JMLR: Workshop and Conference Proceedings vol 49:1–2, 2016

Simple Bayesian Algorithms for Best Arm Identification

Daniel Russo DAN.JOSEPH.RUSSO@GMAIL.COM
Microsoft Research New England

Abstract

This paper considers the optimal adaptive allocation of measurement effort for identifying the best among a finite set of options or designs. An experimenter sequentially chooses designs to measure and observes noisy signals of their quality with the goal of confidently identifying the best design after a small number of measurements. I propose three simple Bayesian algorithms for adaptively allocating measurement effort. One is Top-Two Probability sampling, which computes the two designs with the highest posterior probability of being optimal, and then randomizes to select among these two. One is a variant a top-two sampling which considers not only the probability a design is optimal, but the expected amount by which its quality exceeds that of other designs. The final algorithm is a modified version of Thompson sampling that is tailored for identifying the best design. I prove that these simple algorithms satisfy a strong optimality property. In a frequestist setting where the true quality of the designs is fixed, one hopes the posterior definitively identifies the optimal design, in the sense that that the posterior probability assigned to the event that some other design is optimal converges to zero as measurements are collected. I show that under the proposed algorithms this convergence occurs at an *exponential rate*, and the corresponding exponent is the best possible among all allocation rules.

1. Introduction

This paper considers the optimal adaptive allocation of measurement effort in order to identify the best among a finite set of options or designs. An experimenter sequentially chooses designs to measure and observes noisy signals of their quality. The goal is to allocate measurement effort intelligently so that the best design can be identified confidently after a small number of measurements. Problems of this form have been studied heavily in operations research, statistics, and computer science, as they form a useful abstraction of many practical settings. For example:

- Efficient A/B/C Testing: An e-commerce platform is considering a change to its website and would like to identify the best performing candidate among many potential new designs. To do this, the platform runs an experiment, displaying different designs to different users who visit the site. How should the platform decide what percentage of traffic to allocate to each website design?

- Simulation Optimization: An engineer would like to identify the best performing aircraft design among several proposals. She has access to a realistic simulator through which she can assess the quality of the designs, but each simulation trial is very time consuming and produces only noisy output. How should she allocate simulation effort among the designs?

. Extended abstract. Full version appears as [http://arxiv.org/abs/1602.08448]

This paper proposes three simple algorithms for allocating measurement effort. Each algorithm begins with a prior distribution over the unknown quality of the designs. As measurements are gathered, the experimenter learns about the designs, and beliefs are updated to form a posterior distribution. This posterior distribution gives a principled method for reasoning about the uncertain quality of different designs, and for assessing the probability any given design is optimal.

The first algorithm I propose is called *Top–Two Probability sampling*. It computes at each time-step the two designs with the highest posterior probability of being optimal. It then randomly chooses among them, selecting the design that appears most likely to be optimal with some probability $\beta > 0$, and selecting the second most likely otherwise. Beliefs are updated as observations are collected, so the top two designs change over time. The long run fraction of measurement effort allocated to each design depends on the true quality of the designs, and the distribution of observation noise. The *Top–Two Value sampling* algorithm proceeds in a similar manner, but in estimating the top-two designs it considers not only the probability a design is optimal, but the expected amount by which it exceeds other designs.

The final algorithm I propose is a modification of the *Thompson sampling* algorithm for multi-armed bandits. Thompson sampling has attracted a great deal of recent interest in both academia and industry, but it's designed to maximize the cumulative reward earned while sampling. As a result, in the long run the algorithm allocates almost all effort to measuring the estimated-best design, and it takes a long time to certify that none of the alternative designs would offer better performance. This paper shows, however, that the algorithm can be easily modified for the objective of efficiently identifying the best arm. The variant I propose adds a simple re-sampling step to Thompson sampling, which prevents the algorithm from allocating too much measurement effort to the estimated-best design.

These simple heuristic algorithms are shown to satisfy a strong optimality property. The analysis focuses on frequentist consistency and rate convergence of the posterior distribution, and therefore takes place in a setting where the true quality of the designs is fixed, but unknown to the experimenter. One hopes that as measurements are collected the posterior distribution definitively identifies the true best design, in the sense that the posterior probability assigned to the event that some other design is optimal converges to zero. I show that under the proposed algorithms this convergence occurs at an *exponential rate*, and the corresponding exponent is essentially the best possible among all allocation rules.

In studying the optimality of the proposed algorithms, the paper characterizes what the simulation optimization literature calls an optimal computing budget allocation. This specifies the asymptotically optimal fraction of measurement effort to allocate to each design as a complicated function of the true quality of the designs. Rather than using an *equal allocation* of measurement effort across designs, this calculation shows it is optimal to adjust measurement effort to gather *equal evidence* to rule out each sub-optimal design. By only sampling designs that seem most promising under the posterior, each of the proposed algorithms automatically balances the evidence gathered in this manner.

While the paper's main theoretical results are asymptotic, this is mainly to provide sharp insight. The algorithm design is completely separate from the theoretical analysis, so any approximations in the analysis don't influence practical performance. Numerical experiments suggest the proposed algorithms perform very well over moderate horizons.