BFF10

Book of Abstracts

10th Bayesian, Fiducial, and Frequentist Conference
Salzburg, Austria

Compiled from speaker submissions · organized by session, in presenting order

Table of Contents

Claudia Kirch — Bayesian nonparametric spectral analysis of locally stationary processes
William Peden — Bayesianism, Frequentism, and Imprecise Probabilities

Session 8

Hanti Lin — A Plea for History and Philosophy of Statistics and Machine Learning
Armin Schwartzman — The method of arbitrary functions and frequentist probabilities
Christian Hennig — Some confusing issues with frequentism that a Bayesian or fiducial approach cannot solve

Session 9

Aiyou Chen — Confidence Intervals for Rate Estimation with Importance Sampling in Autonomous Vehicle Evaluation
Frank Konietschke — Advancing Decision-Efficiency in (Pre)-clinical Research via Nonparametric Sequential Frameworks
Christopher Saunders and Peter Vergeer — Fairness Properties for likelihood ratios in the interpretation of forensic evidence

Poster Session

P1. Dominic Edelmann — The Optimal Standardization Factor for the Variance of the Normal Distribution: A Historical Perspective and Recommendations for Practice
P2. Hee-Geon Kang — Generalized Fiducial Inference for Poisson Point Processes
P3. Adriana Gonzalez Sanchez — Finite Sample Inference for Probit Regression
P4. Nouman Qureshi — A Quasi-Bayesian Shrinkage Estimator for Adaptive Cluster Sampling in Rare and Hard-to-Reach Populations
P5. Patrick Langthaler — Nonparametric analysis of right censored data via the Mann-Whitney effect and overlap index
P6. Christian Covington — Extending the generalized Rosenblatt Transform
P7. Francesco De Pretis — Robust Sequential Bernoulli Updating under Latent External Shocks: An Extended Fiducial Perspective
P8. Dennis Christensen — Posterior uncertainty for kernel density estimates
P9. Nicolas Pascal Dietrich — Nonparametric Estimation and Convergence Properties of Archimax Copulas
P10. Junyi Li — Change Point Detection Inference with Unknown Number of Change Points Using Repro-Samples
P11. Kai Schärer — Nonparametric Estimation of Conditional Distribution Functions via Empirical Checkerboard Copula Approximations

Keynote 1

E-Values and BFF: they should be BFF!

Peter Grünwald
CWI and Leiden University · pdg@cwi.nl

E-values are an alternative to p-values, providing a notion of statistical evidence that is simultaneously more robust and flexible. Here I will review and study e-values from the BFF angle. While e-values are frequentist “at heart”, their construction often involves the use of priors. I will extensively compare such e-values to Bayesian methods, contrasting (a) e-value and Bayes factor-based tests, (b) showing how e-based (anytime-valid) confidence intervals get wide rather than wrong if the prior is badly chosen; (c) showing how very Bayes-like quantities arise even for ‘nonparametric’ e-values that superficially do not look like Bayes factors or even likelihood ratios at all. Finally I turn to e-based decision theory, where the “e-posterior” interpolates between a standard Bayes posterior and Martin’s inferential models, thereby establishing a firm connection between e-values and the fiducial tradition as well.

Keynote 2

Outrigger local polynomial regression

Richard J. Samworth
University of Cambridge · r.samworth@statslab.cam.ac.uk

Standard local polynomial estimators of a nonparametric regression function employ a weighted least squares loss function that is tailored to the setting of homoscedastic Gaussian errors. We introduce the outrigger local polynomial estimator, which is designed to achieve distributional adaptivity across different conditional error distributions. It modifies a standard local polynomial estimator by employing an estimate of the conditional score function of the errors and an 'outrigger' that draws on the data in a broader local window to stabilise the influence of the conditional score estimate. Subject to smoothness and moment conditions, and only requiring consistency of the conditional score estimate, we first establish that even under the least favourable settings for the outrigger estimator, the asymptotic ratio of the worst-case local risks of the two estimators is at most 1, with equality if and only if the conditional error distribution is Gaussian. Moreover, we prove that the outrigger estimator is minimax optimal over Hölder classes up to a multiplicative factor $A_{\beta,d}$, depending only on the smoothness $\beta \in (0, \infty)$ of the regression function and the dimension d of the covariates. When $\beta \in (0, 1]$, we find that $A_{\beta,d} \leq 1.69$, with the limit of $A_{\beta,d}$ equal to 1 as β decreases to 0. A further attraction of our proposal is that we do not require structural assumptions such as independence of errors and covariates, or symmetry of the conditional error distribution. Numerical results on simulated and real data validate our theoretical findings; our methodology is implemented in R and available on CRAN.

Session 1

Graph Release with Assured Node Differential Privacy

Tianxi Li
University of Minnesota · tianxili@umn.edu

Differential privacy is a well-established framework for safeguarding sensitive information in data. While extensively applied across various domains, its application to network data — particularly at the node level — remains underexplored. Existing methods for node-level privacy either focus exclusively on query-based approaches, which restrict output to pre-specified network statistics, or fail to preserve key structural properties of the network. In this talk, I will introduce GRAND (Graph Release with Assured Node Differential privacy), which is, to the best of our knowledge, the first network release mechanism that releases entire networks while ensuring node-level differential privacy and preserving structural properties. Under a broad class of latent space models, we show that the released network asymptotically follows the same distribution as the original network. The effectiveness of the approach is evaluated through extensive experiments on both synthetic and real-world datasets.

Inferring model structure from data: MCMC for DAGs

Radu Craiu

University of Toronto · radu.craiu@utoronto.ca

Inferring a directed acyclic graph (DAG) given data is computationally challenging. Current state-of-the-art MCMC methods for graph inference efficiently scan the space by first considering a restricted search space and iteratively expand the space until a stopping criterion is met. We estimate the error introduced from current methods that use restricted spaces compared to the full space and develop a novel MCMC method that reduces this error. The new adaptive algorithm allows for either expansion or contraction of the search space at any iteration. Extensive numerical experiments demonstrate the efficiency of the new algorithm and compare its performance with existing methods. This is joint work with Kieran Campbell and Morris Greenberg.

Combinatorial Extrapolation

Peter Bühlmann

Seminar for Statistics, ETH Zürich · buhlmann@stat.math.ethz.ch

Predicting the effects of unseen combinations from limited observations is a fundamental challenge in domains such as drug discovery and hyperparameter optimization. We study combinatorial extrapolation, where training data contain only axis-aligned samples with a single active covariate, while test inputs involve multiple active covariates. We introduce DExtrI, a method for extrapolating interaction effects beyond the support of the training data. We establish theoretical conditions under which such extrapolation is possible and demonstrate on synthetic and real-world datasets that DExtrI accurately predicts outcomes for previously unseen combinations. We conclude with an outlook on perturbation-based cellular models, where current approaches still leave considerable room for improvement in capturing higher-order interactions. This is joint work with Marin Sola, Xinwei Shen and Michael Vollenweider.

Session 2

The E-measure

Nick Koning
Erasmus University Rotterdam · n.w.koning@ese.eur.nl

We introduce the E-measure: a measure-like generalization of the E-value to a class of hypotheses. Unlike classical measures, E-measures are closed under infimums instead of addition. They arise from a compatibility axiom with logical implications, that there should be at least as much evidence against more specific hypotheses. We show that E-measures are the only non-dominated such objects, if the hypothesis class is closed under intersections. We propose to use the E-measure to present all the relevant evidence for a problem, where the relevance is captured by the choice of hypothesis class. We showcase this by applying the E-measure to decision making, inducing a hypothesis class from the uncertain consequences of decisions. This results in uniform E-consequence bounds on decisions, which nest high-probability loss bounds. Correcting for multiplicity, we consider ‘familywise evidence’ and ‘false evidence rate’ control, generalizing from errors and discoveries to continuous evidence. Remarkably, E-measures control these without multiplicity correction if the hypothesis class is intersection-closed. Moreover, we obtain a ‘frequentist’ notion of updating from E-prior to E-posterior. Abstracting the notion of a ‘hypothesis’, we advocate for using E-measures for any unknown quantity, leading to predictive E-measures.

E-values for Contingency Tables

Thorsten Dickhaus
University of Bremen · dickhaus@uni-bremen.de

Analyzing many contingency tables simultaneously is important in the context of genetic association studies when analyzing associations between categorical genetic markers and a categorical (often binary) disease status. The research question that we are interested in is how to define an e-value for a contingency table, which is easy to compute and powerful. By “easy to compute”, we mean that resource-intensive operations like a loop over all contingency tables with given marginal counts shall be avoided. This requirement refers to the situation that hundreds of thousands of such e-values have to be computed for one and the same dataset in the case of a genome-wide association study (GWAS). By “powerful”, we mean that the e-value should (with high probability) be larger than the p-value based on Fisher's exact test calibrated to the e-value scale by a p-to-e-calibrator. Our primarily intended use case is a GWAS which is either multi-centric (independent patient groups are recruited at different locations) or group-sequential (independent patient groups are recruited at the same location at different time points). These two sampling schemes are realistic for GWAS, and e-values (which are easy to compute and powerful) can allow the data analysts to combine the evidence across centers or across time points, respectively, in a convenient manner (e.g., by multiplication or by averaging of e-values). For the special case of a 2×2 contingency table, we discuss several possibilities to define such e-values, and we characterize sampling schemes under which the usage of each of these e-values is particularly appropriate. This is joint work with Yongqi Wang, Sebastian Arnold, Francesca Giuffrida, and Peter Grünwald.

Multivariate Tail Probability Approximations

John Kolassa
National Science Foundation · jkolassa@nsf.gov

This manuscript extends the univariate Lugannani and Rice saddlepoint tail approximation to multiple dimensions. The resulting approximation uses a multivariate Gaussian approximation to the distribution of the signed roots of the log likelihood statistics. As in the univariate case, the next correction term involves the difference between reciprocals of the signed root of the likelihood statistics and the analogous Wald statistics. Extensions to lattice variables and conditional distributions are provided.

Note: Thanks for the invitation.

Session 3

Uncertainty quantification via cross-validation and its variants under algorithmic stability

Hannes Leeb
University of Vienna · hannes.leeb@univie.ac.at

Recently, there has been substantial interest in statistical guarantees for cross-validation (CV) methods of uncertainty quantification in statistical learning (cf. Barber et al. (2021a), Liang and Barber (2025), Steinberger and Leeb (2023)). These guarantees should hold under minimal assumptions on the data generating process and conditional on the training data, because numerous predictions are usually computed based on one and the same training sample. We push this objective to the limit: We prove asymptotic conditional conservativeness of CV, that is, the probability of the actual coverage probability, conditional on the training data, undershooting its nominal level vanishes asymptotically, under minimal assumptions. In particular, we impose a stability condition, require that the prediction error is stochastically bounded, and show that neither condition can be dropped in general. By way of an asymptotic equivalence result, we also show that the closely related CV+ method of Barber et al. (2021a) provides exactly the same conditional statistical guarantees as CV in large samples, thereby extending the range of applicability of CV+ to the high-dimensional regime. We conclude that, in view of its marginal coverage guarantee, CV+ does indeed improve over simple CV. For our proofs, we introduce a new concept called Lévy gauge, which can be of independent interest.

The Implicit Bootstrap

Maria-Pia Victoria Feser

Department of Statistical Sciences, University of Bologna · maria.victoriafeser@unibo.it

Inference in many complex models is challenging due to analytical and numerical obstacles, often arising from the presence of latent variables introduced in a core model to account for data features such as truncation, censoring, other types of missing (not at random) data, misclassification, complex dependencies, and others. In these instances, deriving consistent estimators, which might require, e.g., methods for approximating integrals, is not only challenging but can be prohibitive when using bootstrap methods for inference. Moreover, numerical approximations might introduce finite sample bias, which can jeopardize the inferential accuracy of methods based on them. We propose an alternative bootstrap-type percentile method, namely the Implicit Bootstrap (IB), that bypasses the need for the specification (and computation) of consistent estimators for models incorporating complex data features. It is based on auxiliary statistics that can be defined implicitly but are computationally easier to obtain than consistent estimators. Typically, these auxiliary statistics are estimators associated with the core model, i.e., ignoring the data features, and the latter are introduced in the inferential procedure using simulations. We demonstrate the first and second order properties of the IB, under similar requirements used to establish the second-order accuracy of the studentized bootstrap, without the necessity of a consistent estimator. We present empirical evidence supporting the theoretical claims through simulation studies involving challenging scenarios.

Uncertainty quantification for functionals in constrained inverse problems

Mikael Kuusela

Carnegie Mellon University · mkuusela@andrew.cmu.edu

Ill-posed inverse problems are characterized by a large set of parameters that are consistent with the observed data. Rigorous and useful finite-sample inference in these problems is possible by inferring functionals of the unknown parameter subject to physical constraints. The presence of the functional, constraints and potential unidentifiability makes this a challenging inference problem. In this talk, I will show that confidence intervals in this setting can be obtained through the inversion of a specific, non-standard likelihood-ratio test. The critical values of this test are typically not available in closed form so we propose calibrating the test over a bounded subset of the parameter space chosen so that it contains the true parameter with a high probability. I will then describe concrete computational strategies based on optimization and sampling for performing this calibration and test inversion in practice. I will first illustrate these ideas with low-dimensional toy examples and then present a realistic application to a moderately high-dimensional particle unfolding problem, where the new confidence intervals are significantly shorter than previous alternatives while maintaining correct finite-sample coverage. Time allowing, I will discuss the potential for carrying out this type of inference in ultra-high-dimensional data assimilation problems as well as recent extensions to the case of multiple functionals of interest.

Session 4

Addressing Small-Sample Challenges in Rare Disease Research Using Nonparametric Methods

Georg Zimmermann

Department of Artificial Intelligence and Human Interfaces, University of Salzburg, Austria · georg.zimmermann@hotmail.de

Rare diseases present unique challenges for the statistical analysis. Relevant outcomes often need to capture multiple aspects of the disease and may include quantitative, validated, and patient-relevant endpoints. In addition, sample sizes are typically small, and missing data can further reduce the available information. Statistical methods therefore need to account appropriately for these characteristics. To address these challenges, we systematically evaluate a range of univariate and multivariate nonparametric methods and assess their empirical performance. Although no universal solution emerges, the developed simulation framework provides detailed insights into the scope of applications of the different approaches. These findings may support methodological recommendations grounded in both statistical evidence and clinical expert opinion.

Note: Requested talk to be scheduled for Friday, July 10.

Two-Layered Logics for Uncertainty

Marta Bílková

Institute of Computer Science of the Czech Academy of Sciences · bilkova@cs.cas.cz

Reasoning about information, its potential incompleteness, uncertainty, and even contradictoriness need to be dealt with adequately. Various logical systems for reasoning in contexts when the available data about observed events are uncertain, unknown, or unreliable, have been considered. They use varied syntactical frameworks to capture semantical aspects like degrees of truth, likelihood, or subjective belief. Separately, incompleteness, uncertainty, and contradictoriness of information have been captured successfully by various logical formalisms and (classical) probability theory. While incompleteness and uncertainty are typically easily accommodated within one formalism, e.g. within various models of imprecise probability, contradictoriness and uncertainty less so — conflict or contradictoriness of information is rather chosen to be resolved before entering the reasoning process than to be reasoned with. One way to reason with conflicting information is to separate positive and negative support — evidence in favour and evidence against a statement — in the semantics and quantify it rather independently. This two-dimensionality gives rise to propositional logics with bi-lateral semantics: they are interpreted over twist-product algebras, bi-lateral frame semantics or bi-lattices, the well known Belnap–Dunn logic of First Degree Entailment being a prominent example. Belnap–Dunn logic with its double-valuation frame semantics can in turn be taken as a base logic for defining various uncertainty measures like probabilities on de Morgan algebras or belief functions. Two-layered logics, a simple two-sorted format developed for reasoning about probabilities of classical events, separate two layers of reasoning: the inner layer consists of a logic chosen to reason about events or evidence, the connecting modalities are interpreted by a chosen uncertainty measure on propositions of the inner layer, typically a probability or a belief function, and the outer layer consists of a logical framework to reason about probabilities or beliefs. Logics originally introduced by Fagin, Halpern, and Megiddo use classical propositional logic on the inner layer to capture events, a probability measure to interpret a modality, and systems of linear inequalities on the outer layer. Hájek, Godo, and Esteva, on the other hand, suggested using many-valued Łukasiewicz logic on the outer layer, to capture the quantitative reasoning about probabilities within a propositional logical language. Our main objective is to extend the apparatus of two-layered modal logics for the formalisation of reasoning with uncertain information, which itself might be non-classical, i.e., incomplete or contradictory. Many-valued logics with a bi-lateral semantics can be used on the outer layer to reason about belief, likelihood or certainty based on potentially incomplete or contradictory evidence, building on Belnap–Dunn logic of First Degree Entailment as an inner logic of the underlying evidence. This results in two-layered logics suitable for various scenarios: expansions of Łukasiewicz logic are adequate in cases when (possibly aggregated) evidence yields a pair of probability measures or a belief function (on a De Morgan algebra), while expansions of Gödel logic are useful to reason about comparative uncertainty, or to capture reasoning about qualitative probability.

References

P. Baldi, P. Cintula, and C. Noguera. Classical and Fuzzy Two-Layered Modal Logics for Uncertainty: Translations and Proof-Theory. International Journal of Computational Intelligence Systems, 13:988–1001, 2020.

N.D. Belnap. How a computer should think. In H. Omori and H. Wansing, editors, New Essays on Belnap-Dunn Logic, volume 418 of Synthese Library. Springer, Cham, 2019.

M. Bilkova, S. Frittella, D. Kozhemiachenko, and O. Majer. Qualitative reasoning in a two-layered framework. International Journal of Approximate Reasoning, 154:84–108, 2023.

M. Bilkova, S. Frittella, D. Kozhemiachenko, and O. Majer. Two-layered logics for probabilities and belief functions over Belnap–Dunn logic. Mathematical Structures in Computer Science, 35:e1, 2025.

J. Michael Dunn. Intuitive semantics for first-degree entailments and ‘coupled trees’. Philosophical Studies, 29(3):149–168, 1976.

R. Fagin, J.Y. Halpern, and N. Megiddo. A logic for reasoning about probabilities. Information and Computation, 87(1–2):78–128, 1990.

P. Hájek, L. Godo, and F. Esteva. Fuzzy logic and probability. In UAI ’95: Proceedings of the Eleventh Annual Conference on Uncertainty in Artificial Intelligence, pages 237–244. Morgan Kaufmann, 1995.

Uncertainty Propagation in System Reliability: From Inference on Components to System Risk

Daniel Krpelik

IT4Innovations, VSB – Technical University of Ostrava, Czech Republic · daniel.krpelik@vsb.cz

Reliability is defined as the probability that a system provides its desired functionality — it is the probability of an event. Applications to risk analysis commonly drive decisions about singular trials with no relation to repeated sampling grounding utility benefits by long-term averages. Such problems are specific by dealing with high stakes (catastrophic failures, loss of lives) and insufficient data (small sample size, interval censoring, missing data). Fully probabilistic models tend to include unaccounted assumptions which could lead to false confidence. Imprecise probability theories emerged to rectify this by embracing ignorance into the mathematical models, leading to more conservative but more trustworthy assessments. Expensive complex systems may be ill-suited for subjection to destructive testing for the sake of statistical analysis. Systems reliability theory decomposes the failure event into sub-events associated with functionalities of subsystems and components, and the original system failure event is subsequently represented with a function dependent on the components' state. Statistical inference is usually conducted for the individual components and needs to be propagated to obtain an integrated assessment about the whole system. This propagation poses challenges both theoretically (to ensure validity of the top-level assessments) and numerically (especially for Imprecise Probabilistic models, which include extremization subproblems). In the talk, we will discuss theoretical and practical aspects of imprecise reliability models, propagation of component-level inferential results, and applications to risk-based decision making.

Session 5: Special Invited Session

Celebrating the 10th BFF Conference

From BFF10 to Quantum Probability 100 (and back): Can Broadening Probability Deepen Statistics?

Xiao-Li Meng

Harvard University · meng@stat.harvard.edu

Perhaps echoing the saying that “Coincidence is God’s way of remaining anonymous,” the COVID-delayed celebration of BFF10 coincides with the centenary of quantum probability, commonly traced to Max Born’s 1926 probabilistic interpretation of the wave function, whose temporal propagation is described by the Schrödinger equation, also published in 1926. Despite this long history, quantum probability remains unfamiliar, even intimidating, to many statisticians. Motivated by the coming era of quantum computing and by foundational questions surrounding the Bayesian, fiducial, and frequentist (BFF) schools, this talk explains the contextual probability framework developed by physicist and philosopher Jacob Barandes to better connect quantum theory with ordinary notions of probability. By relaxing the commitment to a common probability space, Barandes’ framework does not require every probability distribution to arise from a single underlying (possibly multivariate) random variable, but instead relates distributions across different, sometimes incompatible, contexts. In this setting, superposition, interference, and the nonexistence of joint distributions arise naturally while preserving a distinction between mathematical representation and physical reality. The talk then observes that analogous phenomena already exist within classical statistics, ranging from improper priors and incompatible conditional distributions to fiducial representation non-invariance, and invites the audience to contemplate whether these similarities are merely suggestive coincidences or hints of a broader probabilistic framework for comparing, contrasting, and enriching the foundations of the BFF schools.

Session 6

Sequential Bayesian experimental design for prediction in physical experiments informed by computer models

Markus Hainy

Johannes Kepler University Linz · markus.hainy@jku.at

In many scientific and engineering domains, physical experiments are often costly, non-replicable, or time-consuming. The Kennedy & O'Hagan (KOH) model framework has become a widely used approach for combining simulator runs with limited experimental observations. Under a Bayesian implementation, the simulator output, model discrepancy, and observation noise are jointly modeled by coupled Gaussian processes, followed by coherent posterior inference and uncertainty quantification. This work presents a genuinely sequential Bayesian experimental design (BED) framework explicitly aimed at improving the predictive performance of the KOH model. We employ a mutual information (MI)–based criterion and develop a hybrid variant that integrates with measures of local model complexity, leading to significantly more efficient design decisions. We further theoretically establish asymptotic links between the MI-based criterion and the classical integrated mean squared prediction error (IMSPE) minimization criterion. In practice, we find the MI-based criterion is more comprehensive and robust than the IMSPE minimization criterion, especially when the model is highly uncertain in the early stages of the experiment. We demonstrate the effectiveness of the proposed methods through both a synthetic example and a real biochemical case study, and compare the MI- and IMSPE-based criteria against several other classical design criteria under sequential (offline) and adaptive (online) BED settings.

Towards the next generation of multi-analyst experiments

Galin Jones
University of Minnesota · galin@umn.edu

Data analysis is a linchpin of modern scientific research. However, the impact of the data analysis pipeline on the reliability of inferences has been understudied. So-called multi-analyst experiments have begun to address this deficit, resulting in claims that the data analysis pipeline itself may be a primary source of unreliability. Unfortunately, due to the research designs employed, the interpretability of those results is limited. In particular, those designs do not allow for the identification of the sources of unreliability. Hence, it is difficult to use their results to constructively identify ways to improve data analysis practices. We attempt to identify and address these limitations with a new design-of-experiments framework, thereby enabling future multi-analyst studies to draw more specific and actionable conclusions.

Understanding prediction Risk in Large Language Models

Peter Song
University of Michigan · pxsong@umich.edu

Prediction risk analysis in large language models is of great importance for the understanding of trustworthiness of AI tools, in which uncertainty quantification plays a central role. Using a token-level foundational decomposition for the prediction risk, we present a unified framework involving Variability, Confidence and Innateness (VCI) to depict different types of uncertainty. Such decomposition is also extended to agentic systems for prediction risk analysis. The replicability for scientific research is used as a running example to demonstrate the decomposition for risk analysis and uncertainty quantification. This is a joint work with Jiuchen Zhang and Rui Nie.

On Theoretical Foundations of Extended Fiducial Inference

Faming Liang

Purdue University, West Lafayette, IN, USA · fmliang@purdue.edu

Extended fiducial inference (EFI) provides a general paradigm for statistical inference in modern data science. Its defining feature is to formulate statistical inference as a problem of solving data-generating equations under uncertainty: EFI first constructs a fiducial law for the unobserved random errors underlying the observations and then propagates this law to the parameter space, either through an inverse mapping or, more generally, through a compatibility measure defined on the inverse fiber. This formulation enables EFI to address several difficulties that arose in Fisher's original fiducial argument, including issues related to conditioning, marginalization, and noninvertibility. We develop general theoretical foundations for EFI that encompass both single-valued and set-valued inverse mappings and apply to both low- and high-dimensional regimes. We illustrate the theory through linear regression, nonlinear regression, logistic regression, their high-dimensional counterparts, and other analytical examples.

Session 7

Universal ordinal inference with extended ranks

Peter Hoff
Duke University · peter.hoff@duke.edu

The predictive accuracy of a regression model depends largely on the distributional support of the outcome variable and the relationship between the mean and variance of the conditional distribution. In this article we develop a single inferential framework that can represent these aspects for a wide range of conditional distributions. Our approach is built upon a monotonically transformed linear regression model and a pseudo-likelihood based on an extended notion of ranks. This extended rank likelihood approach can accommodate any ordinal data type, including continuous and discrete ordered data, and can represent a wide range of mean-variance relationships. Theoretical results show that for parameter estimation, the maximizer of this likelihood is asymptotically efficient at the two extremes of ordinal data — continuous and binary data. Bayesian parameter estimates and posterior prediction intervals are available via a simple Gibbs sampling algorithm. Prediction intervals with guaranteed frequentist coverage are constructed via conformal calibration of the Bayesian posterior distribution.

Bayesian nonparametric spectral analysis of locally stationary processes

Claudia Kirch

Heinrich-Heine-University Duesseldorf, Germany · claudia.kirch@hhu.de

Many real-world phenomena can well be modelled by locally stationary time series with a slowly changing dependency structure. For such time series the estimation of the time-varying spectral density is of particular interest. In this talk we propose a Bayesian nonparametric method for this purpose based on a suitable dynamic Whittle likelihood for locally stationary time series in combination with a Bernstein-Dirichlet process prior. We prove sup-norm posterior consistency and obtain L2-norm posterior contraction rates. Additionally, this methodology enables model selection between stationarity and non-stationarity based on the Bayes factor. The good finite-sample performance of the method is indicated by means of a simulation study and applications to real-life data-sets. This is joint work with Yifu Tang (University of Otago), Kate Lee and Renate Meyer, both from the University of Auckland.

Bayesianism, Frequentism, and Imprecise Probabilities

William Peden

Johannes Kepler University Linz · william.peden@jku.at

To what extent can imprecise probabilities help us combine the strengths of both Bayesian and frequentist approaches? This paper describes a 5-year collaborative research programme that helps to answer this question, conducted by an interdisciplinary team from philosophy, decision theory, and machine learning. The differences between Bayesianism and frequentism are less important when we have rich information about the data-generating process. Instead, the most interesting differences between Bayesianism and frequentism occur under ambiguity: when there is a lack of information about the true probabilities. To make comparisons under ambiguity, we used a tractable decision problem, the “Bernoulli problem”, based on sequences of Bernoulli trials. Agents know that the trials are exchangeable and binomial, but they do not know the true probabilities or even bounds on those probabilities. We investigated imprecise probability learning systems, which can combine Bayesian and frequentist statistical methods. One popular such system is Imprecise Bayesianism, which uses sets of probability distributions. Its key strength is that it can represent the ambiguity in a decision problem through divergence among the probability distributions in an agent’s set. We found a surprising trade-off that we call the Ambiguity Dilemma. The tools for representing ambiguity have a systematic performance cost, compared to traditional Bayesianism, in the Bernoulli problem. Despite the reliable information that is available even in small samples of Bernoulli trials, an Imprecise Bayesian with divergent sets of beta distributions learns slowly, resulting in inferior decision-making performance in this problem. We subsequently found that the Ambiguity Dilemma is very robust. For instance, it also occurs for Dempster–Shafer updating. However, the Ambiguity Dilemma is avoided by Calibration imprecise probability systems, which combine aspects of frequentism and Bayesianism. A Calibration agent can retain the ambiguity-representation tools of Imprecise Bayesianism while achieving learning performance comparable to the best traditional Bayesian agents in the Bernoulli problem. Thus, Calibration imprecise probability systems exemplify the value of integrating Bayesian and frequentist methodological ideas. Our current focus is on comparisons in more complex and realistic decision problems, such as those with noisy data.

Session 8

A Plea for History and Philosophy of Statistics and Machine Learning

Hanti Lin
University of California, Davis · ika@ucdavis.edu

The integration of the history and philosophy of statistics was initiated at least by Hacking (1975) and advanced by Hacking (1990), Mayo (1996) and Zabell (2005), but it has not received sustained follow-up. Yet such integration is more urgent than ever, as the recent success of artificial intelligence has been driven largely by machine learning—a field historically developed alongside statistics. Today, the boundary between statistics and machine learning is increasingly blurred. What we now need is integration, twice over: of history and philosophy, and of two fields they engage—statistics and machine learning. I present a case study of a philosophical idea in machine learning theory (and in formal epistemology) whose root can be traced back to an often under-appreciated insight in Neyman and Pearson's 1936 work on uniformly most powerful unbiased tests (a follow-up to their 1933 classic). This leads to the articulation of an epistemological principle—largely implicit in, but shared by, the practices of frequentist statistics and machine learning—which I call achievabilism: the thesis that the correct standard for assessing non-deductive inference methods should not be fixed, but should instead be sensitive to what is (provably) achievable in specific problem contexts. I close by putting forward the conjecture that achievabilism also underlies the development of nonparametric Bayesian statistics by Freedman and Diaconis. If so, achievabilism is an epistemological principle shared by many frequentists, Bayesians, and machine learning theorists.

Note: Thanks for organizing the conference :)

The method of arbitrary functions and frequentist probabilities

Armin Schwartzman
University of California, San Diego · armins@ucsd.edu

One important criticism of the frequentist interpretation of probability is that it is logically circular: probability cannot be defined as the limiting frequency of a random sequence via the law of large numbers because the law of large numbers depends on the probability law of the random variables in the sequence. In this talk, I explore how Poincaré's and Hopf's method of arbitrary functions may be used to define probabilities as frequentist limits of deterministic dynamical systems that forget initial conditions. This idea may also help explain how LLMs produce outputs with stable distributions despite variations in prompt inputs.

Some confusing issues with frequentism that a Bayesian or fiducial approach cannot solve

Christian Hennig

University of Bologna · christian.hennig@unibo.it

I will present three rather simple but problematic issues with frequentism, namely (a) the misspecification paradox (checking and accepting a model assumption by a misspecification test actively violates the model), (b) the impossibility to identify many elementary dependence patterns from data (such as constant correlation between Gaussian observations), (c) an estimator that is arguably bad for large n because of consistency, namely the BIC used for estimating the number of clusters in a mixture model. I will then discuss how these issues are not exclusively problems for frequentism, but rather have to do with our use and understanding of probability models more generally (particularly requiring them to be “true” in some sense).

Session 9

Confidence Intervals for Rate Estimation with Importance Sampling in Autonomous Vehicle Evaluation

Aiyou Chen
Waymo LLC · aiyouchen@waymo.com

Accounting for both rare events and complex sampling presents challenges when quantifying uncertainty for rate estimation in autonomous vehicle performance evaluation. In this paper, we introduce a statistical formulation of this problem and develop a unified compound Poisson model framework for unbiased rate estimation through the Horvitz–Thompson estimator. Though asymptotic theory for the model is available, the inference of confidence intervals (CIs) in the presence of rare events requires new investigation. We also advocate for a new monotonicity criterion for rate CIs — summing the rates of disjoint types of events should produce not only a higher point estimate but also higher confidence bounds than for the individual rates — that facilitates interpretability in real applications. We propose a novel exponential bootstrap (EB) method for CI construction based on a fiducial argument; it satisfies the monotonicity property, while novel extensions of some existing methods do not. Comprehensive numerical studies show that EB performs well for a wide range of settings relevant to our applications. Fast implementation of EB based on saddlepoint approximation is also developed, which may be of independent interest.

Advancing Decision-Efficiency in (Pre)-clinical Research via Nonparametric Sequential Frameworks

Frank Konietschke

Charité – Universitätsmedizin Berlin · frank.konietschke@charite.de

Traditional sequential designs in early-stage (pre)-clinical research typically rely on parametric assumptions that are frequently violated by small sample sizes, heavy tails, or extreme outliers. To address these limitations and optimize decision efficiency, we introduce a novel, purely nonparametric sequential framework based on exact and asymptotic rank procedures. This distribution-free framework integrates multi-stage interim analyses with sequential rank-sum statistics, allowing early termination for efficacy or futility without assuming underlying normality. We evaluated the framework's performance using extensive Monte Carlo simulations, benchmarking it against traditional fixed-sample and classical parametric group sequential designs across various non-normal and highly skewed data distributions. The proposed rank-based framework significantly reduced the expected sample size. Crucially, it maintained robust Type I error control and preserved target statistical power where parametric alternatives failed due to model misspecification. By providing a mathematically rigorous, distribution-free approach to real-time data monitoring, this framework offers a highly ethical and resource-efficient paradigm for accelerating go/no-go decisions in modern (pre)-clinical translation.

Fairness Properties for likelihood ratios in the interpretation of forensic evidence

Christopher Saunders and Peter Vergeer

South Dakota State University; Netherlands Forensic Institute · christopher.saunders@sdstate.edu

In the construction of values of forensic evidence, formal methods associating a piece of (trace) evidence from an unknown source to a specified source typically result in Bayes factors. They are constructed by assuming formal priors for the nuisance parameters of the specified source's probability distribution of traces, and the nuisance parameters of the relevant background population of sources' probability distribution of traces. These prior distributions may vary in the degree of information they contain about these nuisance parameters. Since we normally have limited measurements on sources and a limited number of control samples, the following becomes important: we need to pool information across sources; if not, Bayes factors become highly dependent on the choice of prior distributions, due to limited information to balance out the prior information. Our recent work studies the relation of the information present in these prior distributions, and the pooling choices, to biases naturally arising in Bayes factors and also in score-based likelihood ratios. These biases may either lead to negative discrimination of rare suspects, or to positive discrimination of rare suspects. In this presentation, we will define what we see as a fair use of evidence; show what type of priors (related to types of Bayes factors found in the forensic literature) lead to what type of bias; show how pooling for score-based likelihood-ratio leads to bias; and discuss score-based strategies for reducing the dimension of the parameter space and strategies for mitigating the bias against the specified source that arises from pooling information in a score-based approximation to fair values of evidence.

Note: Joint work with Dr. Peter Vergeer of the Netherlands Forensic Institute.

Poster Session

Poster P1

The Optimal Standardization Factor for the Variance of the Normal Distribution: A Historical Perspective and Recommendations for Practice

Dominic Edelmann

German Cancer Research Center (DKFZ), Heidelberg · dominic.edelmann@dkfz-heidelberg.de

Numerous variance estimators exist for the normal distribution, yet the unbiased estimator — commonly known as Bessel's correction — remains the most widely used in practice. This article reviews various alternative estimators, that result from using different standardization factors for the sum of squares. It is demonstrated that there is no compelling reason to favor the unbiased variance estimator over other estimators and that better alternatives are available. Implications of these findings for applications involving small sample sizes are discussed.

Poster P2

Generalized Fiducial Inference for Poisson Point Processes

Hee-Geon Kang

Ewha Womans University · hgkang28@ewha.ac.kr

Poisson point processes are widely used for modeling spatial point patterns. Frequentist and Bayesian inference for these models are well established. However, frequentist uncertainty quantification often relies on large-sample approximations, while Bayesian inference requires prior specification. In this paper, we propose a generalized fiducial inference method for Poisson point processes on bounded observation regions. In our proposal, the fiducial distribution is constructed by exploiting the natural decomposition of a Poisson point process into the total count and the conditional distribution of point locations given the count. The count component is obtained from the generalized fiducial distribution for a univariate Poisson model, and the location component is constructed through a Rosenblatt-type data-generating equation. For homogeneous Poisson point processes, the proposed construction reduces to the classical fiducial distribution for a Poisson mean, which depends only on the total count. For inhomogeneous Poisson point processes, it gives an explicit fiducial density with an L2-Jacobian term. We show that the proposed fiducial density has a local asymptotic normality property under increasing-domain asymptotics, and that it has the same first-order local behavior as the Poisson likelihood when the Jacobian term is locally negligible. The performance of the proposed method is compared numerically with maximum likelihood inference and Bayesian inference. The numerical results show that the proposed method gives stable interval estimates and empirical coverage close to nominal levels, and provides a prior-free alternative for uncertainty quantification in inhomogeneous Poisson point process models.

Poster P3

Finite Sample Inference for Probit Regression

Adriana Gonzalez Sanchez

University of Cincinnati · gonzaai@mail.uc.edu

In this paper, we present a new inference method for constructing confidence intervals for the coefficients of logistic and probit regression models. The proposed approach provides finite sample inference and therefore fills an important gap in the literature by enabling inference for binary regression models when sample sizes are small and likelihood-based methods fail. The proposed approach is developed under the repro samples framework, where statistical inference is carried out by generating and analyzing artificial samples that mimic the sampling mechanism of the observed data. Unlike existing approaches, we formulate the construction of confidence intervals as a linear optimization problem, which allows us to efficiently search for and characterize the range of coefficient values that are consistent with the data. This optimization-based perspective not only provides valid confidence intervals with strong finite-sample guarantees but also offers a principled way to incorporate repro samples into coefficient-level inference while overcoming computational challenges typically associated with constructing exact confidence intervals.

Poster P4

A Quasi-Bayesian Shrinkage Estimator for Adaptive Cluster Sampling in Rare and Hard-to-Reach Populations

Nouman Qureshi

University of Minnesota · qures089@umn.edu

Rare and clustered populations are difficult to estimate because standard sampling methods often produce large variance, unstable weights, and poor coverage when the population is sparsely distributed or concentrated in hidden groups. This study proposes a quasi-Bayesian shrinkage estimator for adaptive cluster sampling designed for rare, hard-to-reach, and highly clustered populations. The method combines adaptive sampling information with shrinkage principles to borrow strength across sampled networks while controlling the instability that commonly appears in classical estimators. Unlike fully Bayesian models that require strict prior specification, the proposed quasi-Bayesian framework uses flexible pseudo-prior structures to stabilize estimation without heavy dependence on strong distributional assumptions. The estimator shrinks extreme network-based estimates toward structurally informed targets, improving precision while preserving sensitivity to local clustering patterns. This approach is especially relevant for studies involving hidden human populations, rare disease cases, endangered species, and socially marginalized groups where traditional estimators may fail because of limited initial captures and unequal cluster expansion. Theoretical properties of the proposed estimator are developed under rare population settings, and its performance is examined through simulation under varying levels of clustering intensity, prevalence, and network growth. Results show consistent gains in efficiency, reduced mean squared error, and narrower interval estimates compared with conventional adaptive cluster sampling estimators and existing design-based methods. By integrating shrinkage with adaptive designs, the proposed framework offers a practical and computationally accessible strategy for uncertainty quantification in complex sampling environments. This work extends the role of quasi-Bayesian methods in survey statistics and provides a useful tool for modern applications where reliable estimation of hidden populations remains a central challenge.

Poster P5

Nonparametric analysis of right censored data via the Mann-Whitney effect and overlap index

Patrick Langthaler

University of Salzburg · patrick.langthaler@plus.ac.at

A staple in nonparametric statistics is the Mann-Whitney effect. Similar functionals, designed not only to capture location but also scale effects, have subsequently been developed. Here we present an extension of a statistical estimation and testing framework based on the Mann-Whitney effect and a so-called overlap index to the setting where data are right censored. The interpretation and performance of the procedure for the commonly assumed proportional hazards setting, as well as other scenarios, including crossing survival curves, are discussed and its performance compared with competitors is assessed in a simulation study and demonstrated in a real data example from cancer research.

Poster P6

Extending the generalized Rosenblatt Transform

Christian Covington
christian.t.covington@gmail.com

The Rosenblatt transform is a widely used tool for converting samples from a multivariate distribution on R^n into independent uniform variables, enabling model checking through tests of product uniformity. Its classical formulation requires absolute continuity, excluding many common classes of models. The generalized Rosenblatt transform (GRT) removes this restriction by introducing auxiliary randomness, but existing elementary treatments leave its existence, uniqueness, and ordering dependence insufficiently formalized. We develop a measure-theoretic theory of the GRT on countable products of standard Borel spaces, substantially broadening its applicability. We prove that our extended GRT, after fixing orders on each component-space and the index set, is the almost-everywhere unique monotone triangular uniformizing map, mirroring the existing uniqueness result for the classical Rosenblatt transform. We then analyze the role of ordering choices, proving invariance when the model is correct and quantitative sensitivity bounds under misspecification. Finally, we show that this ordering dependence can be used constructively as a diagnostic tool for model criticism: failures of uniformity across valid orderings can detect misspecification and localize it to smaller submodels.

Poster P7

Robust Sequential Bernoulli Updating under Latent External Shocks: An Extended Fiducial Perspective

Francesco De Pretis

University of Modena and Reggio Emilia · francesco.depretis@unimore.it

Sequential decision systems are often evaluated under the simplifying assumption that the data-generating process is stationary. In many applications, however, streaming binary outcomes may be affected by unannounced external perturbations, producing transient changes in the underlying success probability. We study this problem in a sequential Bernoulli setting inspired by the Kyburg coin-toss game and recent imprecise-probability comparisons of interval-valued decision rules, where decisions are made from probability intervals after partial observation of the data stream. We compare four uncertainty-updating paradigms — imprecise Bayesian updating, Clopper–Pearson frequentist intervals, Dempster–Shafer belief functions, and Extended Fiducial Inference — against a standard Bayesian benchmark. The data-generating process includes a single latent shock of unknown timing and duration, with shock lengths corresponding to 10% and 30% of the sample path. Across Monte Carlo experiments, decision performance is evaluated under Hurwicz, maximin, minimax-regret, and maximum-entropy criteria. The results show that methods relying on cumulative stationary counts can suffer substantial coverage distortion after regime changes, with consequences for downstream decisions. Extended Fiducial Inference provides a more robust alternative by treating the latent change-point structure as part of the inferential problem rather than as an ignored nuisance, thereby connecting fiducial reasoning with uncertainty quantification for non-stationary data streams. In the simulated scenarios, this yields improved coverage stability and stronger decision performance, especially during and after shock transitions. The poster highlights the relevance of fiducial reasoning for uncertainty quantification in non-stationary automated decision systems, including online experimentation, reliability monitoring, and adaptive clinical or operational decision-making.

Poster P8

Posterior uncertainty for kernel density estimates

Dennis Christensen

dennischristensen95@gmail.com

De Finetti's theorem says that any exchangeable sequence $Y_1, Y_2, \ldots$ of random variables corresponds to a Bayesian model in the sense that there exists a random measure P such that, conditionally on P, the Y_i are i.i.d. In predictive Bayesian inference, we study this correspondence more generally by working directly with a sequence of predictive distributions $\alpha_n(\cdot) = \Pr(Y_{n+1} \in \cdot \mid Y_1, \ldots, Y_n)$. A key question is whether (α_n) converges weakly almost surely. When it does, there exists a random directing measure P such that the sequence (Y_n) is P-asymptotically exchangeable. That is, $(Y_{n+1}, Y_{n+2}, \ldots)$ converges in distribution to $(Z_1, Z_2, \ldots)$ as $n \to \infty$, where (Z_n) is an exchangeable sequence with directing measure P. In other words, when (α_n) converges, then the Y_i are asymptotically i.i.d. conditionally on P. In this work we show that data sequences generated by kernel density estimation are asymptotically exchangeable. Because standard tools from the predictive Bayes literature do not apply in this setting, we prove weak convergence of the predictive distributions directly. Further, for Gaussian kernels we show that the directing measure is absolutely continuous with respect to the Lebesgue measure. This allows us to derive credibility intervals to assess posterior uncertainty of the kernel density estimates.

Poster P9

Nonparametric Estimation and Convergence Properties of Archimax Copulas

Nicolas Pascal Dietrich

University of Salzburg · nicolaspascal.dietrich@plus.ac.at

Archimax copulas constitute a highly flexible class of dependence models, capable of capturing both tail dependence and dependence at intermediate levels. They are characterized by an Archimedean generator and a Pickands dependence function. We first address the question of identifiability within the Archimax class. We then investigate convergence properties of Archimax copulas and show that weak convergence of a sequence of Archimax copulas is equivalent to weak convergence of almost all associated Markov kernels (conditional distributions). Building on these theoretical results, we introduce a novel nonparametric estimator for Archimax copulas, combining an estimator of the Kendall distribution function with a CFG-type estimator. Under mild regularity assumptions, we establish strong consistency of the proposed estimator and evaluate its finite-sample performance using weather data from Austria.

Poster P10

Change Point Detection Inference with Unknown Number of Change Points Using Repro-Samples

Junyi Li

Rutgers University · jl2707@stat.rutgers.edu

We develop a novel inference approach for change-point detection and uncertainty quantification when the number of change points is unknown for uni- and multivariate data. Our approach is developed based on the repro-samples framework, which constructs inference procedures through artificially generated samples or noise that mimic the observed data. When the data-generating distribution belongs to an exponential family, the proposed method achieves exact finite-sample coverage by conditioning on sufficient statistics. When the form of the distribution is unknown, the same approach remains applicable and yields asymptotically valid inference. To address the uncertainty in the unknown number of change points and to reduce the search space, we propose a candidate set construction approach using Fisher's inversion technique. We show that even under weak signal strength conditions, the candidate set constructed can contain the true number of change points with high probability. Overall, the proposed approach provides a unified and flexible framework for valid inference in change-point detection while accounting for uncertainty in the number of change points. Simulation studies and real data analyses on both univariate and multivariate datasets are used to demonstrate the effectiveness of the proposed method, including applications to human activity data and sequential image data such as MNIST and CIFAR, where the goal is to detect changes in the underlying state or class over time.

Poster P11

Nonparametric Estimation of Conditional Distribution Functions via Empirical Checkerboard Copula Approximations

Kai Schärer

University of Salzburg · kai.schaerer@plus.ac.at

This work introduces a computationally efficient nonparametric framework for estimating conditional distribution functions by leveraging Sklar's Theorem, providing a robust foundation for distribution-free inference. By separately estimating the univariate marginal distributions and the conditional distribution of the underlying copula, we provide a flexible estimation framework. Our primary theoretical contribution establishes that empirical checkerboard copula approximations converge uniformly conditional to the true underlying copula under an appropriately scaled resolution and mild continuity assumptions. Crucially, we prove a new density result showing that the class of copulas admitting these continuous Markov kernels is dense within the space of all bivariate copulas under the D_1 metric. Compared to standard kernel-based regression techniques (such as the Nadaraya–Watson estimator), our approach provides a unified framework for a broad class of nonparametric conditional functional regressions. Specifically, we establish strong consistency for the conditional mean, quantile, and expectile functions simultaneously. We complement these theoretical results with an extensive simulation study and a real-world insurance data experiment. Finally, we provide empirical benchmarks demonstrating how our computational complexity scales efficiently with both large sample sizes and increasing numbers of evaluation points.