OPTIMAL COMPUTING BUDGET ALLOCATION FOR
DATA-DRIVEN RANKING AND SELECTION

Yuhao Wang
School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta
yuhaowang@gatech.edu

Enlu Zhou
School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta
enlu.zhou@isye.gatech.edu

ABSTRACT

In a fixed budget ranking and selection (R&S) problem, one aims to identify the best design among a finite number of candidates by efficiently allocating the given computing budget to evaluate design performance. Classical methods for R&S usually assume the distribution of the randomness in the system is exactly known. In this paper, we consider the practical scenario where the true distribution is unknown but can be estimated from streaming input data that arrive in batches over time. We formulate the R&S problem in this dynamic setting as a multi-stage problem, where we adopt the Bayesian approach to estimate the distribution and formulate a stage-wise optimization problem to allocate the computing budget. We characterize the optimality conditions for the stage-wise problem by applying the large deviations theory to maximize the decay rate of probability of false selection. Based on the optimality conditions and combined with the updating of distribution estimates, we design two sequential budget allocation procedures for R&S under streaming input data. We theoretically guarantee the consistency and asymptotic optimality of the proposed procedures. We demonstrate the practical efficiency through numerical experiments in comparison with the equal allocation policy and an extension of the Optimal Computing Budget Allocation algorithm.

Keywords ranking and selection · large deviations theory · optimal computing budget allocation · Bayesian estimation

1 Introduction

In many applications, performance of a complex stochastic system is often evaluated through time-consuming and expensive experimentation or simulation. Comparison of multiple system designs and selection of the best one is referred to as the ranking and selection (R&S) problem. R&S dates back to the 1950s in agricultural and clinical applications ([1, 2]) and has since been used in many application problems arising from a variety of areas such as healthcare and manufacturing.

While R&S has traditionally assumed access to a fixed simulator, the recent application of Digital Twin has been calling for data-driven techniques for R&S. As indicated by the US National Academies report [3], a digital twin is “a set of virtual information constructs that mimics the structure, context, and behavior of a natural, engineered, or social system (or system of systems), is dynamically updated with data from its physical twin, has a predictive capability, and informs decisions that realize value.” The physical system continually provides the digital twin with real-time data for better representation of the real system, whereas the digital twin informs predictions and decisions (through simulation optimization such as R&S) for the real system. This repeated process poses a significant challenge to conventional R&S with a fixed simulation model due to its fast recurrence, requiring efficient data-driven R&S that accommodates dynamic updates of the simulation model.

To define data-driven R&S more specifically, consider that the random factors in the system are modeled by some probability distributions, which are usually called “input distributions” or “input models”. The input distributions are often estimated with input data observed from the real system. While in many existing works of R&S, the input distributions are only estimated once with a fixed set of input data prior to the performance evaluation, in many
application problems (such as digital twin applications), input data are often collected in batches over time. It is then beneficial to incorporate the sequentially arrived input data to get better input distributions that are more accurate estimates of the underlying distributions of the real system, and subsequently refine the performance evaluations that are used to compare potential system designs. We illustrate the problem of data-driven R&S with the following concrete examples.

1. **Selecting the best budget allocation scheme for an activity network:** Activity networks are widely used for project management, computer system security, and digital circuit design [4, 5]. An activity network is represented by a directed acyclic graph with one starting node and one ending node. The directions of arcs in an activity network represent the precedence relationships between different activities, which all have certain duration to complete. However, over the course of the project (i.e., to complete all activities), a random disruption (such as natural disasters, electrical outages, workers strikes) may occur, which makes the project completion time (PCT) stochastic. The goal is to select the best budget allocation scheme, which allocates budget to activities to reduce their completion times, in order to minimize the expected PCT. Because of the complexity of large-scale activity networks, evaluating the PCT of a given scenario is usually computationally expensive. Multiple scenarios are generated from the distribution of random disruption, and PCTs are evaluated in these scenarios to form an estimate for the expected PCT. The distribution of the random disruption can be updated when more data become available, and therefore, more evaluations of PCTs can be done under the updated distributions to refine the selection of the best budget allocation scheme.

2. **Selecting the best COVID policy:** The goal is to compare several COVID policies to select the best one in order to minimize the expected number of deaths over a certain time period. Simulation of the spread of an infectious disease, such as COVID, is often based on the so-called compartment models that are widely used in epidemiology [6]. It is essentially a system of ordinary differential equations (ODEs) or partial differential equations (PDEs) that might have random initial conditions (e.g., initial infected population size) or unknown parameters (e.g., transmission rate, recovery rate). These unknown factors can be estimated with data that arrive frequently (e.g., daily test data) through methods such as Bayesian inference [7]. Evaluation of each COVID policy requires simulating the compartment model on multiple scenarios of the unknown factors (i.e., samples from the distribution on the initial infected population size, or from the estimated distribution of the transmission rate). For a given scenario, simulation of the compartment model is equivalent to numerically solving a system of ODEs or PDEs till the end of the time period, and hence is computationally expensive. Moreover, with new data collected over time, evaluations of the policies should be updated by running more simulations under the updated distributions.

3. **Selecting the best investment strategy:** The goal is to select the best asset portfolio to optimize the trade-off between expected return and risk. Evaluating a portfolio often requires simulating a stochastic model of the asset prices. However, some parameters of the asset price model, such drift and volatility, are unknown and can be estimated from the market data such as the past asset price or interest rates. Given periodically available data, the stochastic model is updated and new simulations are run to update the evaluations of different investment strategies.

These examples above motivate us to consider a fixed budget ranking and selection (R&S) problem with streaming data where new data arrive over time in batches of possibly varying sizes. The computing budget, which is the amount of total computation that can be used for system performance evaluations, between arrivals of two successive batches of input data is usually limited and determined by external factors such as the inter-arrival time of the data and the computational cost of each performance evaluation. The goal of this paper is to identify the best allocation scheme of the limited given computing budget to evaluate candidate designs in the setting of streaming input data.

A procedure for fixed budget R&S aims to achieve a probability of correct selection (PCS) as high as possible with a given computing budget. Our proposed procedure for the streaming data set is built on the Optimal Computing Budget Allocation (OCBA) algorithm, which is one of the most widely applied and studied algorithms for fixed budget R&S with a fixed input distribution. It computes the budget allocation rule by repeatedly maximizing an approximate PCS objective in each iteration with plug-in estimators of design performances. OCBA was first proposed in [8] and was shown to converge asymptotically to the optimal allocation rule in [9]. The statistical validity of OCBA crucially relies on the stationarity of the underlying input distribution, which implies that the performance evaluations are independent and identically distributed (i.i.d.) samples even though they are generated in different iterations of the algorithm. Hence, the performance estimation error diminishes as more samples are generated over iterations, leading to convergence of the allocation policy to the optimal policy. OCBA have been extended in the past years to various problems such as subset selection ([10] and [11]), contextual R&S ([12] and [13]), multi-objective ([14]), finding simplest good designs ([15] and [16]), maximizing opportunity cost ([17]), robust R&S under input uncertainty ([18]), and many others. We refer the reader to [19] for a comprehensive tutorial on OCBA. All of these works either assume known input
distributions or consider an empirical input distribution estimated with a fixed set of input data, where the performance evaluations for the same design follow the same input distribution over iterations.

The setting of streaming data considered in this paper is more challenging than the setting of fixed input distribution in the aforementioned OCBA works. Most notably, the input distribution is no longer fixed but is updated with new input data at each time stage, and hence it breaks the i.i.d. condition of the performance evaluations for each fixed design. To address this challenge, we adopt a Bayesian approach to estimate the unknown parametrized input distribution. We begin with a finite parameter space, where each input parameter represents a simulation scenario, so that we can evaluate the system performance under a fixed input parameter (which will be referred to as “design-input pair”), and aggregate these evaluations according to the current posterior probability for each input parameter. Since the simulation is conducted under fixed input parameter, we are then able to generate i.i.d. performance evaluations (samples) under the same design-input pair. As a result, we can estimate the expected performance for a fixed design through a Bayesian average estimator, which is the weighted sum of the estimated design-input performance multiplied by its corresponding posterior probability. We derive the optimal budget allocation policy among different design-input pairs and propose two fully sequential procedures with provable performance guarantees.

Then, we extend the proposed procedures to the more general setting where the input parameter space can be continuous, by partitioning the entire parameter space into a finite set of subspace. Despite the challenge of non-stationary simulation outputs posed by continuous input parameter subspace as opposed to fixed input parameter, we show the statistical validity of the proposed algorithms holds with the same budget allocation rules given by the proposed algorithms.

We summarize the contributions of this paper as follows.

1. This paper, along with our earlier conference version [20], is the first to consider streaming input data in fixed budget R&S problems and design a data-driven approach. This paper differs from the conference version in three main ways. First, we relax the assumption of a discrete input distribution with finite support, and instead take a Bayesian approach to estimate the parametrized input distribution, which can have continuous support. Second, we develop a new procedure that is designed to solve the original reformulated problem rather than compute an approximate solution. Third, we obtain a stronger convergence result that characterizes the speed at which the allocation policies given by the procedures converge to the optimal solution.

2. We propose a new framework for conducting simulation sequentially with the presence of streaming input data, where we run the simulation for a fixed design-input pair each time and adopt a Bayesian average estimator, defined as the weighted sum of design-input sample means weighted by the corresponding posterior probability, to estimate the expected performance of each design. To calculate the rate function of this performance estimator, we apply the Glartner-Ellis theorem (see [21]). Unlike many other works derived from [22], where they directly use the same rate function, we need to recalculate it due to samples from different distributions in the performance estimator. We formulate a stage-wise rate maximization problem using the recalculated rate function and derive the corresponding optimality conditions. Compared with the optimality conditions in [22], we obtain an additional “Input Balance” condition that characterizes the allocation rule among different input parameters according to the current posterior distribution.

3. We develop two fully sequential procedures of data-driven OCBA (DD-OCBA), namely DD-OCBA-approx(-C) and DD-OCBA-balance(-C), based on different approaches to solve the optimality conditions with either finite or continuous input parameter space. We prove the statistical consistency and asymptotic optimality of both procedures. Specifically, for the setting of finite parameter space, we provide a stronger convergence result by characterizing the asymptotic convergence rate of the allocation policies.

Next, we briefly review the relevant literature with an emphasis on the relation to our work.

**Literature Review**

The research on R&S largely falls into two related yet different categories. The fixed confidence R&S procedures aim to achieve a pre-specified probability of correct selection (PCS) using the least possible amount of simulation effort, whereas the fixed budget R&S procedures typically tend to attain a PCS as high as possible with a given simulation budget. For fixed confidence, a large body of literature goes to the indifference zone (IZ) formulation. An IZ procedure guarantees selecting the best design with at least a pre-specified confidence level, given that the difference between the top-two designs is sufficiently large. Existing IZ procedures in the R&S literature include but are not limited to the KN procedure in [23], the KVP and UVP procedures in [24], and the BIZ procedure in [25]. We refer the reader to [26] for a comprehensive review of IZ formulations. In addition, the Bayesian approaches in [27, 28] and the probably approximately correct (PAC) selection in [29] has also been studied in this stream of works.
In this paper, we focused on the fixed budget R&S. As discussed in Section 1, OCBA was originally derived under a
normality assumption and an approximate PCS objective. The allocation rule of OCBA can be justified from a rigorous
perspective of the large deviations theory in [22], after which lots of works followed this large deviations formulation.
For instance, [30] designed a fully sequential budget allocation algorithm for general distributions using the optimality
conditions in [22], [31] and [32] applied the large deviations theory to constrained R&S; [12] and [33] extended the
large deviations approach to contextual R&S; [34] applied OCBA procedure to optimize the opportunity cost as opposed
to PCS; [13] computed the large deviations rate (LDR) function with respect to a worst-case performance estimator.
In this paper, we compute the LDR function of a performance estimator aggregating samples across different input
distributions. Other well-known fixed budget R&S procedures include the expected value of information (EVI) approach
proposed by [35] and the knowledge-gradient (KG) approach proposed by [36], where EVI is derived by asymptotically
minimizing a bound of the expected loss and KG determines the optimal sampling allocation policy by maximizing the
so-called acquisition function. We refer the reader to [37] for a recent overview of the R&S literature.

All the aforementioned works assume the underlying distribution is known. While there are extensive studies on the
impact of estimated input distributions on the simulation outputs (e.g., [27, 38, 40, 41]), which we refer the reader to
[42] for a recent review, R&S with estimated input distributions have only been studied in recent years. [43, 44] aimed
to eliminate as many inferior designs as possible and return a subset of superior designs with a fixed amount of input
data. [45] formulated a fixed budget problem under OCBA framework to simultaneously allocate the effort to carry
out simulation and the effort to obtain input data; this work is followed by [46], which proposed a general framework
that integrates input data collection and simulation in which the data collection and simulation costs themselves can be
random. [48] took a fixed budget formulation with a robust approach, aiming to select a design with the best
worst-case performance over an uncertainty set of finite distributions that contains the true input distribution; [49] also
used this worst-case criterion but took an IZ formulation. [50] derived confidence bands to account for both estimation
error in input distribution estimation and stochastic error in simulation output in R&S. Despite the assumption of
estimated input distributions, these works focused on a fixed set of input data.

More recently, [51] considered R&S with streaming input data, similar to the setting in this paper, but used a fixed
certainty formulation. They proposed a moving average performance estimator to aggregate simulation outputs from
different input distributions over time stages and designed sequential elimination procedures to screen out the inferior
designs until one is left to be the optimal one with at least a pre-specified confidence level. The methodology in [51] is
fundamentally different from our approach, as their focus is on deriving the valid concentration bound to reach the
given confidence level while we focus on optimizing budget allocation. In [51], each remaining design is simulated
once at every iteration, which allows the usage of common random numbers to reduce the variance of simulation output
and narrow the concentration bound. However, from the perspective of budget allocation, [51] trivially allocates the
simulation budget equally to all designs, which can be inefficient if the simulation budget itself is limited. The fixed
budget R&S with streaming data is first considered in [20] and [52], where [20] is the early conference version of this
paper as discussed in the previous section. [52] considers the setting where input data can be actively collected, which
is similar to the setting of [43] but with periodic data collection and a selecting criterion called “most probable best”.

The rest of the paper is organized as follows. We describe the problem setting and present the overall framework of the
proposed data-driven procedures in Section 2. In Section 3, we explicitly solve the stage-wise budget allocation problem
by applying the large deviations theory to calculate the rate function for the performance estimator and characterize
the stage-wise optimal allocation policy. In Section 4, we propose sequential procedures for the R&S problem under
streaming data. We show the statistical consistency and asymptotic optimality of the two procedures in Section 5. In
Section 6, we extend the framework to continuous input parameter space, while preserving the provable statistical
validity. We present numerical results in Section 7 and conclude in Section 8.

2 Problem Statement

We first give some basic notations. Suppose we have a set of finite number of designs $I = \{1, 2, \cdots, K\}$, and the goal
is to find the design with the highest expected performance. The performance of each design $i \in I$ is evaluated through
repeated simulations. The computing budget, which is the total number of replications we can run on all designs, is
often limited by computational time or expense. The core of the fixed budget R&S problem is to devise procedures that
maximize the probability of correct selection (PCS) of the optimal design when exhausting the computing budget.

In classical R&S, the input distributions, $\{F_i\}_{i \in I}$, that capture various sources of system randomness are assumed to
be known. However, in practice the true underlying distributions are seldom known and need to be estimated from input
data, which are a finite amount of real-world observations. Sources of system randomness, such as disruption delay in
the activity network, are often shared among all designs. Therefore, throughout the paper we assume that all the designs
share the same input distribution $F^c$ (thus, dropping the subscript $i$) and consequently common input data from these

distributions. Here the superscript $c$ stands for “correct”, meaning that the input distribution is exactly the same as the true distribution. Note that here the assumption of common input distribution does not rule out the existence of possibly design-specific distributions, since we can simply incorporate them into the common input distribution even though they may not affect the simulation of all designs.

We assume the underlying input distributions belong to some parametric family.

**Assumption 1**
1. The cumulative density function (cdf) of the true input distribution, $F^c = F_{\theta^c}$, belongs to a known parametric family $\{F_\theta : \theta \in \Theta\}$.
2. Furthermore, the parameter space $\Theta = \{\theta_1, \ldots, \theta_D\}$ is finite.

The parametric family in Assumption 1 enables us to estimate the input distribution $\pi$ by Assumption 2.

For each design "design-input" pair is that the simulation outputs for the fixed design-input pair are i.i.d. across different time stages, despite the simulating the $\pi$ taking the weighted average with weights $\bar{\pi}$. By (1), Bayesian average performance $\bar{\pi}$ satisfies $\bar{\pi}(\theta) = \frac{\pi^0(\theta_j) \prod_{\ell=1}^D f_{\theta_{\ell}}(\xi_{\ell})}{\sum_{j'=1}^D \pi^0(\theta_{j'}) \prod_{\ell=1}^D f_{\theta_{\ell}}(\xi_{\ell})}$, $j = 1, \ldots, D$.

By Assumption 2, $\pi_j > 0$ almost surely.

The posterior distribution provides a density estimate of the unknown parameter $\theta$, and naturally leads to an estimate for the expected performance. Specifically, let $X_i(\theta)$ denote the random performance of design $i$ under input parameter $\theta$ and $\mu_i(\theta) = \mathbb{E}[X_i(\theta) | \theta]$ the expected performance of design $i$ under $\theta$. Moreover, with finite parameter space $\Theta$, we can write $X_{ij} := X_i(\theta_j)$ and $\mu_{ij} := \mathbb{E}[X_{ij}]$. The expected performance under the unknown true input distribution, $\mu_i(\theta^c)$, can be estimated by the Bayesian average performance, which is expected performance under the posterior:

$$\bar{\mu}_i = \mathbb{E}_\pi[\mu_i(\theta)] = \sum_{j=1}^D \pi_j \mu_{ij}. \quad (1)$$

We make the following assumption on the Bayesian average performance as well as the true expected performance, to ensure the uniqueness of the (estimated and true) best design.

**Assumption 3** Let $b := \arg \max_i \bar{\mu}_i$ denote the design that maximizes the Bayesian average performance, which is unique for almost every $\pi$. Moreover, the true best design $b^c := \arg \max_i \mu_i(\theta^c)$ is also unique.

By Assumption 1, Bayesian average performance $\bar{\mu}_i$ allows aggregation of $\mu_{ij}$’s under different input distributions $F_{\theta_j}$’s by simply taking the weighted average with weights $\pi_j$’s. We will refer to simulating design $i$ under input distribution $F_{\theta_j}$ as simulating the $(i, j)$ “design-input” pair throughout the paper. A great advantage of simulating under fixed design-input pair is that the simulation outputs for the fixed design-input pair are i.i.d. across different time stages, despite the varying posterior distribution.

For each design $i$, let $X^\ell_{ij}$ denote the $\ell^{th}$ sample of $X_{ij}$. We make the following assumption of Gaussian simulation output.

**Assumption 4**
1. The simulation output $X_{ij}$ follows a normal distribution with unknown mean $\mu_{ij}$ and unknown variance $\sigma_{ij}^2$.
2. The simulation output $\{X_{ij}^{(\ell)}\}$ are independent for all $i, j$ and $\ell$. 

Data-driven R&S
Data-driven R&S

Assumption 4.1 models the simulation error as Gaussian noise, which is common in the R&S literature as one can use batched simulation output. Assumption 4.2 can be guaranteed since now we simulate on fixed distribution.

2.1 Data-driven Framework

Next, we describe the overall framework of our proposed data-driven budget allocation procedures, where the posterior distribution $\pi$ is updated periodically with streaming data. Specifically, at time stage $t$, new input data of batch size $m(t)$ can be obtained and used to update the estimate of the input distribution, and then we allocate computing budget $n(t)$ to design-input pairs according to the current estimated input distribution. We assume both $n(t)$ and $m(t)$ are given. This process is illustrated in Figure 1, where $M(t) = \sum_{\tau=1}^{t} m(\tau)$ is the total amount of input data collected up to stage $t$.

For input data, we make the following assumption on the input data to help guarantee the Bayesian consistency of the posterior distribution $\pi$.

**Assumption 5** The input data, $\{\xi_s\}_{s=1}^{\infty}$, are identically and independently distributed.

To find the budget allocation rule for each stage, we apply the large deviations theory to formulate an optimization problem under the current estimated input distribution and characterize its optimality condition to derive the stage-wise optimal budget allocation rule in Section 3. Then combining with the updating of posterior distribution, we develop two data-driven budget allocation procedures for the multi-stage setting in Section 4.

3 Rate-Optimal Budget Allocation

In this section, we formulate and solve a static optimal budget allocation problem under the current posterior distribution $\pi$. Let $n$ denote the total simulation budget and $n_{i,j}$ denote the simulation budget allocated to design $i$ under input parameter $j$ (i.e., $\theta_j$). Let $\alpha_i = (\alpha_{i,1}, \ldots, \alpha_{i,D})^T$ be the ratio of budget allocated to design $i$ and across all input parameters, $\alpha_{i,j} = \frac{n_{i,j}}{n}$. Let

$$\hat{\mu}_{i,j}(\alpha_{i,j}n) = \frac{1}{\alpha_{i,j}n} \sum_{s=1}^{\alpha_{i,j}n} X_{i,j}^{(s)}$$

and

$$\hat{\mu}_i(\alpha_i, n) = \sum_{j=1}^{D} \pi_j \hat{\mu}_{i,j}(\alpha_{i,j}n),$$

denote the estimated performance for the $(i,j)$ (design-input) pair and the estimated performance for design $i$, respectively. Ignoring the minor issue of $\alpha_{i,j}n$ not being an integer, we formulate an optimization problem from the large deviations perspective as in [22] and define the rate function

$$G_i(\alpha_b, \alpha_i) := - \lim_{n \to \infty} \frac{1}{n} \log P\left( \hat{\mu}_b(\alpha_b, n) \leq \hat{\mu}_i(\alpha_i, n) \right).$$

That is, $nG_i(\alpha_b, \alpha_i)$ is the exponential rate of the probability of the event $\{\hat{\mu}_b(\alpha_b, n) \leq \hat{\mu}_i(\alpha_i, n)\}$ going to zero as $n$ goes to infinity. Since the probability of false selection (PFS), which is defined as

$$PFS = P\left( \hat{\mu}_b(\alpha_b, n) \leq \max_{i \neq b} \hat{\mu}_i(\alpha_i, n) \right),$$

Figure 1: Illustration of budget allocation with streaming input data
can be bounded by
\[
\max_{i \neq b} P(\hat{\mu}_b(\alpha_b, n) \leq \hat{\mu}_i(\alpha_i, n)) \leq PFS \leq (K-1) \max_{i \neq b} P(\hat{\mu}_b(\alpha_b, n) \leq \hat{\mu}_i(\alpha_i, n)),
\]
we have
\[
\lim_{n \to \infty} \frac{1}{n} \log PFS = - \min_{i \neq b} G_i(\alpha_b, \alpha_i).
\] (2)
That is, \( \min_{i \neq b} G_i(\alpha_b, \alpha_i) \) is the asymptotically exponential decay rate of PFS. To maximize this decay rate of PFS, we consider the following optimization problem:
\[
\begin{align*}
\max_{\alpha_{i,j}, 1 \leq i \leq K, 1 \leq j \leq D} & \quad z \\
\text{s.t.} & \quad G_i(\alpha_b, \alpha_i) - z \geq 0 \quad i \neq b \\
& \quad \sum_{i=1}^{K} \sum_{j=1}^{D} \alpha_{i,j} = 1 \\
& \quad \alpha_{i,j} \geq 0 \quad 1 \leq i \leq K, 1 \leq j \leq D.
\end{align*}
\] (3)
Optimizing the large deviations rate of PFS is first studied in [22], where input distribution is assumed to be known. Their formulation can be seen as a special case of \( D \), which incorporates simulation samples under different input parameter in calculating the rate function. Due to this difference, we cannot directly apply their result. Instead, we take a similar approach using the Gartner-Ellis Theorem (see [21]), and the detailed calculation is shown in the next section.

### 3.1 Calculation of the Rate Function

In this section we give the explicit form of the rate function \( G_i(\alpha_b, \alpha_i) \).

For a fixed \( i \), let \( \Lambda_{i,j}(\cdot) \) denote the log-moment generating function of \( X_{i,j} \) and \( \Lambda_n(\cdot, \cdot) \) denote the log-moment generating function of \( Z_n = (\hat{\mu}_b(\alpha_b, n), \hat{\mu}_i(\alpha_i, n)) \). That is,
\[
\Lambda_n(\lambda_b, \lambda_i) = \log E \left[ e^{\lambda_b \hat{\mu}_b(\alpha_b, n) + \lambda_i \hat{\mu}_i(\alpha_i, n)} \right]
\]
\[
= \log E \left[ \exp \left( \lambda_b \sum_{j=1}^{D} \sum_{s=1}^{n\alpha_{b,j}} X_{b,j} + \lambda_i \sum_{j=1}^{D} \sum_{s=1}^{n\alpha_{i,j}} X_{i,j} \right) \right]
\]
\[
= \sum_{j=1}^{D} n\alpha_{b,j} \Lambda_{b,j} \left( \frac{\lambda_b \pi_j}{n\alpha_{b,j}} \right) + \sum_{j=1}^{D} n\alpha_{i,j} \Lambda_{i,j} \left( \frac{\lambda_i \pi_j}{n\alpha_{i,j}} \right).
\]
The third equality follows from Assumption 4.2 that all simulation outputs are independent across designs and simulation outputs are identically distributed for the same design-input pair. Then, substituting \( \lambda_b \) and \( \lambda_i \) with \( n\lambda_b \) and \( n\lambda_i \), respectively, we obtain
\[
\frac{1}{n} \Lambda_n(n\lambda_b, n\lambda_i) = \sum_{j=1}^{D} \alpha_{b,j} \Lambda_{b,j} \left( \frac{\lambda_b \pi_j}{\alpha_{b,j}} \right) + \sum_{j=1}^{D} \alpha_{i,j} \Lambda_{i,j} \left( \frac{\lambda_i \pi_j}{\alpha_{i,j}} \right).
\]
Since $X_{i,j}$ follows a normal distribution with mean $\mu_{i,j}$ and variance $\sigma_{i,j}^2$, $\Lambda_{i,j}(\lambda) = \lambda \mu_{i,j} + \frac{1}{2} \lambda^2 \sigma_{i,j}^2$. Let $I(x_b, x_i)$ be the Fenchel-Legendre transform of $\Lambda_n$. Then,

$$I(x_b, x_i) = \sup_{\lambda_b, \lambda_i} \left\{ \lambda_b x_b + \lambda_i x_i - \sum_{j=1}^{D} \alpha_{b,j} \Lambda_{b,j} \left( \frac{\lambda_b \pi_j}{\alpha_{b,j}} \right) - \sum_{j=1}^{D} \alpha_{i,j} \Lambda_{i,j} \left( \frac{\lambda_i \pi_j}{\alpha_{i,j}} \right) \right\}$$

$$= \sup_{\lambda_b} \left\{ \lambda_b x_b - \sum_{j=1}^{D} \alpha_{b,j} \Lambda_{b,j} \right\} + \sup_{\lambda_i} \left\{ \lambda_i x_i - \sum_{j=1}^{D} \alpha_{i,j} \Lambda_{i,j} \right\}$$

$$= \lambda_b x_b - \sum_{j=1}^{D} \alpha_{b,j} \Lambda_{b,j} + \lambda_i x_i - \sum_{j=1}^{D} \alpha_{i,j} \Lambda_{i,j} + \frac{1}{2} \sum_{j=1}^{D} \sigma_{i,j}^2 \left( \frac{\lambda_i \pi_j}{\alpha_{i,j}} \right)^2 + \frac{1}{2} \sum_{j=1}^{D} \sigma_{j,i}^2 \left( \frac{\lambda_b \pi_j}{\alpha_{b,j}} \right)^2.$$

By the Gartner-Ellis Theorem, $G_i(\alpha_b, \alpha_i) = \inf_{x_b \leq x_i} I(x_b, x_i)$. It is easy to see that $I_b$ is decreasing for $x_b \leq \bar{\mu}_b$ and increasing for $x_b \geq \bar{\mu}_b$, and $I_i$ is decreasing for $x_i \leq \bar{\mu}_i$ and increasing for $x_i \geq \bar{\mu}_i$. Since $\bar{\mu}_b > \bar{\mu}_i$, we must have

$$G_i(\alpha_b, \alpha_i) = \inf_{\bar{\mu}_i \leq x_i \leq \bar{\mu}_b} I(x, x) = \frac{(\bar{\mu}_b - \bar{\mu}_i)^2}{2 \left( \sum_{j=1}^{D} \sigma_{i,j}^2 + \sum_{j=1}^{D} \frac{\sigma_{j,i}^2 \pi_j^2}{\alpha_{i,j}} \right)}.$$  

When $D = 1$, we recover exactly the same rate function as in [22]. The following lemma summarizes some important properties of $G_i(\alpha_b, \alpha_i)$.

**Lemma 1** Suppose Assumption [2] holds. Then,

1. $G_i(\alpha_b, \alpha_i)$ is strictly increasing in $\alpha_{b,j}$ and $\alpha_{i,j}$ for $\alpha_{b,j}, \alpha_{i,j} > 0$, $j = 1, 2, \ldots, D$. Moreover, $G_i(\alpha_b, \alpha_i) = 0$ if there exists $j_0$ such that $\min(\alpha_{b,j_0}, \alpha_{i,j_0}) = 0$.
2. $G_i(\alpha_b, \alpha_i)$ is concave in $(\alpha_b, \alpha_i)$ for $(\alpha_b, \alpha_i) > 0$.

Lemma [1] implies that any design-input pair must be allocated with a positive ratio of the simulation budget; otherwise, the rate will be zero. Lemma [2] claims the concavity of $G_i$, which guarantees the optimality with the Karush–Kuhn–Tucker (KKT) condition for the optimization problem [3] in the following section.

### 3.2 Optimal Allocation Policy

In this section we derive the optimality conditions for problem [3], shown in the following theorem.

**Theorem 1** Suppose Assumption [2] holds. Let $\alpha \geq 0$ be a feasible allocation policy. Then $\alpha$ is the optimal solution to [3] if and only if the following three conditions hold:

1. (Input Balance) $\frac{\partial G_i(\alpha_b, \alpha_i)}{\partial \alpha_{i,j}} = \frac{\partial G_i(\alpha_b, \alpha_i)}{\partial \alpha_{i,j'}}$ if $i \neq b$ and $1 \leq j < j' \leq D$;  

2. (Total Balance) $\sum_{i \neq b} \frac{\partial G_i(\alpha_b, \alpha_i)}{\partial \alpha_{i,j}}/\partial \alpha_{b,j} = 1$ $1 \leq j \leq D$;  

3. (Local Balance) $G_i(\alpha_b, \alpha_i) = G_{i'}(\alpha_b, \alpha_{i'})$ $i \neq i' \neq b$.

Or equivalently in the explicit form:

1. $\frac{\alpha_{i,j}}{\sigma_{i,j}^2} = \frac{\alpha_{i,j'}}{\sigma_{i,j'}^2}$ $i \neq b$, $1 \leq j \leq D$;  

2. $\left( \frac{\alpha_{b,j}}{\sigma_{b,j}} \right)^2 = \sum_{i \neq b} \left( \frac{\alpha_{i,j}}{\sigma_{i,j}} \right)^2$ $1 \leq j \leq D$;  

3. $\frac{(\bar{\mu}_b - \bar{\mu}_i)^2}{\sum_{j=1}^{D} \frac{\sigma_{i,j}^2 \pi_j^2}{\alpha_{i,j}}} = \frac{(\bar{\mu}_b - \bar{\mu}_{i'})^2}{\sum_{j=1}^{D} \frac{\sigma_{i',j}^2 \pi_j^2}{\alpha_{i',j}}}$ $i \neq i' \neq b$.  

8
Furthermore, the optimal solution $\alpha^*$ to (3) is unique.

Remark: Compared with the optimality condition in [22], in addition to the “total balance” condition that characterizes the relation between the optimal design and the non-optimal designs and the “local balance” conditions that characterize the relation between two non-optimal designs. Here we have the additional optimality condition [2], the “input balance” condition. It states that within the allocation for a certain design $i$, the partial derivative of the rate function $G_i$ with respect to $\alpha_{i,j}$ is the same for all $j$’s. That is, simulation for each fixed input parameter should provide the same improvement to identify that design $b$ is better than $i$. Furthermore, with normally distributed simulation errors, equation (3) indicates that for a fixed design $i$ the optimal allocation ratio $\alpha_{i,j}$ should be proportional to the posterior probability mass $\pi_j$ and the standard deviation $\sigma_{i,j}$, which quantitatively characterizes how input uncertainty affects the optimal allocation policy.

Also notice for fixed $i$, (8) only depends on $i$, which means the relative allocation ratios among different input parameters for a fixed design do not depend on other designs. On the other hand, (10) indicates that the relative allocation ratios among designs under the same input parameter $j$ are affected by all $\pi_j$’s, which implies directly applying OCBA to designs under a fixed input parameter $j$ may perform poorly since it does not take information from other design-input pairs into consideration. Moreover, notice that the three optimality conditions (5)-(7) not only hold for Gaussian simulation noise but also hold as long as the rate function $G_i$ has the properties shown in Lemma 1.

4 Sequential Procedure with Streaming Input Data

In deriving Theorem 1 above, we assume a fixed posterior distribution and full knowledge of simulation output distribution. In this section, by trying to satisfy the optimality conditions in Theorem 1 with the current posterior distribution, we develop two data-driven optimal budget computing budget allocation (DD-OCBA) procedures, namely DD-OCBA-approx and DD-OCBA-balance, for simulation budget allocation in the multi-stage setting with streaming input data. The two procedures mainly differ in how to satisfy the optimality conditions: DD-OCBA-approx solves the optimality conditions approximately, while DD-OCBA-balance tries to balance the two sides of the optimality equations. In the implementation phase, all unknown parameters, such as the posterior probability $\pi_j$, the Bayesian average performance $\mu_i$, and the design-input simulation variance $\sigma_{i,j}$, will be replaced by their respective estimators.

A major difficulty of solving the optimality conditions is that the optimality equations (5)-(10) do not have closed-form solutions, and it is usually computationally expensive to solve them using numerical methods such as gradient descent. To improve computational efficiency, we design the two procedures tackling the optimality conditions in different ways. The DD-OCBA-approx procedure tries to directly solve the optimization problem (3) at each iteration but approximating (10) by assuming that a weighted ratio of allocation budget assigned to the optimal design is much larger than that assigned to other designs, which is a similar assumption taken by [8]. This approximation enables us to compute the solution in a much simpler way. Alternatively, taking a similar approach in [30], the DD-OCBA-balance procedure avoids directly solving the optimality equations and instead balances the two sides of the equations, i.e., reduces the difference between two sides of the equations in each iteration when allocating the budget. Plausibly, DD-OCBA-approx is expected to converge faster since we solve the equations every time, while DD-OCBA-balance may converge slower since we only balance instead of solving the equations. However, DD-OCBA-approx approximates the optimality conditions, meaning that the “optimal solution” we get may not be really optimal in the original problem. DD-OCBA-balance, instead, targets at the original problem and will eventually converge to the true optimal solution as more and more data are collected. The empirical comparison of these two methods will be carried out numerically in Section 7.

4.1 DD-OCBA-approx

In this section we derive the DD-OCBA-approx procedure. Let $\beta_i = \frac{\alpha_{i,j}}{\sigma_{i,j} \pi_j}, 1 \leq i \leq K$, which is independent of $j$ by [8] for optimal $\alpha$. Plugging $\alpha_{i,j} = \beta_i \sigma_{i,j} \pi_j$ into (10), we have

$$\frac{(\bar{\mu}_b - \bar{\mu}_i)^2}{\sum_{j=1}^{D} \frac{\sigma_{i,j} \pi_j}{\beta_i} + \sum_{j'=1}^{D} \sigma_{i,j'} \pi_{j'}} = \frac{(\bar{\mu}_b - \bar{\mu}_{i'})^2}{\sum_{j=1}^{D} \frac{\sigma_{i,j} \pi_j}{\beta_i} + \sum_{j'=1}^{D} \sigma_{i,j'} \pi_{j'}},$$

where $i \neq i' \neq b$.

Assume $\beta_b \gg \beta_i, \forall i \neq b$, i.e., $\frac{\alpha_{i,j}}{\sigma_{i,j}} \gg \frac{\alpha_{b,j}}{\sigma_{b,j}}$, $\forall i \neq b, \forall j$, the simulation budget assigned to the optimal design-input pair divided by its standard deviation is much larger than that of other designs. Then we have
The procedure selects a design-input pair each time to reduce the difference (balance) of at least one of the optimality conditions. If the estimated best design is not selected, unlike DD-OCBA-approx, where we try to directly solve for the optimal solutions, DD-OCBA-balance only requires to influence the Bayesian average estimator.

Furthermore, with (9) and \( \sum_{i=1}^{K} \sum_{j=1}^{K} \alpha_{i,j} = 1 \), we can calculate \( \alpha_{i,j} \) explicitly. On a related note, when input distributions are assumed to be known, i.e., \( D = 1 \) and \( \theta_1 = \theta_0 \), (11) simplifies to:

\[
\frac{\alpha_{i,j}}{\alpha_{i',j'}} = \frac{\pi_j \sigma^2_{i,j} \sum_{k=1}^{D} \sigma_{i,k} \pi_k / (\bar{\mu}_b - \bar{\mu})^2}{\pi_{j'} \sigma^2_{i',j'} \sum_{k=1}^{D} \sigma_{i',k} \pi_k / (\bar{\mu}_b - \bar{\mu}')^2}, \quad i, i' \neq b.
\]

indicating that the allocation rule for a sub-optimal design is proportional to the ratio of its simulation variance to the square of its performance difference from the optimal design. This allocation principle aligns with the OCBA procedure described by [8]. In contrast, when input distributions are unknown, the allocation rule for a sub-optimal design-input pair depends on not only the simulation variance under the same input parameter but also the variances under different input parameters. This is because the simulation outcomes under the same design but varying input parameters jointly influence the Bayesian average estimator.

**DD-OCBA-approx**

1. **Input.** Number of designs \( K \), input parameter space \( \Theta = \{\theta_1, \ldots, \theta_D\} \), initial sample size \( n_0 \), total simulation budget \( n \), input data batch size \( \{n(t)\}_{t=1}^{\infty} \) and stage-wise simulation budget \( \{n(t)\}_{t=1}^{\infty} \), prior distribution \( \pi^0 \).
2. **Initialization.** Time stage counter \( t \leftarrow 0 \), iteration counter \( \ell \leftarrow 0 \), total input data \( M(t) \leftarrow 0 \), posterior distribution \( \pi \leftarrow \pi^0 \). Collect \( n_0 \) initial samples for each design-input pair \( (i,j) \). Set \( N^{(\ell)}_{i,j} = n_0 \). Compute the initial sample mean \( \hat{\mu}^{(\ell)}_{i,j} = \frac{1}{N^{(\ell)}_{i,j}} \sum_{s=1}^{N^{(\ell)}_{i,j}} X_{i,j}^{(s)} \) and sample standard deviation \( \hat{\sigma}^{(\ell)}_{i,j} = \sqrt{\frac{1}{N^{(\ell)}_{i,j}-1} \sum_{s=1}^{N^{(\ell)}_{i,j}} (X_{i,j}^{(s)} - \hat{\mu}^{(\ell)}_{i,j})^2} \).
3. **WHILE** \( \sum_{i=1}^{K} \sum_{j=1}^{D} N^{(\ell)}_{i,j} < n \) **DO**
4. Given input data of batch size \( m(t) \), let \( M(t) = \sum_{\tau=1}^{t} m(\tau) \) and update posterior distribution \( \pi \).
5. Compute \( \hat{\mu}^{(\ell)}_{i} = \sum_{j=1}^{D} \pi_j \hat{\mu}^{(\ell)}_{i,j} \).
6. **REPEAT** \( n(t) \) **TILES**
7. \( \hat{\theta}^{(\ell)} \leftarrow \arg\max_{\theta} \hat{\mu}^{(\ell)}_{i,j} \).
8. Update \( \hat{\alpha}^{(\ell)}_{i,j} \) using (11), \( \theta \) and \( \hat{\sigma}^{(\ell)}_{i,j} = 1 \), with \( \hat{\mu}_i, \sigma_{i,j} \) replaced by \( \hat{\mu}^{(\ell)}_{i} \) and \( \hat{\sigma}^{(\ell)}_{i,j} \), respectively. Calculate \( \bar{N}^{(\ell)}_{i,j} = \hat{\alpha}^{(\ell)}_{i,j} \left( 1 + \sum_{i=1}^{K} \sum_{j=1}^{D} N^{(\ell)}_{i,j} \right), \quad \forall 1 \leq i \leq K, 1 \leq j \leq D \).
9. Find the design-input pair \( (I, J) = \arg\max_{i,j} \left( \bar{N}^{(\ell)}_{i,j} \right) \). Simulate the pair \( (I, J) \) once. Update \( \hat{\mu}_{i,j}^{(\ell+1)} \), \( \hat{\sigma}_{i,j}^{(\ell+1)} \) and \( \hat{\mu}^{(\ell+1)}_{i} \) using the new simulation output, and set \( \hat{\mu}_{i,j}^{(\ell+1)} = \hat{\mu}_{i,j}^{(\ell)} \), \( \hat{\sigma}_{i,j}^{(\ell+1)} = \hat{\sigma}_{i,j}^{(\ell)} \) and \( \hat{\mu}_{i}^{(\ell+1)} = \hat{\mu}_{i}^{(\ell)} \) for \( i \neq I, j \neq J \). Let \( \bar{N}^{(\ell+1)}_{i,j} = N^{(\ell)}_{i,j} + 1 \) and \( \bar{N}^{(\ell+1)}_{i,j} = N^{(\ell)}_{i,j} \) for all \( i \neq I, j \neq J \).
10. \( t \leftarrow t + 1 \). \( l \leftarrow l + 1 \).
11. **END REPEAT**
12. **END WHILE**
13. **Output:** Output \( i_b = \arg\max_i \hat{\mu}^{(\ell)}_i \) as the best design.

**4.2 DD-OCBA-balance**

Unlike DD-OCBA-approx where we try to directly solve for the optimal solutions, DD-OCBA-balance only requires to evaluate both sides of the three optimality equations given the current number of replications for each design-input pair. The procedure selects a design-input pair each time to reduce the difference (balance) of at least one of the optimality equations. In particular, at each iteration, the procedure will first decide whether to simulate the estimated best design or one of the non-optimal designs to balance the “total balance” conditions. If the estimated best design is not selected,
then the procedure selects a non-optimal design to balance the “total balance” conditions. After selecting the design, an input realization is chosen by balancing the “input balance” conditions. Notice that although in (8) the “input balance” conditions are only for non-optimal designs, (8) also holds for \( i = b \) by (9) with \( \frac{\alpha_{i,j}}{\sigma_{i,j}} \pi_{i,j} \), \( \forall j \neq j' \).

The balancing approach utilizes the monotonicity of both sides of all optimality equations in terms of the allocation policy \( \alpha_{i,j} \). For example, if we have in one of the equations in (8) violated by \( \frac{\alpha_{i,j}}{\sigma_{i,j}} \pi_{i,j} \), then we may want to simulate the design-input pair \( (i, j) \) to make the left hand side larger. The DD-OCBA-balance procedure is presented as follows:

**DD-OCBA-balance**

1. **Input.** Number of designs \( K \), input parameter space \( \Theta = \{ \theta_1, \ldots, \theta_D \} \), initial sample size \( n_0 \), total simulation budget \( n \), input data batch size \( \{ m(t) \}_{t=1}^{\infty} \), and stage-wise simulation budget \( \{ n(t) \}_{t=1}^{\infty} \), prior distribution \( \pi^0 \).

2. **Initialization.** Time stage counter \( t \leftarrow 0 \), iteration counter \( \ell \leftarrow 0 \), total input data \( M(t) \leftarrow 0 \), posterior distribution \( \pi \leftarrow \pi^0 \). Collect \( n_0 \) initial samples for each design-input pair \( (i, j) \). Set \( N_{i,j}^{(\ell)} = n_0 \). Compute the initial sample mean \( \hat{\mu}_{i,j}^{(\ell)} = \frac{1}{N_{i,j}^{(\ell)}} \sum_{s=1}^{N_{i,j}^{(\ell)}} X_i(s) \) and sample standard deviation \( \hat{\sigma}_{i,j}^{(\ell)} = \sqrt{\frac{1}{N_{i,j}^{(\ell)}-1} \sum_{s=1}^{N_{i,j}^{(\ell)}} (X_i(s) - \hat{\mu}_{i,j}^{(\ell)})^2} \).

3. **WHILE** \( \sum_{i=1}^{K} \sum_{j=1}^{D} N_{i,j}^{(\ell)} < n \) **DO**

4. Given input data of batch size \( m(t) \), let \( M(t) = \sum_{\tau=1}^{t} m(\tau) \) and update posterior distribution \( \pi \).

5. Compute \( \hat{\mu}_{i}^{(\ell)} = \sum_{j=1}^{D} \pi_{i,j} \hat{\mu}_{i,j}^{(\ell)} \).

6. **REPEAT** \( n(t) \) **TIMES**

7. \( \hat{b}^{(\ell)} \leftarrow \arg\max_i \hat{\mu}_{i}^{(\ell)} \).

8. Let \( j^* \leftarrow \arg\max_j \left( \frac{(N_{i,j}^{(\ell)} \hat{\mu}_{i,j}^{(\ell)})^2}{\sigma_{i,j}^{(\ell)}} - \sum_{i \neq b^{(\ell)}} \left( \frac{N_{i,j}^{(\ell)} \hat{\mu}_{i,j}^{(\ell)})^2}{\sigma_{i,j}^{(\ell)}} \right) \right) \).

9. **IF** \( \frac{(N_{i,j}^{(\ell)} \hat{\mu}_{i,j}^{(\ell)})^2}{\sigma_{i,j}^{(\ell)}} - \sum_{i \neq b^{(\ell)}} \left( \frac{N_{i,j}^{(\ell)} \hat{\mu}_{i,j}^{(\ell)})^2}{\sigma_{i,j}^{(\ell)}} \right) \) \( < 0 \), set \( I = \hat{b}^{(\ell)}, J = \arg\min_j \frac{N_{i,j}^{(\ell)} \hat{\mu}_{i,j}^{(\ell)}}{\sigma_{i,j}^{(\ell)}} \).

10. **ELSE** set \( I = \arg\min_{i \neq b^{(\ell)}} \frac{\sum_{j=1}^{D} (N_{i,j}^{(\ell)} \hat{\mu}_{i,j}^{(\ell)})^2}{N_{i,j}^{(\ell)} \sigma_{i,j}^{(\ell)}}, J = \arg\min_j \frac{N_{i,j}^{(\ell)} \hat{\mu}_{i,j}^{(\ell)}}{\sigma_{i,j}^{(\ell)}} \).

11. **END IF**

12. Simulate the pair \( (I, J) \) once. Update \( \hat{\mu}_{I,I,J}^{(l+1)}, \hat{\sigma}_{I,I,J}^{(l+1)} \) and \( \hat{\mu}_{I,J}^{(l+1)} \) using the new simulation output, and set \( \hat{\mu}_{I,I,J}^{(l+1)} = \hat{\mu}_{I,J}^{(l)} \hat{\sigma}_{I,I,J}^{(l+1)} = \hat{\sigma}_{I,J}^{(l+1)} \) and \( \hat{\mu}_{I,J}^{(l+1)} = \hat{\mu}_{I,J}^{(l)} \) for \( i \neq I, j \neq J \). Let \( N_{I,I,J}^{(l+1)} = N_{I,I,J}^{(l)} + 1 \) and \( N_{I,J}^{(l+1)} = N_{I,J}^{(l)} \) for all \( i \neq I, j \neq J \).

13. \( l \leftarrow l+1, t \leftarrow t+1 \).

14. **END REPEAT**

15. **END WHILE**

### 5 Consistency and Asymptotic Optimality

#### 5.1 Consistency

A R&S algorithm is consistent if it selects the true optimal design \( b^* \) (as defined in Assumption 3) as time stage \( t \) goes to infinity almost surely. This happens if the estimated Bayesian average performance \( \hat{\mu}_i^{(\ell)} \) converges to its true value \( \mu_i(\theta^*) \) for each \( i \) almost surely. Let \( \pi^t = (\pi^t_{j=1}D) \) denote the posterior distribution at stage \( t \). In order to show the consistency, we need to (i) show the convergence of posterior distribution \( \pi^t \) that is estimated from input observations and (ii) the convergence of \( \hat{\mu}_{i,j}^{(\ell)} \), where \( j^* \) is the index such that \( \theta_{j^*} = \theta^* \).

We first make the following assumption on the identifiability on the input parameters.
Assumption 6 (Identifiability) For \( \theta \neq \theta' \in \Theta \), \( F_{\theta} \neq F_{\theta'} \).

Together with Assumption 6, Assumption 6 guarantees the Bayesian consistency of the posterior distribution \( \pi \) by Doob’s consistency theorem (see, e.g., Theorem 10.10 in [53]). The Bayesian consistency states if the number of i.i.d. input data goes to infinity, then the posterior distribution \( \pi \) converges to a Dirac measure centered at \( \theta^c \).

Moreover, we make the following assumption about the input data batch size and simulation budget in each stage for the DD-OCBA-approx and DD-OCBA-balance procedures.

Assumption 7 The stage-wise input data batch size \( m(t) \) and simulation budget \( n(t) \) satisfy

\[
\lim_{T \to \infty} \sum_{t=1}^{T} n(t) = \infty, \quad \lim_{T \to \infty} \sum_{t=1}^{T} m(t) = \infty.
\]

Assumption 7 ensures that both the total amount of input data and total simulation replications go to infinity as time stage \( T \) goes to infinity, which helps guarantee the consistency of the posterior distribution \( \pi^t \) as well as the design-input performance estimate \( \hat{\mu}_{i,j}(\theta^c) \) under the true parameter \( \theta^c \). Recall that \( \ell \) is the iteration counter (i.e. the amount of simulation budget that has been assigned from stage 1 by the algorithm). The following theorem shows the consistency of DD-OCBA-approx and DD-OCBA-balance.

Theorem 2 (Consistency) Suppose Assumptions 4.2 and Assumption 6 hold. Then,

1. DD-OCBA-approx selects the optimal design \( b^c \) almost surely as \( \ell \to \infty \).

2. DD-OCBA-balance selects the optimal design \( b^c \) almost surely as \( \ell \to \infty \).

5.2 Asymptotic Optimality

In the last section, we establish the consistency of the two proposed algorithms, ensuring they select the optimal design as the simulation budget approaches infinity. However, it remains unclear if the algorithm’s allocation policy will converge to the optimal one. Thus, we examine the algorithms’ asymptotic optimality in this section. An algorithm is said to be asymptotically optimal if the allocation policy given by the algorithm converges to the limiting optimal policy as the total simulation budget goes to infinity. Here the allocation policy is defined as \( \{\alpha_{i,j}^{(\ell)} = \frac{N_{i,j}^{(\ell)}}{N_t} \} \), where \( N_{i,j}^{(\ell)} \) is the simulation budget assigned to \( (i, j) \) pair up to iteration \( \ell \) and \( N_t^{(\ell)} \) is the total simulation budget used up to iteration \( \ell \). Notably, the optimal policy is computed by maximizing the decay rate in [4] given a current posterior distribution \( \pi \), which is updated periodically and converges to a Dirac measure, i.e., \( \pi_j \to 1 \). Hence, the limiting optimal policy \( \alpha^* \) is defined to be the limit of that optimal policy that maximizes \( 2 \) as \( \pi_j \to 1 \).

Limiting optimal policy.

From [4], we have

\[
G_i(\alpha_b, \alpha_c) = \frac{(\bar{\mu}_b - \bar{\mu}_c)^2}{2 \left( \sum_{j=1}^{D} \frac{\sigma^2_{b,j} \pi^2_j}{\alpha_{b,j}} + \sum_{j=1}^{D} \frac{\sigma^2_{c,j} \pi^2_j}{\alpha_{c,j}} \right)} \to \frac{(\mu_{b^c} - \mu_{c}^c)^2}{2 \left( \frac{\sigma^2_{b^c,j} \pi_j}{\alpha_{b^c,j}} + \frac{\sigma^2_{c,j} \pi_j}{\alpha_{c,j}} \right)} \quad \text{as} \pi_j \to 1,
\]

where \( \mu_i^c := \mu_i(\theta^c), i \in I \) denote the true expected performance under \( \theta^c \), \( b \) is the design that maximizes the Bayesian average performance and \( b^c \) is the design that maximizes the true expected performance. (12) indicates the limiting rate function only depends on allocation ratio \( \{\alpha_{i,j}\}_{i \in I} \) as the Bayesian average performance converges to the true expected performance, which only depends on the design-input pair under the true input parameter \( \theta^c \). Moreover, the limiting rate function (12) coincides with the classic R&S with known input distribution (e.g., in [8] [22]), since the uncertainty of the input distribution decreases to 0 in the limit. Then, the limiting optimal policy can be easily computed as a special case \( (D = 1) \) of Theorem 1 which we formally states in the following theorem that guarantees the asymptotic optimality of DD-OCBA-approx.

Theorem 3 (Asymptotic Optimality for DD–OCBA-approx) Suppose Assumptions 4.2 and Assumption 6 hold. Then, almost surely, for DD-OCBA-approx,

1. \( \alpha_{i,j}^{(\ell)} \to 0, \forall i \in I, j \neq j^c \) as \( \ell \to \infty \),
As a result, for \( j \)

When all of these parameters varies only in a small neighborhood of their limiting value (consistency), the computed

Assumption 8.1 imposes a stronger assumption on the likelihood function, whereas Assumption 8.2 imposes an

The following lemma states the converge rate of the posterior probability with Assumption 8.

There are several technical reasons why we need the extra Assumption 8 to prove the asymptotic optimality for

Theorem 3 guarantees the asymptotic optimality for DD-OCBA-approx. For DD-OCBA-balance, we introduce the

Assumption 8

1. The strong law of large number (SLLN) holds for the log likelihood ratio sequence \( \left\{ \log \frac{f_{\theta}(\xi_k)}{f_{\theta^c}(\xi_k)} \right\}_{k=1}^{\infty}, \forall \theta \neq \theta^c \in \Theta, \) where \( \xi_k \sim F_{\theta^c}, k = 1, 2, \ldots \) are i.i.d. sequence. That is, almost surely for every sequence \( \left\{ \xi_k \right\}_{k=1}^{\infty} \),

\[
\lim_{m \to \infty} \frac{1}{m} \sum_{k=1}^{m} \log \frac{f_{\theta}(\xi_k)}{f_{\theta^c}(\xi_k)} = E_{\theta^c} \left[ \log \frac{f_{\theta}(\xi)}{f_{\theta^c}(\xi)} \right] = -KL(\theta^c||\theta) < 0,
\]

where \( KL(\cdot||\cdot) \) is the Kullback–Leibler divergence (KL) divergence.

2. There exist \( \bar{n}, \tilde{n} > 0 \), such that \( \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} n(t) = \bar{n}, \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} m(t) = \tilde{n} \) almost surely.

Assumption [8]1 imposes a stronger assumption on the likelihood function, whereas Assumption[8]2 imposes an additional constraint on the data batch size, which ensures that the total input data and simulation replications increase at the same rate \( O(T) \) as time stage \( T \) increases to infinity. The purpose of introducing Assumption[8]is to guarantee a certain (almost sure) convergence rate for both the posterior distribution and the estimators of design-input performance and variance, in addition to consistency of the posterior distribution and the estimators.

There are several technical reasons why we need the extra Assumption 8 to prove the asymptotic optimality for DD-OCBA-balance. We provide the intuition for the most important reason in the following. In DD-OCBA-approx, the stage-wise optimal allocation policy can be computed explicitly as a smooth and bounded function of parameters including the current posterior distribution, design-input performance estimator and design-input variance estimator. When all of these parameters vary only in a small neighborhood of their limiting value (consistency), the computed stage-wise optimal policy also varies little around the limiting optimal policy. This is a key property in the proof of Theorem[3]. Nonetheless, in DD-OCBA-balance, we do not have an explicit-form solution to guide the allocation policy, but instead only evaluate the two sides of the optimality equations in Theorem[1]. Unlike DD-OCBA-approx where the stage-wise optimal policy is a smooth function of the aforementioned parameters (posterior, mean and variance), values of two sides of Input Balance condition is not smooth or bounded in terms of the posterior probability \( \pi_j \) for \( j \neq j^c \). To be specific, as \( \pi_j^T \to 0 \), the limit of \( \frac{\alpha^{(T)}_{i,j}}{\sigma_{i,j}^{2} \pi_j} \) can be either \( 0, \infty \) or a finite value, depending on how fast of both \( \alpha^{(T)}_{i,j} \) and \( \pi_j^T \) converge to zero. Moreover, the dependence on the past allocation policy \( \alpha^{(t)}_{i,j} \) and the complex form of the Local Balance condition in (11) also extremely complicates the proof. Hence, we require some stronger assumptions to guarantee the same convergence rate of the different estimators.

The following lemma states the converge rate of the posterior probability with Assumption 8.

Lemma 2 Suppose Assumptions [2] [5] [8] hold. Then, there exists \( \kappa > 0 \), for almost every sequence of input data \( \left\{ \xi_s \right\}_{s=1}^{\infty} \),

\[
\pi_j^+ = 1 - O \left( \exp \left( -\kappa t \right) \right).
\]

As a result, for \( j \neq j^c \), \( 0 < \pi_j^+ \leq 1 - \pi_j^T = O \left( \exp \left( -\kappa t \right) \right) \).

Lemma[2] implies that the posterior probability converges exponentially fast almost surely, which can help guarantee that any design-input pair under input parameter \( \theta_j \neq \theta^c \) will only be simulated at most finitely many times.

We are now ready to provide the asymptotic optimality for DD-OCBA-balance. In addition, we can further characterize the convergence rate of the allocation policy to the limiting optimal allocation policy, for both DD-OCBA-approx and DD-OCBA-balance..
Theorem 4 Suppose Assumptions 1-3, 4.2, 5 and 8 hold. Then,

1. For DD-OCBA-approx, \( |\alpha_{i,j}^{(f)} - \alpha_{i,j}^{*}| = O \left( \frac{\log \log \ell}{\ell} \right), \quad \forall 1 \leq i \leq K \) almost surely, where \( \alpha_{i,j}^{*} \) satisfies the optimality conditions (13) and (14).

2. For DD-OCBA-balance,

\[
\begin{align*}
&\frac{(\alpha_{i,b,c}^{(f)} - \mu_{i,c}^{*})^2}{\alpha_{i,b,c}^{(f)} + \alpha_{i,j,c}^{(f)}} - \frac{(\mu_{i,b}^{*} - \mu_{i,c}^{*})^2}{\alpha_{i,b,c}^{(f)} + \alpha_{i,j,c}^{(f)}} = O \left( \frac{\log \log \ell}{\ell} \right) \quad \text{almost surely,} \quad (15) \\
&\frac{(\mu_{i,b}^{*} - \mu_{i,c}^{*})^2}{\alpha_{i,b,c}^{(f)} + \alpha_{i,j,c}^{(f)}} - \frac{(\mu_{i,b}^{*} - \mu_{i,c}^{*})^2}{\alpha_{i,b,c}^{(f)} + \alpha_{i,j,c}^{(f)}} = O \left( \frac{\log \log \ell}{\ell} \right) \quad \forall i \neq i' \neq b^c \quad \text{almost surely.} \quad (16)
\end{align*}
\]

6 Extension to Continuous Parameter Space

So far we have assumed that the input parameter space \( \Theta \) is a finite set, which can be regarded as a discretization of the original parameter space. Nonetheless, this may lead to the issue of model mis-specification, which means the true input parameter \( \theta^* \) may not be within the discretized finite set. Although the posterior distribution \( \pi \) still converges to a Dirac delta measure concentrated on \( \theta^* \), which minimizes the Kullback-Leibler divergence from the true distribution \( \pi^* \) to the set of distributions \( \{ \pi_{j,j} \mid j = 1, \ldots, D \} \) (e.g., see \( [54] \)), the best design under \( \pi^* \) may not be the true optimal design under \( \pi^* \) due to the discrepancy between \( \theta^* \) and \( \theta_{j^*} \). To address the issue of model mis-specification, in this section we extend the proposed methods to the general continuous input parameter space by generalizing the discretization approach.

To be specific, let \( \Theta_1, \Theta_2, \ldots, \Theta_D \) be a sub-space of \( \Theta \) (entire parameter space) such that \( \Theta_j \cap \Theta_{j'} = \emptyset \) and \( \bigcup_{j=1}^{D} \Theta_j = \Theta \), i.e., \( \Theta_1, \Theta_2, \ldots, \Theta_D \) is a partition of \( \Theta \). Then, the true parameter \( \theta^* \in \Theta_{j^*} \) for some \( 1 \leq j^* \leq D \). In addition, when \( \Theta_j = \{ \theta_j \} \) is a singleton set, this reduces to the setting in Section 2.

Since now the parameter space \( \Theta \) can be continuous, the posterior (density) distribution is computed as

\[
\pi(\theta | \xi_1, \ldots, \xi_m) = \frac{\pi_0(\theta) \prod_{\ell=1}^{m} f_\Theta(\xi_\ell)}{\int_{\Theta} \pi_0(\theta') \prod_{\ell=1}^{m} f_\Theta(\xi_\ell) d\theta'}.
\]

We further define

\[
\pi_j := \pi(\Theta_j | \xi_1, \ldots, \xi_m) = \frac{\int_{\Theta_j} \pi_0(\theta) \prod_{\ell=1}^{m} f_\Theta(\xi_\ell) d\theta}{\int_{\Theta} \pi_0(\theta) \prod_{\ell=1}^{m} f_\Theta(\xi_\ell) d\theta}
\]

and

\[
\hat{\pi}_j(\theta) = \frac{\pi_0(\theta)}{\pi_j} \prod_{\ell=1}^{m} f_\Theta(\xi_\ell).
\]

That is, \( \pi_j \) is the posterior probability of \( \{ \theta^* \in \Theta_j \} \) and \( \hat{\pi}_j \) is the conditional posterior probability of \( \theta^* \) conditioned on the event \( \theta^* \in \Theta_j \). Then the Bayesian average performance can be expressed as

\[
\bar{\mu}_i = \mathbb{E}_\pi[\mu_i(\theta)] = \sum_{j=1}^{D} \pi_j \mathbb{E}_\pi[\mu_i(\theta)|\theta \in \Theta_j] = \sum_{j=1}^{D} \pi_j \mathbb{E}_{\hat{\pi}_j}[\mu_i(\theta)].
\]

Instead of simulating a design under a fixed input parameter \( \theta_j \), we will then simulate a design under a fixed subspace \( \Theta_j \). To generate one simulation output under the design-input pair \( (i,j) \) (here input \( j \) refers to subspace \( \Theta_j \), as opposed to a single input parameter, \( \theta_j \), previously), we first generate \( \theta \sim \hat{\pi}_j \) and then under the input distribution \( \theta \) we generate one simulation output. Notably, both of the proposed algorithms, DD-OCBA-approx and DD-OCBA-balance, can then be directly applied to allocate simulation budget among the design-input pairs, with the only differences on the posterior updating and generation of simulation outputs. To differentiate with the setting of finite parameter space, we name the two procedures as DD-OCBA-approx-C and DD-OCBA-balance-C, respectively, where “C” stands for continuous.

Remark: In practice, computing the exact posterior probability \( \pi_j \) and sampling from the conditional posterior distribution \( \hat{\pi}_j \) can be computationally expensive. To improve the computation efficiency, given the current posterior
distribution $\pi$, one can first compute the Bayesian mean estimator $\bar{\theta} = E_\pi[\theta]$ and then let $\theta_j = \text{Proj}_{\Theta_j}(\bar{\theta})$, i.e., $\theta_j \in \Theta_j$ is the closest point in $\Theta_j$ to $\bar{\theta}$ (with respect to, e.g., $l_2$ distance). Then, one can run the simulation for $(i, j)$ pair under input parameter $\theta_j$. Furthermore, we approximate the posterior probability $\pi_j$ by the the posterior density at $\theta_j, j = 1, \ldots, D$ as $\pi_j \approx \frac{\pi(\theta_j)}{\sum_{j'=1}^{\ell} \pi(\theta_{j'})}$.

**Convergence Analysis**

In this section, we show that with continuous parameter space, Theorem 2 and Theorem 3 still hold. That is, we can still prove the consistency for DD-OCBA-approx-C and DD-OCBA-balance-C and asymptotic optimality for DD-OCBA-approx-C.

Before jumping into the analysis, we first give some intuitive explanation on the difference of proving the consistency/asymptotic optimality with continuous parameter space. Recall the only differences between DD-OCBA-approx(balance) and DD-OCBA-approx(balance)-C are the way of generating simulation output and updating the posterior distribution, which affects the convergence of posterior probability $\pi_j$, performance estimator $\hat{\mu}_{i,j}$ and variance estimator $\hat{\sigma}_{i,j}$. As a result, the same analysis for consistency and asymptotic optimality can be applied if these aforementioned estimators or posterior probabilities satisfy certain properties as required in the proof of Theorem 2-3, which include the strong consistency of the posterior distribution as well as the performance/variance estimators. With the finite parameter space, for a fixed design-input pair the simulation outputs are i.i.d.. However, with continuous parameter space, as the posterior distribution is being updated, the conditional posterior distribution $\pi_j$ also varies. Consequently, the simulation outputs under a fixed design-input (subspace) pair is not i.i.d.. across stages, requiring us to strengthen the current analysis.

Let $\pi^t, \pi_j^t$ and $\pi_j^t$ be the posterior distribution, posterior probability for $\{\theta^c \in \Theta_j\}$, and conditional posterior distribution for $\{\theta^c : \theta^c \in \Theta_j\}$ at stage $t$, respectively. We make the following assumptions for proving consistency of the algorithms.

**Assumption 9**

1. $\Theta \in \mathbb{R}^d, d < \infty$ is convex and compact. Furthermore, there exists $1 \leq j^c \leq D$, such that $\theta^c \in \text{Int}(\Theta_{j^c})$.

2. $y_i^{(k)}(\theta) := E[X_i^{k}(\theta)]$ is continuous in $\theta$ for $i \in I$ and $k = 1, 2, 4$.

3. At iteration $\ell$, the simulation output for some design-input pair $(i, j)$ is generated by first sampling $\theta^{(t)} \sim \pi_j^{(t)}$ and then simulating design $i$ under $\theta^{(t)}$. Moreover, conditioned on $\pi_j^{(t)}$, the simulation output is independent of the past input data and past simulation outputs.

Assumption 9 together with Assumption 2 and Assumption 6 guarantees the consistency of posterior distribution as well as the consistency of performance and variance estimator, as formally stated in the following Lemma 3.

**Lemma 3** Suppose Assumption [1, 2, 6] and [9] hold.

1. If $\lim_{t \to \infty} M(t) = \infty$, then $\pi_j^{(t)} \to 1$ as $t \to \infty$ for almost every sequence of input data.

2. If also $\lim_{t \to \infty} N_{i,j}^{(t)} = \infty$ for some design $i \in I$, then $\lim_{t \to \infty} \hat{\mu}_{i,j}^{(t)} = \mu_i(\theta^c)$ and $\lim_{t \to \infty} \hat{\sigma}_{i,j}^{(t)} = \sigma_i(\theta^c)$ almost surely.

With Lemma 3 we can then follow the same proofs of Theorem 2 and Theorem 3 to get the following consistency result for DD-OCBA-approx-C and DD-OCBA-balance-C.

**Corollary 1** Suppose Assumption [1, 2, 3, 4, 5, 6, 7, 9] hold. Then, both DD-OCBA-approx-C and DD-OCBA-balance-C selects the true optimal design $b^c := \arg \max \mu_i(\theta^c)$ almost surely. Furthermore, the allocation rule $\alpha^{(t)}_{i,j}, \forall$ given by DD-OCBA-approx-C converges to the limiting optimal allocation policy defined in Theorem 3.
7 NUMERICAL EXPERIMENT

7.1 Comparison Baselines

We test the performance of DD-OCBA-approx(-C) and DD-OCBA-balance(-C) by comparing with (i) Equal Allocation, which allocates an equal simulation budget to all design-input pairs, and (ii) Adaptive OCBA, which is an extension of the OCBA (see [8]) algorithm to our data-driven setting. OCBA sequentially decides the next system to simulate based on the past simulation samples, whereas Adaptive OCBA uses the same budget allocation rule as OCBA, but with simulation samples generated under the current estimated input distribution each time as opposed to a fixed input distribution in OCBA. In our context, with the current posterior distribution $\pi$, a simulation output for design $i$ is generated by first sampling $\theta \sim \pi$ and then running the simulation procedure under this parameter $\theta$ to get a sample of $X_i(\theta)$. In fact, Adaptive OCBA can be viewed as a special case of DD-OCBA-approx-C with number of sub-space $D = 1$ and the entire space $\Theta$ to be finite. On a related note, we do not explicitly compare with the SEIU algorithm by [51], which also considers the R&S with streaming input data but took a fixed confidence formulation. From the perspective of budget allocation, they simply equally allocate the budget to all designs with elimination until only one is left with a given confidence level. Hence, applying their method to the setting of fixed budget R&S is equivalent to using Equal Allocation.

We test the procedures on two different problems, beginning with a quadratic problem with finite input parameter space and then on a portfolio problem with continuous parameter space.

7.2 Quadratic Problem

Consider the following optimization problem.

$$\min_{i \in I} \mathbb{E}_{\theta^c} [(x_i - \zeta)^2] ,$$

where $x_i = \theta^c + i - 1, i \in I = \{1, 2, \ldots, 20\}$ and $\zeta$ follows an exponential distribution with unknown mean $\theta^c = 3$. Hence, the true best design $b^c = 0$. The input parameter space $\Theta = \{\theta_j : \theta_j = 0.5 \cdot j, j = 1, 2, \ldots, 20\}$. In the following we set the initial number of simulation for each design-input pair $n(0) = 10$. The initial prior is uninformative, that is, $\pi^0_j = \frac{1}{20}, j = 1, \ldots, 20$. We first test with constant stage-wise input data batch size and simulation budget. We set the stage-wise simulation budget $n(t) = 100$ and vary the input data batch size $m(t) \in \{5, 10, 15\}$.

Experiment Results

In Figure 2 We plot the empirical PCS of each algorithm at the end of each stage (with a total of 200) along with its 95% confidence interval. The empirical PCS and the confidence interval are calculated by running 500 macro-replications.

![Figure 2: Quadratic example with constant batch size and different choices of $m(t)$](image)

The observations from Figure 2 are summarized as follows:

1. Across all configurations of $m(t)$, DD-OCBA-approx achieves a final empirical PCS around 90% and DD-OCBA-balance achieves a final empirical PCS around 95%, both of which surpass the performance of the other algorithms. The Equal Allocation method, in comparison, is markedly less efficient (achieving a final empirical PCS lower than 60% in all 3 scenarios), primarily due to its overallocation of resources to design-input pairs characterized by incorrect input parameters. Furthermore, among the two highlighted methods, DD-OCBA-balance demonstrates superior performance over DD-OCBA-approx. This advantage

Data-driven R&S
stems from DD-OCBA-balance’s direct approach to satisfy the original optimality conditions, in contrast to DD-OCBA-approx, which relies on an approximation of the "local balance" condition.

2. In the comparison between DD-OCBA-approx and Adaptive OCBA, Adaptive OCBA always achieves a lower final empirical PCS around 80% in all 3 scenarios. Despite the different empirical performance by the two procedures, note that both allocation policies given by the two procedures will converge to the limiting optimal policy defined in Theorem\(^3\). The convergence of Adaptive OCBA is guaranteed by Corollary\(^4\) where Adaptive OCBA can be regarded as a special case of DD-OCBA-approx-C with the number of subspace \(D = 1\). Nonetheless, Adaptive OCBA is outperformed by DD-OCBA-approx-C. This is attributed to how Adaptive OCBA aggregates simulation outputs, which vary across distributions, leading to an initial performance estimator with significant bias. As a result, at early stages, when estimation is inaccurate, the simulation output has a large bias. This early bias diminishes over time with additional input data and simulations (Lemma\(^5\)). On the contrary, DD-OCBA-approx maintains an unbiased estimator for each design input pair, as simulations for these pairs are consistently conducted under the same input parameter, which implies more simulation outputs are always beneficial as it reduces the simulation error and does not introduce larger bias to the design-performance estimator. The only source of bias in DD-OCBA-approx comes from the posterior probability \(\pi_j, j = 1, \ldots, D\), which is independent of the simulation outputs. This fundamental distinction renders DD-OCBA-approx more effective, particularly in scenarios with streaming input data and smaller batch sizes, where Adaptive OCBA’s estimators are prone to larger biases.

7.3 Portfolio Optimization with Continuous Parameter Space and Random Batch Size

In the previous example, we test the performance of DD-OCBA-approx and DD-OCBA-balance with finite parameter space. In this example, we test DD-OCBA-approx-C and DD-OCBA-balance-C, the generalized versions of the proposed procedures, on a more general problem of portfolio optimization where the input parameter space is continuous. An investor invests a certain amount of capital in a riskless asset with interest rate \(r\) and a risky asset, whose price per share at time \(t\) is denoted as \(S_t\). \(\{S_t\}\) is often assumed to follow a Geometric Brownian motion with initial price \(S_0\), which admits the following expression for any fixed \(t\):

\[
S_t = S_0 \exp \left( \left( \theta^c - \frac{\sigma^2}{2} \right) t + \sigma B(t) \right),
\]

where \(\sigma\) is the volatility parameter, \(\theta^c\) is the drift, and \(\{B(t) : t \geq 0\}\) is a standard Brownian motion. At time 0, the investor makes a one-time decision \(x \in [0, 1]\), which is the proportion of investment in the risky asset. Then, the total wealth at time \(t\), denoted by \(W_t\), is

\[
W_t = xW_0 \exp \left( \left( \theta^c - \frac{\sigma^2}{2} \right) t + \sigma B(t) \right) + (1-x)W_0e^{rt}.
\]

A risk-averse investor aims to maximize the mean-variance of the total asset after \(T\) length of time with a risk-averse parameter \(\rho\), as follows:

\[
F(x) = \mathbb{E} \left\{ xW_0 \exp \left( \left( \theta^c - \frac{\sigma^2}{2} \right) T + \sigma B(T) \right) + (1-x)W_0e^{rt} \right\} - \rho x^2 W_0^2 e^{2\theta^c T} (e^{\sigma^2 T} - 1).
\]

Here the variance term is calculated explicitly using the distribution of log-normal random variable, and the expectation term needs to be estimated. Furthermore, we assume the drift \(\theta^c\) (also known as the risky return rate) is unknown but can be estimated with streaming data. Specifically, suppose we have a sequence of observations \(\{S_{k\tau}\}_{k \in \mathbb{N}}\), where \(S_{k\tau}\) is the price of the risky asset at time \(k\tau\). From (17), we know \(Z_k := \log \left( \frac{S_{k+1}\tau}{S_{k}\tau} \right) \sim \mathcal{N} \left( (\theta^c - \frac{\sigma^2}{2})\tau, \sigma^2 \right)\) and \(\{Z_k\}_{k \in \mathbb{N}}\) is serially independent. Hence, we can employ a Normal-Normal conjugate prior to estimate the unknown drift \(\theta^c\).

As for implementation details, we set the initial wealth \(W_0 = 1\), the interest rate of riskless asset \(r = 0.5\), volatility \(\sigma = 1\), drift \(\theta^c = 2.5\), risk-averse parameter \(\rho = 0.1\). The candidate set \(\mathcal{I} = \{0, 1, \ldots, 10\}\) with \(i\)th candidate being the solution \(x_i = 0.1 \cdot i\). We partition the entire parameter space \((\theta^c)\) \(\Theta = \bigcup_{j=-10,10} \Theta_j\) where \(\Theta_j = [\theta^c + 0.2 \cdot j \cdot 0.1, \theta^c + 0.2 \cdot j \cdot 0.1], -9 \leq j \leq 9\), \(\Theta_{-10} = (-\infty, 0.6]\) and \(\Theta_{10} = [4.4, \infty)\). The prior distribution is set to be \(\mathcal{N}(0, \sigma^2\tau)\). Furthermore, the stage-wise input data batch size and simulation budget is set to be random. Specifically, the input data batch size \(m(t) = \bar{n} * Z\) and the simulation budget \(n(t) = \bar{n} * Z\) with \(Z\) being a random variable equally distributed among \(\{1, 2, 3, 4, 5\}\) and \(\bar{n}, \bar{n} \in \mathbb{N}_+\). We set the average stage-wise simulation budget \(\bar{n} = 2.5\bar{n} = 50\) and varies the stage-wise input data batch size \(\bar{m} = 2.5\bar{n}\) from \(\{5, 10, 20\}\). In Figure\(^3\) we plot the empirical PCS of each algorithm at the end of each stage (with a total of 200 stages) along with its 90% confidence interval. The empirical PCS and the confidence interval are calculated by running 500 macro-replications.
8 Conclusion and Future Work

In this paper we consider a fixed budget ranking and selection (R&S) problem, where the common input distribution across designs is unknown but can be estimated with streaming input data that come sequentially in time. We initially assume a finite parametrization of the input distribution and utilize a Bayesian approach for estimation, which is updated at the beginning of each stage as batched data arrives. Subsequently, the stage-wise computing budget is allocated to facilitate new simulations for assessing design performances. We apply the large deviations theory to obtain the optimal stage-wise budget allocation policy for design-input pairs. Based on the optimality equations, we design two fully sequential algorithms that achieve consistency (i.e., select the best design with probability 1 as times go to infinity) and asymptotic optimality (i.e., converge to the optimal budget allocation policy under the true input distribution). We also extend our approach to accommodate a continuous input parameter space, while maintaining provable statistical validity. Our numerical experiments demonstrate superior performance of the proposed procedures over the equal allocation rule and an extension of OCBA when dealing with unknown input distributions with streaming input data.

For future research, it would be intriguing to explore methods for adaptively adjusting the finite input parameter space as additional input data becomes available. This adjustment could help correct any potential discretization errors in the original parameter space. One possible approach could be based on criteria or indices like the Bayesian Information Criterion (BIC) or the Davies-Bouldin Index, which would guide the refinement process. Another promising research direction is to incorporate a non-parametric Bayesian framework. This approach could provide a flexible way to address and correct model mismatches arising from the parametric assumptions initially made about the input distribution. By using non-parametric methods, researchers could potentially capture a wider variety of distribution shapes and complexities, thereby enhancing the robustness and accuracy of the modeling process.
ACKNOWLEDGMENT

The authors gratefully acknowledge the support by the Air Force Office of Scientific Research under Grant FA9550-22-1-0244, the National Science Foundation under Grant NSF-DMS2053489, and AI Institute for Advances in Optimization (AI4OPT).

BIOGRAPHY

YUHAO WANG is a Ph.D. candidate at the H. Milton Stewart School of Industrial and Systems Engineering at Georgia Institute of Technology. He received his B.S. degree from the Department of Mathematics at Nanjing University, China, in 2021. His research interests include simulation, stochastic optimization, and reinforcement learning.

ENLU ZHOU is a Professor in the H. Milton Stewart School of Industrial and Systems Engineering at Georgia Institute of Technology. She received the B.S. degree with highest honors in electrical engineering from Chu Kochen Honors College, Zhejiang University, China, in 2004, and the Ph.D. degree in electrical engineering from the University of Maryland, College Park, in 2009. Prior to joining Georgia Tech in 2013, she was an assistant professor in the Department of Industrial and Enterprise Systems Engineering at the University of Illinois Urbana-Champaign from 2009 to 2013. She is a recipient of the AFOSR Young Investigator award in 2012, NSF CAREER award in 2014, the INFORMS Outstanding Simulation Publication award in 2020, and the Best Theoretical Paper award at the Winter Simulation Conference twice in 2009 and 2022. Her research interests lie in theory, methods, and applications of simulation optimization, stochastic optimization, and stochastic control.

References

[1] Robert E Bechhofer. A single-sample multiple decision procedure for ranking means of normal populations with known variances. The Annals of Mathematical Statistics, pages 16–39, 1954.
[2] S.S. Gupta. On a Decision Rule for a Problem in Ranking Means. University of North Carolina at Chapel Hill, 1956.
[3] Medicine National Academies of Sciences, Engineering et al. Foundational research gaps and future directions for digital twins. 2023.
[4] Haoxiang Yang and David P Morton. Optimal crashing of an activity network with disruptions. Mathematical Programming, 194(1-2):1113–1162, 2022.
[5] Peng Wan, Michael Fu, and Steven Marcus. Crashing of activity networks with disruptions using gradient-based simulation optimization. preprint, 2023.
[6] Fred Brauer. Compartmental models in epidemiology. Mathematical epidemiology, pages 19–79, 2008.
[7] Peng Chen, Keyi Wu, and Omar Ghattas. Bayesian inference of heterogeneous epidemic models: Application to covid-19 spread accounting for long-term care facilities. Computer Methods in Applied Mechanics and Engineering, 385:114020, 2021.
[8] Chun-Hung Chen, Jianwu Lin, Enver Yücesan, and Stephen E Chick. Simulation budget allocation for further enhancing the efficiency of ordinal optimization. Discrete Event Dynamic Systems, 10(3):251–270, 2000.
[9] Yanwen Li and Siyang Gao. On the convergence of optimal computing budget allocation algorithms. In Proceedings of the 2021 Winter Simulation Conference, pages 1–12, 2021.
[10] Chun-Hung Chen, Donghai He, Michael Fu, and Loo Hay Lee. Efficient simulation budget allocation for selecting an optimal subset. INFORMS Journal on Computing, 20(4):579–595, 2008.
[11] Siyang Gao and Weiwei Chen. Efficient subset selection for the expected opportunity cost. Automatica, 59:19–26, 2015.
[12] Siyang Gao, Jianzhong Du, and Chun-Hung Chen. Selecting the optimal system design under covariates. In 2019 IEEE 15th International Conference on Automation Science and Engineering (CASE), 2019.
[13] Xiao Jin, Haobin Li, and Loo Hay Lee. Optimal budget allocation in simulation analytics. In 2019 IEEE 15th International Conference on Automation Science and Engineering (CASE), 2019.
[14] Loo Hay Lee, Ek Peng Chew, Suyan Teng, and David Goldsman. Optimal computing budget allocation for multi-objective simulation models. In Proceedings of the 2004 Winter Simulation Conference, pages 594–602, 2004.
Data-driven R&S

[15] Qing-Shan Jia, Enlu Zhou, and Chun-Hung Chen. Efficient computing budget allocation for finding simplest good designs. *IIE Transactions*, 45(7):736–750, 2013.

[16] Shen Yan, Enlu Zhou, and Chun-Hung Chen. Efficient selection of a set of good enough designs with complexity preference. *Automation Science and Engineering, IEEE Transactions on*, 9:596–606, 07 2012.

[17] Siyang Gao, Weiwei Chen, and Leyuan Shi. A new budget allocation framework for the expected opportunity cost. *Operations Research*, 65(3):787–803, 2017.

[18] Siyang Gao, Hui Xiao, Enlu Zhou, and Weiwei Chen. Robust ranking and selection with optimal computing budget allocation. *Automatica*, 81:30–36, 2017.

[19] Chun-Hung Chen and Loo Hay Lee. *Stochastic simulation optimization: an optimal computing budget allocation*, volume 1. World scientific, 2011.

[20] Yuhaao Wang and Enlu Zhou. Fixed budget ranking and selection with streaming input data. In *2022 Winter Simulation Conference (WSC)*, pages 3027–3038, 2022.

[21] Amir Dembo, Ofer Zeitouni, and IH Dinwoodie. Large deviations techniques and applications. *SIAM Review*, 36(2):303–303, 1994.

[22] Peter Glynn and Sandeep Juneja. A large deviations perspective on ordinal optimization. In *Proceedings of the 2004 Winter Simulation Conference*, pages 489–494, 2004.

[23] Seong-Hee Kim and Barry L Nelson. A fully sequential procedure for indifference-zone selection in simulation. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 11(3):251–273, 2001.

[24] L Jeff Hong. Fully sequential indifference-zone selection procedures with variance-dependent sampling. *Naval Research Logistics (NRL)*, 53(5):464–476, 2006.

[25] Peter I Frazier. A fully sequential elimination procedure for indifference-zone ranking and selection with tight bounds on probability of correct selection. *Operations Research*, 62(4):926–942, 2014.

[26] Seong-Hee Kim and Barry L Nelson. Recent advances in ranking and selection. In *Proceedings of the 2007 Winter Simulation Conference*, pages 162–172, 2007.

[27] Stephen E Chick. Input distribution selection for simulation experiments: accounting for input uncertainty. *Operations Research*, 49(5):744–758, 2001.

[28] Stephen E Chick and Peter Frazier. Sequential sampling with economics of selection procedures. *Management Science*, 58(3):550–569, 2012.

[29] Sijia Ma and Shane G Henderson. An efficient fully sequential selection procedure guaranteeing probably approximately correct selection. In *Proceedings of the 2017 Winter Simulation Conference*, pages 2225–2236, 2017.

[30] Ye Chen and Ilya O Ryzhov. Complete expected improvement converges to an optimal budget allocation. *Advances in Applied Probability*, 51(1):209–235, 2019.

[31] Susan R Hunter and Raghu Pasupathy. Optimal sampling laws for stochastically constrained simulation optimization on finite sets. *INFORMS Journal on Computing*, 25(3):527–542, 2013.

[32] Raghu Pasupathy, Susan R Hunter, Nugroho A Pujowidianto, Loo Hay Lee, and Chun-Hung Chen. Stochastically constrained ranking and selection via score. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 25(1):1–26, 2014.

[33] Sait Cakmak, Siyang Gao, and Enlu Zhou. Contextual ranking and selection with gaussian processes. *arXiv preprint arXiv:2201.07782*, 2022.

[34] Donghai He, Stephen E Chick, and Chun-Hung Chen. Opportunity cost and ocba selection procedures in ordinal optimization for a fixed number of alternative systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 37(5):951–961, 2007.

[35] Stephen E Chick, Jürgen Branke, and Christian Schmidt. Sequential sampling to myopically maximize the expected value of information. *INFORMS Journal on Computing*, 22(1):71–80, 2010.

[36] Peter Frazier, Warren Powell, and Savas Dayanik. The knowledge-gradient policy for correlated normal beliefs. *INFORMS Journal on Computing*, 21(4):599–613, 2009.

[37] L Jeff Hong, Weiwei Fan, and Jun Luo. Review on ranking and selection: A new perspective. *Frontiers of Engineering Management*, 8(3):321–343, 2021.

[38] Faker Zouaoui and James R Wilson. Accounting for input-model and input-parameter uncertainties in simulation. *IIE Transactions*, 36(11):1135–1151, 2004.
[39] Szu Hui Ng and Stephen E Chick. Reducing parameter uncertainty for stochastic systems. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 16(1):26–51, 2006.

[40] Wei Xie, Barry L Nelson, and Russell R Barton. A bayesian framework for quantifying uncertainty in stochastic simulation. *Operations Research*, 62(6):1439–1452, 2014.

[41] Henry Lam and Enlu Zhou. The empirical likelihood approach to quantifying uncertainty in sample average approximation. *Operations Research Letters*, 45(4):301–307, 2017.

[42] Canan G Corlu, Alp Akcay, and Wei Xie. Stochastic simulation under input uncertainty: A review. *Operations Research Perspectives*, page 100162, 2020.

[43] Canan G Corlu and Bahar Biller. A subset selection procedure under input parameter uncertainty. In *Proceedings of the 2013 Winter Simulation Conference*, pages 463–473, 2013.

[44] Canan G Corlu and Bahar Biller. Subset selection for simulations accounting for input uncertainty. In *Proceedings of the 2015 Winter Simulation Conference*, pages 437–446, 2015.

[45] Di Wu and Enlu Zhou. Ranking and selection under input uncertainty: a budget allocation formulation. In *Proceedings of the 2017 Winter Simulation Conference*, pages 2245–2256, 2017.

[46] Jingxu Xu, Zeyu Zheng, and Peter W Glynn. Joint resource allocation for input data collection and simulation. In *2020 Winter Simulation Conference (WSC)*.

[47] Hui Xiao and Siyang Gao. Simulation budget allocation for selecting the top-m designs with input uncertainty. *IEEE Transactions on Automatic Control*, 63(9):3127–3134, 2018.

[48] Hui Xiao, Fei Gao, and Loo Hay Lee. Optimal computing budget allocation for complete ranking with input uncertainty. *IISE Transactions*, 52(5):489–499, 2020.

[49] Weiwei Fan, L Jeff Hong, and Xiaowei Zhang. Distributionally robust selection of the best. *Management Science*, 66(1):190–208, 2020.

[50] Eunhye Song and Barry L Nelson. Input–output uncertainty comparisons for discrete optimization via simulation. *Operations Research*, 67(2):562–576, 2019.

[51] Di Wu, Yuhao Wang, and Enlu Zhou. Data-driven ranking and selection under input uncertainty. *Operations Research*, 2022.

[52] Taeho Kim and Eunhye Song. Optimizing input data acquisition for ranking and selection: A view through the most probable best. In *2022 Winter Simulation Conference (WSC)*, pages 2258–2269, 2022.

[53] Aad W Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.

[54] Heng Lian. On rates of convergence for posterior distributions under misspecification. *Communications in Statistics—Theory and Methods*, 38(11):1893–1900, 2009.

[55] Miklós Csörgő. On the strong law of large numbers and the central limit theorem for martingales. *Transactions of the American Mathematical Society*, 131(1):259–275, 1968.

**Technical Proof**

Recall $\ell$ is the iteration counter. Denote by $t_{\ell}$ the stage where $\ell$th simulation is run. That is

$$t_{\ell} := \arg \max_t \{ t : N(t-1) < \ell \},$$

(18)

where $N(t) = \sum_{\tau=1}^{t} n(t)$ is total simulation budget up to stage $t$.

**1 Proof of Lemma[1]**

**Proof.** Lemma[1] is easily seen from (4). To prove Lemma[2], it suffices to show the concavity of the function for $x > 0$ with form $f(x) = 1/(\sum_{i=1}^{n} a_i x_i)$, where $a_i > 0$ for $i = 1, 2, \ldots, n$. We prove the concavity of the multivariate function by proving the concavity along all lines. For any $y \in \mathbb{R}^n$, let $g(t) = f(x + ty)$ where $t \in \mathbb{R}$ such that $x + ty > 0$. We have

$$g''(t) = \frac{2}{(\sum_{i=1}^{n} a_i x_i + ty_i)^3} \left\{ \sum_{i=1}^{n} a_i y_i \right\}^2 - \frac{n}{(x_i + ty_i)\sum_{i=1}^{n} a_i x_i + ty_i} \leq 0,$$

where the inequality uses the Cauchy inequality. Hence, $f$ is concave in $x > 0$. □
\section{Proof of Theorem 1}

\textbf{Proof.} We first show the existence of \(\alpha\). The existence follows from the continuity of \(G_i\) with respect to \(\alpha \in \Delta^{K^B-1}\), where \(\Delta^n\) denotes the \(n\)-dimensional simplex. Furthermore, by Lemma 1.1, \(G_i\) is strictly increasing in \(\alpha_{i,j}\) and \(\alpha_{i,j'}\). Since \(\alpha_{i,j} = \frac{1}{K^B} \forall i, j\) is a feasible solution and the corresponding objective value is strictly positive, the optimal solution \(\alpha\) must satisfy \(\alpha_{i,j} > 0\) for all \(i, j\).

Now we show the necessity of the three optimality conditions. By Lemma 1.2, the optimization problem (3) is a concave maximization problem, and therefore the KKT conditions are both sufficient and necessary for the optimality. With \(\alpha\) strictly positive, the KKT conditions can be written as

\begin{equation}
1 - \sum_{i \neq b} \lambda_i = 0, \quad (19)
\end{equation}
\begin{equation}
\lambda_i \frac{\partial G_i}{\partial \alpha_{i,j}} (\alpha_b, \alpha_i) = \gamma \quad i \neq b, \quad 1 \leq j \leq D, \quad (20)
\end{equation}
\begin{equation}
\sum_{i \neq b} \lambda_i \frac{\partial G_i}{\partial \alpha_{i,j}} (\alpha_b, \alpha_i) = \gamma \quad 1 \leq j \leq D, \quad (21)
\end{equation}
\begin{equation}
\lambda_i (G_i (\alpha_b, \alpha_i) - z) = 0 \quad i \neq b, \quad (22)
\end{equation}

for some \(\gamma\) and \(\lambda_i \geq 0\), \(i \neq b\). By (19) there exists at least one \(i_0\) such that \(\lambda_{i_0} > 0\). Then since \(G_i\) is increasing in \(\alpha_{i,j}\), we have \(\frac{\partial G_i}{\partial \alpha_{i,j}} (\alpha_b, \alpha_{i_0}) > 0\). This implies \(\gamma > 0\) by (20). Hence, we must have \(\lambda_i > 0\) for all \(i \neq b\). Then we have \(\frac{\partial G_i}{\partial \alpha_{i,j}} = \frac{\partial G_i}{\partial \alpha_{i,j}} (\alpha_b, \alpha_i) = 1 \leq j \leq D, \quad i \neq b\), which proves (6). Since \(\lambda_i > 0\), \(G_i (\alpha_b, \alpha_i) = z\), \(i \neq b\) by (22). Hence, (7) holds. To see why (4) holds, solving for \(\lambda_i = \frac{\gamma}{\frac{\partial G_i}{\partial \alpha_{i,j}} (\alpha_b, \alpha_i)}\) in (20) and substituting \(\lambda_i\) in (21), we get the desired result.

For sufficiency, first let \(\lambda_i = \frac{1}{\frac{\partial G_i}{\partial \alpha_{i,j}} (\alpha_b, \alpha_i)} (\sum_{i \neq b} \frac{1}{\frac{\partial G_i}{\partial \alpha_{i,j}} (\alpha_b, \alpha_i)})\) for \(i \neq b\). Notice that \(\lambda_i > 0\) and does not depend on the choice of \(j\) by (5). Moreover, \(\{i\}_i \neq b\) satisfy condition (19). Further let \(\gamma = \left(\sum_{i \neq b} \frac{1}{\frac{\partial G_i}{\partial \alpha_{i,j}} (\alpha_b, \alpha_i)}\right)^{-1}\), which is also independent of \(j\). We can easily verify that both (20) and (21) hold. (19) also holds by setting \(z = G_i (\alpha_b, \alpha_i)\), which is independent of \(i\) by (7).

Now we are only left to show the uniqueness of \(\alpha\). First notice that from (8) and (9), we have

\begin{equation}
\frac{\alpha_{b,j}}{\pi_j} = \frac{\gamma}{\sigma_j \sigma_{j'}} = \sqrt{\sum_{i \neq b} \frac{\alpha_{i,j}}{\sigma_j \sigma_{j'}}} = \sqrt{\sum_{i \neq b} \frac{\alpha_{i,j'}}{\sigma_j \sigma_{j'}}} = 1 \leq j < j' \leq D.
\end{equation}

Letting \(\beta_i = \alpha_{i,j}/\sigma_j\pi_j\) which is independent of \(j\), we can write \(\alpha_{i,j} = \pi_j \sigma_{i,j} \beta_i\) for all \(i = 1, 2, \cdots, K\) and \(j = 1, 2, \cdots, D\). Since \(\alpha\) and \(\beta = (\beta_b, \cdots, \beta_K)\) are bijective, it is sufficient to show the uniqueness of \(\beta\). Plugging \(\alpha_{i,j}\) into (10) and (9), we have

\begin{equation}
\frac{(\mu_b - \mu_i)^2}{\sum_{j=1}^D \sigma_{b,j} \pi_j + \sum_{j=1}^D \sigma_{i,j} \pi_j} = \frac{(\mu_b - \mu_i)^2}{\sum_{j=1}^D \sigma_{b,j} \pi_j + \sum_{j=1}^D \sigma_{i,j} \pi_j} \quad i \neq b \neq \beta
\end{equation}

with \(\beta^2_b = \sum_{i \neq b} \beta_i^2\). Let \(\eta = \beta / \mu_i\). Then \(\eta\) satisfies

\begin{equation}
\frac{(\mu_b - \mu_i)^2}{\sum_{j=1}^D \sigma_{b,j} \pi_j + \sum_{j=1}^D \sigma_{i,j} \pi_j} = \frac{(\mu_b - \mu_i)^2}{\sum_{j=1}^D \sigma_{b,j} \pi_j + \sum_{j=1}^D \sigma_{i,j} \pi_j} \quad i \neq b \neq \beta
\end{equation}

with \(1 = \sum_{i \neq b} \eta_i^2\). If there exists \(\eta' \neq \eta\) satisfying these two conditions, then there must be \(i \neq k \neq 1\) such that \(\eta_i < \eta_i'\) and \(\eta_k > \eta_k'\). Then, we have

\begin{equation}
\frac{(\mu_b - \mu_k)^2}{\sum_{j=1}^D \sigma_{b,j} \pi_j + \sum_{j=1}^D \sigma_{k,j} \pi_j} < \frac{(\mu_b - \mu_k)^2}{\sum_{j=1}^D \sigma_{b,j} \pi_j + \sum_{j=1}^D \sigma_{k,j} \pi_j} \quad \text{which contradicts (23)}.
\end{equation}

Hence, \(\eta\) is unique, which implies \(\beta = C \eta\) for some constant \(C\). Then if there exists \(\beta' \neq \beta\) and both are optimal, we have \(\beta > (\langle\rangle)\beta'\). This implies the corresponding \(\alpha > (\langle\rangle)\alpha'\), which contradicts \(\sum_{i,j} \alpha_{i,j} = \sum_{i,j} \alpha'_{i,j} = 1\). \hfill \Box
.3 Proof of Theorem 2

Proof: Denote by $\omega$ a sample path of one simulation process which contains all the simulation outputs and input data observation. To prove for either DD-OCA-approx or DD-OCA-balance, it suffices to show $N_{i,j}^{(t)} \to \infty$, as $\ell \to \infty$ almost surely for all $i$. To see this, recall $X_{(r)}$ is the $r$th simulation output for $(i,j)$ pair (the iteration at which $X_{(r)}$ is simulated is still random). By Assumption 3, $\{X^{(r)}_{i,j}\}$ are i.i.d. for $r = 1, 2, \ldots$. Hence $\frac{1}{n} \sum_{r=1}^{n} X^{(r)}_{i,j}$ converges to $\mu_{i,j}$ almost surely by LLN. Take any sample path $\omega$, then almost surely $N_{i,j}^{(t)}|\omega \to \infty$ as $\ell \to \infty$. Under the same sample path $\frac{1}{n} \sum_{r=1}^{n} X^{(r)}_{i,j}|\omega \to \mu_{i,j}$ as $n \to \infty$. Hence, $\frac{1}{N_{i,j}^{(t)}} \sum_{r=1}^{N_{i,j}^{(t)}} X^{(r)}_{i,j}|\omega \to \mu_{i,j}$ as $\ell \to \infty$. Furthermore, $\hat{\sigma}_{i,j}^{(t)} \to \sigma_{i,j}$ almost surely by the same argument and $\sigma_{i,j}^{(t)} \to 1$ almost surely by Doob’s consistency theorem. Therefore, we have $\hat{\mu}_{i,j}^{(t)} \to \mu^{(t)}_i$ as $\ell \to \infty$ almost surely. We fix a sample path $\omega$ in the following proof of showing $N_{i,j}^{(t)} \to \infty$, as $\ell \to \infty$ almost surely. Denote by $A = \{i|N_{i,j}^{(t)} \to \infty\}$. Notice $\hat{\mu}_{i,j}^{(t)}, \hat{\sigma}_{i,j}^{(t)}$ will converge (to some random variable) almost surely, no matter whether $N_{i,j}^{(t)}$ tend to infinity. This is because if $N_{i,j}^{(t)}$ is at most finite, then $\hat{\mu}_{i,j}^{(t)}$ and $\hat{\sigma}_{i,j}^{(t)}$ will remain unchanged after finite iterations. Since $\pi_{i,j}^t$ converges almost surely, $\hat{\mu}_{i,j}^{(t)}$ will also converge almost surely. This implies $\hat{b}_{i,j}^{(t)}$ also converge (to some random design index) almost surely. We use $\mu_{i,j}, \sigma_{i,j}, \hat{b}$ to denote the limit of $\hat{\mu}_{i,j}^{(t)}, \hat{\sigma}_{i,j}^{(t)}$ and $\hat{b}$, respectively.

Proof of Theorem 2

Denote by $N_{i,j}^{(t)} = \sum_{t\neq j} N_{i,j}^{(t)}$. Then there exists an allocation policy $\{\hat{\alpha}_{i,j}\}$ satisfying $\lim_{\ell \to \infty} \frac{N_{i,j}^{(t)}}{N_{i,j}^{(t)}} = \hat{\alpha}_{i,j}$. Furthermore, $\hat{\alpha}_{i,j} = 0$ if $i \neq j^c$ and $\hat{\alpha}_{i,j} > 0$ if $j = j^c$. This is because for $i, i' \neq \hat{b}$, and $j \neq j^c$,

$$N_{i,j}^{(t)} \to \frac{\pi_{i,j}^{t}}{\pi_{i,j}^{t}} \sum_{k=1}^{D} \hat{\sigma}_{i,k}^{(t)} \pi_{k,j}^{t} / (\hat{\mu}_{b,j}^{(t)} - \hat{\mu}_{i,j}^{(t)})^2 \to 0,$$

by the convergence of $\hat{\mu}_{i,j}^{(t)}, \hat{\sigma}_{i,j}^{(t)}$ and the fact that $\pi_{i,j}^{t} \to 0, \pi_{i,j}^{t} \to 1$ by Doob’s consistency theorem. This implies $\hat{\alpha}_{i,j} = 0$. Furthermore, since $N_{i,j}^{(t)} = \hat{\sigma}_{i,j}^{(t)} \sum_{t \neq \hat{b}} \frac{N_{i,j}^{(t)}}{\hat{\alpha}_{i,j}^{(t)}}$, we know $\hat{\alpha}_{i,j} = 0$. Moreover, we have

$$\frac{\hat{\alpha}_{i,j}}{\alpha_{i,j}} = \lim_{\ell \to \infty} \frac{N_{i,j}^{(t)}}{N_{i,j}^{(t)}} \to \frac{\hat{\sigma}_{i,j}^{(t)}}{\sigma_{i,j}^{(t)}} \frac{(\hat{\mu}_{i,j}^{(t)} - \hat{\mu}_{b,j}^{(t)})^2}{(\hat{\mu}_{i,j}^{(t)} - \hat{\mu}_{i,j}^{(t)})^2},$$

This implies $\hat{\alpha}_{i,j} > 0, \forall i \in I$. Suppose there exists $i_0 \notin A$. Let $\varepsilon = \frac{\hat{\alpha}_{i,j} \varepsilon}{K D}$, there exists $L$ large enough, such that $\forall \ell \geq L, \forall i, j, \frac{N_{i,j}^{(t)}}{N_{i,j}^{(t)}} - \hat{\alpha}_{i,j} \leq \frac{\hat{\varepsilon}}{K D}$. Since $N_{i,j}^{(t)}$ remains unchanged after some iteration $\ell_0$, we have $\frac{N_{i,j}^{(t)} - N_{i,j}^{(t)}}{N_{i,j}^{(t)}} > 0$ for $\ell > L$ and for large enough. Notice for any $i, j$, if $N_{i,j}^{(t)} > (\hat{\alpha}_{i,j} + \frac{1}{2})N_{i,j}^{(t)} > \hat{N}_{i,j}^{(t)}$, then $(i, j)$ will not be simulated since $\hat{N}_{i,j}^{(t)} - N_{i,j}^{(t)} < 0 < \hat{N}_{i,j}^{(t)} - N_{i,j}^{(t)}$. Hence, we must have $N_{i,j}^{(t)} \leq (\hat{\alpha}_{i,j} + \frac{1}{2})N_{i,j}^{(t)} + 1 \leq (\hat{\alpha}_{i,j} + \varepsilon)N_{i,j}^{(t)}$ for $\ell > L$ and large enough. Then, we have

$$N^{(t)} = \sum_{i,j} N_{i,j}^{(t)} = N_{i,j}^{(t)} + \sum_{(i,j) \neq (i_0,j_0)} N_{i,j}^{(t)} \leq N_{i,j}^{(t)} + N^{(t)} \sum_{(i,j) \neq (i_0,j_0)} (\hat{\alpha}_{i,j} + \varepsilon) \tag{24}$$

Divided by $N^{(t)}$ on both sides and let $\ell \to \infty$, (24) implies

$$1 \leq \sum_{(i,j) \neq (i_0,j_0)} \hat{\alpha}_{i,j} + (KD - 1)\varepsilon = \sum_{(i,j) \neq (i_0,j_0)} \hat{\alpha}_{i,j} + \frac{KD - 1}{KD} \hat{\alpha}_{i_0,j_0} < 1,$$

a contradiction to $i_0 \notin A$. The proof is complete.

Proof of Theorem 22. It suffices to prove $\lim_{\ell \to \infty} N_{i,j}^{(t)} \to \infty$ almost surely. We first prove $A \neq \emptyset$ by contradiction. Suppose $i \notin A, \forall i \in I$. We know there exists $i_0, j \neq j^c$, such that $N_{i_0,j}^{(t)} \to \infty$. This implies $\frac{N_{i_0,j}^{(t)}}{\hat{\alpha}_{i_0,j} \pi^t_{i_0,j}} \to \infty$, which

Data-driven R&S
can happens only if \( \frac{N_{i,j}^{(t)}}{\hat{\sigma}_{i,j}^{(t)}} \to \infty, \forall j \) otherwise \((i, j)\) cannot be sampled after some iteration. Hence, we have

\[
\frac{N_{i,j}^{(t)}}{\hat{\sigma}_{i,j}^{(t)}} \to \infty, \text{ which further implies } N_{i,j}^{(t)} \to \infty \text{ since } \pi_{j}^{*} \to 1. \text{ Hence we prove } A \neq .
\]

Next, we prove there exists \( j, N_{b,j}^{(t)} \to \infty \). Suppose not, then for all \( j \)

\[
\left( \frac{N_{b,j}^{(t)}}{\hat{\sigma}_{b,j}^{(t)}} \right)^2 - \sum_{i \neq b} \left( \frac{N_{i,j}^{(t)}}{\hat{\sigma}_{i,j}^{(t)}} \right)^2 \leq \left( \frac{N_{b,j}^{(t)}}{\hat{\sigma}_{b,j}^{(t)}} \right)^2
\]

is upper bounded by a constant. Furthermore we know there exists \( i' \in A \), hence

\[
\left( \frac{N_{b,j}^{(t)}}{\hat{\sigma}_{b,j}^{(t)}} \right)^2 - \sum_{i \neq b} \left( \frac{N_{i,j}^{(t)}}{\hat{\sigma}_{i,j}^{(t)}} \right)^2 \to -\infty.
\]

This implies \( i' \) cannot be simulated after some iteration, a contradiction to \( N_{i',j}^{(t)} \to \infty \). Hence, we know there exists \( j, N_{b,j}^{(t)} \to \infty \). This implies

\[
\frac{N_{b,j}^{(t)}}{\pi_{j}^{*} \hat{\sigma}_{b,j}^{(t)}} \to \infty, \text{ which further implies } N_{b,j}^{(t)} \to \infty \text{ wit the similar argument for proving } A \neq . \text{ Hence, we prove } b \in A.
\]

Next, we prove there exists \( i' \neq b, i \in A \). Since \( N_{b,j}^{(t)} \to \infty, N_{i',j}^{(t)} \to \infty, \pi_{j}^{i} \to 0, j \neq j' \), we know

\[
\frac{(\hat{\mu}_{i}^{(t)} - \hat{\mu}_{i}^{(t)')^2}}{N_{b,j}^{(t)}} + \frac{(\hat{\sigma}_{i}^{(t)} + \sigma_{i}^{(t)')^2}}{N_{b,j}^{(t)}} \to \infty.
\]

\[
\sum_{j=1}^{D} \frac{N_{i,j}^{(t)}, j}{} + \sum_{j=1}^{D} \frac{(\hat{\sigma}_{i}^{(t)} + \sigma_{i}^{(t)')^2}}{N_{i,j}^{(t)}} \to \infty.
\]

Finally we prove for all \( i \neq b, i \in A \). Since \( N_{b,j}^{(t)} \to \infty, N_{i',j}^{(t)} \to \infty, \pi_{j}^{i} \to 0, j \neq j' \), we know

\[
\sum_{j=1}^{D} \frac{N_{b,j}^{(t)}}{} + \sum_{j=1}^{D} \frac{(\hat{\sigma}_{i}^{(t)} + \sigma_{i}^{(t)')^2}}{N_{i,j}^{(t)}} \to \infty.
\]

\[
\sum_{j=1}^{D} \frac{N_{i,j}^{(t)}}{} + \sum_{j=1}^{D} \frac{(\hat{\sigma}_{i}^{(t)} + \sigma_{i}^{(t)')^2}}{N_{i,j}^{(t)}} \to \infty. \text{ This implies that } N_{i,j}^{(t)} \to \infty, \forall i \neq b. \text{ So far the proof is complete.}
\]

.4 Proof of Theorem[3]

Proof. Again denote by \( \omega \) any sample path of one simulation process and we fix a sample path \( \omega \). Since \( \pi_{j}^{i} \to 0, j \neq j' \), \( \pi_{j}^{i} \to 0 \) and \( N_{i,j}^{(t)} \to \infty \) as \( \ell \to \infty \), we have \( \hat{\sigma}_{i}^{(t)} \to 0 \) for \( i \in I, j \neq j' \) and \( \hat{\sigma}_{i}^{(t)} \to \sigma_{i}^{j}, i \in I \) as \( \ell \to \infty \).

For simplicity, let \( \sigma_{i}^{j} = 0 \) for \( j \neq j' \). Let \( A = \{(i, j) : N_{i,j}^{(t)} \to \infty \} \). Clearly \((i, j') \in A, \forall i \in I \) by the consistency result. If \( (i, j) \notin A \), then we have \( \sigma_{i,j}^{(t)} \to \sigma_{i,j}^{*} = 0 \). Hence, \( \forall \varepsilon > 0, \exists \tilde{L} \) such that for all \( \ell > \tilde{L}, |\sigma_{i,j}^{(t)} - \sigma_{i,j}^{*} | < \varepsilon \forall i, j, \sigma_{i,j}^{(t)} \leq \varepsilon \forall (i, j) \notin A \) and \( \frac{1}{N_{i,j}^{(t)} \to \varepsilon} \). Let \( \tilde{L}_{i,j} = \min \{ \ell > \tilde{L} : N_{i,j}^{(t)} = N_{i,j}^{(t)} + 1 \} \forall (i, j) \in A \), the first time \((i, j)\) is sampled after \( \tilde{L} \). Let \( L = \max_{\tilde{L}_{i,j} \in A, \tilde{L}_{i,j} \to \varepsilon} \forall (i, j) \in A \), then for any \( \ell > L \), let \( D_{1} = \{(i, j) : N_{i,j}^{(t)} < N_{i,j}^{(t)} \} \). Then if \( (i, j) \in D_{1} \cap A \), let \( L_{i,j} = \max \{ s < \ell : N_{i,j}^{(t)} = N_{i,j}^{(t)} - 1 \} \). Then we have
As a result, where the first inequality follows from $N^{(t)}_{i,j} = N^{(t)}_{i,j+1}$ and $N^{(t)}_{i,j+1} < N^{(t)}$ by the definition of $L_{i,j}$. Since $(i, j)$ is sampled at $L_{i,j}$, we must have $E_1 \geq 0$. Further since $\ell, L_{i,j} \geq \tilde{L}$, we have $E_2 \geq -|\alpha^*_i - \hat{\alpha}^{(t)}_{i,j} - |\alpha^*_i - \hat{\alpha}^{(t)}_{i,j} | > -2\varepsilon$ and $E_3 = N^{(t)}_{i,j} - N^{(t)}_{i,j+1} = N^{(t)}_{i,j} - N^{(t)}_{i,j+1} - N^{(t)}_{i,j} - N^{(t)}_{i,j+1} > -\frac{N^{(t)}_{i,j} - N^{(t)}_{i,j+1}}{N^{(t)}_{i,j} + 1} = -1 > -\varepsilon$.

Hence, we have $\hat{\alpha}^{(t)}_{i,j} - \frac{N^{(t)}_{i,j}}{N^{(t)}} \geq -3\varepsilon$ for $(i, j) \in D_t \cap A$. For $(i, j) \in D_t \setminus A$, we also have $\hat{\alpha}^{(t)}_{i,j} - \frac{N^{(t)}_{i,j}}{N^{(t)}} \geq -\varepsilon > -3\varepsilon$. This implies for all $(i, j) \in D_t$, $\hat{\alpha}^{(t)}_{i,j} - \frac{N^{(t)}_{i,j}}{N^{(t)}} \geq -3\varepsilon$.

As a result, $0 = \sum_{(i, j) \in D_t} \hat{\alpha}^{(t)}_{i,j} - \sum_{(i, j) \in D_t} \frac{N^{(t)}_{i,j}}{N^{(t)}} = \sum_{(i, j) \in D_t} \left( \hat{\alpha}^{(t)}_{i,j} - \frac{N^{(t)}_{i,j}}{N^{(t)}} \right) + \sum_{(i, j) \in D_t} \left( \hat{\alpha}^{(t)}_{i,j} - \frac{N^{(t)}_{i,j}}{N^{(t)}} \right) \leq 3|D_t|\varepsilon \leq 3KD\varepsilon$. By arbitrary $\varepsilon > 0$, we get $\lim_{\varepsilon \to \infty} \max_{i,j} \left( \hat{\alpha}^{(t)}_{i,j} - \frac{N^{(t)}_{i,j}}{N^{(t)}} \right) = 0$. Since for any $(i_0, j_0)$ we have

$$\max_{i,j} \left( \hat{\alpha}^{(t)}_{i,j} - \frac{N^{(t)}_{i,j}}{N^{(t)}} \right) \geq \alpha^*_{i_0,j_0} - \frac{N^{(t)}_{i_0,j_0}}{N^{(t)}} = -\sum_{i \neq i_0, j \neq j_0} \left( \hat{\alpha}^{(t)}_{i,j} - \frac{N^{(t)}_{i,j}}{N^{(t)}} \right) \geq -(KB - 1) \max_{i,j} \left( \hat{\alpha}^{(t)}_{i,j} - \frac{N^{(t)}_{i,j}}{N^{(t)}} \right),$$

we obtain $\lim_{\varepsilon \to \infty} \hat{\alpha}^{(t)}_{i_0,j_0} - \frac{N^{(t)}_{i_0,j_0}}{N^{(t)}} = 0 = \alpha^*_{i_0,j_0} - \lim_{\varepsilon \to \infty} \frac{N^{(t)}_{i_0,j_0}}{N^{(t)}}$ as desired. The proof is complete.

### .5 Proof of Lemma 2

**Proof.** Recall $M(t)$ is the total number of input data up to stage $t$. We have

$$\pi_{j}^{t} = \frac{\pi_{0}(\theta_{j})^{t} \prod_{\ell=1}^{M(t)} f_{\theta, \varepsilon}(\xi_{\ell})}{\sum_{j=D^{t}}^{D} \pi_{0}(\theta_{j})^{t} \prod_{\ell=1}^{M(t)} f_{\theta, \varepsilon}(\xi_{\ell})}.$$

Take the inverse of both sides we obtain

$$\frac{1}{\pi_{j}^{t}} = \frac{\sum_{j=1}^{D} \pi_{0}(\theta_{j})^{t} \prod_{\ell=1}^{M(t)} f_{\theta, \varepsilon}(\xi_{\ell})}{\pi_{0}(\theta_{j})^{t} \prod_{\ell=1}^{M(t)} f_{\theta, \varepsilon}(\xi_{\ell})} = 1 + \sum_{j \neq j} \frac{\pi_{0}(\theta_{j})^{t} \prod_{\ell=1}^{M(t)} f_{\theta, \varepsilon}(\xi_{\ell})}{\pi_{0}(\theta_{j})^{t} \prod_{\ell=1}^{M(t)} f_{\theta, \varepsilon}(\xi_{\ell})}.$$

We first prove $\prod_{\ell=1}^{M(t)} f_{\theta, \varepsilon}(\xi_{\ell}) / f_{\theta, \varepsilon}(\xi_{\ell}) = O(e^{-c_{1}})$ for some $c > 0$. To see this, take log on both sides.

$$\log \left( \prod_{\ell=1}^{M(t)} f_{\theta, \varepsilon}(\xi_{\ell}) / f_{\theta, \varepsilon}(\xi_{\ell}) \right) = \sum_{\ell=1}^{M(t)} \log \left( f_{\theta, \varepsilon}(\xi_{\ell}) / f_{\theta, \varepsilon}(\xi_{\ell}) \right) = M(t) \frac{1}{M(t)} \sum_{\ell=1}^{M(t)} \log \left( f_{\theta, \varepsilon}(\xi_{\ell}) / f_{\theta, \varepsilon}(\xi_{\ell}) \right).$$

By Assumption 8.1, we obtain almost surely, $\frac{1}{M(t)} \sum_{\ell=1}^{M(t)} \log \left( f_{\theta, \varepsilon}(\xi_{\ell}) / f_{\theta, \varepsilon}(\xi_{\ell}) \right) \to -KL(\theta^c || \theta_j) < 0$. Hence, there exists $0 < c_1 < KL(\theta^c || \theta_j)$, such that for all sufficiently large $t$,

$$\log \left( \prod_{\ell=1}^{M(t)} f_{\theta, \varepsilon}(\xi_{\ell}) / f_{\theta, \varepsilon}(\xi_{\ell}) \right) \leq -c_1 M(t).$$

Hence we obtain for all sufficiently large $t$,

$$\prod_{\ell=1}^{M(t)} f_{\theta, \varepsilon}(\xi_{\ell}) \leq e^{-c_1 M(t)}.$$
Since $\frac{M(t)}{1} \to \bar{n}$ by Assumption 8, $\frac{n_0(\theta_j)}{\pi_0(\theta_j)} > 0$ is a constant and the fact that there are only $D_1$ different $j$s such that $j \neq j^c$. We can find $\kappa > 0$, such that
\[
\sum_{j \neq j^c} \frac{n_0(\theta_j)}{\pi_0(\theta_j^c)} \prod_{\ell=1}^{M(t)} \frac{f_{\theta_j}(\xi_\ell)}{f_{\theta_j^c}(\xi_\ell)} = O\left(e^{-\kappa t}\right).
\]
Hence
\[
\pi_{j^c}^t = \frac{1}{1 + O\left(e^{-\kappa t}\right)} = 1 - O\left(e^{-\kappa t}\right).
\]
The proof is complete.

.6 Proof of Theorem 4

Several lemmas are needed to complete the proof. First, Lemma 4 ensures that the design-input pairs under an incorrect input parameter will only be sampled finitely many times. \[8\] ensures that a positive ratio of the budget will be assigned to all design-input pairs under correct input parameter. In the proof of the following, a sample path of the simulation process is fixed. Furthermore, since the consistency hold, we use $b \to b^c$ almost surely hence, we drop the superscript $c$ in the following.

Lemma 4

design-input pair $(i, j), \forall i \in I, j \neq j^c$ will only be simulated finitely many times.

Proof By input balance conditions, we have for any $j \neq j^c, i \in I$ and $\ell$ sufficiently large, almost surely
\[
\frac{\alpha_{i,j}^{(\ell)}}{\sigma_{i,j}^{(\ell)}} \geq \frac{n_0 N^{(\ell)}}{\pi_i^{(t)} \pi_j^{(t)}} \geq \frac{n_0 \bar{n}}{2 \sigma_{i,j}^{(t)}} \times O\left(\frac{1}{t_\ell \pi_i^{(t)}}\right) \to \infty,
\]
where the second inequality is because $\frac{N^{(\ell)}}{\pi_i^{(t)}} \to \bar{n}$ by Assumption 8, which implies for large $\ell$ (or $t_\ell$), $\frac{N^{(\ell)}}{\pi_i^{(t)}} \geq \frac{n_0}{2} t_\ell$. At the same time, for the same design $i$ and sufficiently large $\ell$,
\[
\frac{\alpha_{i,j^c}^{(\ell)}}{\sigma_{i,j^c}^{(\ell)}} \leq \frac{4}{\sigma_{i,j^c}} < \infty,
\]
where the inequality holds since $\alpha^{(\ell)} \leq 1, \pi_j^{(t)} = 1 - O\left(e^{-\kappa t}\right) \geq \frac{1}{2}$ and $\sigma_{i,j} \leq 2 \sigma_{i,j^c}$ as $(i, j^c)$ will be simulated infinitely many times by Theorem 2. This implies that the design-input pair $(i, j)$ with $j \neq j^c$ will only be simulated finitely many times.

Lemma 5 (i) $(i, j^c), i \neq b^c$ is simulated at iteration $\ell$ implies $\left(\frac{\alpha_{b,j^c}^{(\ell)}}{\sigma_{b,j^c}^{(\ell)}}\right)^2 - \sum_{i \neq b} \left(\frac{\alpha_{i,j^c}^{(\ell)}}{\sigma_{i,j^c}^{(\ell)}}\right)^2 \geq -O\left(\frac{1}{t_\ell}\right)$. (ii) Conversely, $(b, j^c)$ is simulated at iteration $\ell$ implies $\left(\frac{\alpha_{b,j^c}^{(\ell)}}{\sigma_{b,j^c}^{(\ell)}}\right)^2 - \sum_{i \neq b} \left(\frac{\alpha_{i,j^c}^{(\ell)}}{\sigma_{i,j^c}^{(\ell)}}\right)^2 \leq -O\left(\frac{1}{t_\ell}\right)$.

Proof By Lemma 4, we know $(i, j) i \in I, j \neq j^c$ will only be simulated finitely many times. Suppose after $\ell_0$, no $(i, j), j \neq j^c$ will be simulated. Then we have
\[
\left|\left(\frac{\alpha_{b,j^c}^{(\ell)}}{\sigma_{b,j^c}^{(\ell)}}\right)^2 - \sum_{i \neq b} \left(\frac{\alpha_{i,j^c}^{(\ell)}}{\sigma_{i,j^c}^{(\ell)}}\right)^2\right| = \frac{1}{(N^{(\ell)})^2} \left|\left(\frac{N_{b,j^c}^{(t)}}{\sigma_{b,j^c}^{(t)}}\right)^2 - \sum_{i \neq b} \left(\frac{N_{i,j^c}^{(t)}}{\sigma_{i,j^c}^{(t)}}\right)^2\right| = O\left(\frac{1}{t_\ell}\right).
\]
Hence, if $(i, j^c)$ is simulated at $\ell$ and $j^* = \arg \max_j \left|\left(\frac{\alpha_{b,j^c}^{(\ell)}}{\sigma_{b,j^c}^{(\ell)}}\right)^2 - \sum_{i \neq b} \left(\frac{\alpha_{i,j^c}^{(\ell)}}{\sigma_{i,j^c}^{(\ell)}}\right)^2\right| \neq j^c$, we know
\[
\left(\frac{\alpha_{b,j^c}^{(\ell)}}{\sigma_{b,j^c}^{(\ell)}}\right)^2 - \sum_{i \neq b} \left(\frac{\alpha_{i,j^c}^{(\ell)}}{\sigma_{i,j^c}^{(\ell)}}\right)^2 \geq \left(\frac{\alpha_{b,j^c}^{(\ell)}}{\sigma_{b,j^c}^{(\ell)}}\right)^2 - \sum_{i \neq b} \left(\frac{\alpha_{i,j^c}^{(\ell)}}{\sigma_{i,j^c}^{(\ell)}}\right)^2 \geq -O\left(\frac{1}{t_\ell}\right).
\]
Otherwise if $j^* = j^c$, we know $\left(\frac{\alpha_{b,j^c}^{(\ell)}}{\sigma_{b,j^c}^{(\ell)}}\right)^2 - \sum_{i \neq b} \left(\frac{\alpha_{i,j^c}^{(\ell)}}{\sigma_{i,j^c}^{(\ell)}}\right)^2 \geq 0$. Hence we prove (i). (ii) can be proved in a similar manner.
Lemma 6

\[ \liminf_{\ell \to \infty} \frac{\alpha_{i,j}^{(\ell)}}{\alpha_{i',j'}^{(\ell)}} > 0, \forall i \neq i', \text{ almost surely.} \]

Proof: Prove by contradiction. Suppose there exists \( i, i' \) such that \( \liminf_{\ell \to \infty} \frac{\alpha_{i,j}^{(\ell)}}{\alpha_{i',j'}^{(\ell)}} = 0 \). For any positive constant \( \varepsilon > 0 \), we can find a sufficiently large \( \ell \) such that \((i', j')\) is sampled at \( \ell \) and \( \frac{\alpha_{i,j}^{(\ell)}}{\alpha_{i',j'}^{(\ell)}} \leq \varepsilon \). Since \( \hat{\mu}_b^{(\ell)}, \hat{\mu}_i^{(\ell)} \) and \( \hat{\mu}_c^{(\ell)} \) all will converge to the true value a.s. and \( \hat{b}(\ell) = b \) for \( \ell \) sufficiently large. Then there exists constants \( a, \varepsilon \) and \( U > L > 0 \), such that for \( \ell \) sufficiently large, \( 0 < \frac{1}{2} < \pi_{i,j}^* < 1 \), \( 0 < r \leq \hat{\sigma}_{i,j}^{(\ell)} \leq h \) and \( U > (\hat{\mu}_b^{(\ell)} - \hat{\mu}_i^{(\ell)})^2 \) and \( L < (\hat{\mu}_b^{(\ell)} - \hat{\mu}_i^{(\ell)})^2 \).

Then

\[
\frac{(\hat{\mu}_b^{(\ell)} - \hat{\mu}_i^{(\ell)})^2}{\sum_{j=1}^{D} \frac{(\hat{\sigma}_{i,j}^{(\ell)})^2}{\alpha_{k,j}^{(\ell)}} + \sum_{j=1}^{D} \frac{(\hat{\sigma}_{i',j'}^{(\ell)})^2}{\alpha_{i',j'}^{(\ell)}}} < \frac{(\hat{\mu}_b^{(\ell)} - \hat{\mu}_i^{(\ell)})^2}{L} \leq \frac{(\hat{\mu}_b^{(\ell)} - \hat{\mu}_i^{(\ell)})^2}{\sum_{j=1}^{D} \frac{(\hat{\sigma}_{i,j}^{(\ell)})^2}{\alpha_{k,j}^{(\ell)}} + \sum_{j=1}^{D} \frac{(\hat{\sigma}_{i',j'}^{(\ell)})^2}{\alpha_{i',j'}^{(\ell)}}} \leq \frac{(\hat{\mu}_b^{(\ell)} - \hat{\mu}_i^{(\ell)})^2}{\sum_{j=1}^{D} \frac{(\hat{\sigma}_{i,j}^{(\ell)})^2}{\alpha_{k,j}^{(\ell)}} + \sum_{j=1}^{D} \frac{(\hat{\sigma}_{i',j'}^{(\ell)})^2}{\alpha_{i',j'}^{(\ell)}}}
\]

(26)

By proof of Lemma 4, we know \( \pi_{i,j}^{*} \to 0 \) for \( j \neq j' \). Hence, for \( \ell \) sufficiently large, we have \( \frac{(\hat{\sigma}_{i,j}^{(\ell)})^2}{\alpha_{i,j}^{(\ell)}} = O \left( \frac{1}{\ell} \right) \) for \( j \neq j' \) by Lemma 2. Then,

\[
\frac{(\hat{\mu}_b^{(\ell)} - \hat{\mu}_i^{(\ell)})^2}{\sum_{j=1}^{D} \frac{(\hat{\sigma}_{i,j}^{(\ell)})^2}{\alpha_{i,j}^{(\ell)}} + \sum_{j=1}^{D} \frac{(\hat{\sigma}_{i',j'}^{(\ell)})^2}{\alpha_{i',j'}^{(\ell)}}} \leq \frac{(\hat{\mu}_b^{(\ell)} - \hat{\mu}_i^{(\ell)})^2}{\sum_{j=1}^{D} \frac{(\hat{\sigma}_{i,j}^{(\ell)})^2}{\alpha_{i,j}^{(\ell)}} + \sum_{j=1}^{D} \frac{(\hat{\sigma}_{i',j'}^{(\ell)})^2}{\alpha_{i',j'}^{(\ell)}}} \leq \frac{(\hat{\mu}_b^{(\ell)} - \hat{\mu}_i^{(\ell)})^2}{\sum_{j=1}^{D} \frac{(\hat{\sigma}_{i,j}^{(\ell)})^2}{\alpha_{i,j}^{(\ell)}} + \sum_{j=1}^{D} \frac{(\hat{\sigma}_{i',j'}^{(\ell)})^2}{\alpha_{i',j'}^{(\ell)}}}
\]

(27)

holds because: \( i', j' \) is sampled at iteration \( \ell \), by Lemma 5 we have \( \left( \frac{\alpha_{i,j}^{(\ell)}}{\alpha_{i',j'}^{(\ell)}} \right)^2 \geq \frac{\alpha_{i,j}^{(\ell)}}{\alpha_{i',j'}^{(\ell)}} \geq O \left( \frac{1}{\ell} \right) \).

Hence, \( \frac{\alpha_{i,j}^{(\ell)}}{\alpha_{i',j'}^{(\ell)}} \geq O \left( \frac{1}{\ell} \right) \), \( \forall i \neq b \). Furthermore, since we also have \( \sum_{i \in \mathbb{Z}} \frac{\alpha_{i,j}^{(\ell)}}{\alpha_{i',j'}^{(\ell)}} \geq \sum_{i \in \mathbb{Z}} \frac{\alpha_{i,j}^{(\ell)}}{\alpha_{i',j'}^{(\ell)}} \geq \frac{1}{\kappa h} - O \left( \frac{1}{\ell} \right) \) and \( (i, j, j) \neq j' \) will only be sampled finitely many times. By pigeon hole principle we obtain \( \frac{\alpha_{b,j'}^{(\ell)}}{\alpha_{b,j}^{(\ell)}} \geq \frac{1}{\kappa h} - O \left( \frac{1}{\ell} \right) \).

\[
\frac{\alpha_{b,j'}^{(\ell)}}{\alpha_{b,j}^{(\ell)}} = O \left( \frac{1}{\ell} \right) \]

27
From (28), if we choose $\varepsilon$ that makes (28) $< 0$. We obtain 

$$
\frac{(\hat{\mu}_b^{(t)} - \hat{\mu}_i^{(t)})^2}{\sum_{j=1}^{D} \frac{(\sigma_{b,j}^{(t)})^2}{\alpha_{b,j}^{(t)}} + \sum_{j=1}^{D} \frac{(\sigma_{i,j}^{(t)})^2}{\alpha_{i,j}^{(t)}}} < 0,
$$

which implies $(i', j^c)$ cannot be sampled at $\ell$, a contradiction.

Lemma 7 (i) $\liminf_{\ell \to \infty} \frac{\alpha_{i,j}^{(t)} c^{(t)}}{\alpha_{i,j}^{(t)}} > 0 \forall i \neq b$ almost surely; (ii) $\liminf_{\ell \to \infty} \frac{\alpha_{i,j}^{(t)} c^{(t)}}{\alpha_{i,j}^{(t)}} > 0, \forall i \neq b$ almost surely.

Proof of (i) Prove by contradiction. By Lemma 6, there exists a positive constant $c > 0$ such that $\frac{\alpha_{i,j}^{(t)} c^{(t)}}{\alpha_{i,j}^{(t)}} \geq c$ for all $k \neq b$ and $\ell$ sufficiently large. Hence, if $\liminf_{\ell \to \infty} \frac{\alpha_{i,j}^{(t)} c^{(t)}}{\alpha_{i,j}^{(t)}} = 0$ for some $i$, then it holds for all $i$. For $\varepsilon < \frac{\sigma}{2K}$, we have there exists $\ell$ sufficiently large, $(i, j^c)$ is simulated at $\ell$ for some $i \neq b$ and $\alpha_{i,j}^{(t)} c^{(t)} < \varepsilon$. Furthermore since $(i, j)$ can only be sampled finitely many times and $1 = \sum_{j=1}^{D} \alpha_{i,j}^{(t)}$. To find $i_0 \neq b$, $\alpha_{i_0,j^c} c^{(t)} \geq 1$. Furthermore, by consistency result, for $\ell$ sufficiently large we have $0 < \sigma \leq \sigma_{i,j}^{(t)} \leq \sigma \forall i \in I$. Then we have 

$$
\frac{(\hat{\alpha}_{b,j}^{(t)} c^{(t)})^2}{\sigma_{b,j}^{(t)}} - \sum_{k \neq b} (\hat{\alpha}_{b,j}^{(t)} c^{(t)})^2 \leq (\hat{\alpha}_{b,j}^{(t)} c^{(t)})^2 - (\hat{\alpha}_{i,j}^{(t)} c^{(t)})^2 \leq \frac{\varepsilon^2}{\sigma^2} - \frac{1}{2K^2 \sigma^2} < \frac{1}{2K^2 \sigma^2}.
$$

This contradicts $(i, j^c)$ is simulated at $\ell$. Hence, we prove (i).

Proof of (ii) (ii) can be proved in a similar way.

Lemma 8 $\limsup_{\ell \to \infty} \alpha_{i,j}^{(t)} c^{(t)} > 0, \forall i, j$ almost surely.

Proof This is a direct result of Lemma 4, 6 and 7. By Lemma 6 for $\ell$ sufficiently large, $\sum_{j=1}^{D} \alpha_{i,j}^{(t)} c^{(t)} \geq 1/2$. By Lemma 6 and 7, $\forall i_0 \in I$, there exists $c > 0$, such that for $\ell$ sufficiently large, $\alpha_{i_0,j} c^{(t)} \geq c\alpha_{i,j} c^{(t)}, i \neq i_0$. Then 

$$
2\alpha_{i_0,j} c^{(t)} \geq \sum_{i,j} \alpha_{i,j} c^{(t)} \geq \frac{\sigma}{R} > 0.
$$

The proof is complete.

Lemma 10 and 11 guarantee the convergence rate of the estimated expected performance and estimated variance, respectively.

Lemma 9 (Inter-changeable $O(\cdot)$ notation). Suppose Assumptions in Theorem 4 hold true. Given a function $f(\ell)$, the following statements are equivalent almost surely:

1. $f(\ell) = O(\sqrt{\log \log \ell})$;
2. $f(\ell) = O(\sqrt{\log \log \ell})$;
3. $f(\ell) = O(\sqrt{\log \log N_{i,j}^{(t)}})$ for some $(i,j)$.

Proof. 1 and 2 are equivalent since by Assumption 8, we know for $\ell$ large enough, $\frac{1}{2} \bar{n} t \leq \ell \leq 2 \bar{n} t$. 1 implies 3 since $N_{i,j}^{(t)} \leq N^{(t)} = \ell + K D_{t_0}$ and for large $\ell$, $\sqrt{\frac{\log \log \ell}{\ell}}$ decreases in $\ell$. To show 3 implies 1, by Lemma 8 we know almost surely, there exists $C > 0, \alpha_{i,j}^{(t)} c^{(t)} \geq C$ for $\ell$ large, which implies $N_{i,j}^{(t)} \geq C N^{(t)} \geq C \ell$. This completes the proof. We will arbitrarily use the three $O(\cdot)$ notations in the following proofs.

Lemma 10 $|\hat{\mu}_i^{(t)} - \mu_i| = O(\sqrt{\log \log \ell})$ almost surely.
Proof Notice that \( \hat{\mu}_{i,j}^{(t)} - \mu_{i,j} = (\pi_{i,j}^{(t)} \hat{\mu}_{i,j}^{(t)} - \mu_{i,j}^{(t)}) + \sum_{j \neq j^*} \pi_{j,j}^{(t)} \hat{\mu}_{i,j}^{(t)} \). For the first term, By LIL, \( |\hat{\mu}_{i,j}^{(t)} - \mu_{i,j}| = O\left(\sqrt{\frac{\log \log N_{i,j}^{(t)}}{N_{i,j}^{(t)}}}\right) = O\left(\sqrt{\frac{\log \log \ell}{\ell}}\right) \). Furthermore, by Lemma 2 \( |\pi_{j,j}^{(t)} - 1| = O\left(\exp(-\kappa \ell t)\right) \) = \( O\left(\sqrt{\frac{\log \log \ell}{\ell}}\right) \). Hence, \( |\pi_{j,j}^{(t)} \hat{\mu}_{i,j}^{(t)} - \mu_{i,j}| = |\pi_{j,j}^{(t)} (\hat{\mu}_{i,j}^{(t)} - \mu_{i,j}) + (\pi_{j,j}^{(t)} - 1) \mu_{i,j}| \leq |\pi_{j,j}^{(t)}| |\hat{\mu}_{i,j}^{(t)} - \mu_{i,j}| + |\pi_{j,j}^{(t)} - 1| |\mu_{i,j}| = O\left(\sqrt{\frac{\log \log \ell}{\ell}}\right) \).

For the second term, \( \pi_{j,j}^{(t)} \hat{\mu}_{i,j}^{(t)} = \hat{\mu}_{i,j}^{(t)} \cdot O\left(\exp(-\kappa \ell t)\right) = O\left(\sqrt{\frac{\log \log \ell}{\ell}}\right) \).

**Lemma 11** \( |(\hat{\sigma}_{i,j}^{(t)})^2 - (\sigma_{i,j})^2| = O\left(\sqrt{\frac{\log \log \ell}{\ell}}\right) \) almost surely. As a result, \( |\hat{\sigma}_{i,j}^{(t)} - \sigma_{i,j}| = O\left(\sqrt{\frac{\log \log \ell}{\ell}}\right) \) almost surely.

**Proof** Since

\[ \sum_{s=1}^{N_{i,j}^{(t)}} (X_{i,j}^{(s)} - \hat{\mu}_{i,j}^{(t)})^2 = \sum_{s=1}^{N_{i,j}^{(t)}} (X_{i,j}^{(s)} - \mu_{i,j}^{(t)})^2 - N_{i,j}^{(t)} (\hat{\mu}_{i,j}^{(t)} - \mu_{i,j}^{(t)})^2, \]

We have

\[ (\hat{\sigma}_{i,j}^{(t)})^2 - (\sigma_{i,j})^2 = \frac{1}{N_{i,j}^{(t)} - 1} \sum_{s=1}^{N_{i,j}^{(t)}} [(X_{i,j}^{(s)} - \mu_{i,j}^{(t)})^2 - \sigma_{i,j}^2] - \frac{N_{i,j}^{(t)}}{N_{i,j}^{(t)} - 1} (\hat{\mu}_{i,j}^{(t)} - \mu_{i,j}^{(t)})^2 + \frac{\sigma_{i,j}^2}{N_{i,j}^{(t)} - 1}. \]

Since \( (X_{i,j}^{(s)} - \mu_{i,j}^{(t)})^2 - \sigma_{i,j}^2 \) are i.i.d. with mean 0, by LIL, we have with probability 1,

\[ \left| \frac{1}{N_{i,j}^{(t)} - 1} \sum_{s=1}^{N_{i,j}^{(t)}} [(X_{i,j}^{(s)} - \mu_{i,j}^{(t)})^2 - \sigma_{i,j}^2] \right| = O\left(\sqrt{\frac{\log \log N_{i,j}^{(t)}}{N_{i,j}^{(t)} - 1}}\right) = O\left(\sqrt{\frac{\log \log \ell}{\ell}}\right). \]

Further since \( |\hat{\mu}_{i,j}^{(t)} - \mu_{i,j}| = O\left(\sqrt{\frac{\log \log \ell}{\ell}}\right) \) and \( |\sigma_{i,j}^2| = O\left(\sqrt{\frac{\log \log \ell}{\ell}}\right) \), we get the desired result.

**Lemma 12** Let \((i, j^c)\) be a fixed design-input pair. Suppose \((i, j^c)\) is sampled at iteration \(r\). Let \( \ell_r = \inf\{\ell > 0 : 1^{(r+\ell)}(i, j^c) = 1\}\). Hence \( r + \ell_r \) is the next iteration \((i, j^c)\) will be sampled after \(r\). Then we have \( r < r + \ell_r = O(r) \) almost surely.

**Proof** Prove by contradiction. Suppose \( \forall C_0 > 0 \), there exists an iteration \(r\) such that \( t_r > C_0 r \). We have

\[ \alpha_{i,j^c}^{(r+\ell)} = \frac{N_{i,j^c}^{(r+\ell)}}{N_{i,j^c}^{(r+\ell)}} = \frac{N_{i,j^c}^{(r)}}{N_{i,j^c}^{(r+\ell)}} + \frac{1}{N_{i,j^c}^{(r+\ell)}} < \frac{2(n_0 + r)}{KBn_0 + (C_0 + 1)r} < \frac{3}{C_0} \]

for large \(r\). The first inequality holds since \( N(r) = KBn_0 + r \) and \( N_{i,j^c}^{(0)} = n_0 \). By the arbitrariness of \(C_0\) and the fact that if \( C_0 \to \infty \), the iteration \(r\) that satisfy \( \ell_r > C_0 r \) must also go to \( \infty \). We have \( \lim \inf_{\ell \to \infty} \alpha_{i,j^c}^{(r)} = 0 \), contradicting Lemma 8(ii).

**.7 Proof of Theorem 4.1.**

**Lemma 13** Suppose the assumptions in Theorem 2 hold true. Then we have

\[ |\hat{\alpha}_{i,j}^{(t)} - \alpha_{i,j}^{*}| = O\left(\sqrt{\frac{\log \log \ell}{\ell}}\right) \] almost surely.

Here \( \alpha_{i,j}^{*} = 0 \) for \( j \neq j^c \).
Proof. Notice both \( \alpha_{i,j}^* \) and \( \hat{\alpha}_{i,j}^{(t)} \), \( i \neq b \) can be computed explicitly as

\[
\alpha_{i,j}^* = \frac{w_{i,j}^*}{\sum_{i',j'} w_{i',j'}^*},
\]

where

\[
w_{i,j}^* = \pi_j^* \sigma_{i,j} \sum_{k=1}^D \sigma_i k \pi_k^*/(\mu_b - \mu_i)^2
\]

if \( i \neq b \) and

\[
w_{b,j}^* = \left( \sum_{i' \neq b} \frac{\sigma_{b,i'}^2 w_{i',j}^2}{\sigma_{i'}^2} \right)^{1/2},
\]

\( \pi_j^c = 0 \) for \( j \neq j^c \).

\[
\hat{\alpha}_{i,j}^{(t)} = \frac{\hat{w}_{i,j}^{(t)}}{\sum_{i',j'} \hat{w}_{i',j'}^{(t)}},
\]

where

\[
\hat{w}_{i,j}^{(t)} = \pi_j^t \sigma_{i,j}^t \sum_{k=1}^D \sigma_{i,k}^t \pi_k^t/ (\hat{\mu}_b - \hat{\mu}_i)^2,
\]

if \( i \neq b \) and

\[
\hat{w}_{b,j}^{(t)} = \left( \sum_{i' \neq b} \frac{(\hat{\sigma}_{b,i'}^{(t)})^2 (\hat{w}_{i',j}^{(t)})^2}{(\hat{\sigma}_{i'}^{(t)})^2} \right)^{1/2}.
\]

Since \( \pi_j^{t\ell} = O \left( \sqrt{\frac{\log \log \ell}{\ell}} \right) \), \( \hat{w}_{i,j}^{*} = w_{i,j}^* + O \left( \sqrt{\frac{\log \log \ell}{\ell}} \right) \). Hence, we also have \( \hat{\alpha}_{i,j}^{(t)} = \alpha_{i,j}^* + O \left( \sqrt{\frac{\log \log \ell}{\ell}} \right) \). ■

Proof of Theorem 4.1. We first prove

\[
\max_{i,j} \{\alpha_{i,j}^{(r)} - \hat{\alpha}_{i,j}^{(r)} \} = O \left( \sqrt{\frac{\log \log \ell}{\ell}} \right).
\]

Let \((i_\ell,j_\ell) = \arg \max_{i,j} \{\alpha_{i,j}^{(r)} - \hat{\alpha}_{i,j}^{(r)} \}\). Then since \( \sum_{i,j} \alpha^{(r)} = \sum_{i,j} \hat{\alpha}^{(r)} = 1 \), we know \( \alpha_{i_\ell,j_\ell}^{(r)} - \hat{\alpha}_{i_\ell,j_\ell}^{(r)} \geq 0 \) and \((i_\ell,j_\ell)\) cannot be simulated at \( \ell \). Then, let \( r \) be the last time \((i_\ell,j_\ell)\) is simulated, where we must have \( \alpha_{i_\ell,j_\ell}^{(r)} - \hat{\alpha}_{i_\ell,j_\ell}^{(r)} \leq 0 \). By Lemma 13, we have there exists \( C_1 > 0 \), such that for both \( r \) and \( \ell \) sufficiently large, \( |\alpha_{i_\ell,j_\ell}^{(r)} - \alpha_{i_\ell,j_\ell}^*| = O \left( \sqrt{\frac{\log \log \ell}{\ell}} \right) \) and \( |\hat{\alpha}_{i_\ell,j_\ell}^{(r)} - \hat{\alpha}_{i_\ell,j_\ell}^*| = O \left( \sqrt{\frac{\log \log \ell}{\ell}} \right) \). Furthermore by Lemma 12, we know \( \ell = O(r) \) and hence \( \sqrt{\frac{\log \log \ell}{\ell}} = O \left( \sqrt{\frac{\log \log r}{r}} \right) \). This implies \( |\alpha_{i_\ell,j_\ell}^{(r)} - \alpha_{i_\ell,j_\ell}^*| = O \left( \sqrt{\frac{\log \log \ell}{\ell}} \right) \) and \( |\hat{\alpha}_{i_\ell,j_\ell}^{(r)} - \hat{\alpha}_{i_\ell,j_\ell}^*| = O \left( \sqrt{\frac{\log \log \ell}{\ell}} \right) \). Hence, there exists \( C_2 > 0 \),

\[
0 \leq \alpha_{i_\ell,j_\ell}^{(r)} - \hat{\alpha}_{i_\ell,j_\ell}^{(r)} \leq \alpha_{i_\ell,j_\ell}^{(r)} + \frac{1}{N^{(r)}} - \alpha_{i_\ell,j_\ell}^* + C_1 \sqrt{\frac{\log \log \ell}{\ell}} \leq \frac{1}{N^{(r)}} + C_1 \sqrt{\frac{\log \log \ell}{\ell}} = O \left( \sqrt{\frac{\log \log \ell}{\ell}} \right),
\]

where \( \alpha_{i_\ell,j_\ell}^{(r)} - \alpha_{i_\ell,j_\ell}^* = \frac{N_{i_\ell,j_\ell}^{(r)} + 1}{N^{(r)}} - \frac{N_{i_\ell,j_\ell}^*}{N^{(r)}} \leq \frac{N_{i_\ell,j_\ell}^{(r)} + 1}{N^{(r)}} - \frac{N_{i_\ell,j_\ell}^*}{N^{(r)}} = \frac{1}{N^{(r)}} \).

Second, we prove \( \forall (i,j), |\alpha_{i,j}^{(r)} - \hat{\alpha}_{i,j}^{(r)}| = O \left( \sqrt{\frac{\log \log \ell}{\ell}} \right) \). Let \( I_1 = \{(i,j), \alpha_{i,j}^{(r)} - \hat{\alpha}_{i,j}^{(r)} \geq 0 \} \) and \( I_2 = \{(i,j), \alpha_{i,j}^{(r)} - \hat{\alpha}_{i,j}^{(r)} \leq 0 \} \). We have if \((i,j) \in I_1, 0 \leq \alpha_{i,j}^{(r)} - \hat{\alpha}_{i,j}^{(r)} \leq \alpha_{i,j}^{(r)} - \alpha_{i,j}^* = O \left( \sqrt{\frac{\log \log \ell}{\ell}} \right) \). Else if \((i,j) \in I_2, \)
Third, noticing \( |\hat{\alpha}_{i,j}^{(t)} - \alpha_{i,j}^{*}| = O \left( \sqrt{\frac{\log \log \ell}{\ell}} \right) \), we obtain
\[
|\alpha_{i,j}^{(t)} - \alpha_{i,j}^{*}| \leq |\alpha_{i,j}^{(t)} - \hat{\alpha}_{i,j}^{(t)}| + |\hat{\alpha}_{i,j}^{(t)} - \alpha_{i,j}^{*}| = O \left( \sqrt{\frac{\log \log \ell}{\ell}} \right) + O \left( \sqrt{\frac{\log \log \ell}{\ell}} \right) = O \left( \sqrt{\frac{\log \log \ell}{\ell}} \right).
\]
This completes the proof. \( \blacksquare \)

### 8 Proof of Theorem 12

We need some more lemmas to prove the result.

The following Lemma bounds the amount of budget allocated to a non-optimal design-input pair \((i, j^c)\) between two successive samples of the best design-input pair \((b, j^c)\) under the same input realization.

**Lemma 14** Suppose \((b, j^c)\) is sampled at iteration \(r\). Let \(\ell_r = \inf \{ \ell > 0 : 1/(r+\ell) = 1 \} \). \(r + \ell_r\) is the next iteration at which \((b, j^c)\) is sampled. Then between the two samples of \((b, j^c)\), the number of samples that can be allocated to \((i, j^c)\), \(i \neq b\) is at most \(O(\sqrt{\log \log r})\) almost surely.

**Proof** Fix a non-optimal design \(i\). Let \(s_r = \sup \{ \ell < \ell_r : 1/(r+\ell) = 1 \} \). \(r + s_r\) is the last time before \(r + \ell_r\) at which \((i, j^c)\) is sampled. If \(s_r < 0\), then the lemma holds true, otherwise, assume \(s_r > 0\). Since \((b, j^c)\) is sampled at \(r\), then
\[
\left( \frac{N_{b,j^c}^{(r)}}{\hat{\sigma}_{b,j^c}^{(r)}} \right)^2 \leq \sum_{k \neq b} \left( \frac{N_{k,j^c}^{(r)}}{\hat{\sigma}_{k,j^c}^{(r)}} \right)^2 + O(1) \leq \sum_{k \neq b} \left( \frac{N_{k,j^c}^{(r)}}{\hat{\sigma}_{k,j^c}^{(r)}} \right)^2 + O \left( \sqrt{\log \log r} \right),
\]
by Lemma 3. Similarly, since \((i, j^c)\) is sampled at \(r + s_r\),
\[
\left( \frac{N_{b,j^c}^{(r+s_r)}}{\hat{\sigma}_{b,j^c}^{(r+s_r)}} \right)^2 \geq \sum_{k \neq b} \left( \frac{N_{k,j^c}^{(r+s_r)}}{\hat{\sigma}_{k,j^c}^{(r+s_r)}} \right)^2 - O \left( \sqrt{\log \log r} \right),
\]
where \(r + s_r = O(r)\) by Lemma 12. Furthermore notice \(\hat{\sigma}_{i,j^c}^{(r)} = \hat{\sigma}_{i,j^c}^{(r+s_r)} + O \left( \frac{\log \log r}{r} \right)\). Then
\[
0 \leq \left( \frac{N_{k,j^c}^{(r+s_r)}}{\hat{\sigma}_{k,j^c}^{(r+s_r)}} \right)^2 - \sum_{k \neq b} \left( \frac{N_{k,j^c}^{(r+s_r)}}{\hat{\sigma}_{k,j^c}^{(r+s_r)}} \right)^2 + O \left( \sqrt{\log \log r} \right)
= \left( \frac{N_{k,j^c}^{(r)}}{\hat{\sigma}_{b,j^c}^{(r)}} \right)^2 + \frac{1}{\hat{\sigma}_{b,j^c}^{(r)}}^2 - \sum_{k \neq b} \left( \frac{N_{k,j^c}^{(r+s_r)}}{\hat{\sigma}_{k,j^c}^{(r+s_r)}} \right)^2 + O \left( \sqrt{\log \log r} \right)
\leq \left( \frac{N_{b,j^c}^{(r)}}{\hat{\sigma}_{b,j^c}^{(r)}} \right)^2 + \frac{1}{\hat{\sigma}_{b,j^c}^{(r)}}^2 - \sum_{k \neq b} \left( \frac{N_{k,j^c}^{(r+s_r)}}{\hat{\sigma}_{k,j^c}^{(r+s_r)}} \right)^2 + O (r)
\leq \sum_{k \neq b} \left( \frac{N_{k,j^c}^{(r)}}{\hat{\sigma}_{k,j^c}^{(r)}} \right)^2 - \sum_{k \neq b} \left( \frac{N_{k,j^c}^{(r+s_r)}}{\hat{\sigma}_{k,j^c}^{(r+s_r)}} \right)^2 + O \left( r \sqrt{\log \log r} \right)
\leq \sum_{k \neq b} \left( \frac{N_{k,j^c}^{(r)}}{\hat{\sigma}_{k,j^c}^{(r)}} \right)^2 - \sum_{k \neq b} \left( \frac{N_{k,j^c}^{(r+)}}{\hat{\sigma}_{k,j^c}^{(r+)}} \right)^2 + O \left( r \sqrt{\log \log r} \right) \quad (29)
\]
Then, since for each \(k \neq b\),
\[
\left( \frac{N_{k,j^c}^{(r)}}{\hat{\sigma}_{k,j^c}^{(r)}} \right)^2 - \left( \frac{N_{k,j^c}^{(r+)}}{\hat{\sigma}_{k,j^c}^{(r+)}} \right)^2 \geq 0.
\]
We obtain
\[
\left( \frac{N_{k,j^c}^{(r)}}{\hat{\sigma}_{k,j^c}^{(r)}} \right)^2 - \left( \frac{N_{k,j^c}^{(r+)}}{\hat{\sigma}_{k,j^c}^{(r+)}} \right)^2 \leq O \left( r \sqrt{\log \log r} \right).
\]
Since
\[ O \left( r \sqrt{r \log \log r} \right) \geq (N_{i,j}^{(r+s_r)})^2 - (N_{i,j}^{(r)})^2 = (N_{i,j}^{(r+s_r)} - N_{i,j}^{(r)}) (N_{i,j}^{(r+s_r)} + N_{i,j}^{(r)}) \geq 2N_{i,j}^{(r)} (N_{i,j}^{(r+s_r)} - N_{i,j}^{(r)}) \]  \hfill (30)

By Lemma 8(ii), we have \( \frac{1}{N_{i,j}^{(r)}} O \left( r \sqrt{r \log \log r} \right) = O \left( \sqrt{r \log \log r} \right) \). This implies \( N_{i,j}^{(r+\ell)} - N_{i,j}^{(r)} = N_{i,j}^{(r+s_r)} - N_{i,j}^{(r)} = O(\sqrt{r \log \log r}) \).  

Conversely, the following Lemma 15 bounds the amount of budget allocated to the optimal design-input pair under the same input realization.

**Lemma 15** Suppose at iteration \( r \) a non-optimal design \( i_1 \) is sampled. Let \( \ell_r = \inf \{ \ell > 0 : \exists i \neq b, 1_{(i,j)}^{(r+\ell)} = 1 \} \). \( r + \ell_r \) is the next iteration at which a non-optimal design is sampled. Then between iteration \( r \) and \( r + \ell_r \), the number of samples that can be allocated to \( (b,j^*) \) is \( O(\sqrt{\log \log r}) \) almost surely.

**Proof** The proof is similar to Lemma 14. Define \( s_r = \sup \{ \ell < \ell_r : 1_{(b,j^*)}^{(r)} = 1 \} \). \( r + s_r \) is the last time before \( r + \ell_r \) the optimal design is sampled. If \( s_r < 0 \), then the lemma holds. Otherwise assume \( s_r > 0 \). Since \( (b,j^*) \) is sampled at \( r + s_r \),
\[
\left( \frac{N_{b,j^*}^{(r)}}{\sigma_{b,j^*}^{(r)}} \right)^2 \leq \sum_{k \neq b} \left( \frac{N_{b,j^*}^{(r+s)}}{\sigma_{b,j^*}^{(r+s)}} \right)^2 + O \left( \sqrt{\log \log r} \right).
\]
Since \( (i_1,j^*) \) is sampled at \( r \), we have
\[
\left( \frac{N_{b,j^*}^{(r)}}{\sigma_{b,j^*}^{(r)}} \right)^2 \geq \sum_{k \neq b} \left( \frac{N_{b,j^*}^{(r+s)}}{\sigma_{b,j^*}^{(r+s)}} \right)^2 - O \left( \sqrt{\log \log r} \right). \]
Further by Lemma 12, we have \( s_r < t_r = O(r) \). We have
\[
0 \geq \left( \frac{N_{b,j^*}^{(r+s)}}{\sigma_{b,j^*}^{(r+s)}} \right)^2 - \sum_{k \neq b} \left( \frac{N_{b,j^*}^{(r+s)}}{\sigma_{b,j^*}^{(r+s)}} \right)^2 - O \left( \sqrt{\log \log r} \right)
= \left( \frac{N_{b,j^*}^{(r+s)}}{\sigma_{b,j^*}^{(r+s)}} \right)^2 - \sum_{k \neq b, i_1=b} \left( \frac{N_{b,j^*}^{(r+s)}}{\sigma_{b,j^*}^{(r+s)}} \right)^2 - \left( \frac{N_{b,j^*}^{(r)}}{\sigma_{b,j^*}^{(r)}} \right)^2 - O \left( \sqrt{\log \log r} \right)
\geq \left( \frac{N_{b,j^*}^{(r+s)}}{\sigma_{b,j^*}^{(r+s)}} \right)^2 - \sum_{k \neq b, i_1=b} \left( \frac{N_{b,j^*}^{(r+s)}}{\sigma_{b,j^*}^{(r+s)}} \right)^2 - \left( \frac{N_{b,j^*}^{(r)}}{\sigma_{b,j^*}^{(r)}} \right)^2 - O \left( \sqrt{\log \log r} \right)
\geq \left( \frac{N_{b,j^*}^{(r+s)}}{\sigma_{b,j^*}^{(r+s)}} \right)^2 - \left( \frac{N_{b,j^*}^{(r)}}{\sigma_{b,j^*}^{(r)}} \right)^2 - O \left( r \sqrt{\log \log r} \right)
\]
With similar reason as in proof of Lemma 14, this implies
\[
N_{b,j^*}^{(r+s)} - N_{b,j^*}^{(r)} \leq O \left( \sqrt{\log \log r} \right)
\]
The proof is complete.

**Lemma 16** \( \left| \frac{\alpha_{b,j^*}^{(r)}}{\bar{\sigma}_{b,j^*}^{(r)}} \right|^2 - \sum_{i \neq b} \left| \frac{\alpha_{b,j^*}^{(r)}}{\bar{\sigma}_{b,j^*}^{(r)}} \right|^2 \right| = O \left( \sqrt{\frac{\log \log \ell}{\ell}} \right) \) a.s.

**Proof** First notice by Lemma 4 when \( \ell \) is large enough, only design-input pair under the true parameter will be simulated. Let \( |\Delta^{(\ell)}| := \left| \frac{\alpha_{b,j^*}^{(r)}}{\bar{\sigma}_{b,j^*}^{(r)}} \right|^2 - \sum_{i \neq b} \left| \frac{\alpha_{b,j^*}^{(r)}}{\bar{\sigma}_{b,j^*}^{(r)}} \right|^2 \right| = \left| \sum_{i \neq b} \left| \frac{\alpha_{b,j^*}^{(r)}}{\bar{\sigma}_{b,j^*}^{(r)}} \right|^2 \right| \. Fix an iteration.
Case 1. $\Delta(\ell) \leq 0$, let $\ell_* = \sup \{ r < \ell : \exists i \neq b, 1_{i,j^c}^{(r)} = 1 \}$. Then $\ell_*$ is the last time before $\ell$ that a non-optimal design, denote by $i_0$, is sampled under $\theta_j$. We have

$$0 \geq \left( \frac{N_{i,j}^{(\ell)}}{\hat{\sigma}_{i,j}^{(\ell)}} \right)^2 - \sum_{i \neq b} \left( \frac{N_{i,j}^{(\ell)}}{\hat{\sigma}_{i,j}^{(\ell)}} \right)^2$$

$$\geq \left( \frac{N_{i,j}^{(\ell)}}{\hat{\sigma}_{i,j}^{(\ell)}} \right)^2 - \sum_{i \neq i_0 \neq b} \left( \frac{N_{i,j}^{(\ell)}}{\hat{\sigma}_{i,j}^{(\ell)}} \right)^2 - \left( \frac{N_{i_0,j}^{(\ell)} + 1}{\hat{\sigma}_{i_0,j}^{(\ell)}} \right)^2$$

$$\geq \left( \frac{N_{i,j}^{(\ell)}}{\hat{\sigma}_{i,j}^{(\ell)}} \right)^2 - \sum_{i \neq i_0} \left( \frac{N_{i,j}^{(\ell)}}{\hat{\sigma}_{i,j}^{(\ell)}} \right)^2 - O(\ell)$$

Divide both sides by $N_i^{(\ell)}$ and notice $\ell = O(\ell_*)$, we have

$$0 > \left( \frac{\alpha_{b,j^c}}{\hat{\sigma}_{b,j^c}} \right)^2 - \sum_{i \neq i_0} \left( \frac{\alpha_{i,j}^{(\ell)}}{\hat{\sigma}_{i,j}^{(\ell)}} \right)^2 - O\left( \frac{1}{\ell} \right)$$

$$\geq \left( \frac{\alpha_{b,j^c}}{\hat{\sigma}_{b,j^c}} \right)^2 - \sum_{i \neq i_0} \left( \frac{\alpha_{i,j}^{(\ell)}}{\hat{\sigma}_{i,j}^{(\ell)}} \right)^2 - O\left( \frac{\log \log \ell}{\ell} \right).$$

Which completes the proof for case 1.

Case 2. $\Delta(\ell) \geq 0$. Let $\ell_* = \sup \{ r < \ell : 1_{(b,j^c)}^{(r)} = 1 \}$. $\ell_*$ is the last time before $\ell$ at which $(b, j^c)$ is sampled. By Lemma 12, $\ell = O(\ell_*)$ since $(b, j^c)$ is not sampled between $\ell_* \leq \ell$; by Lemma 14,

$$N_{i,j^c}^{(\ell)} - N_{i,j^c}^{(\ell_*)} = O\left( \sqrt{\ell \log \log \ell} \right), \forall i \neq b.$$ Then,

$$0 \leq \left( \frac{N_{b,j^c}^{(\ell)}}{\hat{\sigma}_{b,j^c}^{(\ell)}} \right)^2 - \sum_{i \neq i_0} \left( \frac{N_{i,j^c}^{(\ell)}}{\hat{\sigma}_{i,j^c}^{(\ell)}} \right)^2$$

$$\leq \left( \frac{N_{b,j^c}^{(\ell)} + 1}{\hat{\sigma}_{b,j^c}^{(\ell)}} \right)^2 - \sum_{i \neq i_0} \left( \frac{N_{i,j^c}^{(\ell)}}{\hat{\sigma}_{i,j^c}^{(\ell)}} \right)^2 + O\left( \ell \sqrt{\ell \log \log \ell} \right)$$

$$\leq \left( \frac{N_{b,j^c}^{(\ell)}}{\hat{\sigma}_{b,j^c}^{(\ell)}} \right)^2 - \sum_{i \neq i_0} \left( \frac{N_{i,j^c}^{(\ell)}}{\hat{\sigma}_{i,j^c}^{(\ell)}} \right)^2 + O\left( \ell \sqrt{\ell \log \log \ell} \right)$$

$$\leq O\left( \ell \sqrt{\ell \log \log \ell} \right).$$

Divide both sides by $N_i^{(\ell)}$, we obtain $0 \leq \Delta_{j^c}(\ell) \leq O\left( \sqrt{\frac{\log \log r}{\ell}} \right)$. The proof is complete.

Proof of (15) in Theorem 4.2.

Proof. By Lemma 11 and Lemma 16

$$\left| \frac{\alpha_{b,j^c}^{(r)}}{\hat{\sigma}_{b,j^c}^{(r)}} \right|^2 - \sum_{i \neq b} \left( \frac{\alpha_{i,j^c}^{(r)}}{\hat{\sigma}_{i,j^c}^{(r)}} \right)^2 = \left| \frac{\alpha_{b,j^c}^{(r)}}{\hat{\sigma}_{b,j^c}^{(r)}} \right|^2 - \sum_{i \neq b} \left( \frac{\alpha_{i,j^c}^{(r)}}{\hat{\sigma}_{i,j^c}^{(r)}} \right)^2 + O\left( \sqrt{\frac{\log \log r}{\ell}} \right)$$

$$= O\left( \sqrt{\frac{\log \log r}{\ell}} \right)$$
The next Lemma 17 is a little technical, which is used to bound the amount of budget allocated to a non-optimal design-input pair \((b, j^c)\) between two successive samples of a non-optimal design-input pair \((i, j^c)\), as shown in Lemma 18.

**Lemma 17** Under \(j^c\), suppose a non-optimal design \((k, j^c)\) is sampled at iteration \(r\). Define

\[
\ell_r := \inf \{ \ell > 0 : 1_{(k,j^c)}^{(r+\ell)} = 1 \}
\]

\[
s_r' = \sup \{ \ell < \ell_r : 1_{(k,j^c)}^{(r+\ell)} = 1 \}
\]

\[
s_r := \sup \{ \ell < s_r' : 1_{(i,j^c)}^{(\ell)} = 1 \text{ for some } i \neq b \}
\]

\[
a_{i,j^c}^{(r,s)} = N_{i,j^c}^{(r)} - N_{i,j^c}^{(s)}
\]

For all \(C_1 > 0\), if there exists \(C_2\) sufficiently large (depend on \(C_1\) but not on \(r\)), such that \(C_2 \sqrt{r \log \log r} \leq d_{b,j^c}^{(r,s)}\) holds for infinitely many \(r\)'s, then for such sufficiently large \(r\), there exists another sub-optimal design \(i \neq k \neq b\) and a \(u \leq s_r, i\) is sampled at \(r + u\) and

\[
1 + C_1 \sqrt{\frac{\log \log r}{r}} \frac{N_{i,j^c}^{(r)} N_{b,j^c}^{(r+u)}}{N_{i,j^c}^{(r+s)} N_{b,j^c}^{(r+s)}} \leq \frac{N_{i,j^c}^{(r+u)} N_{b,j^c}^{(r+s)}}{N_{i,j^c}^{(r+s)} N_{b,j^c}^{(r+s+1)}} \leq C_3 \sqrt{\frac{\log \log r}{r}}
\]

holds almost surely.

**Proof** By Lemma 12 \(t_r = O(r)\), which implies there exists \(C_0 > 0, d_{b,j^c}^{(r,s)} \leq C_0 r\). Hence, for any fixed \(C_1\), there exists \(C_2\) such that for infinitely many \(r\)'s \(C_2 \sqrt{r \log \log r} \leq d_{b,j^c}^{(r,s)} \leq C_0 r\). Let \(\Delta_j^{(r)} = \left( \frac{\alpha_{b,j^c}^{(r)}}{\delta_{b,j^c}^{(r)}} \right)^2 - \sum_{i \neq b} \left( \frac{\alpha_{i,j^c}^{(r)}}{\delta_{i,j^c}^{(r)}} \right)^2\).

By the definition of \(s_r\), \(\Delta_{j^c}^{(r,s)} \geq -O\left( \sqrt{\frac{\log \log r}{r}} \right)\) and \(\Delta_{j^c}^{(r,s+1)} = \Delta_{j^c}^{(r,s)} + O\left( \sqrt{\frac{\log \log r}{r}} \right)\). Since \((b, j^c)\) is sampled at \(r + s_r', \Delta_{j^c}^{(r+s')} \leq O\left( \sqrt{\frac{\log \log r}{r}} \right)\), then there exists \(C_3 > C_3' > 0\),

\[
\Delta_{j^c}^{(r+s+1)} \leq \Delta_{j^c}^{(r+s')} + C_3' \sqrt{\frac{\log \log r}{r}} \leq C_3 \sqrt{\frac{\log \log r}{r}}.
\]

Then one can choose \(C_4 > C_3\),

\[
|\Delta_{j^c}^{(r+s)}| \leq C_4 \sqrt{\frac{\log \log r}{r}}
\]

Since \(N_{k,j^c}^{(r+s)} - N_{k,j^c}^{(r)} = 1\), (32) implies

\[
\left( \frac{N_{i,j^c}^{(r+s)} / \delta_{i,j^c}^{(r+s)}}{N_{b,j^c}^{(r+s)}} \right)^2 - \sum_{i \neq b, k} \left( \frac{N_{i,j^c}^{(r+s)} / \delta_{i,j^c}^{(r+s)}}{N_{b,j^c}^{(r+s)}} \right)^2 \leq C_4 \sqrt{\frac{\log \log r}{r}}.
\]

By some simple algebraic calculation we get

\[
\sum_{i \neq b, k} \left( \frac{N_{i,j^c}^{(r+s)} / \delta_{i,j^c}^{(r+s)}}{N_{b,j^c}^{(r+s)}} \right)^2 + \left( \frac{N_{k,j^c}^{(r)} / \delta_{k,j^c}^{(r)}}{N_{b,j^c}^{(r+s)}} \right)^2 + C_4 \sqrt{\frac{\log \log r}{r}} \left( \frac{N_{b,j^c}^{(r+s)}}{N_{b,j^c}^{(r+s)}} \right)^2 \geq 1.
\]

By Lemma 7, there exists \(C_5 > 0, C_5 N_{b,j^c}^{(r)} / \delta_{b,j^c}^{(r)} > N_{i}^{(r)}\) for all large \(r\). Hence,

\[
\sum_{i \neq b, k} \left( \frac{N_{i,j^c}^{(r+s)} / \delta_{i,j^c}^{(r+s)}}{N_{b,j^c}^{(r+s)}} \right)^2 \geq 1 - \left( \frac{N_{k,j^c}^{(r)} / \delta_{k,j^c}^{(r)}}{N_{b,j^c}^{(r+s)}} \right)^2 - C_4 C_5 \sqrt{\frac{\log \log r}{r}}.
\]

34
Furthermore, there exists $C_6, C_9, C_{10} \geq 0$, such that

\[
\sum_{i \neq b \neq k} \left( \frac{N_i^{(r+s_r)}}{\sigma_{i,j,c}^{(r+s_r)}} \right)^2 - \sum_{i \neq b \neq k} \left( \frac{N_i^{(r)}}{\sigma_{i,j,c}^{(r)}} \right)^2 \geq 1 - \sum_{i \neq b \neq k} \left( \frac{N_i^{(r)}}{\sigma_{i,j,c}^{(r)}} \right)^2 - \left( \frac{N_i^{(r)}}{\sigma_{i,j,c}^{(r)+}} \right)^2 - C_4C_6^2 \sqrt{\log \log r} \tag{33}
\]

\[
\sum_{i \neq b \neq k} \left( \frac{N_i^{(r)}}{\sigma_{i,j,c}^{(r)+}} \right)^2 - \left( \frac{N_i^{(r)}}{\sigma_{i,j,c}^{(r)+}} \right)^2 - (C_4C_5^2 + C_6) \sqrt{\log \log r} \tag{34}
\]

(33) holds since $(k, j^c)$ is sampled at $r$. By Lemma 7 and $s_r = O(r)$, there exists $C_8, C_9, C_{10} > 0$, $N_{b,j}^{(r)} \geq C_8$, $N_{b,j}^{(r+s_r)} \leq C_{10} r$. We have (34) is lower bounded by

\[
\left( \frac{\sigma_{b,j,c}^{(r)}}{\sigma_{b,j,c}^{(r)+}} \right)^2 \left[ \frac{(N_{b,j})^2 + d_{b,j,c}^{(r,s_r)}}{2(N_{b,j}^{(r+s_r)})^2} - (N_{b,j}^{(r)})^2(2N_{b,j}^{(r+s_r)})^2 \right] - (C_4C_5^2 + C_7) \sqrt{\log \log r} \tag{35}
\]

There exists $C_{12} > 0$

\[
\sum_{i \neq b \neq k} \left( \frac{d_{i,j,c}^{(r+s_r)}}{\sigma_{i,j,c}^{(r+s_r)}} \right)^2 \geq 2C_9 \sqrt{\log \log r}, \quad \text{Choose } 0 < C_{12} \leq \left( \frac{\sigma_{b,j,c}^{(r)}}{\sigma_{b,j,c}^{(r)+}} \right)^2 \left( \frac{C_{12}}{C_{10}} \right)^2 2C_9 \text{ for all large } r.
\]

Furthermore since $d_{b,j,c}^{(r+s_r)} \geq C_2 \sqrt{r \log \log r}$, there exists $C_{13} > 0$

(36) holds for $C_2$ large enough (but not depends on $r$). For example, we can choose $C_{13} = C_{12}/2$ and then (36) holds for all $C_2 \geq 2C_{11} + \frac{2C_4C_7^2 + C_8}{C_{12}}$. Since $C_3, C_4, \ldots, C_{13}$ are all independent of $r$, the $C_2$ here is also independent of $r$. 

35
There exists a non-optimal design $h \neq k \neq b$, such that

$$
\frac{N_{h,j}^{(r+s_r)}}{N_{b,j}^{(r+s_r)}}/\hat{\sigma}_{h,j}^{(r+s_r)} - \sum_{i \neq b \neq k} \frac{N_i^{(r)}}{N_{b,j}^{(r+s_r)}}/\hat{\sigma}_{b,j}^{(r+s_r)} \geq C_{13}C_2 \sqrt{\frac{\log \log r}{r}}.
$$

Or equivalently,

$$
\frac{N_{h,j}^{(r+s_r)}}{N_{b,j}^{(r+s_r)}}/\hat{\sigma}_{h,j}^{(r+s_r)} - \frac{1}{K-2} C_{13} C_2 \sqrt{\frac{\log \log r}{r}}.
$$

By Lemma 7 and the convergence of the sample variance, there exists $0 < C_{14} < \frac{C_2}{K-2} \left( \frac{\hat{\sigma}_{h,j}^{(r+s_r)}}{\hat{\sigma}_{b,j}^{(r+s_r)}} \right)^2 \left( \frac{N_{h,j}^{(r)}}{N_{b,j}^{(r)}} \right)^2$ for all large $r$. We obtain

$$
\frac{N_{h,j}^{(r+s_r)}}{N_{b,j}^{(r+s_r)}}/\hat{\sigma}_{h,j}^{(r+s_r)} \geq 1 + C_{14} C_2 \sqrt{\frac{\log \log r}{r}}.
$$

Hence,

$$
\frac{N_{h,j}^{(r+s_r)}}{N_{b,j}^{(r+s_r)}} \geq \sqrt{1 + C_{14} C_2 \sqrt{\frac{\log \log r}{r}} \frac{N_{h,j}^{(r)}}{N_{b,j}^{(r)}}} \geq 1 + \frac{C_{14} C_2}{4} \sqrt{\frac{\log \log r}{r}} \frac{N_{h,j}^{(r)}}{N_{b,j}^{(r)}}
$$

for all large $r$ by Taylor Expansion. Let $v = \sup \{ \ell \leq s_r : 1_{h,j}^{(r+\ell)} = 1 \}$. $r + v$ is the last time before $r + s_r$ at which $h$ is sampled. We then have

$$
\frac{N_{h,j}^{(r+v)}}{N_{b,j}^{(r+v)}} \geq \frac{N_{h,j}^{(r+s_r)}}{N_{b,j}^{(r+s_r)}} = \frac{N_{h,j}^{(r+s_r)}}{N_{b,j}^{(r+s_r)}} - 1
$$

$$
= (1 - \frac{1}{N_{b,j}^{(r+s_r)}}) \frac{N_{h,j}^{(r+s_r)}}{N_{b,j}^{(r+s_r)}}
$$

$$
\geq (1 - \frac{1}{C_{14} C_2}) (1 + \frac{C_{14} C_2}{4} \sqrt{\frac{\log \log r}{r}}) \frac{N_{h,j}^{(r)}}{N_{b,j}^{(r)}}
$$

$$
\geq (1 - \frac{1}{C_{14} C_2}) (1 + \frac{C_{14} C_2}{4} \sqrt{\frac{\log \log r}{r}}) \frac{N_{h,j}^{(r)}}{N_{b,j}^{(r)}}
$$

$$
\geq (1 + \frac{C_{14} C_2}{8} \sqrt{\frac{\log \log r}{r}}) \frac{N_{h,j}^{(r)}}{N_{b,j}^{(r)}}
$$

for all large $r$. Then, given any $C_1$, there exists $C_2 \geq \max \{ \frac{2C_1}{C_{14}}, 2C_{11} + \frac{2C_{14} C_2}{C_{14}} \}$. The Lemma holds true for the $C_1, C_2$.

**Lemma 18** For a fixed sub-optimal design $k$, between two samples of $(k, j^c)$. Suppose $(k, j^c)$ is sampled at $r$ and let $\ell_r = \inf \{ \ell > 0 : 1_{(k,j^c)} = 1 \}$. $r + \ell_r$ is the next time $(k, j^c)$ being sampled. Then the number of samples that can be allocated to $(b, j^c)$ between $r$ and $r + \ell_r$ is $O(\sqrt{r \log \log r})$ almost surely.
Proof We use the same notation of \( s'_r, s_r, d^{(r,s)}_{b,j} \) as in Lemma 17. Since \( d^{(r,s')}_{b,j} - d^{(r,s')}_{b,j} = d^{(r,s')}_{b,j} - d^{(r,s')}_{b,j} + 1 = O(\sqrt{r \log \log r}) \) by Lemma 15, it is sufficient to prove \( d^{(r,s')}_{b,j} = O(\sqrt{r \log \log r}) \). Prove by contradiction. Suppose the statement does not hold. Then \( \forall C_2 > 0 \), there exists \( r \) such that \( d^{(r,s')}_{b,j} \geq C_2 \sqrt{r \log \log r} \). By Lemma 17, \( \forall C_1 > 0 \) (remain to be specified), there exists an iteration \( r \) at which \((k, j^c)\) is sampled, another non-optimal design \( h \neq k \) and an iteration \( v < s_r \), such that \((h, j^c)\) is sampled at \( v \) and

\[
\frac{\alpha^{(r+v)}_{h,b,j^c}}{\alpha^{(r+s_r)}_{h,b,j^c}} \geq \left(1 + C_1 \sqrt{\frac{\log \log r}{r}}\right) \frac{\alpha^{(r)}_{h,b,j^c}}{\alpha^{(r)}_{h,b,j^c}}
\]

holds. We aim to show \((h, j^c)\) cannot be sampled at \( r + v \) for a contradiction. It is sufficient to show

\[
\sum_{j=1}^{b} \frac{(\hat{\mu}^{(r+v)}_b - \mu^{(r+v)}_b)^2}{\alpha^{(r+v)}_{b,j}} + \sum_{j=1}^{b} \frac{(\hat{\mu}^{(r+v)}_{k,j} - \mu^{(r+v)}_{k,j})^2}{\alpha^{(r+v)}_{k,j}} \geq \sum_{j=1}^{b} \frac{(\hat{\mu}^{(r+v)}_b - \mu^{(r+v)}_b)^2}{\alpha^{(r+v)}_{b,j}} + \sum_{j=1}^{b} \frac{(\hat{\mu}^{(r+v)}_{k,j} - \mu^{(r+v)}_{k,j})^2}{\alpha^{(r+v)}_{k,j}} \tag{38}
\]

Denote by \( \delta^{(r,v)}_i = (\hat{\mu}^{(r)}_b - \mu^{(r)}_b)^2 \). It is equivalent to show

\[
\delta^{(r+v)}_h \left( \sum_{j=1}^{b} \frac{(\hat{\sigma}^{(r+v)}_{b,j})^2 (\pi^{(r+v)}_j)^2}{\alpha^{(r+v)}_{b,j}} + \sum_{j=1}^{b} \frac{(\hat{\sigma}^{(r+v)}_{k,j})^2 (\pi^{(r+v)}_j)^2}{\alpha^{(r+v)}_{k,j}} \right) > \delta^{(r+v)}_k \left( \sum_{j=1}^{b} \frac{(\hat{\sigma}^{(r+v)}_{b,j})^2 (\pi^{(r+v)}_j)^2}{\alpha^{(r+v)}_{b,j}} + \sum_{j=1}^{b} \frac{(\hat{\sigma}^{(r+v)}_{k,j})^2 (\pi^{(r+v)}_j)^2}{\alpha^{(r+v)}_{k,j}} \right) \tag{39}
\]

By Lemma 4, we have \( \frac{\alpha^{(r)}_{b,j}}{\alpha^{(r+v)}_{b,j}} = O\left(\sqrt{\frac{\log \log r}{r}}\right) \) for \( j \neq j^c \). Then, there exist \( C_3 > 0 \)

\[
\text{LHS of (39)} \geq \delta^{(r+v)}_h \left( \sum_{j=1}^{b} \frac{(\hat{\sigma}^{(r+v)}_{b,j})^2 (\pi^{(r+v)}_j)^2}{\alpha^{(r+v)}_{b,j}} + \sum_{j=1}^{b} \frac{(\hat{\sigma}^{(r+v)}_{k,j})^2 (\pi^{(r+v)}_j)^2}{\alpha^{(r+v)}_{k,j}} \right) - C_3 \sqrt{\frac{\log \log r}{r}}
\]

\[
\text{RHS of (39)} \leq \delta^{(r+v)}_k \left( \sum_{j=1}^{b} \frac{(\hat{\sigma}^{(r+v)}_{b,j})^2 (\pi^{(r+v)}_j)^2}{\alpha^{(r+v)}_{b,j}} + \sum_{j=1}^{b} \frac{(\hat{\sigma}^{(r+v)}_{k,j})^2 (\pi^{(r+v)}_j)^2}{\alpha^{(r+v)}_{k,j}} \right) + C_3 \sqrt{\frac{\log \log r}{r}}
\]

Since \((k, j^c)\) is sampled at \( r \),

\[
\delta^{(r)}_h \left( \sum_{j=1}^{b} \frac{(\hat{\sigma}^{(r)}_{b,j})^2 (\pi^{(r)}_j)^2}{\alpha^{(r)}_{b,j}} + \sum_{j=1}^{b} \frac{(\hat{\sigma}^{(r)}_{k,j})^2 (\pi^{(r)}_j)^2}{\alpha^{(r)}_{k,j}} \right) \geq \delta^{(r)}_k \left( \sum_{j=1}^{b} \frac{(\hat{\sigma}^{(r)}_{b,j})^2 (\pi^{(r)}_j)^2}{\alpha^{(r)}_{b,j}} + \sum_{j=1}^{b} \frac{(\hat{\sigma}^{(r)}_{k,j})^2 (\pi^{(r)}_j)^2}{\alpha^{(r)}_{k,j}} \right)
\]

Then we obtain,

\[
\delta^{(r+v)}_h \left( \sum_{j=1}^{b} \frac{(\hat{\sigma}^{(r+v)}_{b,j})^2 (\pi^{(r+v)}_j)^2}{\alpha^{(r+v)}_{b,j}} + \sum_{j=1}^{b} \frac{(\hat{\sigma}^{(r+v)}_{k,j})^2 (\pi^{(r+v)}_j)^2}{\alpha^{(r+v)}_{k,j}} \right) \geq \delta^{(r+v)}_k \left( \sum_{j=1}^{b} \frac{(\hat{\sigma}^{(r+v)}_{b,j})^2 (\pi^{(r+v)}_j)^2}{\alpha^{(r+v)}_{b,j}} + \sum_{j=1}^{b} \frac{(\hat{\sigma}^{(r+v)}_{k,j})^2 (\pi^{(r+v)}_j)^2}{\alpha^{(r+v)}_{k,j}} \right) - 2C_3 \sqrt{\frac{\log \log r}{r}} \tag{40}
\]

Then,

\[
\text{LHS of (40)} = \delta^{(r)}_h \left( \sum_{j=1}^{b} \frac{(\hat{\sigma}^{(r)}_{b,j})^2 (\pi^{(r)}_j)^2}{\alpha^{(r)}_{b,j}} \right) + \delta^{(r)}_h \left( \sum_{j=1}^{b} \frac{(\hat{\sigma}^{(r)}_{k,j})^2 (\pi^{(r)}_j)^2}{\alpha^{(r)}_{k,j}} \right)
\]

\[
+ \delta^{(r)}_h \left( \sum_{j=1}^{b} \frac{(\hat{\sigma}^{(r+v)}_{b,j})^2 (\pi^{(r+v)}_j)^2}{\alpha^{(r+v)}_{b,j}} \right) + \delta^{(r)}_h \left( \sum_{j=1}^{b} \frac{(\hat{\sigma}^{(r+v)}_{k,j})^2 (\pi^{(r+v)}_j)^2}{\alpha^{(r+v)}_{k,j}} \right) - C_3 \sqrt{\frac{\log \log r}{r}} \tag{41}
\]
and

\[
\text{RHS of (40)} = \delta_k^{(r)} \left( \frac{\sigma_{h,j}^{(r)}}{\sigma_{b,j}^{(r)}} \right)^2 (\pi_{j'}^{(r)})^2 \cdot \delta_h^{(r)} \left( \frac{\sigma_{b,j}^{(r+v)}}{\sigma_{b,j}^{(r)}} \right)^2 \delta_h^{(r)} \left( \frac{\sigma_{h,j}^{(r+v)}}{\sigma_{h,j}^{(r)}} \right)^2 (\pi_{j'}^{(r+v)})^2 \\
+ \delta_h^{(r)} \left( \frac{\sigma_{h,j}^{(r)}}{\sigma_{b,j}^{(r)}} \right)^2 (\pi_{j'}^{(r)})^2 \cdot \delta_h^{(r)} \left( \frac{\sigma_{b,j}^{(r+v)}}{\sigma_{b,j}^{(r)}} \right)^2 \delta_h^{(r)} \left( \frac{\sigma_{h,j}^{(r+v)}}{\sigma_{h,j}^{(r)}} \right)^2 (\pi_{j'}^{(r+v)})^2 - C_3 \sqrt{\frac{\log \log r}{r}}. 
\]

(42)

\forall 1 \leq i \leq K, by Lemma 10, \( \delta^{(r+v)} = 1 + O\left(\sqrt{\frac{\log \log r}{r}}\right) \); by Lemma 2 and 11, \( \frac{\sigma_{h,j}^{(r+v)}}{\sigma_{b,j}^{(r)}} = 1 + O\left(\sqrt{\frac{\log \log r}{r}}\right) \), and

\[
\frac{\sigma_{h,j}^{(r+v)}}{\sigma_{b,j}^{(r)}} = 1 + O\left(\sqrt{\frac{\log \log r}{r}}\right); \text{ Furthermore } N_{\sigma_{h,j}^{(r)}} = N_{\sigma_{b,j}^{(r)}} + 1. \text{ Hence, there exist } C_5 > 0, \text{ such that }
\]

\[
\delta_k^{(r)} \left( \frac{\sigma_{b,j}^{(r)}}{\sigma_{b,j}^{(r+v)}} \right)^2 (\pi_{j'}^{(r)})^2 \frac{\sigma_{h,j}^{(r)}}{\sigma_{h,j}^{(r+v)}} \left(1 - C_5 \sqrt{\frac{\log \log r}{r}}\right) + \delta_k^{(r)} \left( \frac{\sigma_{h,j}^{(r)}}{\sigma_{h,j}^{(r+v)}} \right)^2 (\pi_{j'}^{(r)})^2 \frac{\sigma_{b,j}^{(r)}}{\sigma_{b,j}^{(r+v)}} \left(1 + C_5 \sqrt{\frac{\log \log r}{r}}\right) - C_3 \sqrt{\frac{\log \log r}{r}} 
\]

(43)

and

\[
\delta_k^{(r)} \left( \frac{\sigma_{b,j}^{(r)}}{\sigma_{b,j}^{(r+v)}} \right)^2 (\pi_{j'}^{(r)})^2 \frac{\sigma_{h,j}^{(r)}}{\sigma_{h,j}^{(r+v)}} \left(1 + C_5 \sqrt{\frac{\log \log r}{r}}\right) + \delta_k^{(r)} \left( \frac{\sigma_{h,j}^{(r)}}{\sigma_{h,j}^{(r+v)}} \right)^2 (\pi_{j'}^{(r)})^2 \frac{\sigma_{b,j}^{(r)}}{\sigma_{b,j}^{(r+v)}} \left(1 - C_5 \sqrt{\frac{\log \log r}{r}}\right) + C_5 \sqrt{\frac{\log \log r}{r}} 
\]

(44)

Divide by \( \frac{\sigma_{b,j}^{(r+v)}}{\sigma_{h,j}^{(r+v)}} \) by both sides and notice \( 1 \leq \frac{\sigma_{b,j}^{(r+v)}}{\sigma_{h,j}^{(r+v)}} \leq \frac{1}{\alpha_{b,j}^{(r+v)}} \leq C_7 \) for some \( C_7 > 0 \) by Lemma 8. Then \( 43 ) > 44 \) can be implied by

\[
\delta_k^{(r)} \left( \frac{\sigma_{b,j}^{(r)}}{\sigma_{b,j}^{(r+v)}} \right)^2 (\pi_{j'}^{(r)})^2 \frac{\sigma_{h,j}^{(r)}}{\sigma_{h,j}^{(r+v)}} \left(1 - C_5 \sqrt{\frac{\log \log r}{r}}\right) + \delta_k^{(r)} \left( \frac{\sigma_{h,j}^{(r)}}{\sigma_{h,j}^{(r+v)}} \right)^2 (\pi_{j'}^{(r)})^2 \frac{\sigma_{b,j}^{(r)}}{\sigma_{b,j}^{(r+v)}} \left(1 + C_5 \sqrt{\frac{\log \log r}{r}}\right) - C_3 \sqrt{\frac{\log \log r}{r}}
\]

for some \( C_8 \) independent of \( r \) (depends on \( C_5, C_7, C_3 \)). Since \( \frac{\sigma_{b,j}^{(r)}}{\sigma_{h,j}^{(r+v)}} \frac{\sigma_{h,j}^{(r+v)}}{\sigma_{b,j}^{(r+v)}} \) \( < \frac{1}{1 + C_1 \sqrt{\frac{\log \log r}{r}}} < 1 - \frac{C_1}{2} \sqrt{\frac{\log \log r}{r}} \)

\[
\text{for all large } r \text{'s, (45) can be further implied by} \\
\delta_k^{(r)} \left( \frac{\sigma_{b,j}^{(r)}}{\sigma_{b,j}^{(r+v)}} \right)^2 (\pi_{j'}^{(r)})^2 \frac{\sigma_{h,j}^{(r)}}{\sigma_{h,j}^{(r+v)}} \left(1 - C_5 \sqrt{\frac{\log \log r}{r}}\right) + \delta_k^{(r)} \left( \frac{\sigma_{h,j}^{(r)}}{\sigma_{h,j}^{(r+v)}} \right)^2 (\pi_{j'}^{(r)})^2 \frac{\sigma_{b,j}^{(r)}}{\sigma_{b,j}^{(r+v)}} \left(1 + C_5 \sqrt{\frac{\log \log r}{r}}\right) - C_3 \sqrt{\frac{\log \log r}{r}}
\]

\[
= \delta_k^{(r)} \left( \frac{\sigma_{b,j}^{(r)}}{\sigma_{b,j}^{(r+v)}} \right)^2 (\pi_{j'}^{(r)})^2 \frac{\sigma_{h,j}^{(r)}}{\sigma_{h,j}^{(r+v)}} + \left[ C_8 + C_5 \delta_k^{(r)} \left( \frac{\sigma_{h,j}^{(r)}}{\sigma_{h,j}^{(r+v)}} \right)^2 (\pi_{j'}^{(r)})^2 \frac{\sigma_{b,j}^{(r)}}{\sigma_{b,j}^{(r+v)}} \right] - C_3 \sqrt{\frac{\log \log r}{r}}
\]

(45)

Let \( C_9, C_{10} \) be two constants such that \( C_9 \leq \delta_k^{(r)} \left( \frac{\sigma_{b,j}^{(r)}}{\sigma_{b,j}^{(r+v)}} \right)^2 (\pi_{j'}^{(r)})^2 \frac{\sigma_{h,j}^{(r)}}{\sigma_{h,j}^{(r+v)}} \leq C_{10} \). Then by (40), (45) is implied by

\[
2C_8 + 2C_9 + C_5 C_{10} - C_1 \frac{1}{2} C_9 \sqrt{\frac{\log \log r}{r}} < 0.
\]

(46)

Since \( C_3, C_5, C_8, C_9, C_{10} \) are all independent of \( r \), we can choose \( C_1 \) to be large enough such that (46) holds, which gives us the contradiction that \( (h_j, j_c) \) cannot be sampled at \( r + v \).

We can now finally prove the “Local Balance” optimality condition.
Proof of (16) in Theorem 4.2. Proof. Denote by \( \delta_k^{(f)} = (\hat{\mu}_b^{(f)} - \hat{\mu}_i^{(f)})^2 \) and \( \Theta^{(f)}_i = \sum_{j=1}^{D} \frac{\sigma_{ij}^{(f)} (\sigma_{ij}^{(f)})^2}{N_{ij}^{(f)}} \). For any two non-optimal designs \( i \neq k \) and an iteration \( r \). Without loss of generality, assume \( \Theta^{(f)}_i \leq \Theta^{(f)}_k \). Let \( u = \sup \{ \ell : r \in \mathbb{R} \} \) be the last time \( k \) is sampled before \( r \). \( r = O(u) \) by Lemma 12. Also notice \( (\sigma_{ij}^{(f)} (\sigma_{ij}^{(f)})^2) = O(\sqrt{\log \log r}) \) for \( j \neq j^c \), we have

\[
0 \leq \Theta_k^{(r)} - \Theta_i^{(r)} \leq \frac{\delta_k^{(r)}}{(\sigma_{ij}^{(f)} (\sigma_{ij}^{(f)})^2)} + O(\sqrt{\log \log r}) \leq \frac{\delta_k^{(r)}}{(\sigma_{ij}^{(f)} (\sigma_{ij}^{(f)})^2)} + O(\frac{\log \log r}{r}) \leq \frac{\delta_k^{(r)}}{(\sigma_{ij}^{(f)} (\sigma_{ij}^{(f)})^2)} + O\left(\frac{\log \log r}{r}\right) \leq \frac{\delta_k^{(r)}}{(\sigma_{ij}^{(f)} (\sigma_{ij}^{(f)})^2)} + O\left(\frac{\log \log r}{r}\right).
\]

By Lemma 10, \( \delta_i^{(r)} = \delta_i^{(u)} + O(\sqrt{\log \log u}) = \delta_i^{(u)} + O\left(\frac{\log \log r}{r}\right), \forall i \); by Lemma 2 and Lemma 11, \( \pi_i^{(r)} = 1 - O\left(\frac{\log \log r}{r}\right) \). Hence, \( \Theta_i^{(r)} = \Theta_i^{(u)} + O(\sqrt{\log \log r}) \). Furthermore \( \alpha_{i,j}^{(r)} \) is upper and lower bounded by positive constants \( \forall i, j \). Then,

\[
\frac{\delta_k^{(u)}}{(\sigma_{ij}^{(f)} (\sigma_{ij}^{(f)})^2)} + O\left(\frac{\log \log r}{r}\right) \leq \frac{\delta_k^{(u)}}{(\sigma_{ij}^{(f)} (\sigma_{ij}^{(f)})^2)} + O\left(\frac{\log \log r}{r}\right).
\]

Since \( r = O(u) \) and by Lemma 8, \( \lim_{r \to \infty} \Theta_i^{(r)} = 0 \forall i \). Then,

\[
\frac{\delta_k^{(u)}}{(\sigma_{ij}^{(f)} (\sigma_{ij}^{(f)})^2)} + O\left(\frac{\log \log r}{r}\right) \leq \frac{\delta_k^{(u)}}{(\sigma_{ij}^{(f)} (\sigma_{ij}^{(f)})^2)} + O\left(\frac{\log \log r}{r}\right).
\]

The second inequality holds because \( (\sigma_{ij}^{(f)} (\sigma_{ij}^{(f)})^2) = O\left(\frac{\log \log r}{r}\right) \) for all \( i, j \neq j^c \). This finalizes the proof. \(\blacksquare\)
.9 Proof of Lemma 3

Proof of 1. By Assumption 9.2 and 9.3, we know the Bayesian consistency holds by Doob’s consistency theorem. Furthermore by Assumption 9.1, we know there exists $\epsilon > 0$, such that $B_\epsilon(\theta^c) \subset \Theta_{j^c}$, where $B_\epsilon(\theta^c)$ is the ball centered at $\theta^c$ with radius $\epsilon$. Hence, by Bayesian consistency, $\pi_{j^c}^t \rightarrow 1$ almost surely for every sequence of input data.

Proof of 2. Mean: Recall $X_{i,j}^{(r)}$ is the $r$th simulation output for $(i,j)$. Let $\ell_r$ be the iteration at which $X_{i,j}^{(r)}$ is simulated. Here $\ell_r$ is random given $r$.

Let $\hat{y}_{i,j}^{(k)} = \mathbb{E}_{\pi_{j^c}^t}[y_{i,j}^{(k)}(\theta)]$, $Z_{i,j}^{(r)} = X_{i,j}^{(r)} - \hat{y}_{i,j}^{(r)}$. Notice

$$\mu_{i,j}^{(r)} = \mathbb{E}_{\pi_{j^c}^t}[y_{i,j}^{(r)}(\theta)]$$

For (I), we have

$$\mathbb{E}[X_{i,j}^{(r)}|Z_{i,j}^{(1)}, \ldots, Z_{i,j}^{(r-1)}] = \mathbb{E}[\mathbb{E}[X_{i,j}^{(r)}|\pi_{j^c}^{(t_r)}, \ell_r, Z_{i,j}^{(1)}, \ldots, Z_{i,j}^{(r-1)}]|Z_{i,j}^{(1)}, \ldots, Z_{i,j}^{(r-1)}].$$

Notice $\ell_r = \ell_r$, $Z_{i,j}^{(r)}$, $r' = 1, \ldots, r - 1$ are determined by input data and simulation outputs up to iteration $\ell$. By Assumption 9.5,

$$\mathbb{E}[X_{i,j}^{(r)}|\pi_{j^c}^{(t_r)}, \ell_r, Z_{i,j}^{(1)}, \ldots, Z_{i,j}^{(r-1)}] = \mathbb{E}[X_{i,j}^{(r)}|\pi_{j^c}^{(t_r)}] = 0.$$  

This implies $\mathbb{E}[X_{i,j}^{(r)}|Z_{i,j}^{(1)}, \ldots, Z_{i,j}^{(r-1)}] = 0$, which means $\{Z_{i,j}^{(r)}\}_{r=1}^\infty$ is a Martingale difference sequence (MDS) Moreover, let $\theta(\ell_r)$, $\ell_r \sim \pi_{j^c}^{(t_r)}$ be the parameter under which $X_{i,j}^{(r)}$ is simulated. Here $\pi_{j^c}^{(t)}$ is the conditional posterior distribution at stage $t$.

$$\frac{1}{2}\mathbb{E}[(y_{i,j}^{(k)})^2] \leq \mathbb{E}[(X_{i,j}^{(r)})^2] + \mathbb{E}[(\hat{y}_{i,j}^{(r)})^2] \leq \mathbb{E}[(X_{i,j}^{(r)})^2] + \mathbb{E}[(\hat{y}_{i,j}^{(r)})^2] \leq \max\{(y_{i,j}^{(1)}(\theta))^2 + (y_{i,j}^{(2)}(\theta))^2\} < \infty,$$

where the boundedness comes from the fact that $y_{i,j}^{(k)}$, $k = 1, 2$ is continuous and the parameter set is compact by Assumption 9. Hence, by strong law of large number (SLLN) for MDS 55, we obtain $\lim_{\ell \rightarrow \infty} \frac{1}{N_{i,j}^{(t)}} \sum_{r=1}^\ell Z_{i,j}^{(r)} = 0$ almost surely. Notice Since $N_{i,j}^{(t)} \rightarrow \infty$ almost surely as $\ell \rightarrow \infty$, we have $\lim_{\ell \rightarrow \infty} \frac{1}{N_{i,j}^{(t)}} \sum_{r=1}^\ell Z_{i,j}^{(r)} = 0$ almost surely.

For (II), by Bayesian consistency, $\pi_{j^c}^{(t)}$ converges to a Dirac measure at $\theta^c$ almost surely. Further notice $y_{i,j}^{(1)}$ is continuous and bounded. This implies $\hat{y}_{i,j}^{(r)} \rightarrow \mu_{i,j}(\theta^c)$ almost surely. Since $N_{i,j}^{(t)} \rightarrow \infty$ almost surely, we have the average sum (III) $\rightarrow 0$ almost surely. That is, the convergence of sequence $\{\alpha_n\}_{n=1}^\infty$ implies the convergence of sequence $\{S_n\}_{n=1}^\infty$ to the same limit, where $S_n = \frac{1}{n} \sum_{k=1}^n \alpha_k$.

Variance: For $\hat{\sigma}_{i,j}^{(t)}$, let $\hat{y}_{i,j}^{(k)} = \mathbb{E}_{\pi_{j^c}^t}[y_{i,j}^{(k)} (\theta)]$.

$$\frac{N_{i,j}^{(t)}}{N_{i,j}^{(t)} - 1} (\hat{\sigma}_{i,j}^{(t)})^2 - \sigma_{i}(\theta^c)^2 = \frac{1}{N_{i,j}^{(t)}} \sum_{r=1}^{N_{i,j}^{(t)}} (X_{i,j}^{(r)} - \mu_{i,j}^{(r)})^2 - \sigma_{i}(\theta^c)^2$$

$$= \frac{1}{N_{i,j}^{(t)}} \sum_{r=1}^{N_{i,j}^{(t)}} (X_{i,j}^{(r)} - \hat{y}_{i,j}^{(t_r)})^2 + \frac{1}{N_{i,j}^{(t)}} \sum_{r=1}^{N_{i,j}^{(t)}} (\hat{y}_{i,j}^{(t_r)} - (\hat{\mu}_{i,j}^{(t)})^2 - \sigma_{i}(\theta^c)^2$$

(I) can be proved to converge to 0 in a similar manner as for the mean estimator, by regarding $(X_{i,j}^{(r)})^2$, $r = 1, \ldots, N_{i,j}^{(t)}$ as the simulation samples and $\mathbb{E}[(X_{i,j}^{(r)})^2|\pi_{j^c}^{(t_r)}] = \mathbb{E}[\mathbb{E}[(X_{i,j}^{(r)})^2|\pi_{j^c}^{(t_r)}]] = \mathbb{E}[\mathbb{E}[(X_{i,j}^{(r)})^2|\pi_{j^c}^{(t_r)}]] = \mathbb{E}(\hat{y}_{i,j}^{(r)}).$ Hence,
$Z_{i,j}^{(r)} := (X_{i,j}^{(r)})^2 - \tilde{y}_{i,j}^{(t_r)}$ is a MDS sequence for $r = 1, 2, \ldots$. Furthermore we can show $\mathbb{E}[Z_{i,j}^{2r}] \leq 2 \max_{\theta \in \Theta} \{y_{i,j}^{(4)} + (y_{i,j}^{(2)})^2\} < \infty$. Hence, we can invoke the SLLN for MDS to prove $(I) \rightarrow 0$ almost surely.

For (II), again by the Bayesian consistency and the continuity of $y_{i,j}^{(2)}(\theta)$, we can show almost surely, $\tilde{y}_{i,j}^{(t_r)} \rightarrow \mu_i(\theta_c)^2 + \sigma_i(\theta_c)^2$. Since we also have $\hat{\mu}_{i,j}^{(t_r)} \rightarrow \mu_i(\theta_c)$ almost surely, we obtain $\tilde{y}_{i,j}^{(t_r)} - \hat{\mu}_{i,j}^{(t)} - \sigma_i(\theta_c)^2 \rightarrow 0$ almost surely. This implies $(II) \rightarrow 0$ almost surely. The proof is complete. \qed