

Bayesian nonparametric graphical modelling for causal inference

Federico Castelletti

Keywords:

graphical model; causal inference; heterogeneity; mixture model; Bayesian nonparametric

Abstract:

Directed Acyclic Graphs (DAGs) provide an effective framework for estimating causal relations between variables. These are generally inferred from i.i.d. samples and allow to evaluate the effect of an hypothetical intervention on a variable in the system with respect to a response of interest. More realistically however, data are collected from heterogeneous populations, with subject-specific diversities not directly linked to observed individual characteristics. This feature leads to causal effects that might be significantly different across subjects. By consequence, when not accounted, population heterogeneity might determine unreliable causal effect estimates.

We consider a Dirichlet Process (DP) mixture of DAG models as a general setting to account for heterogeneity in the sample. Specifically, each (unknown) component of the mixture corresponds to a Gaussian DAG model, whose parameter satisfies the cluster-specific conditional independencies encoded by the DAG and is assigned a Normal-DAG-Wishart prior. As a consequence, our methodology allows to identify homogeneous sub-groups of individuals with each group having its own DAG structure, parameters and therefore causal effect estimate for any intervention of interest. When applied to genomic data from breast cancer patients, we are therefore able to provide information at subject-specific level on the effect of induced gene-modifications w.r.t. target genes associated to the disease progression.