Sequential Bayes-optimal Policies for Multiple Comparisons with a Control

Jing Xie and Peter I. Frazier

Operations Research & Information Engineering, Cornell University

Sunday November 13, 2011
Junior Faculty Interest Group (JFIG) Paper Competition
INFORMS Annual Meeting
Charlotte, NC
What is Multiple Comparisons with a Control?

- We have a stochastic simulator.
- Given a set of input parameters $x$, it provides a random sample $y(x)$.
- For which inputs $x$ is $E[y(x)] > 0$?

In other words, find the level set of $x \mapsto E[f(x)]$. 
We must allocate ambulances across 11 bases in the city of Edmonton. Which allocations satisfy mandated minimums for percentage of calls answered in time, under a variety of different possible call arrival rates?

Allocations of Ambulances to Bases

[Thanks to Shane Henderson and Matt Maxwell for providing the ambulance simulation]
A researcher who develops a new algorithm would like to know:

- In which problem settings is average-case performance better with Algorithm A than with Algorithm B?
Given Samples, Estimating the Level Set is Well-Understood
Given Samples, Estimating the Level Set is Well-Understood
If our simulator is complex and takes a long time to run, the number of samples we can take is limited.

This makes accurate MCC more difficult.

**Where should we place our limited samples to estimate the level set as accurately as possible?**

Our contribution: We provide an answer to this question with Bayes-optimal performance.
The Optimal Policy Puts Samples Where They Help Most
The Optimal Policy Puts Samples Where They Help Most

\[ y(x) \]

\[ x=1 \] \[ x=10 \]
The Optimal Policy Puts Samples Where They Help Most
Mathematical Model

- We have alternatives $x = 1, \ldots, k$.
- Samples from alternative $x$ are $\text{Normal}(\mu_x, \sigma_x^2)$.
- $\mu_x$ is unknown, while $\sigma_x^2$ is assumed known (can be relaxed).
- We have an independent normal Bayesian prior on each $\mu_x$.
- Sampling continues until an external deadline requires it to stop.
- We assume this deadline is unknown and geometrically distributed.
- When sampling stops, we estimate the level set $\{x : \mu_x > 0\}$ based on the samples. The reward is the number of alternatives correctly classified.
A policy $\pi$ is a rule for choosing where to sample next, based on previous observations.

Let $R$ be the number of alternatives classified correctly when sampling stops.

$E^\pi[R|\vec{\mu}]$ is the performance under policy $\pi$ and true mean vector $\vec{\mu}$.

$\int E^\pi[R|\vec{\mu}]P(d\vec{\mu}) = E^\pi[R]$ is the Bayes- or average-case performance.

We wish to find the policy that maximizes this.
Finding the Optimal Policy Means Solving a Dynamic Program

- We wish to find the policy $\pi$ that solves:

$$\sup_\pi \mathbb{E}^\pi [R]$$

- The solution is characterized theoretically via dynamic programming.
- The *curse of dimensionality* usually makes computing the solution to such dynamic programs intractable.
We Rewrite the Problem as a Bandit Problem

- The expected reward is the expected number of alternatives correctly classified at the end.
- We decompose this expected reward into an infinite sum of discounted expected one-step rewards

\[ \mathbb{E}^{\pi}[R] = R_0 + \mathbb{E}^{\pi} \left[ \sum_{n=1}^{\infty} \alpha^{n-1} R_n \right]. \]

Here,
- \( \alpha \) is the parameter of the geometric distribution of the deadline.
- \( R_0 \) is the expected reward if we stop after taking no samples.
- \( R_n \) is the expected one-step improvement, due to sampling, of the probability of correctly classifying the alternative sampled.
We Can Compute the Optimal Policy

- Written in this way, the problem becomes a **multi-armed bandit** problem.
- (Gittins & Jones 1974) shows the optimal solution is

\[
\arg \max_x \nu_x(S_{nx}),
\]

where \( S_{nx} \) is a parameterization of the Bayesian posterior on \( \mu_x \).

- The Gittins index \( \nu_x(\cdot) \) is defined in terms of a single-alternative version of the problem

\[
\nu_x(s) = \sup_{\tau > 0} \mathbb{E} \left[ \frac{\sum_{n=1}^{\tau} \alpha^{n-1} R_n}{\sum_{n=1}^{\tau} \alpha^{n-1}} \bigg| S_{0x} = s, x_1 = \cdots = x_\tau = x \right].
\]

- We can compute Gittins indices efficiently because the single-alternative problem is much smaller than the full DP.
The Optimal Policy Improves Accuracy in Ambulance Positioning

PE = pure exploration (sample at random);

MV = max variance (equal allocation);

OPT = optimal policy.

Time ↓
Conclusion: The Optimal Policy Saves Time

- The **Multiple Comparisons with a Control** problem appears in many different simulation applications.
- **We found the optimal method** for deciding where to sample.
- This allows accurately characterizing level sets more quickly and with fewer simulation samples.
Thank You

Any questions?
Much of the literature on this topic has focused on providing non-Bayesian statistical guarantees on accuracy.

- One-stage policies: Tukey 1953, …
- Two-stage policies: Dudewicz and Ramberg 1972, …

We focus on sampling sequentially in a Bayes-optimal way.