Balancing optimal large deviations in sequential selection

Ye Chen Ilya O. Ryzhov

January 24, 2019

Abstract

In the ranking and selection problem, a sampling budget is allocated among a finite number of designs with the goal of efficiently identifying the best. Allocations of this budget may be static (with no dependence on the random values of the samples) or adaptive (decisions are made based on the results of previous decisions). A popular methodological strategy in the simulation literature is to first characterize optimal static allocations, by using large deviations theory to derive a set of optimality conditions, and then to use these conditions to guide the design of adaptive allocations. We propose a new methodology that can be guaranteed to adaptively learn the solution to these optimality conditions in a computationally efficient manner, without any tunable parameters, and under general assumptions on the sampling distribution.

1 Introduction

This paper studies ranking and selection, a mathematical framework for the formal study of information collection that drives many methodological developments in simulation and stochastic optimization. In this framework, there are $M > 2$ alternatives (“systems,” “populations,” etc.), and for any given alternative $x = 1, ..., M$, we have the ability to collect independent random samples from a common distribution $F_x$ with mean $\mu_x$, where we suppose for convenience that $\mu_x \neq \mu_y$ for $x \neq y$. The values $\mu_x$ are unknown to us, but may be inferred from the samples. Our goal is to identify, with high probability, the “best” alternative $x^* = \arg \max_x \mu_x$ after a learning period in which $N$ samples of various alternatives are collected. Applications arise when simulation is used to compare, e.g., configurations of manufacturing systems (Boesel et al., 2003), policies for inventory management (Xu et al., 2010), or traits for selective plant breeding (Hunter & McClosky, 2016).

We collect samples from individual alternatives in a sequential manner; formally, let $\{x^n\}_{n=0}^{N-1}$ be a sequence of alternatives (also referred to as the “allocation”), with each $x^n \in \{1, ..., M\}$. In each time stage $n$, we observe $W_{x^n}^{n+1} \sim F_{x^n}$. Denote by $\mathcal{F}_n$ the sigma-algebra generated by $x^0, W_{x^0}^1, ..., x^{n-1}, W_{x^{n-1}}^n$. Crucial to this paper is the following distinction between two kinds of
allocations: \( \{x^n\} \) is said to be adaptive if \( x^n \in \mathcal{F}^n \) for all \( n \), and static if \( x^n \in \mathcal{F}^0 \) for all \( n \). In words, under a static allocation, the sampling decision \( x^n \) (for any value of \( n \)) can be precomputed at time 0, and does not depend on the observed values of any samples, whereas under an adaptive allocation, \( x^n \) becomes known at time \( n \) based on the results of previous decisions.

Static allocations have been widely studied in the literature on statistical design of experiments (Pronzato & Müller, 2012), which often places less emphasis on sampling noise (in some cases, samples may even be deterministic) and more on large decision spaces that have to be “filled” in order to reduce statistical uncertainty (Qian et al., 2008; Zhang & Qian, 2013). By contrast, the simulation community (and the broader community studying stochastic optimization) has preferred to focus on adaptive allocations. Such an allocation can be obtained from a decision rule that, at time \( n \), maps a set of sufficient statistics calculated from the first \( n \) samples (e.g., sample means) to a value in \( \{1, \ldots, M\} \). In this way, although it is not known in advance exactly which alternative will be sampled at some future time stage, the decision rule encodes the logic by which this decision will be made based on the information that will be observed. There are many “schools of thought” for how such rules should be designed: the list includes, but is not limited to, indifference-zone methods that successively screen out alternatives based on pairwise comparisons (Kim & Nelson, 2001; Andradóttir & Kim, 2010); Bayesian methods based on various measures of expected improvement (Jones et al., 1998; Chick et al., 2010; Salemi et al., 2014); posterior sampling techniques (Agrawal & Goyal, 2012; Russo & Van Roy, 2014); upper confidence bound methods (Bubeck et al., 2009; Audibert & Bubeck, 2010; Ma & Henderson, 2017); and others, e.g., Fan et al. (2016). Many of these techniques have performed well in experiments and enjoy various types of performance guarantees; usually, each school of thought has its own approach to theoretical analysis that is tractable for the algorithmic class under consideration and tends to produce a certain type of result.

In this paper, we focus on one particular school of thought, which originated from the simulation community (specifically, from the seminal paper by Glynn & Juneja, 2004), and which is based on the idea of using static allocations to provide guidance for adaptive ones. This method uses large deviations theory to obtain very strong characterizations of optimality for a static allocation under fairly general distributional assumptions. More precisely, Glynn & Juneja (2004) found that, under a static allocation, and in the asymptotic regime where \( N \to \infty \), the probability of correctly identifying \( x^* \) can be made to converge to 1 at the fastest possible rate if the allocation satisfies certain conditions. These optimality conditions cannot be satisfied in practice without full
knowledge of $\mu_x$; however, one can then design an adaptive allocation that attempts to satisfy these conditions (i.e., to learn their solution) asymptotically. In recent years, there has been an entire stream of literature based on precisely this idea, with some representative papers being Pasupathy et al. (2014), Hunter & Feldman (2015), Hunter & McClosky (2016), Gao et al. (2017), and Shin et al. (2018); closely related is the methodology of optimal computing budget allocation (OCBA), which is based on a very similar idea but with various approximations of the optimality conditions (Chen & Lee, 2010; Chen et al., 2015; Zhang et al., 2016). These papers primarily focus on characterizing optimal static allocations in various settings of interest, then propose simple heuristics for adaptively approximating these allocations. Many of these heuristics require considerable computational effort due to, e.g., having to solve systems of nonlinear equations.

We propose a new methodology called Balancing Optimal Large Deviations (BOLD), which provably learns the solution to the optimality conditions obtained using large deviations theory. BOLD has several advantages. First, it is quite computationally efficient and easy to implement under a variety of sampling distributions (by contrast, many of the above-cited papers are restricted to specific distributional families, such as normal). Our method does not need to numerically solve systems of nonlinear equations, but rather adaptively learns the solution by iteratively “balancing” the two sides of each equation; since it is much easier to evaluate either side than to solve for the value that makes them equal, the calculation of $x^n$ can be done much more quickly.

Second, BOLD does not require any tuning. This is of particular interest because of several recent theoretical advances that provably enable partial learning of the solution. The large deviations analysis of Glynn & Juneja (2004) essentially shows that the decision to sample the optimal alternative $x^*$ must be considered separately from the decision to sample a particular $x \neq x^*$. In other words, certain logic is used to decide whether $x^*$ should be sampled, and if we decide not to do so, then other criteria are used to choose from among the suboptimal alternatives (a stark illustration of the exploration/exploitation tradeoff). Recent work by Russo (2019) and Qin et al. (2017) has developed a methodology where the first decision is passed back to the decision-maker in the form of a tunable parameter: we sample $x^*$, or our best guess of what $x^*$ is, with some fixed probability $\rho$. If we do not sample $x^*$, the second decision of which suboptimal alternative to sample is then made by the method in an asymptotically optimal manner. Essentially, BOLD removes the tuning requirement, and satisfies all of the optimality conditions asymptotically.

Third, BOLD can be viewed a new algorithmic concept for simulation optimization. Each school
of thought in the existing literature is based on a particular computational principle – for example, sampling from a posterior distribution, or maximizing an expected improvement criterion – and then this computational structure is connected to a certain theoretical result. The literature on large deviations-based methods has long had the result (optimality conditions for static allocations), but thus far has not accepted a single formal decision criterion that can be linked back to it. Our paper provides this criterion, which is a plug-in estimate of the large deviations rate function itself. The potential value of this development goes beyond the scope of the present paper because, as can be seen from Pasupathy et al. (2014) and the other papers cited above, large deviations rate functions can be derived in many more general simulation optimization settings, not just ranking and selection. This suggests that the algorithmic ideas of BOLD may also generalize to those settings; in other words, BOLD is not only an algorithm for ranking and selection, but also a “template” for creating similar algorithms in other problem classes. We present this concept rigorously and give a full proof that the optimality conditions are satisfied.\footnote{Some very recent papers have explored algorithmic ideas that resemble BOLD: for example, the “adaptive Welch divergence” algorithm of Shin et al. (2018) is identical to BOLD within the special case of normal sampling distributions, while the “optimal EOC allocation” algorithm of Gao et al. (2017) uses similar logic in a more general setting. However, these algorithms were presented as heuristics; the present paper is the first to provide a rigorous convergence proof in a general setting.}

Our theoretical analysis takes place in a general setting, where $F_x$ can be essentially any distribution that has a large deviations rate function. This generality introduces substantial technical challenges. Our main result in this paper is that BOLD asymptotically solves the large deviations optimality conditions w.p. 1. The first stepping stone toward this result is to show that BOLD is consistent, i.e., samples each alternative infinitely often as $N \to \infty$, but even this basic property (which may be easy to show under specific distributional assumptions) becomes non-trivial to prove in the general setting of BOLD. There are also many other intermediate steps characterizing various aspects of the sampling behaviour of BOLD, some of which may be of stand-alone theoretical interest. In short, our paper not only presents a set of guarantees for a new algorithm, but also develops the theoretical machinery needed to analyze future algorithms of this type.

We note one technical subtlety which the reader should bear in mind. Although Glynn & Juneja (2004) explicitly characterizes the convergence rate of the optimal static allocation, an adaptive allocation that converges to this allocation asymptotically (as BOLD does) is not guaranteed to achieve the same convergence rate as the static allocation. This issue has been known since Glynn & Juneja (2011); see also Glynn & Juneja (2018) and Wu & Zhou (2018) for some discussion and
analysis. Nonetheless, the large deviations optimality conditions are still important to the performance of an adaptive procedure; for example, Russo (2019) finds that these same conditions are an important prerequisite for rate-optimality of another type of performance criterion developed in that paper. Whatever the limitations of Glynn & Juneja (2004) may be, as of this writing this framework continues to be the foundation for virtually all of the work on convergence rates for ranking and selection, and our contribution in this paper is to demonstrate that a simple, computationally efficient adaptive procedure can provably solve the optimality conditions, for general sampling distributions, without the need for tuning. In and of itself, this result is one of the strongest available for any modern ranking and selection method. We further note that, if future researchers were to derive a new, as yet undiscovered class of large deviations rate functions for adaptive procedures, the BOLD logic could still be used to balance those rates in a similar way.

2 Preliminaries

Section 2.1 gives additional definitions and notation that will be needed for our analysis. Section 2.2 gives an exposition of the large deviations framework developed by Glynn & Juneja (2004) to characterize optimal static allocations. Section 2.3 gives several specific examples for a variety of distributional families.

2.1 Definitions and notation

We continue building on the notation introduced in Section 1. Let $N^n_x = \sum_{m=0}^{n-1} 1_{\{x^n=x\}}$ be the number of times that $x$ is chosen for sampling up to time $n \geq 1$. At any time $n \geq 1$, we can calculate the sample mean

$$\theta^n_x = \frac{1}{N^n_x} \sum_{m=0}^{n-1} 1_{\{x^n=x\}} W_{x}^{n+1}$$

for any $x$ (we can also compute $\theta^{n+1}_x$ recursively from $\theta^n_x$). The quantity $\theta^0_x$ can be set to any arbitrary constant. Let $x^{*,n} = \arg \max_x \theta^n_x$ denote the index of the alternative that is believed to be the best, given the time-$n$ sample means; if $\arg \max_x \theta^n_x$ is not unique, let $x^{*,n}$ be the alternative that has the smallest sample size $N^n_y$ among $y \in \arg \max_x \theta^n_x$, with further ties broken arbitrarily.

We say that correct selection occurs at time $n$ if $x^{*,n} = x^*$. The probability of correct selection (PCS), denoted by $P(x^{*,n} = x^*)$, depends on the allocation $\{x^n\}$ in a complicated way.
Glynn & Juneja (2004) obtained the following theoretical characterization of PCS for static (deterministic) allocations that satisfy $\lim_{n \to \infty} \frac{N^n_x}{n} = \alpha_x$ with $\alpha_x > 0$ for all $x$. For such allocations, the probability $P(x^* \neq x)$ of incorrect selection is shown to satisfy

$$- \lim_{n \to \infty} \frac{1}{n} \log P(x^* \neq x) = \Gamma(\alpha; \mu),$$

where $\Gamma$ depends on the limiting allocation $\alpha = (\alpha_1, \ldots, \alpha_M)$ as well as on the population parameters $\mu = (\mu_1, \ldots, \mu_M)$. In words, the probability of incorrect selection converges to zero at an exponential rate with exponent $\Gamma$. One can then reverse-engineer (1) to obtain an optimal static allocation. This can be done by solving the optimization problem

$$\max_{\alpha \in \mathbb{R}^M} \Gamma(\alpha; \mu) \quad \text{s.t.} \quad \sum_{x=1}^{M} \alpha_x = 1, \quad \alpha \geq 0.$$  

It can be shown that (2) is a convex program, whence it follows that the optimal solution $\alpha^*$ can be characterized by the KKT conditions.

Of course, $\alpha^*$ depends on the unknown population parameters $\mu$ and thus cannot be computed in practice; nonetheless, the formulation in (2) provides useful guidance for developing more practical procedures. We will discuss this in more detail further down, but it can already be seen that $\alpha^*$ can be approximated, for example, by replacing $\mu$ in (2) by the sample means $\theta^n = (\theta^n_1, \ldots, \theta^n_M)$.

### 2.2 Derivation of the exponential convergence rate

We now summarize the theoretical analysis of Glynn & Juneja (2004), which will be vital for understanding subsequent developments. Basically, we sketch the key steps leading to (1); to reduce citation clutter, all arguments in this subsection can be assumed to originate from Glynn & Juneja (2004). For simplicity, suppose that $N^n_x = \alpha_x n$ for all $n$. While this will not be the case in finite time, it does not matter for the analysis since we will let $n \to \infty$, and only minor technical nuisances are needed to show that the same results hold for static allocations satisfying $\frac{N^n_x}{n} \to \alpha_x$.

First, it can be shown that $\Gamma(\alpha; \mu) = \min_{x \neq x^*} \Gamma_x(\alpha^*_x, \alpha_x; \mu)$, where

$$\Gamma_x(\alpha^*_x, \alpha_x; \mu) = - \lim_{n \to \infty} \frac{1}{n} \log P(\theta^n_{x^*} < \theta^n_x),$$

provided that the limit in (3) exists. In words, the convergence rate of the probability of false selection can be related to the convergence rate of the probability of error in the pairwise comparison.
between $x^*$ and individual $x \neq x^*$. In turn, this probability is essentially an integral over the tail of the joint distribution of $(\theta^0_n, \theta^0_n)$, and thus is governed by the large deviations behaviour of this joint distribution.

To characterize this behaviour, we study the sampling distributions of $x$ and $x^*$. For any $x = 1, \ldots, M$, define $\Psi_x (\gamma) = \log \mathbb{E} \left( e^{\gamma W_x} \right)$, where $W_x$ denotes a generic sample from $F_x$. Let

$$I_x (u; \mu_x) = \sup_{\gamma} \gamma u - \Psi_x (\gamma)$$

be the Fenchel-Legendre transform of $\Psi_x$. By observing that

$$\log \mathbb{E} \left( e^{\gamma_1 \theta^0_n + \gamma_2 \theta^0_n} \right) = \log \mathbb{E} \left( e^{\gamma_1 \theta^0_n} \right) + \log \mathbb{E} \left( e^{\gamma_2 \theta^0_n} \right)$$

and applying the Gärtner-Ellis theorem (Dembo & Zeitouni, 2009), it is shown\(^2\) that

$$\Gamma_x (\alpha_x, \alpha_x^*; \mu_x) = \inf_{u \in [\mu_x, \mu_x^*]} \alpha_x I_x^* (u; \mu_x^*) + \alpha_x I_x (u; \mu_x)$$

so the convergence rate of the probability of error in the pairwise comparison is governed by the individual rate functions $I_x^*, I_x$. We now state without proof some additional facts about the rate functions $I_x$ that will be useful later. For notational convenience, we define $D_x (u; \mu) = \frac{d}{du} I_x (u; \mu).

**Lemma 2.1.** The following facts hold:

a) For any $x$, $I_x (\mu_x; \mu_x) = 0$.

b) Both $I_x^* (u; \mu_x^*)$ and $I_x (u; \mu_x)$ are decreasing in $u$ for $u < \mu_x$ and increasing in $u$ for $u > \mu_x^*$.

Therefore, the minimizer of the function

$$u \mapsto \alpha_x I_x^* (u; \mu_x^*) + \alpha_x I_x (u; \mu_x)$$

must lie in the interval $[\mu_x, \mu_x^*]$.

c) Both $I_x^* (u; \mu_x^*)$ and $I_x (u; \mu_x)$ are strictly convex on the interval $[\mu_x, \mu_x^*]$. Furthermore, both $\frac{d^2}{du^2} I_x^* (u; \mu_x^*)$ and $\frac{d^2}{du^2} I_x (u; \mu_x)$ are strictly positive for $u \in [\mu_x, \mu_x^*]$.

d) The function $I_x^* (u; \mu_x^*)$ is decreasing on the interval $[\mu_x, \mu_x^*]$ with

$$D_x^* (\mu_x^*; \mu_x^*) = 0, \quad D_x^* (\mu_x; \mu_x^*) < 0.$$
e) The function $I_x(u; \mu_x)$ is increasing on the interval $[\mu_x, \mu_{x*}]$ with

$$D_x(\mu_{x*}; \mu_x) > 0, \quad D_x(\mu_x; \mu_{x*}) = 0.$$  

Continuing our exposition, the equation

$$\alpha_x D_{x*} (u; \mu_{x*}) + \alpha_x D_x (u; \mu_x) = 0 \quad (5)$$

can be shown to have a unique solution $u_{x*,x} (\alpha_{x*}, \alpha_x)$, whence

$$\Gamma_x (\alpha_{x*}, \alpha_x; \mu) = \alpha_x I_x (u_{x*,x} (\alpha_{x*}, \alpha_x); \mu_x),$$

with the partial derivatives

$$\frac{\partial}{\partial \alpha_{x*}} \Gamma_x (\alpha_{x*}, \alpha_x; \mu) = I_x (u_{x*,x} (\alpha_{x*}, \alpha_x); \mu_x), \quad (6)$$

$$\frac{\partial}{\partial \alpha_x} \Gamma_x (\alpha_{x*}, \alpha_x; \mu) = I_x (u_{x*,x} (\alpha_{x*}, \alpha_x); \mu_x). \quad (7)$$

From this, it follows that the optimality conditions of problem (2) consist of two parts:

- Total balance condition:

$$\sum_{x \neq x*} \frac{I_x (u_{x*,x} (\alpha_{x*}, \alpha_x); \mu_{x*})}{I_x (u_{x*,x} (\alpha_{x*}, \alpha_x); \mu_x)} = 1. \quad (8)$$

- Individual balance conditions:

$$\Gamma_x (\alpha_{x*}, \alpha_x; \mu) = \Gamma_y (\alpha_{x*}, \alpha_y; \mu), \quad \text{for all } x, y \neq x*.$$  

Condition (8) arises due to the equality constraint in (2) combined with (6)-(7), while (9) arises after applying a standard technique for linearizing the minimum in the objective of (9).

### 2.3 Examples

If we are willing to impose additional assumptions on the sampling distributions, it is possible to obtain much more explicit forms of (8)-(9). Four examples are given below.

**Normal distributions.** Suppose that $W_x \sim N (\mu_x, \lambda_x^2)$ for all $x$. Then, it is shown in Example 1 of Glynn & Juneja (2004) that $I_x (u; \mu_x) = \frac{(u-\mu_x)^2}{2\lambda_x^2}$. By plugging this into (5) and working through
some algebra, we find that (8) becomes
\[ \left( \frac{\alpha x^*}{\lambda x^*} \right)^2 = \sum_{x \neq x^*} \left( \frac{\alpha x}{\lambda x} \right)^2, \tag{10} \]
while (9) becomes
\[ \frac{(\mu_x - \mu_{x^*})^2}{\lambda x^* + \lambda_x^2} = \frac{(\mu_y - \mu_{x^*})^2}{\lambda y^2 + \lambda y}, \quad x, y \neq x^*. \tag{11} \]

**Bernoulli distributions.** Suppose that \( W_x \sim \text{Bernoulli} (\mu_x) \). Then, it is shown in Example 2 of Glynn & Juneja (2004) that
\[ I_x (u; \mu_x) = u \log \frac{u}{\mu_x} + (1 - u) \log \frac{1 - u}{1 - \mu_x}, \tag{12} \]
and (8) can be checked by plugging
\[ u_{x^*} (\alpha x^*, \alpha x) = \frac{\frac{\mu_{x^*}}{1 - \mu_{x^*}} \lambda x^* + \alpha x}{1 + \frac{\mu_{x^*}}{1 - \mu_{x^*}} \lambda x^* + \alpha x}, \]
in (12). To check (9), we use
\[ \Gamma_x (\alpha x^*, \alpha x; \mu) = -(\alpha x^* + \alpha x) \log \left[ (1 - \mu_{x^*}) \frac{\alpha x}{\alpha x^* + \alpha x} + (1 - \mu_x) \frac{\alpha x^*}{\alpha x + \alpha x^*} + \mu \frac{\alpha x^*}{\alpha x + \alpha x^*} \right]. \]

**Exponential distributions.** Suppose that \( W_x \sim \text{Exp} (\lambda x) \), so that \( \mu_x = \frac{1}{\lambda x} \). Gao & Gao (2016) derived \( I_x (u; \mu_x) = \lambda x u - 1 - \log (\lambda x u) \), whence (8) becomes
\[ \sum_{x \neq x^*} \frac{\lambda x^* (\alpha x^* + \alpha x)}{\alpha x^* + \alpha x + \alpha x^* + \alpha x \lambda x} - 1 - \log \frac{\lambda x^* (\alpha x^* + \alpha x)}{\alpha x^* + \alpha x + \alpha x^* + \alpha x \lambda x} = 1, \tag{13} \]
and (9) becomes
\[ \alpha x^* \log \frac{\lambda x^* (\alpha x^* + \alpha x)}{\alpha x^* + \alpha x} + \alpha x \log \frac{\lambda x (\alpha x + \alpha x)}{\alpha x + \alpha x^* + \alpha x^* + \alpha x \lambda x} = \alpha x^* \log \frac{\lambda x^* (\alpha x^* + \alpha y)}{\alpha x^* + \alpha x^* + \alpha x y} + \alpha y \log \frac{\lambda y (\alpha x^* + \alpha y)}{\alpha x^* + \alpha x^* + \alpha y \lambda y}. \tag{14} \]

**Noncentral chi-squared distributions.** Ryzhov (2018) introduced the following “targeting and selection” problem. As in the normal case, suppose that \( W_x \sim \mathcal{N} (\mu_x, \lambda x^2) \), but the best alternative is now given by \( x^* = \arg \min_x (\mu x - c)^2 \) for some prespecified constant \( c \), while the selection decision is given by \( x^* = \arg \min_x (\theta_n^x - c)^2 \). In other words, we are now trying to select an alternative whose value matches the “target” \( c \) as closely as possible.
In this case, the analysis of PCS has a few extra nuances because now the selection criterion is not linear in the sample mean, but overall one can redo the derivations in Glynn & Juneja (2004) for the noncentral chi-squared distribution to obtain similar optimality conditions. Equation (8) again becomes (10), as in the normal example. Equation (9) becomes

\[
\frac{(|\mu_x - c| - |\mu_{x^*} - c|)^2}{\frac{\lambda^2_x}{\alpha_x} + \frac{\lambda^2}{\alpha_x}} = \frac{(|\mu_y - c| - |\mu_{x^*} - c|)^2}{\frac{\lambda^2_y}{\alpha_{x^*}} + \frac{\lambda^2_y}{\alpha_{y^*}}}, \quad x, y \neq x^*.
\]

3 The BOLD (Balancing Optimal Large Deviations) algorithm

The examples in Section 2.3 demonstrate that the optimality conditions are highly nonlinear; that the functional forms used to state them are considerably different for different sampling distributions; that solving these equations is quite computationally burdensome for moderately large \( M \); and that, of course, the optimal proportions \( \alpha \) depend on the unknown distributional parameters, and thus are also unknown to the decision-maker.

The standard approach adopted in the literature is to use (8)-(9) as guidance for the development of adaptive procedures. For example, one could first assume a particular distributional family (much of the simulation literature focuses on normal distributions) to obtain more explicit conditions such as (10)-(11). Then, one could replace the unknown distributional parameters with plug-in estimates (for example, using \( \theta^n_x \) instead of \( \mu_x \)), solve the resulting nonlinear equations numerically to obtain an approximate allocation \( \alpha^n \), then use this allocation to distribute part or all of the sampling budget. Alternately, one could assign the next experiment to alternative \( x \) with probability \( \alpha^n_x \), then update the sample means, calculate a new approximation \( \alpha^{n+1} \), and repeat the process. This is essentially the approach adopted in the implementation of optimal computing budget allocation methods (Chen & Lee, 2010; Pasupathy et al., 2014). We may observe, however, that even if the true values are known, it is much more difficult to balance the rate functions (i.e., to find \( \alpha \) such that (9) holds) than to evaluate them (i.e., to compute \( \Gamma_x \) for some fixed \( \alpha \)). This is the main insight behind our proposed BOLD (Balancing Optimal Large Deviations) method, which evaluates the individual rate functions under the most recent set of sample means, then seeks to balance them by assigning simulations to alternatives whose rates appear to be too small.

In order to make the presentation of the algorithm more compact, we introduce some additional
notational shorthand. For all \( x \), let
\[
I^n_x (\cdot) = I_x (\cdot; \theta^n_x), \quad D^n_x (\cdot) = D_x (\cdot; \theta^n_x)
\]
denote time-\( n \) approximations of the rate functions and their derivatives. Now, in the \( n \)th stage of sampling, we first consider the equation
\[
N^n_x D^n_x (u) + N^n_y D^n_y (u) = 0, \quad x \neq y. \tag{16}
\]
If we specifically set \( x \) or \( y \) equal to \( x^{*,n} \), (16) can be viewed as a substitute for (5). Essentially, we plug in the sample means instead of the true values, and replace the unknown best alternative \( x^* \) by the currently-estimated best alternative \( x^{*,n} \). Second, we essentially multiply both sides of (5) by \( n \), recasting the equation in terms of the current sample sizes rather than the sampling proportions; this modification does not affect the solution to (16), which we denote by \( u^n_{x,y} \). We then define
\[
\Gamma^n_{x,y} = N^n_x I^n_x (u^n_{x,y}) + N^n_y I^n_y (u^n_{x,y}), \tag{17}
\]
with the specific case \( \Gamma^n_{x^{*,n}, x} \) being an approximation of the rate function of \( x \neq x^{*,n} \).

Informally, suppose for the sake of argument that \( N^n_x \rightarrow \infty \) for all \( x \) (as will be proved later), and that \( n \) is large enough such that \( x^{*,n} = x^* \). Then, from (17), we expect that \( \Gamma^n_{x^{*,n}, x} \) will also increase to infinity. This suggests that we can balance the rate functions of suboptimal alternatives by simply assigning the next sample to the alternative with the smallest value of (17); intuitively, this should force \( \Gamma^n_{x^{*,n}, x} \) to go to infinity at the same rate for all \( x \neq x^* \). In a similar way, we can approximate the left-hand side of (8) and check whether this quantity is above or below the desired value of 1 to determine whether the next sample should be assigned to \( x^{*,n} \). The precise statement of the BOLD algorithm is given in Figure 1.

Several observations are in order. First, the BOLD algorithm has no tunable parameters. Second, BOLD is nonrandomized and does not introduce any additional noise into the sampling decision beyond what is already contained in the sample mean. Third, if the decision-maker knows the distributional family from which the samples are generated, BOLD becomes very easy to implement; closed-form expressions for \( I^n \) and \( \Gamma^n \) are relatively easy to derive for many common families, some of which were discussed in Section 2.3. To give one example, if the sampling distributions are \( \mathcal{N} (\mu_x, \lambda^2_x) \) with known \( \lambda_x \), the condition in (18) becomes
\[
\left( \frac{N^n_{x^{*,n}}}{\lambda_{x^{*,n}}} \right)^2 < \sum_{x \neq x^{*,n}} \left( \frac{N^n_x}{\lambda_x} \right)^2, \tag{20}
\]
**Step 0:** Initialize $n = 0$ and $N_x^n = 0$ for all $x$.

**Step 1:** If $\arg\max_x \theta^n_x$ is not unique, assign $x^n = x^{*n}$ according to the definition in Section 2.1, and proceed directly to Step 4.

**Step 2:** If $\arg\max_x \theta^n_x$ is unique, check whether

$$\sum_{x \neq x^{*n}} I^n_{x^{*n}, x} \left( u^n_{x^{*n}, x} \right) > 1.$$  

(18)

**Step 3:** If (18) holds, assign $x^n = x^{*n}$. Otherwise, assign

$$x^n = \arg\min_{x \neq x^{*n}} \Gamma^n_{x^{*n}, x}.$$  

(19)

**Step 4:** Collect new information $W^n_{x,n+1}$, update sample means. Increment $n$ by 1 and return to step 1.

---

Figure 1: Description of BOLD algorithm.

so BOLD will assign $x^n = x^{*n}$ if (20) holds. If (20) does not hold, the calculation in (19) becomes

$$x^n = \arg\min_{x \neq x^{*n}} \frac{(\theta^n - \theta^n_{x^{*n}})^2}{\lambda_x^{2, n} x^{*n} + \lambda_x^{2, n}}.$$  

(21)

and if the variances are unknown, one could simply plug in estimators of these quantities where necessary.\(^3\) From a computational standpoint, the minimization in (21) is actually simpler than the calculations used in some other popular heuristics, such as expected improvement (Ryzhov, 2016). At the same time, BOLD is flexible: if the sampling distribution is, e.g., exponential rather than normal, one simply substitutes the appropriate closed forms of $I^n, \Gamma^n$ into (18)-(19). BOLD is also much easier to implement than the standard plug-in approach to large deviations allocations (Pasupathy et al., 2014), since it does not have to solve a system of nonlinear equations.

Essentially, BOLD exploits the structure of the optimality conditions to adaptively learn the optimal static allocation without the need for tuning. While this does not mean that it achieves the same convergence rate as the static allocation, nonetheless, if one accepts that the static allocation is a valuable goal to work towards, BOLD offers a principled computational framework for achieving this goal. We should note that, if one has no knowledge of the distributional family (that is, one cannot evaluate $I(u; \theta)$ for arbitrary $u, \theta$), implementing BOLD may become more difficult. In this

\(^3\)We again note that (20)-(21) are identical to the “adaptive Welch divergence” heuristic of Shin et al. (2018). However, BOLD only has this particular computational form in the special case of normality.
paper, we do not consider this case, and assume that the distributional family is known; we note, however, that this assumption is already comparable to, or more general than, the assumptions underlying many existing methods in the simulation literature.

4 Theoretical analysis

Our main theoretical result is that BOLD can be guaranteed to learn the solution to (8)-(9) with probability 1 as $n \to \infty$. The proof of this result relies on a number of more fundamental properties, such as statistical consistency. Thus, to make our presentation more systematic, we have organized this section as follows. Section 4.1 proves that BOLD is consistent, i.e., that $N^n_x \to \infty$ for all $x$. Section 4.2 proves that $N^n_x$ is $\Theta(n)$ for all $x$. Section 4.3 proves that BOLD asymptotically satisfies the total balance condition (8). Finally, Section 4.4 proves that BOLD asymptotically satisfies all of the individual balance conditions (9). Later subsections refer to results from earlier ones. All of these results involve various intermediate steps, which are stated separately as lemmas; for readability, the proofs of these lemmas are deferred to the Appendix, but some informal discussion is given to provide a sense of how these lemmas fit into the overall argument.\(^4\)

All limits in this section should be assumed to hold almost surely. For convenience, one can discard a suitable set of measure zero from the probability space $(\Omega, \mathcal{F}, P)$ to avoid having to repeat the qualification “almost surely.” To make the notation less cumbersome, various random quantities can be assumed to be evaluated on the same fixed sample path $\omega \in \Omega$ unless explicitly stated otherwise. Thus, for example, if we refer to “a fixed positive constant $C,$” it is implied that $C$ can depend on the sample path, but that we are considering the constant value $C(\omega)$ for some fixed $\omega$ that is not written explicitly.

4.1 Consistency of BOLD

We first show that BOLD is consistent, that is, $N^n_x \to \infty$ for all $x$. Although this basic property is satisfied by many algorithms and heuristics, it will be helpful for the rest of our analysis, as it will allow us to study a simplified form of the BOLD algorithm in Sections 4.2-4.4 without loss of\(^4\)The structure of our argument builds on Chen & Ryzhov (2017, 2019), which studied a BOLD-like algorithm in the special case of normal distributions. The general case considered here is much more challenging: first, consistency is easy to show for normal distributions, but not in general; second, when $F_x$ is normal, the total balance condition (10) has no explicit dependence on the sample means and becomes much easier to handle; third, under normality, many of the intermediate steps in Section 4.4 are easier to show or may be taken for granted.
generality.

Before proceeding, we state three assumptions. Assumptions 4.1-4.2 are regularity conditions on the rate functions $I_x$, essentially ensuring that $I_x(u; \theta)$ has a unique zero at $u = \theta$ and is well-behaved around this value. Both assumptions can be verified for all of the examples given in Section 2.3.

**Assumption 4.1.** For any $x$ and arbitrarily small $\varepsilon > 0$,

\[
\inf_{\theta \in H, |u-\theta| > \varepsilon} I_x(u; \theta) > 0, \tag{22}
\]

\[
\inf_{\theta \in H, |u-\theta| > \varepsilon} |D_x(u; \theta)| > 0, \tag{23}
\]

where $H$ is any closed interval with non-empty interior containing $\mu_x$.

**Assumption 4.2.** Both $I_x(u; \theta)$ and $D_x(u; \theta)$ are continuous at every pair $(u, \theta)$ with $\theta \in H$, where $H$ is as in Assumption 4.1. Furthermore, for any $x$, and any $\varepsilon > 0$, there exists some $\delta > 0$ such that, for all pairs $(u, \theta)$ satisfying $\theta \in H$ and $|u - \theta| < \delta$, we have $I_x(u; \theta) < \varepsilon$ and $|D_x(u; \theta)| < \varepsilon$.

Next, Assumption 4.3 essentially ensures that $x$ and $y$ cannot be “mistaken” for each other if the sample mean for $x$ is close to the true value of $y$ or vice versa. This assumption is primarily used to address the pathological situation where $\mu_x = \theta^n_y$ for $x \neq y$, which cannot happen with non-zero probability when the sampling distributions are continuous. In the continuous setting, Assumption 4.3 can be omitted entirely, but it enables us to handle discrete distributions (for example, it can be verified for Bernoulli distributions).

**Assumption 4.3.** For any $x, y$, there exists a fixed positive constant $C$ such that

\[
\sup_{u \neq \theta} \left| \frac{D_x(u; \theta)}{D_y(\theta; u)} \right| \leq C.
\]

Now, we give a brief sketch of the overall argument. We prove that $N^n_x \to \infty$ by contradiction, i.e., by showing that it is impossible to have $\lim_{n \to \infty} N^n_z < \infty$ for any $z$. This is broken down into several steps. First, in Lemma 4.1, we show that, if such a $z$ exists, it must be the case that $x^{*,n} \neq z$ for all large enough $n$. It follows that, for any large enough $n$ when condition (18) holds, the sampling decision can only be made among those alternatives that are measured infinitely often. Since the law of large numbers will apply to all such alternatives, it follows that $x^{*,n}$ will eventually take the same value (i.e., $\arg \max_x \theta^n_x$ will eventually yield the same alternative index) for all large
enough \( n \). (Of course, this limiting value of \( x^{*,n} \) depends on the sample path; we again remind the reader that all arguments in this analysis implicitly occur on the same fixed sample path.)

**Lemma 4.1.** Let Assumptions 4.1-4.3 hold. Suppose that there exists some alternative \( z \) that is only sampled finitely often, i.e., \( \lim_{n \to \infty} N^n_z < \infty \). Then, \( x^{*,n} \neq z \) for all large enough \( n \).

Next, Lemma 4.2 proves that the BOLD algorithm does not get stuck on Steps 1-2 in Figure 1. That is, \( \arg \max_x \theta^n_x \) will be unique for all large enough \( n \) (this has to be verified for discrete sampling distributions), which means that condition (18) will always be checked; at the same time, there will be infinitely many time stages \( n \) at which this condition fails. Combining this with Lemma 4.1, it follows that, if there exists \( z \) satisfying \( \lim_{n \to \infty} N^n_z < \infty \), there must nonetheless have been infinitely many chances to sample that \( z \); that is, if \( x^{*,n} \neq z \) for all large enough \( n \), then \( z \) must be one of the alternatives being compared in (19) whenever (18) has failed.

**Lemma 4.2.** Let Assumptions 4.1-4.2 hold. For all large enough \( n \), condition (18) will be checked (i.e., Step 1 in Figure 1 will be invoked only finitely many times). Furthermore, there are infinitely many times \( n \) at which condition (18) does not hold.

Finally, the contradiction can be obtained. The calculation in (19) assigns \( x^n = \arg \min_{x \neq x^{*,n}} \Gamma^n_{x^{*,n},x} \). However, \( \Gamma^n_{x^{*,n},x} \) tends to increase over time when both \( x^{*,n} \) and \( x \) are sampled infinitely many times, whereas \( \Gamma^n_{x^{*,n},z} \) will be bounded from above. It follows that, for arbitrarily large \( n \), the argmin of (19) must correspond to an alternative that can no longer be sampled, which yields the conclusion that there is no \( z \) for which \( \lim_{n \to \infty} N^n_z < \infty \).

**Theorem 4.1.** Let Assumptions 4.1-4.3 hold. Under the BOLD algorithm, \( N^n_x \to \infty \) for all \( x \).

**Proof:** Suppose that there exists \( z \) with \( \lim_{n \to \infty} N^n_z < \infty \). From Lemma 4.2, we can see that for all large \( n \), (18) is always checked, and that there are infinitely many time stages in which (18) fails. Second, from Lemma 4.1, we know that, for all large enough \( n \), \( x^{*,n} \neq z \), which implies that \( N^n_{x^{*,n}} \to \infty \). Furthermore, for all large enough \( n \), \( x^{*,n} = \bar{x} \) where \( \bar{x} = \arg \max_{x \in A} \mu_x \), where \( A = \{ x : N^n_x \to \infty \} \). Therefore, for any large enough \( n \) where condition (18) fails, we will sample some \( y \neq x^{*,n} \) satisfying \( y \in A \).

On one hand, since \( x^{*,n} = \bar{x} \) for all large enough \( n \), we write

\[
\Gamma^n_{\bar{x},z} = \inf_u \left( N^n_{\bar{x}} I^n_{\bar{x}}(u) + N^n_z I^n_z(u) \right)
\]
\[ \leq N^n_x I^n_x (\theta^n_x) \\
= N^n_x I^n_x (\theta^n_x; \bar{\theta}_z). \quad (24) \]

Since \( \theta^n_x \to \mu_x \), it follows from Lemma 2.1 that \( \Gamma^n_{x,z} \) must be bounded by some fixed constant.

On the other hand, by Lemma 2.1, both \( I^n_x (u) \) and \( I^n_y (u) \) are positive. Let \( d^n_{x,y} = \frac{\theta^n_x + \theta^n_y}{2} \) and \( d_{x,z} = \frac{\mu_x + \mu_y}{2} \). We know that

\[ \Gamma^n_{x,y} = N^n_x I^n_x (u^n_{x,y}) + N^n_y I^n_y (u^n_{x,y}) \],

where \( u^n_{x,y} \) is the desired minimizer. Two cases must be considered: if \( d^n_{x,y} \leq u^n_{x,y} \leq \theta^n_x \), then

\[ \Gamma^n_{x,y} \geq N^n_y I^n_y (u^n_{x,y}) \geq N^n_y I^n_y (d^n_{x,y}) \],

and if \( \theta^n_y \leq u^n_{x,y} \leq d^n_{x,y} \), then

\[ \Gamma^n_{x,y} \geq N^n_x I^n_x (u^n_{x,y}) \geq N^n_x I^n_x (d^n_{x,y}) \].

Thus, in either case we have

\[ \Gamma^n_{x,y} \geq \min \{ N^n_x I^n_x (d^n_{x,y}) , N^n_y I^n_y (d^n_{x,y}) \} \).

Obviously, since both \( N^n_y \to \infty \) and \( N^n_x \to \infty \), we have \( \theta^n_x \to \mu_x \), \( \theta^n_y \to \mu_y \) and \( d^n_{x,y} \to d_{x,y} \). Then, for all large enough \( n \), we have \( |d^n_{x,y} - \theta^n_x| > C_1 \) and \( |d^n_{x,y} - \theta^n_y| > C_1 \), where \( C_1 \) is a fixed positive constant. By Assumption 4.1, this implies that, for all large enough \( n \), we have \( I^n_x (d^n_{x,y}) > C_2 \) and \( I^n_y (d^n_{x,y}) > C_2 \), where \( C_2 \) is another fixed positive constant. However, both \( N^n_y \to \infty \) and \( N^n_x \to \infty \), while \( \sup_n N^n_x \) is finite. Then, from (24), for all large enough \( n \), we must have

\[ \Gamma^n_{x,y} \geq N^n_x I^n_x (d^n_{x,y}) + N^n_y I^n_y (d^n_{x,y}) \]

\[ > N^n_x I^n_x (\theta^n_x) \\
\geq \Gamma^n_{x,z}. \]

This means that, for any large enough \( n \) where any alternative \( y \neq \bar{x}, z \) is sampled while condition (18) does not hold (recall that this occurs infinitely many times), the BOLD algorithm actually cannot sample alternative \( y \) since \( \Gamma^n_{x,y} > \Gamma^n_{x,z} \), thus leading to the desired contradiction. Therefore, every alternative must be sampled infinitely often. \( \square \)
**Step 0:** Initialize $n = 0$ and $N^n_x = 0$ for all $x$.

**Step 1:** Check whether
\[ \sum_{x \neq x^*} \frac{I^n_x(u^n_{x^*,x})}{I^n_x(u^n_{x^*,x})} > 1. \]  
(25)

**Step 2:** If (25) holds, assign $x^n = x^*$. Otherwise, assign
\[ x^n = \arg \min_{x \neq x^*} \Gamma^n_{x,x^*}. \]  
(26)

**Step 3:** Collect new information $W_{x^n+1}$, update sample means. Increment $n$ by 1 and return to step 1.

Figure 2: Description of simplified BOLD algorithm used in the theoretical analysis.

Theorem 4.1 implies that $\theta_n^x \to \mu_x$ for all $x$, and so the PCS converges to 1. For our purposes, this result also implies, e.g., that $I_x(u;\theta^n_x) \to I_x(u;\mu_x)$ pointwise. In other words, our approximations of the rate functions will eventually be arbitrarily close to the true rate functions.

### 4.2 Equivalent order of sampling rates

From Theorem 4.1, we see that $x^{*,n} = x^*$ for all large enough $n$. Thus, when analyzing the asymptotic behaviour of BOLD on a fixed sample path, it is sufficient to study the simpler algorithm in Figure 2. Of course, this simplified version is only used for theoretical analysis, and cannot be implemented in practice; however, the results that we prove for it will also hold for the original statement of the BOLD algorithm. To reduce notational clutter in our analysis, we define

\[ \Delta^n = \sum_{x \neq x^*} \frac{I^n_x(u^n_{x^*,x})}{I^n_x(u^n_{x^*,x})}. \]

The next result shows that $N^n_x$ have the same order of magnitude for all $x$, that is, every alternative is sampled with “sufficient” frequency. This is known to be an important prerequisite for good performance in ranking and selection; some methods in the literature, such as expected improvement (Ryzhov, 2016), have the consistency property of Theorem 4.1 but allocate disproportionately few samples to suboptimal alternatives.
Theorem 4.2. Let Assumptions 4.1-4.3 hold. For any two alternatives \( x \) and \( y \),
\[
\limsup_{n \to \infty} \frac{N^n_x}{N^n_y} < \infty.
\]

Proof: We divide the proof into several cases, depending on whether \( x^* \) is one of the alternatives considered in the comparison.

Case 1: \( y = x^* \). Suppose that the result does not hold; then, there exists \( x \neq x^* \) such that
\[
\limsup_{n \to \infty} \frac{N^n_x}{N^n_{x^*}} = \infty.
\]
Then, there exists a subsequence \( \{n_k\}_{k=0}^{\infty} \) satisfying \( \lim_{k \to \infty} \frac{N^n_{x_k}}{N^n_{x^*}} = \infty \). Now define a different subsequence \( \{m_k\} \) by
\[
m_k = \sup \left\{ l \leq n_k : x^l = x \right\}.
\]
We must also have \( \lim_{k \to \infty} \frac{N^{m_k}_x}{N^{m_k}_{x^*}} = \infty \) for this subsequence since \( N^{m_k}_{x^*} \to \infty \) and \( N^n_{x^*} \) is increasing in \( n \). Furthermore, we have \( x^{m_k} = x \) for all \( k \). As \( k \to \infty \), we also have
\[
\frac{D_{x^k}^{m_k} \left( u^{m_k}_{x^k}, x \right)}{D_{x^*}^{m_k} \left( u^{m_k}_{x^*}, x \right)} = \frac{N^{m_k}_{x^*}}{N^{m_k}_{x^*}},
\]
where the right-hand side of (27) vanishes to zero as \( k \to \infty \). Since \( u^{m_k}_{x^*, x} \in [\theta^{m_k}_{x^*}, \theta^{m_k}_{x^*}] \) by construction, we must have \( D_{x^*}^{m_k} \left( u^{m_k}_{x^*, x} \right) \to 0 \) due to (27) and Lemma 2.1. From Assumption 4.1, we have \( u^{m_k}_{x^*, x} - \theta^{m_k}_{x^*} \to 0 \). Furthermore, from Assumption 4.2, we obtain \( I_{x^k}^{m_k} \left( u^{m_k}_{x^k}, x \right) \to 0 \). At the same time, since \( \theta^{m_k} \to \mu \) by Theorem 4.1, there exists a fixed positive constant \( C_1 \) such that, for large enough \( k \),
\[
I_{x^k}^{m_k} \left( u^{m_k}_{x^k}, x \right) > I_{x^k}^{m_k} \left( \frac{\theta^{m_k}_{x^k} + \theta^{m_k}_{x^*}}{2} \right) > C_1,
\]
where the last inequality follows from Assumption 4.1. Therefore, for all large enough \( k \), we have
\[
\Delta^{m_k} \geq \frac{I_{x^k}^{m_k} \left( u^{m_k}_{x^k}, x \right)}{I_{x^k}^{m_k} \left( u^{m_k}_{x^*, x} \right)}> 1,
\]
which means that alternative \( x^* \) should be sampled at time \( m_k \) for large enough \( k \). However, this contradicts the fact that only \( x \) can be sampled along this subsequence. Therefore, for all \( x \neq x^* \),
\[
\limsup_{n \to \infty} \frac{N^n_x}{N^n_{x^*}} < \infty.
\]

Case 2: \( x, y \neq x^* \). Suppose that the result does not hold; then, there exist \( x, y \neq x^* \) for which
\[
\limsup_{n \to \infty} \frac{N^n_x}{N^n_y} = \infty.
\]
By repeating the argument from above, there must exist a subsequence \( \{n_k\} \)}
such that \( \lim_{k \to \infty} \frac{N_n^k}{N_y^k} = \infty \) and only alternative \( x \) is sampled along this subsequence. However, for large enough \( k \), we have

\[
\Gamma_{x^*,y}^{n_k} = \inf_u N_{x^*}^{n_k} I_{x^*}^{n_k} (u) + N_y^{n_k} I_y^{n_k} (u) \leq N_y^{n_k} I_y^{n_k} (\theta_x^{n_k}) \leq C_2 N_y^{n_k},
\]

where \( C_2 \) is a fixed positive constant and the last inequality in (28) holds by Assumption 4.2 since \( \theta^{n_k} \to \mu \). At the same time, for all large enough \( k \), we have

\[
\Gamma_{x^*,x}^{n_k} = N_x^{n_k} I_x^{n_k} (u_{x^*,x}) + N_x^{n_k} I_x^{n_k} (u_{x^*,x}) \geq C_3 N_x^{n_k} I_x^{n_k} (u_{x^*,x}) + N_y^{n_k} I_y^{n_k} (u_{x^*,x}) \geq N_x^{n_k} \max \{ C_3 I_x^{n_k} (u_{x^*,x}), I_y^{n_k} (u_{x^*,x}) \} \geq C_4 N_x^{n_k},
\]

where \( C_3, C_4 \) are positive constants. Inequality (29) holds because \( \limsup_{n \to \infty} \frac{N_n}{N_x^*} < \infty \) for \( x \neq x^* \), as was shown before, and inequality (30) holds by Assumption 4.1 since \( u_{x^*,x}^{n_k} \in [\theta_x^{n_k}, \theta_x^{n_k}] \) and \( \theta^{n_k} \to \mu \). Then, for all large enough \( k \), it follows that

\[
\frac{\Gamma_{x^*,y}^{n_k}}{\Gamma_{x^*,x}^{n_k}} \leq \frac{C_2}{C_4} \cdot \frac{N_y^{n_k}}{N_x^{n_k}},
\]

with the right-hand side of (31) vanishing to zero. Consequently, for all large enough \( k \), we will have

\[
\Gamma_{x^*,y}^{n_k} < \Gamma_{x^*,x}^{n_k},
\]

which contradicts the fact that only alternative \( x \) is sampled along the subsequence \( \{n_k\} \). Thus, for all \( x, y \neq x^* \), \( \limsup_{n \to \infty} \frac{N_n}{N_y^*} < \infty \).

**Case 3: \( x = x^* \).** Suppose that the result does not hold; then, we can find \( y \neq x^* \) satisfying \( \limsup_{n \to \infty} \frac{N_n}{N_y^*} = \infty \). By repeating the argument from above, there must exist a subsequence \( \{n_k\} \) such that \( \lim_{k \to \infty} \frac{N_y^{n_k}}{N_x^{n_k}} = \infty \) and only alternative \( x^* \) is sampled along this subsequence. Furthermore, since we have just shown that

\[
0 < \liminf_{n \to \infty} \frac{N_y}{N_y} \leq \limsup_{n \to \infty} \frac{N_n}{N_y} < \infty
\]

for any suboptimal \( z \neq y \), we can see that in fact \( \limsup_{n \to \infty} \frac{N_n}{N_x^*} = \infty \) and \( \lim_{k \to \infty} \frac{N_y^{n_k}}{N_x^{n_k}} = \infty \) for any suboptimal alternative \( z \). However, for any suboptimal \( z \), as \( k \to \infty \) we also have

\[
\frac{D_{x^*}^{n_k} (u_{x^*,x})}{D_{x^*}^{n_k} (u_{x^*,x})} = \frac{N_y^{n_k}}{N_x^{n_k}},
\]

(32)
with the right-hand side of (32) vanishing to zero. Since \( u_{x^*,z}^n \in [\theta_x^{n_k}, \theta_x^{n_k}] \) by construction, we must have \( D_x^{n_k} \left( u_{x^*,z}^{n_k} \right) \to 0 \) due to (32) and Lemma 2.1. From Assumption 4.1, we have \( u_{x^*,z}^n - \theta_x^{n_k} \to 0 \). Furthermore, from Assumption 4.2, we obtain \( I_x^{n_k} \left( u_{x^*,z}^n \right) \to 0 \). At the same time, since \( \theta^{n_k} \to \mu \) by Theorem 4.1, there exists a fixed positive constant \( C_5 \) such that, for large enough \( k \),

\[
I_x^{n_k} \left( u_{x^*,z}^n \right) \to 0
\]

where the last inequality follows from Assumption 4.1. Therefore, for large enough \( k \), we have

\[
\frac{I_x^{n_k} \left( u_{x^*,z}^n \right)}{I_x^{n_k} \left( u_{x^*,z}^n \right)} \to 0,
\]

whence \( \Delta^{n_k} < 1 \), which means that \( x^* \) cannot be sampled. This contradicts the construction of \( n_k \), along which only \( x^* \) is sampled; therefore, we conclude that \( \lim sup_{n \to \infty} \frac{N_x^n}{N_x^\theta} < \infty \), completing the proof.

From Theorem 4.2, by symmetry it follows that \( N_x^n = \Theta (n) \) for all \( x \). We can also obtain the following result as a corollary. Essentially, this result means that, for large enough \( n \), the value \( u_{x^*,x}^n \) used to calculate \( \Gamma_x^n \) in (17) will be bounded away from the endpoints of its feasible interval, allowing us to place bounds on various quantities derived from the empirical rate functions \( I_x^n \).

**Corollary 4.1.** Let Assumptions 4.1-4.3 hold. For all \( x \), there exists some fixed positive constant \( \delta \) such that, for all large enough \( n \),

\[
|u_{x^*,x}^n - \mu_x| > \delta, \quad |u_{x^*,x}^n - \mu| > \delta.
\]

**Proof:** From Theorem 4.2, we have

\[
\lim inf_{n \to \infty} \left| \frac{D_x^n \left( u_{x^*,x}^n \right)}{D_x^{n_k} \left( u_{x^*,x}^n \right)} \right| = \lim inf_{n \to \infty} \frac{N_x^n}{N_x^\theta} > 0
\]

and, symmetrically,

\[
\lim sup_{n \to \infty} \left| \frac{D_x^n \left( u_{x^*,x}^n \right)}{D_x^{n_k} \left( u_{x^*,x}^n \right)} \right| = \lim sup_{n \to \infty} \frac{N_x^n}{N_x^\theta} > 0.
\]

Combining this with Assumption 4.2 leads to the desired conclusion.

### 4.3 Total balance condition

Our next goal is to prove \( \Delta^n \to 1 \), thus satisfying the total balance condition (8) asymptotically. Before proceeding, we impose one additional regularity condition on \( I_x (u; \theta) \) and \( D_x (u; \theta) \), viewed here as bivariate functions of both \( u \) and \( \theta \).
**Assumption 4.4.** For all $x$, both $\frac{\partial I}{\partial \theta}(u; \theta)$ and $\frac{\partial D}{\partial \theta}(u; \theta)$ are continuously differentiable with respect to both $u$ and $\theta$.

We also define the notation

$$L^{n,n+m}_x = N^{n+m}_x - N^n_x$$

to be the number of samples allocated to a fixed alternative $x$ in between two fixed times $n$ and $n + m - 1$. The remainder of our analysis revolves around understanding the behaviour of $L^{n,n+m}_x$ for various choices of $n$, $m$ and $x$. Let us informally discuss how this behaviour is important for verifying the total balance condition.

From the previous analysis, we know that condition (25) both succeeds and fails infinitely many times; in other words, $\Delta_n$ crosses the level 1 infinitely often from both above and below. Consider some large $n$ for which $x^n = x^*$, i.e., $\Delta^n > 1$, and choose $m$ such that $n + m$ is the next time stage at which $x^*$ is sampled, whence $\Delta^{n+s} \leq 1$ for all $0 < s < m$. We will show that, first of all, $m$ is “small” in a certain sense to be defined. We will then argue that the “small” change from $n$ to $n + m$ also produces “small” incremental changes in $\Delta^{n+s}$ for $0 < s < m$, such that the difference $\Delta^{n+s} - \Delta^n$ vanishes to zero as $n \to \infty$. Similar results can be obtained for the case where $x^n \neq x^*$ and $\Delta^n$ crosses the level 1 from below. Essentially, as $\Delta^n$ repeatedly crosses the level 1, it must stay closer and closer (converge) to that level as $n \to \infty$.

We now state the steps in this argument more formally. First, we establish a rate bound that essentially governs the change in $\Gamma^{n}_{x^*,x}$ over “small” time increments of length $O (\sqrt{\frac{\log \log n}{n}})$. (Again, we note that our analysis studies the simplified algorithm given in Figure 2; however, due to the consistency of BOLD, the limiting results proved for this simplified version will also hold for the original version.)

**Lemma 4.3.** Let Assumptions 4.1-4.4 hold. For all $x \neq x^*$ and all large enough $n$, we have

$$\left| u^{n}_{x^*,x} - u^{n+m}_{x^*,x} \right| = O \left( \sqrt{\frac{\log \log n}{n}} \right)$$

if $m = O (\sqrt{\frac{\log \log n}{n}})$.

Next, we show that, in between two samples of $x^*$, the number of samples that can be assigned to suboptimal alternatives also follows the rate $O (\sqrt{\frac{\log \log n}{n}})$, i.e., is “small” in the sense of
Lemma 4.3. Similarly, the number of samples that can be allocated to \( x^* \) in between two samples of any two suboptimal alternatives follows the same rate.

**Lemma 4.4.** Let Assumptions 4.1-4.4 hold. In between two samples of the optimal alternative, the number of samples that can be allocated to any suboptimal alternatives is \( O \left( \sqrt{\frac{\log \log n}{n}} \right) \). Symmetrically, between two samples of any suboptimal alternatives (that is, two time stages when (25) fails), the number of samples that could be allocated to the optimal alternative is also \( O \left( \sqrt{\frac{\log \log n}{n}} \right) \).

We can now complete the proof. From Lemma 4.4, we know that at most \( O \left( \sqrt{\frac{\log \log n}{n}} \right) \) time stages can transpire between two samples of \( x^* \), allowing Lemma 4.3 to be applied. The rate \( O \left( \sqrt{\frac{\log \log n}{n}} \right) \) obtained from Lemma 4.3 can then be used to show that, in between two samples of \( x^* \), \( \Delta^n \) is bounded below by \( 1 - O \left( \sqrt{\frac{\log \log n}{n}} \right) \). A symmetric upper bound can be obtained for time stages transpiring in between two samples of suboptimal alternatives, thus establishing convergence.

**Theorem 4.3.** Let Assumptions 4.1-4.4 hold. Then,

\[
\lim_{n \to \infty} \Delta^n = 1.
\]

**Proof:** If \( x^* \) is sampled at some large enough time \( n \), we have \( \Delta^n > 1 \) from the definition of the BOLD algorithm. Let

\[
m = \inf \left\{ l > 0 : \Delta^{n+l} > 1 \right\},
\]

whence \( \Delta^{n+m} > 1 \). By Lemma 4.4, we know that \( m = O \left( \sqrt{\frac{\log \log n}{n}} \right) \). Then, for any \( 0 \leq s \leq m \), Lemma 4.3 implies that

\[
\left| u_{x^*,x}^n - u_{x^*,x}^{n+s} \right| = O \left( \sqrt{\frac{\log \log n}{n}} \right).
\]

Similarly as in the proof of Lemma 4.4, for any \( 0 \leq s \leq m \) and all \( x \neq x^* \), we have

\[
\frac{I_{x^*}^{n+s} \left( u_{x^*,x}^{n+s} \right)}{I_{x^*}^{n+s} \left( u_{x^*,x}^n \right)} - \frac{I_x^n \left( u_{x^*,x}^n \right)}{I_x^n \left( u_{x^*,x}^n \right)} = I_{x^*} \left( u_{x^*,x}^{n+s} ; \mu_{x^*} \right) - I_{x^*} \left( u_{x^*,x}^n ; \mu_{x^*} \right) + O \left( \sqrt{\frac{\log \log n}{n}} \right) = \Theta \left( 1 \right) \left( u_{x^*,x}^{n+s} - u_{x^*,x}^n \right) + O \left( \sqrt{\frac{\log \log n}{n}} \right) = O \left( \sqrt{\frac{\log \log n}{n}} \right).
\]
It follows that $\Delta^{n+s} - \Delta^n = O\left(\sqrt{\log \log n} \frac{n}{\log n}\right)$. Therefore, there exists some fixed positive constant $C_1$ such that, for all $n$ and any $0 \leq s \leq m$, we have $\Delta^{n+s} \geq 1 - C_1\sqrt{\log \log n}$, which implies that $\lim \inf_{n \to \infty} \Delta^n = 1$. Symmetrically, one can show using very similar arguments that $\lim \sup_{n \to \infty} \Delta^n = 1$, which completes the proof. □

4.4 Individual balance conditions

Our final result is that $\lim_{n \to \infty} \Gamma_{x^*,y}^{n,n+\tau} = 1$ for any $y, z \neq x^*$, thus asymptotically satisfying the individual balance conditions (9). This proof also has several parts establishing bounds on $L_{x,x^*,y}^{n,n+\tau}$ for different combinations of $n$, $m$ and $x$. Specifically:

- In Lemma 4.5, we consider a large enough $n$ where $z \neq x^*$ is sampled, and choose $m$ such that $n + m$ is the next time that this same $z$ is sampled. We then show that $x^*$ can receive at most $O\left(\sqrt{n \log \log n}\right)$ samples between times $n$ and $n + m$.

- In Lemma 4.6, we define $n$ and $m$ in the same way, but now establish a bound of $O\left(\sqrt{n \log \log n}\right)$ on the number of samples that can be assigned to any $y \neq z$ between times $n$ and $n + m$.

- Analogously to the argument in Section 4.3, we combine these rates with Lemma 4.3 and show that the ratios $\Gamma_{x^*,y}^{n,n+\tau}$ are bounded below by $1 - O\left(\sqrt{\log \log n \frac{n}{\log n}}\right)$ and above by $1 + O\left(\sqrt{\log \log n \frac{n}{\log n}}\right)$, resulting in the desired convergence.

Note that the intermediate rate results are successively stronger; for instance, Lemma 4.4 considers samples of $x^*$ in between two samples of any two suboptimal alternatives, while Lemma 4.5 considers samples of $x^*$ in between two samples of the same $z \neq x^*$. Lemma 4.6 then generalizes this to samples of any $y \neq z$. Later proofs build on earlier ones, so it is not possible to omit any of these steps.

We now give formal statements of these steps and prove the final convergence result.

**Lemma 4.5.** For a time stage $n$ in which some suboptimal $z \neq x^*$ is sampled, define

$$m_n = \inf \left\{l > 0 : x^{n+l} = z \right\}.$$ 

Let Assumptions 4.1-4.4 hold. There exists a fixed positive constant $C < \infty$ and a time stage $n_0 \geq 3$ such that, for any $n \geq n_0$ in which some $z \neq x^*$ is sampled, we have $L_{x,x^*,y}^{n,n+\tau} \leq C\sqrt{n \log \log n}$. 

23
Lemma 4.6. For a time stage $n$ in which some suboptimal $z \neq x^*$ is sampled, define

$$m_n = \inf \left\{ l > 0 : x^{n+l} = z \right\}.$$  

Let Assumptions 4.1-4.4 hold. There exists a fixed positive constant $C < \infty$ and a time stage $n_0 \geq 3$ such that, for any $n \geq n_0$ in which some $z \neq x^*$ is sampled, we have $L_y^{n,n+m_n} \leq C\sqrt{n \log \log n}$ for any $y \neq z$.

**Theorem 4.4.** Let Assumptions 4.1-4.4 hold. For any $y,z \neq x^*$,

$$\lim_{n \to \infty} \frac{\Gamma_n^{y,z}}{\Gamma_n^{x^*,z}} = 1.$$  

**Proof:** For notational convenience, let $r_{y,z} = \frac{\Gamma_n^{y,z}}{\Gamma_n^{x^*,z}}$. Suppose that $z$ is sampled at stage $n$, and let $m_n = \inf \left\{ l > 0 : x^{n+l} = z \right\}$ as in previous lemmas. Obviously, we have $r_{y,z}^n \geq 1$ and $r_{y,z}^{n+m_n} \geq 1$. By Lemmas 4.5 and 4.6, we know that $m_n = O\left(\sqrt{n \log \log n}\right)$. Then, by Lemma 4.3, for any $0 \leq t \leq m_n$ we have

$$\left| u_{x^*,y}^n - u_{x^*,y}^{n+t} \right| = O\left(\sqrt{\frac{\log \log n}{n}}\right), \quad \left| u_{x^*,z}^n - u_{x^*,z}^{n+t} \right| = O\left(\sqrt{\frac{\log \log n}{n}}\right).$$

Now, for any $0 \leq t \leq m_n$,

$$r_{y,z}^{n+t} - r_{y,z}^n = \frac{\Gamma_n^{x^*,y} \Gamma_n^{x^*,z} - \Gamma_n^{x^*,z} \Gamma_n^{x^*,y}}{\Gamma_n^{x^*,z} \Gamma_n^{x^*,y}} = \frac{\left(\Gamma_n^{x^*,y} - \Gamma_n^{x^*,y}\right) \Gamma_n^{x^*,z} + \left(\Gamma_n^{x^*,z} - \Gamma_n^{x^*,z}\right) \Gamma_n^{x^*,y}}{\Gamma_n^{x^*,z} \Gamma_n^{x^*,y}} \geq -\frac{\Gamma_n^{x^*,y} - \Gamma_n^{x^*,y}}{\Gamma_n^{x^*,z} \Gamma_n^{x^*,y}} \left| \Gamma_n^{x^*,z} \right| + \frac{\Gamma_n^{x^*,z} - \Gamma_n^{x^*,z}}{\Gamma_n^{x^*,z} \Gamma_n^{x^*,y}} \left| \Gamma_n^{x^*,y} \right|.$$  

Note that $\Gamma_n^{x^*,y}, \Gamma_n^{x^*,z}, \Gamma_n^{x^*,y}, \Gamma_n^{x^*,z}$ are all $\Theta(n)$ by Theorem 4.2, Assumption 4.1 and Corollary 4.1.

We can further see that

$$\left| \Gamma_n^{x^*,y} - \Gamma_n^{x^*,y} \right| \leq \left| N_y \delta_y \left( u_{x^*,y}^{n+t} - u_{x^*,y}^n \right) \right| + \left| N_{x^*,y} \delta_y \left( u_{x^*,y}^{n+t} - u_{x^*,y}^n \right) \right| + \Theta(n) \left| \Gamma_n^{x^*,y} \left( u_{x^*,y}^{n+t} \right) - \Gamma_n^{x^*,y} \left( u_{x^*,y}^n \right) \right| + O\left(\sqrt{n \log \log n}\right).$$
We also have

\[
\left| I_{y}^{n+t}(u_{x,y}^{n+t}) - I_{y}^{n}(u_{x,y}^{n}) \right|
\leq \left| I_{y}(u_{x,y}^{n+t}; \mu_{y}) - I_{y}(u_{x,y}^{n}; \mu_{y}) \right| + \left| I_{y}(u_{x,y}^{n+t}; \rho_{y}^{n+t}) - I_{y}(u_{x,y}^{n}; \rho_{y}) \right|
\leq \left| I_{y}(u_{x,y}^{n+t}; \mu_{y}) - I_{y}(u_{x,y}^{n}; \mu_{y}) \right| + O\left( \sqrt{\frac{\log \log n}{n}} \right)
\]

(35)

(36)

where (34) is due to the triangle inequality, (35) holds by Assumption 4.4 and the law of the iterated logarithm, and (36) holds by Lemma 4.3. Using similar arguments, we can obtain

\[
\left| I_{y}^{n+t}(u_{x,y}^{n+t}) - I_{y}^{n}(u_{x,y}^{n}) \right| = O\left( \sqrt{\frac{\log \log n}{n}} \right).
\]

It follows that \( \left| \Gamma_{x,y}^{n+t} - \Gamma_{x,y}^{n} \right| = O\left( \sqrt{n \log \log n} \right) \), and, analogously, \( \left| \Gamma_{x,z}^{n+t} - \Gamma_{x,z}^{n} \right| = O\left( \sqrt{n \log \log n} \right) \).

Returning to (33), we obtain

\[
\frac{\Gamma_{x,y}^{n+t} - \Gamma_{x,y}^{n}}{\Gamma_{x,z}^{n+t} - \Gamma_{x,z}^{n}} \geq \frac{O\left( \sqrt{n \log \log n} \right) \Theta(n) + O\left( \sqrt{n \log \log n} \right) \Theta(n)}{\Theta(n) \Theta(n)} = O\left( \frac{\sqrt{\log \log n}}{n} \right).
\]

Thus, there must exist some fixed positive constant \( C \) such that, for all such large enough \( n \) and any \( 0 \leq t \leq m_{n} \),

\[
r_{y,z}^{n+t} \geq r_{y,z}^{n} - C \sqrt{\frac{\log \log n}{n}} \geq 1 - C \sqrt{\frac{\log \log n}{n}}.
\]

Since this holds for \( 0 \leq t \leq m_{n} \), and since \( r_{y,z}^{n+m_{n}} \geq 1 \) and the same arguments can be applied to the next “cycle” of time stages from \( n + m_{n} \) to the next time \( z \) is sampled, it follows that \( r_{y,z}^{n} \geq 1 - C \sqrt{\frac{\log \log n}{n}} \) for all large enough \( n \). It follows that \( \lim inf_{n \to \infty} r_{y,z}^{n} = 1 \). Symmetrically, we can show that \( \lim sup_{n \to \infty} r_{y,z}^{n} = 1 \), completing the proof.

\[ \square \]

5 Numerical illustrations

We briefly present some numerical illustrations of our theoretical results for instances of the example settings listed in Section 2.3 based on simulated data. The goal of these simulations is to
demonstrate the ability of BOLD to learn the optimal static allocation without tuning, as compared to several benchmarks. The ranking and selection literature has put forth many algorithms and heuristics, many of which are closely specialized to particular families of sampling distributions; however, as our focus in this paper is theoretical, we do not extensively investigate finite-time empirical performance. It is not possible to guarantee that BOLD will outperform every algorithm for every distribution and problem instance, but its generality and ease of implementation may make it an appealing computational benchmark to researchers working on specific problem classes.

All examples have $M = 30$ alternatives. Figure 3 considers the setting where $F_x$ is $\mathcal{N}(\mu_x, \lambda^2_x)$. The true means $\mu_x$ and sampling variances $\lambda^2_x$ are instantiated according to Sec. 5.1 in Ryzhov (2018). The sampling variances are made known to all algorithms, while the means are unknown. In addition to BOLD, we implemented the traditional expected improvement (EI) algorithm (Jones et al., 1998; Ryzhov, 2016), the Thompson sampling algorithm (Russo & Van Roy, 2014) with normal priors, and the top-two expected improvement method of Qin et al. (2017). These are included for illustrative purposes: EI and Thompson sampling are known to satisfy $\frac{N \lambda^2_x}{n} \rightarrow 1$, meaning that they do not sample suboptimal alternatives as often as prescribed by large deviations analysis. Top-two EI assigns a flat proportion $\frac{1}{2}$ of the budget to $x^\ast,n$ and uses a variant of expected improvement to choose among the suboptimal alternatives the rest of the time.\footnote{If this proportion is tuned to $\alpha^\ast_x$, top-two EI will learn the other proportions $\alpha^\ast_x$. Chen & Ryzhov (2017, 2019) shows how this tuning process can be automated; the resulting procedure is essentially identical to BOLD in the special case of normal distributions. Qin et al. (2017) recommends using the value $\frac{1}{2}$.}

We also solved (2) via brute force and obtained the optimal static allocation $\alpha^\ast$. Figure 3(a)
(a) Empirical allocations, $N = 10^5$.

(b) Values of $\Gamma \left( \frac{N_n}{n}; \mu \right)$.

Figure 4: Illustration with Bernoulli distributions.

compares $\alpha^*$ with the empirical budget allocations made by the algorithms after $N = 10^5$ samples; for better visualization, only the alternatives with the five largest values of $\mu_x$ are shown. Figure 3(b) shows the trajectory $\Gamma \left( \frac{N_n}{n}; \mu \right)$ over $n = 1, ..., N$. This quantity represents the theoretical rate exponent that would be attained if $\frac{N_n}{n}$ were a static allocation; recall that (2) aims to maximize this exponent, so we wish to increase it over time.

We see that, as expected, BOLD is the only method to learn $\alpha^*$ with high precision. EI and Thompson sampling spend virtually the entire budget on $x^*$; as this proportion converges to 1, the theoretical rate exponent in Figure 3(b) converges to zero. Top-two EI is forced to over-sample $x^*$, leading it to underestimate the remaining proportions. It is the only method other than BOLD for which the theoretical rate exponent increases over time; however, this growth eventually slows down as the predetermined allocation to $x^*$ means that the method will never be able to satisfy all of the optimality conditions.

In Figure 4, we let $F_x$ be Bernoulli ($\mu_x$) with $\mu_x$ drawn from a normal distribution with mean 0.5 and standard deviation 0.1, thus yielding many alternatives with rewards close to 0.5 (maximum noise) with some positive and negative outliers. As EI and TT-EI are specialized to normal sampling distributions, we do not include them here, but instead implement the top-two probability sampling (TTPS) method for Bernoulli rewards given in Appendix B of Russo (2019), which, like TT-EI, assigns a flat proportion $\frac{1}{2}$ of the budget to $x^*, n$ and has similar guarantees for the other proportions. In our experience, TTPS (and the closely related method of top-two Thompson sampling) proved to be much more computationally expensive than either BOLD or Thompson sampling, because it
requires numerical approximation of a difficult probability. Due to random variation in $x^*, n$, the empirical proportion assigned by TTPS to $x^*$ has not yet converged to $\frac{1}{2}$.

In Figure 5, we let $F_x$ be $Exp(\lambda_x)$ with $\lambda_x$ drawn from a uniform distribution on $[0.3, 0.7]$. We implement Thompson sampling (with gamma priors) and the specialized method of Ryzhov & Powell (2011) based on expected value of information. Finally, Figure 6 considers the noncentral chi-squared case from Section 2.3, with target level $c = 5$ and $(\mu_x, \lambda_x)$ set as in the normal case, and compares BOLD with Thompson sampling (with normal priors on $\mu$) and the specialized local time method (LTM) of Ryzhov (2018). The results are consistent with the other cases.
We also give a short discussion of finite-time performance. For brevity, we show (in Figure 7) results only for the instances with normal and noncentral chi-squared distributions. In the former case, we report the value $\mu_{x^*,n}$ of the alternative believed to be the best at time $n$, and in the latter case, we report $(\mu_{x^*,n} - c)^2$; thus, we aim to maximize in the former case and minimize in the latter. As shown in Gao et al. (2017), these metrics are closely related to PCS, and we display them here because they produce smoother trajectories visually. Under normal distributions, BOLD and EI are the most robust for small budgets (up to 100 samples), followed by a regime in which Thompson sampling and TT-EI pull slightly ahead, and finally ending with a regime in which the differences between methods cannot be discerned visually. However, in the noncentral chi-squared case, BOLD is clearly leading by $n = 15$, while LTM is a close second and Thompson sampling lags behind somewhat.

From these examples, we will not draw far-reaching conclusions about finite-time empirical performance, which is largely beyond the scope of this paper. These examples simply illustrate that BOLD is the only algorithm that can learn $\alpha^*$ without tuning, and that our sequential approach to this asymptotic goal does not conflict, in principle, with competitive finite-time performance. We also note that, in many cases, BOLD was the fastest of the algorithms we implemented, which, together with its generality, should make it useful as a benchmark in future methodological research.
6 Conclusion

We have proposed a simple, fast, and general algorithmic strategy for learning optimal static allocations in ranking and selection problems. While some similar heuristics can be found in recent papers (under additional distributional assumptions), we have given the first rigorous proof that this strategy asymptotically solves the optimality conditions obtained from large deviations theory. Notably, our BOLD algorithm does this with no tuning required.

Our approach provides a computational foundation for the growing literature, originating from Glynn & Juneja (2004), that derives optimal static allocations in a variety of settings (more broad than only ranking and selection) and then uses them as guidance for adaptive ones. Past work of this kind focused on characterizing the optimality conditions and subsequently relied on ad-hoc heuristics (which may require tuning or be computationally costly) to actually solve them in an adaptive setting. Our approach, however, is guaranteed to learn the solution in a general setting while remaining tuning-free and computationally efficient. We hope that BOLD will come to be viewed as the standard algorithmic strategy for methods based on large deviations analysis, in ranking and selection and also in the broader context where such methods have been proposed.

References


7 Appendix: proofs

In this section, we provide full proofs for all results that were stated in the main text.
7.1 Proof of Lemma 4.1

Let $\tilde{\theta}_z = \lim_{n \to \infty} \theta^n_z$ denote the final value of the sample mean for alternative $z$. There must exist some $y \neq z$ that is sampled infinitely often, i.e. $N^n_y \to \infty$ and $\theta^n_y \to \mu_y$. Obviously, if $\mu_y > \tilde{\theta}_z$, then $x^{*,n} \neq z$ for large enough $n$.

Suppose that $\mu_y < \tilde{\theta}_z$. This implies that $\theta^n_y < \tilde{\theta}_z$ for all large enough $n$. We will show by contradiction that $x^{*,n} \neq z$ for all large enough $n$. Suppose that there are infinitely many $n$ for which $x^{*,n} = z$. Recall that $u^n_{z,y}$ satisfies

$$N^n_z D^n_z (u^n_{z,y}) + N^n_y D^n_y (u^n_{z,y}) = 0,$$

whence

$$\frac{D^n_y (u^n_{z,y})}{D^n_z (u^n_{z,y})} = \frac{N^n_z}{N^n_y},$$

and the right-hand side of (37) vanishes as $n \to \infty$. From Lemma 2.1, it follows that $D^n_y (u^n_{z,y}) \to 0$, whence, by Assumption 4.1, it follows that $u^n_{z,y} \to 0$. By Assumption 4.2, we obtain $I^n_y (u^n_{z,y}) \to 0$. At the same time, since $\theta^n_y \to \mu_y$, we also have $u^n_{z,y} \to \mu_y$ due to Lemma 2.1. Consequently, for all large enough $n$, we have

$$I_z (u^n_{z,y}; \tilde{\theta}_z) > I_z \left( \frac{\tilde{\theta}_z + \mu_y}{2}; \tilde{\theta}_z \right) > 0.$$  

Therefore, for all large enough $n$, we have $\frac{I_z (u^n_{z,y}; \tilde{\theta}_z)}{I^n_y (u^n_{z,y})} > 1$. For any large enough $n$ where $x^{*,n} = z$, we have

$$\sum_{x \neq x^{*,n}} \frac{I^n_{x^{*,n}} (u^n_{x^{*,n},z})}{I^n_x (u^n_{x^{*,n},z})} \geq I_z (u^n_{z,y}; \tilde{\theta}_z) \frac{I^n_y (u^n_{z,y}; \theta^n_y)}{I^n_y (u^n_{z,y})} > 1,$$

which means that BOLD will sample alternative $z$. This contradicts the assumption that alternative $z$ is only sampled finitely many times; thus, for all large enough $n$, $x^{*,n} \neq z$.

Finally, we consider the pathological case where $\mu_y = \tilde{\theta}_z$, which can only arise for discrete sampling distributions. Again, we proceed by contradiction; suppose that there are infinitely many time stages $n$ such that $x^{*,n} = z$. For any such $n$ large enough, we have $\theta^n_y < \tilde{\theta}_z$. Now, we claim that, for all large enough $n$ where $x^{*,n} = z$, we must have

$$\frac{I_z (u^n_{z,y}; \tilde{\theta}_z)}{I^n_y (u^n_{z,y})} > 1.$$  

(38)
Suppose that (38) does not hold. Then, it follows that \( I_z(u^n_{z,y}; \bar{\theta}_z) \leq I^n_y(u^n_{z,y}) \). On one hand, we have

\[
\Gamma^n_{z,y} = N^n_z I^n_z(u^n_{z,y}) + N^n_y I^n_y(u^n_{z,y}) \\
\geq (N^n_z + N^n_y) I^n_z(u^n_{z,y}) \\
\geq N^n_y I^n_z(u^n_{z,y}).
\]

On the other hand, we have

\[
\Gamma^n_{z,y} = \inf_u N^n_z I^n_z(u) + N^n_y I^n_y(u) \leq N^n_z I^n_z(\theta^n_y),
\]

From this, we conclude that \( N^n_y I^n_z(u^n_{z,y}) \leq \Gamma^n_{z,y} \leq N^n_z I^n_z(\theta^n_y) \), whence

\[
\frac{I^n_y(\theta^n_y)}{I^n_z(u^n_{z,y})} \geq \frac{N^n_y}{N^n_z},
\]

where the right-hand side converges to infinity. Then,

\[
\frac{I^n_z(\theta^n_y) - I^n_z(u^n_{z,y})}{I^n_z(u^n_{z,y}) - I^n_z(\bar{\theta}_z)} \geq \frac{N^n_y - N^n_z}{N^n_z},
\]

where the right-hand side also converges to infinity. Because \( \theta^n_y \to \bar{\theta}_z \) by assumption, both \( \theta^n_y - u^n_{z,y} \) and \( u^n_{z,y} - \bar{\theta}_z \) will be small terms. By Taylor’s expansion,

\[
\frac{I^n_z(\theta^n_y) - I^n_z(u^n_{z,y})}{I^n_z(u^n_{z,y}) - I^n_z(\bar{\theta}_z)} \approx \frac{D^n_z(u^n_{z,y}) (\theta^n_y - u^n_{z,y})}{D^n_y(u^n_{z,y}) (u^n_{z,y} - \bar{\theta}_z)} = \frac{u^n_{z,y} - \theta^n_y}{\theta_z - u^n_{z,y}}.
\]

Thus,

\[
\frac{u^n_{z,y} - \theta^n_y}{\theta_z - u^n_{z,y}} \to \infty.
\]

At the same time, recall from (37) that \( \left| \frac{D^n_y(u^n_{z,y})}{D^n_z(u^n_{z,y})} \right| = \frac{N^n_y}{N^n_z} \), which converges to infinity. However, by Assumption 4.3, we have \( |D^n_y(\theta_z)| \geq C |D^n_z(\theta^n_y)| \), where \( C \) is a constant. Then,

\[
\frac{D^n_y(V)}{D^n_z(u^n_{z,y})} \geq C \frac{N^n_y}{N^n_z},
\]

which implies

\[
\frac{D^n_y(V) - D^n_y(u^n_{z,y})}{D^n_y(u^n_{z,y}) - D^n_y(\theta^n_y)} \geq \frac{CN^n_y - N^n_z}{N^n_z} \to \infty.
\]
Through a similar analysis to the above, we will have

\[
\frac{\bar{\theta}_z - u^n_{z,y}}{u^n_{z,y} - \bar{\theta}^n_y} \to \infty,
\]

which contradicts (39). Therefore, we must have (38), which means alternative \( z \) will be sampled if \( x^{*,n} = z \) at these stages. This contradicts the assumption that alternative \( z \) is only sampled finitely many times, completing the proof.

7.2 Proof of Lemma 4.2

Denote by \( A = \{ x : N^n_x \to \infty \} \) the set of all alternatives that are sampled infinitely many times. Obviously \( A \) is non-empty. Note that \( \min_{x \neq y} |\mu_y - \mu_x| > 0 \), whence it follows that, if there is more than one alternative in \( A \), then \( \min_{x,y \in A} |\theta^n_y - \theta^n_x| > 0 \) for \( x \neq y \) and all large enough \( n \).

Therefore, we can take \( n \) to be large enough such that no alternative in \( A^c \) will be sampled from the \( n \)th stage onwards, and \( \min_{x \in A} N^n_x > \max_{x' \in A^c} N^n_{x'} \). For such \( n \), the only possible way in which (18) is not checked is if there exist \( x \in A \) and \( x' \in A^c \) such that \( \theta^n_x = \theta^n_{x'} \). By construction, in Step 1 of Figure 1, we will sample \( x' \) in such a situation, but this will contradict the fact \( x' \) cannot be sampled for \( n \) large enough. Consequently, (18) will be checked for all large enough \( n \).

The second part is proved by contradiction. First, for all \( x' \in A^c \), define \( \bar{\theta}_{x'} = \lim_{n \to \infty} \theta^n_{x'} \).

Suppose that, condition (18) holds for all large enough \( n \) (i.e., it fails only finitely many times). From this it follows that there must be at least two alternatives in \( A \). To prove this (again by contradiction), suppose first that there is only one alternative, which we denote by \( z \). Then, \( A^c = \{ x' : x' \neq z \} \). Since (18) holds for all large enough \( n \), and \( z \) is the only alternative that is sampled infinitely often, it follows that, for all large enough \( n \), we have \( x^{*,n} = z \). In order for this to happen, it is necessary to have \( \mu_z > \max_{x' \in A^c} \bar{\theta}_{x'} \).

However, by repeating the analysis in the proof of Lemma 4.1, we can also see that for any alternative \( x' \neq z \), i.e., for any \( x' \in A^c \), we have \( u^n_{z,x'} \to \mu_z \) and \( I^n_z(u^n_{z,x'}) \to 0 \), and at the same time, for all large enough \( n \), we have \( I^n_{x'}(u^n_{z,x'}) > I_{x'}(\frac{\theta_{x'} + \mu_z}{2}; \bar{\theta}_{x'}) > 0 \). Therefore, for any alternative \( x' \neq z \) and for all large enough \( n \), we must have

\[
\frac{I^n_z(u^n_{z,x'})}{I^n_{x'}(u^n_{z,x'})} < \frac{1}{M},
\]
whence
\[
\sum_{x' \neq x^*, n} \frac{I_{x^*, n}^n (u_{x^*, n, x'}^n)}{I_x^n (u_{x^*, n, x'}^n)} = \sum_{x' \neq z} \frac{I_z^n (u_{x^*, z, x'}^n)}{I_x^n (u_{x^*, z, x'}^n)} < \sum_{x' \neq z} \frac{1}{M} < 1,
\]
so condition (18) does not hold, producing the desired contradiction. Thus, there must be at least two alternatives in \(A\).

Consequently, we can take \(y, z \in A\) with \(y \neq z\) and \(\lim_{n \to \infty} \theta_{y}^n = \mu_z\) and \(\lim_{n \to \infty} \theta_{y}^n = \mu_y\). Without loss of generality, suppose that \(\mu_y < \mu_z\). Then, for all large enough \(n\), we will have \(\theta_{y}^n < \theta_{z}^n\), whence \(x^{*, n} \neq y\). At the same time, from the supposition that (18) holds for all large enough \(n\), we know that only \(x^{*, n}\) is sampled after some time. Consequently, there is a time after which \(y\) is never sampled, contradicting the fact that \(y \in A\). We conclude that there are infinitely many time stages at which (18) does not hold.

7.3 Proof of Lemma 4.3

Recall that
\[
N_{x^*} D_{x^*} (u_{x^*, x}^n) + N_x D_x (u_{x^*, x}^n) = 0, \tag{40}
\]
\[
N_{x^*} D_{x^*} (u_{x^*, x}^{n+m}) + N_x D_x (u_{x^*, x}^{n+m}) = 0, \tag{41}
\]
with \(N_{x^*} = N_x + L_{x^*}^{n+m}\) (and similarly for \(x^*\)). Then, from (41), we have
\[
N_{x^*} D_{x^*} (u_{x^*, x}^{n+m}) + N_x D_x (u_{x^*, x}^{n+m}) = - \left( L_{x^*}^{n+m} D_{x^*} (u_{x^*, x}^{n+m}) + L_x D_x (u_{x^*, x}^{n+m}) \right) = O(1) \cdot \left( L_x^{n+m} + L_{x^*}^{n,m} \right) \tag{42}
\]
\[
= O \left( \sqrt{n \log \log n} \right), \tag{43}
\]
where (42) holds since \(u_{x^*, x}^{n+m} \in [\theta_{x^*}^{n+m}, \theta_{x^*}^{n+m}]\) and \(\theta^n \to \mu\), which bounds both \(D_{x^*} (u_{x^*, x}^{n+m})\) and \(D_x (u_{x^*, x}^{n+m})\), and (43) holds because \(\max \{ L_x^{n,m}, L_{x^*}^{n,m} \} = O(m) = O \left( \sqrt{n \log \log n} \right)\). Also note that
\[
D_{x^*} (u_{x^*, x}^{n+m}) = D_{x^*} (u_{x^*, x}^{n+m}, \theta_{x^*}^{n+m})
\]
where (44) holds by Assumption 4.4 together with the mean value theorem, and (45) holds due to the law of the iterated logarithm as long as \( n \) is large enough. In a similar way, we also have

\[
D_{x^*}^{n+m} \left( u_{x^*, x}^{n+m} \right) = D_x \left( u_{x^*, x}^n; \mu_x \right) + O \left( \sqrt{\frac{\log n \log \log n}{n}} \right), \tag{46}
\]

\[
D_x^n \left( u_{x^*, x}^n \right) = D_x \left( u_{x^*, x}^n; \mu_x \right) + O \left( \sqrt{\frac{\log n \log \log n}{n}} \right), \tag{47}
\]

\[
D_x^n \left( u_{x^*, x}^n \right) = D_x \left( u_{x^*, x}^n; \mu_x \right) + O \left( \sqrt{\frac{\log n \log \log n}{n}} \right). \tag{48}
\]

Denote by \( J_x (u; \theta) = \frac{d}{du} D_x (u; \theta) = \frac{d^2}{du^2} I_x (u; \theta) \). From (43), we have

\[
O \left( \sqrt{n \log n \log \log n} \right) = N_x^n \left( D_{x^*}^{n+m} \left( u_{x^*, x}^{n+m} \right) - D_{x^*}^n \left( u_{x^*, x}^n \right) \right) + N_x^n \left( D_x^{n+m} \left( u_{x^*, x}^{n+m} \right) - D_x^n \left( u_{x^*, x}^n \right) \right), \tag{49}
\]

\[
= N_x^n \left( O \left( \sqrt{\frac{\log n \log \log n}{n}} \right) + D_{x^*}^n \left( u_{x^*, x}^{n+m}; \mu_x^* \right) - D_{x^*}^n \left( u_{x^*, x}^n; \mu_x^* \right) \right) \tag{50}
\]

\[
+ N_x^n \left( O \left( \sqrt{\frac{\log n \log \log n}{n}} \right) + D_x^n \left( u_{x^*, x}^{n+m}; \mu_x \right) - D_x^n \left( u_{x^*, x}^n; \mu_x \right) \right) \tag{51}
\]

\[
= O \left( \sqrt{n \log n \log \log n} \right) + N_x^n J_{x^*} \left( \xi_{x^*, x}^{n,m+m}; \mu_x^* \right) \left( u_{x^*, x}^{n+m} - u_{x^*, x}^n \right) \tag{52}
\]

\[
+ N_x^n J_x \left( \eta_{x^*, x}^{n,m+m}; \mu_x \right) \left( u_{x^*, x}^{n+m} - u_{x^*, x}^n \right) \tag{53}
\]

\[
= O \left( \sqrt{n \log n \log \log n} \right) + \Theta \left( n^{m+m} \right) \left( u_{x^*, x}^{n+m} - u_{x^*, x}^n \right), \tag{54}
\]

where (49) is due to (40), equation (50) follows from (45)-(48), and both (51) and (54) hold because \( N_x^n = \Theta \left( n \right) \) for all \( x \) as shown in Theorem 4.2. Equation (52) holds by the mean value theorem, with

\[
\xi_{x^*, x}^{n,m+m}, \eta_{x^*, x}^{n,m+m} \in \left[ \min \left\{ u_{x^*, x}^{n}, u_{x^*, x}^{n+m} \right\}, \max \left\{ u_{x^*, x}^{n}, u_{x^*, x}^{n+m} \right\} \right].
\]
Finally, (53) holds since both $J_{x^*}(u; \mu_{x^*})$ and $J_x(u; \mu_x)$ are strictly positive for $u \in [\mu_x, \mu_{x^*}]$ by Lemma 2.1. The desired conclusion then follows.

### 7.4 Proof of Lemma 4.4

Suppose that, at some large enough time stage $n$, we have $\Delta^n > 1$ and $\Delta^{n+1} < 1$, which means that $x^*$ is sampled at stage $n$ and then a suboptimal alternative is sampled at stage $n + 1$. Let

$$s_n = \inf \left\{ l > 0 : \Delta^{n+l} > 1 \right\},$$

so that stage $n + s_n$ is the first time that $x^*$ is sampled after time $n$. We wish to show that $s_n = O\left(\sqrt{n \log \log n}\right)$. (The symmetric case where we consider the number of samples of $x^*$ in between samples of any two $x, y \neq x^*$ can be proved in an identical manner.)

For any stage $n + m$ where $0 \leq m \leq s_n$, we denote $\kappa = \frac{m}{\sqrt{n \log \log n}}$ for simplicity. First, for all large enough $n$, we have

$$\Delta^n = \sum_{x \neq x^*} \frac{I_{x^*}(u^n_{x^*,x}; \theta^n_{x^*})}{I_x(u^n_{x^*,x}; \theta^n_x)} = \sum_{x \neq x^*} \frac{I_{x^*}(u^n_{x^*,x}; \mu_{x^*}) + I_{x^*}(u^n_{x^*,x}; \theta^n_{x^*}) - I_x(u^n_{x^*,x}; \mu_x)}{I_x(u^n_{x^*,x}; \mu_x)}$$

$$= \sum_{x \neq x^*} \frac{I_{x^*}(u^n_{x^*,x}; \mu_{x^*}) + \Theta_x(u^n_{x^*,x}; \mu_{x^*}) - I_x(u^n_{x^*,x}; \mu_x)}{I_x(u^n_{x^*,x}; \mu_x)} = \sum_{x \neq x^*} \frac{I_{x^*}(u^n_{x^*,x}; \mu_{x^*}) + \Theta_x(u^n_{x^*,x}; \mu_{x^*})}{I_x(u^n_{x^*,x}; \mu_x)}$$

in which (55) holds because of Assumption 4.4 together with the mean value theorem, (56) holds due to the law of the iterated logarithm, and (57)-(58) hold by Assumption 4.1 and Corollary 4.1. Since $\Delta^n > 1$, there must exist a fixed positive constant $C_1$ such that

$$\sum_{x \neq x^*} \frac{I_{x^*}(u^n_{x^*,x}; \mu_{x^*})}{I_x(u^n_{x^*,x}; \mu_x)} > 1 - C_1 \sqrt{\frac{\log \log n}{n}}.$$  

39
Similarly, as before, from Assumption 4.4 and the law of the iterated logarithm, (61)-(62) yield
\[
\Delta^{n+m} = \sum_{x \neq x^*} \frac{I_x(u^{n+m}_{x,x}; \mu_x^*)}{I_x(u^{n+m}_{x,x}; \mu_x^*)} + O\left(\sqrt{\frac{\log \log n}{n}}\right)
\]
\[
> \sum_{x \neq x^*} \frac{I_x(u^{n+m}_{x,x}; \mu_x^*)}{I_x(u^{n+m}_{x,x}; \mu_x^*)} - C_2 \sqrt{\frac{\log \log n}{n}},
\]

where \(C_2\) is another fixed positive constant. Then, in order to have \(\Delta^{n+m} > 1\), from (59), it is sufficient to have
\[
\sum_{x \neq x^*} \frac{I_x(u^{n+m}_{x,x}; \mu_x^*)}{I_x(u^{n+m}_{x,x}; \mu_x^*)} - C_2 \sqrt{\frac{\log \log n}{n}} \geq \sum_{x \neq x^*} \frac{I_x(u^n_{x,x}; \mu_x^*)}{I_x(u^n_{x,x}; \mu_x^*)} + C_1 \sqrt{\frac{\log \log n}{n}},
\]
which could be written as
\[
\sum_{x \neq x^*} \frac{I_x(u^{n+m}_{x,x}; \mu_x^*)}{I_x(u^{n+m}_{x,x}; \mu_x^*)} - \sum_{x \neq x^*} \frac{I_x(u^n_{x,x}; \mu_x^*)}{I_x(u^n_{x,x}; \mu_x^*)} \geq C_3 \sqrt{\frac{\log \log n}{n}} \quad (60)
\]

where \(C_3 = C_1 + C_2\) is another positive constant. As this would lead to a contradiction \((\Delta^{n+m} > 1 \implies \text{is sampled, which cannot occur until time } s_n)\), we show that \(\kappa\) must be bounded above by a constant in order to prevent (60) from occurring, which will yield the desired result.

Note that, since \(m = \kappa \sqrt{n \log \log n}\), there must be at least one alternative \(y \neq x^*\) such that \(L_y^{n+m} \geq \frac{\kappa}{\theta^m} \sqrt{n \log \log n}\). We know that \(u^n_{x,y}\) and \(u^{n+m}_{x,y}\) satisfy
\[
- \frac{D_x^* (u^n_{x,y}, \theta^n_x)}{D_y(u^n_{x,y}, \theta^n_y)} = \frac{N^n_y}{N^n_x}, \quad (61)
\]
\[
- \frac{D_x^* (u^{n+m}_{x,y}, \theta^{n+m}_{x,y})}{D_y(u^{n+m}_{x,y}, \theta^{n+m}_{x,y})} = \frac{N^{n+n+m}_y}{N^{n+m}_x} = \frac{N^n_y + L^{n+n+m}_y}{N^n_x + 1}. \quad (62)
\]

Similarly, as before, from Assumption 4.4 and the law of the iterated logarithm, (61)-(62) yield
\[
- \frac{D_x^* (u^n_{x,y}, \mu_x^*)}{D_y(u^n_{x,y}, \mu_y)} = \frac{N^n_y}{N^n_x} + O\left(\sqrt{\frac{\log \log n}{n}}\right), \quad (63)
\]
\[
- \frac{D_x^* (u^{n+m}_{x,y}, \mu_x^*)}{D_y(u^{n+m}_{x,y}, \mu_y)} = \frac{N^n_y + L^{n+n+m}_y}{N^n_x + 1} + O\left(\sqrt{\frac{\log \log n}{n}}\right). \quad (64)
\]
By subtracting (63) from (64), we obtain

$$
\frac{D_x^* \left( u_{x^*,y}^{n+m}; \mu_x^* \right)}{D_y \left( u_{x^*,y}^{n+m}; \mu_y \right)} + \frac{D_x^* \left( u_{x^*,y}^n; \mu_x^* \right)}{D_y \left( u_{x^*,y}^n; \mu_y \right)} = \frac{N_y^n + L_y^{n,m}}{N_x^n + 1} - \frac{N_y^n}{N_x^n} + O \left( \sqrt{\log \log n} \right)
$$

$$\geq \frac{N_y^n + L_y^{n,m}}{N_x^n + 1} - \frac{N_y^n}{N_x^n} - C_4 \sqrt{\log \log \frac{n}{n}}, \quad (65)
$$

where $C_4$ is another fixed positive constant.

Equation (65) can be further manipulated to obtain

$$
\frac{N_y^n + L_y^{n,m}}{N_x^n + 1} - \frac{N_y^n}{N_x^n} - C_4 \sqrt{\log \log \frac{n}{n}} = \frac{I_y^{n,m}}{N_x^n + 1} - \left( \frac{N_y^n}{N_x^n} - \frac{N_y^n}{N_x^n + 1} \right) - C_4 \sqrt{\log \log \frac{n}{n}}
$$

$$\geq \frac{\kappa}{M} \sqrt{\log \log \frac{n}{n}} - \frac{N_y^n}{N_x^n + 1} - 1 - C_4 \sqrt{\log \log \frac{n}{n}} \quad (66)
$$

$$\geq \left( \frac{\kappa}{M} - C_4 \right) \sqrt{\log \log \frac{n}{n}} - \frac{C_5}{n} \quad (67)
$$

$$\geq \left( \frac{\kappa}{M} - C_6 \right) \sqrt{\log \log \frac{n}{n}} \quad (68)
$$

where $C_5, C_6$ are positive constants. Inequality (66) holds because $L_y^{n,m} \geq \frac{\kappa}{M} \sqrt{n \log \log n}$, while (67) holds since $N_x^n = \Theta(n)$ by Theorem 4.2, and (68) holds since $\frac{1}{n} \leq \sqrt{\log \log \frac{n}{n}}$ for large enough $n$.

Without loss of generality, suppose that $\kappa \geq M (C_6 + 1)$; if $\kappa$ is smaller than this constant, the desired result already follows. Then, from (65) and (68), we have

$$
- \frac{D_x^* \left( u_{x^*,y}^{n+m}; \mu_x^* \right)}{D_y \left( u_{x^*,y}^{n+m}; \mu_y \right)} + \frac{D_x^* \left( u_{x^*,y}^n; \mu_x^* \right)}{D_y \left( u_{x^*,y}^n; \mu_y \right)} > 0. \quad (69)
$$

We then derive

$$
- \frac{D_x^* \left( u_{x^*,y}^{n+m}; \mu_x^* \right)}{D_y \left( u_{x^*,y}^{n+m}; \mu_y \right)} + \frac{D_x^* \left( u_{x^*,y}^n; \mu_x^* \right)}{D_y \left( u_{x^*,y}^n; \mu_y \right)} = - \frac{D_x^* \left( u_{x^*,y}^{n+m}; \mu_x^* \right)}{D_y \left( u_{x^*,y}^{n+m}; \mu_y \right)} + \frac{D_x^* \left( u_{x^*,y}^n; \mu_x^* \right)}{D_y \left( u_{x^*,y}^n; \mu_y \right)}
$$

$$= - \frac{D_x^* \left( u_{x^*,y}^{n+m}; \mu_x^* \right)}{D_y \left( u_{x^*,y}^{n+m}; \mu_y \right)} D_y \left( u_{x^*,y}^n; \mu_y \right) + D_y \left( u_{x^*,y}^{n+m}; \mu_y \right) D_y \left( u_{x^*,y}^n; \mu_y \right) - D_y \left( u_{x^*,y}^{n+m}; \mu_y \right) D_y \left( u_{x^*,y}^n; \mu_y \right) D_y \left( u_{x^*,y}^{n+m}; \mu_y \right) D_y \left( u_{x^*,y}^n; \mu_y \right) - D_y \left( u_{x^*,y}^{n+m}; \mu_y \right) D_y \left( u_{x^*,y}^n; \mu_y \right) D_y \left( u_{x^*,y}^{n+m}; \mu_y \right) D_y \left( u_{x^*,y}^n; \mu_y \right)
$$
\[ -\Theta (1) \left( D_{x^*} \left( u_{x^*, y}^{n+m}; \mu_{x^*} \right) - D_{x^*} \left( u_{x^*, y}^{n}; \mu_{x^*} \right) \right) + \Theta (1) \left( D_y \left( u_{x^*, y}^{n}; \mu_y \right) - D_y \left( u_{x^*, y}^{n+m}; \mu_y \right) \right) \]

(70)

\[ = \Theta (1) \left( u_{x^*, y}^{n} - u_{x^*, y}^{n+m} \right), \]

(71)

where (70) holds by Assumption 4.1 and Corollary 4.1, while (71) holds by Lemma 2.1 and the mean value theorem. Then, from (69) we have

\[ u_{x^*, y}^{n} - u_{x^*, y}^{n+m} \geq C_7 \left( \frac{D_{x^*} \left( u_{x^*, y}^{n+m}; \mu_{x^*} \right) + D_y \left( u_{x^*, y}^{n+m}; \mu_y \right)}{D_y \left( u_{x^*, y}^{n}; \mu_y \right)} \right), \]

(72)

where \( C_7 \) is a fixed positive constant.

Now, combining (65) with (68) and (72), we have

\[ u_{x^*, y}^{n} - u_{x^*, y}^{n+m} \geq C_7 \left( \frac{I_{x^*} \left( u_{x^*, y}^{n+m}; \mu_{x^*} \right) + I_{y} \left( u_{x^*, y}^{n+m}; \mu_y \right)}{I_{y} \left( u_{x^*, y}^{n}; \mu_y \right)} \right), \]

(73)

Similarly to the derivation of (71), we derive

\[ = -\Theta (1) \left( I_{x^*} \left( u_{x^*, y}^{n+m}; \mu_{x^*} \right) - I_{x^*} \left( u_{x^*, y}^{n}; \mu_{x^*} \right) \right) I_{y} \left( u_{x^*, y}^{n+m}; \mu_y \right) - I_{y} \left( u_{x^*, y}^{n}; \mu_y \right) I_{x^*} \left( u_{x^*, y}^{n+m}; \mu_{x^*} \right) \]

(74)

where \( C_8 \) is a fixed positive constant, and (74) follows from (73).

The same analysis can be repeated for a different suboptimal \( x \neq y \), leading to

\[ \geq C_8 \left( \frac{\kappa}{M} - C_6 \right) \sqrt{\frac{\log \log n}{n}}, \]

(75)
as in (71). We then have

\[ C \]

with \( C \) analogously to (65) and (68), with positive constants \( C_9, C_{10} \) and \( C_{11} \). We also obtain

\[
\frac{D_{x^*} \left( u_{x^*,x}^{n+m}; \mu_{x^*} \right)}{D_x \left( u_{x,x}^{n+m}; \mu_x \right)} + \frac{D_{x^*} \left( u_{x^*,x}^{n}; \mu_{x^*} \right)}{D_x \left( u_{x,x}^{n}; \mu_x \right)} = \Theta \left( 1 \right) \left( u_{x^*,x}^{n} - u_{x,x}^{n+m} \right)
\]

as in (71). We then have

\[
u_{x^*,x}^{n} - u_{x,x}^{n+m} \geq C_{12} \min \left\{ 0, -\frac{D_{x^*} \left( u_{x^*,x}^{n+m}; \mu_{x^*} \right)}{D_x \left( u_{x,x}^{n+m}; \mu_x \right)} + \frac{D_{x^*} \left( u_{x^*,x}^{n}; \mu_{x^*} \right)}{D_x \left( u_{x,x}^{n}; \mu_x \right)} \right\}.
\]

with \( C_{12} \) being a fixed positive constant.

Now combining (75) and (76), we obtain

\[
u_{x^*,x}^{n} - u_{x,x}^{n+m} \geq -C_{11} C_{12} \sqrt{\frac{\log \log n}{n}}.
\]

Then, similarly as in (74),

\[
\frac{I_{x^*} \left( u_{x^*,x}^{n+m}; \mu_{x^*} \right)}{I_x \left( u_{x,x}^{n+m}; \mu_x \right)} - \frac{I_{x^*} \left( u_{x^*,x}^{n}; \mu_{x^*} \right)}{I_x \left( u_{x,x}^{n}; \mu_x \right)} = \Theta \left( 1 \right) \left( u_{x^*,x}^{n} - u_{x,x}^{n+m} \right)
\]

\[
\geq C_{13} \min \left\{ 0, u_{x^*,x}^{n} - u_{x,x}^{n+m} \right\}
\]

\[
\geq -C_{14} \sqrt{\frac{\log \log n}{n}}.
\]

with \( C_{13} \) and \( C_{14} = C_{11} C_{12} C_{13} \) being fixed positive constants.

Finally, combining (74) and (78), we return to the LHS of (60) and derive

\[
\sum_{x \neq x^*} \frac{I_{x^*} \left( u_{x^*,x}^{n+m}; \mu_{x^*} \right)}{I_x \left( u_{x,x}^{n+m}; \mu_x \right)} - \sum_{x \neq x^*} \frac{I_{x^*} \left( u_{x^*,x}^{n}; \mu_{x^*} \right)}{I_x \left( u_{x,x}^{n}; \mu_x \right)} = \sum_{x \neq x^*, y} \frac{I_{x^*} \left( u_{x^*,x}^{n+m}; \mu_{x^*} \right)}{I_x \left( u_{x,x}^{n+m}; \mu_x \right)} - \sum_{x \neq x^*, y} \frac{I_{x^*} \left( u_{x^*,x}^{n}; \mu_{x^*} \right)}{I_x \left( u_{x,x}^{n}; \mu_x \right)}
\]

\[
+ \frac{I_{x^*} \left( u_{x^*,y}^{n+m}; \mu_{x^*} \right)}{I_y \left( u_{x,y}^{n+m}; \mu_y \right)} - \frac{I_{x^*} \left( u_{x^*,y}^{n}; \mu_{x^*} \right)}{I_y \left( u_{x,y}^{n}; \mu_y \right)}
\]

\[
\geq -MC_{14} \sqrt{\frac{\log \log n}{n}} + C_8 \left( \frac{\kappa}{M} - C_6 \right) \sqrt{\frac{\log \log n}{n}}
\]
\[ C_8 \left( \frac{\kappa}{M} - C_6 \right) - MC_{14} \left( \log \log \frac{n}{n} \right). \]  

(79)

Since \( M, C_3, C_6, C_8, C_{14} \) are all fixed positive constants, (79) implies that if

\[ \kappa \geq \max \left\{ M (C_6 + 1), M \left( \frac{MC_{14} + C_3}{C_8} + C_6 \right) \right\}, \]

then (60) is satisfied. This implies that, at stage \( n + m \), where \( m = \kappa \sqrt{n \log \log n} \), we actually have \( \Delta^{n+m} > 1 \), whence \( x^* \) should be sampled at stage \( n + m \). This, however, contradicts the definition of \( m \), as \( x^* \) cannot be sampled until time \( s_n \). Therefore, \( \kappa \) must be bounded above by a constant, whence \( s_n = O \left( \sqrt{n \log \log n} \right) \), as required.

7.5 Proof of Lemma 4.5

Before proceeding, we first state a technical lemma that establishes a relationship between \( L_{x^*,n}^{n+m} \) and samples allocated to suboptimal alternatives. This result is proved in a separate subsection of the Appendix.

**Lemma 7.1.** For a time stage \( n \) in which some suboptimal \( z \neq x^* \) is sampled, define

\[ m_n = \inf \left\{ l > 0 : x^{n+l} = z \right\}, \quad t_n = \sup \left\{ l < m_n : x^{n+l} \neq x^* \right\}, \]

and \( \kappa_n = \frac{L_{x^*,n}^{n+m}}{\sqrt{n \log \log n}} \).

Let Assumptions 4.1-4.4 hold. Then, there exist fixed positive constants \( C_1, C_2 \) such that, for any large enough \( n \) where \( x^n = z \), if \( \kappa_n \geq C_1 \), there exists some suboptimal alternative \( y \neq z \) and a time stage \( n + s_n \), such that \( x^{n+s_n} = y \) and

\[ \frac{N_y^{n+s_n}}{N_{x^*}^{n+s_n}} > \frac{N_y^n}{N_{x^*}^n} \left( 1 + \kappa_n C_2 \left( \log \log \frac{n}{n} \right) \right), \]

where \( s_n \leq t_n \).

In the following, Lemma 7.1 will be used to show the desired result by contradiction, i.e., if \( L_{x^*,n}^{n+m} \) is not \( O \left( \sqrt{n \log \log n} \right) \), then alternative \( y \) specified in the statement of Lemma 7.1 cannot be sampled at stage \( n + s_n \), which contradicts the definition of \( s_n \).

Define \( t_n = \sup \left\{ l < m_n : x^{n+l} \neq x^* \right\} \) as in Lemma 7.1. If \( t_n = 0 \), then, by Lemma 4.4 we have \( L_{x^*,n}^{n+m_n} = O \left( \sqrt{n \log \log n} \right) \). Now, suppose that \( t_n > 0 \); then, in order to show that \( L_{x^*,n}^{n+m_n} = \)
\[ O(\sqrt{n \log \log n}) \text{, it is sufficient to show that } L^{n_t+n_{m}}_{x^*} = O(\sqrt{n \log \log n}) \text{ since } L^{n_t+n_{m}}_{n, n+m} = O(\sqrt{n \log \log n}) \text{ by Lemma 4.4.} \]

Denote by \( \kappa_n = \frac{L^{n_t+n_{m}}_{n, n+m}}{\sqrt{n \log \log n}} \) and let \( C_1, C_2 \) be the constants obtained from Lemma 7.1. From this lemma, we know that, if \( \kappa_n \geq C_1 \) for large enough \( n \), then we can find a suboptimal alternative \( y \neq z \) and a time stage \( n + s_n \) such that \( x^{n+s_n} = y \) and

\[
\frac{N_y^{n+s_n}}{N_{x^{n+s_n}}} \geq \frac{N_y^n}{N_{x^n}} (1 + \kappa_n C_2) \sqrt{\frac{\log \log n}{n}},
\]

where \( s_n \leq t_n \). This also leads to

\[
\frac{L^{n,n+s_n}_{x, x^*}}{L^{n,s+n}_{x^*, x}} \geq \frac{N_y^n}{N_{x^n}} (1 + \kappa_n C_2) \sqrt{\frac{\log \log n}{n}} + \frac{N_y^n}{L^{n,n+s_n}_{x^*, x}} \kappa_n C_2 \sqrt{\frac{\log \log n}{n}}.
\]

(81)

Since \( x^n = z \), we must have \( \Gamma_{x^*, z}^n \leq \Gamma_{x^*, y}^n \) by construction of BOLD, meaning that

\[
N_y^n L_{x^*, z} (u_{x^*, z}^n) + N_{x^n} L_{x^*, z} (u_{x^*, z}^n) \leq N_y^n I_{y^n} (u_{x^*, y}^n) + N_{x^n} I_{x^n} (u_{x^*, y}^n).
\]

(82)

At time stage \( n + s_n \), we have

\[
\Gamma_{x^*, z}^{n+s_n} \leq N_y^{n+s_n} L_{x^*, z} (u_{x^*, z}^n) + N_{x^n} L_{x^*, z} (u_{x^*, z}^n)
\]

(83)

\[
= N_y^n L_{x^*, z} (u_{x^*, z}^n) + N_{x^n} L_{x^*, z} (u_{x^*, z}^n) + L_{x^*, z} L_{x^*, z} (u_{x^*, z}^n) + L_{x^*, z} L_{x^*, z} (u_{x^*, z}^n) + L_{x^*, z} L_{x^*, z} (u_{x^*, z}^n)
\]

(84)

\[
\leq N_y^n I_{y^n} (u_{x^*, y}^n) + N_{x^n} I_{x^n} (u_{x^*, y}^n) + C_3 \sqrt{n \log \log n}
\]

(85)

\[
+ L_{x^*, z} L_{x^*, z} (u_{x^*, z}^n) + L_{x^*, z} L_{x^*, z} (u_{x^*, z}^n)
\]

(86)

\[
\leq N_y^n I_{y^n} (u_{x^*, y}^n) + N_{x^n} I_{x^n} (u_{x^*, y}^n) + C_3 \sqrt{n \log \log n}
\]

(87)

\[
+ L_{x^*, z} L_{x^*, z} (u_{x^*, z}^n) + L_{x^*, z} L_{x^*, z} (u_{x^*, z}^n)
\]

\[
\leq N_y^n I_{y^n} (u_{x^*, z}^n) + N_{x^n} I_{x^n} (u_{x^*, z}^n) + C_4 \sqrt{n \log \log n}
\]

(88)

where \( C_3, C_4 \) are fixed positive constants. Inequality (83) holds by definition of \( u_{x^*, z}^n \) and Corollary 4.1; inequality (84) holds by Assumption 4.4 and the law of the iterated logarithm; inequality (85) holds due to (82); inequality (86) holds by definition of \( u_{x^*, y}^n \) and Corollary 4.1; finally, inequality
(87) is again obtained from Assumption 4.4 and the law of the iterated logarithm. In order to have \( \Gamma^{n+s_n}_{x^*,z} < \Gamma^{n+s_n}_{x^*,y} \) and thus obtain the desired contradiction, it is sufficient to show

\[
C_4 \sqrt{n \log \log n} + L^{n+s_n}_z I^{n+s_n}_y (u^{n+s_n}_{x^*,z}) + L^{n+s_n}_x I^{n+s_n}_x (u^{n+s_n}_{x^*,z}) < L^{n+s_n}_y I^{n+s_n}_y (u^{n+s_n}_{x^*,y}). \tag{89}
\]

Note that \( L^{n+s_n}_z = 1 \) by construction of \( s_n \). To show (89), we need to consider two possible cases for the ordering of \( u^{n+s_n}_{x^*,z} \) and \( u^{n+s_n}_{x^*,y} \).

**Case 1:** \( u^{n+s_n}_{x^*,y} \leq u^{n+s_n}_{x^*,z} \). In this case, there must exist a fixed positive constant \( C_5 \) such that, for all large enough \( n \), if \( \kappa_n \geq \max \left\{ \frac{2C_4}{C_5} + 1, C_1 \right\} \), we have

\[
L^{n+s_n}_y I^{n+s_n}_y (u^{n+s_n}_{x^*,y}) + L^{n+s_n}_x I^{n+s_n}_x (u^{n+s_n}_{x^*,z}) \geq L^{n+s_n}_x I^{n+s_n}_x (u^{n+s_n}_{x^*,z}) + L^{n+s_n}_y I^{n+s_n}_y (u^{n+s_n}_{x^*,y}) \geq N^n y \kappa_n \sqrt{n \log \log n} + L^{n+s_n}_x I^{n+s_n}_x (u^{n+s_n}_{x^*,z}) \geq 2C_4 \sqrt{n \log \log n} \left( L^{n+s_n}_x I^{n+s_n}_y (u^{n+s_n}_{x^*,z}) \right) \geq C_4 \sqrt{n \log \log n} + L^{n+s_n}_x I^{n+s_n}_x (u^{n+s_n}_{x^*,z}), \tag{90}
\]

where (90) holds by Lemma 2.1, (91) follows from (81), (92) holds due to Theorem 4.2, Assumption 4.1 and Corollary 4.1, and finally (93) holds by Assumption 4.4 and Lemma 2.1. Thus (89) holds and the desired contradiction follows.

**Case 2:** \( u^{n+s_n}_{x^*,y} > u^{n+s_n}_{x^*,z} \). In this case, showing (89) is equivalent to showing

\[
\left| L^{n+s_n}_x I^{n+s_n}_y (u^{n+s_n}_{x^*,z}) + L^{n+s_n}_x I^{n+s_n}_x (u^{n+s_n}_{x^*,z}) - L^{n+s_n}_y I^{n+s_n}_y (u^{n+s_n}_{x^*,y}) \right| < \frac{L^{n+s_n}_y I^{n+s_n}_y (u^{n+s_n}_{x^*,y})}{L^{n+s_n}_x}. \tag{94}
\]

We first consider the LHS of (94) and observe that, when \( n \) is large enough, we have

\[
\left| L^{n+s_n}_x I^{n+s_n}_y (u^{n+s_n}_{x^*,z}) + L^{n+s_n}_x I^{n+s_n}_x (u^{n+s_n}_{x^*,z}) - L^{n+s_n}_y I^{n+s_n}_y (u^{n+s_n}_{x^*,y}) \right| \leq C_6 \sqrt{n \log \log n} + L^{n+s_n}_x I^{n+s_n}_y (u^{n+s_n}_{x^*,z}) - L^{n+s_n}_y I^{n+s_n}_y (u^{n+s_n}_{x^*,y}). \tag{95}
\]
\[ \leq \frac{C_6 \sqrt{n \log \log n}}{L_{x^*,y}} + \frac{I^n_{x^*} (u^n_{x^*,z}) - I^n_{x^*} (u^n_{x^*,y})}{I^n_y (u^n_{x^*,y})} + C_7 \sqrt{\frac{\log \log n}{n}}, \]  

(96)

where \( C_6, C_7 \) are positive constants, (95) holds by Assumption 4.1, Assumption 4.4 and Corollary 4.1, while (96) is obtained by repeating the arguments in (58). For the RHS of (94), when \( n \) is large enough we have

\[ L_{n,n} + s_n x^* \geq N_{n,n} + s_n x^* (u^n_{x^*,y}) - I^n_{x^*} (u^n_{x^*,y}) + C_7 \sqrt{\frac{\log \log n}{n}}, \]

(97)

where \( C_8, C_9 \) are fixed positive constants. Inequality (97) holds by (81), while the remaining inequalities follow by Theorem 4.2.

Combining (96) and (98), we can see that, if \( \kappa_n \geq \max \left\{ \frac{C_6}{C_8} + 1, \frac{C_7}{C_9} + 1, C_1 \right\} \), then (94) will hold if

\[ \frac{I^n_{x^*} (u^n_{x^*,z}) - I^n_{x^*} (u^n_{x^*,y})}{I^n_y (u^n_{x^*,y})} \leq \frac{N^n_y}{N^n_{x^*}}. \]

(99)

We then observe that

\[ N^n_y I^n_{x^*} (u^n_{x^*,y}) + N^n_{x^*} I^n_{x^*} (u^n_{x^*,y}) \geq N^n_y I^n_{x^*} (u^n_{x^*,y}) + N^n_{x^*} I^n_{x^*} (u^n_{x^*,y}) \]

(100)

\[ \geq N^n_y I^n_{x^*} (u^n_{x^*,z}) + N^n_{x^*} I^n_{x^*} (u^n_{x^*,z}) \]

(101)

\[ > N^n_y I^n_{x^*} (u^n_{x^*,z}), \]

(102)

where (100) holds by definition of \( u^n_{x^*,y} \) and (101) holds by (82). This shows that (99) holds, providing the desired contradiction. It follows that there exists some fixed constant \( C \) such that \( \kappa_n < C \) for all \( n \), which means that \( I^n_{x^*,y} \) is \( O(\sqrt{n \log \log n}) \), as required.

### 7.6 Proof of Lemma 4.6

For \( y = x^* \), the conclusion obviously holds by Lemma 4.5. For \( y \neq x^* \), define

\[ t_n = \sup \left\{ l < m_n : x^{n+l} = y \right\}. \]
It is sufficient to show \( L_{y}^{n,n+t_n} = O\left(\sqrt{n \log \log n} \right) \) since \( L_{y}^{n,m+n} = L_{y}^{n,n+t_n} + 1 \). Since \( x^{n+t_n} = y \), we must have \( \Gamma_{x,y}^{n+t_n} \leq \Gamma_{x,y}^{n+t_n} \), that is,

\[
N_{y}^{n+t_n} I_{y}^{n+t_n} \left( u_{x,y}^{n+t_n} \right) + N_{x,y}^{n+t_n} I_{x,y}^{n+t_n} \left( u_{x,y}^{n+t_n} \right) \leq N_{x,y}^{n+t_n} I_{y}^{n+t_n} \left( u_{x,y}^{n+t_n} \right) + N_{x,y}^{n+t_n} I_{x,y}^{n+t_n} \left( u_{x,y}^{n+t_n} \right). \tag{103}
\]

Then, for all large enough \( n \),

\[
L_{y}^{n,n+t_n} I_{y}^{n+t_n} \left( u_{x,y}^{n+t_n} \right) \leq N_{y}^{n+t_n} I_{y}^{n+t_n} \left( u_{x,y}^{n+t_n} \right) + N_{x,y}^{n+t_n} I_{x,y}^{n+t_n} \left( u_{x,y}^{n+t_n} \right) - N_{y}^{n+t_n} I_{y}^{n+t_n} \left( u_{x,y}^{n+t_n} \right) - N_{x,y}^{n+t_n} I_{x,y}^{n+t_n} \left( u_{x,y}^{n+t_n} \right) \leq N_{x,y}^{n+t_n} I_{x,y}^{n+t_n} \left( u_{x,y}^{n+t_n} \right) + N_{y}^{n+t_n} I_{y}^{n+t_n} \left( u_{x,y}^{n+t_n} \right) - N_{y}^{n+t_n} I_{y}^{n+t_n} \left( u_{x,y}^{n+t_n} \right) - N_{x,y}^{n+t_n} I_{x,y}^{n+t_n} \left( u_{x,y}^{n+t_n} \right) = L_{x,y}^{n} I_{x,y}^{n} \left( u_{x,y}^{n} \right) - L_{x,y}^{n} I_{x,y}^{n} \left( u_{x,y}^{n} \right) + N_{x,y}^{n} I_{x,y}^{n} \left( u_{x,y}^{n} \right) + N_{y}^{n} I_{y}^{n} \left( u_{x,y}^{n} \right) - N_{y}^{n} I_{y}^{n} \left( u_{x,y}^{n} \right) - N_{x,y}^{n} I_{x,y}^{n} \left( u_{x,y}^{n} \right) = C_{1} \sqrt{n \log \log n} \tag{104}
\]

\[
L_{x,y}^{n} I_{x,y}^{n} \left( u_{x,y}^{n} \right) + L_{y}^{n} I_{y}^{n} \left( u_{x,y}^{n} \right) - L_{y}^{n} I_{y}^{n} \left( u_{x,y}^{n} \right) - L_{x,y}^{n} I_{x,y}^{n} \left( u_{x,y}^{n} \right) + C_{1} \sqrt{n \log \log n} \tag{105}
\]

\[
L_{x,y}^{n} I_{x,y}^{n} \left( u_{x,y}^{n} \right) + L_{y}^{n} I_{y}^{n} \left( u_{x,y}^{n} \right) - L_{y}^{n} I_{y}^{n} \left( u_{x,y}^{n} \right) - L_{x,y}^{n} I_{x,y}^{n} \left( u_{x,y}^{n} \right) = C_{2} \sqrt{n \log \log n} + C_{1} \sqrt{n \log \log n} \tag{106}
\]

\[
= C_{3} \sqrt{n \log \log n}, \tag{107}
\]

where \( C_{1}, C_{2} \) and \( C_{3} = C_{1} + C_{2} \) are fixed positive constants. Inequality (103) holds by definition of \( u_{x,y}^{n+t_n} \); inequality (104) holds by Assumption 4.4 and the law of the iterated logarithm; inequality (105) holds because \( z \) is sampled at stage \( n \); inequality (106) holds by Assumption 4.4 and Lemma 4.5. Thus, for all large enough \( n \), we have

\[
L_{y}^{n,n+t_n} I_{y}^{n+t_n} \left( u_{x,y}^{n+t_n} \right) \leq C_{3} \sqrt{n \log \log n} \leq C_{4} \sqrt{n \log \log n},
\]

where \( C_{4} \) is a fixed positive constant and the last inequality holds by Assumption 4.1 and Corollary 4.1.

### 7.7 Proof of Lemma 7.1

Suppose that \( x^{n} = z \). At stage \( n+t_n \), a suboptimal alternative is sampled. Repeating the arguments in the proof of Theorem 4.3, we must also have

\[
\Delta_{n+t_n} \geq 1 - C_{3} \sqrt{\frac{\log \log n}{n}}, \tag{108}
\]
where $C_3$ is a fixed positive constant. Since $\Delta^n \leq 1$ by construction of the BOLD algorithm, (108) yields
\[
\Delta^{n+t_n} \geq \Delta^n - C_3 \sqrt{\frac{\log \log n}{n}},
\]
which leads to
\[
\sum_{x \neq z, x^*} \frac{I_x^{n+t_n} (u_{x^*, x}^{n+t_n})}{I_x^{n+t_n} (u_{x^*, z}^{n+t_n})} - \sum_{x \neq z, x^*} \frac{I_x^n (u_{x^*, x}^n)}{I_x^n (u_{x^*, z}^n)} \geq \frac{C_3}{n} \sqrt{\frac{\log \log n}{n}}. \tag{109}
\]

Repeating the arguments we established in (61)-(74), there must exist two fixed positive constants $C_4, C_5$ such that, if $\kappa_n \geq C_4 + 1$, then for all large enough $n$ we have
\[
\frac{I_x^{n+t_n} (u_{x^*, z}^{n+t_n})}{I_x^{n+t_n} (u_{x^*, z}^{n+t_n})} \geq C_5 (\kappa_n - C_4) \sqrt{\frac{\log \log n}{n}}.
\]

Then, (109) becomes
\[
\sum_{x \neq z, x^*} \frac{I_x^{n+t_n} (u_{x^*, x}^{n+t_n})}{I_x^{n+t_n} (u_{x^*, x}^{n+t_n})} - \sum_{x \neq z, x^*} \frac{I_x^n (u_{x^*, x}^n)}{I_x^n (u_{x^*, x}^n)} \geq (C_5 \kappa_n - C_6) \sqrt{\frac{\log \log n}{n}},
\]
where $C_6 = C_4 C_5 + C_3$ is a fixed positive constant. Therefore, there must exist some suboptimal $y \neq z$ such that
\[
\frac{I_x^{n+t_n} (u_{x^*, y}^{n+t_n})}{I_x^{n+t_n} (u_{x^*, y}^{n+t_n})} - \frac{I_y^n (u_{x^*, y}^n)}{I_y^n (u_{x^*, y}^n)} \geq \frac{C_5 \kappa_n - C_6}{M - 2} \sqrt{\frac{\log \log n}{n}}.
\]

Repeating the arguments of (74), we then have
\[
\Delta^2 \geq C_5 \kappa_n - C_6 \sqrt{\frac{\log \log n}{n}},
\]
where $C_7$ is also a fixed positive constant. Similarly as in (61)-(73), we then have
\[
\frac{N_x^{n+t_n}}{N_x^n} - \frac{N_y^{n+t_n}}{N_y^n} \geq \frac{D_x^{n+t_n} (u_{x^*, y}^{n+t_n})}{D_x^n (u_{x^*, y}^n)} + \frac{D_y^{n+t_n} (u_{x^*, y}^{n+t_n})}{D_y^n (u_{x^*, y}^n)} - \Theta (1) \left( u_{x^*, y}^{n+t_n} - u_{x^*, y}^n \right) - C_8 \sqrt{\frac{\log \log n}{n}}
\]
\[
\geq \Theta (1) \left( u_{x^*, y}^{n+t_n} - u_{x^*, y}^n \right) - C_8 \sqrt{\frac{\log \log n}{n}}
\]
\[
\geq (C_9 \kappa_n - C_{10}) \sqrt{\frac{\log \log n}{n}},
\]
where $C_8, C_9, C_{10}$ are fixed positive constants.
Define \( s_n = \sup \{ l \leq t_n : x^{n+l} = y \} \). Then, for all large enough \( n \), if \( \kappa_n \geq \max \left\{ C_4 + 1, \frac{C_{10} \ln n}{\ln \ln n} + 1 \right\} \), we have

\[
\frac{N_{y}^{n+s_n}}{N_{x^*}^{n+s_n}} - \frac{N_{y}^{n}}{N_{x^*}^{n}} \geq \frac{N_{y}^{n+t_n} - 1}{N_{x^*}^{n+t_n}} - \frac{N_{y}^{n}}{N_{x^*}^{n}} \geq (C_9 \kappa_n - C_{10}) \sqrt{\frac{\log \log n}{n}} - \frac{C_{11}}{n} \quad \text{(110)}
\]

\[
\geq \frac{C_9 \kappa_n - C_{10}}{2} \sqrt{\frac{\log \log n}{n}}, \quad \text{(111)}
\]

where \( C_{11} \) is a positive constant, and (110) follows by Theorem 4.2.

Again applying Theorem 4.2, for all large enough \( n \), there exists a positive constant \( C_{12} \) such that

\[
\frac{N_{y}^{n+s_n}}{N_{x^*}^{n+s_n}} > \frac{N_{y}^{n}}{N_{x^*}^{n}} \left( 1 + \frac{C_9 \kappa_n - C_{10}}{C_{12}} \right) \sqrt{\frac{\log \log n}{n}}.
\]

Now, we can let \( C_1 = \max \left\{ C_4 + 1, \frac{2C_{10} \ln n}{C_9} + 1 \right\} \) and \( C_2 = \frac{C_9}{2C_{12}} \). Then, for large enough \( n \), if \( \kappa_n \geq C_1 \) we have

\[
\frac{N_{y}^{n+s_n}}{N_{x^*}^{n+s_n}} > \frac{N_{y}^{n}}{N_{x^*}^{n}} \left( 1 + \kappa_n C_2 \right) \sqrt{\frac{\log \log n}{n}},
\]

which completes the proof.