The local time method for targeting and selection

Ilya O. Ryzhov

February 8, 2016

Abstract

We propose a framework for “targeting and selection” (T&S), a new problem class in simulation optimization where the objective is to select a simulation alternative whose mean performance matches a pre-specified target as closely as possible. T&S resembles the more well-known problem of ranking and selection, but presents unexpected challenges: for example, a one-step look-ahead method may produce statistically inconsistent estimates of the values, even under very standard normality assumptions. We create a new and fundamentally different approach, based on a Brownian local time model, that exhibits characteristics of two widely-studied methodologies, namely expected improvement and optimal computing budget allocation. We characterize the asymptotic sampling rates of this method and relate them to the convergence rates of metrics of interest. The local time method outperforms benchmarks in experiments, including problems where the modeling assumptions of T&S are violated.

1 Introduction

In simulation optimization problems, a decision-maker seeks to optimize a performance measure that can only be estimated by running an expensive stochastic simulation. Often, the goal is to optimize the simulation itself, by choosing a design or a set of input parameters (e.g., for a factory layout, call center, or manufacturing operation) that produce the best mean performance inside the simulator. The budget for experimenting with different settings is typically limited, and the literature on simulation optimization develops ways to use the available simulations efficiently. See Hong & Nelson (2009) for an introduction to several important schools of thought in this literature. In particular, the ranking and selection (R&S) framework assumes that the set of competing designs is finite and usually small, and also that the decision-maker always wishes to implement the design with the maximum (or minimum) mean performance.

This paper presents a new and different class of simulation optimization problems, which we call targeting and selection by analogy. As in R&S, we consider a finite, small set of design alternatives. However, we suppose that the decision-maker’s objective is not to maximize the mean performance,
but rather to select the system whose performance matches a pre-specified target as closely as possible. This problem naturally arises when validating or calibrating a simulation model. In the deployment of simulation optimization systems, it is necessary to ensure that the simulator presents an accurate model of the underlying real-world processes, e.g., by matching performance metrics, such as the number of customers served or the number of delayed orders, based on historical data. To that end, one may need to compare multiple competing simulation models, or experiment with parameter settings within a single model. See Frazier et al. (2009) for an example of such an application (calibrating a business simulator for a trucking company).

Targeting and selection may also arise in other areas of application. Several additional examples include the following:

1. **Designing competitive events.** Online gaming services such as Microsoft’s Xbox Live have millions of subscribers. The server learns a user’s skill level based on past performance (Dangauthier et al., 2007), then uses a match-making algorithm (Herbrich et al., 2006) to choose an opponent for the player in each new game. To ensure fair and competitive games, we wish to match opponents whose skill levels are as close together as possible.

2. **Manufacturing and quality control.** We have the ability to control the number of defective items produced by a manufacturing process (Lambert, 1992). It is too expensive to minimize the number of defects, but undesirable to have too many. Therefore, we prefer a design where this metric simply matches an acceptable tolerance level.

3. **Polling strategies.** A political candidate would like to campaign in key “swing” regions that could influence the outcome of the election. Before committing vital resources to an event, the candidate’s team can poll a certain area to improve estimates of the candidate’s rating. The goal is to select the region where the candidate’s rating is as close to 50% as possible.

At first glance, targeting and selection (T&S) greatly resembles R&S. Indeed, in this paper, we use a standard Bayesian learning model from R&S (see e.g. Ch. 2 of Powell & Ryzhov, 2012), which assumes independent and normally distributed simulation output with independent normal priors. These assumptions are fairly standard in R&S, appearing in both Bayesian (Gupta & Miescke, 1996; Chick & Gans, 2009) and frequentist (Hartmann, 1991; Kim & Nelson, 2001) formulations. However, even with this standard model, T&S turns out to pose problems for the direct application of standard R&S techniques. To give one example, the expected improvement
EI methodology produces good practical performance, and is guaranteed to identify the optimal alternative asymptotically, in numerous variants of R&S (Frazier et al., 2008; Chick et al., 2010; Ryzhov et al., 2012). Surprisingly, this is not true in T&S: we derive one widely-studied form of EI, known as a knowledge gradient policy, and prove that it may undervalue some alternatives so severely that they may never be measured. This motivates the need for new approaches to the targeting problem.

We propose one such approach, called the local time method (LTM). First, we approximate the learning process (i.e., the sequence of sample means obtained by simulating a fixed alternative) by a continuous-time diffusion, a technique pioneered by Brezzi & Lai (2002). Using facts from Brownian motion theory, the diffusion can be rewritten as a time-changed Wiener process; this “standardization” property can make continuous-time learning problems more tractable than their discrete-time counterparts (Yao, 2006; Chick & Gans, 2009). In our case, the Brownian connection leads to a natural idea for the T&S problem: the local time of the Wiener process (Cinlar, 2011) can be interpreted as a measure of the proximity of the true value of the alternative to the target (if the true mean performance is closer to the target, the sample mean spends more time around it as well). Our proposed algorithm then simulates the alternative with the highest expected local time, calculated adaptively based on the most recent set of sample means.

In addition to its intuitive appeal, the method turns out to exhibit interesting parallels to another widely-studied simulation optimization methodology known as optimal computing budget allocation (or OCBA; see Chen & Lee, 2010 for an overview). We explicitly characterize the asymptotic rates at which different alternatives are simulated under the LTM, and find that these rates match those of the OCBA policy developed by Chen et al. (2008) for a subset selection problem; our result also helps to explain how the approach in that work essentially converts subset selection into an instance of T&S. We provide an interpretation of these rates, showing that they optimize the rate at which the posterior likelihood of exactly matching the target vanishes to zero. Finally, we use a large deviations analysis (Glynn & Juneja, 2004) to derive the theoretically optimal allocation for the T&S problem, and relate it to the LTM rates in a limiting case. Thus, one may choose to view the LTM as a relative of OCBA specifically designed for fully sequential targeting and selection.

More broadly, our work on T&S can be placed within a recent stream of literature focusing on expanding the optimization framework of simulation optimization beyond basic R&S. For instance,
Ryzhov & Powell (2012) interprets the alternatives in R&S as objective coefficients in a linear program. Other problems that optimize some economic objective taking the unknown values as inputs include the newsvendor problem (Lariviere & Porteus, 1999), dynamic pricing (Harrison et al., 2012), and projects with multi-attribute performance metrics (Morrice & Butler, 2006). The simulation literature has also explored constrained optimization (Andradóttir & Kim, 2010; Xu et al., 2010; Pasupathy et al., 2014) and feasibility determination (Batur & Kim, 2010), as well as multi-objective problems where the goal is to estimate the Pareto frontier (Ryu et al., 2009; Lee et al., 2010; Hunter & McClosky, 2015). However, none of these various developments addresses the targeting problem directly. To our knowledge, our work is the first to formulate the general T&S problem and propose a new method explicitly designed to solve it.

This paper makes the following contributions: 1) We formulate the targeting and selection problem, and motivate its further study by proving that the knowledge gradient method for this problem may be inconsistent. 2) We create the local time method for T&S, based on the Brownian local time of a continuous-time interpolation of the sequence of estimated values for an alternative. The underlying idea is fundamentally different from any of the predominant R&S methodologies, relying on the intuitive interpretation of the local time as a measure of proximity to the target. 3) We show that the LTM retains consistency, explicitly characterize its asymptotic sampling rates, and relate them to the convergence rate of the overall algorithm. 4) We present numerical experiments demonstrating the effectiveness of the LTM against other natural benchmarks. We deliberately consider examples where our modeling assumptions are violated (for example, the simulation output may be non-normal, and its variance unknown), and show numerically that the LTM performs robustly in these settings, while the other benchmarks do not.

We proceed as follows. Section 2 describes the Bayesian model used for learning in T&S and discusses the inconsistency of the EI approach. Section 3 introduces the local time method. Section 4 contains our asymptotic analysis, and Section 5 presents experimental results. Section 6 concludes.

2 Targeting and selection

Section 2.1 gives a mathematical framework for targeting and selection, based on a Bayesian learning model with independent normal priors and observations. This model forms the basis for many methods in R&S; yet, even with the same model, T&S is structurally different. In Section 2.2
we show that the nonlinear objective of the targeting problem causes inconsistent behaviour in knowledge gradient methods, a widely-studied class of heuristic learning algorithms.

2.1 Learning model for T&S

Suppose that there are \( M \) design alternatives. For each alternative \( x \in \{1, ..., M\} \), we can run a stochastic simulation and observe the output \( W_x \sim \mathcal{N}(\mu_x, \lambda_x^2) \). We assume that the sampling variance \( \lambda_x^2 \) is known, whereas the mean performance \( \mu_x \) is unknown. The decision-maker begins with a prior distribution \( \mu_x \sim \mathcal{N}(\theta_0 x, (\sigma_0^x)^2) \) for each \( x \). These prior distributions are assumed to be independent.

As in the literature on fully-sequential Bayesian R&S, we assume that the decision-maker simulates one alternative at a time. The \((n + 1)\)st decision \( x^n \) depends on the previous decisions \( x^0, ..., x^{n-1} \) and the independent observations \( W_x^1, ..., W_x^{n-1} \) resulting from those decisions. We let \( \mathcal{F}^n \) be the sigma-algebra generated by this information. We use the convention that any random variable indexed by \( n \) in the superscript is measurable with respect to \( \mathcal{F}^n \). Also, we let \( P^n \) and \( E^n \) denote conditional probabilities and expectations given \( \mathcal{F}^n \).

Standard results from Bayesian analysis (DeGroot, 1970) tell us that the conditional distribution of \( \theta_x \) given \( \mathcal{F}^n \) is still normal with mean \( \theta^n_x \) and variance \( (\sigma^n_x)^2 \). These parameters are updated recursively according to the equations

\[
\theta^{n+1}_x = \left\{ \begin{array}{ll}
\frac{(\sigma^n_x)^2 \theta^n_x + \lambda_x^{-2} W_x^{n+1}}{(\sigma^n_x)^2 + \lambda_x^{-2}} & \text{if } x^n = x \\
\theta^n_x & \text{if } x^n \neq x,
\end{array} \right.
\]  

(1)

and

\[
(\sigma^{n+1}_x)^2 = \left\{ \begin{array}{ll}
(\sigma^n_x)^2 + \lambda_x^{-2} & \text{if } x^n = x \\
(\sigma^n_x)^2 & \text{if } x^n \neq x.
\end{array} \right.
\]

(2)

It can also be shown (Ch. 2 of Powell & Ryzhov, 2012) that the conditional distribution of \( \theta^{n+1}_x \) given \( \mathcal{F}^n \) and given that \( x^n = x \) is also normal with mean \( \theta^n_x \) and variance

\[
(\tilde{\sigma}^n_x)^2 = (\sigma^n_x)^2 - (\sigma^{n+1}_x)^2.
\]

(3)

We see that running a simulation on alternative \( x \) provides no information about any \( y \neq x \). Our beliefs after \( n \) observations are completely characterized by the vectors \( \theta^n = (\theta^n_1, ..., \theta^n_M) \)
and $\sigma^n = (\sigma^n_1,...,\sigma^n_M)$. The decision-maker chooses to simulate $x^n$ adaptively, based on all the information observed up to that point. To reflect this dependence on $\theta^n$ and $\sigma^n$ in the notation, we write the $(n+1)$st simulation decision as $X^n(\theta^n, \sigma^n)$.

Let $c$ be a pre-determined target value, e.g. the desired performance level. Suppose that we are given a budget of $N$ simulations. At time $N$, after the entire budget has been used, the decision-maker’s selection decision is given by

$$X^N(\theta^N, \sigma^N) = \arg\min_x E^N \left[(\mu - c)^2\right]$$

$$= \arg\min_x \left(\theta^N_x - c\right)^2 + \left(\sigma^N_x\right)^2. \quad (4)$$

One possible objective of T&S may be written as

$$\sup_\pi E^\pi \min_x E^N \left[(\mu_x - c)^2\right], \quad (5)$$

where $\pi$ is a policy defined by decision rules $X^\pi,n (\theta^n, \sigma^n)$ for $n = 0,...,N-1$, and $E^\pi$ is a conditional expectation given that every decision will be made according to the policy $\pi$. We may also consider an alternate objective

$$\sup_\pi E^\pi \mathbb{1}\left\{\arg\min_x E^N[(\mu_x - c)^2] = \arg\min_x (\mu_x - c)^2\right\}, \quad (6)$$

which is widely known in the literature as the “probability of correct selection” (Kim & Nelson, 2001). Recent work by Gao & Shi (2014) suggests that (5) and (6) exhibit similar convergence properties for large $N$, and can be optimized by the same allocation. Since neither objective can be tractably optimized, our discussion will mostly consider suboptimal algorithms, and we will use whichever objective is most convenient for our presentation, as necessary.

In ordinary R&S, (4) would be replaced by the simple decision

$$X^{RS,N}(\theta^N, \sigma^N) = \arg\max_x E^N \mu_x = \arg\max_x \theta^N_x. \quad (7)$$

In targeting and selection, however, $x$ is preferred if $\mu_x$ is probabilistically more likely to be close to the target $c$. It is not enough for the point-estimate $\theta^N_x$ to be close to $c$. If the posterior variance of $\mu_x$ is too high, we may prefer another alternative whose posterior mean is farther from $c$, but whose posterior variance is lower.

The target $c$ is assumed to be a fixed number; however, it is relatively straightforward to accommodate noise in the target. For instance, suppose that $c$ is replaced by a stochastic target
\( C \sim \mathcal{N}(c, \sigma_c^2) \) which is independent of any \( \mu_x \); it is then straightforward to show that matching the mean \( c \) is sufficient.

**Proposition 2.1.** For \( C \sim \mathcal{N}(c, \sigma_c^2) \) independent of \( \mu \), the objective

\[
\sup_{\pi} \mathbb{E}^\pi \min_x \mathbb{E}^N \left[ (\mu_x - C)^2 \right]
\]

is equivalent to the objective in (5).

**Proof:** We calculate

\[
\mathbb{E}^N \left[ (\mu_x - C)^2 \right] = \mathbb{E}^N \mathbb{E}^N \left[ (\mu_x - C)^2 \mid \mu \right] = \mathbb{E}^N \left[ (\mu_x - c)^2 \right] + \sigma_c^2.
\]

Since \( \sigma_c^2 \) does not depend on \( x \) or \( \pi \),

\[
\sup_{\pi} \mathbb{E}^\pi \min_x \mathbb{E}^N \left[ (\mu_x - C)^2 \right] = \sigma_c^2 + \sup_{\pi} \mathbb{E}^\pi \min_x \mathbb{E}^N \left[ (\mu_x - c)^2 \right],
\]

and \( \sigma_c^2 \) can be dropped from (8) without affecting the optimal choice of \( \pi \). \( \square \)

If the value \( C \) is also being learned, the problem could be modified to match the unknown value \( (\mu_x - C) \) to the fixed target \( c = 0 \). In Section 5.2, we conduct numerical experiments on a problem where a similar issue arises. In the base case (known, fixed \( c \)), the key feature distinguishing (4) from (7) is that T&S incorporates the posterior variance \( \left( \sigma_x^N \right)^2 \) into the implementation decision. By contrast, in most variants of R&S, the implementation decision typically only depends on \( \theta^N \).

### 2.2 Inconsistency of the knowledge gradient method

In this section, we give additional motivation for our work by showing that T&S presents difficulties for a class of heuristics known as “knowledge gradient” methods. For R&S with normality assumptions, these methods are statistically consistent, meaning that they lead us to learn the value of the best alternative in the limit as the simulation budget becomes large. Normality ensures that any alternative has some potential to improve the selection decision, since the simulation output can be arbitrarily high or low with some non-zero probability. In T&S, this is no longer true, because extreme-valued observations are no longer helpful to the decision-maker. Consequently, procedures that perform very well in R&S may produce poor results in T&S, and may not even retain their usual consistency properties.
Knowledge gradient methods (Powell & Ryzhov, 2012) always simulate alternatives for which a single additional observation is expected to produce the greatest single-period improvement in our objective function. In the context of T&S, formulated as in (4), this decision is expressed as

\[ X_{KG,n}^n (\theta^n, \sigma^n) = \arg \max_x \nu_{x}^{KG,n} \]  

(9)

where

\[ \nu_{x}^{KG,n} = \mathbb{E}_{x}^n \left[ \min_y \left( (\theta^n_y - c)^2 + (\sigma^n_y)^2 \right) - \min_y \left( (\theta^{n+1}_y - c)^2 + (\sigma^{n+1}_y)^2 \right) \right], \]  

(10)

where \( \mathbb{E}_{x}^n \) is an expectation given \( \mathcal{F}^n \) and given that \( x^n = x \). At time \( n \), the future beliefs \( \theta^{n+1} \) are unknown, but there are some possible values of \( \theta^{n+1}_x \) for which our valuation of \( x \) improves. The expected improvement in our estimate of the highest-valued alternative, obtained as a result of simulating \( x \), is given in (10).

Such procedures are widely studied in Bayesian R&S (see Chick, 2006, for an overview). They have the advantage of being adaptable to many optimal learning problems beyond R&S: one simply has to calculate the expected improvement of the appropriate objective function, as in (10). Frazier et al. (2008) shows that, in Bayesian R&S with independent normal priors and observations, this approach produces consistent estimators of the true values. However, in targeting and selection, this property no longer holds, even with the same learning model.

**Proposition 2.2.** For all \( x \), \( \nu_{x}^{KG,n} \geq 0 \). Furthermore, \( \nu_{x}^{KG,n} = 0 \) if

\[ \min_y \left( (\theta^n_y - c)^2 + (\sigma^n_y)^2 \right) \leq (\sigma^{n+1}_x)^2. \]  

(11)

**Proof:** Observe that

\[ M^n_x = (\theta^n_x - c)^2 + (\sigma^n_x)^2 \]  

(12)

is an \( \mathcal{F}^n \)-martingale, independent of the policy used for making simulation decisions. We can see this by interpreting (12) as \( \mathbb{E}^n \left[ (\mu_x - c)^2 \right] \). The positivity of \( \nu_{x}^{KG,n} \) follows by applying Jensen’s inequality for concave functions to obtain

\[ \mathbb{E}_{x}^n \min_y M^{n+1}_y \leq \min_y \mathbb{E}_{x}^n M^{n+1}_y = \min_y M^n_y. \]

Substituting this into (10), we obtain \( \nu_{x}^{KG,n} \geq 0 \).

Suppose now that \( x^n = x \), and define

\[ C^n_x = \min_{y \neq x} \left( (\theta^n_y - c)^2 + (\sigma^n_y)^2 \right). \]
Then, $\sigma_{n+1}^x$ can be obtained deterministically from $\sigma^n_x$ via (2), whereas $\theta_{n+1}^x$ follows a normal distribution with mean $\theta^n_x$ and variance $(\tilde{\sigma}_x^n)^2$ as in (3). Furthermore, for $y \neq x$, we have $\theta_{n+1}^y = \theta^n_y$ and $\sigma_{n+1}^y = \sigma^n_y$ because the priors are independent. Then,

$$E^n_x \left[ \min_y (\theta_{n+1}^y - c)^2 + (\sigma_{n+1}^y)^2 \right] = E \min \left\{ C^n_x, (\theta_{n+1} - c)^2 + (\sigma_{n+1}^x)^2 \right\}.$$ 

We consider the conditions under which

$$C^n_x \leq (\theta_{n+1}^x - c)^2 + (\sigma_{n+1}^x)^2, \quad (13)$$

rewriting (13) as

$$C^n_x - (\sigma_{n+1}^x)^2 \leq (\theta_{n+1}^x - c)^2. \quad (14)$$

Observe that (14) will be true for any value of $\theta_{n+1}^x$ if the left-hand side is negative, that is, if $C^n_x \leq (\sigma_{n+1}^x)^2$. If this is the case, however, it follows that

$$C^n_x \leq (\sigma_{n+1}^x)^2 \leq (\sigma^n_x)^2 \leq (\theta^n_x - c)^2 + (\sigma^n_y)^2,$$

whence

$$\min_y (\theta^n_y - c)^2 + (\sigma^n_y)^2 = C^n_x$$

and $\nu_{KG,n}^x = 0$, as required.

We derive a closed-form expression for $\nu_{KG,n}^x$ in the Appendix (Section 7). For now, we focus on the fact that it is possible to have $\sigma^n_x > 0$, meaning that we are uncertain about the true value of $x$, and yet $\nu_{KG,n}^x = 0$, suggesting that a single simulation of $x$ has no value. As a result, there may be a non-zero probability that $x$ will never be measured, which means that $\theta^n_x$ will not converge to the true value $\mu_x$.

**Theorem 2.1.** There exists a T&S problem for which (9) has a non-zero probability of never measuring a particular alternative.

**Proof:** Consider a problem with two alternatives, $x$ and $y$. For simplicity, take $\theta^n_x = \theta^n_y = c$ and choose $\lambda$ and $\sigma^0$ such that

$$(\sigma^0_y)^2 \leq (\sigma^0_x)^2.$$
It follows by (11) that $\nu_{x}^{KG,n} = 0$ and the KG method will prefer alternative $y$ to alternative $x$. Furthermore, if
\begin{equation}
(\theta_{y}^{n} - c)^2 + (\sigma_{y}^{n})^2 \leq (\sigma_{x}^{1})^2
\end{equation}
for all $n$, then alternative $y$ will always be preferred, and alternative $x$ will never be measured. Observe that
\begin{equation}
(\theta_{y}^{n} - c)^2 + (\sigma_{y}^{n})^2 \leq (\theta_{y}^{n} - c)^2 + (\sigma_{y}^{0})^2,
\end{equation}
so if
\begin{equation}
(\theta_{y}^{n} - c)^2 \leq (\sigma_{x}^{1})^2 - (\sigma_{y}^{0})^2
\end{equation}
holds for all $n$, then (15) will hold as well. We rewrite (16) as
\begin{equation}
-\sqrt{(\sigma_{x}^{1})^2 - (\sigma_{y}^{0})^2} \leq \theta_{y}^{n} - c \leq \sqrt{(\sigma_{x}^{1})^2 - (\sigma_{y}^{0})^2}
\end{equation}
and let $A$ be the event that (17) holds for all $n \geq 0$ under the KG policy.

Now, define a continuous-time process $B$ to be a Brownian motion with volatility $\sigma_{0}^{y}$ and initial value $B_{0} = 0$. This process should be interpreted as a continuous-time interpolation of the values $\theta_{y}^{n} - c$ that we would observe if we were to make $n = 0, 1, 2, \ldots$ measurements of $y$ starting at time 0. The random variable $\theta_{y}^{n} - c$ has the same distribution as $B_{t_{n}}$, where
\begin{equation}
t_{n} = \frac{\text{Var}(\theta_{y}^{n})}{(\sigma_{y}^{0})^2}.
\end{equation}
Recall that the conditional distribution of $\theta_{y}^{n+1}$ given $\theta_{y}^{n}$ is normal with mean $\theta_{y}^{n}$ and variance $(\tilde{\sigma}_{y}^{n})^2$. Also, $\theta_{y}^{n+1}$ is conditionally independent of $\theta_{y}^{n'}$ for $n' < n$ given $\theta_{y}^{n}$. The discrete-time processes $(\theta_{y}^{n} - c)_{n=0}^{\infty}$ and $(B_{t_{n}})_{n=0}^{\infty}$ are both Markov with the same distribution, and $\lim_{n \to \infty} \theta_{y}^{n} - c = \mu_{y} - c$ corresponds to $B_{1}$.

We can use the scaling properties of Brownian motion to write $B_{t} = \sigma_{0}^{y}W_{t}$ where $W_{t}$ is a Wiener process. It follows that
\begin{equation}
P(A) \geq P\left( -\sqrt{\frac{(\sigma_{x}^{1})^2 - (\sigma_{y}^{0})^2}{(\sigma_{y}^{0})^2}} \leq W_{t} \leq \sqrt{\frac{(\sigma_{x}^{1})^2 - (\sigma_{y}^{0})^2}{(\sigma_{y}^{0})^2}}, \ 0 \leq t \leq 1 \right).
\end{equation}
The right-hand side of (18) is simply the probability that, by time 1, the Wiener process has not yet exited the interval $[-d, d]$ where
\begin{equation}
d = \sqrt{\frac{(\sigma_{x}^{1})^2 - (\sigma_{y}^{0})^2}{(\sigma_{y}^{0})^2}}.
\end{equation}
In other words, if $T_{[-d,d]}$ is the exit time, then $P(A) \geq P(T_{[-d,d]} > 1)$. This probability is strictly positive; see Revuz & Yor (2005) or Borodin & Salminen (2012) for the distribution of $T_{[-d,d]}$. □

A consequence of Theorem 2.1 is that the KG policy is no longer guaranteed to discover the best alternative in the limit. Given that we measure $y$ infinitely often, but never measure $x$, the value $\mu_x$ will remain unknown and the conditional probability that $(\mu_x - c)^2 \leq (\mu_y - c)^2$ will be strictly positive. Although KG is just one of many possible algorithms, this result is somewhat surprising, since KG-type methods have been shown to be consistent in nearly every variant of R&S with normally distributed observations (many of these are surveyed in Powell & Ryzhov, 2012), to the point where consistency has become a kind of default sanity check. The fact that consistency is violated here motivates the design of new methods specifically for T&S, rather than the direct application of methods from the previous literature.

3 The local time method for T&S

Our approach to T&S is based on the following intuition. If we measure alternative $x$ infinitely often, we will have $\theta^n_x \rightarrow \mu_x$ a.s. by the strong law of large numbers. Thus, if $\mu_x$ is close to the target $c$, the sequence $(\theta^n_x)_{n=0}^\infty$ of estimates will also spend more time around $c$. We first create a continuous-time probabilistic interpolation of the sequence, which can be expressed as a time-changed Wiener process (Section 3.1). We then leverage the concept of Brownian local time as a natural way to measure closeness to a fixed target; Sections 3.2-3.3 show how the local time of the interpolation can be used to develop an efficient learning criterion for the discrete-time problem.

3.1 Brownian interpolation

We show how to construct the interpolation for a single alternative. The index $x$ is dropped from all quantities in this discussion, but we implicitly assume that a fixed $x$ is considered. The approximation procedure follows the work on approximating Gittins indices by Brezzi & Lai (2002), Yao (2006) and Chick & Gans (2009), but we give a brief summary.

First, let $X$ be a Brownian motion with unknown drift $\mu$ and known volatility $\lambda$. That is, $X$ is conditionally Brownian given $\mu$. We assume that $\mu \sim \mathcal{N}(\theta_0, s_0^2)$. Thus, the process $X$ is a probabilistic interpolation of the cumulative sum of observations made by simulating the alternative repeatedly. The increment $X_{t+1} - X_t$ is conditionally normal with mean $\mu$ and variance $\lambda^2$, just as
in our sampling model. Let $F_t = \sigma (X_u, 0 \leq u \leq t)$ be the filtration generated by the process. The conditional distribution of $\mu$ given $F_t$ is $\mathcal{N} (\theta_t, s_t^2)$ where

$$
\theta_t = \frac{s_0^{-2} \theta_0 + \lambda^{-2} X_t}{s_0^{-2} + \lambda^{-2} t},
$$

$$
s_t^2 = \left( s_0^{-2} + \lambda^{-2} t \right)^{-1}.
$$

This is the continuous-time analog of (1) and (2).

The process $(\theta_t)_{t \geq 0}$ is a continuous $F_t$-martingale. We calculate its quadratic variation as follows. For any given value of $\mu$, we can apply Itô’s formula and write

$$
d\theta_t = A_t dt + \frac{\lambda^{-2}}{s_0^{-2} + \lambda^{-2} t} dX_t,
$$

where $A_t = -\frac{\lambda^{-4}}{(s_0^{-2} + \lambda^{-2} t)^2} X_t$, and $W_t$ is a Wiener process. The decomposition in (19) has no third term because $\theta_t$ is linear in $X_t$. Observe that $\mu$ does not appear in the second term in (19), so the quadratic variation of $\theta_t$ is simply

$$
d(\theta_t, \theta_t) = \frac{\lambda^{-2}}{(s_0^{-2} + \lambda^{-2} t)^2} dt,
$$

with no dependence on $W_t$ or $\mu$. Integrating this expression yields

$$
\int \frac{\lambda^{-2}}{(s_0^{-2} + \lambda^{-2} t)^2} dt = -\frac{1}{s_0^{-2} + \lambda^{-2} t}.
$$

Applying Theorem III.4.6 of Karatzas & Shreve (1991), we can write $\theta_t = Z(\theta \theta_t)$, where $Z$ is another Wiener process. The negative sign in (20) indicates that $Z$ is a Wiener process moving backwards in time. That is, we let

$$
Z(v) = \theta_0 - \theta_t, \quad v = \frac{1}{s_0^{-2} + \lambda^{-2} t},
$$

and $Z_v$ defines a Wiener process in the $-v$ scale on the interval $[0, v_0]$, with $v_0 = s_0^2$ and $Z_{v_0} = 0$.

### 3.2 Connection to Brownian local time

For convenience, let us consider a Wiener process $W_t$ moving forward in time, with $W_0 = 0$. The local time of $W$ at zero is a process $L$ that increases only when $W_t = 0$. That is,

$$
L_{t-\varepsilon}(\omega) < L_{t+\varepsilon}(\omega) \quad \forall \varepsilon > 0 \iff W_t(\omega) = 0.
$$

(21)
The random variable $L_t$ is a measure of how frequently $W$ has visited zero by time $t$. In fact, the quantity $L_t$ is determined by the number of such visits. Because

$$Y_t = \max_{0 \leq u \leq t} W_u - W_t$$

has the same distribution as $|V_t|$, where $V$ is another Wiener process, it follows that the number of visits to zero by $W$ (which determines $L$) has the same distribution as the number of visits to zero by $Y$. However, $\max_{0 \leq u \leq t} W_u$ is also determined by the number of such visits (since it only increases at those times $t$ when $Y_t = 0$), leading to the result (Theorem VIII.5.15 of Cinlar, 2011) that $L_t$ has the same distribution as $\max_{0 \leq u \leq t} W_u$.

The process $L_t$ can be used to analyze the local time $L^a_t$ of $W$ at an arbitrary level $a > 0$. Similarly to (21), we have

$$L^a_{t-\varepsilon}(\omega) < L^a_{t+\varepsilon}(\omega) \forall \varepsilon > 0 \iff W_t(\omega) = a.$$

We have $L^a_t = 0$ if $a$ has never been visited by time $t$; otherwise, we have $L^a_t = L_{t-T_a}$ where $T_a$ is the hitting time of $a$. It then follows (p. 412 of Cinlar, 2011) that

$$P(L^a_t \in du) = \frac{2}{\sqrt{2\pi t}} e^{-\frac{(u+a)^2}{2t}} du, \quad u > 0.$$

**Proposition 3.1.** The expected local time at $a$ by time $t$ is

$$\mathbb{E}L^a_t = 2\sqrt{t} f\left(-\frac{a}{\sqrt{t}}\right)$$

where $f(z) = z\Phi(z) + \phi(z)$ and $\Phi, \phi$ are the standard normal cdf and pdf.

**Proof:** We calculate

$$\mathbb{E}L^a_t = 2 \int_0^\infty \frac{1}{\sqrt{2\pi t}} u e^{-\frac{(u+a)^2}{2t}} du = 2 \int_0^\infty \frac{1}{\sqrt{2\pi t}} (u + a) e^{-\frac{(u+a)^2}{2t}} du - 2a \int_0^\infty \frac{1}{2\pi t} e^{-\frac{(u+a)^2}{2t}} du$$

$$= \frac{1}{\sqrt{2\pi t}} \int_{a^2}^\infty e^{-\frac{v}{2t}} dv - 2a \Phi\left(-\frac{a}{\sqrt{t}}\right) = 2\sqrt{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{a^2}{2t}} - 2a \Phi\left(-\frac{a}{\sqrt{t}}\right) = 2\sqrt{t} f\left(-\frac{a}{\sqrt{t}}\right),$$

as required. □
We now relate this result to our continuous-time interpolation of the learning process in Section 3.1. Recall that $Z$ is a Wiener process moving backwards from time $v_0$ to time 0. Furthermore, $Z_v = \theta_0 - \theta_t$, so if $\theta_t = c$, then $Z_v = \theta_0 - c$. We are thus interested in the expected local time of $Z$ at $\theta_0 - c$ after a time interval of length $v_0$. The local time $L^0$ has the same distribution as $L^{-a}$ by the symmetry of the Wiener process, so the desired quantity is

$$
\ell = E[L^0_{v_0}] = 2\sqrt{v_0}f\left(-\frac{\theta_0 - c}{\sqrt{v_0}}\right) = 2s_0f\left(-\frac{\theta_0 - c}{s_0}\right). \tag{22}
$$

Larger values of $\ell$ suggest that the process $\theta$ will spend more time at the target level over an infinite time horizon.

### 3.3 The local time method

It remains to connect the analysis in Section 3.2 back to the targeting and selection problem. Intuitively, if $\mu_x$ is close to $c$, we should also see values of $\theta^n_x$ that are closer to $c$. At each time step $n$, we use the expected local time as a forecast of how closely $(\theta^n_x)_{n'=n}$ will approach $c$. We also use our time-$n$ beliefs about $x$ to initialize the Brownian approximation.

The local time method (LTM) makes simulation decisions by calculating

$$
X^{LTM,n}(\theta^n_x, \sigma^n_x) = \arg\max_x \ell^n_x, \tag{23}
$$

where

$$
\ell^n_x = 2\sigma^n_x f\left(-\frac{|\theta^n_x - c|}{\sigma^n_x}\right) \tag{24}
$$

is the relevant analog of (22). The LTM makes a tradeoff between $|\theta^n_x - c|$, our point estimate of the distance between $\mu_x$ and the target, and the uncertainty $\sigma^n_x$. The closed-form expression in (24) is computationally efficient and allows us to make simulation decisions quickly.

Surprisingly, the LTM formula in (24) exhibits structural parallels to the well-known expected improvement (EI) formula of Jones et al. (1998). The EI method calculates

$$
X^{EI,n}(\theta^n_x, \sigma^n_x) = \arg\max_x E^\theta_x\left[\max_y \left(\mu_x, \max_{y \neq x} \theta^n_x - \max_{y \neq x} \theta^n_y\right)\right] = \arg\max_x \sigma^n_x f\left(-\frac{|\theta^n_x - \max_{y \neq x} \theta^n_y|}{\sigma^n_x}\right). \tag{25}
$$

The KG policy of Frazier et al. (2008) has a similar form, but uses $\max_{y \neq x} \theta^n_y$ in (25). In R&S, these methods calculate the value of information for alternative $x$ by comparing its estimated value.
with some reference value. When the goal is to find the largest value, the reference is typically related to the largest of the estimated values, as in (25). In T&S, however, the appropriate reference is simply the target \( c \).

**Proposition 3.2.** The LTM formula has the equivalent representations

\[
\ell^n_x = 2E^n [\max (\mu_x, c) - \max (\theta^n_x, c)]
\]  

(26)

and

\[
\ell^n_x = E^n [|\mu_x - c| - |\theta^n_x - c|].
\]  

(27)

**Proof:** The representation in (26) follows directly from (25). To obtain (27), observe that

\[
E^n [\max (\mu_x, c) - \max (\theta^n_x, c)] = E^n [\min (\theta^n_x, c) - \min (\mu_x, c)]
\]

whence

\[
\ell^n_x = E^n [\max (\mu_x, c) - \min (\mu_x, c)] - [\max (\theta^n_x, c) - \min (\theta^n_x, c)]
\]

\[
= E^n [|\mu_x - c| - |\theta^n_x - c|],
\]

as required.

The representation in (27) is somewhat counter-intuitive. We would like to discover an alternative for which \( |\mu_x - c| \) is small, but the LTM actually prefers to measure alternatives for which \( |\mu_x - c| \) is greater (on average) than the point estimate \( |\theta^n_x - c| \). The intuition for this is formalized in Section 4.2, but we make the following brief comment. If none of the true values is equal to \( c \), the posterior density \( P^n (\mu_x \in dc) \) will converge to zero with more measurements of \( x \); that is, our estimate of \( \mu_x \) is “pushed away” from \( c \) as we learn more. However, this happens more slowly for alternatives that are probabilistically closer to \( c \). Thus, by attempting to push alternatives away from \( c \) as much as possible (maximizing \( E^n |\mu_x - c| \)), we will naturally discover and assign more effort to alternatives that are closer to the target.

Unlike EI, the LTM is a true index policy. The quantity \( \ell^n_x \) depends only on \( \theta^n_x \) and \( \sigma^n_x \), and does not involve our beliefs about other alternatives \( y \neq x \). This allows us to establish the following result for the LTM, which is not known to hold for EI methods.

**Proposition 3.3.** For each \( x \), the process \( (\ell^n_x)_{n=0}^{\infty} \) is a positive supermartingale.
**Proof:** We have $\ell^n_x \geq 0$ a.s. because $f$ is a strictly positive function. To show the supermartingale property, we can use Proposition 3.2. If $x^n \neq x$, then $\ell^{n+1}_x = \ell^n_x$ and the supermartingale property holds trivially at time $n$. If $x^n = x$, then

$$
\mathbb{E}_x^{n+1} \ell^n_x = 2\mathbb{E}_x^n (\mu_x, c) - 2\mathbb{E}_x^n (\theta^n_x, c) \\
= \ell^n_x + 2 \max (\theta^n_x, c) - 2 \max (\theta^n_x, c) - 2\bar{\sigma}_x^n f \left( -\frac{|\theta^n_x - c|}{\bar{\sigma}_x^n} \right) \\
\leq \ell^n_x.
$$

The first equality applies the tower property to (26), and the second follows by repeating the EI calculation with $\mu_x \sim \mathcal{N} \left( \theta^n_x, (\sigma_x^n)^2 \right)$ replaced by $\theta^n_x + 1 \sim \mathcal{N} \left( \theta^n_x, (\bar{\sigma}_x^n)^2 \right)$. The last line is again due to the positivity of $f$.

\[\blacksquare\]

## 4 Asymptotic behaviour of the local time method

The connection between the LTM and EI formulas suggests that the LTM is a more suitable representation of the value of information for T&S than the KG policy. In particular, the properties of the EI formula allow us to show that the LTM is statistically consistent (unlike KG), as well as characterize the sampling rates of the LTM. The latter result, discussed in Section 4.2, enables an interesting connection to another widely-studied class of methods known as optimal computing budget allocation (OCBA).

### 4.1 Consistency of the LTM

Earlier, we saw that the KG policy could violate consistency if it assigned a zero value to information about an alternative even though we were still uncertain about the value of the alternative. It is fairly straightforward to show that the LTM avoids this issue. All proofs in this section assume an infinite time horizon.

**Proposition 4.1.** For all $x$ and $n$, $\ell^n_x = 0$ if and only if $\sigma^n_x = 0$, and $\lim_{n \to \infty} \ell_n^x$ exists a.s.

**Proof:** The limit exists because every positive supermartingale converges. For all finite $z$, $f(z)$ is
strictly positive and finite, so \( \ell^n_x = 0 \) implies \( \sigma^n_x = 0 \). Furthermore, \( f'(z) = \Phi(z) \), whence
\[
\sigma^n_x f \left( -\frac{\theta^n_x - c}{\sigma^n_x} \right) \leq \sigma^n_x f(0),
\] (28)
so \( \sigma^n_x = 0 \) implies \( \ell^n_x = 0 \).

**Proposition 4.2.** The event that \( \ell^n_x \to 0 \) is equivalent to the event that \( x \) is measured infinitely often.

**Proof:** The process \((\theta^n_x)_{n=0}^\infty\) is a uniformly integrable martingale because \( \theta^n_x = \mathbb{E}^n \mu_x \). Therefore, \( \theta^n_x \) converges to a finite limit.

Let \( N_x = \sum_{n=0}^{\infty} 1_{\{x^n_x = x\}} \) be the number of times \( x \) is measured. If \( N_x(\omega) = \infty \), \( \sigma^n_x(\omega) \to 0 \) by (2), whence \( \ell^n_x(\omega) \to 0 \) by (28). If \( N_x(\omega) < \infty \), then \( \sigma^n_x(\omega) \to \sigma_x^\infty(\omega) \) where \( \sigma_x^\infty(\omega) > 0 \). Because \( f \) is continuous and strictly positive, and both \( \theta^n_x(\omega) \) and \( \sigma^n_x(\omega) \) converge to finite values, it follows that \( f \left( -\frac{\theta^n_x(\omega) - c}{\sigma^n_x(\omega)} \right) \) converges to some strictly positive limit. Therefore, \( \ell^n_x(\omega) \) also converges to a strictly positive limit. \( \square \)

**Theorem 4.1.** The event \( \bigcap_x \{N_x = \infty\} \) has probability 1 under the probability measure induced by the LTM.

**Proof:** We first remove an appropriate set of measure zero from the outcome space. We fix \( \omega \) and suppose that \( N_x(\omega) < \infty \) for some \( x \). Let \( y \) be another alternative for which \( N_y(\omega) = \infty \). By Proposition 4.2, \( \ell^n_x(\omega) \to \ell^\infty_x(\omega) \) with \( \ell^\infty_x(\omega) > 0 \), whereas \( \ell^n_y(\omega) \to 0 \). For some \( 0 < \varepsilon < \ell^\infty_x(\omega) \), we can find \( K_\varepsilon(\omega) \) such that \( \ell^n_y(\omega) < \frac{\varepsilon}{2} \) and \( |\ell^n_y(\omega) - \ell^\infty_x(\omega)| < \frac{\varepsilon}{2} \) for all \( n \geq K_\varepsilon(\omega) \).

It follows that \( \ell^n_y(\omega) < \ell^n_x(\omega) \) for all \( n \geq K_\varepsilon(\omega) \), which means that the LTM prefers \( x \) to \( y \) for all \( n \geq K_\varepsilon(\omega) \). It is therefore impossible for \( y \) to be measured i.o. while \( x \) is measured finitely many times on the sample path \( \omega \). We conclude that \( N_x(\omega) = \infty \) for all \( x \) and almost every \( \omega \). \( \square \)

**Theorem 4.2.** \( \lim_{N \to \infty} \mathbb{E}^{LT M} \min_x \mathbb{E}^N (\mu_x - c)^2 = \mathbb{E} \min_x (\mu_x - c)^2 \).

**Proof:** By Theorem 4.1, \( \theta^n_x \to \mu_x \) a.s. under the LTM. It follows that
\[
\min_x \mathbb{E}^N (\mu_x - c)^2 \to \min_x (\mu_x - c)^2
\]
almost surely. Furthermore, \( M^n_x = \mathbb{E}^n (\mu_x - c)^2 \) defines a uniformly integrable martingale, whence
\[
M^n = \min_x \mathbb{E}^n (\mu_x - c)^2
\] (29)
is a positive supermartingale. Furthermore, for any fixed $x$, $|M^n| \leq |M^n_x|$, so the supermartingale defined in (29) is uniformly integrable as well (see p. 75 of Cinlar, 2011). It follows that

$$M^n \to \min_x (\mu_x - c)^2$$

in $L^1$ as well. By Proposition III.4.7 of Cinlar (2011),

$$\lim_{N \to \infty} \mathbb{E}_{LTM}^{N} \min_x \mathbb{E}^N \left[ (\mu_x - c)^2 \right] = \mathbb{E}_{LTM}^{N} \min_x \mathbb{E}^N \left[ (\mu_x - c)^2 \right] = \mathbb{E} \min_x (\mu_x - c)^2,$$

where the superscript $LTM$ can be dropped from the last expectation because the true values $\mu_x$ do not depend on the policy.

Theorem 4.2 is a form of asymptotic optimality. Recall our objective function in (5). For any fixed policy $\pi$,

$$\mathbb{E}^{\pi} \min_x \mathbb{E}^N \left[ (\mu_x - c)^2 \right] \geq \mathbb{E}^{\pi} \mathbb{E}^N \min_x (\mu_x - c)^2 = \mathbb{E}^{\pi} \min_x (\mu_x - c)^2,$$

due to Jensen’s inequality for concave functions combined with the tower property of conditional expectation. Consequently, the asymptotic objective value of the LTM is precisely the global lower bound on the objective value of any policy.

4.2 Asymptotic sampling rates of the LTM

While the consistency of the LTM is reassuring, it only occurs in the limit and is easy to guarantee (e.g., by a round-robin sampling policy). In this section, we give a more detailed characterization of the asymptotic behaviour of the LTM by studying the rates at which different alternatives are sampled. Let $N^n_x = \sum_{n'=0}^{n-1} \mathbb{1}_{\{x' = x\}}$ be the number of simulations allocated to alternative $x$ by time $n$. The sampling ratio $\lim_{n \to \infty} \frac{N^n_x}{N^n_y}$ characterizes (if this limit exists) the rate at which $x$ is sampled relative to $y$.

In fact, the sampling ratios exist for all $x, y$. Their characterization can be found by exploiting the structural similarities between the LTM and EI formulas. Ryzhov (2015b, 2016) derives the sampling ratios for EI in R&S, and a straightforward modification of this analysis (using the target $c$ as the “reference value” in the EI formula) yields the following result.

**Theorem 4.3.** Under the LTM,

$$\lim_{n \to \infty} \frac{N^n_x}{N^n_y} = \frac{\lambda^2_x (\mu_y - c)^2}{\lambda^2_y (\mu_x - c)^2}$$

almost surely for any $x, y$. 

18
Thus, in the limit, each alternative receives a non-zero proportion of the simulation budget, but the relative magnitudes of the proportions are governed by (30). Alternatives receive more simulations if they have higher noise variance (i.e., we need more simulations to learn about them), but also if their true values are closer to the target (i.e., they are better).

The interpretation of the sampling ratios is perhaps more interesting than their precise algebraic form. It is known (Glynn & Juneja, 2004) that the sampling ratios are closely related to the convergence rates of offline learning algorithms. We now develop a similar connection. In this analysis, we adopt a frequentist view of the T&S problem, in which the true values \( \mu \) are treated as fixed. We assume that \( \mu_x \neq c \) for all \( x \), which holds w.p. 1 in the original Bayesian formulation of the problem. Our initial beliefs are given by \( \theta_0^x = 0 \) and \( \sigma_0^x = \infty \) for all \( x \). Consequently, the posterior update leads to the usual frequentist statistics

\[
\theta_n^x = \frac{1}{N_n^x} \sum_{n'=0}^{n-1} 1\{x_n'=x\} W_{x_n'}^{n'+1}, \quad (\sigma_n^x)^2 = \frac{\lambda_x^2}{N_n^x}.
\]

Consider an algorithm that assigns a fixed proportion \( p_x > 0 \) of the total simulation budget to alternative \( x \). Then, the total number of measurements allocated to \( x \) is \( p_x N \). Since we will consider an asymptotic analysis where \( N \to \infty \), we will treat this quantity as an integer throughout.

We now observe that, for any alternative \( x \),

\[
\lim_{N \to \infty} \frac{1}{N} \log P \left( \theta_N^x \in dc \right) = -G_x(p_x),
\]

where the rate function

\[
G_x(p_x) = \frac{(\mu_x - c)^2}{2\lambda_x^2} p_x.
\]

Consequently,

\[
\lim_{N \to \infty} \frac{1}{N} \log \max_x P \left( \theta_N^x \in dc \right) = -\min_x G_x(p_x).
\]

Below, we prove that the limiting ratios in (30) characterize the allocation \( p_1, ..., p_M \) that maximizes \( \min_x G_x(p_x) \), thus causing \( \max_x P \left( \theta_N^x \in dc \right) \) to vanish to zero at the fastest possible rate.

Before proving this statement, we first explain the intuition. For any \( x \), the likelihood that the sample mean \( \theta_N^x \) is equal to the target \( c \) vanishes to zero at an exponential rate (since \( \mu_x \neq c \)). However, if \( x \) is probabilistically closer to the target, this convergence should occur more slowly (that is, the exponent \( G_x \) should be smaller). We can speed up this convergence by increasing \( p_x \).
Then, by choosing $p_x$ to maximize the smallest of the exponents, we will allocate more simulations
alternatives that are closer to the target, which provides us with more precise estimates of their
values and helps us determine which of them has the closest value. To put it another way, if the
likelihood of equaling $c$ is reduced for the closest alternatives (those with the smallest rate functions
$G_x$), it will be reduced even more for the less suitable alternatives. Maximizing $\min_x G_x (p_x)$ will
then help us identify the closest alternatives.

**Theorem 4.4.** The allocation characterized by the ratios

$$
\frac{p_x}{p_y} = \frac{\lambda^2_x (\mu_y - c)^2}{\lambda^2_y (\mu_x - c)^2}
$$

(32)
is the optimal solution to the problem

$$
\begin{align*}
\max & \quad \min_x G_x (p_x) \\
\text{s.t.} & \quad \sum_x p_x = 1 \\
& \quad p_x \geq 0.
\end{align*}
$$

(33)

**Proof:** By observing that (31) is linear in $p_x$ and applying a standard technique for linearizing the
maximum of finitely many linear functions, (33) can be reformulated as a linear program. If we
can find $p_x, q_x \geq 0$ and $w, z \in \mathbb{R}$ satisfying

$$
\begin{align*}
\sum_x p_x &= 1, \quad z \leq G_x (p_x) \quad \text{for all } x, \\
\sum_x q_x &= 1, \quad w \geq G_x (q_x) \quad \text{for all } x, \\
q_x (z - G_x (p_x)) &= 0, \quad p_x (w - G_x (q_x)) = 0 \quad \text{for all } x,
\end{align*}
$$

(34)

then the values $p_x$ obtained from this system will be the optimal solution of (33). Since (33) is
feasible, the system must have a solution by strong duality.

Observe, however, that such a solution should satisfy $p_x > 0$ for all $x$, as $G_x (0) = 0$ for any
and $G_x (p) > 0$ for $p > 0$. The complementary slackness conditions in (34) then imply that
$G_x (q_x) = G_y (q_y)$ for all $x, y$. Since the values $q_x$ add up to 1, it follows that $q_x > 0$ for all $x$.
Applying (34) again, we find that $G_x (p_x) = G_y (p_y)$ for all $x, y$, yielding (32). □

We now discuss a connection between the LTM and the optimal computing budget allocation
(OCBA) methodology (Chen & Lee, 2010). OCBA can be thought of as an optimal deterministic
policy along the lines of the above analysis, where the proportions $p_x$ are designed to optimize some
objective of interest. In fact, the ratios in (32) are identical to those calculated in Chen et al. (2008)
for an OCBA method. Curiously, that method was not developed for a targeting problem; rather, it studies a version of R&S in which the goal is to find the alternatives with the $K$ highest values, rather than just the single best. The analytical approach, however, assumes that the decision-maker knows a fixed value $c$ that is assumed to be between the $K$th- and $(K+1)$st-highest true values. Essentially, this version of OCBA converts the subset selection problem into a targeting problem: by finding alternatives whose values are probabilistically close to $c$ and assigning more the budget to them, we are more likely to separate the $K$th and $(K+1)$st-best, as long as $c$ satisfies the main modeling assumption.

This parallel suggests a deeper connection between the LTM and OCBA. In the Appendix (Section 8), we use a large deviations analysis to characterize the theoretical budget allocation that optimizes the convergence rate of (6). Again, this approach is frequentist in nature; we assume that the true values $\mu_x$ are fixed and define $x^* = \arg\min_x (\mu_x - c)^2$ to be the true best alternative, with $\mu_* = \min_x (\mu_x - c)^2$ denoting its value and $p_*$ denoting the proportion of the budget assigned to it. It turns out that (6) converges at the optimal rate when

$$\frac{(|\mu_x - c| - |\mu_* - c|)^2}{\frac{\lambda_x^2}{p_*} + \frac{\lambda_x^2}{p_x}} = \frac{(|\mu_y - c| - |\mu_* - c|)^2}{\frac{\lambda_y^2}{p_*} + \frac{\lambda_y^2}{p_y}}$$

for any $x, y \neq x^*$. Additionally,

$$p_* = \lambda_* \sqrt{\sum_{x \neq x^*} \frac{p_x^2}{\lambda_x^2}}.$$  \hspace{1cm} (36)

In general, the LTM does not arrive at the rate-optimal asymptotic allocation. However, the theoretically optimal allocation is not practical to implement. Note that (35) depends on $|\mu_* - c|$, while (36) requires us to know $x^*$. However, if these were known, we would also know the solution to the T&S problem. Even if we use plug-in estimators to replace $\mu_x$ by $\theta^n_x$, and $x^*$ by $\arg\min_x (\theta^n_x - c)^2 + (\sigma^n_x)^2$, equations (35)-(36) represent a non-trivial computational challenge. For these reasons, recent literature on large deviations-based methods (Pasupathy et al., 2014) recommends using tractable approximations to the optimal allocation.

The LTM can be interpreted as one such approximation. Consider a limiting case where $\mu_* \rightarrow c$, i.e., there is one alternative that matches the target exactly (although it may still be the case that $\lambda_*>0$). Then, (30) implies that $p_* \rightarrow 1$ under the LTM, but the ratios $\frac{p_x}{p_y}$ from (30) remain fixed for $x, y \neq x^*$. It can be shown that these ratios are optimal for this limiting case (the proof is given in the Appendix).
Proposition 4.3. If $\mu_s = c$, the LTM ratios $\frac{p_x}{p_y} = \frac{\lambda^2_x(\mu_y-c)^2}{\lambda^2_y(\mu_x-c)^2}$ are optimal for $x, y \neq x^*$. 

By a straightforward extension of the analysis in Ryzhov (2016), it can be shown that, for a.e. $\omega$ and $x \neq x^*$, the mapping $n \mapsto N^n_x(\omega)$ is $O(\log n)$ in the limiting case where $\mu_s = c$ (from (30) we know that it is $O(n)$ in all other cases). Thus, in the limiting case, most of the budget is used to accurately estimate $\mu_s$ (as this determines whether or not the limiting case occurs), but the optimal ratios are preserved for all other alternatives.

In closing this section, we would like to emphasize that the LTM is designed for sequential learning, and thus may offer practical advantages over the standard way of implementing OCBA-like allocations in a sequential setting. In the OCBA literature, one typically first derives the desired sampling ratios in terms of $\mu$, and then implements a randomized policy where $p_x$ is directly approximated using plug-in estimators. Glynn & Juneja (2011) observes that such estimators may perform poorly since small errors in the estimated values may produce large errors in the estimated ratios. In the next section, we present numerical evidence in support of the LTM by comparing against the plug-in OCBA approach of Chen et al. (2008), which also converges to the limiting ratios but exhibits different short-term performance.

5 Experimental results

We present experimental results demonstrating the effectiveness of the LTM in several applications of T&S. Section 5.1 considers synthetic data designed to satisfy the modeling assumptions of this paper (for example, observations are normally distributed and the noise variance is known to all competing algorithms). The other examples explore settings where these assumptions do not hold, but the LTM formula (24) can still be applied to allocate the experimental budget. While these examples also use simulated data, the settings are motivated by practical applications: Section 5.2 considers an example motivated by the problem of efficiently matching players in competitive online gaming (Herbrich et al., 2006; Dangauthier et al., 2007), whereas Section 5.3 considers the problem of calibrating a manufacturing process (Lambert, 1992).
5.1 Experiments on synthetic data

We consider a synthetic problem with $M = 30$ alternatives. For each alternative $x$, the parameter $\lambda_x^2$ was generated from a uniform distribution on $[100, 150]$. The prior variance was set according to $(\sigma_0^x)^2 = 50$, and the prior mean was chosen as $\theta_0^x = 0$. We conducted 1000 independent simulation runs (macro-replications), each containing $N = 1000$ measurements and starting from the same set of priors. Each macro-replication uses a different set of true values $\mu_x$, generated from the prior distributions $\mathcal{N}\left(\theta_0^x, (\sigma_0^x)^2\right)$. These true values were used to generate observations, but were not revealed to any of the algorithms. Thus, the prior is accurate on average, but the decision-maker is not able to distinguish between the alternatives. The target was set to $c = 5$.

The values $\lambda_x^2$ were revealed to all competing algorithms. In this way, all of our modeling assumptions are satisfied and the experiments proceed as in Section 2.1. Common random numbers were used, so that two algorithms see the same value of $W_{x+1}$ if they both choose the same alternative $x$ at time $n$. The following policies were implemented:

**Local time method (LTM).** The LTM is defined by equations (23)-(24).

**Thompson sampling (Thompson).** The Thompson sampling policy has been the subject of recent attention (Chapelle & Li, 2011; Russo & Van Roy, 2014) because it is easy to generalize and can be guaranteed to possess desirable convergence rates in many settings. In T&S, we implement this policy as follows. Given $(\theta^n, \sigma^n)$, we generate a single sample $\hat{\mu}_x^n \sim \mathcal{N}\left(\theta_x^n, (\sigma_x^n)^2\right)$ for each $x$. We then choose $x^n$ to minimize $(\hat{\mu}_x^n - c)^2$. The use of random samples instead of sample averages or point estimates promotes exploration.

**Plug-in OCBA.** This policy is inspired by the plug-in approach typically used in the practical implementation of OCBA methods (Chen & Lee, 2010), which first derive a set of desired sampling ratios in terms of the true values $\mu$ and then sequentially approximate these ratios. Given $(\theta^n, \sigma^n)$ we randomly choose $x^n$ with $p^n_{x^n} = P^n(x^n = x)$ denoting the pmf. The probabilities satisfy

$$\frac{p^n_x}{p^n_y} = \frac{\lambda_x^2 (\theta_x^n - c)^2}{\lambda_y^2 (\theta_y^n - c)^2},$$

an approximation of (32) with the unknown true values replaced by plug-in estimates. By construction, this policy asymptotically achieves the same sampling ratios as the LTM, but may exhibit different short-term behaviour.
**DrawChance policy.** Given $(\theta^n, \sigma^n)$, the DrawChance policy chooses $x^n$ to maximize

$$P^n (W_x^{n+1} \in dc) = \frac{1}{\sqrt{2\pi (\sigma_x^n)^2 + \lambda_x^2}} e^{-\frac{(\theta^n - c)^2}{2[(\sigma_x^n)^2 + \lambda_x^2]}} dc,$$

the likelihood that the next observation of $x$ will be approximately equal to the target. The motivation for this policy originates from the application in Section 5.2 and will be discussed in more detail there.

**Greedy policy.** This simple greedy policy is given by $X^G (\theta^n, \sigma^n) = \arg \min_x \mathbb{E}^n [(\mu_x - c)^2]$.

**Knowledge gradient (KG).** This policy was discussed in Section 2.2; see the Appendix for the complete derivation. We included KG in our experimental study because it is a leading algorithmic approach in the simulation literature, and experiments allow us to see whether the theoretical issues shown in Theorem 2.1 are likely to have any practical impact.

Figure 1 presents several metrics of interest, aggregated over 1000 independent simulation runs starting from the same conditions. At each time step $n$, we calculate the selection decision $\bar{x}^n = \arg \min_x \mathbb{E}^n [(\mu_x - c)^2]$ and report the quantity $(\mu_{\bar{x}^n} - c)^2$, the true squared distance of the selected alternative from the target. Figure 1(a) shows the trajectory of these values over time, averaged across 1000 runs. Figure 1(b) shows the standard error of the squared distance across the same set of runs. Figure 1(c) simply reports the average value of $\mu_{\bar{x}^n}$ and compares this quantity to the target. Finally, Figure 1(d) reports the empirical probability that $\bar{x}^n = \arg \max_x \mu_x$, known as the “probability of correct selection” or PCS.

It is evident that only three of the policies perform well, namely the LTM, Thompson, and OCBA methods (the good performance of OCBA is not surprising, as it is designed to achieve the same asymptotic allocation as the LTM). Figure 1(c) shows that, after an initial period of heavy exploration lasting 10-15 measurements, these three methods find alternatives whose values are close to the target, and subsequently narrow in on $c$. However, while all three methods are able to match the target on average, the quality of the selected alternative is subject to significant variation across runs. Thus, in Figure 1(a), we find that LTM chooses alternatives that are closer to the target on average, and Figure 1(b) shows that this also happens more consistently (with lower standard error compared to the Thompson policy). Finally, Figure 1(d) indicates that the

---

1 Please note that this refers to the standard error of the random variable $(\mu_{\bar{x}^n} - c)^2$, not the standard error of the mean of this quantity; the latter can be found by dividing the standard errors in Figure 1(b) by 1000. Our goal is to quantify the possible variation in the performance of the policy within a single run.
LTM also leads to a higher probability of selecting the single best alternative.

The other three policies tend to get “stuck” on suboptimal alternatives very early in the time horizon (leading to very low PCS). The KG policy achieves the worst performance (even worse than the greedy policy), indicating that looking ahead to the next observation simply does not provide enough information to ensure a sufficient degree of exploration. The DrawChance policy also gets stuck, but does so a bit later than the greedy policy.

5.2 Application to competitive online gaming

This example is motivated by the problem of automated match-making in a competitive online gaming service. While our experiments use simulated data, the underlying learning model is drawn from Herbrich et al. (2006) and Dangauthier et al. (2007), which describe a large-scale practical
implementation on Microsoft’s Xbox Live platform. The setting does not satisfy all of the modeling assumptions from Section 2.1; in particular, observations in this problem are heavily non-normal. Nonetheless, the problem bears enough similarity to T&S that the methods considered in Section 5.1 can all be easily adapted and implemented here. In our opinion, the match-making example is made more interesting by its departures from T&S, as the performance of T&S algorithms here may provide insight into their robustness against their own assumptions.

We assume that player $x$ has a “skill level” $\mu_x$, an unknown but stationary unitless quantity. When $x$ plays a game, a performance value $W_x \sim \mathcal{N}(\mu_x, \lambda^2)$ is generated independently of any other random variables in the problem. However, $W_x$ is never observed by the game master. Instead, if $x$ wins a game against $y$, this is interpreted as a signal that $W_x > W_y$; conversely, $x$ loses when $W_x < W_y$. These are the only two possible outcomes (i.e., either $x$ or $y$ must win). The game master uses the observed outcomes to learn players’ skill values and create games that are more evenly matched (by selecting opponents with similar skills).

Dangauthier et al. (2007) proposes the following approximate Bayesian learning model. We begin with independent priors $\mu_x \sim \mathcal{N}\left(\theta_x^0, (\sigma_x^0)^2\right)$. Suppose that $x$ and $y$ are matched in the $(n+1)$st game, and produce performance values $W_x^{n+1}, W_y^{n+1}$. Our beliefs about player $x$ are updated recursively using the equations

$$
\theta_x^{n+1} = \begin{cases} 
\theta_x^n + \frac{(\sigma_x^n)^2}{\sqrt{(\sigma_x^n)^2 + (\sigma_y^n)^2 + 2\lambda^2}} \cdot v \left( \frac{\theta_x^n - \theta_y^n}{\sqrt{(\sigma_x^n)^2 + (\sigma_y^n)^2 + 2\lambda^2}} \right) & \text{if } W_x^{n+1} > W_y^{n+1}, \\
\theta_x^n - \frac{(\sigma_x^n)^2}{\sqrt{(\sigma_x^n)^2 + (\sigma_y^n)^2 + 2\lambda^2}} \cdot v \left( \frac{\theta_x^n - \theta_y^n}{\sqrt{(\sigma_x^n)^2 + (\sigma_y^n)^2 + 2\lambda^2}} \right) & \text{if } W_x^{n+1} < W_y^{n+1},
\end{cases}
$$

(37)

and

$$
(\sigma_x^{n+1})^2 = \begin{cases} 
(\sigma_x^n)^2 \left( 1 - \frac{(\sigma_y^n)^2}{(\sigma_x^n)^2 + (\sigma_y^n)^2 + 2\lambda^2} \right) \cdot w \left( \frac{\theta_x^n - \theta_y^n}{\sqrt{(\sigma_x^n)^2 + (\sigma_y^n)^2 + 2\lambda^2}} \right) & \text{if } W_x^{n+1} > W_y^{n+1}, \\
(\sigma_x^n)^2 \left( 1 - \frac{(\sigma_y^n)^2}{(\sigma_x^n)^2 + (\sigma_y^n)^2 + 2\lambda^2} \right) \cdot w \left( \frac{\theta_x^n - \theta_y^n}{\sqrt{(\sigma_x^n)^2 + (\sigma_y^n)^2 + 2\lambda^2}} \right) & \text{if } W_x^{n+1} < W_y^{n+1},
\end{cases}
$$

(38)

where

$$
v(z) = \frac{\phi(z)}{\Phi(z)},
$$

$$
w(z) = v(z) (v(z) + z).
$$

For the opponent $y$, we apply (37)-(38) with $x$ and $y$ reversed. The updating equations are derived using the moment-matching technique (Ryzhov, 2015a). This learning model is an approximation,
as the normal prior is not conjugate with the binary observation. Essentially, the model forces the posterior to be normal, and (37)-(38) sets the parameters of that artificial normal distribution in a way that optimally approximates the exact, non-normal posterior. In addition, the model does not consider correlations in our beliefs about \( x \) and \( y \) that may be induced by the observation. However, although the model relies on approximations, it has shown itself to be valuable in practice. Here, we do not attempt to analyze the approximations in the learning model; for our purposes, it is interesting precisely because it violates our modeling assumptions from Section 2.1, thus allowing us to test the robustness of the LTM with respect to those assumptions.

The experiments are set up as follows. There are \( M = 30 \) players. We set \( \theta_n^x = 0 \) for all \( x \), while \( (\sigma_n^x)^2 \) is drawn from a uniform distribution on \([100, 150]\). The noise of the performance values was set to \( \lambda^2 = 100 \). We conducted 2000 independent simulation runs (macro-replications), each of which contained \( N = 1000 \) games and used a different set of true skills \( \mu_x \) generated from the prior. For the \( n \)th game in a given macro-replication, a “contestant” player \( y_n^x \) is first selected uniformly at random (representing a user who has submitted a new query to the service). The game master then selects an opponent \( x_n^x \) based on the current beliefs \( (\theta_n^x, \sigma_n^x) \), performance values are simulated, and the beliefs are updated. The match is better if the difference \( (\mu_{x_n} - \mu_{y_n})^2 \) is smaller.

We adapted the same policies from Section 5.1 to this problem. The \( n \)th game can be interpreted as an instance of T&S where the unknown value \( \mu_{x_n} - \mu_{y_n} \) is matched with the target \( c = 0 \). Thus, the LTM becomes

\[
X^{LTM,n}(\theta^n, \sigma^n) = \arg \max_x 2 \sqrt{\frac{(\sigma_n^x)^2 + (\sigma_n^y)^2}{(\sigma_n^x)^2 + (\sigma_n^y)^2}} \frac{|\theta_n^x - \theta_n^y|}{\sqrt{(\sigma_n^x)^2 + (\sigma_n^y)^2}} \]

that is, the Brownian approximation is applied to the process \( (\theta_n^{x'} - \theta_n^{y'})_{n' = n}^\infty \) with \( y_n^x \) fixed. The other policies are adapted similarly. In particular, we note that the DrawChance policy

\[
X^{DC,n}(\theta^n, \sigma^n) = \arg \max_x \frac{1}{\sqrt{2\pi \left( (\sigma_n^x)^2 + (\sigma_n^y)^2 + 2\lambda^2 \right)}} e^{-\frac{(\sigma_n^y)^2}{2\left( (\sigma_n^x)^2 + (\sigma_n^y)^2 + 2\lambda^2 \right)}}
\]

is equivalent to the algorithm originally proposed by Herbrich et al. (2006) for this problem. Since the game master wishes to promote competitive games, it is intuitive to choose the opponent to maximize the likelihood of a “draw.” In this way, we are comparing against a benchmark that has previously been implemented in practice.
Figure 2: Experimental results for match-making problem.

Figure 2 shows how several metrics of game quality evolve over time. Figure 2(a) reports the values \((\mu_x^n - \mu_y^n)^2\) averaged over 2000 macro-replications, while Figure 2(b) reports the true likelihood of a draw, given by \(\frac{1}{\sqrt{4\pi\lambda^2}}e^{-\frac{(\mu_x^n - \mu_y^n)^2}{4\lambda^2}}\). In Figure 2(c), we first calculate the “win rate” for each player (that is, the proportion of games won by that player, out of all games where that player participated) and average these rates across all players. Figure 2(d) reports the absolute deviation of these averaged win rates from 0.5, the idea being that the game master would prefer all players to win exactly 50% of their games. The KG policy is omitted from these figures, as we found that it significantly underperformed every other policy.

In Figure 2(c), we see that multiple policies achieve the desired average win rate of 0.5; however, Figure 2(d) shows that these policies produce different levels of variation in the win rates of individual players. The Thompson and OCBA policies achieve near-perfect win rates on average, but are
far less effective at consistently selecting opponents with similar skill levels (Figures 2(a)-2(b)) in any given match. Overall these two policies are much less effective in this example than in Section 5.1, where they provided the closest competition to the LTM.

On other hand, the DrawChance and greedy policies, which were ineffective in Section 5.1, now perform very well, although the LTM still offers some improvement in every metric of interest (closer skill levels, higher draw likelihoods, and win rates that are consistently closer to 0.5). The key point, however, is not simply that the LTM achieves the best performance, but also that it does so consistently in both the synthetic example of Section 5.1 as well as here. The other policies perform well in one of these examples, but not the other. This can be viewed as a form of robustness of the LTM with respect to the modeling assumptions of the T&S problem.

5.3 Application to manufacturing

Lambert (1992) proposes the following zero-inflated Poisson (ZIP) model for predicting the number of defective items produced by a manufacturing process. Let \( X \) and \( Y \) be independent random variables such that \( X \) follows a Bernoulli distribution with success probability \( \rho \), and \( Y \) follows a Poisson distribution with parameter \( \zeta \). Now, define \( W = (1 - X)Y \), so that \( X = 1 \) implies \( W = 0 \). Thus, \( W \) is non-normal and may have a bimodal pmf (Figure 3(a)).

We now introduce a control \( x \) and suppose that \( \zeta(x) = e^{\beta_0 + \beta_1 x} \) and \( \rho(x) = \frac{1}{1 + \zeta(x)} \). For our example, suppose that \( \beta_0 = 1.5, \beta_1 = -2, \) and \( x \) can take on \( M = 30 \) uniformly spaced values between 0 and 1. Figure 3(b) shows the means and variances of the ZIP distribution for each
of these values of $x$. Suppose that the decision-maker wishes to find $x$ that makes the average number of defects match the target $c = 0.75$ as closely as possible (i.e., it is too expensive to reduce the number of defects below this value, but undesirable to let it go higher). The decision-maker does not know the exact relationship between $x$ and the mean, or even that the output follows a ZIP distribution. However, noisy observations from the distribution can be obtained, e.g., from expensive field tests.

We apply the T&S framework to this problem. The learning model from Section 2.1, and all of the algorithms from Section 5.1, can be implemented as in standard T&S (i.e., with the implicit assumption that the output is normal), but the actual observations $W_{x_n+1}$ can be drawn from the appropriate ZIP model. However, the Bayesian updating equations assume that the noise variance $\lambda^2_x$ is known, which is not the case in the problem we are considering. In this example, whenever we apply the learning model and algorithms, we replace $\lambda^2_x$ by a frequentist plug-in estimator $\hat{\lambda}_n^x$, set equal to the empirical standard error of all observations of $x$ collected up to time $n$, with the default value $\left(\hat{\lambda}_n^x\right)^2 = 10$ if $N_n^x < 2$. Note that none of the algorithms we have discussed explicitly handles the fact that the variance is unknown; however, the plug-in method adopted here is standard in the general simulation optimization literature (Jones et al., 1998; Chen & Lee, 2010) as well as the literature on Brownian approximations for learning problems (Brezzi & Lai, 2002; Chick & Gans, 2009). For our purposes, this approach allows us to test the sensitivity of the policies to the known-variance assumption.

We ran each policy for $N = 1000$ iterations, repeated for 1000 macro-replications starting from the same prior $\theta_0^x \equiv 1$, $\left(s_0^x\right)^2 \equiv 2$ (representing a setting where the decision-maker has little prior information about the alternatives other than a rough range for their values). As in Section 5.1, common random numbers were used to promote more reliable comparisons between policies. The results are summarized in Figure 4; we found that, in this example, only the LTM, Thompson and OCBA policies performed well, and the other algorithms are omitted for brevity of the exposition.

As before, the LTM achieves a lower average distance from $c$, and does so more consistently (achieving lower standard errors by the end of the time horizon), due to greater exploration in the early iterations (Figure 4(c)). The most notable difference from the synthetic example is that the Thompson and OCBA policies experience some performance degradation with respect to PCS (relative to the LTM). The need to estimate the noise variance, combined with the randomized nature of these policies, appears to make it more difficult to identify the best alternative. By
contrast, the LTM appears to be more robust against the noise-variance assumption.

Our experiments indicate that, while the LTM performs well in “standard” T&S problems, it retains its advantages in related problem classes where the normality and known-variance assumptions do not hold. The other policies appear to be more sensitive: thus, the DrawChance and greedy policies perform well in Section 5.2, but not in the other examples, whereas the opposite is true of the Thompson and OCBA policies. In our view, a key point in favour of the LTM is its ability to consistently achieve the best performance in all of these examples.
6 Conclusion

We have proposed a framework for a class of simulation optimization problems where the goal is to identify a system whose performance value closely matches a given target. This “targeting and selection” problem can be formulated with a simple modification of the more common ranking and selection problem; nonetheless, it may not be amenable to the direct application of R&S methods such as the knowledge gradient.

We have also proposed a fundamentally new approach for T&S based on the Brownian local time of a continuous-time interpolation of the sequence of estimated values for each alternative. The algorithm resulting from this argument is computationally efficient and exhibits parallels to both EI and OCBA methodologies. The asymptotic sampling ratios achieved by this method match those of a recently proposed OCBA policy, and can be viewed as an approximation of the theoretically optimal rates. One important practical advantage of the LTM is its apparent insensitivity to violations in the underlying modeling assumptions, as shown numerically in our experiments. We believe that the local time argument is a novel and potentially useful idea for sequential selection problems that involve targeting.

References


Hartmann, M. (1991), ‘An improvement on Paulson’s procedure for selecting the population with the largest mean from k normal populations with a common unknown variance’, Sequential Analysis 10(1-2), 1–16.


Ryzhov, I. O. (2016), ‘On the convergence rates of expected improvement methods’, *Submitted for publication*.


### 7 Appendix: derivation of KG for targeting and selection

We have already seen that \( \nu_x^{KG,n} = 0 \) if (11) holds. Again, we define

\[
C^n_x = \min_{y \neq x} \left( \theta^n_y - c \right)^2 + \left( \sigma^n_y \right)^2.
\]

and consider the case where \( C^n_x \geq (\sigma^{n+1}_x)^2 \). We will compute \( \nu_x^{KG,n} \) in closed form by referring to the following fact from Johnson & Kotz (1969) on the incomplete second moment of the standard normal distribution.

**Lemma 7.1.** For \( Z \sim \mathcal{N}(0,1) \) and \( z \geq 0 \), \( E \left( Z^2 1_{\{Z \leq z\}} \right) = \Phi(z) - z\phi(z) \).

We calculate

\[
E^n_x \min \left( C^n_x, \left( \theta^{n+1}_x - c \right)^2 + \left( \sigma^{n+1}_x \right)^2 \right) = C^n_x P \left( \left( \theta^{n+1}_x - c \right)^2 \geq C^n_x - (\sigma^{n+1}_x)^2 \right) \\
+ (\sigma^{n+1}_x)^2 P \left( \left( \theta^{n+1}_x - c \right)^2 \leq C^n_x - (\sigma^{n+1}_x)^2 \right) \\
+ E \left( \left( \theta^{n+1}_x - c \right)^2 1_{\{ \left( \theta^{n+1}_x - c \right)^2 \leq C^n_x - (\sigma^{n+1}_x)^2 \}} \right) \tag{39}
\]

Because \( \theta^{n+1}_x \sim \mathcal{N} \left( \theta^n_x, (\tilde{\sigma}_x^n)^2 \right) \) given \( \mathcal{F}^n \), we can write

\[
P \left( \left( \theta^{n+1}_x - c \right)^2 \leq C^n_x - (\sigma^{n+1}_x)^2 \right) = \Phi(d_2) - \Phi(d_1)
\]
where
\[
\begin{align*}
    d_1 &= \frac{(c - \theta_n^x) - \sqrt{C_n^x - (\sigma_x^{n+1})^2}}{\tilde{\sigma}_n^x}, \\
    d_2 &= \frac{(c - \theta_n^x) + \sqrt{C_n^x - (\sigma_x^{n+1})^2}}{\tilde{\sigma}_n^x}.
\end{align*}
\]

The last term of (39) can be rewritten as
\[
\int_{d_1}^{d_2} ((\theta_n^x - c) + \tilde{\sigma}_n^x z)^2 \phi(z) dz = (\theta_n^x - c)^2 (\Phi(d_2) - \Phi(d_1))
+ 2\tilde{\sigma}_n^x (\theta_n^x - c) \int_{d_1}^{d_2} z \phi(z) dz
+ (\tilde{\sigma}_n^x)^2 \int_{d_1}^{d_2} z^2 \phi(z) dz.
\tag{40}
\]

We use Lemma 7.1, together with the fact that \(\int_{-\infty}^{\infty} z \phi(z) dz = \phi(z)\), to simplify (40) and obtain
\[
\int_{d_1}^{d_2} ((\theta_n^x - c) + \tilde{\sigma}_n^x z)^2 \phi(z) dz = (\theta_n^x - c)^2 (\Phi(d_2) - \Phi(d_1))
+ 2\tilde{\sigma}_n^x (\theta_n^x - c) (\phi(d_1) - \phi(d_2))
+ (\tilde{\sigma}_n^x)^2 (\Phi(d_2) - \Phi(d_1) + d_1 \phi(d_1) - d_2 \phi(d_2)).
\tag{41}
\]

Because \((\sigma_x^{n+1})^2 + (\tilde{\sigma}_n^x)^2 = (\sigma_x^n)^2\), we can combine terms in (39) and (41) to obtain
\[
\mathbb{E}_x^n \min \left( C_n^x, (\theta_x^{n+1} - c)^2 + (\sigma_x^{n+1})^2 \right) = C_n^x [1 - (\Phi(d_2) - \Phi(d_1))]
+ \left[ (\theta_n^x - c)^2 + (\sigma_x^n)^2 \right] \cdot (\Phi(d_2) - \Phi(d_1))
+ 2\tilde{\sigma}_n^x (\theta_n^x - c) (\phi(d_1) - \phi(d_2))
+ (\tilde{\sigma}_n^x)^2 (d_1 \phi(d_1) - d_2 \phi(d_2)).
\]

With some algebra, the last two terms can be combined to yield
\[
\mathbb{E}_x^n \min \left( C_n^x, (\theta_x^{n+1} - c)^2 + (\sigma_x^{n+1})^2 \right) = C_n^x [1 - (\Phi(d_2) - \Phi(d_1))]
+ \left[ (\theta_n^x - c)^2 + (\sigma_x^n)^2 \right] \cdot (\Phi(d_2) - \Phi(d_1))
+ (\tilde{\sigma}_n^x)^2 (d_1 \phi(d_1) - d_2 \phi(d_2)).
\tag{42}
\]

The KG factor is computed by subtracting (42) from \(\min_y (\theta_y^n - c)^2 + (\sigma_y^n)^2\).
8 Derivation of the rate-optimal allocation for T&S

Define $x^* = \arg\min_x (\mu_x - c)^2$ to be the true best alternative, with value $\mu_*$ and allocation $p_*$. We make an incorrect selection decision at time $N$ if

$$(\theta^*_N - c)^2 + \frac{\lambda^2}{p_* N} > \min_{x \neq x^*} (\theta^*_x - c)^2 + \frac{\lambda^2_x}{p_x N}.$$ 

Let $E$ denote the event that this happens. Then, following Glynn & Juneja (2004),

$$(M - 1) \max_{x \neq x^*} P \left( (\theta^*_N - c)^2 + \frac{\lambda^2}{p_* N} > (\theta^*_x - c)^2 + \frac{\lambda^2_x}{p_x N} \right) \leq P(E)$$

and

$$P(E) \leq \max_{x \neq x^*} P \left( (\theta^*_N - c)^2 + \frac{\lambda^2}{p_* N} > (\theta^*_x - c)^2 + \frac{\lambda^2_x}{p_x N} \right).$$

Therefore, if there exists a function $G_x$ satisfying

$$\lim_{N \to \infty} \frac{1}{N} \log P \left( (\theta^*_N - c)^2 + \frac{\lambda^2}{p_* N} > (\theta^*_x - c)^2 + \frac{\lambda^2_x}{p_x N} \right) = -G_x(p_*, p_x),$$

it follows that

$$\lim_{N \to \infty} \frac{1}{N} \log P(E) = - \min_{x \neq x^*} G_x(p_*, p_x).$$

We can then solve an optimization problem to optimize the rate at which $P(E)$ vanishes to zero. In order to derive $G_x$, we require the limiting rate function of the pair $((\theta^*_N - c)^2 + \frac{\lambda^2}{p_* N}, (\theta^*_x - c)^2 + \frac{\lambda^2_x}{p_x N})$. The first step is to find the log-mgf (moment-generating function) of this pair.

Due to the independence of the alternatives, the log-mgf of the pair is the sum of the individual log-mgfs of $x^*$ and $x$. It is then sufficient to derive the log-mgf $\Psi_N$ of $((\theta^*_N - c)^2 + \frac{\lambda^2}{p N})$ for a generic alternative with mean $\mu$, noise variance $\lambda^2$, and allocation $p$. Because the quantity of interest is a nonlinear function of the sample average $\theta^N$, we cannot directly apply the derivations in Glynn & Juneja (2004). Instead, we directly verify Assumption 2.3.2 in Dembo & Zeitouni (2009) by deriving a function $\Psi$ satisfying $\Psi(\gamma) = \lim_{N \to \infty} \frac{1}{N} \Psi_N(\gamma N)$.

**Proposition 8.1.** The limiting log-mgf is given by

$$\Psi(\gamma) = \frac{\gamma}{1 - 2\sigma^2/p} (\mu - c)^2.$$
Proof: We first write
\[
\Psi_N(\gamma) = \log \mathbb{E} \left( e^{\gamma \gamma^N - c} \right) = \gamma \frac{\lambda^2}{p N} + \log \mathbb{E} \left( e^{\gamma \gamma^N - c} \right),
\]
whence
\[
\lim_{N \to \infty} \frac{1}{N} \Psi_N(\gamma N) = \lim_{N \to \infty} \frac{1}{N} \gamma \frac{\lambda^2}{p} + \lim_{N \to \infty} \frac{1}{N} \log \mathbb{E} \left( e^{\gamma \gamma^N - c} \right),
\]
assuming that the limit of the second term exists. The first term clearly vanishes to zero, so we focus on the second term.

Observe that, under the frequentist view, \( \theta^N_x \) follows a normal distribution with mean \( \mu \) and variance \( \frac{\lambda^2}{p N} \). Consequently, the random variable
\[
Y = \frac{\left( \theta^N_x - c \right)^2}{\lambda^2} p N
\]
follows a non-central chi-squared distribution with parameter \( \frac{(\mu - c)^2}{\lambda^2} p N \) and 1 degree of freedom. We then apply the mgf of this distribution to obtain
\[
\mathbb{E} \left( e^{\gamma^N (\theta^N_x - c)^2} \right) = \mathbb{E} \left( e^{\gamma \frac{\lambda^2}{2} Y} \right) = \frac{1}{\sqrt{1 - 2 \frac{\lambda^2}{p}}} e^{\gamma \frac{N}{1 - 2 \frac{\lambda^2}{p}} (\mu - c)^2},
\]
whence
\[
\log \mathbb{E} \left( e^{\gamma^N (\theta^N_x - c)^2} \right) = -\frac{1}{2} \log \left( 1 - 2 \frac{\lambda^2}{p} \right) + \frac{\gamma N}{1 - 2 \frac{\lambda^2}{p}} (\mu - c)^2,
\]
leading to the result
\[
\lim_{N \to \infty} \frac{1}{N} \Psi_N(\gamma N) = \frac{\gamma}{1 - 2 \frac{\lambda^2}{p}} (\mu - c)^2,
\]
as required. \( \square \)

The next task is to find the Fenchel-Legendre transform \( I(u) = \sup_{\gamma} \gamma u - \Psi(\gamma) \) for the generic alternative. We can then apply the Gärtner-Ellis theorem (Thm. 2.3.6 in Dembo & Zeitouni, 2009) to these functions to obtain the rates \( G_x \).

**Proposition 8.2.** The Fenchel-Legendre transform of \( \Psi \) is given by
\[
I(u) = \frac{p}{2 \lambda^2} (\sqrt{u} - |\mu - c|)^2.
\]
Proof: From Proposition 8.1,
\[ I(u) = \sup_{\gamma} \gamma u - \frac{\gamma}{1 - 2^{2\lambda^2} \mu^2} (\mu - c)^2. \]

For fixed \( u \), we differentiate the expression inside the supremum and set it equal to zero, which yields the equation
\[ u = \frac{(\mu - c)^2}{\left(1 - 2^{2\lambda^2} \mu^2 \right)^2}, \]
whence
\[ \gamma^* = \frac{p}{2\lambda^2} \left( 1 - \sqrt{\frac{(\mu - c)^2}{u}} \right) \]
and
\[ I(u) = \frac{p}{2\lambda^2} \left( u - 2\sqrt{u(\mu - c)^2 + (\mu - c)^2} \right) = \frac{p}{2\lambda^2} (\sqrt{u} - |\mu - c|)^2, \]
as required. \( \square \)

Theorem 8.1. The rate function \( G_x \) for \( x \neq x^* \) is given by
\[ G_x(p_x, p_x) = \frac{\left(|\mu_x - c| - |\mu^* - c|\right)^2}{2 \left( \frac{\lambda_x^2}{p_x} + \frac{\lambda^2}{p_x} \right)}. \] (44)

Furthermore, the simulation allocation that optimizes the rate of convergence of \( P(E) \) is the solution to the system of equations given by
\[ \frac{\left(|\mu_x - c| - |\mu^* - c|\right)^2}{\frac{\lambda_x^2}{p_x} + \frac{\lambda^2}{p_x}} = \frac{\left(|\mu_y - c| - |\mu^* - c|\right)^2}{\frac{\lambda_y^2}{p_y} + \frac{\lambda^2}{p_y}} \]
for any \( x, y \neq x^* \). Additionally,
\[ p_x = \lambda^* \sqrt{\sum_{x \neq x^*} \frac{p^2}{\lambda^2}}. \]

Proof: By the Gärtner-Ellis theorem,
\[ G_x(p_x, p_x) = \inf_{u_x \geq u_x \geq 0} I_x(u_x) + I_x(u_x), \]
where \( I_x, I^*_x \) are obtained by plugging the parameters of \( x^* \) and \( x \) into (43). From this point on, the arguments of Glynn & Juneja (2004) can be applied, since they hold under general assumptions. First, it can be shown that
\[ G_x(p_x, p_x) = \inf_{u \geq 0} I_x(u) + I_x(u). \] (45)
The minimizer of (45) is given by
\[ u^* = \left( \frac{p_* \lambda^2}{2} \left| \frac{\mu_*}{\lambda^2} \right| + \frac{p_* \lambda^2}{2} \left| \mu_* - c \right| \right)^2, \]
which yields (44). The optimal allocation then follows by Example 1 of Glynn & Juneja (2004). □

Finally, we prove Proposition 4.3 and show that, in the limiting case where \( \mu_* = c \), the optimal allocation will preserve the LTM ratios derived in (30) for suboptimal alternatives.

**Proof of Proposition 4.3:** Suppose that \( \mu_* = c \). Then, from Proposition 8.1, it follows that \( \Psi_* (\gamma) = 0 \), whence
\[
I_* (u) = \begin{cases} 
0 & u = 0 \\
\infty & u > 0.
\end{cases}
\]
Then, the infimum in (45) is achieved by letting \( u = 0 \), yielding
\[
G_x (p_*, p_x) = I_x (0) = \frac{(\mu_x - c)^2}{2 \lambda_x^2} p_x,
\]
which is identical to (31). The result then follows by Theorem 4.4. □