Brummer & Partners MathDataLab Postdoc Proposal

Liam Solus

December 15, 2020

Proposed topic: Graphical Models: Causality and Algebraic Aspects

Project area: *Probabilistic graphical models*, which encode the conditional independence structure of a data-generating distribution via the absence of edges in a graph, are now widely used to study complex systems arising in biology, epidemiology, the social sciences and artificial intelligence. In the field of *causality* such models are further used to represent *cause-effect systems*, where the arrows of the graphs are interpreted as representing causal implications [3]. Current trends in causality center around two key problems: the first is the problem of *causal discovery*, where we aim to identify efficient and reliable algorithms for learning the causal structure of the system from data. The second is the problem of *causal inference*, where the aim is to make quantitative predictions on the effects of an intervention within the system. Recent developments in the problem of causal discovery use a mix of observational and interventional data to estimate a class of graphs that contain the true underlying causal structure of the system [5, 6]. Among these algorithms are a number that make use of the geometric and algebraic structure underlying parameterizations of graphical models [4, 5, 6]. However, much work is still needed to increase the reliability of the state-of the art and make causal discovery algorithms a standard tool for practitioners in fields such as biology and the medical sciences.

Project description: This project has several possible directions for exploration, and is open to anyone with experience in the field of causality and/or algebraic statistics. The first possible direction is the exploration of new methods for causal structure learning that incorporates *context-specific* conditional independencies. Recent work [1] generalized the family of discrete interventional directed acyclic graph (DAG) models to a family of models that encode both soft and hard interventions within diverse *context-specific* settings. Such models use a combination of interventional data and context-specific information to infer refined causal structure. Characterizations of model equivalence and a proof of the consistency of the Bayesian information criterion for such models was given in [2]. As such, the stage is set for the development of causal discovery algorithms that encode context-specific information about the causal system. Of particular interest would be the development of such causal discovery algorithms that are tailored to data sets arising in epidemiology. The genesis of these new models is the work of [1], which introduced them as a tool for generalizing some results in algebraic geometry and algebraic statistics. A second direction of exploration in this project could be to further our algebraic understanding of these models, exploring the combinatorics of their defining ideals in the case that the models are toric, or investigating algebraic measures on the complexity of maximum likelihood estimation. A postdoc with interests in the applications and/or algebraic statistics of graphical models would have numerous people to talk to at KTH in both the Mathematics and Mathematical Statistics departments as well as the newly formed department for the Mathematics of Data and AI.

References

[1] E. Duarte and L. Solus. *Algebraic geometry of discrete interventional models.* Preprint available at `https://arxiv.org/abs/2012.03593` (2020).

[2] E. Duarte and L. Solus. *Characterizing and learning context-specific causal models with observational and interventional data.* Preprint forthcoming (2020).

[3] J. Pearl. *Causality: models, reasoning and inference.* Vol. 29. Cambridge: MIT press, 2000.

[4] L. Solus, Y. Wang, and C. Uhler. *Consistency guarantees for permutation-based causal inference algorithms.* To appear in Biometrika. (2020).

[5] Y. Wang, L. Solus, K.D. Yang, and C. Uhler. *Permutation-based causal inference algorithms with interventions.* Advances in Neural Information Processing Systems. 2017.

[6] K. D. Yang, A. Katcoff, and C. Uhler. *Characterizing and learning equivalence classes of causal dags under interventions.* 2018 International Conference on Machine Learning ICML (2018).