

Statistique et santé

- Antoine Chambaz (MAP5, Université Paris Cité)

Title: Apprendre, évaluer et analyser une règle d'aide à la prise de décision pour la transfusion précoce de produits sanguins en réanimation

Abstract: Les patients gravement blessés souffrant de choc hémorragique nécessitent souvent une transfusion massive. La transfusion précoce de produits sanguins (plasma, plaquettes et globules rouges) est courante et associée à de meilleures issues de l'hospitalisation. Cependant, déterminer la quantité adéquate de produits sanguins à transfuser demeure un sujet de débat scientifique. Le conférencier présentera une méthodologie pour apprendre, évaluer et analyser une règle d'aide à la prise de décision pour la transfusion précoce de sang en réanimation. L'étude utilise des données de la Traumabase française, un observatoire français pour les traumatismes majeurs.

Travail joint avec Nicolas Gatulle (APHP), Julie Josse (Inria), Pan Zhao (Inria) et le Groupe Traumabase.

- El Mehdi Issouani (LMAC, Université de Compiègne)

Title: Aptamer Embedding: using a Continuous vector representation for aptamer design, analysis and deeper understanding

Abstract: Aptamers are DNA or RNA molecules selected for their high affinity and specificity to bind target molecules, similar to antibodies. They are identified through the SELEX process, which involves exposing a random sequence library to a target and retaining sequences that bind well. Employing machine learning techniques, influenced by the findings of Andrea Di Gioacchino et al. (2022), and incorporating insights from text analysis methods like those introduced by Mikolov et al. (2013) with word2vec, we aim at refining the selection of aptamers for Lyme disease detection by proposing an aptamer embedding method. Previous methods, for selecting aptamers that bind specific targets, relied on thermodynamics, but the trend now is machine learning, especially with the availability of large datasets. We employ Restricted Boltzmann Machines to create a continuous vector representation of sequences, enabling better analysis and prediction of aptamer binding. Additionally, we consider a high-dimensional discrete representation of sequences and compare it to the continuous representation. This approach combines biological insights with statistical models aiming to enhance the SELEX process's effectiveness.

Joint work with Miraine Davila Felipe and Ghislaine Gayraud

- Mélanie Zetlaoui (MODAL'X, Université Paris-Nanterre)

Title: Factorisation en matrices positive, efficacité sous contrainte d'identifiabilité : application au risques alimentaires

Abstract: Ce travail s'inscrit dans l'évaluation des risques alimentaires pour des groupes d'individus afin de produire des recommandations alimentaires. Afin d'étudier

les comportements alimentaires des individus, nous utilisons un modèle à variables latentes, la factorisation en matrices positives (NMF), qui permet de modéliser les régimes alimentaires des individus par des systèmes alimentaires constitués de combinaisons d'aliments et qui sont facilement interprétables. Ce modèle présente néanmoins un problème d'identifiabilité, imposant de définir les paramètres sous des contraintes. Dans ce type de modèle, l'information est dégénérée. Nous montrons comment il est possible de définir des bornes d'efficacité sous des contraintes d'identifiabilité dans des modèles (semi)-paramétriques et réguliers (au sens de Le Cam). En pratique, nous présentons un estimateur one-step dans de tels modèles et nous montrons son efficacité. Enfin, la méthode NMF est appliquée à une base de données de consommations alimentaires issue d'une enquête française.

- Patrice Bertail (MODAL'X, Université Paris-Nanterre)

Title: Statistical problems in food risk assessments (linked to nutrients and contaminants)

Abstract: The purpose of this talk is to review some recent approaches for assessing risks in food essentially from a dynamical perspective. Several single or multiple compartment models have been proposed to model the long term behavior of contaminants in the body. The basic model is closely related to a reversed Spaare-Andersen model (used in insurance) and PDMP models, with exponential elimination of contaminants. The purpose is to try to evaluate the probability to be simultaneously over a safe (low) threshold for nutrients and below a high threshold for contaminants in the long term. For multiple compartments models, modelling the stock of nutrients/contaminants in several parts of the body is quite a challenging task. We will also give a few ideas and how it is possible to study the efficiency of nutritional logos (nutriscore as well as other proposals) .