Kernel Balancing with Aggregate Profiles for Generalization and Transportability

Kwangho Kim

Department of Statistics, Korea University, Korea

Abstract

Human participation is often the bottleneck in data collection, making respondent samples scarce, selective, and difficult to link to the populations for which inference is intended. These challenges, compounded by declining survey response rates and restricted access to population level microdata, increase the risk of bias from nonrepresentative samples. Kernel balancing addresses this problem by balancing covariate functions in a reproducing kernel Hilbert space, but standard formulations typically require individual level covariates from the population used as the balancing target. In many applications, only aggregate population profiles, such as margins, moments, or cell counts, are available. We propose a profile based kernel balancing framework that uses such summaries to construct an estimated target kernel mean and assigns stable calibration weights to align respondent samples with this target. The framework covers descriptive and causal estimands, including population means and average treatment effects for source and target populations. We show that estimator error decomposes into weighted sampling error and profile approximation error, reflecting how accurately the available summaries recover the ideal target kernel mean.