L-Università ta' Malta
Faculty of Science

Department of
Statistics &
Operations Research

Statistics and Operations Research

Final Year Presentation of Dissertations
B.Sc. (Hons.) SCI Yr IV

2024

Programme of Dissertation Presentations

Time	Speaker	Page
09:00	**Opening Speech by Dr Fiona Sammut**	
09:05	Mr Benjamin Grillo	3
09:25	Ms Maya Azzopardi Holland	4
09:45	Mr Adam Buckle	5
10:05	**Coffee Break**	
10:25	Ms Giulia Maria Tabone	6
10:45	Mr Haykel Mansour	7
11:05	Ms Giulia Cachia	8
11:25	Ms Deborah Camilleri	9
11:45	**Lunch Break**	
12:30	Ms Noleen Grima	10
12:50	Ms Melanie Cachia	11

Class of 2024

A Comparative Analysis of Hyperparameter Effects on CNN Architectures for Facial Emotion Recognition

By

Mr Benjamin Grillo

This dissertation investigates facial emotion recognition, an area of computer vision that involves identifying human emotions from facial expressions. It approaches facial emotion recognition as a classification task using labelled images from the FER2013 dataset, employing Convolutional Neural Networks for their capacity to process and extract hierarchical features from image data efficiently. This research utilises custom network architectures to conduct a comparative analysis of the impact of various hyperparameters—such as the number of convolutional layers, regularisation parameters, and learning rates—on model performance. Hyperparameters are systematically tuned to determine their effects on accuracy and overall performance. Notably, the best-performing model developed during this research surpassed human-level performance, established as being somewhere between 65% and 68% on the FER2013 dataset, according to various studies. These findings provide a foundational understanding of hyperparameter optimisation for facial emotion recognition, demonstrating the impact of different configurations on model performance.

The Fractional Compound Poisson Process with Insurance Claims Application

By

Maya Azzopardi Holland

Through the years, the fractional Poisson and the fractional compound Poisson process have became increasingly more popular within mathematical finance, due to their capability of modelling jumps. This dissertation aims to present the mathematical theory of the fractional compound Poisson process, by first discussing the theory of a Lévy process, followed by introducing the term subordinator, along with the sub-classes of a subordinator. The property of long-range dependency and the Mittag-Leffler function are also reviewed as they are fundamental in understanding the demeanour of the fractional models. The fractional Poisson process is characterised through Mittag-Leffler waiting times and by analysing the power-like decay of the process, to confirm the presence of long-range dependency for the jumps of the claims. Moreover, the dissertation studies a Poisson process time-changed by an inverse α-stable subordinator, known as the fractional Poisson process. Furthermore, the fractional compound Poisson process is introduced as an application of the fractional Poisson process. The moment generating
function of the fractional extension of the compound Poisson process is derived, through which the method of moments estimators for the parameters of the fractional models are determined. Additionally, the method of moments estimators for the parameters are implemented on an insurance claims dataset, to estimate the parameters of the complex models. Finally, the robustness of the estimated parameters is investigated, by re-evaluating the parameters for different values in the unit of time, in order to examine whether the estimated parameters are sensitive to a change in the parameter t.

Learning from Experienced Delivery Drivers using Inverse Optimization

By

Mr Adam Buckle

Most studies relevant to the optimization of delivery routes focus on minimizing criteria such as travelling time and the cost of the journey. However, experienced delivery drivers who will have performed such journeys multiple times would learn to formulate convenient routes which they would use on a regular basis. These might not be the routes that minimize the aforementioned factors. This is because such drivers would want to use routes that are free from roadworks, regular traffic congestion, and narrow roads among other criteria. This logic is rather time consuming, if at all possible, to incorporate in traditional optimization frameworks.

In this dissertation, we use Inverse Optimization (IO) to model the preferences of such drivers. Optimization frameworks feature an objective function and a number of constraints. The aim of these frameworks is to find the values of the decision variables that give the best objective value. IO is based on the idea that the values of the objective function and/or constraint coefficients are not always known. Instead, several values for the decision variables are provided. The goal is to find values for the coefficients of the objective function and/or constraints such that the provided values of the decision variables constitute the optimal solution to the optimization problem. For large datasets like the one we use, finding coefficients that make all the routes optimal is rather impractical. Therefore, the IO framework aims to find the objective function which renders as many routes taken by such drivers as possible to be the optimal routes.

Analysing DASS through Structural Equation Modeling

By

Ms Giulia Maria Tabone

Structural Equation Modeling (SEM) is a statistical technique that investigates causal relationships between observed and latent variables. This dissertation provides the theory underlying the SEM framework, with specific attention directed to Muth´en (1984)'s categorical variable methodology (CVM) that makes use of the weighted least squares (WLS) estimator to estimate the model parameters of a SEM fitted on observed categorical and ordinal data. SEM is applied to a dataset, obtained from Open-Source Psychometrics Project (2018) which consists of 42-ordinal items related to the Depression, Anxiety, Stress Scale (DASS) where item value takes values from 1 to 4. The DASS is a clinical assessment tool composed of three self-report scales designed to measure the severity of symptoms common to depression, anxiety, and stress. One of the objectives was to use the DASS data to examine the mental health status in Europe focusing on four countries, namely, Germany, France, Great Britain and Poland. SEM was used to investigate any causal relationship between Depression, Anxiety, and Stress, and to examine the
impact of gender and sexual orientation on these latent variables. Initially the study was conducted without considering the different countries present in the data. Subsequently, following testing for measurement invariance, a Multiple-Group Analysis SEM was fitted, to determine whether the causal relationships between Depression, Anxiety, Stress, Gender and Sexual Orientation vary across the four countries. Interestingly, In France and Great
Britain, LGBTQIA+ affiliation increased the likelihood of experiencing depression, less so in Germany, while in Poland, LGBTQIA+ individuals showed a lower likelihood of experiencing depression.

Bayesian Hierarchical Modelling in Motorsports

By

Mr Haykel Mansour

In this dissertation, a thorough analysis of MotoGP race results between 2016 and 2021 is presented by using Bayesian hierarchical models (BHMs) to obtain estimates for skill variables for riders, teams, and manufacturers. This research in MotoGP, which has yet to be explored with Bayesian models, was inspired by the existing research on other motorsports such as Formula 1. The dataset used for this dissertation includes all race results, filtered to only include riders who have participated in a substantial amount of races to ensure reliability in the results. Two models were implemented using the rjags package in RStudio; one defined on the riders and the teams, whilst the other on the riders and manufacturers. To estimate the skill parameters for the entities in the hierarchy, being rider and team or rider and manufacturer, a logit function on the proportion of riders beaten was used.

The results show the abilities of Bayesian models within the context of MotoGP with both models achieving MCMC convergence and outputting reliable estimates for the different skill parameters. Both models also showed good results when it came to predicting rider standings, particularly when considering the top and bottom riders. The team and manufacturer standings were also considered for each model respectively and both showed accurate outputs predicting most of them perfectly or just one place off. The similarity between the results and the Bayesian RMSE of both models is an indication that most of the variability in the results of MotoGP comes from the riders themselves.

Regression Random Forests for the Identification of Salient Drivers of Tourist Expenditure in Malta

By

Ms Giulia Cachia

This dissertation investigates tourist spending patterns in Malta using regression random forests, focusing on accommodation and other expenditures. Utilising a dataset from the National Statistics Office's Tourstat survey covering the period 2010-2023, this study analyses various demographic and trip-related variables affecting tourist spending. Unsurprisingly, the total number of nights is a significant predictor of spending, however other key predictors also included the type of accommodation and country of residence. Notably, the accommodation expenditures model highlighted the importance of the tourist's country of origin underscoring economic disparities and preferences that affect spending patterns. These insights were crucial for the models' success in predicting spending behaviours related to general tourism and specific accommodation expenses.

In this study we look at two models: one fitting general tourist spending and another focused specifically on accommodation expenses. The accommodation expenditure model exhibited superior predictive accuracy, with lower root mean square error (RMSE) and higher R-squared values, highlighting its utility in capturing spending variations more effectively than the model related to other expenditure. This research has the potential of not only enhancing understanding of spending behaviours among tourists in Malta but also provides actionable insights for stakeholders in the hospitality sector to tailor services and marketing strategies effectively.

A Multivariate Heston-Hawkes Jump Diffusion with Application to High-Frequency Big Tech Stock Prices

By

Ms Deborah Camilleri

The tech industry has witnessed significant growth and disruption in recent years. Indeed, tech giants such as Apple, Meta and Microsoft play pivotal roles in the reshaping of traditional tech services. Investments in these tech giants are continuously evolving, where market dynamics are constantly influenced by real-time information. Understanding these dynamics is essential as these companies are rather significant in the market, due to their large market capitalisation. Furthermore, changes in the stock prices can signal shifts in technological trends and market perceptions of future developments.

Taking this into consideration, this dissertation investigates the dynamic relationships between the stock prices of these big tech companies using a combination of the multivariate Hawkes process and the multi-asset Heston model. By utilising intraday data, the study explores the notion of jumps and volatility of an asset affecting not only the future values of the asset price itself, however also those for other assets. Theoretical foundations are laid out in chapters focusing on the d-dimensional Hawkes process, the d-asset Heston model, and a non-parametric jump-identification technique called the L-estimator.

Furthermore, stock prices are analysed pairwise, by initially identifying the occurrences of jumps and fitting a bivariate Hawkes model on these jumps, followed by disentangling said jumps to obtain the continuous part of the data, in order to fit a two-asset Heston model on this data, thereby studying the stochastic volatility endured by the assets in question. The models fitted highlight that jumps are indeed mutually exciting, stochastic volatility is at its peak when the price is at its minimum, and that each of the asset price and volatility processes for AAPL, META and MSFT are moderately correlated, as anticipated.

Quantile Regression Methods for Modelling Indoor PAH Levels from Lifestyle and Indoor Environment Information

By

Ms Noleen Grima

Quantile regression (QR) methods are employed in this dissertation to capture the effect that indoor activities, such as smoking inside and cooking using a gas cooker, might have on the levels of polycyclic aromatic hydrocarbons (PAH), particularly across different quantiles of the distribution. PAHs are combustion-driven pollutants that can be carcinogenic, posing potential health risks. Unlike the traditional linear regression model, the linear QR model can find a relationship between the quantiles of the response variable and the predictors, providing information on the whole distribution of the dependent variable and not just at the mean. The extension of QR to penalised QR, specifically using the LASSO penalty, is studied and applied to the dataset that was previously used in Harrison et al. (2009). Considering this regularisation technique gave us the opportunity to look at pairwise interaction terms, which allows for a more comprehensive exploration of the complex dynamics underlying PAH exposure, addressing the limitations associated with classical QR models in handling high-dimensional data. A comparison of the fitted models is also presented. The results of this study reveal that concentration levels of PAHs demonstrate a tendency to decrease during the Summer months. This is evident since the respective variable became significant in almost all of the models which were fitted. Furthermore, based on performance measures such as the RMSE and MAE, the LASSO QR model outperforms the classical QR model, particularly for higher quantiles. The LASSO QR model with interaction terms was then found to have a better fit of the model than the LASSO QR on the main effects only.

The Use of Canonical Correlation Techniques to Explore the Relationships Between Customer Satisfaction and Loyalty

By

Ms Melanie Cachia

In this dissertation, a dataset related to consumer satisfaction and behaviour was acquired with the goal of understanding the complex relationships between customer satisfaction and loyalty. Canonical Correlation Analysis (CCA) is a multivariate technique designed to examine the linear relationships between two sets of variables, making it suitable for exploring the connection between different aspects of customer satisfaction and loyalty in our data.

The dissertation commenced with a detailed exploration of CCA, highlighting both its theoretical underpinnings and applicability to consumer behaviour studies. This initial examination demonstrated that CCA could identify significant correlations between variables representing customer satisfaction and loyalty, providing valuable insights into the factors influencing consumer behaviour. To tackle the complexities of multicollinearity and high dimensionality, this dissertation also delves into the theoretical framework of Sparse CCA (SCCA) which uses penalty constraints to incorporate sparsity.

Traditional CCA and SCCA are both applied to the data. In traditional CCA, the significance of the resulting canonical correlations was tested using Wilk's Lambda. In SCCA, we employed a combination of permutation testing and grid search to derive optimal regularisation parameters and identified the significant resulting canonical correlations. The results demonstrated that both traditional CCA and SCCA successfully identified significant canonical correlations. However, SCCA provided a broader view of the relationships. By comparing the effectiveness of CCA and SCCA, the dissertation underscores the potential of advanced statistical methods in uncovering deep insights into the dynamics of consumer behaviour, which are crucial for strategic decision-making in customer relationship management.