prudence les estimations quand on tient compte de la période Covid.

Apprentissage statistique et grande dimension / 20

Détection des patients transplantés rénaux à risque de perdre leur greffon par long short-term memory

Auteurs: Clément Benoist¹; Jean-Baptiste Woillard²; Marc Labriffe²; Pierre Marquet²

¹ Université de Limoges/CHU de Limoges

² Inserm U1248, pharmacologie et transplantation

Auteur correspondant clement.benoist@unilim.fr

Introduction : Le suivi de la fonction rénale est indispensable pour assurer le succès de la transplantation rénale et éviter la perte du greffon. Le débit de filtration glomérulaire (DFG) est la mesure biologique classiquement utilisée pour évaluer la fonction rénale; celle-ci est usuellement estimée à partir de la créatinine sérique. Prédire le DFG en avance permettrait de détecter les patients à risque pour leur apporter une prise en charge adaptée. Par ailleurs, les approches de machine learning et deep learning sont souvent considérées comme des boîtes noires et les cliniciens ont besoin d'une explication de la prédiction.

Matériels et méthodes : La prédiction s'effectue par LSTM (long short-term memory). Nous avons procédé à une séparation entre une base d'apprentissage (50 % des transplantations), une base de test (25 % des transplantations) et une base de validation (25 % de transplantations). Sur la base des connaissances des experts, 7 variables longitudinales et 8 variables non longitudinales ont été utilisées. Une optimisation des hyperparamètres est effectuée par échantillonnage aléatoire (250 jeux d'hyperparamètres). Néanmoins, les sorties d'un réseau de neurones sont difficilement interprétables. Les méthodes classiques telles que les valeurs de Shapley ou LIME (local interpretable model agnostic explanations) se fondent sur l'importance des variables. Cependant, les cliniciens ont l'habitude de comparer leurs patients à des patients précédents en fonction de leur expérience. Généralement, les Case-Based Reasoning sont classiquement employés, néanmoins l'algorithme du SimplEx (<https://arxiv.org/abs/2110.15355>) demande moins d'hypothèses sur les données et donc peut s'appliquer à une grande variété de données ; nous allons donc utiliser l'algorithme du SimplEx. Le SimplEx propose des exemples similaires extrait d'un corpus précédemment constitué. En

médecine, nous préférons la sensibilité à la spécificité. De la même façon qu'en théorie des sondages, nous avons utilisé un système de quota pour surreprésenter les cas associés à une fonction rénale dégradée (notamment les cas d'insuffisance rénale sévère).

Résultats : Le RMSE (root mean square error) du LSTM est $10.4 \text{ mL/min/1.73m}^2$. La prédiction d'une insuffisance rénale sévère (DFG inférieur à 30) est associée à une sensibilité de 73.1 % et à une spécificité de 99.1 %. Pour le SimplEx, des exemples seront proposés dans la présentation. Les exemples proposés par le SimplEx peuvent confirmer ou infirmer l'estimation du SimplEx, ce qui joue le rôle de double sécurité pour la prédiction. La plupart du temps, les exemples du SimplEx confirment la prédiction du LSTM.

Discussion : Dans ce travail, nous avons implémenté l'algorithme du SimplEx sur le LSTM pour l'explicabilité, ce qui permet la comparaison avec les prédictions de DFG pour des patients similaires. Cela permet d'accroître éventuellement la confiance du clinicien dans la prédiction ou de ne pas suivre aveuglément une mauvaise prédiction.

Données longitudinales / 40

Inférence variationnelle structurée pour les modèles de Markov caché

Auteur: Pierre Gloaguen¹

¹ *AgroParisTech*

Inférence causale et essais cliniques / 41

Risk ratio, odds ratio, risk difference... Which causal measure is easier to generalize?

Auteur: Julie Josse¹

¹ *INRIA*

Essais cliniques / 42

The importance of jointly analyzing treatment administration and toxicity associated with targeted therapies: a case study of regorafenib in soft tissue sarcoma patients

Auteur: Thomas Filleron¹

¹ *Oncopole Toulouse*

Apprentissage statistique et grande dimension / 43

Individualized treatment rule characterization via Gaussian process surrogate

Auteur: Nikki Freeman¹

¹ *university of North Carolina*