

## **Inference from Nonrandom Samples Using Bayesian Machine Learning**

Yutao Liu

Boehringer Ingelheim

---

We consider inference from nonrandom samples in data-rich settings where high-dimensional auxiliary information is available both in the sample and the target population, with survey inference being a special case. We propose a regularized prediction approach that predicts the outcomes in the population using a large number of auxiliary variables such that the ignorability assumption is reasonable and the Bayesian framework is straightforward for quantification of uncertainty. Besides the auxiliary variables, we also extend the approach by estimating the propensity score for a unit to be included in the sample and also including it as a predictor in the machine learning models. We find in simulation studies that the regularized predictions using soft Bayesian additive regression trees yield valid inference for the population means and coverage rates close to the nominal levels. We demonstrate the application of the proposed methods using two different real data applications, one in a survey and one in an epidemiologic study.

**Co-authors:** Qixuan Chen and Andrew Gelman