Proceedings of the 2023 Winter Simulation Conference
C. G. Corlu, S. R. Hunter, H. Lam, B. S. Onggo, J. Shortle, and B. Biller, eds.

TOP-TWO THOMPSON SAMPLING FOR SELECTING CONTEXT-DEPENDENT BEST DESIGNS

Xinbo Shi
Yijie Peng
Gongbo Zhang

Guanghua School of Management
Peking University
5 Yiheyuan Road
Beijing 100871, P. R. CHINA

ABSTRACT

We consider a contextual ranking and selection problem which aims to identify the best-performing alternative for each context. The performance is measured by an arbitrary identifiable statistical characteristic. Under a Bayesian framework, we establish the posterior large deviation ratios for general adaptive sampling policies. We propose an efficient sampling policy based on top-two Thompson sampling, which is proven to be consistent. Numerical experiments demonstrate that the proposed algorithm outperforms existing algorithms under both Gaussian and non-Gaussian settings.

1 INTRODUCTION

Ranking and Selection (R&S) has been constantly receiving research interest in simulation for decades (Bechhofer 1954; Bechhofer et al. 1995; Kim and Nelson 2006). Its goal is to identify the best alternative among a finite number of options, where the performances are modeled by statistical characteristics of simulation outputs, typically in terms of means. In this work, we focus on context-dependent R&S (CR&S), a variant of classic R&S that aims to find the best alternative for each context. To exemplify this, a factory manufacturing seasonal products may want to make annual production schedules for the four upcoming seasons. The seasons and potential production schedules can be deemed as contexts and alternatives, respectively. Other applications include patient-specific cancer treatment (Kim et al. 2011; Li et al. 2022), personalized assortment optimization (Miao and Chao 2022), context-aware recommendation system (Woerndl et al. 2007), etc.

There are two main streams in the formulation of R&S problems. In the fixed-budget formulation, the total sample size is predetermined, and the decision maker strives to maximize the probability of correct selection (PCS) by allocating the simulation budget. In contrast, in the fixed-precision formulation, samples are required and allocated sequentially until a specified PCS level is reached. For a recent review of the R&S problems, see Hong et al. (2021). The common and central issue therein is how to efficiently allocate the costly simulation budget to improve sampling efficiency. An efficient sampling policy well balances the *exploitation* of the empirically most promising alternatives and the *exploration* of the undersampled ones. This idea is described as the exploitation-exploration tradeoff (Hong and Nelson 2006). In CR&S, the best design is context-dependent and takes different sampling efforts to be distinguished from other alternatives under each context. Therefore, this tradeoff exists not only among alternatives but also among contexts, making the problem more demanding.

The focus of research on the fixed budget R&S problem is gradually shifting from a static formulation to a dynamic one. Following the works of Chen et al. (2000) and Chen and Lee (2011), an extensive body

of literature on the optimal computing budget allocation (OCBA) procedure seeks to achieve sampling ratios that satisfy certain asymptotic optimality conditions. However, asymptotically optimal sampling ratios depend on unknown distribution parameters. These algorithms are based on a static optimization formulation while overlooking the information gain brought about by sequential sampling decisions. In view of this, Peng et al. (2018) formulate the sequential sampling decision as a Markov decision process (MDP) and propose a dynamic scheme called approximately optimal allocation policy (AOAP), based on a myopic approximation of the MDP. More recently, Russo (2020) considers allocating the simulation budget in an adaptive way and introduces an efficient algorithm known as top-two Thompson sampling (TTTS). TTTS generalizes Thompson sampling (TS), a well-known heuristic algorithm designed for the multi-armed bandit problem, by forcing enough exploration. TS balances the exploitation-exploration tradeoff by a stationary randomized strategy that assigns the simulation budget to arms in proportion to the posterior probability of being optimal (Russo et al. 2018). However, when the posterior belief is concentrated around the true value, TS behaves largely like pure exploitation, and thus the exploration of the potential optimal arm is typically insufficient. Both AOAP and TTTS deliver optimal sampling ratios asymptotically. In this work, we propose a sampling scheme for CR&S that fits into the dynamic setting as well. We consider general adaptive sampling policies under a Bayesian framework, which successively update posterior belief on the underlying distributions and allocate simulation replications based on observed sample information.

CR&S has recently attracted increasing research attention. To the best of our knowledge, the study by Shen et al. (2021) is the first in this field. It develops a sampling policy based on indifferent zone (IZ) under a fixed-precision formulation, where a linear response function is used to map contexts to the mean performance of each alternative. Li et al. (2018) extend the work of Shen et al. (2021) by allowing the response function to be non-linear. Cakmak et al. (2021) model sample performance with a Gaussian process (GP) under a Bayesian framework, where the prior distribution is constructed on the distances between contexts. These works associate contexts with continuous covariates, mitigating the difficulty arising from multiple contexts. Moreover, feasible alternatives must be identical across contexts. In contrast, our work assumes no extra information other than the context, which is a categorical variable, and the set of feasible alternatives may vary across contexts. The most related work to ours is by Du et al. (2022), who characterize asymptotically optimal sampling ratios for a CR&S problem and derive a sampling policy by sequentially balancing the two sides of equations of asymptotic optimality conditions. More recently, Li et al. (2022) utilize the clustering structure of both alternatives and contexts to enhance efficient learning of the best designs and develop a dynamic sampling policy to maximize the worst-case PCS.

The aforementioned studies assume Gaussian sample distributions. In our work, we consider selecting the best alternative under each context for generally parameterized families of distributions. Besides identifying the best alternative with the largest (or the smallest) mean, our framework allows for a broader range of statistical characteristics in the presence of nuisance parameters. Inspired by Russo (2020), we properly control the sampling effort on the most promising alternatives with a hyperparameter. With a careful choice of the hyperparameter, our method is theoretically consistent and rate-optimal asymptotically. Our contributions are two-fold. First, we derive posterior large deviation ratios under mild assumptions on the sampling distribution, which depict the convergence rate of posterior belief on the parameter set leading to correct selection for general adaptive sampling policies. Second, we propose an efficient heuristic, top-two Thompson sampling for CR&S (TTTS-C), which yields better performance than other compared algorithms. Specifically, TTTS-C leads to a significantly larger probability of simultaneous correct selection in all contexts than the compared algorithms.

The rest of the paper is organized as follows. Section 2 formulates the CR&S problem under a Bayesian framework. An efficient sampling policy is proposed and its theoretical properties are shown in Section 3. Section 4 presents the numerical results of the proposed algorithm under both Gaussian and non-Gaussian settings. The last section concludes the paper.

2 PROBLEM FORMULATION

We consider the problem of identifying the best alternative among $|\mathscr{D}_c|$ alternatives for each context $c \in \mathscr{C}$, where $|\cdot|$ is the cardinality of a set, $\mathscr{D}_c$ is the set of all competitive alternatives under a context c, and $\mathscr{C}$ is the set of all possible contexts. Let $Y_{(c,d)}$ denote the simulation sample of alternative $d \in \mathscr{D}_c$ under context c, which follows a regular parametric family of distributions with density $p(y|\theta_{(c,d)})$, where $\theta_{(c,d)} = (\mu_{(c,d)}, \eta_{(c,d)})$ is a finite-dimensional parameter, $\mu_{(c,d)} \in M \subseteq \mathbb{R}$ is the parameter of interest, $\eta_{(c,d)} \in H \subseteq \mathbb{R}^k$ is the nuisance parameter, and M and H are parameter spaces of μ and η, respectively. To be specific, the performance of each context-design pair is measured by $\mu_{(c,d)}$, and the best context-dependent alternative is $d^*(c) := \arg\max_{d \in \mathscr{D}_c} \mu_{(c,d)}$ for each context c. Taking the normal distribution as an example, we have $p(\cdot|\theta) = \phi(\cdot|\mu, \sigma^2)$, where σ^2 is the nuisance parameter that can only be estimated by simulation.

Let $\boldsymbol{Y} = \left(Y_{(c,d)}\right)$ denote the vector of simulation samples, and $\boldsymbol{\theta} = \left(\theta_{(c,d)}\right) \in \Theta$ be the parametric vector for all context-design pairs. We make some regularity assumptions on the parametric family.

Assumption 1 The parametric family of sampling distributions $\{p(\cdot|\theta) : \theta \in M \times H\}$ satisfies:

(1) The parameter space $M \times H$ is compact and Θ has a boundary with Lebesgue measure zero;
(2) The parametric family $\{p(\cdot|\theta) : \theta \in M \times H\}$ is continuous in θ;
(3) The parametric family $\{p(\cdot|\theta) : \theta \in M \times H\}$ is identifiable, i.e., for $\theta \neq \theta'$, the set $\{y \in \mathbb{R} : p(y|\theta) \neq p(y|\theta')\}$ has positive Lebesgue measure;
(4) The set of log-likelihood ratio tests $\mathscr{L} = \{\ln\{p(y|\theta')/p(y|\theta'')\} : \theta', \theta'' \in M \times H\}$ constitutes a universal Glivenko-Cantelli (GC) class, i.e., for any $\theta \in M \times H$, $\sup_{f \in \mathscr{L}} \left|\frac{1}{T}\sum_{t=1}^{T} f(Y^{(t)}) - \mathbb{E}[f(Y)]\right| \to 0$, *a.s.*, where $Y^{(1)}, \ldots, Y^{(T)}$ are independent and identically distributed samples of $Y \sim p(\cdot|\theta)$.

Both (1) and (2) in Assumption 1 are common regular conditions in Bayesian R&S procedures. (3) ensures that the inference for the optimal context-dependent design is not influenced by unknown nuisance parameters and plays a role in proving the consistency of the proposed algorithm. It's also common in statistics. (4) is key to the asymptotic analysis of the sampling policy. We assume that $Y_{(c,d)}$ are independent across contexts as well as alternatives. Then the joint density of $\boldsymbol{Y}$ belongs to a parametric model $\mathscr{F} = \{p(\boldsymbol{Y}|\boldsymbol{\theta}) = \prod_{c \in \mathscr{C}} \prod_{d \in \mathscr{D}_c} p(Y_{(c,d)}|\theta_{(c,d)}) : \boldsymbol{\theta} \in \Theta\}$. Moreover, we assume that this model is correctly specified and that the true distribution of $\boldsymbol{Y}$ passes through $\mathscr{F}$ at the ground truth $\boldsymbol{\theta}^*$.

Suppose that the total number of simulation budgets is T. Let $\{\boldsymbol{Y}^{(t)} : t \in \mathbb{N}^+\}$ be a sequence of independent simulation replications of $\boldsymbol{Y}$. At each step $t \leq T$, $t \in \mathbb{N}^+$, the decision maker chooses a context-design pair (C_t, D_t) and observes the realization of the corresponding coordinate $Y^{(t)}_{(C_t,D_t)}$. We focus on randomized sampling policies $\boldsymbol{\psi} := (\psi^{(t)})_{t \in \mathbb{N}}$, where $\psi^{(t)}$ is an $\mathscr{E}_{t-1}$-measurable random distribution over all context-design pairs and $\mathscr{E}_{t-1}$ is a σ-field generated by the observed sample information $\{(C_1, D_1), Y^{(1)}_{(C_1,D_1)}, \ldots, (C_{t-1}, D_{t-1}), Y^{(t-1)}_{(C_{t-1},D_{t-1})}\}$ up to time $(t-1)$. The probability of observing a context-design pair (c,d) conditioned on $\mathscr{E}_{t-1}$ is given by $\psi^{(t)}(c,d) := \mathbb{E}[\mathbf{1}\{(C_t, D_t) = (c,d)\}|\mathscr{E}_{t-1}]$. Then, an independent source of randomness is utilized to realize the context-design pair (C_t, D_t) according to ψ_t. These policies are known as *adaptive sampling policies*, including a wide range of legitimate non-anticipatory policies that rely only on the observed information.

We design the proposed algorithm from a Bayesian perspective. Suppose that $\Pi_0(\boldsymbol{\theta})$ is a prior distribution of $\boldsymbol{\theta}$. We highlight that this prior does not necessarily reflect the true generating mechanism of the model parameter and the underlying parameter $\boldsymbol{\theta}^*$ is considered fixed as in a frequentist perspective. We merely require the support of the prior distribution to at least include an open neighborhood containing $\boldsymbol{\theta}^*$. Then, the information on the model parameter $\boldsymbol{\theta}$ is accumulated through the recursively updated posterior distribution given by:

$$\Pi_t(\boldsymbol{\theta}) = \frac{\Pi_{t-1}(\boldsymbol{\theta}) L_t(\boldsymbol{\theta})}{\int_\Theta \Pi_{t-1}(\boldsymbol{\theta}') L_t(\boldsymbol{\theta}') d\boldsymbol{\theta}'} ,$$

where $L_t(\boldsymbol{\theta}) = \mathbb{E}\left[p(\boldsymbol{Y}|\boldsymbol{\theta})\middle|Y^{(t)}_{(C_t,D_t)}\right] = p\left(Y^{(t)}_{(C_t,D_t)}\middle|\theta_{(C_t,D_t)}\right)$ is the expected likelihood function conditioned on the partial observation $Y^{(t)}_{(C_t,D_t)}$. With a slight abuse of notation, we refer to $\Pi_t(\cdot)$ as both the density and the probability measure of the posterior distribution. We assume that the prior density $\Pi_0(\boldsymbol{\theta})$ is bounded both above and below so that it contributes to $\Pi_t(\boldsymbol{\theta})$ by at most a multiplicative constant. For simplicity, we use the uninformative prior $\Pi_0(\boldsymbol{\theta}) = 1$, and then the likelihood function and posterior belief become

$$L_T(\boldsymbol{\theta}) = L_T(\boldsymbol{\theta}^*)\exp\left\{-\ln\frac{L_T(\boldsymbol{\theta}^*)}{L_T(\boldsymbol{\theta})}\right\} = L_T(\boldsymbol{\theta}^*)\exp\{-TW_T(\boldsymbol{\theta})\}\ ,$$

and

$$\Pi_T(\boldsymbol{\theta}) = \frac{L_T(\boldsymbol{\theta})}{\int_\Theta L_T(\boldsymbol{\theta}')d\boldsymbol{\theta}'} = \frac{\exp\{-TW_T(\boldsymbol{\theta})\}}{\int_\Theta \exp\{-TW_T(\boldsymbol{\theta}')\}d\boldsymbol{\theta}'}\ ,$$

respectively, where

$$W_T(\boldsymbol{\theta}) = \frac{1}{T}\ln\frac{L_T(\boldsymbol{\theta}^*)}{L_T(\boldsymbol{\theta})} = \sum_{c\in\mathscr{C},d\in\mathscr{D}_c}\frac{1}{T}\sum_{t=1}^{T}\xi_t(c,d)\lambda\left(\theta^*_{(c,d)},\theta_{(c,d)};Y^{(t)}_{(c,d)}\right)\ ,$$

is the average exponential rate at which $\Pi_T(\boldsymbol{\theta})$ decays, $\xi_t(c,d) = \mathbf{1}\{(C_t,D_t) = (c,d)\}$, and $\lambda(\theta^*,\theta;y) = \ln\{p(y|\theta^*)/p(y|\theta)\}$ is the log-likelihood ratio.

We aim to maximize the probability of simultaneous correct selection in all contexts. The selection policy $\hat{d}:\mathscr{C}\to\bigcup_{c\in\mathscr{C}}\mathscr{D}_c$ can be modeled as a Bayesian decision based on the posterior belief Π_T. Consider the loss function $\ell(\hat{d}(\cdot),d^*(\cdot)) = \max_{c\in\mathscr{C}}\mathbf{1}\{\hat{d}(c)\neq d^*(c)\}$, which equals to one as long as the best alternative is incorrectly selected under any context. Then, the Bayesian selection decision that minimizes the posterior expected loss $\mathbb{E}[\ell(\hat{d}(\cdot),d^*(\cdot))|\mathscr{E}_T]$ can be derived explicitly as:

$$\hat{d}_B(\cdot) := \arg\max_{d(\cdot)}\Pi_T\left(\bigcap\nolimits_{c\in\mathscr{C}}\Theta^{(c,d(c))}\right)\ , \tag{1}$$

where $\Theta^{(c,d)} = \{\boldsymbol{\theta}\in\Theta:\mu_{(c,d)}\geq\mu_{(c,d')},\ \forall d'\in\mathscr{D}_c\}$, and the subscript B stands for 'Bayesian'. It would be reasonable to maximize, if possible, the objective in (1) for correct selection $d^*(\cdot)$, i.e., $\Pi_T(\bigcap_{c\in\mathscr{C}}\Theta^{(c,d^*(c))})$. However, it's impracticable to do so due to the dependence of the posterior belief on random samples. Therefore, we propose to maximize an asymptotic proxy of $\Pi_T(\bigcap_{c\in\mathscr{C}}\Theta^{(c,d^*(c))})$, as we will discuss in Section 3.1.

3 TOP-TWO THOMPSON SAMPLING

In this section, we propose the top-two Thompson sampling algorithm for CR&S problem, which allocates the simulation budget in a similar way to TS but controls the effort devoted to the best alternative a posteriori. This control allows for sufficient exploration and ensures that the algorithm performs well.

We first introduce a strategy parameterized by vector $\boldsymbol{\gamma} = (\gamma(c))$, where $\gamma(c)\in(0,1)$, $\forall\ c\in\mathscr{C}$. At each step t, we first sample $\hat{\boldsymbol{\theta}}^{(t)}_1$ from the posterior distribution $\Pi_{t-1}(\cdot)$, and find the corresponding best alternative for each context $\hat{d}^{(t)}_1(c) := \arg\max_{d\in\mathscr{D}_c}\hat{\mu}^{(t)}_{1,(c,d)}$, $c\in\mathscr{C}$. Then we successively re-sample $\hat{\boldsymbol{\theta}}^{(t)}_2$ from $\Pi_{t-1}(\cdot)$, and find the best alternative for each context $\hat{d}^{(t)}_2(c) := \arg\max_{d\in\mathscr{D}_c}\hat{\mu}^{(t)}_{2,(c,d)}$, $c\in\mathscr{C}$, until $\hat{d}^{(t)}_2(\cdot)$ and $\hat{d}^{(t)}_1(\cdot)$ are unequal as a function of c. Let Δ_t denote the set of contexts where $\hat{d}^{(t)}_1(c)$ differs from $\hat{d}^{(t)}_2(c)$, i.e., $\Delta_t := \{c\in\mathscr{C}:\hat{d}^{(t)}_1(c)\neq\hat{d}^{(t)}_2(c)\}$. Then, the observed context C_t is uniformly chosen from Δ_t. The observed alternative D_t is equal to $\hat{d}^{(t)}_1(C_t)$ with probability $\gamma(C_t)$ and is equal to $\hat{d}^{(t)}_2(C_t)$ with

complementary probability. Therefore, in the TTTS-C, $\psi^{(t)}(c,d) = \mathbb{E}[\zeta^{(t)}(c,d)|\mathscr{E}_{t-1}]$, where

$$\zeta^{(t)}(c,d) = \mathbb{E}[\mathbf{1}\{(C_t,D_t) = (c,d)\}|\mathscr{E}_{t-1}, \hat{d}_1^{(t)}(\cdot), \hat{d}_2^{(t)}(\cdot)] = \begin{cases} \gamma(c)/|\Delta_t| & , \quad c \in \Delta_t,\ d = \hat{d}_1^{(t)}(c)\ , \\ (1-\gamma(c))/|\Delta_t| & , \quad c \in \Delta_t,\ d = \hat{d}_2^{(t)}(c)\ , \\ 0 & , \quad \textit{otherwise}\ . \end{cases}$$

We note that TTTS-C is unique in its ability to simultaneously allocate simulation budgets to both contexts and alternatives. When $|\mathscr{C}| = 1$, TTTS-C reduces to the TTTS algorithm proposed by Russo (2020) for selecting the context-free best alternative. Notice that in the TTTS-C, $\zeta^{(t)}(c,d)$ can be rewritten as $\zeta^{(t)}(c,d) = \alpha^{(t)}(c)\beta^{(t)}(c,d)$, where

$$\alpha^{(t)}(c) = \mathbf{1}\{c \in \Delta_t\}/|\Delta_t|\ , \quad \text{and} \quad \beta^{(t)}(c,d) = \begin{cases} \gamma(c) & , \quad d = \hat{d}_1^{(t)}(c)\ , \\ 1-\gamma(c) & , \quad d = \hat{d}_2^{(t)}(c)\ , \\ 0 & , \quad \textit{otherwise}\ , \end{cases}$$

and $\sum_{c\in\mathscr{C}} \alpha^{(t)}(c) = 1$, $\sum_{d\in\mathscr{D}_c} \beta^{(t)}(c,d) = 1$, $\forall\ c \in \mathscr{C}$. Let $\boldsymbol{\alpha}^{(t)} = \big(\alpha^{(t)}(c)\big) \geq 0$ and $\boldsymbol{\beta}^{(t)} = \big(\beta^{(t)}(c,d)\big) \geq 0$. At each step t, TTTS-C uses $\boldsymbol{\alpha}^{(t)}$ and $\boldsymbol{\beta}^{(t)}$ to balance the simulation budget allocated to contexts and alternatives, respectively. The vector $\boldsymbol{\beta}^{(t)}$ balances the trade-off between exploiting the most promising best alternatives and exploring others according to their posterior probability of being the optimal design.

The expected time required to sample $\hat{\boldsymbol{\theta}}_2^{(t)}$ from the posterior distribution $\Pi_{t-1}(\cdot)$ increases as the number of contexts $|\mathscr{C}|$ grows. However, we can show that the computational complexity of TTTS-C grows sub-linearly in $|\mathscr{C}|$, which results in negligible additional computation asymptotically compared to the TTTS. Suppose the posterior belief of $\Theta_{(c,d^*)}$ is upper bounded by $(1-\delta)$ for each context c till time step T, where $\delta > 0$ is small enough. Also, assume that t is sufficiently large such that with high probability, the first sample $\hat{\boldsymbol{\theta}}_1^{(t)}$ leads to $\hat{d}_1^{(t)}(\cdot) = d^*(\cdot)$. Then, the probability of stopping re-sampling from the posterior distribution in each attempt is approximately $1-(1-\delta)^{|\mathscr{C}|}$, resulting in an expected time of repetitions of $1/\big(1-(1-\delta)^{|\mathscr{C}|}\big) \approx 1+(1-\delta)^{|\mathscr{C}|}$ and asymptotic computational complexity of order $O(|\mathscr{C}|+|\mathscr{C}|(1-\delta)^{|\mathscr{C}|})$. We note that this complexity is derived under a fixed budget setting. Our focus is on the number of $|\mathscr{C}|$ rather than δ.

The proposed TTTS-C is proven to be consistent in Theorem 1, implying that the estimated best designs will converge to the true best designs when the sampling budget tends to infinity.

Theorem 1 For any hyperparameter $\boldsymbol{\gamma} > 0$, the proposed TTTS-C is consistent, i.e., $\sum_{t=1}^{T} \xi_t(c,d) \to \infty$, *a.s.*, as $T \to \infty$, $\forall\ c \in \mathscr{C}$, $\forall\ d \in \mathscr{D}_c$.

We leave the proof of Theorem 1 to future work. An immediate consequence is that under TTTS-C, the Bayesian decision $\hat{d}_B(\cdot)$ is a consistent estimation of $d^*(\cdot)$, if $|\mu_{(c,d)} - \mu_{(c',d')}| \geq \delta > 0$ for $(c,d) \neq (c',d')$.

3.1 Posterior Large Deviations Ratios

In this section, we establish posterior large deviation ratios (pLDR) for general identifiable distribution families. Denote $D(\theta^*\|\theta) := \mathbb{E}_{Y\sim p(\cdot|\theta^*)}[\lambda(\theta^*,\theta;Y)]$ as the Kullback–Leibler (KL) divergence between two probability measures with parameters θ^* and θ. Let $D_{\bar{\Psi}_T}(\boldsymbol{\theta}^*\|\boldsymbol{\theta}) := \sum_{c\in\mathscr{C},d\in\mathscr{D}_c} \bar{\psi}_T(c,d) D(\theta^*_{(c,d)}\|\theta_{(c,d)})$ be the overall KL divergence with respect to weights $\bar{\psi}_T = \sum_{t=1}^{T} \psi_t/T$. We make a regularity assumption on the KL divergence.

Assumption 2 The KL divergence $D(\theta^*\|\theta)$ is well defined and continuously differentiable with respect to θ.

Theorem 2 establishes pLDR for adaptive sampling policies, which are probably randomized. It is a general result for all identifiable parameterized distributions. It implies that the posterior belief on any

open set $\tilde{\Theta}$ bounded away from $\boldsymbol{\theta}^*$ decays exponentially, at a rate that depends on the distance between $\boldsymbol{\theta}^*$ and the closest point to it in $\tilde{\Theta}$.

Theorem 2 If Assumptions 1-2 hold, then for any adaptive sampling policy $\boldsymbol{\psi}$ and open set $\tilde{\Theta} \subseteq \Theta$,

$$\lim_{T\to\infty} -\frac{1}{T}\ln \Pi_T(\tilde{\Theta}) - \inf_{\boldsymbol{\theta}\in\tilde{\Theta}} D_{\bar{\psi}_T}(\boldsymbol{\theta}^* \| \boldsymbol{\theta}) = 0, \quad a.s..$$

We leave the proof of Theorem 2 to future work. Theorem 2 provides an asymptotic approximation to (1), which is impossible to maximize due to the dependence of the posterior belief on random samples. To be specific, note that maximizing $\Pi_T\left(\bigcap_{c\in\mathscr{C}} \Theta^{(c,d^*(c))}\right)$ is equivalent to minimizing $\Pi_T\left(\bigcup_{c\in\mathscr{C}}\bigcup_{d\in\mathscr{D}_c\setminus\{d^*(c)\}} \Theta_{(c,d)}\right)$, where $\Theta_{(c,d)} = \{\boldsymbol{\theta}\in\Theta : \mu_{(c,d)} \geq \mu_{(c,d^*(c))}\}$, $\forall\ c\in\mathscr{C}$, $d\in\mathscr{D}_c\setminus\{d^*(c)\}$. Following Theorem 2 and

$$\max_{c\in\mathscr{C}} \max_{d\in\mathscr{D}_c\setminus\{d^*(c)\}} \Pi_T\left(\Theta_{(c,d)}\right) \leq \Pi_T\left(\bigcup_{c\in\mathscr{C}}\bigcup_{d\in\mathscr{D}_c\setminus\{d^*(c)\}} \Theta_{(c,d)}\right) \leq \sum_{c\in\mathscr{C}} |\mathscr{D}_c| \cdot \max_{c\in\mathscr{C}} \max_{d\in\mathscr{D}_c\setminus\{d^*(c)\}} \Pi_T\left(\Theta_{(c,d)}\right)\ ,$$

we have

$$\lim_{T\to\infty} -\frac{1}{T}\ln \Pi_T\left(\bigcup_{c\in\mathscr{C}}\bigcup_{d\in\mathscr{D}_c\setminus\{d^*(c)\}} \Theta_{(c,d)}\right) - \min_{c\in\mathscr{C}} \min_{d\in\mathscr{D}_c\setminus\{d^*(c)\}} \inf_{\boldsymbol{\theta}\in\Theta_{(c,d)}} D_{\bar{\psi}_T}(\boldsymbol{\theta}^* \| \boldsymbol{\theta}) = 0, \quad a.s.\ . \tag{2}$$

Therefore, the posterior belief on $\bigcup_{c\in\mathscr{C}}\bigcup_{d\in\mathscr{D}_c\setminus\{d^*(c)\}} \Theta_{(c,d)}$ decays at a rate that is only relevant to sampling ratio $\bar{\psi}_T$ asymptotically. Note that $\Pi_T(\bigcup_{c\in\mathscr{C}}\bigcup_{d\in\mathscr{D}_c\setminus\{d^*(c)\}} \Theta_{(c,d)})$ differs from the probability of incorrect selection (PICS) defined in Gao et al. (2019) for the context-dependent R&S problem in two ways. First, the former posterior probability does not rely on any selection policy. Second, it is defined a posteriori while PICS is defined a priori. However, equation (2) implies it decreases at a similar rate as PICS. We consider an underlying static optimization problem as follows:

$$\begin{aligned} \Gamma^* := \max_{\psi\geq 0} \quad & \min_{c\in\mathscr{C}} \min_{d\in\mathscr{D}_c\setminus\{d^*(c)\}} \inf_{\boldsymbol{\theta}\in\Theta_{(c,d)}} D_{\psi}(\boldsymbol{\theta}^* \| \boldsymbol{\theta}) \\ s.t. \quad & \textstyle\sum_{c\in\mathscr{C}} \sum_{d\in\mathscr{D}_c\setminus\{d^*(c)\}} \psi(c,d) = 1\ , \end{aligned} \tag{3}$$

where

$$\begin{aligned} \inf_{\boldsymbol{\theta}\in\Theta_{(c,d)}} D_{\psi}(\boldsymbol{\theta}^* \| \boldsymbol{\theta}) &= \inf_{\mu_{(c,d)}\geq\mu_{(c,d^*(c))}} \sum_{c'\in\mathscr{C}} \sum_{d'\in\mathscr{D}_{c'}} \psi(c',d') D(\theta^*_{(c',d')} \| \theta_{(c',d')}) \\ &= \inf_{\mu_{(c,d)}\geq\mu_{(c,d^*(c))}} \psi(c,d) D(\theta^*_{(c,d)} \| \theta_{(c,d)}) + \psi(c,d^*(c)) D(\theta^*_{(c,d^*(c))} \| \theta_{(c,d^*(c))}) \\ &=: G_{(c,d)}(\psi(c,d^*(c)), \psi(c,d))\ . \end{aligned}$$

The bivariate function $G_{(c,d)}(\cdot,\cdot)$ reflects the complexity of differentiating alternative d with $d^*(c)$ under context c. According to (2), Γ^* is an asymptotic upper bound on the exponential rate of convergence of the posterior belief under any adaptive sampling policy. Following $G_{(c,d)}(hx,hy) = hG_{(c,d)}(x,y)$, $\forall\ h\geq 0$, $x,y\in\mathbb{R}^+$, the static optimization problem (3) can be rewritten as the following optimization problem:

$$\begin{aligned} \Gamma^* = \max_{\boldsymbol{\alpha},\boldsymbol{\beta}\geq 0} \quad & \min_{c\in\mathscr{C}} \alpha(c) \min_{d\in\mathscr{D}_c\setminus\{d^*(c)\}} G_{(c,d)}\left(\beta(c,d^*(c)), \beta(c,d)\right) \\ s.t. \quad & \sum_{c\in\mathscr{C}} \alpha(c) = 1, \quad \sum_{d\in\mathscr{D}_c} \beta(c,d) = 1, \quad \forall c\in\mathscr{C}\ . \end{aligned} \tag{4}$$

The Karush-Kuhn-Tucker (KKT) conditions provide necessary and sufficient conditions for the optimal solutions to the optimization problem (4), leading to

$$\sum_{d\in\mathscr{D}_c\backslash\{d^*(c)\}} \frac{\partial G_{(c,d)}(\beta(c,d^*(c)),\beta(c,d))/\partial x_1}{\partial G_{(c,d)}(\beta(c,d^*(c)),\beta(c,d))/\partial x_2} = 1, \quad \forall c\in\mathscr{C}\ , \tag{5}$$

$$\alpha(c)G_{(c,d)}(\beta(c,d^*(c,d)),\beta(c,d)) = z, \quad \forall c\in\mathscr{C}, d\in\mathscr{D}_c\backslash\{d^*(c)\},\ \exists\, z\in\mathbb{R}\ , \tag{6}$$

where $\partial/\partial x_i$ denotes the derivative operator with respect to the i-th argument. Equation (5) balances the simulation budget allocated to the best alternative and other alternatives for each context, and equation (6) establishes certain balances between the simulation budget allocated to context-design pairs from a single context and from different contexts. It's worth noting that z coincides with Γ^* at the optimum, and that our conditions coincide with those in Gao et al. (2019) by eliminating z.

It can be shown that TTTS-C allocates a $\gamma(c)$ fraction of the simulation budget to measure $(c,d^*(c))$ under context c in the long run, i.e., $\lim_{T\to\infty}\bar{\psi}_T(c,d^*(c))/\sum_{d\in\mathscr{D}_c}\bar{\psi}_T(c,d)=\gamma(c)$. Therefore, the asymptotic sampling ratio $\bar{\boldsymbol{\psi}}_T$ of TTTS-C may not exactly achieve the optimal solutions to the optimization problem (4). To characterize the asymptotic properties of TTTS-C for a fixed hyperparameter $\boldsymbol{\gamma}$, we consider a modified optimization problem by adding a constraint $\beta(c,d^*(c))=\gamma(c)$ to the optimization problem (4), which captures the largest rate of convergence of the posterior distribution when the conditional sampling ratio for $(c,d^*(c))$ is fixed as $\gamma(c)$. Equivalently, the static optimization problem (4) becomes

$$\Gamma^*_{\boldsymbol{\gamma}} := \max_{\boldsymbol{\alpha},\boldsymbol{\beta}\geq 0}\min_{c\in\mathscr{C}}\alpha(c)\min_{d\in\mathscr{D}_c\backslash\{d^*(c)\}} G_{(c,d)}(\gamma(c),\beta(c,d)), \quad c\in\mathscr{C}\ . \tag{7}$$

The asymptotic sampling ratio $\bar{\boldsymbol{\psi}}_T$ of TTTS-C can be shown to satisfy (6), implying that TTTS-C asymptotically maximizes the exponential convergence rate given by (7). We leave the proof to future work.

3.2 Tuning of hyperparameter γ

TTTS-C requires a smart choice of the hyperparameter $\boldsymbol{\gamma}$. The optimal choice of $\boldsymbol{\gamma}$, which guarantees TTTS-C to be asymptotically optimal, relies on the true distribution parameters and is unknown a priori. A natural choice for the parameter is a predetermined constant that yields satisfactory allocation for all instances. In the case where there is only one context, i.e., $\mathscr{C}=\{c\}$, Russo (2020) suggests using $\gamma(c)=1/2$. This simple choice leads to a near-optimal rate of convergence, at least half of the optimal rate. Numerical experiments show that such a choice also performs robustly well in CR&S. In this work, we also provide a modification of TTTS-C by allowing $\boldsymbol{\gamma}$ to be learned sequentially as simulation samples are gathered. Algorithm 1 presents the pseudo-code of the TTTS-C algorithm with tuning hyperparameter $\boldsymbol{\gamma}$. At each step t, the hyperparameter $\boldsymbol{\gamma}$ is updated based on the estimated optimal conditional sampling ratios for context-dependent best alternatives if a user-specified updating rule $\mathscr{R}$ is satisfied. Our numerical experiments show that updating $\boldsymbol{\gamma}$ at certain predetermined steps suffices to deliver asymptotic optimality.

4 NUMERICAL EXPERIMENTS

In this section, we compare our proposed TTTS-C method with existing methods for the context-dependent selection problem under both Gaussian and non-Gaussian settings. Both conjugate prior distributions and non-conjugate prior distributions such as the uninformative prior are tested in the experiments. We compare TTTS-C with the following three algorithms for Gaussian settings:

- *Equal allocation* (EA): This method assigns the same number of samples to each context and to each alternative within a context.
- *Contextual optimal computing budget allocation* (C-OCBA, see Du et al. 2022): This algorithm allocates simulation replications to sequentially balance the two sides of equations (5) and (6).

Algorithm 1 TTTS-C with tuning hyperparameter $\boldsymbol{\gamma}$.

Input: updating rule $\mathscr{R}$ for $\boldsymbol{\gamma}$
Initialize $t=0$, $\boldsymbol{\gamma} \in (0,1)^{\mathscr{C}}$, Π_0
repeat
  if $\mathscr{R}$ is satisfied **then**
    Consistently estimate $\hat{\boldsymbol{\theta}}$;
    Plug in $\hat{\boldsymbol{\theta}}$ to estimate $\hat{G}_{(c,d)}$ and the context-dependent best alternative $\hat{d}(c)$;
    Solve $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ from (5) and (6) with $G_{(c,d)} = \hat{G}_{(c,d)}$ and $d^*(c) = \hat{d}(c)$;
    Update $\gamma(c)$ with $\hat{\beta}(c, \hat{d}(c))$;
  end if
  Decide (C_t, D_t) by TTTS-C with $\boldsymbol{\gamma}$ and Π_t;
  Observe $Y^{(t)}_{(c,d)}$ and update posterior distribution Π_{t+1};
  $t \leftarrow t+1$;
until simulation budget is exhausted

- *Dynamic sampling scheme for context-dependent optimization* (DSCO, see Li et al. 2020): This method chooses the context-alternative pair that has the highest value function, which quantifies the increase in the probability of making the correct final decision with an additional sample.

For non-Gaussian settings, contextual R&S is rarely studied. Therefore, we only compare TTTS-C with EA.

In our analysis, we focus on PCS as a measure of the probability of correct selection across all contexts. We also measure the minimum and average probability of correct selection in each context, denoted by PCSM and PCSE, respectively. To be specific, the efficiency of sampling policies is evaluated using three metrics:

$$\begin{aligned} \text{PCS} &= \mathrm{P}\left(\hat{d}(c) = d^*(c), \forall\, c \in \mathscr{C}\right), \\ \text{PCSM} &= \min_{c \in \mathscr{C}} \mathrm{P}(\hat{d}(c) = d^*(c)), \\ \text{PCSE} &= \sum_{c \in \mathscr{C}} \mathrm{P}(\hat{d}(c) = d^*(c))/|\mathscr{C}|. \end{aligned}$$

We plot these metrics using a logarithmic scale to clearly show the exponential rate at which PCS_ increases, where PCS_ denotes anyone of the three metrics. We test three versions of TTTS-C, including TTTS-C-coin, TTTS-C-true, and TTTS-C-update. TTTS-C-coin uses $\gamma(c) = 1/2$ for all c. TTTS-C-true is implemented with the optimal choice of $\boldsymbol{\gamma}$. TTTS-C-update starts like TTTS-C-coin but updates $\boldsymbol{\gamma}$ according to Algorithm 1 at predetermined time steps κ^l, $l = 1, 2, 3, \ldots$. We implement TTTS-C-update with $\kappa = 5$.

4.1 Gaussian Settings

Known variance. In the first experiment, we test TTTS-C under a Gaussian setting with known variances. A synthetic instance is constructed with $|\mathscr{C}| = 10$ and $|\mathscr{D}_c| = 10$, $\forall\, c \in \mathscr{C}$. The mean performance and variance of alternatives in each context are randomly chosen from Table 1, with a perturbation uniform in $[0, 0.1]$ added to each alternative. Once the distribution parameters are generated, they are fixed for all macro-replications.

All compared algorithms use the true value of variances so that the only parameter is the unknown mean. The prior distribution of the mean parameters is set as $N(0, 100)$, which is approximately uninformative. A warm-up period assigns 10 initial samples equally to each context-design pair.

Figure 1 illustrates the three metrics. Consistent with our analysis, the proposed TTTS-C algorithm outperforms other benchmarks significantly in terms of PCS. TTTS-C algorithms yield a satisfactory probability of correct selection larger than 0.95 when samples are exhausted, while other algorithms end

Table 1: A library of true distribution parameters.

Mean	Standard deviation
$(0.3,0.3,0.3,0.3,0.3,0.3,0.3,0.3,0.3,0.5)$ $(-0.5,-0.5,-0.5,-0.5,-0.5,-0.5,-0.25,0,0.25,0.5)$ $(0.05,0.1,0.15,0.2,0.25,0.3,0.35,0.4,0.45,0.5)$ sort(10 replications Unif($[-0.5,0.5]$))	0.6 to 1.5 with increment 0.1 $(1,1,1,1,1,1,1,1,1,1)$ 1.5 to 0.6 with decrement 0.1

up with PSC_ lower than 0.9 ($\log(1-0.9) \approx -2.3$). It is surprising that TTTS-C-coin achieves a similar performance as the other two versions of TTTS-C that use the global optimal sampling ratios, within a limited budget of 60,000 samples. In terms of PCSM and PCSE, DSCO and C-OCBA performs slightly better than TTTS-C at the first few samples when PCS_ is less than 0.5. However, their performances fall behind TTTS-C as more sampling budgets become available. This phenomenon is partially attributed to the different way in which sampling policies allocate samples to contexts. While C-OCBA and DSCO always allocate the sample to the vaguest context, in which the alternatives are least distinguished, TTTS-C adopts a randomized strategy that allows for more exploration across contexts. As a result, the compared algorithms are slightly more efficient at initial steps in terms of PCSM by focusing sampling effort on the vaguest context. However, they are inferior to TTTS-C in the long run or in terms of PCS.

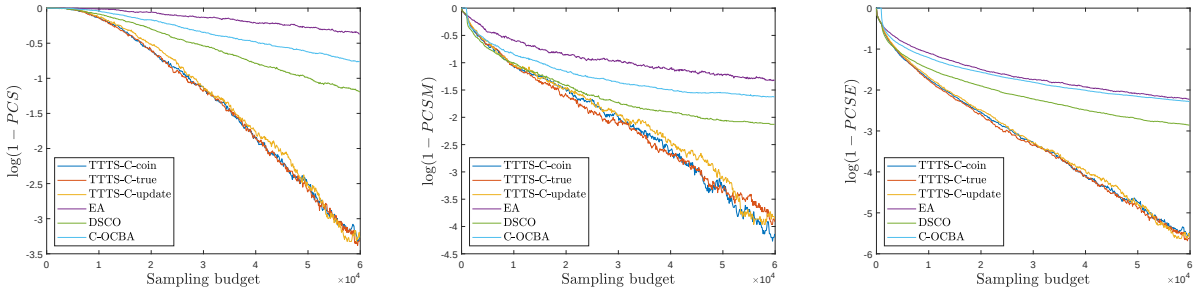

Figure 1: Log probability of incorrect selection under the Gaussian setting with known variance, estimated via 1,000 macro-replications.

The mean computing times when algorithms first achieve the probability of incorrect selection of 0.5, 0.25, 0.1, and 0.05 are reported in Table 2. The hyphens represent instances where the corresponding algorithm did not achieve the given PCSE level within 60,000 samples. Although EA and C-OCBA are more computationally economic than TTTS-C, they demand a lot more simulation replications and do not attain the same level of PCS_ with the fixed budget. We are also interested in the mean computing time of TTTS-C in achieving a certain level of PCS_ with respect to the number of contexts. We generate three more instances with 5, 20, and 40 contexts respectively, where the underlying distributions are given the same way as before. Table 2 reports the mean computing time in PCSE surpassing 0.9 within 60,000 samples. Our results show that the computing time in need grows sub-linearly in the number of contexts.

Unknown variance. Second, we test TTTS-C under a Gaussian setting with unknown variance. A synthetic instance of larger size is formed with $|\mathscr{C}| = 30$ and $|\mathscr{D}_c| = 30$, $\forall\ c \in \mathscr{C}$.

The true distribution of alternative d under context c is $N(\mu_{(c,d)}, \sigma^2_{(c,d)})$ where $\sigma^2_{(c,d)} = \eta_{(c,d)}$ is the nuisance parameter. The mean performance $\mu_{(c,d)}$ and standard deviation $\sigma_{(c,d)}$ are generated from $N(0,5^2)$ and Unif($[4,6]$), respectively, for each context-design pair. To facilitate implementation, the prior distribution is chosen as the Normal-Gamma distribution with parameters $m_{0,(c,d)}, n_{0,(c,d)}, \alpha_{0,(c,d)}, \beta_{0,(c,d)}$, the conjugate prior of Gaussian distribution with unknown mean and variance, such that

$$\mu_{(c,d)}|\sigma^2_{(c,d)} \sim N\left(m_{0,(c,d)}, \frac{\sigma^2_{(c,d)}}{n_{0,(c,d)}}\right) \quad \text{and} \quad 1/\sigma^2_{(c,d)} \sim \text{Gamma}\left(\alpha_{0,(c,d)}, \beta_{0,(c,d)}\right).$$

Table 2: Columns 2-5 represent the mean computing time (in seconds) of algorithms when PCSE first surpasses given levels. Columns 6-9 represent the mean computing time (in seconds) when PCSE first surpasses 0.9 for an increasing number of contexts. Data is estimated via 1,000 macro-replications.

	PCSE Level				**Contexts**			
	0.5	0.75	0.9	0.95	5	10	20	40
TTTS-C-coin	0.07	0.31	0.83	1.32	0.50	0.83	2.23	2.86
TTTS-C-true	0.07	0.30	0.77	1.67	0.43	0.77	2.30	2.91
TTTS-C-update	2.12	2.79	3.68	4.13	2.06	3.68	10.72	34.63
EA	0.07	0.32	-	-	0.81	-	-	-
DSCO	0.18	1.95	8.34	-	2.45	8.34	-	-
C-OCBA	0.03	0.22	-	-	0.47	-	-	-

We choose $m_{0,(c,d)} = 0, n_{0,(c,d)} = 10^{-3}, \alpha_{0,(c,d)} = 1, \beta_{0,(c,d)} = 10$ such that the prior is approximately non-informative. Similarly, 10 samples are allocated to each context-design pair at the beginning.

In this experiment, TTTS-C algorithms result in significantly higher PCS and PCSE than other algorithms. In terms of PCSM, TTTS-C algorithms perform similarly to DSCO and C-OCBA. This is attributed to the insufficient sampling budget, as PCSM is less than 0.6 for all algorithms within 300,000 samples. Also notice that TTTS-C has a tendency to surpass DSCO and C-OCBA as samples are gathered, as shown in Figure 2.

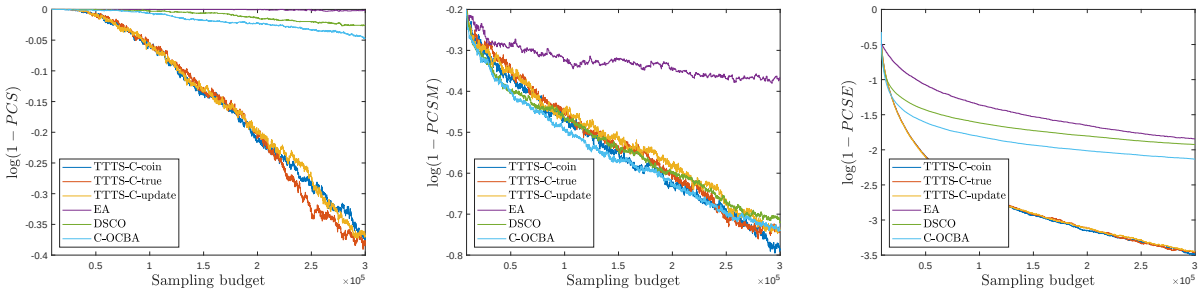

Figure 2: Log probability of incorrect selection under the Gaussian setting with unknown variance, estimated via 2,000 macro-replications.

4.2 Non-Gaussian Setting

Selecting the most durable product for diversified demands. Suppose a capacitor manufacturer has recently developed a variety of production processes with different costs. They plan to offer differentiated products within several price ranges while reducing the number of production lines. A product testing phase is conducted to select the most durable alternatives in each range to be put into production.

In our experiment, we generate 20 production processes that are divided into 5 ranges according to the unit cost. These ranges are understood as contexts. The manufacturer wants to select in each range a design that produces the most durable capacitors, where durability is defined as the mean time before a capacitor fails. The manufacturer employs sequential accelerated life tests (ALT) to evaluate the durability of capacitors. In each round, a production process is selected and a capacitor is produced accordingly. The capacitor is then exerted with extreme voltage and thermal conditions in the ALT and the time-to-failure is recorded if a failure occurs before the end of the test. The objective is to identify the best production process in each cost range within a limited budget of T ALT trials.

As commonly used in survival and reliability analysis (Kapur and Lamberson 1977), the lifespan of capacitors is modeled by the two-parameter Weibull distribution $W(k,\rho)$ with shape parameter k

and scale parameter ρ. The probability density with respect to the Lebesgue measure is $f_W(y;\rho,k) = k/\rho \cdot (y/\rho)^{k-1} \cdot e^{-(y/\rho)^k}$ and thus the mean time-to-failure is $\mu = \rho\,\text{Gamma}(1+1/k)$, where $\text{Gamma}(\cdot)$ is the Gamma function. Let Y be the observed failure time during the test. Then it follows a Weibull distribution truncated at time τ when the test is finished, with the likelihood function

$$p(Y|k,\rho) = f_W(Y;\rho,k)^{\mathbf{1}\{Y<\tau\}} \times S_W(\tau;\rho,k)^{\mathbf{1}\{Y=\tau\}},$$

where $S_W(\tau;\rho,k) := \int_\tau^\infty f_W(y;\rho,k)dy$ is the survival function. Note that the mathematical expectation of Y is not μ since the ALT lasts only for a finite period of time. Therefore, mean-based R&S procedures do not directly apply. To be coherent with our framework, we reparameterize the Weibull distribution by $\mu = \rho\,\text{Gamma}(1+1/k)$ and $\eta = k$. Then the likelihood can be rewritten as

$$p(Y|\mu,\eta) = f_W(Y;\mu/\text{Gamma}(1+1/k),\eta)^{\mathbf{1}\{Y<\tau\}} \times S_W(\tau;\mu/\text{Gamma}(1+1/k),\eta)^{\mathbf{1}\{Y=\tau\}}.$$

In a synthetic instance, $\tau = 120$ is fixed for all ALTs, $\mu_{(c,d)}$ and $\eta_{(c,d)}$ are generated from $\text{Unif}([90,110])$ and $\text{Unif}([2,4])$, and the prior of $(\mu_{(c,d)},\eta_{(c,d)})$ is chosen as $\text{Unif}([0,200]\times[0,10])$. As always, 10 initial samples are equally allocated to each context-design pair. Figure 3 demonstrates that TTTS-C-coin has a significant advantage over EA.

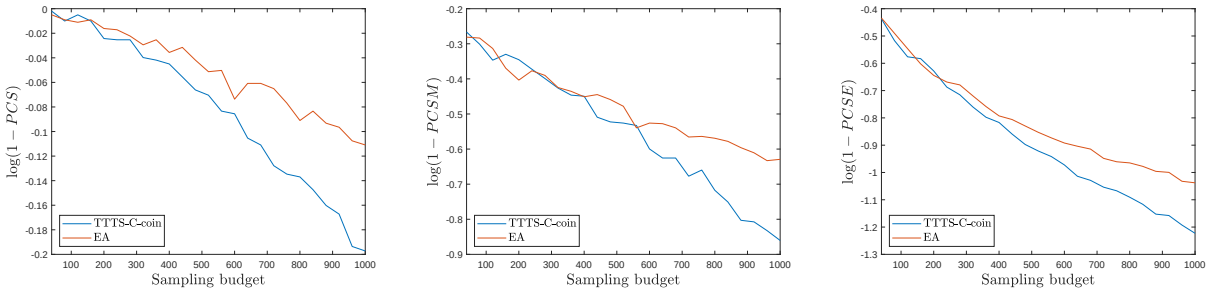

Figure 3: Log probability of incorrect selection under the truncated Weibull setting, estimated via 1,000 macro-replications.

5 CONCLUSION AND FUTURE RESEARCH

In this work, we consider CR&S from a Bayesian perspective. The goal is to efficiently allocate costly samples to correctly identify the best alternative for each independent context. We establish general posterior large deviation ratios for adaptive sampling policies and propose a randomized sampling policy TTTS-C based on Thompson sampling by forcing sufficient exploration. Experiments corroborate the asymptotic results and also demonstrate the capability of TTTS-C with finite samples. Future research may include an analysis of the asymptotic rate optimality of the proposed algorithm and real-world applications.

ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grants 72250065, 72022001, and 71901003, and by the China Scholarship Council (CSC) under Grant CSC202206010152.

REFERENCES

Bechhofer, R. E. 1954. "A Single-Sample Multiple Decision Procedure for Ranking Means of Normal Populations with known Variances". *The Annals of Mathematical Statistics* 25(1):16–39.

Bechhofer, R. E., T. J. Santner, and D. M. Goldsman. 1995. "Design and Analysis of Experiments for Statistical Selection, Screening and Multiple Comparisons". *Computational Statistics & Data Analysis* 22(5):568–567.

Cakmak, S., E. Zhou, and S. Gao. 2021. "Contextual Ranking and Selection with Gaussian Processes". In *Proceedings of the 2021 Winter Simulation Conference*, edited by S. Kim, B. Feng, K. Smith, S. Masoud, Z. Zheng, C. Szabo, and M. Loper, 1–12. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Chen, C.-H., and L. H. Lee. 2011. *Stochastic Simulation Optimization: An Optimal Computing Budget Allocation*, Volume 1. World scientific.

Chen, C.-H., J. Lin, E. Yücesan, and S. E. Chick. 2000. "Simulation Budget Allocation for Further Enhancing the Efficiency of Ordinal Optimization". *Discrete Event Dynamic Systems* 10:251–270.

Du, J., S. Gao, and C.-H. Chen. 2022. "Rate-Optimal Contextual Ranking and Selection". https://arxiv.org/abs/2206.12640, accessed 25th June 2022.

Gao, S., J. Du, and C.-H. Chen. 2019. "Selecting the Optimal System Design under Covariates". In *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)*, 547–552.

Hong, L. J., W. Fan, and J. Luo. 2021. "Review on Ranking and Selection: A New Perspective". *Frontiers of Engineering Management* 8(3):321–343.

Hong, L. J., and B. L. Nelson. 2006. "The Tradeoff Between Sampling and Switching: New Sequential Procedures for Indifference-Zone Selection". *INFORMS Journal on Computing* 18(1):20–36.

Kapur, K. C., and L. R. Lamberson. 1977. *Reliability in Engineering Design*. New York: John Wiley & Sons, Inc.

Kim, E. S., R. S. Herbst, I. I. Wistuba, J. J. Lee, G. R. Blumenschein, Jr., A. Tsao, D. J. Stewart, M. E. Hicks, J. Erasmus, Jr., S. Gupta, C. M. Alden, S. Liu, X. Tang, F. R. Khuri, H. T. Tran, B. E. Johnson, J. V. Heymach, L. Mao, F. Fossella, M. S. Kies, V. Papadimitrakopoulou, S. E. Davis, S. M. Lippman, and W. K. Hong. 2011. "The BATTLE Trial: Personalizing Therapy for Lung Cancer". *Cancer Discovery* 1(1):44–53.

Kim, S.-H., and B. L. Nelson. 2006. "Chapter 17 Selecting the Best System". In *Simulation*, edited by S. G. Henderson and B. L. Nelson, Volume 13 of *Handbooks in Operations Research and Management Science*, 501–534. Elsevier.

Li, H., H. Lam, Z. Liang, and Y. Peng. 2020. "Context-Dependent Ranking and Selection under a Bayesian Framework". In *Proceedings of the 2020 Winter Simulation Conference*, edited by K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, and Z. Zheng, 2060–2070. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Li, H., H. Lam, and Y. Peng. 2022. "Efficient Learning for Clustering and Optimizing Context-Dependent Designs". *Operations Research* 0(0).

Li, X., X. Zhang, and Z. Zheng. 2018. "Data-Driven Ranking and Selection: High-Dimensional Covariates and General Dependence". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, 1933–1944. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Miao, S., and X. Chao. 2022. "Online Personalized Assortment Optimization with High-Dimensional Customer Contextual Data". *Manufacturing & Service Operations Management* 24(5):2741–2760.

Peng, Y., E. K. Chong, C.-H. Chen, and F. Michael. 2018. "Ranking and Selection as Stochastic Control". *IEEE Transactions on Automatic Control* 63(8):2359—2373.

Russo, D. 2020. "Simple Bayesian Algorithms for Best-Arm Identification". *Operations Research* 68(6):1625–1647.

Russo, D., B. V. Roy, A. Kazerouni, I. Osband, and Z. Wen. 2018. "A Tutorial on Thompson Sampling". *Foundations and Trends® in Machine Learning* 11(1):1–96.

Shen, H., L. J. Hong, and X. Zhang. 2021. "Ranking and Selection with Covariates for Personalized Decision Making". *INFORMS Journal on Computing* 33(4):1500–1519.

Woerndl, W., C. Schueller, and R. Wojtech. 2007. "A Hybrid Recommender System for Context-aware Recommendations of Mobile Applications". In *2007 IEEE 23rd International Conference on Data Engineering Workshop*, 871–878. Institute of Electrical and Electronics Engineers, Inc.

AUTHOR BIOGRAPHIES

XINBO SHI is a Ph.D. candidate in Guanghua School of Management at Peking University, Beijing, China. His research interest includes simulation optimization and machine learning. His email address is 2101111021@stu.pku.edu.cn.

YIJIE PENG is an Associate Professor in Guanghua School of Management at Peking University, Beijing, China. His research interests include stochastic modeling and analysis, simulation optimization, machine learning, data analytics, and healthcare. His email address is pengyijie@pku.edu.cn.

GONGBO ZHANG is a Ph.D. candidate in Guanghua School of Management at Peking University, Beijing, China. His research interests include stochastic modeling and analysis, simulation optimization and reinforcement learning. His email address is gongbozhang@pku.edu.cn.