Programme and Abstracts

CFE-CMStatistics 2024

18th International Conference on

Computational and Financial Econometrics (CFE 2024)

and

Computational and Methodological Statistics (CMStatistics 2024)

https://www.cmstatistics.org/CFECMStatistics2024

King's College London, UK

14 – 16 December 2024

ISBN 978-9925-7812-8-7

C1704: **Dependence maps: A graphical tool for visualizing multivariate dependence**
Presenter: **Ivan Medovikov**, Brock University, Canada

A graphical tool for visualizing complex relationships between the variables is demonstrated that is based on the difference of empirical copulas of the underlying variables that we call a "dependence map". We also present a version of the tool designed to assess comonotonicity of vectors, that is - to highlight vector dependence. We demonstrate the applicability of the tool by visually studying the dependence between hourly wholesale electricity prices and grid features such as load and weather in New York. We demonstrate how the tool can uncover instances of non-linear (e.g. tail) dependence.

CC413 Room K2.41 SURVIVAL ANALYSIS **Chair: Kathrin Moellenhoff**

C1625: **Testing the effect of multiple covariates on the cure rate using martingale difference correlation**
Presenter: **Blanca Monroy-Castillo**, Universidade da Coruna, Spain
Co-authors: M Amalia Jacome, Ricardo Cao

In survival analysis, there are situations in which not all subjects are susceptible to the final event. Methods are now well-developed to deal with this kind of data, generally known as cure model analysis. Mixture cure models allow for the estimation of the probability of cure and the survival function for uncured subjects. One of the main goals when working with these types of models is to determine if the covariates have any effect on the cure rate. In the nonparametric sense, a test was proposed in a past study. To propose a new test, the focus is on a new measure. Distance correlation, proposed by another study, has been extensively studied and extended. One of these extensions is the martingale difference correlation, which measures the departure from conditional mean independence between a scalar response variable V and a vector predictor variable U. Moreover, martingale difference correlation and its empirical counterpart inherit several desirable features from distance correlation and sample distance correlation. A new significance test for multiple covariates is proposed using martingale difference correlation and its extensions.

C1429: **Modeling the restricted mean survival time as a function of time horizons with pseudo-value regression trees**
Presenter: **Alina Schenk**, University of Bonn, Germany
Co-authors: Matthias Schmid

The restricted mean survival time (RMST) has become an increasingly important estimand in time-to-event studies. Defined as the restricted area under the survival function over $[0,\tau]$, the RMST represents the average event-free survival time up to the time horizon τ. In practice, directly modeling the RMST conditional on covariates is particularly useful for assessing the impact of a treatment or exposure on the expected lifetime. However, most direct modeling approaches for the RMST focus on a single fixed time horizon $\tau > 0$. Choosing an appropriate value for τ can be challenging and has been widely discussed in the literature. The purpose is to introduce an alternative approach to modeling the RMST as a function of τ using pseudo-value regression trees (PRT). PRT are characterized by a multivariate regression tree built on a pseudo-value outcome and by successively fitting a set of regularized additive models to the data in the nodes of the tree using gradient boosting. Like previously published approaches, PRT models RMST values at various time horizons simultaneously and incorporates time-varying covariate effects. A simulation study and a real-world application are presented to demonstrate the properties of the proposed method.

C0361: **Estimation in a three-state model with interval-censored data**
Presenter: **Luis Machado**, University of Minho, Portugal
Co-authors: Marta Azevedo, Gustavo Soutinho

In many fields, including medical research, engineering, and the social sciences, analyzing time-to-event data is essential for uncovering underlying processes and facilitating decision-making. A common challenge in this analysis arises when events are confirmed to have taken place within specific time intervals, yet the exact timing within those intervals remains unknown, a phenomenon known as interval censoring. The focus is on a three-state progressive multi-state survival model where the intermediate state and/or the final state may be interval-censored. The primary aim is to estimate state occupation probabilities, which are crucial for understanding the dynamics of state transitions over time. Additionally, the estimation of the bivariate distribution of the gap times is considered. New estimation methods are introduced based on the Turnbull estimator of survival to address the challenges posed by interval-censored events. Imputation-based methods are also explored for estimating event times within the interval, such as using the midpoint, left-point, and right-point of the interval. Findings contribute to filling the gap in the literature regarding interval-censored multi-state models, providing valuable insights for researchers and practitioners dealing with such data.

CC431 Room Safra Lec. Theatre FUNCTIONAL DATA ANALYSIS **Chair: Alessia Pini**

C1300: **Additive regression for Riemannian functional responses**
Presenter: **Germain Van Bever**, Universite de Namur, Belgium
Co-authors: Jeong Min Jeon

Additive regression is explored for a functional response whose values lie on a general Riemannian manifold. Euclidean predictors that may not be directly observable but are estimable are also addressed, such as component scores obtained from dimension reduction. The smooth backfitting method is employed to estimate the additive model, and its asymptotic properties are derived. Additionally, novel dimension reduction techniques are discussed for general predictors in the presence of the Riemannian functional response. The usefulness of the approach is demonstrated through a real data application.

C1523: **A fully functional approach for statistical shape analysis**
Presenter: **Issam-Ali Moindjie**, University of Quebec in Montreal, Canada
Co-authors: Marie-Helene Descary, Cedric Beaulac

The shape $\tilde{\mathbf{X}}$ of a random planar curve, $\mathbf{X}$, is what remains when the deformation variables (scaling, rotation, translation, and parametrization) are removed. Previous studies in statistical shape analysis have focused on analyzing $\tilde{\mathbf{X}}$ through discrete observations of $\mathbf{X}$. While this approach has some computational advantages, it overlooks the continuous nature of variables: $\tilde{\mathbf{X}}$, $\mathbf{X}$, and it ignores the potential dependence of deformation variables on each other and $\tilde{\mathbf{X}}$, which results in a loss of information in the data structure. The approach uses functional data analysis to introduce a new framework for studying $\mathbf{X}$. Basis expansion techniques are employed to find analytic solutions for deformation variables such as rotation and parametrization deformations. Then, the generative model of $\mathbf{X}$ is investigated using a joint-principal component analysis approach. Numerical experiments on synthetic data and the *2dshapesstructures* datasets demonstrate how this new approach performs better at analyzing random planar curves than traditional functional data methods.

C1573: **Generalized functional probabilistic principal component analysis for longitudinal microbiome data analysis**
Presenter: **Xiangnan Xu**, Humboldt University of Berlin, Germany

Longitudinal microbiome studies are essential for understanding the dynamic microbial communities that inhabit various body sites and their interactions with host health, offering valuable insights for precision medicine. However, the analysis of longitudinal microbiome data presents challenges due to its high dimensionality, compositionality, overdispersion, and typically sparse sampling at irregular time points. Existing methods for dimension reduction, such as functional data analysis and tensor decomposition, often assume uniform sampling across individuals or focus on Gaussian-distributed variables. To address these, a novel generalized functional probabilistic principal component analysis (GFPPCA) framework is proposed that extends functional tensor decomposition to the setting of the exponential family distribution. GFPPCA integrates functional tensor

