

Causality: a special issue of Journal de la Société Française de Statistique (editorial)

Titre: Causalité : un numéro spécial du Journal de la Société Française de Statistique (éditorial)

David Benkeser¹, Antoine Chambaz² and Mark J. van der Laan³

Abstract: Invited by Gilles Celeux, editor of the *Journal de la Société Française de Statistique*, to edit a special issue of the journal devoted to **causality**, we conceived the special issue as a structured collection of articles covering a large spectrum of this exciting topic. Renowned experts of the field contribute nine articles. The special issue unfolds in three parts. Part 1 discusses causality from an epistemological stance. Part 2 discusses the notions of causal models, causal quantities and identifiability. Part 3 discusses the inference of causal quantities.

Résumé : Invités par Gilles Celeux, éditeur du *Journal de la Société Française de Statistique*, à orchestrer la préparation d'un numéro spécial dédié au thème de la **causalité**, nous l'avons conçu comme une collection structurée d'articles couvrant un large spectre de cet excitant domaine. Des experts de renommée internationale se sont associés au projet et ont offert neuf articles. Le numéro spécial se déploie en trois parties. La première partie discute de la causalité en adoptant un point de vue épistémologique. La seconde partie discute des notions de modèles causaux, de quantités causales et d'identifiabilité. La troisième partie aborde enfin le thème de l'inférence de quantités causales.

Keywords: causal models, causality, identifiability, inference, instrumental variables, semiparametric models, targeted learning

Mots-clés : apprentissage ciblé, causalité, identifiabilité, inférence, modèles causaux, modèles semi-paramétriques, variables instrumentales

AMS 2000 subject classifications: 60-01, 62A01, 62F12, 62F35, 62G08, 62G20

In 2017 Gilles Celeux, editor of the *Journal de la Société Française de Statistique*, invited us to edit a special issue of the journal devoted to **causality**. We want to thank you, Gilles, for giving us this opportunity. Three years later we are happy to present, at long last, the fruit of a collective endeavour.

We conceived the special issue as a structured collection of articles covering a large spectrum of this exciting topic then invited renowned experts of the field who, by and large, accepted enthusiastically to contribute the piece we requested from them. Laura Balzer (University of Massachusetts-Amherst), Rhian Daniel (Cardiff University), Karla DiazOrdaz (The London School of Hygiene & Tropical Medicine), Isabelle Drouet (Sorbonne Université), Peter B. Gilbert (Fred Hutchinson Cancer Research Center and University of Washington), Michal Juraska (Fred

¹ Department of Biostatistics and Bioinformatics, Emory University, Rollins School of Public Health, 1518 Clifton Road, Atlanta, Georgia 30322.

E-mail: benkeser@emory.edu

² MAP5 (UMR CNRS 8145), Université de Paris, 45 rue des Saints-Pères, 75270 Paris cedex 06.

E-mail: antoine.chambaz@u-paris.fr

³ Division of Biostatistics, University of California, Berkeley, Berkeley Way West, 2121 Berkeley Way, #5311 Berkeley, CA 94720-7360.

E-mail: laan@berkeley.edu

Hutchinson Cancer Research Center), Edward H. Kennedy (Carnegie Mellon University), Noemi Kreif (University of York), Alexander R. Luedtke (Fred Hutchinson Cancer Research Center and University of Washington), Sonia Memetea, Hachem Saddiki (University of Massachusetts-Amherst), Ilya Shpitser (Johns Hopkins University), Dylan S. Small (University of Pennsylvania), Richard Starmans (Utrecht University and Tilburg University), Jiacheng Wu (University of Washington), we want to thank you warmly for your excellent contributions.

The special issue contains nine articles and unfolds in three parts.

Part 1 discusses causality from an epistemological stance. It consists of two articles. *Prometheus unbound or paradise regained: the concept of causality in the contemporary AI-data science debate* (Starmans, 2020) highlights some aspects, core themes and controversies regarding causality from a historical-philosophical perspective with special attention to their role in the AI-data science debate. *Simpson's paradox, a tale of causality* (Chambaz et al., 2020) presents Simpson's paradox and discusses its nature based on three examples.

Part 2 discusses the notions of causal models, causal quantities and identifiability. It consists of two articles. *A primer on causality in data science* (Saddiki and Balzer, 2020) reviews the so-called “causal roadmap”, a formal framework to augment traditional statistical analyses in an effort to answer the causal questions driving the research. *Identification in Causal Models With Hidden Variables* (Shpitser, 2020) reviews identification in causal models with hidden variables for common targets that arise in causal inference applications, including causal effects, direct, indirect, and path-specific effects, and outcomes of dynamic treatment regimes.

Part 3 discusses the inference of causal quantities. It consists of five articles. *Paradoxes in instrumental variable studies with missing data and one-sided noncompliance* (Kennedy and Small, 2020) sheds light on two apparent paradoxes that arise in so-called single consent designs where there is one-sided noncompliance, i.e., where unencouraged units cannot access treatment. *Data-adaptive doubly robust instrumental variable methods for treatment effect heterogeneity* (DiazOrdaz et al., 2020) studies two doubly-robust estimators of the average treatment effect in the treated as a function of baseline covariates when there is a valid (conditional) instrument. *Assessing trends in vaccine efficacy by pathogen genetic distance* (Benkeser et al., 2020) proposes a nonparametric method for estimating and testing the trend in the effect of a vaccine across genetic distance, with an application to a recent preventive malaria vaccine efficacy trial. *Efficient principally stratified treatment effect estimation in crossover studies with absorbent binary endpoints* (Luedtke and Wu, 2020) reviews identifiability results for a parameter of the distribution of the data observed under a crossover design with the principally stratified treatment effect of interest and provides an estimator for the setting that makes no assumptions on the distribution that generated the data. Finally, *A Ride in Targeted Learning Territory* (Benkeser and Chambaz, 2020) presents a gentle introduction to the field of targeted learning, weaving together two main threads, one theoretical and the other computational. The article relies on a companion R package and contains exercises.

We hope you will enjoy this special issue.

References

- Benkeser, D. and Chambaz, A. (2020). A ride in targeted learning territory. *Journal de la Société Française de Statistique*, 161(1):201–286.

- Benkeser, D., Juraska, M., and Gilbert, P. B. (2020). Assessing trends in vaccine efficacy by pathogen genetic distance. *Journal de la Société Française de Statistique*, 161(1):164–175.
- Chambaz, A., Drouet, S., and Memetea, S. (2020). Simpson’s paradox, a tale of causality. *Journal de la Société Française de Statistique*, 161(1):42–66.
- DiazOrdaz, K., Daniel, R., and Kreif, N. (2020). Data-adaptive doubly robust instrumental variable methods for treatment effect heterogeneity. *Journal de la Société Française de Statistique*, 161(1):135–163.
- Kennedy, E. H. and Small, D. S. (2020). Paradoxes in instrumental variable studies with missing data and one-sided noncompliance. *Journal de la Société Française de Statistique*, 161(1):120–134.
- Luedtke, A. R. and Wu, J. (2020). Efficient principally stratified treatment effect estimation in crossover studies with absorbent binary endpoints. *Journal de la Société Française de Statistique*, 161(1):176–200.
- Saddiki, H. and Balzer, L. (2020). A primer on causality in data science. *Journal de la Société Française de Statistique*, 161(1):67–90.
- Shpitser, I. (2020). Identification in causal models with hidden variables. *Journal de la Société Française de Statistique*, 161(1):91–119.
- Starmans, R. (2020). Prometheus unbound or paradise regained: the concept of causality in the contemporary AI-data science debate. *Journal de la Société Française de Statistique*, 161(1):4–41.