

Introduction aux événements récurrents en grande dimension

Juliette Murris

Inserm, Centre de recherche des Cordeliers, Université de Paris, Sorbonne Université, Paris, France ;

HeKA, Inria, Paris, France ;

RWE&Data, Pierre Fabre, Boulogne-Billancourt, France.

Introduction – Les critères de jugement composites sont fréquemment utilisés pour répondre aux objectifs principaux des essais cliniques ou des études en vie réelle. Ils permettent en effet de capturer l'information d'événements fatals (comme le décès) et non fatals (comme les hospitalisations, les rechutes, etc.). Les analyses classiques de survie se concentrent généralement sur le délai avant le premier événement. Ces approches ne sont alors plus optimales pour les maladies caractérisées par plusieurs hospitalisations, comme le cancer colorectal. Les données relatives aux événements subséquents sont en effet ignorées. Différentes méthodes statistiques permettent d'analyser l'ensemble de ces événements, dits récurrents. Nous nous intéresserons par ailleurs au cadre de la grande dimension : quand le nombre de variables explose par rapport au nombre d'individus. Ce type de données est de plus en plus fréquent, notamment en raison des technologies modernes, en génomique, sur les bases de données médico-administratives, par le suivi des maladies par des dispositifs médicaux intelligents, etc. Nous savons déjà que les modèles classiques ne convergent plus dans un tel contexte, et nous nous penchons sur le cas des événements récurrents.

Méthode – Dans un premier temps, nous illustrerons les différentes méthodes d'analyse des événements récurrents, en particulier les modèles d'Andersen-Gill, de Prentice, William et Peterson, de Wei-Lin-Weissfeld, de Poisson et négatif-binomiaux. Les données utilisées sont en vie réelle. Pour chaque méthode, l'effet du traitement estimé et la signification statistique seront décrits. Aussi, nous dressons un état des lieux par une revue systématique de la littérature des approches innovantes traitant des événements récurrents en grande dimension. Les méthodes identifiées sont alors appliquées à des données simulées.

Résultats – D'abord, nous verrons sur les données réelles que les méthodes qui prennent en compte les événements récurrents montrent un effet traitement plus important. De plus, nous montrons le gain en termes de puissance statistique par rapport aux autres méthodes. Puis, la revue de littérature nous indique que très peu de méthodes existent pour répondre à la problématique des événements récurrents en grande dimension. Deux méthodes identifiées seront néanmoins appliquées aux données simulées : une méthode de sélection de variables et un réseau de neurones profond. Tandis que les modèles pénalisés rapportent des meilleurs résultats en termes d'indice de concordance (C-index de Harrell, C-index de Kim), le réseau de neurones ne semble pas tout à fait s'adapter aux événements récurrents.

Discussion – Par l'analyse sur données réelles, l'effet du traitement ne se limiterait pas à l'occurrence du premier événement pour certaines pathologies. Cela induit que les événements récurrents devraient finalement systématiquement être intégrés dans les protocoles d'étude. La revue de littérature et l'analyse sur données simulées met également en évidence l'absence de recommandations scientifiques en grand dimension.

Conclusion – Dans l'ensemble, cette présentation soulèvera de nombreuses questions relatives à l'analyse des événements récurrents. Aussi, ce travail souligne la nécessité de développer de nouvelles approches dans un contexte de grande dimension.

Mots-clés

Analyse de survie, événements récurrents, grande dimension