Probabilistic Programming in R with Bruno

Robert Zinkov[1,⋆], **Chung-Chieh Shan**[1]

1. Department of Computer Science, Indiana University
⋆Contact author: zinkov@indiana.edu

Keywords: Machine learning, Graphical models, Bayesian modeling, MCMC

Markov Chain Monte Carlo (MCMC) methods have proven to be very effective for estimating posterior distributions. Their utility has led to the creation of general-purpose libraries for automatically performing this sampling. Unfortunately, many existing libraries have difficulty modeling distributions which vary in dimension. In addition, they have problems representing nonparametric Bayesian models as their dimensionality is defined to grow with the amount of data available [2].

We present *Bruno*. A universal probabilistic programming language for computing probability distributions over programs. By providing a programming language where every term can denote a probability distribution, we are able to model domains with changing dimensionality and have nonparametric components. We also provide a *R* interface called **rbruno** in the same spirit as **rjags** [1] to access the samples produced by these probabilistic programs. Lastly, our software allows users to specify how they wish to perform inference on their model. This flexibility makes it possible to finely tune the MCMC sampler if the problem domain demands it.

References

[1] Plummer, M. (2012). rjags: Bayesian graphical models using mcmc. http://CRAN.R-project.org/package=rjags.

[2] Teh, Y. W., M. I. Jordan, M. J. Beal, and D. M. Blei (2006). Hierarchical Dirichlet processes. *Journal of the American Statistical Association 101*(476), 1566–1581.