Date: Tuesday, 03/Sept/2024

8:15am - 9:00am
Registration Desk

9:00am - 9:15am
Blue Lecture Room
Opening
Location: Blue Lecture Room
Session Chair: **Wolfgang Trutschnig**
Session Chair: **Sebastian Fuchs**
Session Chair: **Jonathan Ansari**

9:15am - 10:10am
Blue Lecture Room
Plenary 1: Plenary Speaker Ingrid Van Keilegom
Location: Blue Lecture Room
Session Chair: **Wolfgang Trutschnig**

10:10am - 10:35am
Coffee Break

10:35am - 12:15pm
Blue Lecture Room
Invited Session 1: Probability Theory Without Probability Measures
Location: Blue Lecture Room
Session Chair: **Pedro Terán**

10:35am - 12:15pm
Green Lecture Room
Invited Session 2: Classification and Clustering with Copulas
Location: Green Lecture Room
Session Chair: **Fabrizio Durante**

12:15pm - 1:15pm
Lunch in the canteen

1:15pm - 3:15pm
Blue Lecture Room
Contributed Session 1: Fuzzy Theory and Applications
Location: Blue Lecture Room
Session Chair: **Maria Angeles Gil**

1:35pm - 3:15pm
Green Lecture Room
Contributed Session 2: Dependence Modeling I
Location: Green Lecture Room
Session Chair: **F. Marta L. Di Lascio**

3:15pm - 3:35pm
Coffee Break

3:35pm - 4:50pm
Blue Lecture Room
Invited Session 3: Dependence Uncertainty
Location: Blue Lecture Room
Session Chair: **Alfred Müller**

3:35pm - 4:50pm
Green Lecture Room
Invited Session 4: Machine Learning and Decision Making Under Weakly Structured Information
Location: Green Lecture Room
Session Chair: **Thomas Augustin**
Session Chair: **Christoph Jansen**

4:50pm - 5:05pm
Group photo

5:05pm - 6:00pm
Blue Lecture Room
Plenary 2: Plenary Speaker Siegfried Hörmann
Location: Blue Lecture Room
Session Chair: **Wolfgang Trutschnig**

Date: Wednesday, 04/Sept/2024

9:00am - 9:55am
Blue Lecture Room
Plenary 3: Plenary Speaker Steven Vanduffel
Location: Blue Lecture Room
Session Chair: **Jonathan Ansari**

9:55am - 10:20am
Coffee Break

10:20am - 12:00pm
Blue Lecture Room
Invited Session 5: Optimal Transport for Statistics and Machine Learning
Location: Blue Lecture Room
Session Chair: **Johannes Wiesel**

10:20am - 12:00pm
Green Lecture Room
Invited Session 6: Theoretical and Applied Aspects of Imprecise Probabilities
Location: Green Lecture Room
Session Chair: **Enrique Miranda**
Session Chair: **Arthur Van Camp**

12:00pm - 1:30pm
Lunch in the canteen

1:30pm - 2:50pm
Blue Lecture Room
Contributed Session 3: Dependence Modeling and Applications
Location: Blue Lecture Room
Session Chair: **Songkiat Sumetkijakan**

1:30pm - 2:50pm
Green Lecture Room
Contributed Session 4: Dealing with Complex Data
Location: Green Lecture Room
Session Chair: **Patrick Benjamin Langthaler**

2:50pm - 3:10pm
Coffee Break

3:10pm - 5:15pm Blue Lecture Room	**Invited Session 7: Latest Trends in Clustering and Classification Methods** Location: Blue Lecture Room Session Chair: **Marta Nai Ruscone**
3:10pm - 5:40pm Green Lecture Room	**Invited Session 8: Robust Statistical Methods I** Location: Green Lecture Room Session Chair: **Beatriz Sinova Fernández**
7:00pm - 8:00pm	**Social Event 1: Classical Concert** Location: Wiener Saal in the Mozarteum building Meeting Time: 6:30 pm Meeting Point: Schwarzstraße 26, 5020 Salzburg, Mozarteum Foundation Start of Concert: 7:00 pm (1 hour)

Date: Thursday, 05/Sept/2024

9:00am - 9:55am Blue Lecture Room	**Plenary 4: Plenary Speaker Susanne Saminger-Platz** Location: Blue Lecture Room Session Chair: **Sebastian Fuchs**
9:55am - 10:20am	**Coffee Break**
10:20am - 12:00pm Green Lecture Room	**Invited Session 10: Robustness for Dependent Data** Location: Green Lecture Room Session Chair: **Una Radojicic**
10:20am - 12:00pm Blue Lecture Room	**Invited Session 9: Adaptive Computing and Decision Making** Location: Blue Lecture Room Session Chair: **Simon Hirlaender**
12:00pm - 1:30pm	**Lunch in the canteen**
1:30pm - 5:30pm	**Social Event 2: Trip to the Untersberg** Meeting Time: 1:15 pm Departure of rented Bus: 1:30 pm Meeting Point: Hellbrunnerstrasse 34, right in front of the Faculty of Natural Sciences Departure Time return journey by bus 5:00 pm Sturdy footwear is recommended, but hiking boots are not necessary. It can be windy and the weather can change quickly. Please bring appropriate clothing. In case of bad weather, the trip to the Untersberg will be cancelled and we will register another group for the sightseeing tour.
2:00pm - 4:00pm	**Social Event 3: Sightseeing Tour** Meeting Time: 1:45 pm Meeting Point: Mozartplatz, monument Start of Tour: 2:00 pm (2 hours) Guided sightseeing tour through Salzburg
7:00pm - 10:00pm	**Dinner: Conference Dinner** Location: M32 restaurant Meeting Time: 6:30 pm (picturesque walk with 100m elevation) Meeting Point: Unipark Nonntal, Erzabt-Klotz-Strasse 1 Lift to M32 (instead of walking): Gstättengasse 13

Date: Friday, 06/Sept/2024

9:00am - 10:40am Blue Lecture Room	**Contributed Session 5: Statistical Learning and Applications** Location: Blue Lecture Room Session Chair: **Simon Hirlaender**
9:00am - 10:40am Green Lecture Room	**Contributed Session 6: Dependence Modeling II** Location: Green Lecture Room Session Chair: **Elisa Perrone**
10:40am - 11:05am	**Coffee Break**
11:05am - 12:00pm Blue Lecture Room	**Plenary 5: Plenary Talk Christian Wagner** Location: Blue Lecture Room Session Chair: **Sebastian Fuchs**
12:00pm - 1:00pm	**Lunch in the canteen**
1:00pm - 1:50pm Blue Lecture Room	**Invited Session 11: Random Sets and Fuzzy Random Variables** Location: Blue Lecture Room Session Chair: **Pedro Terán**

Presentations

Plenary 1: Plenary Speaker Ingrid Van Keilegom

Time: Tuesday, 03/Sept/2024: 9:15am - 10:10am · *Location:* Blue Lecture Room
Session Chair: Wolfgang Trutschnig

Dependent censoring based on copulas

Ingrid Van Keilegom
KU Leuven, Belgium; ingrid.vankeilegom@kuleuven.be

The paper considers survival models in which the survival time and the censoring time are stochastically dependent, which is referred to as dependent censoring. The non-identifiability of a fully nonparametric dependent censoring model leads to challenging problems. A common approach to handle this dependence is based on copulas. To overcome the non-identifiability of the model, the copula can be considered fully known. This is however a heavy assumption in practice, since the strength of the dependence is rarely known. Hence, it results in estimators that can be used for sensitivity analyses but rarely for point estimation of unknown quantities. Recently, a new approach to handle dependent censoring has been proposed, in which the copula is not fully known. The marginal distributions of the survival and censoring time can be modelled parametrically or semiparametrically. The paper describes the literature on these two streams of copula based models.

Invited Session 1: Probability Theory Without Probability Measures

Time: Tuesday, 03/Sept/2024: 10:35am - 12:15pm · *Location:* Blue Lecture Room
Session Chair: Pedro Terán

10:35am - 11:00am

Causal Markov categories and possibility theory

Tobias Fritz[1], Pedro Terán[2]

[1]University of Innsbruck, Austria; [2]University of Oviedo, Spain; tobias.fritz@uibk.ac.at

Markov categories are a recent approach to probability theory that is category-theoretical rather than measure-theoretical. It takes (an abstract version of) Markov kernels, rather than measurable mappings, as primitive. Our aim is to bring researchers in possibility theory and other uncertainty formalism into contact with Markov categories by explaining how the categorical formalism applies in this context as well. We note that, under any continuous triangular norm, possibilistic transition matrices between finite sets form a causal Markov category, just like probabilistic transition matrices do.

11:00am - 11:25am

Limit theorems for convex expectations

Jonas Blessing, Michael Kupper

University of Konstanz, Germany; kupper@uni-konstanz.de

Based on the Chernoff approximation, we provide a general approximation result for convex monotone semigroups. Starting with a family $(I(t))_{t\geq 0}$ of operators, the semigroup is constructed as the limit $S(t)f:=\lim_{n\to\infty}I(\frac{t}{n})^n f$ and is uniquely determined by the time derivative $I'(0)f$ for smooth functions. We identify explicit conditions for the generating family $(I(t))_{t\geq 0}$ that are transferred to the semigroup $(S(t))_{t\geq 0}$ and can easily be verified in applications. Furthermore, there is a structural link between Chernoff type approximations for nonlinear semigroups and law of large numbers type results for convex expectations.

11:25am - 11:50am

Large deviation theory without probability measures

José Miguel Zapata García

Universidad de Murcia, Spain; jmzg1@um.es

Large deviation theory studies the asymptotic tail behaviour of sequences of probability distributions. A series of recent results have generalized the basic principles and theorems of this theory to maxitive translation invariant previsions, striping away the probabilistic aspects of this theory. In this note we survey these results, and explain their connection to large deviation theory.

11:50am - 12:15pm

Possibilistic variables and Akian semirings

Pedro Terán

Universidad de Oviedo, Spain; teranpedro@uniovi.es

The relationship between possibility measures and Akian's semiring-valued idempotent probabilities is studied. The semiring operations in $[0,1]$ for which both theories overlap are shown to be the maximum and an arbitrary 1-Lipschitz triangular norm. Sufficient conditions are then given for the notion of convergence in law to be independent of the choice of a specific triangular norm.

Invited Session 2: Classification and Clustering with Copulas

Time: Tuesday, 03/Sept/2024: 10:35am - 12:15pm · *Location:* Green Lecture Room
Session Chair: Fabrizio Durante

10:35am - 11:00am

Dissimilarity-based clustering with soft proximity constraints

Fabrizio Durante[1], Roberta Pappadà[2]

[1]Università del Salento, Lecce, Italy; [2]University of Trieste, Italy; rpappada@units.it

We provide a general framework for the definition of dissimilarity measures to cluster dependent random variables. A copula-based pairwise dissimilarity measure that takes into account soft proximity constraints is introduced and its properties are illustrated. Furthermore, some recent issues that are particularly of interest for the analysis of geo-referenced data are discussed.

11:00am - 11:25am

Hierarchical clustering and CoClust algorithm: a nested procedure to analyse sustainable heating data

F. Marta L. Di Lascio[1], Roberta Pappadà[2]

[1]Free University of Bozen-Bolzano, Italy; [2]University of Trieste, Italy; marta.dilascio@unibz.it

In this work we combine two different clustering methods to investigate district heating data by taking into account both static and dynamic information concerning the buildings energy profile. The idea is to use the hierarchical clustering algorithm based on the Gower's index to find a first partition of buildings based on their static characteristics, such as age class, energy class, and heating surface, and, next, to investigate the within-cluster multivariate dependence of thermal energy demand among buildings. The two-step procedure we propose aims at assessing the usefulness of static information to support the management of energy demand in the urban area. We show the procedure on data concerning the district heating system of the Italian city Bozen-Bolzano.

11:25am - 11:50am

Copula-based fuzzy clustering of count data with Total Variation distance

Pierpaolo D'Urso[1], Livia De Giovanni[2], Lorenzo Federico[2], Vincenzina Vitale[1]

[1]Sapienza University; [2]LUISS University, Italy; lfederico@luiss.it

This paper presents a novel fuzzy clustering technique designed specifically for count data, referred to as the Fuzzy C-medoids algorithm based on Total Variation Distance. We evaluate its performance against a benchmark relying on Shannon divergence, commonly employed in scenarios involving discrete probability distributions, through simulation analysis. A comprehensive evaluation of the proposed approach's effectiveness is carried out, revealing promising results. The study's findings emphasize the potential of the proposed fuzzy method, particularly in scenarios where discrete probability distributions are involved.

11:50am - 12:15pm

Hierarchical variable clustering based on measures of predictability

Yuping Wang, Sebastian Fuchs
University of Salzburg, Austria; yuping.wang@plus.ac.at

Within a portfolio of random variables, complex structures and associations can occur that influence the overall behaviour of the portfolio. For revealing and identifying such interrelationships, a variable cluster analysis can be performed as a pre-processing step in a data analysis. Motivated by recent developments in the field of hierarchical variable clustering, we here present a short and concise summary of the article Hierarchical variable clustering based on the predictive strength between random vectors [6], in which a variable clustering procedure is introduced that is based on a measure of predictability.

Contributed Session 1: Fuzzy Theory and Applications

Time: Tuesday, 03/Sept/2024: 1:15pm - 3:15pm · *Location:* Blue Lecture Room
Session Chair: Maria Angeles Gil

1:15pm - 1:35pm

Fuzzy clustering implementations for big data in R

Vincenzo Di Perna, Maria Brigida Ferraro
Sapienza University of Rome, Italy; mariabrigida.ferraro@uniroma1.it

The exponential growth in data volume, speed, and variety presents both unprecedented opportunities and challenges across diverse domains. Hence, it is imperative to refine the methodologies and to address the intricacies inherent in the analysis of massive datasets. While implementations of the fuzzy k-means algorithm and its variants are provided by numerous R packages, their computation for extensive datasets demand a considerable amount of time. This inefficiency is not unique to such algorithms, as numerous statistical techniques in R lack the ability to leverage modern resources for computational time reduction. The proposed implementations are designed to enhance the efficiency of the fuzzy k-means type clustering algorithms within the R environment through the integration of parallel computing techniques.

1:35pm - 1:55pm

New decision rules for fuzzy statistical inferences

Julien Rosset, Laurent Donzé
University of Fribourg, Switzerland; julien.rosset@unifr.ch

We profit from the concept of the fuzzy confidence interval to implement two decision rules to perform fuzzy inferences. The procedure can make a decision when both the data and hypotheses are fuzzy. Through a numerical application, we illustrate our proposed method. We compare the levels of acceptance and rejection of different fuzzy null hypotheses at a 5% significant level for a confidence interval at a 95% confidence level.

1:55pm - 2:15pm

Two-Sample Depth-Based Test for Dispersion in Fuzzy Environment

Przemyslaw Grzegorzewski, Anna Kozak
Warsaw University of Technology, Poland; przemyslaw.grzegorzewski@pw.edu.pl

The problem of comparing the variability of two populations in the presence of fuzzy data is considered. We show that applying the modified band depth to specifically transformed membership functions, combined with a permutation procedure, led us to construct a new interesting test for dispersion.

2:15pm - 2:35pm

Aggregation of Inhomogeneous Non-stationary Time Series using P-boxes

Olgierd Hryniewicz, Katarzyna Kaczmarek-Majer
Systems Research Institute, Poland; hryniewi@ibspan.waw.pl

Monitoring of inhomogeneous non-stationary processes has been considered. The application of well-known statistical methods for the analysis of such processes may be questionable, and in the case of long streams of data even infeasible. In the paper, we consider processes consisting of segments and subsegments. The data from subsegments belonging to respective segments are represented by histograms. For consecutive segments, they are aggregated using probability boxes (p-boxes) and a simple probabilistic method. As a result of this aggregation, consecutive segments of the monitored process are represented by triangular fuzzy numbers. These fuzzy numbers may be used for process monitoring using statistical process control (SPC) methods, such as, e.g., control charts, for fuzzy data.

2:35pm - 2:55pm

A Bayesian modeling approach to fuzzy data analysis

Antonio Calcagni[1], Przemysław Grzegorzewski[2]
[1]University of Padova, Italy; [2]Warsaw University of Technology, Poland; antonio.calcagni@unipd.it

Statistical data analysis often entails various uncertainties. Epistemic fuzzy numbers help address these complexities, enabling generalized statistical methods. We propose enhancing fuzzy estimators by integrating a general epistemic mechanism. Our approach, validated through simulation studies, offers a flexible solution for fuzzy data analysis.

2:55pm - 3:15pm

Compound Conditionals and Fuzzy Sets

Lydia Castronovo[1], Giuseppe Sanfilippo[2]
[1]University of Messina, Italy; [2]University of Palermo, Italy; lydia.castronovo@studenti.unime.it

Following the approach proposed by Coletti and Scozzafava, we propose a slightly different interpretation of fuzzy set theory in terms of conditional events and coherent conditional probabilities, examining it within the framework of the betting scheme. We propose a revised definition of fuzzy sets and introduce operations such as complement and intersection. To achieve this, we employ logical operations among conditional events within the framework of conditional random quantities. Specifically, we interpret the membership function of a fuzzy set as a coherent conditional prevision assessment on suitable compound conditionals. Furthermore, we provide examples and examine the class of Frank t-norms to describe coherent conditional previsions.

Contributed Session 2: Dependence Modeling I

Time: Tuesday, 03/Sept/2024: 1:35pm - 3:15pm · *Location:* Green Lecture Room
Session Chair: F. Marta L. Di Lascio

1:35pm - 1:55pm

Dependence properties of bivariate copula families

Jonathan Ansari[1], Marcus Rockel[2]

[1]Paris Lodron Universität Salzburg, Austria; [2]Albert-Ludwigs-Universität Freiburg, Germany; marcus.rockel@finance.uni-freiburg.de

Motivated by recently investigated results on dependence measures and robust risk models, we provide an overview of dependence properties of many well known bivariate copula families, where the focus is on the Schur order for conditional distributions, which has the fundamental property that minimal elements characterize independence and maximal elements characterize perfect directed dependence. We give conditions on copulas that imply the Schur ordering of the associated conditional distribution functions. For extreme-value copulas, we prove the equivalence of the lower orthant order, the Schur order for conditional distributions, and the pointwise order of the associated Pickands dependence functions. For Chatterjee's rank correlation, which is consistent with the Schur order for conditional distributions, we give some new closed-form formulas in terms of the parameter of the underlying copula family.

1:55pm - 2:15pm

Missing in Survival Analysis

Sam Efromovich, Lirit Fuksman

University of Texas at Dallas; lirit.fuksman@utdallas.edu

Typical data in survival analysis is right censored. Right censoring implies that some observations of the lifetimes of interest are not available. Additionally, survival data is often incomplete due to missing observations. Accordingly, it is important to develop optimal procedures that deal with missing censored data. Optimal estimation of bivariate density based on missing censored data is considered, and the theory is complemented by numerical simulations and an example.

2:15pm - 2:35pm

Comparing bivariate random vectors by means of statistical preference

Julián Ros, Raúl Pérez-Fernández, Ignacio Montes

University of Oviedo, Spain; perezfernandez@uniovi.es

Stochastic orders are probabilistic tools used for comparing random quantities. Statistical preference is a stochastic ordering with two main features: firstly, the comparison is based on the joint distribution of the random variables, thus considering potential dependence between them. Secondly, it is accompanied by a winning probability, which measures the strength of the preference. This contribution proposes several ways to extend statistical preference for comparing (bivariate) random vectors. These proposals follow three different approaches: (i) aggregating the winning probabilities of each component, (ii) computing the winning probabilities between the aggregated random vectors, or (iii) defining a purely bivariate winning probability. Besides investigating these properties, the relationships between them and the componentwise median are studied.

2:35pm - 2:55pm

On topologically typical bi-variate Extreme Value copulas

Nicolas Pascal Dietrich[1], Wolfgang Trutschnig[2]

[1]University of Salzburg, Austria; [2]University of Salzburg, Austria; nicolaspascal.dietrich@plus.ac.at

Motivated by the fact that (in the sense of Baire categories and working with the uniform distance d_infty) typical bivariate copulas are completely dependent, we prove that even the subclass of all mutually completely dependent copulas with full support are typical (co-meager). Additionally, considering the subclass of Extreme Value copulas, working with so called Pickands-measures, i.e., univariate probability measures with expected value $frac{1}{2}$, allows not only to determine the support of the Extreme Value copula via the afore-mentioned measure, but also, working with Markov-kernels, to characterize the discrete component of the Extreme Value copula by the atoms of the Pickands-measure. Establishing a homeomorphism between the spaces of all Pickands-measures and the family of Extreme Value copulas and using Markov-kernels, we derive the surprising result that, in contrast to the family of all bi-variate copulas, a typical Extreme Value copula has degenerate discrete component, is not absolutely continuous and has full support.

2:55pm - 3:15pm

Constructing Measures of Dependence via Sensitivity of Conditional Distributions

Patrick Benjamin Langthaler, Jonathan Ansari, Sebastian Fuchs, Wolfgang Trutschnig

Paris-Lodron University Salzburg, Austria; patrick.langthaler@salk.at

An essential objective in statistics is to determine the relationship, specifically the dependence, between two random variables X and Y . To this end many classic and novel methods have been described. Here we show that some of these methods can be seen as special cases of a broad class of measures of dependence, which essentially measure the sensitivity of the distribution of Y conditional on X to changes in X. We introduce some new members of this class which characterise dependence in the sense that they are equal to 0 if and only if X and Y are independent and are equal to 1 if and only Y is a measurable function of X. We establish some further useful properties of these measures and discuss their estimation. In particular, for continuous (X, Y), the introduced measures of dependence rely only on the underlying copula of (X, Y) and not on the marginal distributions. This allows the use of Checkerboard Copulas to obtain a strongly consistent estimator.

Invited Session 3: Dependence Uncertainty

Time: Tuesday, 03/Sept/2024: 3:35pm - 4:50pm · *Location:* Blue Lecture Room
Session Chair: Alfred Müller

3:35pm - 4:00pm

Conditions for the Multivariate Stochastic Order under Dependence Uncertainty

Alfred Müller[2], Marco Scarsini[1]

[1]Luiss University, Italy; [2]Universität Siegen, Germany; mueller@mathematik.uni-siegen.de

The usual stochastic order for random variables can be easily checked in the univariate case. In the multivariate case this is much more difficult. In this paper we consider conditions for multivariate stochastic order for the case that we only know the marginal distributions and have no or only partial knowledge of the dependence structure.

4:00pm - 4:25pm

Pricing Insurance Contracts with an Existing Portfolio as Background Risk

Matthias Scherer, Corrado De Vecchi

Technical University of Munich, Germany; scherer@tum.de

We develop and investigate a premium principle that explicitly takes into account the impact of a new risk on some insurer's existing portfolio. Specifically, we propose the notion of indifference premium for a new risk conditioned on an existing portfolio acting as background risk. The resulting premium rule, which in our case depends on the joint distribution of the new risk and the existing portfolio, is analyzed in detail with respect to its mathematical properties. In order to underline the differences between our approach and the literature on law-invariant premium rules, special attention is given to the indifference premium behaviour with respect to some well-known dependence concepts. Axiomatic and continuity properties of the proposed indifference premium rule are also investigated. To demonstrate the practical relevance of our approach, we consider a portfolio of exchangeable risks and investigate the role of the portfolio's dimension on the price of a risk to be added. This illustrates the (limits of) diversification benefits under this flexible assumption on the joint distribution of a sequence of risks.

4:25pm - 4:50pm

Stochastic orders under uncertainty

Alfred Müller

Universität Siegen, Germany; mueller@mathematik.uni-siegen.de

We study stochastic order relations under uncertainty. This topic has hardly been considered in the literature so far. It is shown that we cannot expect any reasonable robustness of classical first order stochastic dominance, whereas such properties hold for some versions of almost stochastic dominance, as almost stochastic dominance is naturally related to robustness with respect to the Wasserstein distance.

Invited Session 4: Machine Learning and Decision Making Under Weakly Structured Information

Time: Tuesday, 03/Sept/2024: 3:35pm - 4:50pm · *Location:* Green Lecture Room
Session Chair: Thomas Augustin
Session Chair: Christoph Jansen

3:35pm - 4:00pm

Risk versus Profitability in Predictive Maintenance

Florian Sobieczky[1], Ivo Bukovsky[2], Ondřej Budík[2], Maqbool Khan[3]
[1]Software Competence Center Hagenberg, Austria; [2]University of South Bohemia, Dpt. of Computer Science; [3]Pak-Austria Fachhochschule, Institute of Applied Sciences and Technology; florian.sobieczky@scch.at

For a renewal theoretic cost-model which we introduced earlier [1] and which is able to deliver criteria for a higher profitability of Predictive Maintenance over Reactive Maintenance, we present an extension enabling the control of the risk associated to high fluctuations of the profit-function to its asymptotic behaviour. The method relies on bounds of the distribution function of the associated reward-renewal process and allows predictions of recommended minimal initial wealth.

Acknowledgement: This work has been supported by IPMAI (AT-CZ-Interreg Project ATCZ00060). I.B. was supported by the Research Partnership "Process optimization during production in chemical industry by artificial intelligence", project LUABA22069 (MSMT-22065/2022-10).

[1] F. Sobieczky, Ivo Bukovsky, Ondrej Budik, Maqbool Khan, Predictive Maintenance Strategy Cost-Model, ENBIS (European Network of Industry and Industrial Statistics), 2023, https://conferences.enbis.org/event/32/contributions/423/

4:00pm - 4:25pm

Best-possible upper bound on the variance in polynomial time

Marco de Angelis
University of Strathclyde, United Kingdom; marco.de-angelis@strath.ac.uk

We show that the best-possible upper bound on the variance of an interval data set can be found by maximising a concave function, prompting a polynomial-time algorithm in the general case. We look at the variance upper bound from an imprecise probability's perspective. There are two alternative and non-competing interpretations of an interval: (a) a set of real values and (b) a credal set of all possible probability measures in the given interval. These two interpretations do not precipitate any quantitative discrepancies for interval arithmetic. In fact, interval arithmetic provides a means to compute with these imprecise probabilistic objects.

In this work, we show that looking at intervals from an imprecise probabilistic angle can help resolving the repeated-variables problem of interval arithmetic. The main idea is looking at the variance of the stochastic mixture of all bivariate discrete distributions located at the endpoints of each interval as the objective function to maximise. We discover that such a variance admits a global maximum when searching the space of probability masses of the bivariate distributions. We prove that this global maximum is an upper-bound of the real-valued interval variance and that such a maximum can coincide with the maximum of the real-valued variance in a number of cases.

4:25pm - 4:50pm

Creation of Probabilistic Models by Experts: A Case Study

Radek Švadlenka[1], Radim Jiroušek[2]
[1]Prague University of Economics and Business, Czech Republic; [2]Czech Academy of Sciences, Inst. Information Th. and Automation, Prague; radek.svadlenka@alef.com

Several approaches for the data-based construction of probabilistic models have been described in the literature. The present paper presents an alternative approach based on experts' knowledge. Naturally, the knowledge must be somehow transformed into probabilistic notions, which are, in our case, two-dimensional contingency tables. The basic idea is that even non-mathematicians can easily understand simple probabilistic elements (like two-dimensional distributions), from which one can assemble non-trivial complex models. For this, we use decomposable compositional models. The resulting models may be equivalent to a classical Bayesian network or a causal model of Perl's type if the experts can determine which variables are causes and which are their consequences.

Plenary 2: Plenary Speaker Siegfried Hörmann

Time: Tuesday, 03/Sept/2024: 5:05pm - 6:00pm · *Location:* Blue Lecture Room
Session Chair: Wolfgang Trutschnig

Measuring dependence between a scalar response and a functional covariate

Siegfried Hörmann
Graz University of Technology, Austria; shoermann@tugraz.at

Measuring the degree of dependence between a response Y and a covariate X is a

classical and important problem in statistics, which has recently seen a new upsurge.

Motivated by applications in functional data analysis, we consider a dependence coefficient which accommodates the case where X takes values in a general metric space. The asymptotic behavior of the estimator for this dependence coefficient is delicate, as it crucially depends on the nearest neighbor structure of the covariate sample. Some relevant bounds on the degrees of the nearest neighbor graphs available in multivariate settings do not exist in function spaces. The purpose of this talk is to give some insight into this matter and to advise a way how to overcome the problem. As an important application of our results, we consider an independence test.

Plenary 3: Plenary Speaker Steven Vanduffel

Time: Wednesday, 04/Sept/2024: 9:00am - 9:55am · *Location:* Blue Lecture Room
Session Chair: Jonathan Ansari

Risk bounds under uncertainty

Steven Vanduffel
VUB, Belgium; steven.vanduffel@vub.be

We study for any given distortion risk measure its robustness to distributional uncertainty by deriving its largest (smallest) value when the underlying loss distribution lies within a ball – specified through the Wasserstein distance - around a reference distribution. We employ the technique of isotonic projections to provide for any distortion risk measure a complete characterization of sharp bounds. We generalize our results by deriving sharp bounds when distributional uncertainty is described via Bregman-Wasserstein balls and also deal with the case of distortion riskmetrics. Applications to model risk assessment as well as to portfolio choice under ambiguity are discussed. The talk is based on joint works with Carole Bernard, Silvana Pesenti, Peng Liu and Yi Xia.

Invited Session 5: Optimal Transport for Statistics and Machine Learning

Time: Wednesday, 04/Sept/2024: 10:20am - 12:00pm · *Location:* Blue Lecture Room
Session Chair: Johannes Wiesel

10:20am - 10:45am

Sensitivity of multiperiod optimization problems

Daniel Bartl[1], Johannes Wiesel[2]
[1]University of Vienna, Austria; [2]Carnegie Mellon University, USA; daniel.bartl@univie.ac.at

We analyze the effect of small changes in the underlying probabilistic model on the value of multiperiod stochastic optimization problems and optimal stopping problems. We work in finite discrete time and measure these changes with the adapted Wasserstein distance. We prove explicit first-order approximations for both problems. Expected utility maximization is discussed as a special case.

10:45am - 11:10am

Regularised sparse optimal transport for manifold learning

Gilles Mordant
Universität Göttingen, Germany; mordantgilles@gmail.com

In this talk, we discuss a method for manifold learning that relies on a symmetric version of the optimal transport problem with a quadratic regularisation.

We show that the solution of such a problem yields a sparse and adaptive affinity matrix that can be interpreted as a generalisation of the bistochastic kernel normalisation.

We prove that the resulting kernel is consistent with a Laplace-type operator in the continuous limit, discuss geometric interpretations and establish robustness to heteroskedastic noise.

We will show a link to maximum likelihood estimation in Gaussian Mixture Model and the Porous Medium Equation.

The performance on certain simulated and real data examples will be shown.

11:10am - 11:35am

Minimum intrinsic dimension scaling for entropic optimal transport

Austin J. Stromme
ENSAE / CREST, France; austin.stromme@ensae.fr

Motivated by the manifold hypothesis, which states that data with a high extrinsic dimension may yet have a low intrinsic dimension, we develop refined statistical bounds for entropic optimal transport that are sensitive to the intrinsic dimension of the data. Our bounds involve a novel notion of intrinsic dimension, measured at only a single distance scale depending on the regularization parameter, and show that it is only the minimum of these single-scale intrinsic dimensions which governs the rate of convergence. We call this the Minimum Intrinsic Dimension scaling (MID scaling) phenomenon, and establish MID scaling with no assumptions on the data distributions so long as the cost is bounded and Lipschitz, and for various entropic optimal transport quantities beyond just values, with stronger analogs when one distribution is supported on a manifold. Our results significantly advance the theoretical state of the art by showing that MID scaling is a generic phenomenon, and provide the first rigorous interpretation of the statistical effect of entropic regularization as a distance scale.

11:35am - 12:00pm

Empirical martingale projections via the adapted Wasserstein distance

Johannes Wiesel
Carnegie Mellon University, United States of America; wiesel@cmu.edu

Given a collection of multidimensional pairs $\{(X_i,Y_i):1\leq i\leq n\}$, we study the problem of projecting the associated suitably smoothed empirical measure onto the space of martingale couplings (i.e. distributions satisfying $\mathbb{E}[Y|X]=X$) using the adapted Wasserstein distance. We call the resulting distance the smoothed empirical martingale projection distance (SE-MPD), for which we obtain an explicit characterization. We also show that the space of martingale couplings remains invariant under the smoothing operation. We study the asymptotic limit of the SE-MPD, which converges at a parametric rate as the sample size increases if the pairs are either i.i.d. or satisfy appropriate mixing assumptions. Additional finite-sample results are also investigated. Using these results, we introduce a novel consistent martingale coupling hypothesis test, which we apply to test the existence of arbitrage opportunities in recently introduced neural network-based generative models for asset pricing calibration.

Invited Session 6: Theoretical and Applied Aspects of Imprecise Probabilities

Time: Wednesday, 04/Sept/2024: 10:20am - 12:00pm · *Location:* Green Lecture Room
Session Chair: Enrique Miranda
Session Chair: Arthur Van Camp

10:20am - 10:45am

Reversing inhomogeneous and imprecise Markov chains

Damjan Škulj
University of Ljubljana, Slovenia; damjan.skulj@fdv.uni-lj.si

Reversal of a Markov chain means observing the same process backwards in time. A class of Markov chains whose reverses adhere to the same probabilistic laws as the original processes are called reversible Markov chains. In this article, a reversal of imprecise Markov chains is proposed. This is done by reversing a set of compatible inhomogeneous Markov chains. The reversed processes can be explicitly characterized if the interpretation of strong independence is adopted. However, the reversed processes generally lack some important properties such as the Markov property.

10:45am - 11:10am

Entropy-Based Search for Most Informative Belief Functions

Radim Jiroušek, Václav Kratochvíl
ÚTIA AV ČR, Czech Republic; radim@utia.cas.cz

The paper deals with the problem studied in our previous paper published in Int. J. Approx. Reasoning, which raised new questions rather than brought solutions. Thus, the current contribution also tries to answer the ever-lasting question: Which belief function entropies described in the literature can detect optimal models? Nevertheless, here, we approach the problem in a different way. We try to find out the entropy functions that are indirectly proportional to the informative content of belief functions, i.e., the more informative the belief function, the lower its entropy.

11:10am - 11:35am

Aggregation of the distortion models induced by the KL divergence and Euclidean distance

Ignacio Montes
University of Oviedo, Spain; imontes@uniovi.es

Distortion or neighbourhood models are tools within the imprecise probability theory that allow to robustify a probability measure. These are built by considering the closed ball around a probability measure with a given radius and using a distorting function to compare probability measures. These include well-known models such as the linear vacuous, pari-mutuel or total variation models. In this contribution we focus on the distortion models that arise from considering the Euclidean distance or the Kullback-Leibler divergence as distorting functions, and analyse their behaviour under different aggregation rules: conjunction, disjunction or convex mixtures.

11:35am - 12:00pm

A comparative analysis of aggregation rules for coherent lower previsions

Juan Jesús Salamanca, Ignacio Montes, Enrique Miranda
University of Oviedo, Spain; mirandaenrique@uniovi.es

We consider the problem of aggregating information provided by experts, when this information is expressed by means of coherent lower previsions. These constitute a framework general enough so as to include as particular cases not only probability measures but also the majority of models from the imprecise probability theory. Although the aggregation problem has already been tackled in the literature, our contribution provides a unified view by putting together a number of rationality criteria and aggregation rules studied in different papers. Specifically, we consider five aggregation rules, twelve rationality criteria and provide a detailed analysis of the properties satisfied by each rule.

Contributed Session 3: Dependence Modeling and Applications

Time: Wednesday, 04/Sept/2024: 1:30pm - 2:50pm · *Location:* Blue Lecture Room
Session Chair: Songkiat Sumetkijakan

1:30pm - 1:50pm

Estimation of Conditional Value-at-Risk in Linear Model

Jana Jurečková[1], Jan Picek[2], Jan Kalina[3]

[1]Institute of Information Theory and Automation AV ČR, Czech Republic (Senior Research Fellow); [2]Technical University in Liberec, Czech Republic; [3]Institute of Computer Science AV ČR, Czech Republic; jurecko@karlin.mff.cuni.cz

We monitor the risk of a loss Z of a portfolio; the risk is measured by the conditional value-at-risk (CVaR). However, instead of Z we can only observe its image Y = β0 + Xβ + Z in the linear model with the covariates whose intensities are not under our control. Our proposed estimator of CVaR(Z) is based on the averaged two-step regression quantile combined with an R-estimate of regression parameters. This method also enables a joint estimation of CVaR(Z) and value-at-risk VaR(Z).

1:50pm - 2:10pm

Applied Statistics in industry: defining an appropriate target variable and analysing factors affecting aluminium ingot quality

Manuela Schreyer[1], Marco Tschimpke[2], Alexander Gerber[3], Steffen Neubert[3], Wolfgang Trutschnig[2]

[1]AMAG Austria Metall GmbH, Austria; [2]Department for Artificial Intelligence & Human Interfaces, University of Salzburg, Austria; [3]AMAG Casting GmbH, Austria; manuela.schreyer@amag.at

Methods from the field of data science are becoming increasingly important in various areas of the economy. One of the main applications in industry is the continuous improvement of product quality, which is becoming increasingly challenging due to growing quality demands. A variety of statistical and machine learning methods are available for this purpose. However, to use these methods for a specific application, it is necessary to define an appropriate target variable. This can be a significant challenge due to the extensive product portfolio. The target variable must cover many cases and remain comparable for all cases under consideration. Limiting it to a single case would often result in a database too small for any analysis. This work aims at improving the quality of aluminium plates for the aerospace industry by analysing casting process data. The focus is on the development of a target value for this issue. Additionally, a brief outlook is given on the subsequent analysis of possible influencing variables and the development of a predictive model of product quality.

2:10pm - 2:30pm

Minimizing the expected absolute difference with a fixed continuous copula

Juan Baz, Irene Díaz, Susana Montes

Universidad de Oviedo, Spain; bazjuan@uniovi.es

Given a fixed random variable X and a fixed continuous copula C, the existence of a random variable Y such that (X,Y) has copula C and the quantity $E[|X-Y|]$ is minimum is proved. The proof, based on the Berge Maximum Theorem, also gives the expression of the distribution function of Y and implies a monotonicity property. In addition, the explicit solution for some of the most relevant copulas is stated.

2:30pm - 2:50pm

Singular components of shock model copulas

Tomaž Košir, Matjaž Omladič

University of Ljubljana, Slovenia; tomaz.kosir@fmf.uni-lj.si

We describe singular components for three families of shock induced copulas that are of importance in many applications. These are Marshall (also called Marshall–Olkin), reflected maxmin (RMM for short), and maxmin families of copulas. Although it is generally believed that shock model copulas are singular, both RMM and maxmin families contain nontrivial cases of members that are absolutely continuous.

Contributed Session 4: Dealing with Complex Data

Time: Wednesday, 04/Sept/2024: 1:30pm - 2:50pm · *Location:* Green Lecture Room
Session Chair: Patrick Benjamin Langthaler

1:30pm - 1:50pm

Sensitivity analysis on the choice of the metric on Cronbach's α coefficient for interval-valued data in questionnaires

José García-García[1], María Ángeles Gil[1,2], María Asunción Lubiano[1]
[1]University of Oviedo, Spain; [2]Spanish Royal Academy of Sciences, Spain; garciagarjose@uniovi.es

Interval-valued scales have become an efficient alternative to conventional single-point scales for capturing richer information in questionnaires measuring imprecise human traits. Thus, new statistical techniques are being developed to analyze this type of data. For instance, to quantify internal consistency reliability of interval-valued scale-based items in a construct, the definition of Cronbach's α coefficient can be extended by considering the family of θ-metrics between intervals. Since theoretical results are shown not to be informative enough, the impact of metric choice is discussed through a Monte Carlo study concerning different scenarios of a simulation procedure mimicking human rating.

1:50pm - 2:10pm

Addressing Incomplete Data in Two-way Contingency Tables with Three-Level Variables

Pitchayanin Makapawee, Kanyawee Kamkongkaew, Monchai Kooakachai
Department of Mathematics and Computer Science, Faculty of Science, Chulalongkorn University, Thailand; Monchai.k@chula.ac.th

Handling incomplete, imprecise, or missing observations in two-way contingency tables presents a challenge in accurately assigning uncertain data points to appropriate groups. Various approaches have been proposed to address this issue. Traditionally, the Expectation and Maximization (EM) algorithm has been utilized to estimate the group of incomplete observations. Subsequently, two alternative methods, derived from modifications to the formula used in the E-step of the algorithm, were introduced. However, these methods have mainly been evaluated in scenarios where categorical variables have two levels. This study aims to broaden the comparison of these methods to contingency tables with categorical variables featuring three levels. Through a simulation study, our findings indicate that while the classical EM algorithm generally per- forms well, there are specific parameter configurations where alternative methods demonstrate superior performance.

2:10pm - 2:30pm

Robustness of statistical methods for modelling paired count data

Marta Nai Ruscone
University of Genova, Italy; marta.nairuscone@unige.it

Bivariate Poisson models are appropriate for modeling paired count data. However the bivariate Poisson model does not allow for negative dependence structure, therefore it is necessary to consider alternatives, which can produce both positive and negative dependence. A natural way is to consider copulas to generate various bivariate discrete distributions. While such models exist in the literature, the issue of choosing a suitable copula has been overlooked so far. Different copulas lead to different structure, any copula misspecification can render the inference useless. In this work, we consider bivariate Poisson models generated with a copula and investigate its robustness under outliers contamination and model misspecification. Particular focus is given on the robustness of copula related parameters.

2:30pm - 2:50pm

Multi-way contingency tables with uniform margins

Elisa Perrone[1], Roberto Fontana[2], Fabio Rapallo[3]
[1]Eindhoven University of Technology, The Netherlands; [2]Politecnico di Torino, Italy; [3]Università di Genova, Italy; e.perrone@tue.nl

We study the problem of transforming a multi-way contingency table into an equivalent table with uniform margins and same dependence structure. Such a problem relates to recent developments in copula modeling for discrete random vectors. Here, we focus on three-way binary tables and show that, even in such a simple case, the situation is quite different than for two-way tables. Many more constraints are needed to ensure a unique solution to the problem. Therefore, the uniqueness of the transformed table is subject to arbitrary choices of the practitioner. We illustrate the theory through some examples, and conclude with a discussion on the topic and future research directions.

Invited Session 7: Latest Trends in Clustering and Classification Methods

Time: Wednesday, 04/Sept/2024: 3:10pm - 5:15pm · *Location:* Blue Lecture Room
Session Chair: Marta Nai Ruscone

3:10pm - 3:35pm

An EM Stopping Rule for Avoiding Degeneracy in Gaussian-based Clustering with Missing Data

Christophe Biernacki[1], Vincent Vandewalle[1,2]
[1]Inria, France; [2]Univ. Côte d'Azur, France; christophe.biernacki@inria.fr

Missing data frequency increases with the growing size of multivariate modern datasets. In Gaussian model-based clustering, the EM algorithm easily takes into account such data but the degeneracy problem is dramatically aggravated during the EM runs: parameter degeneracy is quite slow and also more frequent than with complete data. Consequently, parameter degenerated solutions may be confused with valuable parameter solutions and, in addition, computing time may be wasted through wrong runs. In this work, a simple and low informational condition on the latent partition allows to propose a very simple partition-based stopping rule of EM which shows good behavior on numerical experiments.

3:35pm - 4:00pm

Variable selection for clustering three-way data

Mackenzie Rose Neal, Paul David McNicholas
McMaster University, Canada; nealm6@mcmaster.ca

Ample work on dimension reduction for multivariate model-based clustering has been conducted; however, to date, relatively few dimension reduction methods have been presented in the matrix-variate paradigm. Such work is, for example, useful for modelling data arising from longitudinal studies with multiple responses, multivariate repeated measures data, or image data. Similar to the multivariate paradigm, issues persist when clustering data with noisy and uninformative variables within the matrix-variate paradigm. Thus, a variable selection algorithm for the matrix-variate paradigm is presented and tested on real datasets.

4:00pm - 4:25pm

Clustering on the d-dimensional sphere: Latest trends and the role of Poisson kernel-based density models

Marianthi Markatou
SUNY University at Buffalo, United States of America; markatou@buffalo.edu

Many data sets can be analyzed as unit vectors on a d-dimensional sphere. Specific examples include clustering of documents, the study of comets, and data that arise in the analysis of microarray experiments where one is interested in identifying groups of genes that are functionally related.

We present a discussion of the latest algorithmic trends for clustering data on the d-dimensional unit sphere; we then discuss the role that Poisson kernel-based models play in clustering d-dimensional spherical data. The relationship of these models with the Brownian motion several other classes of densities is illustrated. Furthermore, we present sampling algorithms for data generation from the Poisson kernel-density model and illustrate the usefulness of Poisson kernel-based clustering on several data examples.

4:25pm - 4:50pm

Clustering ordinal data via parsimonious models

Roberto Rocci, Monia Ranalli
Sapienza University of Rome, Italy; roberto.rocci@uniroma1.it

This review presents some parsimonious models to cluster two-way and three-way ordinal data. They are formulated has a reparameterization of a finite mixture of Gaussians that is partially observed through a discretization of its variates. Model parameters are estimated

using a composite likelihood approach in order to reduce the numerical complexity. The parsimony is obtained by reducing the dimensionality of the variable's space within or between the components

4:50pm - 5:15pm

Correlation-based hierarchical clustering with spatial constraints

Fabrizio Durante
University of Salento, Italy; fabrizio.durante@unisalento.it

Clustering algorithms aim at identifying association and dependence within a vector of n ≥ 3 random variables and creating sub-vectors of variables that are considered to be similar according to a given criterion. Such methods are particularly useful as a pre-processing step in the building of a multivariate stochastic model and can be interpreted as an unsupervised learning method. In this talk, we present some novel algorithms that use some additional information about the involved random variables in order to guide the clustering process in a semi-supervised learning setting. In particular, we exploit the geometric structure of the space of correlation matrices in order to adopt a convenient way to merge different sources of information.

Invited Session 8: Robust Statistical Methods I

Time: Wednesday, 04/Sept/2024: 3:10pm - 5:40pm · *Location:* Green Lecture Room
Session Chair: Beatriz Sinova Fernández

3:10pm - 3:35pm

Low-rank Approximation of Data Matrices Using Robust Sparse Principal Component Analysis

Pia Pfeiffer, Peter Filzmoser
TU Wien, Austria; peter.filzmoser@tuwien.ac.at

The estimation of principal components can be influenced by outlying observations, so-called row-wise outliers. For high-dimensional data, it becomes more and more likely that an observation contains outlying cells, which would lead to many row-wise outliers and a breakdown of traditional robust methods. In this case, it is preferable to achieve protection against cell-wise outliers. We present various approaches for principal component analysis that lead to row-wise and cell-wise robustness. Moreover, we focus on sparse methods that

enforce zeros in the loadings matrix and thus simplify the interpretation.

3:35pm - 4:00pm

Common Models of Errors in Variables

Henrik Kaiser
Paris Lodron University of Salzburg, Austria; henrik.kaiser@plus.ac.at

Blurred observations can be described in multiple ways. If the errors occur as a scaling or shifting variable, we find ourselves in the frequently employed multiplicative or additive model of errors in variables. In these, the determination of distribution or density of the unblurred variable, under the assumption of independent errors, corresponds to some kind of deconvolution problem. We briefly present both models and point out their parallels. Special focus lies on the additive model, which was first studied in the late 1980s and is still subject of further research. After a discussion of the traditional approach by virtue of Fourier transforms, we outline its developments until today. A distinction between known and unknown errors is made.

4:00pm - 4:25pm

Robust Sparse PCA with R

Valentin Todorov
UNIDO, Austria; valentin@todorov.at

Numerous methods have been proposed for improving the interpretability of PCA results by setting to zero coefficients in the components and thus creating sparse solutions. Few of these methods are capable for identifying atypical observations in the data and providing sparse and robust to outliers solutions at the same time. And finally, from a practical point of view, the availability of software is of utmost importance for the use of these methods in data analysis. We compare three methods for sparse and robust PCA for which R implementations are available at the Comprehensive R Archive Network (CRAN) and illustrate them on real data examples.

4:25pm - 4:50pm

Aggregation of random initializations for robust linear clustering

Pedro Cesar Álvarez-Esteban, Luis Angel García-Escudero, Agustín Mayo-Iscar
University of Valladolid, Spain; lagarcia@uva.es

Outliers can be very harmful to clustering methods, to the extent that even a small fraction of outliers can significantly impact their performance. Consequently, the use of robust clustering procedures has been advocated. A robust linear clustering procedure will be reviewed here, where robustness is pursued by applying a trimming approach. The proposed algorithm for its practical implementation is based on concentration steps, analogous to those applied in well-known high-breakdown robust multivariate procedures.

However, correct initialization is key for the algorithm's adequate performance. A large number of random initializations is often required to guarantee avoidance of local minima of the robust linear clustering target function. An alternative "ensemble initialization" method will be given to try to overcome this problem.

4:50pm - 5:15pm

Fuzzy S-estimators

Beatriz Sinova[1], Stefan Van Aelst[2]
[1]Universidad de Oviedo, Spain; [2]KU Leuven, Belgium; stefan.vanaelst@kuleuven.be

The central tendency and scale of the distribution of a random fuzzy number have been estimated through several measures in the literature. The best-known measures, the Aumann-type mean and the standard deviation, lack robustness and become unreliable under data contamination. With respect to the other approaches, with a more robust behaviour, scale estimators corresponding to the fuzzy medians and fuzzy trimmed means are available, but there is not any natural choice regarding M-estimators of location. In the real-valued setting, S-estimators of location present a similar behaviour to M-estimators and they are associated with a scale estimator. Therefore, the notion of S-estimator will be adapted to the fuzzy-valued case and its practical interest will be illustrated by means of some simulation studies.

5:15pm - 5:40pm

Fuzzy medoids based on an L1-type distance

Beatriz Sinova Fernández[1], Sergio Palacio Vega[2], María Ángeles Gil[1,3]

[1]University of Oviedo, Spain; [2]McKinsey & Company; [3]Spanish Royal Academy of Sciences, C/ Valverde, 22, 28004, Madrid, Spain; sinovabeatriz@uniovi.es

The first robust central tendency measures for fuzzy number-valued data introduced in the literature were extensions of the notion of median in the real-valued settings. In particular, they were defined as the fuzzy numbers that minimize the mean distance to the sample fuzzy number-valued observations. In order to solve the minimization problem, only L1-type metrics were considered in the fuzzy-valued settings, and the corresponding measures were proven to present some interesting properties. However, these approaches do not necessarily keep the shape of the sample data, and they do not have to coincide with a sample observation like it happens in real-valued scenarios. The aim is to propose new location measures (fuzzy medoids) by restricting the previous minimization problem to the set of fuzzy observations. Their practical interest is highlighted by means of a real-life example.

Plenary 4: Plenary Speaker Susanne Saminger-Platz

Time: Thursday, 05/Sept/2024: 9:00am - 9:55am · *Location:* Blue Lecture Room
Session Chair: Sebastian Fuchs

Copulas and Triangular Norms: Selected Commonalities and Differences

Susanne Saminger-Platz
Johannes Kepler University Linz, Austria; susanne.saminger-platz@jku.at

Copulas and triangular norms share a common history in the context of probabilistic metric space, but both have roots way back in different mathematical disciplines and continued to develop further in different fields of applications as well as theory. In this contribution, we discuss some questions posed independently in each respective fields and highlight related results as such providing a small tour on selected commonalities and differences of copulas and t-norms.

Invited Session 10: Robustness for Dependent Data

Time: Thursday, 05/Sept/2024: 10:20am - 12:00pm · *Location:* Green Lecture Room
Session Chair: Una Radojicic

10:20am - 10:45am

Order determination in second-order source separation models using data augmentation

Una Radojicic[1], Klaus Nordhausen[2]

[1]TU Wien Austria; [2]University of Jyväskylä, Finland; klaus.k.nordhausen@jyu.fi

We propose a robust estimator for the number of latent components in an internal noise model within the second-order source separation (SOS) framework. Our approach utilizes a data augmentation strategy in conjunction with the robust SOS approach eSAM-AMUSE, which combines information from eigenvalues and variations of eigenvectors of eSAM-AMUSE. The resulting dimension estimate can be visualized using a ladle plot. Through a simulation study, we demonstrate the superior properties of the new estimator, which outperforms the bootstrap-based AMUSEladle estimator.

10:45am - 11:10am

Functional Outlier Detection

Jeremy Oguamalam, Una Radojicic, Peter Filzmoser
TU Wien, Austria; una.radojicic@tuwien.ac.at

Functional data analysis is a sub-field of statistics concerned with data generated by stochastic processes defined on an (in-)finite interval. Unlike traditional multivariate observations, functional data are observed at individual time points throughout the domain. This unique data structure necessitates novel approaches for analysis. The robustness of these methods is crucial to ensure reliable results, particularly in the presence of anomalies arising from measurement errors or high data variability. This study explores various frameworks designed to address these challenges, demonstrating their efficacy through simulations and real-world data analysis.

11:10am - 11:35am

Robust ARIMA versus LSTM

Christophe Croux
KU Leuven, Belgium; christophe.croux@kuleuven.be

NA

11:35am - 12:00pm

The Joint Weighted Average (JWA) Operator

Stephen Broomell[1], Christian Wagner[2]

[1]Purdue University, United States of America; [2]Lab for Uncertainty in Data and Decision Making, University of Nottingham, Nottingham, UK; broomell@purdue.edu

Information aggregation is a vital tool for human and ma- chine decision making in the presence of uncertainty. Traditionally, approaches to aggregation broadly diverge into two categories, those which attribute a worth or weight to information sources and those which at- tribute said worth to the evidence arising from said sources. The latter is pervasive in the physical sciences, underpinning linear order statistics and enabling non-linear aggregation. The former is popular in the social sciences, providing interpretable insight on the sources. While prior work has identified the need to apply both approaches simultaneously, it has yet to conceptually integrate both approaches and provide a semantic in- terpretation of the arising aggregation approach. Here, we conceptually integrate both approaches in a novel joint weighted averaging operator. We leverage compositional geometry to underpin this integration, show- ing how it provides a systematic basis for the combination of weighted aggregation operators–which has thus far not been considered in the lit- erature. We proceed to show how the resulting operator systematically integrates a priori beliefs about the worth of both sources and evidence, reflecting the semantic integration of both weighting strategies. We con- clude and highlight the potential of the operator across disciplines, from machine learning to psychology.

Invited Session 9: Adaptive Computing and Decision Making

Time: Thursday, 05/Sept/2024: 10:20am - 12:00pm · *Location:* Blue Lecture Room
Session Chair: Simon Hirlaender

10:20am - 10:45am

Exploring Deep Reinforcement Learning algorithms for enhanced HVAC control

Antonio Manjavacas, Alejandro Campoy-Nieves, Miguel Molina-Solana, Juan Gómez-Romero
University of Granada, Spain; manjavacas@ugr.es

Heating, Ventilation, and Air Conditioning (HVAC) systems are one of the major sources of energy consumption in buildings. Typically, HVAC control has relied on reactive controllers, which often lack the ability to adapt to specific building dynamics. In recent years, Deep Reinforcement Learning (DRL) algorithms have emerged as a potential alternative to reactive controllers. However, these solutions are still immature, with a lack of standardisation and difficulties in real-world deployment. This paper presents an empirical evaluation of several state-of-the-art DRL algorithms for HVAC control, highlighting their main strengths and limitations. We emphasize the importance of using standard frameworks for comparative analysis, enabling a more comprehensive assessment of these innovative approaches to HVAC control.

10:45am - 11:10am

Python-Based Reinforcement Learning on Simulink Models

Georg Schäfer[1,2,3], Max Schirl[2,3], Jakob Rehrl[1,2], Stefan Huber[1,2], Simon Hirlaender[3]
[1]Josef Ressel Centre for Intelligent and Secure Industrial Automation; [2]Salzburg University of Applied Sciences; [3]Paris Lodron University of Salzburg; georg.schaefer@fh-salzburg.ac.at

This paper proposes a framework for training Reinforcement Learning agents using Python in conjunction with Simulink models. Leveraging Python's superior customization options and popular libraries like Stable Baselines3, we aim to bridge the gap between the established Simulink environment and the flexibility of Python for training bleeding edge agents. Our approach is demonstrated on the Quanser Aero 2, a versatile dual-rotor helicopter. We show that policies trained on Simulink models can be seamlessly transferred to the real system, enabling efficient development and deployment of Reinforcement Learning agents for control tasks. Through systematic integration steps, including C-code generation from Simulink, DLL compilation, and Python interface development, we establish a robust framework for training agents on Simulink models. Experimental results demonstrate the effectiveness of our approach, surpassing previous efforts and highlighting the potential of combining Simulink with Python for Reinforcement Learning research and applications.

11:10am - 11:35am

Multi-Agent Reinforcement Learning and its application to wireless network communication

Sabrina Pochaba[1,2], Peter Dorfanger[1], Matthias Herlich[1], Roland Kwitt[2], Simon Hirlaender[2]
[1]Salzburg Research Forschungsgesellschaft mbH, Austria; [2]Paris-Lodron-University, Salzburg, Austria;
sabrina.pochaba@salzburgresearch.at

In real world problems, there are often many players interacting with each other while following their own goals.
Although the subject of Machine Learning (ML) makes great progress in picturing good learning strategies, it lacks in considering many player strategies.
Thus, the subject of Multi-Agent Reinforcement Learning (MARL) tries to address this need. Here, the learning strategies of Reinforcement Learning (RL), where one agent interact with an environment to solve a task, is extended to multiple players trying to solve their own tasks properly.
To do so, algorithms from RL are used and applied to multi-agent settings, paired with game theoretical aspects to forecast multi-player behaviour.
After explaining the main concepts of MARL with its challenges and advantages, we apply MARL to a setting of wireless communication. Here, the multi-agent setting can take the different amount of communicating devices into account that communicate in real-world communication scenarios.

11:35am - 12:00pm

Deep Meta Reinforcement Learning for Rapid Adaptation in Linear Markov Decision Processes: Applications to CERN's AWAKE Project

Simon Hirlaender[1], Sabrina Pochaba[1], Lukas Lamminger[1], Andrea Santamaria Garcia[2], Annika Eichler[3], Chenran Xu[2], Jan Kaiser[3], Verena Kain[4]
[1]IDA Lab, Austria; [2]KIT, Karlsruhe; [3]DESY, Hamburg; [4]CERN, Geneva; simon.hirlaender@plus.ac.at

Real-world reinforcement learning (RL) applications face challenges, such as the need for numerous interactions and achieving stable training under dynamic conditions. Meta-RL emerges as a solution, particularly in environments where simulations cannot perfectly mimic real-world conditions. This study demonstrates Meta-RL's potential in the CERN's AWAKE project, focusing on the electron line's control. By incorporating Model-Agnostic Meta-Learning (MAML), we showcase how Meta-RL facilitates rapid adaptation to environmental changes with minimal interaction steps. Our findings indicate Meta-RL's efficacy in managing Partially Observable Markov Decision Processes (POMDPs) with evolving hidden parameters, underlining its significance in high-dimensional control challenges prevalent in particle physics experiments and beyond.

Contributed Session 5: Statistical Learning and Applications

Time: Friday, 06/Sept/2024: 9:00am - 10:40am · *Location:* Blue Lecture Room
Session Chair: Simon Hirlaender

9:00am - 9:20am

Optimal change-point sequential detection in autoregressive time series

Serge Pergamenchtchikov, Roman Tenzin
Université de Rouen Normandie, France; mr.roman.tenzin@gmail.com

In this study, we consider a sequential detection problem in Bayesian setting for autoregressive times series based on a bounded number of observations under the condition that the post-change parameters are unknown. To this end we propose a new truncated sequential detection
method through the theory developed in Pchelintsev, Pergamenchtchikov and Tenzin (2024) for the statistical model with known post-change distributions. Based on the developed method, the quickest detection algorithm is proposed, that is, optimal in terms of the minimum mean time delay with the probability of a false alarm limited by some fixed known threshold.

9:20am - 9:40am

Smoke Emission Detection by Deep Learning

Jo-Huan Lee, Chi-Bin Cheng
Tamkang University, Taiwan; cbcheng@mail.tku.edu.tw

Air pollution has long been a pressing public concern. Before the advent of fixed monitoring equipment, authorities relied solely on human monitoring and citizen reports to collect evidence of pollution events. In an effort to address this issue, the cities of Taichung and Kaohsiung in Taiwan implemented cloud-based smart surveillance systems for automatically detecting and recognizing smoke emission incidents. However, these systems were based on traditional image processing techniques, which limited their detection accuracy. To improve performance, this study introduced a YOLOv7 model into the existing surveillance systems. The YOLO series, based on deep learning techniques, have gained widespread adoption for real-time video analysis tasks due to their high accuracy and efficiency. Moreover, the transfer learning capability of YOLO models allows for fine-tuning on new training samples, enabling adaptation to specific application domains. In this study, the YOLOv7 model was trained on 60,076 labeled photos and subsequently tested on real-world videos collected from the monitoring sites in Taichung and Kaohsiung. The results demonstrated a significant improvement in performance compared to the traditional image processing techniques. Specifically, the overall accuracy of smoke detection increased from 37% to 73%, while the average F1 score improved from 0.33 to 0.64.

9:40am - 10:00am

Learning in Reproducing Kernel Hilbert Spaces for Orbits of Iterated Function Systems

Priyanka Roy[1,2], Susanne Saminger-Platz[2]
[1]Software Competence Center Hagenberg, A-4232 Hagenberg, Austria; [2]Johannes Kepler University Linz, Institute for Math. Methods in Medicine and Data-Based Modeling, A-4020 Linz, Austria; priyanka.roy@jku.at

One of the problems in learning theory is to approximate a function f that underlies the relationship between (x,y), i.e., y=f(x) based on sample points (x_t,y_t)_{t=1}^{n}. Given the sample points, the function can be approximated in Reproducing Kernel Hilbert Spaces through various learning algorithms. However, it is usually customary to consider the sampling nature as independent and identically distributed (i.i.d.) in the context of learning theory. We leverage the i.i.d. assumption by considering an input sample trajectory (x_t)_{t in N} obtained via an Iterated Function System that is a particular Markov Chain, with (y_t)_{t in N} corresponding to an observation sequence when the model is in the corresponding state x_t. We discuss learning bounds for approximation for such a process.

10:00am - 10:20am

Data augmentation with ChatGPT for assessing subject alignment

Louisa Kontoghiorghes[1], Ana Colubi[2]
[1]King's College London, United Kingdom; [2]King's College London, United Kingdom; louisa@kontoghiorghes.com

As a statistical tool, topic modeling requires replication. Topics are identified as distributions over a set of words, and documents are described as mixtures of topics. Sometimes, text information is available through a unique document that covers well-defined subjects, but without replication, the intuitive representation of the document as a mixture of latent topics cannot be derived. Nevertheless, the available document can potentially generate more knowledge (not yet available) that could assist in employing compelling text mining tools. The proposal is to use ChatGPT to mimic the process of generating such knowledge. To illustrate the approach, a research proposal is used as the initial document. The aim is to verify if a given piece of research aligns with the subjects of the research proposal or not combining text mining and statistical tools.

10:20am - 10:40am

The impact of sentiment in social network communication

Soveatin Kuntur, Anna Krystyna Cena
Warsaw University of Technology, Poland; anna.cena@pw.edu.pl

The pervasive influence of social media in shaping public discourse and informing societal and policy matters cannot be denied. Several attempts were made in order to investigate the nature / dynamics of this process, both, from social networks and sentiment analysis perspective. Few more has set a goal of combining aforementioned approaches. Contribution of this paper continues further in this direction. We will take a closer look into the network specific measures that allow to aggregate the impact of a given user. The main focus of this contribution, however, is to investigate the interaction of content with varied polarities on to the social network structure itself, and the final assessment of each user's importance. We analyzed a Twitter dataset, including various contentious socio-political topics, assessing the structure of engagement networks (including retweets, mentions, etc.) and the sentiment of the tweets. We explored multiple methods of integrating sentiment into the network, such as considering it as a weight factor for connections.

Contributed Session 6: Dependence Modeling II

Time: Friday, 06/Sept/2024: 9:00am - 10:40am · *Location:* Green Lecture Room
Session Chair: Elisa Perrone

9:00am - 9:20am

Characterizations of Multivariate Implicit Dependence Copulas

Tippawan Santiwipanont, Songkiat Sumetkijakan, Noppawit Yanpaisan
Chulalongkorn University, Thailand; songkiat.s@chula.ac.th

The copula C of continuously distributed random variables $X_1,dots,X_d$ is said to be an emph{implicit dependence copula} if there are Borel functions $alpha_1,dots,alpha_d$
such that $alpha_1(X_1),dots,alpha_d(X_d)$ are equal almost surely and continuously distributed, that is their common distribution function is continuous.
Bivariate implicit dependence copulas have recently been characterized in terms of a generalized Markov product.
In this manuscript, the characterizations are extended to the multivariate case in terms of a product of d copulas, called $mathscr{A}$-product where $mathscr{A}$ is a class of copulas A_t, $tin[0,1]$.
The class of implicit dependence d-copulas are characterized as $mathscr{A}$-products of d complete dependence copulas.
Explicit forms of the joining copulas A_t are obtained when the functions $alpha_i$ are countably piecewise monotonic surjections.

9:20am - 9:40am

Patchwork methods for multivariate quasi-copulas

Matjaz Omladic, Tomaz Kosir
University of Ljubljana, Faculty of Mathematics and Physycs, Slovenia; matjaz@omladic.net

In this contribution, we exhibit the work presented in a series of recent papers of the authors jointly with some other members of the Ljubljana Copula Group. These results appear to be new since the patchwork approach has been limited to copulas so far. In the bivariate case, we can find actual
constructions of desired patches. The multivariate case seems much more involved, so these constructions cannot be extended to higher dimensions. Our main result is to present the necessary and sufficient conditions on the patch's domain and the values on its border for the desired multivariate quasi-copula to exist. We also give all possible solutions to the problem.

9:40am - 10:00am

Quantifying directed dependence with Kendall's tau

Carsten Limbach, Sebastian Fuchs
Pairs Lodron University of Salzburg, Austria; carsten.limbach@plus.ac.at

The Markov product of copulas is a powerful tool for modeling directed dependence between two random variables X and Y, and its evaluation via certain concordance measures such as Spearman's rho and Spearman's footrule allows for quantifying different aspects of directed dependence, and thus, different types of influence that X can exert on Y. Motivated by the recent publications "Quantifying directed dependence via dimension reduction" by Fuchs (2024) and "On the copula correlation ratio and its generalization" by Shih and Emura (2021), we apply Kendall's tau to the Markov product which leads to a dependence measure determining the degree of agreement among the conditional distributions of Y given $X=x$.

10:00am - 10:20am

Some notes on bivariate lower semilinear copulas

Lea Maislinger, Wolfgang Trutschnig
University of Salzburg, Austria; lea.maislinger@plus.ac.at

We revisit the class of bivariate lower semilinear copulas (LSL) as introduced by Durante et al. in 2008 (see [2]), view LSL copulas from the Markov kernel perspective, and use the kernel to derive a simple formula for the singular mass of arbitrary LSL copulas. Complementing these results we derive simple formulas for Kendall's τ and Spearman's ρ and study monotonicity/ordering properties like positive quadrant dependence (PQD), left tail dependence (LTD) and stochastic increasingness (SI).

10:20am - 10:40am

Robust estimation for step-stress experiments with non-destructive one-shot devices

NARAYANASWAMY BALAKRISHNAN[1], MARÍA JAENADA[2], LEANDRO PARDO[2]
[1]McMaster University; [2]Complutense University of Madrid, Spain; mjaenada@ucm.es

Dealing with censored data is an important concern in reliability and survival analysis. Interval censored data usually arises in reliability tests when failure times are known only to lie within an interval instead but cannot be observed exactly. That is the case of non-destructive one-shot devices. Besides, most of nowadays devices are frequently highly reliable with large lifetimes, and then long experimentation times would be needed for inference under normal operating conditions. Alternatively, accelerated life tests (ALTs) shorten the lifetime of the devices by increasing one or more stress factors causing failure. Then, after suitable inference, results can be extrapolated to normal conditions. Step-stress ALT designs increase the stress level at which devices are tested throughout the experiment at some fixed times. Classical estimation methods based on the likelihood function of the lognormal lifetime distribution may get highly influenced by data contamination. In this work we propose a family of divergence-based robust estimators for step-stress experiments with non-destructive one-shot devices.

Plenary 5: Plenary Talk Christian Wagner

Time: Friday, 06/Sept/2024: 11:05am - 12:00pm · *Location:* Blue Lecture Room
Session Chair: Sebastian Fuchs

Intervals are tomorrow's numbers. An interdisciplinary view of why, and how we may get there.

Christian Wagner
Lab for Uncertainty in Data and Decision Making, University of Nottingham, Nottingham, UK; christian.wagner@nottingham.ac.uk

Intervals provide a convenient means to model uncertainty in data and play an important role across much of soft computing research, whether it is focussing on the development of aggregation or statistical operators for interval-valued data, or on approaches which leverage intervals to derive or analyse more complex models such as fuzzy sets.

I will start the talk by highlighting the broader inter- and trans-disciplinary potential intervals have in the context of explainable AI and specifically in the communication of uncertainty. Reflecting primarily on research within the Lab for Uncertainty in Data and Decision Making (LUCID), I will explain how we came to the conclusion that while our soft computing research in the sense outlined above was important, we had neglected to reflect on where such intervals actually come from in the real world. I will discuss what as a result has become a major research avenue at LUCID – an interdisciplinary body of work focussing on the elicitation and processing of vague and uncertain information from people. I will cover initial failures of our research and the inter-disciplinary approaches we took to address them, flagging open-source software tools in the process. Finally, I will highlight some of the (many) remaining challenges and research avenues which sit 'between disciplines' and which are impeding the much broader adoption of intervals as a primary data type in research and applications.

Invited Session 11: Random Sets and Fuzzy Random Variables

Time: Friday, 06/Sept/2024: 1:00pm - 1:50pm · *Location:* Blue Lecture Room
Session Chair: Pedro Terán

1:00pm - 1:25pm

Intersections of randomly translated sets

Tommaso Visonà
Universität Bern, Switzerland; tommaso.visona@unibe.ch

Let $\Xi_n=\{\xi_1,\dots,\xi_n\}$ be a sample of n independent points distributed on a regular closed element K of the extended convex ring in $\mathbb{R}^d$ according to a probability measure μ on K, admitting a density function. We consider random sets generated from the intersection of the translations of K by elements of Ξ_n, as

\begin{displaymath}
X_n=\bigcap_{i=1}^n (K-\xi_i).
\end{displaymath}

This work aims to show that scaled X_n as $n\to\infty$ converges in distribution to the zero cell of a Poisson hyperplane tessellation whose distribution is determined by the curvature measure of K and the behaviour of the density of μ near the boundary of K.

1:25pm - 1:50pm

On a linear regression model for star-shaped set-valued random elements

José Graña-Colubi[1], Gil Gonzalez-Rodriguez[1,2], Ana Belen Ramos-Guajardo[1,2]
[1]University of Oviedo, Spain; [2]INDUROT; ramosana@uniovi.es

The aim of this work is to propose a linear regression model between two random elements that take values in a particular star-shaped sets space. Random star-shaped sets will be introduced, as well as the corresponding framework to deal with them. This framework is based on a center-radial characterization in order to take advantage of the well-known properties of Hilbert spaces. The main properties of the proposed linear regression model will be analyzed.