AI SSAI

AI for science, science for AI

CNRS

When Causal Inference Meets Statistical Analysis

April 17-21, 2023

National Conservatory of Arts and Crafts (CNAM), Paris, France

Contents

About

> “To find out what happens to a system when you interfere with it, you have to interfere with it.” – George E. P. Box - Use and Abuse of Regression

A Thematic Quarter on Causality

This thematic quarter aims to underline the interactions between Statistical, Probability Theory, Causal Inference theory, and theoretical computer science methods for causal inference.

When Causal Inference meets Statistical Analysis

When Causal Inference Meets Statistical Analysis conference aims to enable researchers, students, and practitioners to explore the rich and cross-disciplinary relationship between causal inference and statistical analysis and highlight novel approaches related to the challenges of causal inference, causal discovery, and causal analysis. The conference will explore topics related to, but not limited to:

- Causal discovery
- Causal learning and control problems
- Theoretical foundation of causal inference
- Causal inference and active learning
- Causal learning in low data regime
- Reinforcement learning
- Causal machine learning
- Causal generative models
- Benchmark for causal discovery and causal reasoning

Organizing committee

Marianne Clausel	University of Lorraine
Alessandro Leite	TAU, INRIA Saclay, Paris-Saclay University
Georges Oppenheim	Paris-Saclay University

Timetable

CT: Contributed Talk, KL: Keynote Lecture, IT: Invited Talk.

Monday, April 17th

9:00–9:30	**Arrival & welcome coffee**		
9:30–10:00	**Welcome & Opening remarks**		
10:00–11:00	KL	**Bin Yu** University of California, Berkeley	Veridical data science with a case study to seek genetic drivers of a heart disease
11:00–11:30	**Coffee break**		
11:30–12:00	CT	**Víctor Elvira** University of Edinburgh, UK	Graphs in State-Space Models for Granger Causality in Climate Science
12:00–12:30	CT	**Mouad El Bouchattaoui** Paris-Saclay University, France	CDVAE: Estimating causal effects over time under unobserved adjustment variables
12:30–14:00	**Lunch**		
14:00–15:00	KL	**Eric Gaussier** University of Grenoble Alpes, France	Causal discovery and inference in time series
15:00–15:30	**Coffee break**		
15:30–16:00	**Discussion group**		
16:00–16:30	IT	**Martini Cinquini** University of Pisa, Italy	Boosting Synthetic Data Generation with Effective Nonlinear Causal Discovery
16:30-17:00	CT	**Benjamin Heymann** Criteo AI Lab, France	Causal Inference with Information Fields
17:00-17:30	IT	**Kirtan Padh** Helmholtz AI, Germany	Stochastic Causal Programming for Bounding Treatment Effects
17:30–18:00	**Discussion group**		
18:00–19:30	**Conference reception**		

Tuesday, April 18th

9:00–9:30	**Arrival & welcome coffee**		
9:30-10:30	KL	**Yingzhen Li** Imperial College London, England	Towards Causal Deep Generative Models for Sequential Data
10:30–11:00	**Coffee**		
11:00–11:30	**Discussion group**		
11:30–12:00	IT	**Atalanti Mastakouri** Amazon Research Tuebingen, Germany	Causal feature selection in time series data
12:00–12:30	CT	**Antonin Arsac** Paris-Saclay University & CEA List, France	Causal discovery for time series with constraint-based model and PMIME measure
12:30–14:00	**Lunch**		
14:00–15:00	KL	**Daniel Malinsky** Columbia University, USA	A Cautious Approach To Constraint-Based Causal Model Selection Based on Equivalence Tests
15:00–15:30	**Coffee break**		
15:30–16:00	**Discussion group**		
16:00–16:30	CT	**Anouar Meynaoui** Rennes University, France	Identifiability in time series extended summary causal graphs
16:30–17:00	IT	**Pan Zhao** PreMeDICaL, Inria-Inserm, France	Efficient and robust transfer learning of optimal individualized treatment regimes with right-censored survival data
17:00-17:30	IT	**Shiyang Yan** INRIA Saclay, France	Out-of-distribution from a causal perspective: from domain generalization to OOD problems in ocean plankton analysis
17:30–18:00	CT	**Martin Spindler** University of Hamburg	Uniform Inference in High-Dimensional Generalized Additive Models

Wednesday, April 19th

9:00–9:30	**Arrival & welcome coffee**		
9:30-10:30	KL	**Dominik Janzing** Amazon Research, Germany	Causal insights from merging data sets and merging data sets via causal insights
10:30–11:00	**Coffee**		
11:00–11:30	**Discussion group**		
11:30–12:00	CT	**Simon Ferreira** ENS de Lyon & EasyVista, France	Challenges of Root Cause Identification for Collective Anomalies in Time Series given a Summary Causal Graph
12:00–12:30	IT	**Yuchen Zhu** Yale University, USA	Relaxing Observability Assumption in Causal Inference with Kernel Methods
12:30–14:00	**Lunch**		
14:00–15:00	KL	**Chandler Squires** MIT, USA	Beyond ICA: Causal Disentanglement via Interventions
15:00–16:00	**Poster session**		
16:30–18:00	**Social event: Visit the Museum of Arts and Crafts**		

Thursday, April 20th

9:00–9:30	**Arrival & welcome coffee**		
9:30-10:30	KL	**Jonas Peters** ETH Zurich, Switzerland	Statistical Testing under Distributional Shifts and its Application to Causality
10:30–11:00	**Coffee**		
11:00–11:30	**Discussion group**		
11:30–12:00	IT	**Junhyung Park** Max Planck Institute for Intelligent Systems, Germany	Towards a Measure-Theoretic Axiomatisation of Causality
12:00–12:30	IT	**Alexander G. Reisach** Université Paris Cité, France	Simple Sorting Criteria Help Find the Causal Order in Additive Noise Models
12:30–14:00	**Lunch**		
14:00–15:00	KL	**Jason Hartford** Recursion Pharmaceutical, USA	Toward causal inference from high-dimensional observations
15:00–15:30	IT	**Sorawit Saengkyongam** University of Copenhagen, Denmark	Effect-Invariant Mechanisms for Policy Generalization
15:30–16:30	KL	**Krikamol Muandet** CISPA Helmholtz Center for Information Security, Germany	Reliable Machine Learning with Instruments
16:30–18:00	**Poster session**		
19:30–22:30	**Conference dinner at Le Procope restaurant**		

Friday, April 21st

9:00–9:30	**Arrival & welcome coffee**		
9:30-10:30	KL	**Antoine Chambaz** Université Paris Cité, France	Learning, evaluating and analyzing a recommendation rule for early blood transfer in the ICU
10:30–11:00	**Coffee**		
11:00–11:30	**Discussion group**		
11:30–12:00	IT	**Lucas de Lara** Toulouse Mathematics Institute, France	When Causality Meets Optimal Transport
12:00–12:30	IT	**Hugo Henri Joseph Senetaire** Technical University of Denmark, Denmark	Explainability as statistical inference
12:30–14:00	**Lunch**		
14:00–15:00	KL	**Elina Robeva** University of British Columbia, Canada	Learning Linear Non-Gaussian Causal Models via Higher Moment Relationships
15:00–15:30	CT	**Claire Theobald** Lorraine University, France	Clarity: an improved gradient method for producing quality visual counterfactual explanations
15:30–16:00	IT	**Carles Balsells Rodas** Imperial College London, England	Causal discovery from conditionally stationary time series
16:00–16:30	IT	**Pardis Semnani** University of British Columbia, Canada	Causal Inference in Directed, Possibly Cyclic, Graphical Models
16:30–17:30	**Wrap-up & cocktail**		

Keynote Lectures

Learning, evaluating and analyzing a recommendation rule for early blood transfer in the ICU

Antoine Chambaz KL

Paris Cité University, Paris, France

Severely injured patients experiencing hemorrhagic shock often require massive transfusion. Early transfusion of blood products (plasma, platelets and red blood cells) is common and associated with improved outcomes in the hospital. However, determining a right amount of blood products is still a matter of scientific debate. The speaker will present and discuss a methodology to learn, evaluate and analyse a recommendation rule for early blood transfer in the ICU. The study uses data from the French Traumabase, a French observatory for Major Trauma. This is a joint work with Pan Zhao (Inria), Julie Josse (Inria), Nicolas Gatulle (APHP) and the Traumabase Group.

Causal discovery and inference in time series

Eric Gaussier KL

University of Grenoble Alpes, Grenoble, France

Time series are present in various forms in many different domains, as healthcare, Industry 4.0, monitoring systems or energy management to name but a few. Despite the presence of temporal information, discovering causal relations and identifying interventions in time series are not easy tasks, in particular when one is dealing with "incomplete" graphs. We will first discuss which causal graphs are relevant to time series and present methods to extract them from observational data and identify interventions in them.

Toward causal inference from high-dimensional observations

Jason Hartford KL

Recursion Pharmaceutical, Utah, USA

Over the last decade, biologists have developed a multitude of tools with which we can generate high-dimensional observations of biological systems. These tools either directly measure causal variables (e.g. the expression levels of genes) or they measure unstructured proxies for the causal variables (e.g. the pixels in an image of a cell). When these high-dimensional causal variables are directly measured, the relationship with outcomes of interest typically remains confounded so we need to rely on experimentation to identify causal effects. The challenge is (1) we often cannot intervene directly on the causal variables, and instead perturb them indirectly, and (2) the number of experiments that we can practically run is typically far fewer than the number of causal variables of interest. I will discuss how we can use instrumental variable methods to estimate causal effects in this regime, and show when we can estimate the causal effect with relatively few experiments despite being underspecified. The second part of my talk will focus on representation learning which is needed when we measure unstructured proxies for the causal variables. These modalities, such as images or sensor data can often be collected cheaply in experiments, but they are challenging to use in a causal inference pipeline without extensive feature engineering or labelling to extract underlying latent factors. I will present results from a series of recent papers that describe when we can disentangle latent variables with identifiability guarantees.

Causal insights from merging data sets and merging data sets via causal insights

Dominik Janzing KL

Amazon Research, Germany

While humans often draw causal conclusions from putting observations into the broader context of causal knowledge, AI still needs to develop these techniques. I show how causal insights can be obtained from the synergy of datasets referring to different sets of variables and argue that causal hypotheses then predict joint properties of variables that have never been observed together. This way, causal discovery becomes a prediction task in which additional variable sets play the role of additional data points in traditional iid learning. For instance, a causal DAG can be seen as a binary classifier that tells us which conditional independences are valid, which then enables a statistical learning theory for learning DAGs. I describe "Causal MaxEnt" (a modified version of MaxEnt that is asymmetric with respect to causal directions) as one potential approach to infer DAGs and properties of the joint distribution from a set of marginal distributions of subsets of variables and derive causal conclusions for toy examples.

Further reading

[1] Dominik Janzing: Merging joint distributions via causal model classes with low VC dimension, arxiv:1804.03206.
[2] Sergio Garrido Mejia, Elke Kirschbaum, Dominik Janzing: Obtaining causal information by merging data sets with MaxEnt. AISTATS 2022.
[3] Dominik Janzing: Causal versions of Maximum Entropy and Principle of Insufficient Reason, Journal of Causal Inference 2021.

Towards Causal Deep Generative Models for Sequential Data

Yingzhen Li

KL

Imperial College London, London, England

One of my research dreams is to build a high-resolution video generation model that enables granularity controls in e.g., the scene appearance and the interactions between objects. I tried, and then realised the need of me inventing deep learning tricks for this goal is due to the issue of non-identifiability in my sequential deep generative models. In this talk I will discuss our research towards developing identifiable deep generative models in sequence modelling, and share some recent and on-going works regarding switching dynamic models. Throughout the talk I will highlight the balance between causality "Theorist" and deep learning "Alchemist", and discuss my opinions on the future of causal deep generative modelling research.

A Cautious Approach To Constraint-Based Causal Model Selection Based on Equivalence Tests

Daniel Malinsky

KL

Columbia University, New York City, USA

Causal graphical models are used many scientific domains to represent important causal assumptions about the processes that underlie collected data. The focus of this work is on graphical structure learning (a.k.a. causal discovery or model selection) for the "downstream" purpose of using the estimated graph for subsequent causal inference tasks, such as establishing the identifying formula for some causal effect of interest and then estimating it. An obstacle to having confidence in existing procedures in applied health science settings is that they tend to estimate structures that are overly sparse, i.e., missing too many edges. However, statistical "caution" (or "conservativism") would err on the side of more dense graphs rather than more sparse graphs. This paper proposes to reformulate the conditional independence hypothesis tests of classical constraint-based algorithms as equivalence tests: test the null hypothesis of association greater than some (user-chosen, sample-size dependent) threshold, rather than test the null of no association. We argue this addresses several important statistical issues in applied causal model selection and leads to procedures with desirable behaviors and properties.

Reliable Machine Learning with Instruments

Krikamol Muandet KL

CISPA – Helmholtz Center for Information Security, Germany

Society is made up of a set of diverse individuals, demographic groups, and institutions. Learning and deploying algorithmic models across these heterogeneous environments face a set of various trade-offs. In order to develop reliable machine learning algorithms that can interact successfully with the real world, it is necessary to deal with such heterogeneity. In this talk, I will focus on how to employ an instrumental variable (IV) to alleviate the impact of unobserved confounders on the credibility of algorithmic decision-making and the reliability of machine learning models that are learned from observational and heterogeneous data. In particular, I will present how we can leverage tools from machine learning, namely, kernel methods and deep learning, to solve potentially ill-posed non-linear IV regression and proxy variable problems. Lastly, I will argue that a better understanding of the ways in which our data are generated and how our models can influence them will be crucial for reliable human-machine interactions, especially when gaining full information about data may not be possible.

Statistical Testing under Distributional Shifts and its Application to Causality

Jonas Peters KL

ETH Zurich, Zurich, Switzerland

We discuss the problem of statistical testing under distributional shifts. In such problems, we are interested in the hypothesis P in H0 for a target distribution P, but observe data from a different distribution Q. We assume that P is related to Q through a known shift tau and formally introduce hypothesis testing in this setting. We propose a general testing procedure that first resamples from the observed data to construct an auxiliary data set and then applies an existing test in the target domain. This proposal comes with theoretical guarantees. Testing under distributional shifts allows us to tackle a diverse set of problems. We argue that it may prove useful in reinforcement learning and covariate shift, we show how it reduces conditional to unconditional independence testing and we provide example applications in causal inference. This is joint work with Nikolaj Thams, Sorawit Saenkyongam, and Niklas Pfister.

Learning Linear Non-Gaussian Causal Models via Higher Moment Relationships

Elina Robeva KL

University of British Columbia, Vancouver, Canada

In this talk we will discuss the problem of learning the directed graph for a linear non-Gaussian causal model. We will specifically address the cases where the graph may have cycles or there might be hidden variables. While ICA methods are able to recover the correct graph, they do not always guarantee to have found the optimal solution. Our methods are based on using specific relationships that hold among the 2nd and 3rd moments of the random vector and can help us characterize the graph. While in the acyclic case, the directed graph can be found uniquely, when cycles are allowed the graph can be learned up to an equivalence class. We give a description of all the graphs that each equivalence class consists of. We then give an algorithm based on relationships among the second and third order moments of the random vector that recovers the equivalence class of the graph, assuming the graph lies in a specific family of cyclic graphs. This is joint work in progress with Mathias Drton, Marina Garrote-Lopez, and Niko Nikov.

Beyond ICA: Causal Disentanglement via Interventions

Chandler Squires KL

MIT, Massachusetts, USA

Traditional methods for representation learning, such as independent component analysis (ICA), seek representations of observed data in terms of a set of independent variables. However, human reasoning relies on variables which are not statistically independent, such as height and weight. The emerging field of causal representation learning takes the view that the variables in a representation should not be statistically independent, but rather that they should be causally autonomous. From a generative modeling perspective, we may view learning such a representation as an inference problem, which we call "causal disentanglement". In this talk, I will outline a research program which aims to build a computational and statistical theory of causal disentanglement. I will discuss initial results in this program, including our own work proving identifiability of causal representations from interventions in a linear setting. I will conclude with an overview of future applications for these methods in biology, healthcare, and human-AI interaction.

Veridical data science with a case study to seek genetic drivers of a heart disease

Bin Yu

KL

University of California, Berkeley, USA

“AI is like nuclear energy–both promising and dangerous.” – Bill Gates, 2019.

Data Science is a pillar of AI and has driven most of recent cutting-edge discoveries in biomedical research and beyond. Human judgment calls are ubiquitous at every step of a data science life cycle, e.g., in problem formulation, choosing data cleaning methods, predictive algorithms and data perturbations. Such judgment calls are often responsible for the "dangers" of AI. To mitigate these dangers, we introduce in this talk a framework based on three core principles: Predictability, Computability and Stability (PCS). The PCS framework unifies and expands on the ideas and best practices of statistics and machine learning. It emphasizes reality check through predictability and takes a full account of uncertainty sources in the whole data science life cycle including those from human judgment calls such as those in data curation/cleaning. PCS consists of a workflow and documentation and is supported by our software package veridical or v-flow. Moreover, we illustrate the usefulness of PCS in the development of iterative random forests (iRF) for predictable and stable non-linear interaction discovery (in collaboration with the Brown Lab at LBNL and Berkeley Statistics). Finally, in the pursuit of genetic drivers of a heart disease called hypertrophic cardiomyopathy (HCM) as a CZ Biohub project in collaboration with the Ashley Lab at Stanford Medical School and others, we use iRF and UK Biobank data to recommend gene-gene interaction targets for knock-down experiments. We then analyze the experimental data to show promising findings about genetic drivers for HCM.

Invited and Contributed Talks

Boosting Synthetic Data Generation with Effective Nonlinear Causal Discovery

***Martina Cinquini*[1], *Fosca Giannotti*[2], *and Riccardo Guidotti*[1]** IT

[1] University of Pisa
[2] ISTI-CNR

Synthetic data generation has been widely adopted in software testing, data privacy, imbalanced learning, and artificial intelligence explanation. In all such contexts, it is crucial to generate plausible data samples. A common assumption of approaches widely used for data generation is the independence of the features. However, typically, the variables of a dataset depend on one another, and these dependencies are not considered in data generation leading to the creation of implausible records. The main problem is that dependencies among variables are typically unknown. In this paper, we design a synthetic dataset generator for tabular data that can discover nonlinear causalities among the variables and use them at generation time. State-of-the-art methods for nonlinear causal discovery are typically inefficient. We boost them by restricting the causal discovery among the features appearing in the frequent patterns efficiently retrieved by a pattern mining algorithm. We design a framework for generating synthetic datasets with known causalities to validate our proposal. Broad experimentation on many synthetic and real datasets with known causalities shows the effectiveness of the proposed method.

Causal discovery from conditionally stationary time series

Carles Balsells Rodas IT

Imperial College London, England

Causal discovery, i.e., inferring underlying causal relationships from observational data, has been shown to be highly challenging for AI systems. In time series modeling context, traditional causal discovery methods mainly consider constrained scenarios with fully observed variables and/or data from stationary time series. We develop a causal discovery approach to handle a wide class of non-stationary time series that are conditionally stationary, where the non-stationary behaviour is modeled as stationarity conditioned on a set of (possibly hidden) state variables. Named State-Dependent Causal Inference (SDCI), our approach is able to recover the underlying causal dependencies, provably with fully-observed states and empirically with hidden states. The latter is confirmed by experiments on synthetic linear system and nonlinear particle interaction data, where SDCI achieves superior performance over baseline causal discovery methods. Improved results over non-causal RNNs on modeling NBA player movements demonstrate the potential of our method and motivate the use of causality-driven methods for forecasting.

Causal discovery for time series with constraint-based model and PMIME measure

Antonin Arsac, Aurore Lomet, and Jean-Philippe Poli **CT**

Paris-Saclay University, CEA List, Palaiseau, France

Causality defines the relationship between cause and effect. In multivariate time series field, this notion allows to characterize the links between several time series considering temporal lags. These phenomena are particularly important in medicine to analyze the effect of a drug for example, in manufacturing to detect the causes of an anomaly in a complex system or in social sciences... Most of the time, these complex systems do not rely on the assumption of linearity required in many machine learning methods. To circumvent this problem, we present in this paper a new approach for discovering causality in time series data that combines a causal discovery algorithm with an information theoretic-based measure. Hence the proposed method allows inferring both linear and non- linear relationships and building the underlying causal graph. We evaluate the performance of our approach on several simulated datasets, showing promising results.

Causal feature selection in time series data

Atalanti Mastakouri **IT**

Amazon Research Tuebingen, Germany

The knowledge of the causal graph that described the variables of a system is of major importance in several domains, such as biology, earth science and finance both for explainability, as well as, decision making reasons. Nevertheless, in the absence of randomized controlled trials and sufficient domain knowledge, causal discovery of the full-time graph in time series settings with latent confounders is a rather challenging problem without a solution so far. In such scenarios we suggest to focus on a smaller yet significant task, that of identification of causal features of a target node, not aiming to recover the relationships between the candidate causes themselves, yet not constraining them. In this talk, I will present a conditional independence based method which, by treating the time series as discreet time stamps of a Markov chain, and by assuming that the target cannot cause the potential causes, allows us to identify the set of true unconfounded causes of a target time series. I will explain the fundamentals of the method and provide results from application on two real datasets.

Causal Inference in Directed, Possibly Cyclic, Graphical Models

Pardis Semnani **IT**

University of British Columbia, Vancouver, Canada

In this talk, we consider the problem of learning a directed graph Markov equivalent to an unknown directed graph G from observational data sampled from a distribution in the graphical model of G. We assume that the distribution is faithful to the graph G and that there are no unobserved variables, but do not rely on any further assumptions regarding the graph or the distribution of the variables. In particular, we allow for directed cycles in G and work in the fully non-parametric setting. Given the set of conditional independence statements satisfied by the distribution, we aim to find a directed graph which satisfies the same d-separation statements as G. We propose a hybrid approach consisting of two steps. We first find a partially ordered partition of the vertices of G by optimizing a certain score in a greedy fashion. We prove that any optimal partition uniquely characterizes the Markov equivalence class of G. Given an optimal partition, we propose an algorithm for constructing a graph in the Markov equivalence class of G whose strongly connected components correspond to the elements of the partition, and which are partially ordered according to the partial order of the partition. Our algorithm comes in two versions – one which is provably correct and another one which performs fast in practice. This is joint work with Elina Robeva.

Causal Inference with Information Fields

***Benjamin Heymann*[1]*, Michel De Lara*[2]*, and Jean-Philippe Chancelier*[1]** **CT**

[1] Criteo AI Lab, Paris, France
[2] École des Ponts, CERMICS Marne-la-Vallée, France

Inferring the potential consequences of an unobserved event is a fundamental scientific question. To this end, Pearl's celebrated do-calculus provides a set of inference rules to derive an interventional probability from an observational one. In this framework, the primitive causal relations are encoded as functional dependencies in a Structural Causal Model (SCM), which maps into a Directed Acyclic Graph (DAG) in the absence of cycles. In this paper, by contrast, we capture causality without reference to graphs or functional dependencies, but with information fields. The three rules of do-calculus reduce to a unique sufficient condition for conditional independence: the topological separation, which presents some theoretical and practical advantages over the d-separation. With this unique rule, we can deal with systems that cannot be represented with DAGs, for instance systems with cycles and/or 'spurious' edges.

CDVAE: Estimating causal effects over time under unobserved adjustment variables

***Mouad El Bouchattaoui*[1], *Myriam Tami*[1], *Benoit Lepetit*[2], *and Paul-Henry Cournède*[2]** **CT**

[1] Paris-Saclay University, CentraleSupélec, MICS Lab, France
[2] $Saint - Gobain, France$

Estimating treatment effects over time is relevant in many real-world appli- cations nowadays (precision medicine, epidemiology, economy, marketing). Many state-of-the-art methods either assume unconfoundedness or seek to infer unobserved confounders. However, in this work, we aim to predict the causal quantities and the response of interest, allowing more precise and individual response trajectories. To this end, we assume the existence of unobserved risk factors, i.e., adjustment variables that affect only the out- comes. We introduce CDVAE, a model developed using Dynamic Variational Autoencoders (DVAE), which infers the unobserved sources of heterogeneity and adjusts for them in the outcome model. In a longitudinal study, we apply a counterfactual regression to estimate Heterogeneous Treatment Effects (HTE) by balancing the reconstruction term in the Evidence Lower Bound (ELBO) of the DVAE with a propensity score function. We show that the trade-off of estimating well both the factual outcome and HTEs is improved further by balancing the Kullback-Leibler regularization term by a hyper-parameter β, scheduled to vary in cycles. Evaluation of CDVAE on synthetic datasets achieves superior performance over state-of-the-art methods.

Challenges of Root Cause Identification for Collective Anomalies in Time Series given a Summary Causal Graph

***Simon Ferreira*[1,2] *and Charles K. Assaad*[1]** **CT**

[1] ENS de Lyon, Lyon, France
[2] EasyVista, Lyon, France

This paper discusses the challenges of identifying the root causes of collective anomalies given observational time series and a summary causal graph which depicts an abstraction of causal relations present in a dynamic system at its normal regime. The paper first depicts the three main steps of the EasyRCA algorithm which can identify the root causes under the assumption that the summary causal graph is acyclic but loops on vertices are allowed. Furthermore, it gives some insight on difficulties when the acyclicity assumption is relaxed by discussing the applicability and inapplicability of the EasyRCA algorithm in case of general summary causal graphs.

Clarity: an improved gradient method for producing quality visual counterfactual explanations

***Claire Theobald*[1], *Frédéric Pennerath*[2], *Brieuc Conan-Guez*[2], *Miguel Couceiro*[1], and Amedeo Napoli*[1]**

CT

[1] Lorraine University, CNRS, LORIA, F-54000 Nancy France
[2] CentraleSupélec, CNRS, LORIA, F-57000 Metz France

Visual counterfactual explanations identify modifications to an image that would change the prediction of a classifier. We propose a set of techniques based on generative models (VAE) and a classifier ensemble directly trained in the latent space, which all together, improve the quality of the gradient required to compute visual counterfactuals. These improvements lead to a novel classification model, Clarity, which produces realistic counterfactual explanations over all images. We also present several experiments that give insights on why these techniques lead to better quality results than those in the literature. The explanations produced are competitive with the state-of-the-art and emphasize the importance of selecting a meaningful input space for training.

Effect-Invariant Mechanisms for Policy Generalization

Sorawit Saengkyongam

IT

University of Copenhagen, Denmark

Policy learning is a crucial component of many interactive learning systems such as personalized healthcare. A major challenge in policy learning is how to adapt effectively to unseen environments or tasks. Recently, it has been suggested to exploit invariant conditional distributions to learn models that generalize better to unseen environments. However, assuming invariance of entire conditional distributions (to which we refer as full invariance) may be too strong an assumption in practice. In this paper, we introduce a relaxation of the full invariance called effect invariance (e-invariance) and prove that it is sufficient, under suitable assumptions, for zero-shot generalization. We also propose an approach to incorporate e-invariance when we have a sample from the test environment, enabling few-shot generalization. Finally, we present empirical results using a mobile health intervention dataset to demonstrate the effectiveness of our approach.

Efficient and robust transfer learning of optimal individualized treatment regimes with right-censored survival data

Pan Zhao[1], *Julie Josse*[1], *and Shu Yang*[2]

IT

[1]PreMeDICaL, Inria-Inserm, Montpellier, France
[2] Department of Statistics, North Carolina State University, US

An individualized treatment regime (ITR) is a decision rule that assigns treatments based on patients' characteristics. The value function of an ITR is the expected outcome in a counterfactual world had this ITR been implemented. Recently, there has been increasing interest in combining heterogeneous data sources, such as leveraging the complementary features of randomized controlled trial (RCT) data and a large observational study (OS). Usually, a covariate shift exists between the source and target population, rendering the source-optimal ITR unnecessarily optimal for the target population. We present an efficient and robust transfer learning framework for estimating the optimal ITR with right-censored survival data that generalizes well to the target population. The value function accommodates a broad class of functionals of survival distributions, including survival probabilities and restrictive mean survival times (RMSTs). We propose a doubly robust estimator of the value function, and the optimal ITR is learned by maximizing the value function within a pre-specified class of ITRs. We establish the $N^{-1/3}$ rate of convergence for the estimated parameter indexing the optimal ITR, and show that the proposed optimal value estimator is consistent and asymptotically normal even with flexible machine learning methods for nuisance parameter estimation. We evaluate the empirical performance of the proposed method by simulation studies and a real data application of sodium bicarbonate therapy for patients with severe metabolic acidaemia in the intensive care unit (ICU), combining a RCT and an observational study with heterogeneity.

Explainability as statistical inference

Hugo Henri Joseph Senetaire[1], *Damien Garreau*[2], *Jes Frellsen*[1], *Pierre-Alexandre Mattei*[2] IT

[1]Technical University of Denmark, Denmark
[2] Université Cote d'Azur, Inria, LJAD, UMR CNRS 7351, France

A wide variety of model explanation approaches have been proposed in recent years, all guided by very different rationales and heuristics. In this paper, we take a new route and cast interpretability as a statistical inference problem. We propose a general deep probabilistic model designed to produce interpretable predictions. The model parameters can be learned via maximum likelihood, and the method can be adapted to any predictor network architecture and any type of prediction problem. Our method is a case of amortized interpretability models, where a neural network is used as a selector to allow for fast interpretation at inference time. Several popular interpretability methods are shown to be particular cases of regularised maximum likelihood for our general model. We propose new datasets with ground truth selection which allow for the evaluation of the features importance map. Using these datasets, we show experimentally that using multiple imputation provides more reasonable interpretations.

Graphs in State-Space Models for Granger Causality in Climate Science

Víctor Elvira[1], *Émilie Chouzenoux*[2], *Jordi Cerdà*[3], *and Gustau Camps-Valls*[3] **CT**

[1]University of Edinburgh, UK
[2] Inria Saclay, France
[3] Universitat de València, Spain

Granger causality (GC) is often considered not an actual form of causality. Still, it is arguably the most widely used method to assess the predictability of a time series from another one. Granger causality has been widely used in many applied disciplines, from neuroscience and econometrics to Earth sciences. We revisit GC under a graphical perspective of state-space models. For that, we use GraphEM, a recently presented expectation-maximisation algorithm for estimating the linear matrix operator in the state equation of a linear-Gaussian state-space model. Lasso regularisation is included in the M-step, which is solved using a proximal splitting Douglas-Rachford algorithm. Experiments in toy examples and challenging climate problems illustrate the benefits of the proposed model and inference technique over standard Granger causality methods.

Identifiability in time series extended summary causal graphs

Anouar Meynaoui[1], *Charles K. Assaad*[2], *Emilie Devijver*[3], *Eric Gaussier*[3], *and Gregor Gossler*[3] **CT**

[1]Rennes University, IRMAR, 35000 Rennes, France
[2] EasyVista, 38000 Grenoble, France
[3] Grenoble Alpes University, CNRS, Grenoble INP, LIG, 38000 Grenoble, France

Several studies have been devoted to causality in time series. In these studies, the causal graph can either explicitly represent all causal relations or only an abstraction of them. An example of the former is the window causal graph under consistency through time, whereas an example of the latter is the so-called ex- tended summary causal graph. If both window and extended summary causal graphs have been studied in the context of causal discovery, the identifiability problem, which consists in finding a do-free formula for an intervention, has only been solved for window causal graphs. We address here this problem on extended summary causal graphs and propose an efficient algorithm to compute, when there is no hidden common causes, the do-free formulas corresponding to a given intervention. The proposed algorithm guarantees identifiability in any causal graph provided that the causal sufficiency assumption is fulfilled.

Inference in High-Dimensional Generalized Additive Models

CT

[1]Rennes University, IRMAR, 35000 Rennes, France
[2] EasyVista, 38000 Grenoble, France
[3] Grenoble Alpes University, CNRS, Grenoble INP, LIG, 38000 Grenoble, France

Several studies have been devoted to causality in time series. In these studies, the causal graph can either explicitly represent all causal relations or only an abstraction of them. An example of the former is the window causal graph under consistency through time, whereas an example of the latter is the so-called ex- tended summary causal graph. If both window and extended summary causal graphs have been studied in the context of causal discovery, the identifiability problem, which consists in finding a do-free formula for an intervention, has only been solved for window causal graphs. We address here this problem on extended summary causal graphs and propose an efficient algorithm to compute, when there is no hidden common causes, the do-free formulas corresponding to a given intervention. The proposed algorithm guarantees identifiability in any causal graph provided that the causal sufficiency assumption is fulfilled.

Out-of-distribution from a causal perspective: from domain generalization to OOD problems in ocean plankton analysis

Shiyang Yan

IT

Inria Saclay, France

A wide variety of model explanation approaches have been proposed in recent years, all guided by very different rationales and heuristics. In this paper, we take a new route and cast interpretability as a statistical inference problem. We propose a general deep probabilistic model designed to produce interpretable predictions. The model parameters can be learned via maximum likelihood, and the method can be adapted to any predictor network architecture and any type of prediction problem. Our method is a case of amortized interpretability models, where a neural network is used as a selector to allow for fast interpretation at inference time. Several popular interpretability methods are shown to be particular cases of regularised maximum likelihood for our general model. We propose new datasets with ground truth selection which allow for the evaluation of the features importance map. Using these datasets, we show experimentally that using multiple imputation provides more reasonable interpretations.

Relaxing Observability Assumption in Causal Inference with Kernel Methods

Yuchen Zhu IT

Yale University, USA

Causal Inference is necessary in many social science domains for understanding the effects of interventions such as that of a new drug, or that of educational policy changes. A fundamental obstacle in achieving the consistent estimation of such effects is the existence of latent variables. Often practitioners have to deal with such latency with observed covariates which can be seen as corrupted records of the latent variable. Moreover, departing from traditional statistical methods, which often exhibit consistency guarantees at the cost of restrictive modeling assumptions, kernel methods are a flexible approach for nonparametric estimation but where guarantees can still be achieved. With these goals in mind, in this talk I will describe ways to formalize the problem and outline kernel-based methods to solve them.

Simple Sorting Criteria Help Find the Causal Order in Additive Noise Models

***Alexander G. Reisach*[1], *Myriam Tami*[2], *Christof Seiler*[3,4], *Antoine Chambaz*[1], *and Sebastian Weichwald*[5,6]** IT

[1]Paris Cité University, CNRS, MAP5, F-75006 Paris, France
[2] CentraleSupélec, Paris-Saclay University, France
[3] Department of Advanced Computing Sciences, Maastricht University, The Netherlands
[4] Mathematics Centre Maastricht, Maastricht University, The Netherlands
[5] Department of Mathematical Sciences, University of Copenhagen, Denmark
[6] Pioneer Centre for AI, Copenhagen, Denmark

Additive Noise Models (ANM) encode a popular functional assumption that enables learning causal structure from observational data. Due to a lack of real-world data meeting the assumptions, synthetic ANM data are often used to evaluate causal discovery algorithms. Reisach et al. (2021) show that, for common simulation parameters, a variable ordering by increasing variance is closely aligned with a causal order and introduce var-sortability to quantify the alignment. Here, we show that not only variance, but also the fraction of a variable's variance explained by all others, as captured by the coefficient of determination R^2, tends to increase along the causal order. Simple baseline algorithms can use R^2-sortability to match the performance of established methods. Since R^2-sortability is invariant under data rescaling, these algorithms perform equally well on standardized or rescaled data, addressing a key limitation of algorithms exploiting var-sortability. We characterize and empirically assess R^2-sortability for different simulation parameters. We show that all simulation parameters can affect R^2-sortability and must be chosen deliberately to control the difficulty of the causal discovery task and the real-world plausibility of the simulated data. We provide an implementation of the sortability measures and sortability-based algorithms in our library CausalDisco.

Stochastic Causal Programming for Bounding Treatment Effects

***Kirtan Padh*[1], *Jakob Zeitler*[2], *David Watson*[3], *Matt Kusner*[2], *Ricardo Silva*[2], *and Niki Kilbertus*[1]**

IT

[1]Helmholtz AI, Helmholtz Munich & Technical University Munich
[2] University College London, England
[3] King's College London, England

Causal effect estimation is important for many tasks in the natural and social sciences. We design algorithms for the continuous partial identification problem: bounding the effects of multivariate, continuous treatments when unmeasured confounding makes identification impossible. Specifically, we cast causal effects as objective functions within a constrained optimization problem, and minimize/maximize these functions to obtain bounds. We combine flexible learning algorithms with Monte Carlo methods to implement a family of solutions under the name of stochastic causal programming. In particular, we show how the generic framework can be efficiently formulated in settings where auxiliary variables are clustered into pre-treatment and post-treatment sets, where no fine-grained causal graph can be easily specified. In these settings, we can avoid the need for fully specifying the distribution family of hidden common causes. Monte Carlo computation is also much simplified, leading to algorithms which are more computationally stable against alternatives.

Towards a Measure-Theoretic Axiomatisation of Causality

Junhyung Park

IT

Max Planck Institute for Intelligent Systems, Germany

Despite the fact that causality is a major area of research in a wide range of domains, there is still no universally agreed axiomatisation of causality. We view causality both as an extension of probability theory and as a study of what happens when one intervenes on the system, and argue in favour of taking the undisputed, measure-theoretic axiomatisation of probability as the starting point towards an axiomatisation of causality. To that end, we propose causal spaces, which consists of a probability space along with a collection of transition probability kernels, called causal kernels, that encode the causal information of the space. Our proposed framework is not only rigorously grounded in measure theory, but it also overcomes many limitations of existing frameworks in a natural way, for example cycles, latent variables and stochastic processes.

Uniform Inference in High-Dimensional Generalized Additive Models

Philipp Bach, Sven Klaassen, Jannis Kueck, Martin Spindler **CT**

University of Hamburg

We develop a method for uniform valid confidence bands of a nonparametric component f_1 in the general additive model $Y = f_1(X_1) + \cdots + f_p(X_p) + \epsilon$ in a high-dimensional setting. We employ sieve estimation and embed it in a high-dimensional Z-estimation framework allowing us to construct uniformly valid confidence bands for the first component f1. As usual in high-dimensional settings where the number of regressors p may increase with sample, a sparsity assumption is critical for the analysis. We also run simulations studies which show that our proposed method gives reliable results concerning the estimation properties and coverage properties even in small samples. Finally, we illustrate our procedure with an empirical application demonstrating the implementation and the use of the proposed method in practice.

When Causality Meets Optimal Transport

Lucas De Lara **IT**

Toulouse Mathematics Institute, France

Counterfactual frameworks have grown popular in machine learning for both explaining algorithmic decisions but also defining individual notions of fairness, more intuitive than typical group fairness conditions. However, state-of-the-art models to compute counterfactuals are either unrealistic or unfeasible. In particular, while Pearl's causal inference provides appealing rules to calculate counterfactuals, it relies on a model that is unknown and hard to discover in practice. We address the problem of designing realistic and feasible counterfactuals in the absence of a causal model. We define transport-based counterfactual models as collections of joint probability distributions between observable distributions, and show their connection to causal counterfactuals. More specifically, we argue that optimal-transport theory defines relevant transport-based counterfactual models, as they are numerically feasible, statistically-faithful, and can coincide under some assumptions with causal counterfactual models. Finally, these models make counterfactual approaches to fairness feasible, and we illustrate their practicality and efficiency on fair learning. With this paper, we aim at laying out the theoretical foundations for a new, implementable approach to counterfactual thinking.

Posters

Stochastic causal programming for bounding treatment effects

Kirtan Padh

TU Munich, Helmholtz AI

Causal effect estimation is important for numerous tasks in the natural and social sciences. However, identifying effects is impossible from observational data without making strong, often untestable assumptions. We consider algorithms for the partial identification problem, bounding treatment effects from multivariate, continuous treatments over multiple possible causal models when unmeasured confounding makes identification impossible. We consider a framework where observable evidence is matched to the implications of constraints encoded in a causal model by norm-based criteria. This generalizes classical approaches based purely on generative models. Casting causal effects as objective functions in a constrained optimization problem, we combine flexible learning algorithms with Monte Carlo methods to implement a family of solutions under the name of stochastic causal programming. In particular, we present ways by which such constrained optimization problems can be parameterized without likelihood functions for the causal or the observed data model, reducing the computational and statistical complexity of the task.

Uniform Inference in High-Dimensional Generalized Additive Models

Philipp Bach, Sven Klaassen, Jannis Kueck, Martin Spindler

University of Hamburg

We develop a method for uniform valid confidence bands of a nonparametric component f_1 in the general additive model $Y = f_1(X_1) + \cdots + f_p(X_p) + \epsilon$ in a high-dimensional setting. We employ sieve estimation and embed it in a high-dimensional Z-estimation framework allowing us to construct uniformly valid confidence bands for the first component f1. As usual in high-dimensional settings where the number of regressors p may increase with sample, a sparsity assumption is critical for the analysis. We also run simulations studies which show that our proposed method gives reliable results concerning the estimation properties and coverage properties even in small samples. Finally, we illustrate our procedure with an empirical application demonstrating the implementation and the use of the proposed method in practice.

Sequential Counterfactual Risk Minimization

Houssam Zenati

Inria Grenoble, Criteo AI Lab

Counterfactual Risk Minimization (CRM) is a framework for dealing with the logged bandit feedback problem, where the goal is to improve a logging policy using offline data. In this paper, we explore the case where it is possible to deploy learned policies multiple times and acquire new data. We extend the CRM principle and its theory to this scenario, which we call “Sequential Counterfactual Risk Minimization (SCRM).” We introduce a novel counterfactual estimator and identify conditions that can improve the performance of CRM in terms of excess risk and regret rates, by using an analysis similar to restart strategies in accelerated optimization methods. We also provide an empirical evaluation of our method in both discrete and continuous action settings, and demonstrate the benefits of multiple deployments of CRM.

A Guide for Practical Use of ADMG Causal Data Augmentation

Audrey Poinsot

INRIA

Data augmentation is essential when applying machine learning (ML) in small-data regimes. It generates new samples following the observed data distribution while increasing their diversity and variability to help researchers and practitioners improve their models’ robustness and, thus, deploy them in the real world. Nevertheless, its usage in tabular data still needs to be improved, as prior knowledge about the underlying data mechanism is seldom considered, limiting the fidelity and diversity of the generated data. Causal data augmentation strategies have been pointed out as a solution to handle these challenges by relying on conditional independence encoded in a causal graph. In this context, this paper experimentally analyzed the acyclic-directed mixed graph (ADMG) causal augmentation method considering different settings to support researchers and practitioners in understanding under which conditions prior knowledge helps generate new data points and, consequently, enhances the robustness of their models. The results highlighted that the studied method (a) is independent of the underlying model mechanism, (b) requires a minimal number of observations that may be challenging in a small-data regime to improve an ML model’s accuracy, (c) propagates outliers to the augmented set degrading the performance of the model, and (d) is sensitive to its hyperparameter’s value.

A Neural Mean Embedding Approach for Back-door and Front-door Adjustment

Liyuan Xu

UCL

We consider the estimation of average and counterfactual treatment effects, under two settings: back-door adjustment and front-door adjustment. The goal in both cases is to recover the treatment effect without having an access to a hidden confounder. This objective is attained by first estimating the conditional mean of the desired outcome variable given relevant covariates (the “first stage" regression), and then taking the (conditional) expectation of this function as a “second stage" procedure. We propose to compute these conditional expectations directly using a regression function to the learned input features of the first stage, thus avoiding the need for sampling or density estimation. All functions and features (and in particular, the output features in the second stage) are neural networks learned adaptively from data, with the sole requirement that the final layer of the first stage should be linear. The proposed method is shown to converge to the true causal parameter, and outperforms the recent state-of-the-art methods on challenging causal benchmarks, including settings involving high-dimensional image data.

Effect-Invariant Mechanisms for Policy Generalization

Sorawit Saengkyongam

ETH Zurich

Policy learning is a crucial component of many interactive learning systems, such as recommender systems, autonomous driving, and personalized healthcare. One of the major challenges in policy learning is how to adapt effectively to unseen environments or tasks. Recently, it has been suggested to exploit invariant conditional distributions to learn models that generalize better to unseen environments. However, assuming invariance of entire conditional distributions (to which we refer as full invariance) may be too strong an assumption in practice. In this paper, we introduce a relaxation of the full invariance called effect invariance (e-invariance) and prove that it is sufficient, under suitable assumptions, for zero-shot generalization. We also propose an approach to incorporate e-invariance when we have a sample from the test environment, enabling few-shot generalization. Finally, we present empirical results using a mobile health intervention dataset to demonstrate the effectiveness of our approach.

Causal Inference with Information Fields

***Benjamin Heymann*[1] *and Michel De Lara*[2]**

[1] Criteo
[2] CERMICS, Ecole des Ponts

Inferring the potential consequences of an unobserved event is a fundamental scientific question. To this end, Pearl's celebrated do-calculus provides a set of inference rules to derive an interventional probability from an observational one. In this framework, the primitive causal relations are encoded as functional dependencies in a Structural Causal Model (SCM), which maps into a Directed Acyclic Graph (DAG) in the absence of cycles. In this paper, by contrast, we capture causality without reference to graphs or functional dependencies, but with information fields.The three rules of do-calculus reduce to a unique sufficient condition for conditional independence: the topological separation, which presents some theoretical and practical advantages over the d-separation. With this unique rule, we can deal with systems that cannot be represented with DAGs, for instance systems with cycles and/or 'spurious' edges.

Reactmine: a statistical search algorithm for inferring chemical reactions from time series data

Francois Fages

Inria Saclay

Inferring chemical reaction networks (CRN) from concentration time series is a challenge encouraged by the growing availability of quantitative temporal data at the cellular level. This motivates the design of algorithms to infer the preponderant reactions between the molecular species observed in a given biochemical process, and build CRN structure and kinetics models. Existing ODE-based inference methods such as SINDy resort to least square regression combined with sparsity-enforcing penalization, such as Lasso. However, we observe that these methods fail to learn sparse models when the input time series are only available in wild type conditions, i.e. without the possibility to play with combinations of zeroes in the initial conditions. We present a CRN inference algorithm which enforces sparsity by inferring reactions in a sequential fashion within a search tree of bounded depth, ranking the inferred reaction candidates according to the variance of their kinetics on their supporting transitions, and re-optimizing the kinetic parameters of the CRN candidates on the whole trace in a final pass. We show that Reactmine succeeds both on simulation data by retrieving hidden CRNs where SINDy fails, and on two real datasets, one of fluorescence videomicroscopy of cell cycle and circadian clock markers, the other one of biomedical measurements of systemic circadian biomarkers possibly acting on clock gene expression in peripheral organs, by inferring preponderant regulations in agreement with previous model-based analyses.

Relaxing Observability Assumption in Causal Inference with Kernel Methods

Yuchen Zhu[1] *Limor Gultchin*[2] *Arthur Gretton*[1] *Anna Korba*[3] *Matt Kusner*[1] *Afsaneh Mastouri*[1] *Krikamol Muandet*[4] *Ricardo Silva*[1]

[1] UCL
[2] University of Oxford
3 ENSAE/CREST
[4] CISPA

Useful Information

Talks' Location

Talks will be held at the **Jean-Baptiste-Say Auditorium** of the **National Conservatory of Arts and Crafts**, Paris, France. It is situated on the first underground floor and has independent access through entrance number one, as depicted in Fig. 1.

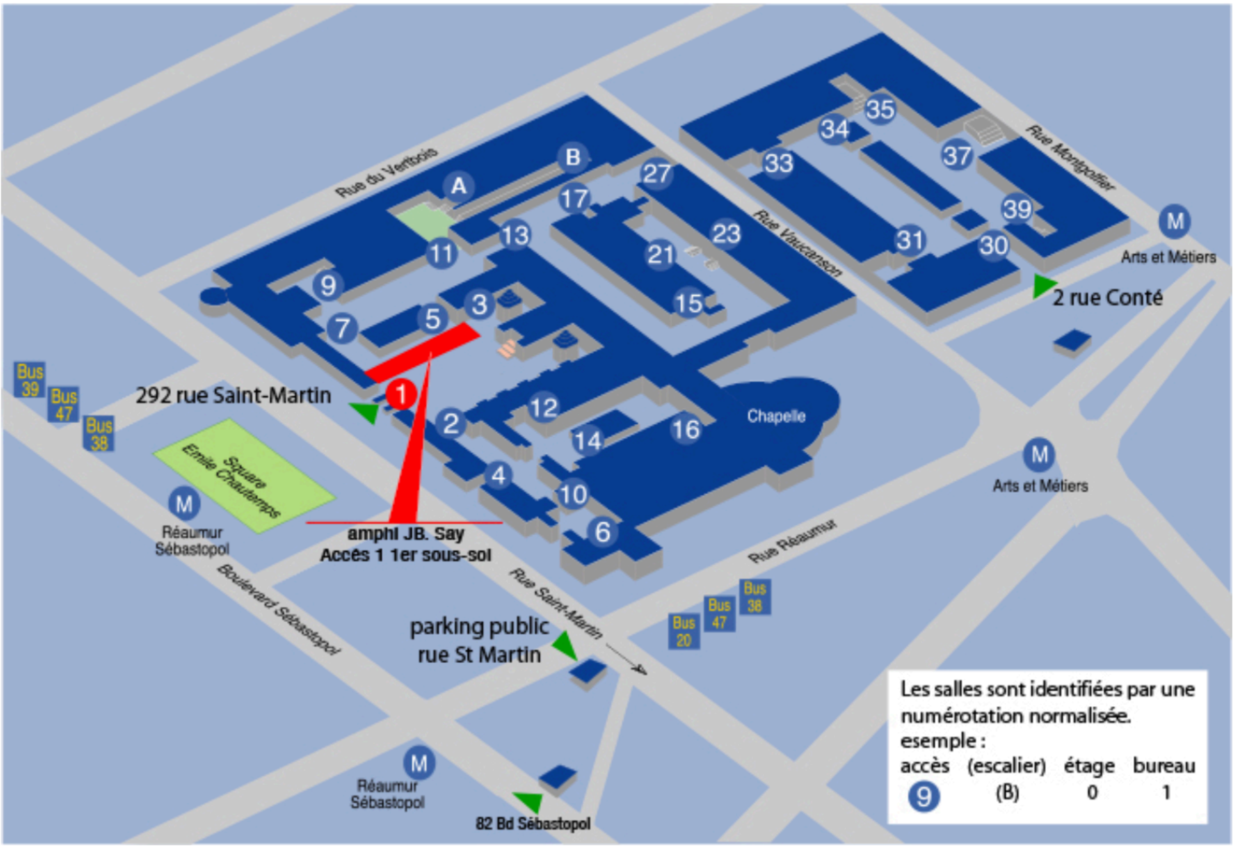

Figure 1: Access to CNAM's Jean-Baptiste-Say Auditorium

Invited and contributed talks will have a 25-minute slot followed by five minutes for questions. On the other hand, keynote lectures will have a 45-minute space, followed by 10 minutes for questions.

Discussion Groups

discussion groups represent an opportunity for attendees to present their research works through a lightning talk ($\approx$ 5 minutes) or to propose some subject for further discussion.

Coffee Breaks and Lunches

Coffee breaks and lunches will be offered in the space in front of the auditorium Jean-Baptiste-Say.

Poster Sessions

Poster sessions will be held on **Wednesday** and **Thursday** afternoon in the space in front of the auditorium Jean-Baptiste-Say.

Internet Access

A **Wi-Fi network** named **cnam-visiteurs** will be available during the conference. The **username** and **password** are **cnrs** and **vrsm0285**. The CNAM also provides access to the **eduroam network**.

Conference Dinner

Conference dinner will be held at the "**Le Procope restaurant**", at 13 Rue de l'Ancienne Comédie, 75006 Paris, France.

How to get to CNAM?

The National Conservatory of Arts and Crafts (CNAM) is located at the **292 Rue Saint-Martin, 75003 Paris, France**.

Many commute options exist to get there from Paris and its neighborhoods. You can check the best one through the Ile de France Mobilité's website (*www.iledefrance-mobilites.fr/en*) or its mobile application.

Partner Institutions and Sponsors

The When Causal Inference Meets Statistical Analysis conference is part of the Thematic Quarter on Causality organized by the The Artificial intelligence for science, science for artificial intelligence (AISSAI) center of the French National Center for Scientific Research (CNRS).

Organized by

AI for science, science for AI

With additional support from