Proceedings of the International Conference on Statistics: Theory and Applications (ICSTA'19)
Lisbon, Portugal – August 13 – 14, 2019
Paper No. 30
DOI: 10.11159/icsta19.30

Bayesian Experimental Design: Efficient Exploration of the Design Space

Lida Mavrogonatou, Vladislav Vyshemirsky
University of Glasgow
United Kingdom
l.mavrogonatou.1@research.gla.ac.uk; Vladislav.Vyshemirsky@glasgow.ac.uk

Extended Abstract

We consider experiments as processes consisting of three main components: a set of experimental conditions ($\delta \in D$), commonly known as the experimental design, a set of parameters ($\theta \in \Theta$) and an observable outcome ($y \in Y$). Collection of datasets, often consisting of experiment outcomes y is essential when performing statistical analyses such as parameter inference, model selection and prediction. The field of experimental design is concerned with identifying the experimental designs δ that provide datasets with the highest amount of information, commonly referred to as the optimal design, while taking into account certain restrictions.

Bayesian experimental design is formed on the basis of a decision theoretic approach, summarised by (Lindley, 1972), which relies on definition of a utility function, representing the benefit from each choice of design, experimental parameters and induced dataset. The optimal design is identified as the one incurring the highest expected utility, otherwise expressed as the marginal utility function over the joint space $\Theta \times Y$. However, problems arise with evaluation of the expected utility, even in simple problems, as it is often not possible to be performed analytically and as a result, approximation of it is required.

Standard approaches entail approximation of the expected utility using nested Monte- Carlo integration in a pre-determined set of design points, however this is shown to still be problematic as it relies on evaluation of the utility for a large sample from $\Theta \times Y$ which is in most cases found to be computationally intensive. (Müller, 2005) consider the expected utility as an unnormalised probability density function by placing a joint distribution on the space $D \times \Theta \times Y$ thus defining an augmented probability model, whose marginal distribution over $\Theta \times Y$ is proportional to the expected utility. They further employ sampling methods such as MCMC to explore the utility surface by obtaining observations from the augmented model. Alternatively, (Müller & Parmigiani, 1995) developed methodology which focuses on "learning" the relationship between the expected utility evaluated at a small initial number of design sets and utilising it to make predictions for new sets thus avoiding further evaluation of the expected utility.

We consider methodologies for estimation of the expected utility, examining both paramateric and non- parametric approaches. Comparison of different methods is presented through application to model selection problems. We proceed to discussing methodology for efficient exploration of the design space D. Standard approaches often rely on evaluation of the expected utility on a fixed, discrete set of experiments and consequently choosing the one incurring the maximum utility. As a result, each experiment is treated independently and so useful information is often disregarded. We discuss online approaches that allow us to incorporate already obtained knowledge into desicions on future experiments to be considered.

References

[1] D. V. Lindley, Bayesian statistics, a review. SIAM, 1972.
[2] P. Müller, "Simulation based optimal design" in Handbook of Statistics, vol. 25, pp. 509–518, 2005.
[3] P. Müller & G. Parmigiani, "Optimal design via curve fitting of Monte Carlo experiments," *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1322–1330, 1995.