ΠΑΝΕΠΙΣΤΗΜΙΟ
ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ
UNIVERSITY OF WEST ATTICA

35ο Πανελλήνιο και 1ο Διεθνές Συνέδριο Στατιστικής

Στατιστική στις Επιστήμες Υγείας

25-28 Μαΐου 2023

Πανεπιστήμιο Δυτικής Αττικής
Πανεπιστημιούπολη Αρχαίου Ελαιώνα

Πληροφορίες
https://gsi-conference.uniwa.gr

ΔΙΟΡΓΑΝΩΣΗ

ΕΛΛΗΝΙΚΟ ΣΤΑΤΙΣΤΙΚΟ ΙΝΣΤΙΤΟΥΤΟ

ΠΑΝΕΠΙΣΤΗΜΙΟ
ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ
UNIVERSITY OF WEST ATTICA

Τόμος Περιλήψεων του 35ου Πανελληνίου Συνεδρίου και πρώτου Διεθνούς Συνεδρίου Στατιστικής / Book of Abstracts of 35th Panhellenic Statistics Conference and First International Conference of Statistics Ελληνικό Στατιστικό Ινστιτούτο/ Greek Statistical Institute

Συντάκτες/Editors: Elefterios Angelis, Miltiadis Chalikias, Alexandros Karagrigoriou, Georgios Papageorgiou Τόπος έκδοσης/ Place of publication: Αθήνα, Ελλάδα/Athens, Greece

Έτος έκδοσης/ Year of publication: 2023

35ο Πανελλήνιο και 1ο Διεθνές Συνέδριο Στατιστικής

35th Panhellenic and 1st International Statistics Conference

ΠΡΟΓΡΑΜΜΑ & ΠΕΡΙΛΗΨΕΙΣ (ΜΕΡΟΣ Β)

Οι περιλήψεις που περιλαμβάνονται στον παρόντα Τόμο είναι αυτές που εμφανίζονται έγχρωμες, στο πρόγραμμα πιο κάτω.

PROGRAM & ABSTRACTS (PART B)

The abstracts included in the present Volume appear in color in the program below.

Ελληνικό Στατιστικό Ινστιτούτο/ Greek Statistical Institute

ΠΡΟΓΡΑΜΜΑ ΣΥΝΕΔΡΙΟΥ/ CONFERENCE SCHEDULE

ΠΕΜΠΤΗ 25 ΜΑΪΟΥ 2023/ THURSDAY MAY 25 2023

Αίθουσα Α/ Hall A: Κεντρικό Αμφιθέατρο/ Central Amphitheater

13:30 - 14:30	Εγγραφή Συνέδρων/ Registration of the Participants
Αίθουσα Α / Hall A 14:30 – 15:00	**ΕΝΑΡΞΗ ΤΟΥ ΣΥΝΕΔΡΙΟΥ – ΧΑΙΡΕΤΙΣΜΟΙ** **CONFERENCE OPENING – WELCOME ADDRESSES** **Προεδρεύων/Chair: Μ. Χαλικιάς, Ε. Παπαγεωργίου/ M. Chalikias, E. Papageorgiou**
	- Χαιρετισμοί – Welcome Addresses
Αίθουσα Α/ Hall A 15:00 - 17:30	**ΕΝΑΡΚΤΗΡΙΕΣ ΚΕΝΤΡΙΚΕΣ ΟΜΙΛΙΕΣ / PLENARY TALKS** **Προεδρεύουσα/Chair : Μ. Βαμβακάρη/M. Vamvakari**
15:00-15:30	**N. Balakrishnan** Linear Prediction
15:30-16:00	**S. Bar-Lev** An infinite class of exponential dispersion models for count data - a survey of current research
16:00-16:30	**S. Datta** Regression Analysis of a Future State Entry Time Distribution Conditional on a Past State Occupation
16:30-17:00	**N. Papadatos** The characteristic function of the discrete Cauchy distribution and the Cauchy-Cacoullos family of discrete distributions
17:00-17:30	**Ch. Charalambides** A class of power series q-distributions
17:30-18:00	**ΔΙΑΛΕΙΜΜΑ – ΚΑΦΕΣ – COFFEE BREAK**

Αίθουσα Α/ Hall A 18:00-19:20	**ΕΙΔΙΚΗ ΣΥΝΕΔΡΙΑ / SPECIAL SESSION** **Βραβείο Καλύτερου Νέου Στατιστικού/Best Young Statistician Award** **Προεδρεύων/Chair: Γ. Ψαρράκος / G. Psarrakos**
18:00 – 18:20	**M. Jaenada, N. Balakrishnan and L. Pardo** Step-stress test experiments under interval censored data with lognormal lifetime distribution
18:20 – 18:40	**G. Tzoumerkas and D. Fouskakis** Objective shrinkage priors via imaginary data
18:40 – 19:00	**V. E. Papageorgiou and G. Tsaklidis** An extended epidemiological particle filter for assessing infectious disease transmission. Application to covid-19 data in Italy
19:00 – 19:20	**E. N. Kalligeris, A. Karagrigoriou and Ch. Parpoula** On Stochastic Dynamic Modeling of Incidence Data

Προθάλαμος Συνεδριακού Κέντρου/ Conference Center Lobby **19:30**	***WELCOME COCKTAIL***

ΠΑΡΑΣΚΕΥΗ 26 ΜΑΪΟΥ 2023/ FRIDAY MAY 26 2023

Αίθουσα Α/ Hall A: Κεντρικό Αμφιθέατρο/ Central Amphitheater

Αίθουσα Β: / Hall B: Σ109Α

Αίθουσα Α/ Hall A	**«Ελένειο» Βραβείο Καλύτερης Διδακτορικής Διατριβής 2021-2022** **«Eleneio» Award for Best Doctoral Thesis 2021-2022** **Προεδρεύων/Chair: Δ. Φουσκάκης, Μ. Κατέρη / D. Fouskakis, M. Kateri**
08:45 – 09:15	**Ch. Thomadakis** Joint modeling of longitudinal and competing-risks data using cumulative incidence functions for the failure submodels and accounting for potential misclassified cause of failure through double sampling
Αίθουσα Α/ Hall A **09:20 - 11:00**	**ΠΡΟΣΚΕΚΛΗΜΕΝΕΣ ΟΜΙΛΙΕΣ / INVITED TALKS** **Stochastic Modeling and Statistical Methods: Advances and Applications Part I Προεδρεύων/Chair: Ε. Παπαγεωργίου / E. Papageorgiou**
09:20-09:45	**E. N. Kalligeris, V. S. Barbu, G. Hacques, L. Seifert and N. Vergne** Assessing the Dynamic of Visual-Motor Skill in Climbing via Drifting Markov Modeling
09:45-10:10	**P. M. Menendez, A. Felipe, M. Jaenada, L. Pardo** Applying Rényi's pseudodistance for robust Model Selection for independent not identically distributed observations
10:10-10:35	**D. Bagkavos, M. Guillen and J.P. Nielsen** Nonparametric conditional survival function estimation with plug-in bandwidth and robust model selection
10:35-11:00	**Ch. T. Nakas, L. E. Bantis, B. Brewer and B. Reiser** Statistical inference for ROC curves after the Box-Cox transformation with the use of the package `rocbc' in R
Αίθουσα Β/ Hall B **09:20 - 11:00**	**ΕΙΔΙΚΗ ΣΥΝΕΔΡΙΑ / SPECIAL SESSION** **Techniques & Applications of Data Analytics & Machine Learning I** **Προεδρεύων/Chair : Π. Οικονόμου / P. Economou**

09:20-09:40	**K. Skarlatos, P. E. Maravelakis, S. Bersimis** Techniques for detecting change points in multivariate data streams with an application in shipping
09:40-10:00	**E. Skamnia, P. Economou, S. Bersimis** Detecting outliers, break points and level shifts with an application in shipping industry
10:00-10:20	**G. Papageorgiou, P. Economou, S. Bersimis** Detecting hidden trends and patterns in shipping industry using text visualization
10:20-10:40	**G. Arvanitopoulos, P. Economou, S. Bersimis** Applications of text mining in health
10:40-11:00	**Z. Bartsioka, A. Charalabidis, P. E. Maravelakis and S. Bersimis** Monitoring learning outcomes using runs and scans
11:00-11:30	***ΔΙΑΛΕΙΜΜΑ – ΚΑΦΕΣ – COFFEE BREAK***

Αίθουσα Α/ **Hall A** **11:30 - 13:30**	**ΠΡΟΣΚΕΚΛΗΜΕΝΕΣ ΟΜΙΛΙΕΣ / INVITED TALKS** **Statistical Methods and Applications in Systems Assurance & Quality Part I** **Προεδρεύων/Chair : Π. Οικονόμου / P. Economou**
11:30-12:00	**S. Knoth** Some Stylized Facts of the Conditional Expected Delay (CED)
12:00-12:30	**C. Weiss** Stein EWMA Control Charts for Count Processes
12:30-13:00	**M. Stehlik** Data transformations and transfer functions
13:00-13:30	**N. Pavlidis** Dimensionality reduction and clustering
Αίθουσα Β/ **Hall B** **11:30 – 13:30**	**ΣΥΝΕΔΡΙΑ / CONTRIBUTED SESSION** **Actuarial and Financial Mathematics** **Προεδρεύων / Chair: Δ. Κωνσταντινίδης / D. Konstantinides**
11:30-11:50	**Ch. Skiadas, Y. Dimotikalis** Expanding the Life Tables to Include the Healthy Life Expectancy. The Case of Norway
11:50-12:10	**A. Papachristos, A. Bozikas** Actuarial Aspects of Subjective Survival Probabilities

12:10-12:30	**M.V. Boutsikas, D. J. Economides. and E. Vaggelatou** On the time and aggregate claim amount until the surplus drops below zero or reaches a safety level in a jump diffusion risk model
12:30-12:50	**~~D. Ioannides~~ CANCELLED** ~~Estimation the Probability of Default Cascades in Financial Markets~~
12:50-13:10	**S. Tzaninis** Martingales, change of measures and ruin probabilities for an inhomogeneous renewal risk model
13:10-13:30	**Ming Cheng, Dimitrios G. Konstantinides, Dingcheng Wang** Multivariate regularly varying insurance and financial risks in multi-dimensional risk models

13:30-14:30	***ΜΕΣΗΜΒΡΙΝΗ ΔΙΑΚΟΠΗ – MIDDAY BREAK***

Αίθουσα Α/ Hall A **14:30 - 15:30**	**ΠΡΟΣΚΕΚΛΗΜΕΝΕΣ ΟΜΙΛΙΕΣ / INVITED TALKS** **Εφαρμοσμένη Βιοστατιστική / Applied Biostatistics** **Προεδρεύων/Chair : Δ. Παναγιωτάκος/D. Panagiotakos**
14:30 – 15:00	**V. Sypsa** Modelling transmission of infectious diseases and assessing the impact of interventions: From HIV and antibiotic resistant bacteria to SARS-CoV-2
15:00 – 15:30	**F. F. Caballero** Elastic Net Regression models and their use to obtain a continuous measure of multimorbidity

Αίθουσα Α/ Hall A **15:40 – 17:00**	**ΕΙΔΙΚΗ ΣΥΝΕΔΡΙΑ / SPECIAL SESSION** **Asymptotic Statistics** **Προεδρεύων / Chair : Σ. Τρέβεζας / S. Trevezas**
15:40 – 16:00	**P. Karkalakis, P.-H. Cournède, G. Hermange, S. Trevezas** A Moment-type Estimator for the Generalized Gamma Model with Complete and Interval Censored Data –Application to First Division Time of HSC

16:00 – 16:20	**Ι. Οικονομίδης, Σ. Τρέβεζας** The Method of Moments vs Maximum Likelihood -An overestimated difference?
16:20 – 16:40	**Λ. Κορδαλής, Σ. Τρέβεζας** The Convolution Algebra of Multi-time Markov Renewal Chains and elements of Statistical Estimation
16:40 – 17:00	**S. Trevezas, G. Gavrilopoulos, I. Votsi** Nonparametric Maximum Likelihood Estimation for discrete time Denumerable Markov Chains

Αίθουσα Β/ Hall B **15:40 – 17:00**	**ΣΥΝΕΔΡΙΑ / CONTRIBUTED SESSION** **Χρονοσειρές – Προβλέψεις** **Προεδρεύων / Chair : Χ. Κίτσος / C. Kitsos**
15:40 – 16:00	**Γ. Κοσμά, Β. Καρυώτη** Ανάλυση και Πρόβλεψη Χρονοσειρών με δεδομένα εποχικής γρίπης
16:00 – 16:20	**Μ. Μάρκου, Β. Καρυώτη** Πρόβλεψη χρονοσειρών και μελλοντικών ακροτάτων σε αγχώδεις διαταραχές στον πληθυσμό της Ελλάδας.
16:20 – 16:40	**Σ. Μαλεφάκη, Π. Οικονόμου** Παρακολούθηση μακροπρόθεσμης σχέσης μεταξύ συνολοκληρωμένων χρονοσειρών
16:40 – 17:00	**Μ. Χαλικιάς** Βελτιστοποίηση Σχεδιασμών Επαναλαμβανόμενων μετρήσεων 2 αγωγών εκτίμηση άμεσων επιδράσεων (direct effect)

18:00 - 19:30	***ΙΣΤΟΡΙΚΟΣ ΠΕΡΙΠΑΤΟΣ ΣΤΗΝ ΑΘΗΝΑ – HISTORIC WALKING TOUR IN ATHENS***
20:30	***ΕΠΙΣΗΜΟ ΔΕΙΠΝΟ – CONFERENCE DINNER***

ΣΑΒΒΑΤΟ 27 ΜΑΪΟΥ 2023/ SATURDAY MAY 27 2023

Αίθουσα Α/ Hall A: Σ109Α

Αίθουσα Β/ Hall B: Σ109Β

Αίθουσα Α/ Hall A **09:00 - 11:00**	**ΣΥΝΕΔΡΙΑ / CONTRIBUTED SESSION** **Applied Probability – Applied Statistics** **Προεδρεύων/Chair : Α. Μπατσίδης / A. Batsidis**
09:00 – 9:20	**K. Bourazas, K. Fokianos, Ch. Panayiotou, M. Polycarpou** An Adaptive Kernel-based Multivariate CUSUM for Location Shifts
09:20-09:40	**K. Fountoukidis, D. L. Antzoulakos, A. C. Rakitzis** The Variable Sample Size and Sampling Interval Run Sum Max Chart
09:40-10:00	**T. Gkelsinis, V. S. Barbu** Recent advances in reliability indexes for semi-Markov repairable systems with applications to financial credit scoring
10:00-10:20	**K. Tasias, P. Mpistintzanos, F. Pekridis** A mission reliability redundancy model for an aircraft fleet with cold standby spare parts
10:20-10:40	**I. Triantafyllou** Combined m-consecutive-k-out-of-n: F & consecutive kc-out-of-n: F structures with cold standby redundancy
10:40-11:00	~~**D. Lyberopoulos, N. Macheras**~~ **CANCELLED** Martingale characterizations of mixed compound Poisson processes with ~~applications in risk theory~~
Αίθουσα Β/ Hall B **09:00-11:00**	**ΣΥΝΕΔΡΙΑ / CONTRIBUTED SESSION** **Εφαρμοσμένη Στατιστική – Εφαρμογές Στατιστικής** **Προεδρεύουσα : Θ. Μοσχονά/ Th. Moschona**
09:00 - 09:20	**Τ. Δάρας, Β. Παντούλα, Β. Μάνδικας** Τρόποι εφαρμογής της Γραμμικής Διαχωριστικής Ανάλυσης σε δεδομένα από το χώρο της Υγείας

09:20 – 09:40	**Σ. Δαφνής, Χ. Ε. Ζώτος, Α. Ραυτοπούλου, Γ. Κ. Παπαδόπουλος** Η Επίδραση Των Περιόδων Ανεργίας Στην Ατομική Υγεία
09:40 – 10:00	**Σ. Φόρτης, Σ. Ντελίκου, Α. Γιαννάκη, Α. Ξυδάκη, Κ. Μαγγανάς, Χ. Φούντζουλα, Ε. Παπαγεωργίου, Α. Κριεμπάρδης** Επικοινωνία κυττάρων του αίματος μέσω μικροκυστιδίων: Ο ρόλος τους στη δρεπανοκυτταρική αναιμία
10:00 – 10:20	**Π. Δρόσος, Ε. Παύλου, Σ. Φόρτης, Ε. Παπαγεωργίου, Κ. Σταμούλης, Σ. Βαλσάμη, Μ. Πολίτου, Α. Κριεμπάρδης** Αιμοστατικό δυναμικό αιμοπεταλίων αποθηκευμένων στο ψύχος
10:20 – 10:40	**Ε. Παύλου, Σ. Φόρτης, Μ. Σιουμάλα, Σ. Δασκαλάκη, Α. Γιαννιώτη, Β. Μπίρτσας, Ε. Παπαγεωργίου, Δ. Πετράς, Ε. Νομικού, Α. Κριεμπάρδης** Η επίδραση της αιμοκάθαρσης στην αιμόσταση
10:40 - 11:00	**Π. Κ. Ρεβέλου, Χ. Παππάς, Γ. Παπαδόπουλος, Π. Ταραντίλης** Αναγνώριση της βοτανικής προέλευσης του ελληνικού ελαιολάδου με υπέρυθρη φασματοσκοπία και εφαρμογή μοντέλων επιβλεπόμενης μάθησης
11:00-11:30	**ΔΙΑΛΕΙΜΜΑ – COFFEE BREAK**

Αίθουσα Α/ **Hall A** **11:30 – 13:30**	**ΠΡΟΣΚΕΚΛΗΜΕΝΗ ΣΥΝΕΔΡΙΑ / INVITED SESSION** **Statistical Methods and Applications in Systems Assurance & Quality Part II** **Προεδρεύων/Chair: A. Ρακιτζής/A. Rakitzis**
11:30-12:00	**C. Heuchenne** Anomaly detection for compositional data using support vector data description
12:00-12:30	**K. P. Tran** Secure and Robust Federated Learning with Explainable Artificial Intelligence for Healthcare Systems
12:30-13:00	**P. Otto** A Dynamic Spatiotemporal Stochastic Volatility Model with an Application to Environmental Risks
13:00-13:30	**T. Perdikis** Distribution-free Control Charts for Monitoring Dispersion in Finite Horizon Productions

Αίθουσα Β/ Hall B 11:30 – 13:30	**ΣΥΝΕΔΡΙΑ / CONTRIBUTED SESSION Στατιστική - Statistics Προεδρεύων/Chair : Σ. Δαφνής / S. Dafnis**
11:30-11:50	**V. Chasiotis , D. Karlis** Subdata selection for big data regression based on leverage scores
11.50-12:10	**G. Seitidis, S. Nikolopoulos, I. Ntzoufras, D. Mavridis** Inconsistency identification in network meta-analysis via stochastic search variable selection
12:10-12:30	**K. Kontouli, O. Koutsiouroumpa, Ch. Christogiannis, S. Nikolakopoulos, D. Mavridis** Single-Arm Trials In Evidence Synthesis
12:30-12:50	**F. Milienos, N. Balakrishnan, S. Pal** A multiple stage destructive cure rate model
12:50-13:10	**A. Anastasiou, J. Tsimikas** Optimal threshold selection on a two-cut point optimal ROC curve
13:10-13:30	**P. Papastamoulis** Estimating mixtures of multinomial logistic regressions

13:30 - 15:45 ΜΕΣΗΜΒΡΙΝΗ ΔΙΑΚΟΠΗ – MIDDAY BREAK

Αίθουσα Α/ Hall A 15:45 - 17:45	**ΠΡΟΣΚΕΚΛΗΜΕΝΗ ΣΥΝΕΔΡΙΑ / INVITED SESSION Stochastic Modeling and Statistical Methods: Advances and Applications Part II Προεδρεύων / Chair: Α. Καραγρηγορίου/A. Karagrigoriou**
15:45 – 16:15	**V. Bisht and S. B. Singh** A Study on Dynamic Reliability Measures of Multi-state k-out-of-n systems
16:15 – 16:45	**V. P. Koutras, S. Malefaki, P. Psomas and A. N. Platis** Modelling wind intensity effects on wind turbine maintenance policies
16:45 – 17:15	**V. S. Barbu, F. Mokhtari**, **C. Ayhar** and **S. Rahmnani** Nonparametric estimation of semi-Markov processes and applications to reliability
17:15 – 17:45	**N. Eleftheroglou** Adaptive Prognostics of Engineering Assets Utilizing Markov Models

15:45-17:45	**POSTER SESSION**
	~~**T. Tsiampalis, D. Panagiotakos**~~ CANCELLED ~~The AFFINITY method: A methodological Framework for Imputing missing spatial data at an aggregate level and guaranteeing personal data privacy; implementation in the context of the official spatial Greek census data~~
	N. Nikolaou, M. Valizadeh, S. Behzadi, J. Staab, M. Dallavalle, A. Peters, A. Schneider, H. Taubenböck, K. Wolf A machine learning framework for cardiovascular health prediction modeling the interplay between various environmental, neighborhood and socio-economic features: a German-wide application
	T. Moysiadis, D. Koparanis, K. Liapis, M. Ganopoulou, G. Vrachiolias, I. Katakis, Ch. Moyssiadis, I. Vizirianakis, L. Angelis, K. Fokianos, I. Kotsianidis A novel personalized stepwise dynamic predictive algorithm in Chronic Lymphocytic Leukemia
	M. Ganopoulou, D. Koparanis, K. Liapis, E. Lamprianidou, S. Papadakis, K. Fokianos, I. Kotsianidis, L. Angelis, T. Moysiadis Causal Structure assessment in Health-Related Quality of Life questionnaires
	K. Gourgoura, P. Rivadeneyra, E. Stanghellini, Ch. Caroni, F. Bartolucci, G. Pucci, R. Curcio, M. Cavallo, L. Sanesi, G. Morgana, S. Bartoli, R. Ferranti, M. B. Pasticci, G. Vaudo Modeling the Health Impact of COVID-19 using Mixed Interaction Models and Chain Graph Models
	Z. Kyrana, E. Pratsinakis, N. Papafilippou, A. Markos, G. Menexes Comparison of dimensionality reduction and clustering methods on the multidimensional dataset "Forest Cover Type" with mixed-type data
	Μ. Τριανταφύλλου Συγκριτική ανάλυση κόστους νοσηλείας χειρουργικού τομέα σε ένα δημόσιο νοσοκομείο την πενταετία 2016-2020: Η περίπτωση του Γενικού Νοσοκομείου Αθηνών «Ιπποκράτειο»
	Δ.-Δ. Βάρσου, Α. Τσουμάνης, Μ. Αρτεμίου, Ν. Χειμαριός, Α. Παπαδιαμάντης, Α. Αφαντίτης Isalos Analytics Platform: Ένα λογισμικό μηχανικής μάθησης για μη-προγραμματιστές

17:45-18:15	***ΔΙΑΛΕΙΜΜΑ – COFFEE BREAK***

Αίθουσα Α/ Hall A 18:15-20:15	**ΠΡΟΣΚΕΚΛΗΜΕΝΗ ΣΥΝΕΔΡΙΑ / INVITED SESSION** **Stochastic Modeling and Statistical Methods: Advances and Applications Part III Προεδρεύων/Chair: I. Τριανταφύλλου / I. Triantafyllou**
18:15 – 18:45	**E. Polyzos, L. Pyl** Stochastic modeling of the elastic properties of carbon-fiber-reinforced 3D printed filaments using polynomial chaos expansion
18:45 – 19:15	**A. Kumar, P. Kumar** Understanding a system's performance in the Presence of k-out-of-n: F and Standby redundancy: A Reliability approach through Markov process
19:15 – 19:45	**G. D'Amico, F. Gismondi, F. Petroni** On some generalization of the ROCOF for general multi-state systems with applications
19:45 – 20:15	**P. C. Trandafir, Ch. P. Kitsos** Applying LSI to Probability and Statistics

Αίθουσα Β/ Hall B 18:15-20:35	**ΣΥΝΕΔΡΙΑ / CONTRIBUTED SESSION** **Εφαρμοσμένη Στατιστική/ Applied Statistics** **Chair/Προεδρεύων: Δ. Μαυρίδης / D. Mavridis**
18:15 – 18:35	**K. Georgiou, K. Charmanas, N. Mittas, L. Angelis** Thematic analysis and research trends on medical diagnosis patents using state-of-the-art topic modelling
18:35 – 18:55	**V. Georgakis, P. Xenos** Healthcare Risk Management: Machine learning in Cardiovascular intensive care unit (CICU)
18:55-19:15	**A. Nalpantidi, D. Kallis** Multinomial mixture model for spatial data with an application to demographics
19:15-19:35	**E. Panagiotopoulos, I. Oikonomides, S. Trevezas** A Comparison of Generalized Linear Mixed-Effects Models and Random Forest Models on Crop Progress Monitoring

19:35-19:55	**L. Champezou, D. Kallis** Spatiotemporal clustering with an application on COVID-19 deaths in the provinces of the Netherlands
19:55-20:15	**D. Kouloumpou, G. Anastassiou** Brownian Motion Approximation by Neural Networks
20:15-20:35	**M. Ravani, K. Georgiou, G. Liantas, I. Chatzigeorgiou, L. Angelis, G. K. Ntinas** Carbon footprint and human toxicity study of greenhouse products from aquaponic cultivation
20:35-20:55	**M. Ravani, K. Georgiou, G. Liantas, L. Angelis, G. K. Ntinas** A comparative study of carbon footprint of sheep milk production using archetypal analysis

ΚΥΡΙΑΚΗ 28 ΜΑΪΟΥ 2023/ SUNDAY MAY 28 2023

Αίθουσα Α/ Hall A: Σ109Α

Αίθουσα Β/ Hall B: Σ109Β

Αίθουσα Α/ Hall A **09:00 - 11:00**	**ΠΡΟΣΚΕΚΛΗΜΕΝΗ ΣΥΝΕΔΡΙΑ / INVITED SESSION** **Stochastic Modeling and Statistical Methods: Advances and Applications Part IV Προεδρεύουσα / Chair: Σ. Μαλεφάκη/ S. Malefaki**
09:00 – 09:30	**O. Jones, L. Hayes, J. Cable** Fitting a detailed stochastic population model using Approximate Bayesian Computation
09:30 – 10:00	**M. Limnios, N. R. Hansen** Nonparametric Modeling of Event Processes with Applications to Conditional Local Independence Testing
10:00 – 10:30	**S. Georgiadis, D. Di Giuseppe, A. Scherer, M. Lund Hetland, K. Pavelka, J. Vencovsky, Z. Rotar, K. Perdan Pirkmajer, G. T. Jones, B. Glintborg, A. G. Loft,** **B. Gudbjornsson, O. Palsson, B. Michelsen, E. Klami Kristianslund, H. Relas,** **J. Huhtakangas, A. Ciurea, M. J. Nissen, J. K. Wallman, A. Yazici, M. Birlik,** **L. Ørnbjerg** A registry-based simulation study evaluating interchangeability of patient reported outcomes in axial spondyloarthritis
10:30 – 11:00	**A. Burnetas** Recursive Computation of Equilibrium and Optimal Strategies in Unobservable Feed-Forward Queueing Networks with Delay-Sensitive Customers.
Αίθουσα Β/ Hall B **09:00 - 11:00**	**ΣΥΝΕΔΡΙΑ / CONTRIBUTED SESSION** **Εφαρμοσμένες Πιθανότητες – Εφαρμοσμένη Στατιστική** **Προεδρεύων / Chair: Φ. Μηλένιος/ F. Milenios**
09:00 – 09:20	**Α. Ρακιτζής, Ε. Μαμζερίδου** Άνω μονόπλευρα Διαγράμματα Ελέγχου με Εκτιμημένες Παραμέτρους για Μηδενοδιογκωμένες Διεργασίες

09:20 – 09:40	**Γ. Παπαγεωργίου, Μ. Χαλικιάς** Σύγκριση εκτιμητών για την πρόβλεψη εκλογικών αποτελεσμάτων
09:40 -10:00	**Χ. Ευαγγελάρας, Β. Τραπουζανλής** Σχεδιασμοί για definitive screening με ελάχιστη συσχέτιση μεταξύ των τετραγωνικών επιδράσεων
10:00 – 10:20	**Σ. Τζανίνης** Η πιθανότητα χρεοκοπίας σε ένα γενικευμένο ανανεωτικό μοντέλο κινδύνου
10:20 – 10:40	**Π. Βλιώρα, Γ. Ψαρράκος, A. Toomaj** Μια οικογένεια μέτρων μεταβλητότητας που βασίζεται στην αθροιστική υπολειπόμενη εντροπία και σε στρεβλές συναρτήσεις
10:40 - 11:00	**Σ. Δαφνής, Μ. Β. Κούτρας** Μελέτη των ιδιοτήτων μιας ευρείας οικογένειας συνεχών κατανομών

11:00 - 11:30	***ΔΙΑΛΕΙΜΜΑ – COFFEE BREAK***

Αίθουσα Α/ **Hall A** **11:30 - 12:00**	**ΚΕΝΤΡΙΚΗ ΟΜΙΛΙΑ / PLENARY TALK** **e-Health** **Προεδρεύουσα/Chair : Γ. Τουλούμη / G. Touloumi**
11:30 – 12:00	**S. Denaxas** Methods for using electronic health records to study human diseases

Αίθουσα Α/ **Hall A** **12:10-13:30**	**ΕΙΔΙΚΗ ΣΥΝΕΔΡΙΑ / SPECIAL SESSION** **Το Βήμα των Υποψηφίων Διδακτόρων** **Προεδρεύων /Chair: Α. Μπουρνέτας, Γ. Τουλούμη**
12:10 – 12:30	**Σ. Ζαφειράτου, M. Stafoggia, Ε. Σαμόλη, Α. Αναλυτής, Χ. Γιαννακόπουλος, Κ. Β. Βαρώτσος, A. Schneider, Κ. Κατσουγιάννη** Επιδράσεις υψηλών θερμοκρασιών στη θνησιμότητα στο νομό Αττικής, εφαρμόζοντας μη γραμμικό μοντέλο κατανεμημένης χρονικής υστέρησης στο πλαίσιο του προγράμματος EXHAUSTION
12:30 – 12:50	**Σ. Ρούσσος, Β. Σύψα** Μεθοδολογία εκτίμησης του μεγέθους δύσκολα προσεγγίσιμων πληθυσμών με εφαρμογή στον πληθυσμό των χρηστών ενέσιμων ναρκωτικών στην Αθήνα

12:50 – 13:10	**Β. Μπαραλού, Χ. Θωμαδάκης, Ν. Δεμίρης, Γ. Τουλούμη** Στατιστικές μέθοδοι ανίχνευσης σε πραγματικό χρόνο της έναρξης και της εξέλιξης HIV επιδημικών εκρήξεων σε άτομα που κάνουν χρήση ενδοφλέβιων ναρκωτικών
13:10 – 13:30	**Ν. Καλπουρτζή, Γ. Τουλούμη** Προβλέψεις αριθμού θανάτων από καρδιαγγειακά νοσήματα μέχρι το 2035 στην Ελλάδα, χρησιμοποιώντας μοντέλο ατομικών προσομοιώσεων σε περιπτώσεις με έλλειψη εθνικών δεδομένων κοορτής

Αίθουσα Β/ Hall B **12:10-13:50**	**ΕΙΔΙΚΗ ΣΥΝΕΔΡΙΑ / SPECIAL SESSION** **Techniques & Applications of Data Analytics & Machine Learning II** **Προεδρεύων / Chair : Π. Μπομποτάς / P. Bobotas**
12:10 – 12:30	**A. Karaminas, P. Economou, S. Bersimis** Improving healthcare services using machine learning techniques
12:30 – 12:50	**F. Bersimis, P. Bagos** Polygenic risk scores and applications in clinical practice
12:50 – 13:10	**E. Nika, T. Tsiampalis, D. Georgakellos** Health budgets using data analytical techniques
13:10 – 13:30	**Ch. Boudoulis, O. Antonopoulou, A. Fousteris, P. Economou and S. Bersimis** The use of data analytics in Insurance: an overview
13:30 - 13:50	**Ch. Spyropoulos, P. Economou, S. Bersimis** The value of data exploitation in sports industry: an overview

13:50 - 16:00	***ΜΕΣΗΜΒΡΙΝΗ ΔΙΑΚΟΠΗ – MIDDAY BREAK***

Αίθουσα Α/ Hall A **16:00 - 18:00**	**ΠΡΟΣΚΕΚΛΗΜΕΝΗ ΣΥΝΕΔΡΙΑ / INVITED SESSION** **Missing data in longitudinal health studies: Methods to analyze cohort studies and surveillance data.** **Προεδρεύουσα/Chair : Γ. Τουλούμη / G. Touloumi**

16:00-16:30	**N. Pantazis** Missing data and reporting delay in surveillance data
16:30-17:00	**G. Touloumi, Ch. Thomadakis, L. Meligkotsidou, N. Pantazis** Longitudinal and Time-to-Drop-out Joint Models under at Random Drop-out Mechanism
17:00-17:30	**Ch. Thomadakis, N. Pantazis and G. Touloumi** Issues with the expected information matrix of linear mixed models provided by popular statistical packages under MAR dropout
17:30-18:00	**G. Bakoyannis** Semiparametric regression for competing risks data with missing not at random cause of failure

Αίθουσα Β/ Hall B **16:00-18:00**	**ΣΥΝΕΔΡΙΑ / CONTRIBUTED SESSION** **Στοχαστικές και Στατιστική** **Προεδρεύων / Chair: Π. Παπασταμούλης / P. Papastamoulis**
16:00 – 16:20	**Γ. Ψαρράκος** Στοχαστική διάταξη και ταυτότητες συνδιακύμανσης
16:20 – 16:40	~~**Ρ. Αύκου, Γ. Βασιλειάδης, Γ. Τσακλίδης**~~ **CANCELLED** Φίλτρο Κλειστού Κρυφού Ομογενούς Μαρκοβιανου Συστήματος Με ~~Πεπερασμένες Χωρητικότητες Στις Καταστάσεις~~
16:40 – 17:00	**Α. Μπατσίδης, M. D. Jiménez Gamero και B. Milošević** Έλεγχος καλής προσαρμογής για τη γενικευμένη Poisson κατανομή
17:00 – 17:20	**Π. Οικονόμου** Ένας νέος αλγόριθμος συσταδοποίησης παρουσία επικαλυπτόμενων Γκαουσιανών υποπληθυσμών
17:20 - 17:40	**Π. Μπομποτάς, Σ. Κουρούκλης** Βελτιωμένη εκτίμηση παραμέτρων σε inverse Gaussian κατανομή
17:40 - 18:00	**Δ. Πανάρετος, Α. Σαχλάς, Γ. Τζαβελλάς, Μ. Βαμβακάρη, Δ. Παναγιωτάκος** Αξιολόγηση της ακρίβειας των παραγόντων στην διερευνητική παραγοντική ανάλυση: ορισμός προβλήματος
18:00 - 18:30	***ΔΙΑΛΕΙΜΜΑ - COFFEE BREAK***

Αίθουσα Α/ Hall A: Κεντρικό αμφιθέατρο/ Central Amphitheater	
Αίθουσα Α/ **Hall A** **18:30-20:30**	**ΠΡΟΣΚΕΚΛΗΜΕΝΗ ΣΥΝΕΔΡΙΑ / INVITED SESSION** **Statistical Methods and Applications in Systems Assurance & Quality Part III Προεδρεύων/Chair: Σ. Μπερσίμης / S. Bersimis**
18:30 – 19:00	**J. M. Poggi** Statistics and Machine Learning in industry: combining heterogeneous or multi-scale model outputs
19:00 – 19:30	**C. Ley** Sport Statistics – When Figures are more than Numbers
19:30 – 20:00	**M. Bourguinon** A parametric quantile beta regression for modeling case fatality rates of COVID-19
20:00 - 20:30	**D. Jeske** Design and Inference for a RCT when Treatment Observations Follow a Two-Component Mixture Model
Αίθουσα Α/ **Hall A** **20:30-21:00**	**ΛΗΞΗ ΣΥΝΕΔΡΙΟΥ – CONFERENCE CLOSING** **Προεδρεύων/Chair: Ε. Αγγελής/ E. Angelis**

Linear Prediction

N. Balakrishnan

McMaster University, Hamilton, Ontario, Canada

In this talk, I will discuss the problem of optimal linear prediction of censored order statistics. I will provide its connection to case deletion diagnostics in general linear model. Next, I will discuss the problem of joint linear prediction and present some optimality results under both unbiasedness and biasedness conditions of the predictors. Throughout, I will present some examples to illustrate the theoretical results established.

Nonparametric estimation of semi-Markov processes and applications to reliability

Vlad Stefan Barbu*

LMRS, University of Rouen-Normandy, France & Centre for Demographic Research "Vladimir Trebici", "Costin C. Kiritescu" National Institute of Economic Research of Romanian Academy, Romania
vladstefan.barbu@univ-rouen.fr

Our presentation is dedicated to kernel estimation of the main characteristics of a continuous-time semi-Markov process, like conditional and unconditional sojourn times in a state, semi-Markov kernel, etc. Our main objective is to establish asymptotic properties of the semi-Markov kernel estimators and of the sojourn time distribution estimators (conditional and unconditional), as well as of the estimators of the associated Radon-Nikodym derivatives, when the sample size becomes large. Corresponding reliability indicators are also estimated. The approach is illustrated by considering a numerical example.

REFERENCES

C. Ayhar, V. S. Barbu, F. Mokhtari, S. Rahmani, On the asymptotic properties of some kernel estimators for continuous semi-Markov processes, Journal of Nonparametric Statistics, 34(2), 299-318, 2022

*This is a joint work with **Fatiha MOKHTARI**, **Chafiâa AYHAR** and **Saâdia RAHMANI** (LSMSA, University of Saida–Doctor Moulay Taher, Algeria).

An infinite class of exponential dispersion models for count data - a survey of current research

Shaul K. Bar-Lev[1] and Ad Ridder[2]

[1]Faculty of Industrial Engineering and Technology Management, Holon Institute of Technology, Holon, Israel

[2]School of Business and Economics, VU University, Amsterdam, Netherlands

shaulb@hit.ac.il

Modeling count data by parametric families of counting distributions were initially based on the Poisson and the negative binomial distributions. Later on, many other count models were developed in four paths: a) Generalizing existing models (as generalized Poisson or negative binomial distributions); b) Mixing distributions (as Poisson-inverse Gaussian); c) Discretization of continuous distributions (as discrete Gamma distribution); d) Adding extra probability mass on zero (as zero-inflated Poisson). All these models share one thing in common- their probability functions can be expressed explicitly in a closed form - thus making them computable. All such parametric models generate, however, a limited set. In contrast, we introduce an infinite set of new classes of counting natural exponential families (NEFs) and their associated exponential dispersion models (EDMs), whose probability distributions cannot be expressed explicitly. Still, they can be calculated numerically by the availability of existing powerful computing software. These classes are characterized by their respective variance functions (VFs). Particularly, we show that any function of the type $f(m) = 1 + \sum_{n=0}^{\infty} a_n m^n$, $a_n \geq 0$, with a positive radius of convergence, can be used to generate a VF of an NEF of the form $V(m) = m\, f(m)$. The set of such VFs (or NEFs/EDMs) is, indeed, infinite as any function of such a form can be used to construct a VF and thus an NEF or EDM. Thus, this set includes an enormous number of explicit classes of NEFs/EDMs supported on the nonnegative integers. The vast majority of the constructed classes have not been used before - simply, as they have not been known before. Accordingly, these new classes significantly enrich the set of probability models for modeling count data as well as other applications (as for insurance risk aspects). Furthermore, we show that the distributions of all these classes are infinitely divisible, a property equivalent to that the dispersion parameter of the corresponding EDM ranges over the whole positive line. Also, we show that these classes are skewed to the right, leptokurtic, over-dispersed, and zero-inflated (relative to the Poisson NEF). All of the above results appear in five papers - three already published and two are forthcoming. This lecture is an overview of the five papers.

A Study on Dynamic Reliability Measures of Multi-state *k*-out-of-*n* systems

Vaibhav Bisht[1] and S. B. Singh[2]

Department of Mathematics, Statistics and Computer Science,
G. B. Pant University of Agriculture and Technology, Pantnagar, India
[1]bishtvickybng@gmail.com, [2]drsurajbsingh@yahoo.com

Reliability analysis of complex systems is an important area of research in engineering and applied mathematics. Multi-state *k*-out-of-*n* systems are widely used in practice to model the reliability of systems with redundant components. The deterioration and restoration of an individual component have an impact on the dynamic behavior of these systems. Traditional techniques for evaluating the reliability of these systems are based on the assumption that the states of the system are always constant, which may not be appropriate in realistic situations. To tackle this issue, a method based on the L_z-transform is presented to examine the dynamic behavior of multi-state *k*-out-of-*n* systems. The L_z-transform is a powerful tool for analyzing stochastic processes with time-varying rates. The dynamic reliability indices like reliability, availability, mean expected performance, and mean performance deficiency have been computed here by applying this transform to the multi-state *k*-out-of-*n* system. Numerical simulations on a multi-state *k*-out-of-*n* system with different states and *n* components are used to find the efficacy of the suggested method. The outcomes demonstrate that obtaining the dynamic reliability indices by the L_z-transform method is easier than those produced by conventional techniques. The suggested method offers a versatile and effective tool for evaluating the reliability indices of multi-state *k*-out-of-*n* systems in real-world engineering applications.

Keywords: L_z-transform, *k*-out-of-*n* system, Failure and Repair Rate, Markov process.

A parametric quantile beta regression for modeling case fatality rates of COVID-19

Marcelo Bourguignon, Diego I. Gallardo and Helton Saulo

Department of Statistics, Universidade Federal do Rio Grande do Norte, Natal, Brazil
Department of Mathematics, Facultad de Ingenieria, Universidad de Atacama, Copiapó, Chile
Department of Statistics, University of Brasília, Brasília, Brazil
m.p.bourguignon@gmail.com

Motivated by the case fatality rate (CFR) of COVID-19, in this paper, we develop a fully parametric quantile regression model based on the generalized three-parameter beta (GB3) distribution. Beta regression models are primarily used to model rates and proportions. However, these models are usually specified in terms of a conditional mean. Therefore, they may be inadequate if the observed response variable follows an asymmetrical distribution, such as CFR data. In addition, beta regression models do not consider the effect of the covariates across the spectrum of the dependent variable, which is possible through the conditional quantile approach. In order to introduce the proposed GB3 regression model, we first reparameterize the GB3 distribution by inserting a quantile parameter and then we develop the new proposed quantile model. We also propose a simple interpretation of the predictor-response relationship in terms of percentage increases/decreases of the quantile. A Monte Carlo study is carried out for evaluating the performance of the maximum likelihood estimates and the choice of the link functions. Finally, a real COVID-19 dataset from Chile is analyzed and discussed to illustrate the proposed approach.

A class of power series *q*-distributions

Charalambos A. Charalambides
Department of Mathematics, University of Athens
ccharal@math.uoa.gr

A class of power series *q*-distributions, generated by considering a *q*-Taylor expansion of a parametric function into powers of the parameter, is discussed. Its *q*-factorial moments are obtained in terms of *q*-derivatives of its series (parametric) function. Also, it is shown that the convolution of power series *q*-distributions is also a power series *q*-distribution. Furthermore, the *q*-Poisson (Heine and Euler), *q*-binomial of the first kind, negative *q*-binomial of the second kind, and *q*-logarithmic distributions are shown to be members of this class of distributions and their *q*-factorial moments are deduced. In addition, convolution properties of these distributions are examined.

REFERENCES

Benkherouf, L. and Bather, J. A. (1988) Oil exploration: sequential decisions in the face of uncertainty. *J. Appl. Probab*., **25**, 529-543.

Charalambides, Ch. A. (2010) The *q*-Bernstein basis as a *q*-binomial distribution, *J. Statist. Plann. Inference*, **140**, 2184-2190.

Charalambides, Ch. A. (2016) *Discrete q-Distributions*, John Wiley & Sons, Hoboken, New Jersey.

Charalambides, Ch. A and Papadatos, N. (2005) The *q*-factorial moments of discrete *q*-distributions and a characterization of the Euler distribution. In N. Balakrishnan, I. G. Bairamov, and G. L. Gebizlioglu, Eds, *Advances on Models, Characterization and Applications*, Chapman & Hall/CRC, Boca Raton, Fl, pp. 57-71.

Dunkl, C. F. (1981) The absorption distribution and the *q*-binomial theorem. *Commun. Statist. Theory Methods*, **10**, 1915-1920.

Ernst, T. (2012) *A Comprehensive Treatment of q-Caluclus*, Birkhauser, Springer, Basel.

Jackson, F. H. (1909) A *q*-form of Taylor's theorem, *Messenger of Math*., **38**, 62-64.

Jackson, F. H. (1942) On basic double hypergeometric functions, *Quarterly Journal of Math*., **13**, 69-82.

Kemp, A. (1992) Heine and Euler extensions of the Poisson distribution, *Commun. Statist. Theory Methods*, **21**, 791-798.

Kemp, A. and Kemp, C. D. (1991) Weldon's dice data revisited, *Amer. Statist*., **45**, 216-222.

Kemp, A. and Newton, J. (1990) Certain state-dependent processes for dichotomized parasite population. *J. Appl. Probab.*, **27**, 251-258.

Kemp, C. D. (1997) A *q*-logarithmic distribution. In Balakrishnan, N. (Ed.) Advances in *Combinatorial Methods and Applications to Probability and Statistics*, Birkhauser, Boston, MA, pp. 465-470.

Khatri, C. G. (1959) On certain properties of power-series distributions, *Biometrika*, **46**, 486-490.

Kyriakoussis, A. and Vamvakari, M. (2013) A *q*-analogue of the Stirling formula and a continuous limiting behaviour of the *q*-binomial distribution-Numerical calculations, *Methodol. Comput. Appl. Probab*., **15**, 187-213.

Noack, A. (1950) A class of random variables with discrete distributions, *Ann. Math. Statist.*, **21**, 127-132.

Patil, G. P. (1962) Certain properties of the generalized power series distribution, *Ann. Inst. Statist. Math*., **14**, 179-182.

Multivariate regularly varying insurance and financial risks in multi-dimensional risk models

Ming Cheng, Dimitrios G. Konstantinides, Dingcheng Wang

School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, China

Department of Statistics and Actuarial - Financial Mathematics, University of the Aegean, Karlovassi, Greece

School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, China

wangdc@uestc.edu.cn

Multivariate regular variation is a key concept that has found its use in insurance, finance, and risk management. This paper proposes a new assumption (see Assumption 3.1) via a framework of multivariate regular variation. Assumption 3.1 includes a wide type of dependence structures, such as the Asimit-Jones Dependence. Under the condition that insurance and financial risks follow Assumption 3.1, asymptotic analyses for multi-dimensional risk models are studied in the discrete-time and continuous-time cases, respectively. Also, a 2-dimensional numerical example satisfying Assumption 3.1 is presented, through which we show the accuracy of the asymptotic result for the discrete-time multi-dimensional insurance risk model.

REFERENCES

Chen, Y., 2017. Interplay of subexponential and dependent insurance and financial risks. Insurance Math. Econom. 77: 78-83

Chen, Y., Wang J., Zhang W., 2021. Tail distortion risk measure for portfolio with multivariate regularly variation. Commun. Math. Stat.

Ji, L., 2020. On the cumulative Parisian ruin of multi-dimensional Brownian motion risk models. Scand. Actuar. J. 9: 819-842.

Tang, Q., Yang, Y., 2019. Interplay of insurance and financial risks in a stochastic environment. Scand. Actuar. J. (5): 432-451.

A Moment-type Estimator for the Generalized Gamma Model with Complete and Interval Censored Data – Application to First Division Time of HSC

P.-H. Cournède[1], G. Hermange[1], P. Karkalakis[2], S. Trevezas[2]

[1]CentraleSupélec, Paris, France

[2]National and Kapodistrian University of Athens

pascal@math.uoa.gr

In this talk we revisit parameter estimation for the Generalized Gamma model, which was introduced by Stacy (1962) and includes as special cases many commonly used models, such as the Exponential, Gamma, and Weibull. By a suitable reparameterization, Prentice (1974) also incorporated the Lognormal model. As is common, the Maximum Likelihood Estimator (MLE) is not available in closed form and numerical algorithms are necessary to approximate the MLE, often needing multiple initializations. Here we present a novel moment-type estimator that can serve as a proxy to the MLE and we investigate its asymptotic properties. We also discuss parameter estimation in the case of interval censored data (with a fixed censoring mechanism), using a simple iterative imputation algorithm. An application is given with real data, which are interval censored due to the specific experimental protocol. These observations describe the time needed for the first division of Hematopoietic Stem Cells (HSC), which is of primary importance when studying blood formation.

On some generalization of the ROCOF for general multi-state systems with applications

Guglielmo D'Amico [1,*], Fulvio Gismondi [2], Filippo Petroni [3]

[1]Department of Economics, University "G. d'Annunzio" of Chieti-Pescara, Pescara, Italy; g.damico@unich.it
[2]Department of Economic and Business Science, "Guglielmo Marconi" University, Italy; f.gismondi@unimarconi.it
[3]Department of Management, Marche Polytechnic University, Ancona, Italy; f.petroni@univpm.it
*****Correspondence: g.damico@unich.it

The rate of occurrence of failures (ROCOF) is an important indicator in the reliability analysis of complex random systems. In general, it gives information on whether there are a lot of failures or only a few within a given time frame. The ROCOF was studied for Markov processes (MP) in [3] and [5]. Extension to semi-Markov processes (SMP) was advanced in [4], where the authors derived a formula for the ROCOF and a statistical estimator with an investigation of its asymptotic properties. Recently, a generalization of the ROCOF, named the ROCOF of higher order, was calculated for MP in [1] and successively extended by [2] to a SMP framework with mixed probability distribution for the initial law of the system, taking into account the possible random starting from any state of the system with any duration. In this study, we provide further generalizations of the ROCOF of higher order motivated by practical applications, especially in mathematical finance. The results include previous investigations as particular cases and demonstrate the usefulness of the indicator not only for reliability purposes but for more general scopes beyond the reliability framework.

Keywords: semi-Markov processes; reliability; high-frequency finance.

REFERENCES

[1] D'Amico G. (2015) Rate of Occurrence of Failures (ROCOF) of Higher-Order for Markov Processes: Analysis, Inference and Application to Financial Credit Ratings. *Methodol. Comput. Appl. Probab.* 17, 929-949.

{2] D'Amico, G. and F. Petroni. "ROCOF of higher order for semi-Markov processes." *Applied Mathematics and Computation* 441 (2023): 127719.

{3] Yeh, L. (1997) The Rate of Occurrence of Failures. *J. Appl. Probab.* 34(1), 234-247.

[4] Ouhbi B. and Limnios N. (2002) The rate of occurrence of failures for semi-Markov processes and estimation. *Stat. Probab. Lett.* 59, 245-255.

[5] Shi D. (1985) A new method for calculating the mean failure numbers of a repairable system during (0; t]. *Acta Math. Appl. Sin.* 8(1), 101-110.

Regression Analysis of a Future State Entry Time Distribution Conditional on a Past State Occupation

Somnath Datta

Department of Biostatistics, University of Florida

somnath.datta@ufl.edu

We present a nonparametric method for estimating the conditional future state entry probabilities and distributions of state entry time conditional on a past state visit when data are subject to dependent censorings in a progressive multistate model where Markovity of the system is not assumed. These estimators are constructed using the competing risk techniques with risk sets consisting of fractional observations and inverse probability of censoring weights. The fractional observations correspond to estimates of the numbers of persons who ultimately enter a state from which the future state in question can be reached in one step. We then address the corresponding regression problem by combining these marginal estimators with the pseudo-value approach. Performance of our regression scheme is studied using a comprehensive simulation study. Time permitting, we will discuss extensions of these methodology to the case of current status multistate data.

REFERENCES

Anyaso-Samuel, S. and Datta, S. Adjusting for informative cluster size in pseudo-value-based regression approaches with clustered time to event data, Statistics in Medicine (2023), in press.

Datta, S. and Sundaram, R. Nonparametric marginal estimation in a multistage model using current status data, Biometrics, 62, 829-837 (2006).

Adaptive Prognostics of Engineering Assets Utilizing Markov Models

Nick Eleftheroglou

Aerospace Structures & Materials Department, Aerospace Engineering Faculty, Delft University of Technology, 2629 HS Delft, the Netherlands n.eleftheroglou@tudelft.nl

Prognostic methodologies have found increasing use the last decade and provide a platform for remaining useful life (RUL) predictions of engineering systems utilizing condition monitoring data. Of particular interest is the reliable RUL prediction of engineering assets that either underperform or outperform due to unexpected phenomena that might occur during the operational life. These assets are often referred as outliers and the prediction of their RUL is a challenging task. The challenge is to accurately predict the RUL of an outlier without taking into account outlier's condition monitoring data in the training process but just in the testing process. As a result, the lifetime of the testing asset is shorter (left outlier) or longer (right outlier) than the training process' lifetimes. This study addresses this challenge by proposing a new prognostic model; the Adaptive Non-Homogenous Hidden Semi Markov Model (ANHHSMM), which is an extension of the Non-Homogenous Hidden Semi Markov Model (NHHSMM). The ANHHSMM uses diagnostic measures that are predicted based on the training and testing condition monitoring data and it adapts the trained parameters of the NHHSMM. To demonstrate the effectiveness of the proposed adaptive methodology, composite structures have been used as a validation engineering asset. In particular, the training data set consists of strain data collected from open-hole carbon–epoxy specimens, which were subjected to fatigue loading only, while the testing data set consists of strain data collected from specimens that were subjected to fatigue and in-situ impact loading, which can be considered as an unexpected phenomenon and unseen event regarding the training process. Utilizing the aforementioned strain data the ANHHSMM RUL predictions and the NHHSMM RUL predictions were compared, so as to verify that the ANHHSMM provides better prognostics than the NHHSMM. ANHHSMM provides better predictions in comparison to the NHHSMM for all the test cases, demonstrating its capability to adapt to unexpected phenomena and integrate unforeseen data to the prognostics course.

Keywords: remaining useful life, adaptive prognostics, Markov models, outlier analysis

Applying Rényi's pseudodistance for robust Model Selection for independent not identically distributed observations

Angel Felipe, María Jaenada, Pedro Miranda, Leandro Pardo

Dept. of Statistics and O. R,

Complutense University of Madrid (Spain)

pmiranda@ucm.es

We propose a new model selection criterion based on Rényi's pseudodistance (RP) for the case of independent and not identically distributed observations. Model selection tries to find the best model considering at the same time complexity and goodness of fit. This new criterion recovers as a particular case the classical Akaike's criterion. The choice of RP responds to the goal of obtaining a procedure aimed to be robust to outliers. In this sense, RP has been applied to other domains leading to robust estimators and test statistics. RP depends on a tuning parameter cotrolling the trade-off between efficiency and robustness. We show that this new criterion is unbiased. Next, we develop the case of comparing a model to a restricted model, establishing the asymptotic distribution of the restricted estimator. We also derive a formula to decide if the restricted model better fits the data and obtain its asymptotic distribution. A simulation study shows that this new criterion seems to have a good behavior in the presence of outliers. To better explain the results, we use the multiple linear regression model as a common thread, obtaining explicit expressions of the model selection criterion and the case of restricted models.

REFERENCES

Jones. M.C., Hjort, N.L., Harris, I.R. and Basu, A. (2001). A comparison of related density-based minimum divergence estimators. Biometrika, 88, 865-873.

A. Toma and S. Leoni-Auban (2010). Robust tests based on dual divergence estimators and saddle points approximation. Journal of Multivariate Analysis, 101 1143-1155.

Felipe, A., Jaenada, M., Miranda, P. and Pardo, L. Model Selection for independent not identically distributed observations based on Rényi's pseudodistances. Arxiv.

A registry-based simulation study evaluating interchangeability of patient reported outcomes in axial spondyloarthritis

Stylianos Georgiadis, Daniela Di Giuseppe, Almut Scherer, Merete Lund Hetland, Karel Pavelka, Jiri Vencovsky, Ziga Rotar, Katja Perdan Pirkmajer, Gareth T. Jones, Bente Glintborg, Anne Gitte Loft, Bjorn Gudbjornsson, Olafur Palsson, Brigitte Michelsen, Eirik Klami Kristianslund, Heikki Relas, Johanna Huhtakangas, Adrian Ciurea, Michael J. Nissen, Johan K. Wallman, Ayten Yazici, Merih Birlik, Lykke Ørnbjerg

EuroSpA Research Collaboration, on behalf of ATTRA (Czech Republic), Biorx.si (Slovenia), BSRBR-AS (United Kingdom), DANBIO (Denmark), ICEBIO (Iceland), NOR-DMARD (Norway), ROB-FIN (Finland), SCQM (Switzerland), SRQ (Sweden), TURKBIO (Turkey)

stylianos.georgiadis@regionh.dk

Patient reported outcomes (PROs) are key measures in the clinical assessment of patients with axial spondyloarthritis (axSpA). Both single- and multiple-item PROs have been developed. Missing data are a challenge in observational studies and particularly when data are registered in clinical practice. An unexplored potential option to deal with missing data in PROs is to apply a patient-level single imputation approach, where PROs that express similar contents or associated PROs are interchangeable. We define the interchangeability between two variables as the substitution of one variable whose value is missing at a particular timepoint with the second variable whose value is available at the same timepoint (and vice versa). In the current study, we evaluate the potential interchangeability of PROs by implementing a simulation study on anonymized prospectively collected data of patients registered with axSpA in 10 European registries participating in the European Spondyloarthritis (EuroSpA) Research Collaboration Network. The performance of the imputation method is assessed in terms of coverage and bias. Sample size, level of missingness and missing data pattern are included as parameters in the simulations.

REFERENCES

Morris, T.P., White, I.R. and Crowther, M.J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, 38(11), 2074-2102.

Design and Inference for a RCT when Treatment Observations Follow a Two-Component Mixture Model

Daniel R. Jeske

University of California, Riverside

daniel.jeske@ucr.edu

A mixture of a distribution of responses from untreated patients and a shift of that distribution is a useful model for the responses from a group of treated patients. The mixture model accounts for the fact that not all the patients in the treated group will respond to the treatment and their responses follow the same distribution as the responses from untreated patients. The treatment effect in this context consists of both the fraction of the treated patients that are responders and the magnitude of the shift in the distribution for the responders. In this talk, we investigate the design and analysis of a RCT that uses a two-component mixture model for the observations in the treatment group.

This is joint work with colleagues at the University of Calfiornia, Bradley Lubich (PhD student) and Weixin Yao (Professor).

Assessing the Dynamic of Visual-Motor Skill in Climbing via Drifting Markov Modeling

Owen Jones, Laura Hayes, Jo Cable

Cardiff University

In many practical situations of interest, we can construct stochastic models that concisely reflect our understanding of the system's dynamics, but which are none-the-less too complicated to investigate analytically. Approximate Bayesian Computation (ABC) can provide a route for estimation and inference in such situations as it only requires the ability to simulate from the model as opposed to being able to calculate the likelihood. Even so successful application of ABC requires adaptation to the problem at hand-in particular regards how we measure the distance between real and simulated data-and we illustrate this with a stochastic population model. The CRIPES lab (Cardiff Research into Infection and Parasites in Ecological Systems) has been maintaining populations of different species of the *Gyrodactylus* parasite for many years. The populations are closely managed so that they do not go extinct nor place an inhumane burden on their fish hosts. They are also periodically exploited to provide parasites for various experiments. In this work we use ABC to estimate the virulence of each species, using lab records of parasite population numbers and the deaths of host fish. Our model includes a novel mechanism that captures population management interventions.

Fitting a detailed stochastic population model using Approximate Bayesian Computation

Emmanouil-Nektarios Kalligeris[1], Vlad Stefan Barbu[2], Guillaume Hacques[1], Ludovic Seifert[1] and Nicolas Vergne[2]

[1]Center for the Study of Transformations in Physical and Sporting Activities, University of Rouen Normandy, Boulevard Siegfried, 76821 Mont-Saint-Aignan, Rouen,France
[2]Laboratory of Mathematics Raphaël Salem, University of Rouen Normandy, Avenue de l'Université, BP.12, 76801 Saint Étienne du Rouvray, Rouen, France

The technological advances on climbing tracking equipment, have provided new opportunities in several areas such as the one of visual-motor skill i.e., the capacity to detect relevant optical information from the environment and to coordinate movements in order to achieve an outcome. Drifting Markov models constitute a powerful tool for modeling the heterogeneities of sequences in a more flexible fashion as compared to the homogeneous Markov chains or hidden Markov models. In this work by mixing 1) an eye-tracking system to collect gaze behavior and 2) instrumented holds to collect the hand contact time on holds, we seek to identify the dynamic of skill acquisition by capturing the behavior of visual-motor skill through Drifting Markov modeling. To that end, a real case is conducted based on 10 climbing sessions of 11 individuals.

REFERENCES

Barbu, V.S., Vergne N., Lothodé C. and Seiller, A. (2021). drimmR: An R Package for Estimation, Simulation and Reliability of Drifting Markov Models, *CRAN*.

Nguyen T.N., Seifert, S., Hacques, G., Hammami, M. and Chahir, Y. (2023). Vision-based Global Localization of Points of Gaze in Sport Climbing, *International Journal of Pattern Recognition and Artificial Intelligence*.

On Stochastic Dynamic Modeling of Incidence Data

E.-N. Kalligeris[1], A. Karagrigoriou[1] and C. Parpoula[3]

[1]Lab of Mathematics Raphaël Salem, University of Rouen Normandy, France

[2]Lab of Statistics and Data Analysis, University of the Aegean

[3]Department of Psychology, Panteion University of Social and Political Sciences

emmanouil.kalligeris@univ-rouen.fr

In this paper, a Markov Regime Switching Model of Conditional Mean with covariates, is proposed and investigated for the analysis of incidence rate data. The components of the model are selected by both penalized likelihood techniques in conjunction with the Expectation Maximization algorithm, with the goal of achieving a high level of robustness regarding the modeling of dynamic behaviors of epidemiological data. In addition to statistical inference, Changepoint Detection Analysis is performed for the selection of the number of regimes, which reduces the complexity associated with Likelihood Ratio Tests. Within this framework, a three-phase procedure for modeling incidence data is proposed and tested via real and simulated data.

Applying LSI to Probability and Statistics

Christos P. Kitsos[1], Paula Camelia Trandafir[2]

[1]University of West Attica

[2]Public University of Navarre

*

In this paper we are referring briefly to the Logarithm Sobolev Inequalities (LSI). Due to LSI, a new probability distribution was emerged, the Generalized Normal, the classical Multivariate Normal with an extra "shape" parameter, let us say γ (gamma). The main characteristics of the Generalized Normal are discussed and simulation studies provide the corresponding graphs with different values of γ, while for γ=2 the typical Multivariate Normal is approached. Under the Generalized Normal the Gramer – Rao inequality is extended, and a new Generalized G-R bound (GG-RB) can be obtained, among othe Probabilistic implementations. The discussed distribution is a Kotz type distribution. As far as the Statistical Information Theory concerns, Entropy measures are evaluated for the Generalized Normal as well as the Kullback – Leibler (K-L) divergence (or relative entropy) are investigated. More applications and extensions are also discussed and investigated.

Some Stylized Facts of the Conditional Expected Delay (CED)

Sven Knoth

Department of Mathematics and Statistics, Helmut Schmidt University Hamburg, Germany

knoth@hsu-hh.de

The popular zero-state average run length (ARL) is just the mean of the random run length, which is the core element of a control chart. However, more appropriate measures for evaluating the detection power make use of the conditional expected delay (CED), which is the mean of the detection delay for a given change point position $\tau = 1, 2, \ldots$ under the condition that no false alarm was triggered. Originally, it was only a prerequisite for building sophisticated delay measures. Aiming to optimize the latter, interesting patterns of the CED series were found. While the more theoretical strand of monitoring research struggles with some open optimality problems, did the more applied one not recognize the CED stylized facts so far. This contribution tries to close the gap. More importantly, it emphasizes and affirms the importance of an appropriate CED analysis.

Modelling wind intensity effects on wind turbine maintenance policies

Vasilis P. Koutras[1], Sonia Malefaki[2], Panagiotis M. Psomas[1] and Agapios N. Platis[1]

[1]University of the Aegean; v.koutras@aegean.gr; ppsomas@aegean.gr; platis@aegean.gr

[2]University of Patras; smalefaki@upatras.gr

Wind turbines are one of the most important electricity production systems, and a lot of research effort is paid on enhancing their performance and availability. To this direction, turbines' maintenance policies should be carefully designed to reduce the probability of failures that interrupt electricity production. However, designing turbines' maintenance is not an easy task since it can be carried out only under specific weather conditions. Beyond this, turbines themselves can experience internal damages that result in performance deterioration or even failures. This work provides an indicative model that incorporates wind turbines output performance deterioration due to internal damages such as fatigue or wear-out of turbine's parts as well as due to external weather conditions. Note that low wind intensity does not allow the turbine to operate but it is considered as an occasion to carry out opportunistic maintenance (OM) actions. Considering that wind speed can randomly change, the time that wind speed remains at certain levels can be modeled by exponential distributions, though the holding times at any of the wind turbines internal condition can be modelled by general distributions. A semi-Markov process is used to model the evolution of the entire system in time and the asymptotic probability distribution, and the availability are evaluated. To investigate the benefits of OM, the availability is evaluated and compared on systems with and without turbine's OM. Although such a model seems to be satisfying, it suffers from a modelling issue. When the state of the system changes due to wind speed changes, the system enters a new state where the internal turbine's condition remains the same. But under the semi-Markov model, the holding time at this certain internal condition state is regenerated, a fact that from a modelling point of view is not correct. To counteract this, a simulation is carried out using the proper distributions for the holding times at each internal condition state to estimate systems' asymptotic probability distribution and availability.

Understanding a system's performance in the Presence of *k*-out-of-*n: F* and Standby redundancy: A Reliability approach through Markov process

[1]Amit Kumar and [2]Pardeep Kumar

[1]Symbiosis Institute of Technology,

Symbiosis International (Deemed University) (SIU) Lavale, Pune, Maharashtra, India.

[2]Department of Mathematics,

Lovely Professional University, Punjab, India

E-mail-amit303singh@gmail.com

It is a well-established fact that redundancy techniques are widely used to increase the performance of many structures at the time of their design. In the present paper author tried to investigate a system in which two types of redundancy namely warm standby and k-out-of-n redundancies are used. The system comprises of two types of unit namely type "P" and "Q". Type "P" unit is having two sub units in parallel configuration while type "Q" unit is having tree sub unit in warm standby redundancy. The whole structure is working on the principal of 3-out-of-*(P+Q): F* redundancy. Markov process and failures/repair of the components of the system are used to develop a Mathematical model. SVT and LT are used to find out various reliability measures of the same. Sensitive failure associated to reliability and MTTF is obtained by aid of sensitivity analysis.

Keywords: 3-out-of-(*P+Q): F* system; sensitivity analysis; warm standby; Markov process; Mathematical modeling.

Sport Statistics – When Figures are more than Numbers

Christophe Ley

University of Luxembourg

Christophe.ley@uni.lu

This talk will provide a gentle introduction to the growing world of sport analytics. After talking about its genesis and giving some striking examples from professional sports, I will describe how one can use probability distributions to model the outcomes of football matches, and how this can be combined with machine learning procedures to predict big tournaments and hereby even outperform bookmakers. I will conclude with an outlook on how these findings can be translated to sports medicine and, in particular, the estimation of injury risks.

REFERENCES

Ley, C. and Dominicy, Y. (eds) (2020). *Science meets Sports: when Statistics are more than Numbers*. Cambridge Scholars Publishing.

Dominicy, Y. and Ley, C. (eds) (2023) *Statistics Meets Sports: What We Can Learn from Sports Data.* Cambridge Scholars Publishing.

Nonparametric Modeling of Event Processes with Applications to Conditional Local Independence Testing

Myrto Limnios and Niels R. Hansen

Department of Mathematical Sciences, University of Copenhagen,
Universitetsparken 5, 2100 Copenhagen Ø (Denmark)

In the context of disease progression analysis, estimating the causal effect of a time continuous treatment assigned to a given population is an important problem. This is particularly relevant for understanding (the causal) underlying phenomena in many applications gathering massive high-dimensional and time-dependent data structures, ranging from biomedicine to financial markets. In practice, existing learning algorithms inferring the underlying causal graph consider the progression of the recorded markers as time continuous event processes, for which it is required to specify a model, see e.g., algorithms for directed mixed graphs [4], and the Causal Analysis algorithm [5]. We propose, in this work, a nonparametric model for testing if, a process directly influences another when conditioned on the history of others. Known as the (asymmetric) conditional local independence test, we use a version of the Local Covariance Measure from [1], where we model both the unknown intensity process and the test statistic using their finite order Volterra expansion. This results in a linear combination of tensor products of kernel functions composed of stochastic integrals w.r.t. the event processes observed up to that time. Under some assumptions of sparse decomposition and adequate regularity, the optimal parameters involved in the Volterra expansions are solution of a LASSO penalized method. Finite-sample concentration bounds for the estimation and prediction errors are derived, yielding data-driven optimal weights for the LASSO penalty term, that allow for sparse and heterodastic expansions. These results are obtained by investigating non asymptotic probabilistic Bernstein bounds for time-dependent martingales, following the works of [2, 3]. Once the optimal parameters are estimated for a particular node of the graph, a practical implementation of the statistical test based on cross-validation and sample splitting is proposed and adapted from [1], that can be plugged in the aforementioned learning graphs algorithms.

REFERENCES

[1] Christgau A. M., Petersen L., Hansen N. R., Nonparametric Conditional Local Independence Testing, arxiv.2203.13559, 2022.

[2] Bacry E., Bompaire M., Gaiffas S., Muzy J.-F., Sparse and low-rank multivariate Hawkes processes, In: Journal of Machine Learning Research, 2020, 21(50), pp. 1-32.

[3] Hansen N. R., Reynaud-Bouret P., Rivoirard V., Lasso and probabilistic inequalities for multivariate point processes, In: Bernoulli, 2015, 21(1), pp. 83-143.

[4] Mogensen S. W., Malinsky D., Hansen N. R., Causal learning for partially observed stochastic dynamical systems, In: Proceedings of the 34 th conference on Uncertainty in Artificial Intelligence, 2018, pp. 350-360.

[5] Meek C., Toward learning graphical and causal process models, In: Proceedings of the UAI 2014 workshop causal inference: Learning and prediction, 2014, pp. 43-48.

Anomaly detection for compositional data using support vector data description

Marianthi Markatou*

Department of Biostatistics, University at Buffalo

markatou@buffalo.edu

Safety of medical products is a major public health concern worldwide; continual monitoring of medical products after their introduction to the market is essential for patient and consumer safety. Furthermore, likelihood ratio test approaches constitute a class of rigorous statistical tools that permit objective identification of adverse events associated with a specific drug or a drug class. In this talk, we briefly review existing likelihood ratio test approaches and propose a novel class of likelihood ratio tests for adverse event identification that perform well when the underlying Poisson model assumption is violated. Our approach uses an alternative parametrization to formulate a unified framework with a common test statistic that can handle both, Poisson and zero-inflated Poisson models. We discuss the construction and algorithmic implementation of the proposed methods and present simulations and real data examples to illustrate their performance and compare with existing in the literature likelihood ratio tests.

**Joint work with Anran Liu and Saptarshi Chakraborty*

Likelihood Methods in Pharmacovigilance

Thi Thuy Van Nguyen[1,2], Cédric Heuchenne[3,1] and Kim Phuc Tran[4]

[1] HEC Liège - Management School of the University of Liège, Belgium,

[2] International Research Institute for Artificial Intelligence and Data Science, Dong A University, Danang, Vietnam,

[3] UCLouvain Brussels, Belgium,

[4] University of Lille, ENSAIT, GEMTEX, Lille, France

C.Heuchenne@uliege.be

Traditional analysis methods cannot be used directly on compositional data (CoDa) due to their special structure: CoDa are vectors for which components are strictly positive, sum to a constant and they often represent proportions, percentages, concentrations, or frequencies of some whole. In this talk, we will summarize the most popular machine learning techniques that treat CoDa, including principal component analysis, clustering, classification, and regression. Although these techniques have been developed and widely used for many years, most of them are applied to normal data, and research related to these techniques on CoDa is still limited. Based on some specific ideas developed into those methodologies, we will introduce an efficient transformation method based on the Dirichlet density estimation to transform CoDa into real data and apply these transformed data in anomaly detection using support vector data description (SVDD). A simulation example to illustrate this method will also be provided.

A Dynamic Spatiotemporal Stochastic Volatility Model with an Application to Environmental Risks

Philipp Otto, Osman Doğan, Süleyman Taşpınar

Leibniz University Hannover, Istanbul Technical University, City University New York

Corresponding author: philipp.otto@ikg.uni-hannover.de

This article introduces a dynamic spatiotemporal stochastic volatility (SV) model with explicit terms for the spatial, temporal, and spatiotemporal spillover effects. Moreover, the model includes time-invariant site-specific constant log-volatility terms. Thus, this formulation allows to distinguish between spatial and temporal interactions, while each location may have a different volatility level. We study the statistical properties of an outcome variable under this process and show that it introduces spatial dependence in the outcome variable. Further, we present a Bayesian estimation procedure based on the Markov Chain Monte Carlo (MCMC) approach using a suitable data transformation. After providing simulation evidence on the proposed Bayesian estimator's performance, we apply the model in a highly relevant field, namely environmental risk modeling. Even though there are only a few empirical studies on environmental risks, previous literature undoubtedly demonstrated the importance of climate variation studies. For example, for local air quality in Northern Italy in 2021, we show pronounced spatial and temporal spillovers and larger uncertainties/risks during the winter season compared to the summer season.

REFERENCES

Otto, P., Doğan, O., & Taşpınar, S. (2022). A Dynamic Spatiotemporal Stochastic Volatility Model with an Application to Environmental Risks. *arXiv preprint arXiv:2211.03178*.

The characteristic function of the discrete Cauchy distribution and the Cauchy-Cacoullos family of discrete distributions

Nickos Papadatos

Department of Mathematics, National and Kapodistrian University of Athens,

Panepistemiopolis, 157 84, Athens, Greece

npapadat@math.uoa.gr

We investigate a discrete distribution with integer values that was proposed by Cacoullos some years ago. In particular we give closed forms for the normalizing constant and the characteristic function. Moreover, the family is extended to a larger one, the Cauchy-Cacoullos family, that includes a scale-type parameter and we investigate inference problems as well as properties like infinite divisibility.

Dimensionality reduction and clustering

Nicos G. Pavlidis[1], Panagiotis Anagnostou[2] and Sotiris Tasoulis[2]

[1] Lancaster University, [2] University of Thessaly

n.pavlidis@lancaster.ac.uk

We consider the problem of clustering high dimensional data focusing on linear dimensionality reduction methods designed to reveal cluster structure. We show how such approaches can be understood as instantiations of a common framework, and discuss their theoretical properties. We illustrate the effectiveness of derived algorithms on a range of applications. We highlight the trade-off between more sophisticated but computationally expensive methods, compared to simpler algorithms that are computationally (very) efficient. We describe extensions to obtain interpretable clusters, and perform clustering interactively. Our most recent work deals with the issue of ensemble clustering and specifically uncertainty quantification.

References

Pavlidis, N. G., Hofmeyr, D. and Tasoulis, S. (2016). Minimum Density Hyperplanes. *Journal of Machine Learning Research*, **17** (156), 1-33.

Hofmeyr, D., Pavlidis N. G, and Eckley, I. (2019). Minimum Spectral Connectivity Projection Pursuit: Divisive clustering using optimal projections for spectral clustering. *Statistics and Computing*, **29**, 391-414.

Anagnostou, P., Tasoulis, S. Plagianakos, V., and Tasoulis, D. (2023). HiPart: Hierarchical divisive clustering toolbox. *The Journal of Open Source Software*, **8** (84), 5024.

Distribution-free Control Charts for Monitoring Dispersion in Finite Horizon Productions

Theodoros Perdikis[a] , Giovanni Celano[b*], Subhabrata Chakraborti[c]

[a]Athens University of Economics and Business Athens, Greece

[b]Università di Catania, Catania, Italy,

[c]University of Alabama, Tuscaloosa, AL, USA

giovanni.celano@unict.it

Distribution-free, (also known as nonparametric), control charts have been shown to be efficient tools for on-line monitoring of lots of parts manufactured within a finite production horizon (FHP). Despite the partial process knowledge at the beginning of production in a FHP process, a distribution-free control chart can be started without any restrictive assumption about the quality parameter distribution. In this work, three nonparametric Shewhart-type control charts are investigated to monitor the scale parameter in FHP processes. The violation of equality of medians between reference and test samples is investigated and a fourth control chart based on Moses test is introduced. An extensive numerical analysis is performed to select the best performing control chart for the investigated problem. Metrics of performance suitable to the FHP scenario have been considered. A real industrial example is also presented to show a practical implementation of the investigated charts.

Statistics and Machine Learning in industry: combining heterogeneous or multi-scale model outputs

Jean-Michel Poggi

LMO, University Paris-Saclay, Orsay, France & University Paris Cité, Dept SD, IUT Paris-Rives de Seine, Paris, France

jean-michel.poggi@universite-paris-saclay.fr

This paper proposes some applied considerations about the prominent role of data and the necessity to focus on the data produced by the algorithms. Applied statisticians could benefit a lot to experiment more systematically some machine learning (ML) methods. Sequential learning proposes to optimally combine, and therefore also to optimally reject, data produced by algorithms, whether or not based on statistical methods, or by experts. As a statistician, let me highlight this important contribution to applied statistics methodology by considering two examples of forecasting in industry. These examples use ML algorithms to combine "second-order" data, i.e., data from model outputs coming from heterogeneous experts (first example, see Auder et al. 2016) and experts related to multiscale aggregates (second example, see Goehry et al. 2020). We will use the sequential prediction (see Cesa-Bianchi, Lugosi 2006) paradigm which allows to combine several models of internal structure, known or unknown, built on very different assumptions in a unified approach. A very attractive feature is that it does not require any prior knowledge about the internal way used by each expert to generate predictions. The crucial ingredient is to not select the best model or the best resolution but at the contrary by leveraging diversity and optimally organize the cooperation of the different models or different resolutions.

REFERENCES

Auder B., Bobbia M., Poggi J-M. and Portier P. (2016). Sequential Aggregation of Heterogeneous Experts for PM10 Forecasting. *Atmospheric Pollution Research*, 7, 1101-1109.

Cesa-Bianchi N. and Lugosi G. (2006). *Prediction, learning, and games*, Cambridge University Press.

Goehry B., Goude Y., Massart P. and Poggi J-M. (2020). Aggregation of Multi-scale Experts for Bottom-up Load Forecasting, *IEEE Transactions on Smart Grid*, 11, 3, 1895-1904.

Stochastic modeling of the elastic properties of carbon-fiber-reinforced 3D printed filaments using polynomial chaos expansion

Efstratios Polyzos[1,*] and Lincy Pyl[1]

Department of Mechanics of Materials and Constructions, Vrije Universiteit Brussel (VUB), 1050 Brussels, Belgium,

Efstratios.Polyzos@vub.be

This article introduces a novel stochastic analytical approach for modeling the elastic properties of 3D printed filaments reinforced with short carbon fibers. The approach incorporates aspect ratio data from a few fibers extracted from micro-CT images to estimate the stiffness contribution tensor (***N***-tensor) of thousands of fibers found in the filament. A data-driven polynomial chaos expansion method is used to approximate fibers with varying aspect ratios. The ***N***-tensor is then incorporated into three effective field methods, the Non-Interaction, Mori-Tanaka, and Maxwell, to calculate the elastic properties of the filament. The effectiveness and accuracy of the approach are demonstrated by applying it to filaments made from recycled poly(ethylene terephthalate) reinforced with recycled short carbon fibers for additive manufacturing of composite parts. The results of the stochastic approach are compared to experimental data and illustrate a good approximation of the elastic properties for samples with volume fractions up to 30%.

Secure and Robust Federated Learning with Explainable Artificial Intelligence for Healthcare Systems

Ali Raza, Kim Phuc Tran*, Ludovic Koehl

University of Lille, ENSAIT, ULR 2461 - GEMTEX - Génie et Matériaux Textiles, F-59000 Lille, France

*kim-phuc.tran@ensait.fr

The demand for resources in the healthcare sector, such as clinical doctors, nurses, and medical equipment is steadily increasing which is the aging population around the globe. In order to address such challenges researchers and stakeholders are proposing smart healthcare systems that can assist the healthcare sector while being effective and less costly. In recent decades, deep learning-based smart healthcare systems have outperformed existing approaches in various applications. However, such deep learning models usually require many good quality training data, which brings data privacy concerns among the data owners. In order to the above mentioned challenges, we proposed a privacy-preserving, efficient, and interpretable / explainable Artificial Intelligence -based end-to-end framework with Federated Learning (FL) to address the limitations of deep learning applications for EEG signal classification and anomaly detection. The key idea of the proposed framework is to train a joint model using updates from local clients who train local models using their local training data, without sharing their raw data directly with other parties. This enhances the privacy of data owners and helps train robust deep-learning models [1] [2]. However, FL alone does not meet the desired criteria for smart healthcare systems. Hence, in order to address such challenges we are developing methods to detect poisoning attacks in FL without ever accessing the local raw data directly, using techniques such as clustering of model updates and majority voting.

REFERENCES

[1] Raza, Ali, et al. "Designing ECG monitoring healthcare system with federated transfer learning and explainable AI." Knowledge-Based Systems 236 (2022): 107763.

[2] Raza, Ali, et al. "AnoFed: Adaptive anomaly detection for digital health using transformer-based federated learning and support vector data description." Engineering Applications of Artificial Intelligence 121 (2023): 106051.

Data transformations and transfer functions

Milan Stehlík

Linz Institute of Technology, Johannes Kepler University in Linz, Austria

Institute of Statistics, Universidad de Valparaíso, Valparaíso, Chile

mlnstehlik@gmail.com, Milan.Stehlik@jku.at

During the talk we speak on learning mechanisms of data transformation and aggregation. We will introduce SPOCU transfer function and provide some of its unique properties for processing of complex data. Statistical learning will be also discussed.

REFERENCES

J. Kiselak, Y. Lu, J. Svihra, P. Szepe, M. Stehlik (2020) "SPOCU": scaled polynomial constant unit activation function, Neural Computing and Applications, DOI: 10.1007/s00521-020-05182-1

Hermann, P.; Mrkvicka, T.; Mattfeldt, T., Minarova, M.; Helisova, K.; Nicolis, O.; Wartner, F.; Stehlik M. (2015), Fractal and stochastic geometry inference for breast cancer: a case study with random fractal models and Quermass-interaction process. Statistics in Medicine, 34, 2636-2661

Stehlik, M. (2016) On convergence of topological aggregation functions, Fuzzy Sets and Systems, 287, 48-56.

Stein EWMA Control Charts for Count Processes

Christian H. Weiß
Helmut Schmidt University, Dep. Mathematics & Statistics, Hamburg, Germany.
weissc@hsu-hh.de

The monitoring of serially independent or autocorrelated processes of unbounded or bounded counts is considered, having a Poisson or (negative) binomial marginal distribution under in-control conditions. Traditional control charts for such count processes, like the Shewhart-type c- or np-charts, or the memory-type exponentially weighted moving-average (EWMA) charts, are well-suited to detect sole shifts in the process mean. But they may fail in "more sophisticated" out-of-control scenarios, where the mean changes together with further distributional properties, or where purely distributional changes (not affecting the mean) happen. These might be increases or decreases in dispersion compared to the in-control model (overdispersion or underdispersion, respectively), or an excessive number of zero counts (zero inflation), to mention those being most relevant in practice. For such cases, novel EWMA-type control charts are proposed, which are based on so-called Stein identities, and which can be flexibly adapted to uncover zero inflation, over- or underdispersion. Stein identities are moment-based equations to uniquely characterize a specified distribution family, but they were also utilized for statistical approaches such as goodness-of-fit tests. In this research, diverse Stein EWMA control charts are developed, the average run length performance of which is investigated by simulations. Their usefulness is demonstrated by a real-world data example on registrations in the emergency department of a children's hospital, which illustrates the application and interpretation of the novel Stein EWMA charts in practice.

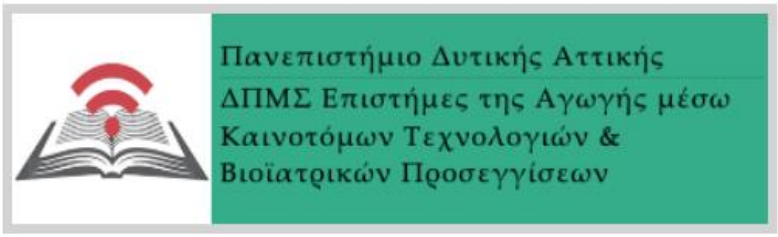
REFERENCES

Sudheesh, K.K. and Tibiletti, L. (2012) Moment identity for discrete random variable and its applications. *Statistics*, **46**, 767–775.

Weiß, C.H. (2008) Thinning operations for modelling time series of counts—a survey. *AStA Advances in Statistical Analysis*, **92**, 319–341.

Weiß, C.H., Puig, P. and Aleksandrov, B. (2023) Optimal Stein-type goodness-of-fit tests for count data. *Biometrical Journal*, **65**, 2200073.

ΧΟΡΗΓΟΙ / SPONSORS




ΠΑΝΕΠΙΣΤΗΜΙΟ
ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ
UNIVERSITY OF WEST ATTICA
35th Panhellenic and
1st International Statistics
Conference
Statistics in Health Sciences
25-28 May 2023
University of West Attica
CAMPUS ANCIENT OLIVE GROVE
Informations
https://gsi-conference.uniwa.gr
Organization
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ
UNIVERSITY OF WEST ATTICA