Bayesian Inference over Data Distributions: A Gaussian Process Approach

Suyog H. Chandramouli (suyoghc@princeton.edu)[1] & Richard M. Shiffrin [2]
[1]Department of Psychology, Princeton University
[2]Department of Psychological and Brain Sciences, Indiana University

Abstract

We present BI*, a Bayesian inference framework that places beliefs over data and data patterns rather than model parameters. Unlike conventional Bayesian approaches that require specifying parametric models, BI* operates directly on the space of possible data-generating distributions—one of which represents the true state of the world from which observed data are sampled. The framework infers the probability that each candidate distribution is the true one, given the observed sample. BI* is coherent, broadly applicable across scientific settings, and provides general methods for model selection. In addition, it can incorporate aspects of science often left to human judgment, such as the effects of experimenter bias and measurement error. We make this general theory computationally feasible through fully Bayesian Gaussian Processes. We illustrate the framework's workings with an artificial dataset, demonstrating how beliefs update from prior to posterior, and how data priors influence model selection.

Keywords: Bayesian model selection, data priors, Gaussian Processes

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*