Tackling Explosive Likelihood and Non-Identifiability in Beta Mixtures

Tom Tang[1], Mingxing He[2], Daniel J. McDonald[1], **Jiahua Chen**[1]

[1]Department of Statistics, the University of British Columbia, Canada.

[2]Key Laboratory of Applied Statistics, Kunming University of Science and Technology, China.

ABSTRACT

This work explores statistical inference under beta mixture models. A key challenge with finite beta mixtures is that the likelihood can explode, and the model itself can be non-identifiable. To address the problem of explosive likelihood, we propose a penalized maximum likelihood estimator that adds a penalty term to the log-likelihood. We also introduce a moment-based alternative, which is more computationally efficient. For order selection in the presence of non-identifiability, we investigate order reduction using composite transportation divergence. We demonstrate the proposed methods through simulation studies and apply them to DNA methylation data. All implementations are available in the https://github.com/jhchen-stat-ubc-ca/MixtureInf2.0, an R package MixtureInf2.0 on GitHub.

Keywords: EM algorithm; Machine Learning; Model-based clustering; Optimal transport

Back to Sessions List

Robust Inverse Normal Transformation-Based Tests under Linear Mixed Effects Models

Elif Acar[1], Sonja Friesen[1]

[1]*Department of Mathematics and Statistics, University of Guelph*

ABSTRACT

Many biomedical applications generate vast amounts of data that cannot be managed under human oversight alone, necessitating automated analysis pipelines. These pipelines require robust statistical methods to handle frequent assumption violations in biological settings. Our research is motivated by the International Mouse Phenotyping Consortium (IMPC), which examines mammalian gene function using genetically edited knockout mice. The IMPC employs a linear mixed effects model (LMM) that considers both fixed effects, such as gene, sex, and weight, and random effects that may arise from shared experimental batch or litter. Using simulations based on the IMPC experiment setup, we assess the robustness of the LMM under violations of the normality assumption for error and random effects. We further investigate whether rank-based inverse normal transformations could enhance the detection power of gene effects in these scenarios.

Keywords: Gene effect; Inverse normal transformation; Mixed effects model; Robust inference

Back to Sessions List

Using Statistical Modelling to Inform Public Health Decision Making in Hepatitis B

William W.L. Wong[1]

School of Pharmacy, University of Waterloo

ABSTRACT

Managing chronic hepatitis B (CHB) is difficult because the majority of those infected have clinical silent disease. The asymptomatic nature of CHB means that the disease often remains undiagnosed, leaving its prevalence highly uncertain. This generates significant uncertainty for policy makers in the planning of hepatitis eradication programs to meet the goals set by the WHO. The objective of this work is to establish a statistical framework for the estimation of CHB prevalence and the undiagnosed proportion. A state-transition model describing infection, disease progression and treatment response was mathematically formulated and developed. We then back-calculated the historical prevalence of CHB and the undiagnosed proportion through a calibration process based on a Bayesian Markov chain Monte Carlo (MCMC) algorithm. The algorithm constructs posterior distributions of the historical prevalence of CHB by comparing the model-generated predictions of the annual numbers of health events related to CHB infection and its sequelae against observed calibration targets. These predicted estimates offer crucial insights for policymakers, aiding in the formulation of effective CHB screening and treatment policies.

Keywords: chronic hepatitis B; Markov chain Monte Carlo; prevalence; public health policy

Back to Sessions List

Variable Selection in Mixture Regression Models

Zeny Feng[1], Grace Stelter[2], Lorna Deeth[2], Alysha Cooper[2]

Department of Mathematics and Statistics, University of Guelph

ABSTRACT

The selection of relevant variables through regularization has been widely used in many studies. Finite mixture regressions have also been extensively employed to explore and model the heterogeneous covariate effects on the response. Despite their complementary nature, the issue of variable selection in mixture regression models has received little attention in the literature so far, resulting in limited methodologies available for fitting regularized mixture regression models. In this talk, we introduce a novel algorithm for optimizing regularized mixture regression models with different choices of penalties. Our proposed method is desirable for retaining only relevant covariates in each of the subpopulations and allowing these relevant covariates to have heterogenous effects in different subpopulations. We applied our proposed method to Chiroptera data for relevant and heterogeneous environmental and biological effects on the evolutional development of bat's forearm.

Keywords: Mixture regression models; variable selection; penalized likelihood; optimization

Back to Sessions List