Proceedings of the 2024 Winter Simulation Conference
H. Lam, E. Azar, D. Batur, S. Gao, W. Xie, S. R. Hunter, and M. D. Rossetti, eds.

THOMPSON SAMPLING PROCEDURES FOR RANKING AND SELECTION WITH PAIRWISE COMPARISONS AND BINARY OUTCOMES

Dongyang Li[1], Enver Yücesan[2], and Chun-Hung Chen[3]

[1]Dept. of Industrial Systems Eng. and Mgmt., National University of Singapore, SINGAPORE
[2]Technology and Operations Mgmt. Area, INSEAD, Fontainebleau, FRANCE
[3]Dept. of Systems Eng. and Operations Research, George Mason University, Fairfax, VA, USA

ABSTRACT

We consider ranking and selection (R&S) problems where the performance of competing designs or alternatives can only be assessed by evaluating two alternatives simultaneously with a binary outcome; the true performance of an alternative is quantified by its average probability of outperforming other alternatives. We discuss the challenges associated with applying conventional R&S techniques to this particular setting and propose heuristic algorithms based on Thompson sampling to overcome those difficulties. Through a series of simple numerical experiments, we assess the effectiveness of these heuristics and highlight open research questions.

1 INTRODUCTION

Ranking and selection (R&S) aims at identifying the best design among a possibly large, but finite, set of competing designs by simulating each design through a sufficient number of replications.

In this paper, we consider a specific variant of R&S where the performance of competing designs or alternatives can only be assessed by evaluating two alternatives simultaneously whereby the sampling outcome is binary; the true performance of a design is then defined as its average probability of outperforming the other competing designs. More specifically, given a set of k competing designs or alternatives with some unknown underlying ranking, our objective is to determine the top-ranked design in the set through pairwise comparisons. Identifying the top design through pairwise comparisons is a challenge encountered in many applications, including player ranking in games, sports tournaments, recommender systems, image-based search, public choice models such as voting schemes or decision rules in committees, and market research (Eriksson 2013). Other applications include selection in evolutionary algorithms, optimizing search engine result relevance, and preference elicitation in decision making (Groves and Branke 2019). Pairwise comparisons are also used in multi-criteria decision-making among alternatives as human judgment is impaired by the lack of transitivity due to such factors as threshold effect, disturbance in concentration or, simply, errors in input data (Xiao et al. 2023).

The method of paired comparisons, whose roots can be traced back to psychometrics, studies general preferences based on all possible pairwise comparisons between alternatives (e.g., A/B testing). However, as the number of alternatives, k, grows large, the consideration of all possible pairwise comparisons (i.e., total enumeration of $k(k-1)/2$ cases) becomes increasingly inefficient (Lu and Boutilier 2011). This complexity is magnified when operating under a constrained sampling budget, making it highly desirable to devise procedures that efficiently identify the most informative pairs to evaluate, rather than non-adaptively or exhaustively evaluating all possible comparisons. While the development of traditional R&S procedures has achieved huge success, in this paper, we discuss some of the difficulties in applying them to our setting where the pairwise comparison outcomes are modeled as Bernoulli random variables. To address these challenges, we resort to Thompson sampling, a popular approach in the realm of multi-armed bandit problems, and develop two easy-to-implement heuristic policies to guide the sampling process. The numerical results show that our simple heuristics deliver quite efficient and robust performance compared with several benchmarks, which highlights promising avenues for further research.

The remainder of the paper is organized as follows: Section 2 provides a light overview of the vast literature about R&S and pairwise comparison settings. The problem is formally formulated in Section 3 where the heuristic algorithms are also presented. Section 4 illustrates the performance of the proposed heuristics under various metrics. Promising venues for further research are discussed in Section 5.

2 LITERATURE REVIEW

R&S literature features two main streams of research: fixed precision and fixed budget. The former stream, which dates back to the pioneering works of Bechhofer (1954), Paulson (1964), and Rinott (1978), treats R&S as a statistical problem and aims at providing frequentist guarantees for the estimated best design. To achieve a pre-specified confidence level, an indifference-zone (IZ) parameter is often introduced to indicate the smallest difference in mean performance that is worth detecting; as a result, this stream is labeled as IZ procedure in the literature. More recent notable works in this stream include the fully sequential procedure of Kim and Nelson (2001), the IZ-free procedure of Fan et al. (2016), and the parallel computing procedure of Zhong et al. (2022).

The latter stream of literature addresses R&S from an efficient budget allocation perspective, which aims at allocating a finite simulation budget to competing designs to maximize some quality metric for the estimated best design. This stream can be further categorized into two sub-streams. The first sub-stream originates from the adaptive optimal computing budget allocation (OCBA) approach (Chen et al. 2000), which is shown to be asymptotically optimal using large deviations theory (Glynn and Juneja 2004). These procedures dynamically sample designs to learn and satisfy the optimal sampling ratios, with representative papers including Gao et al. (2017), Shin et al. (2018), and Chen and Ryzhov (2023). The second sub-stream adopts the Bayesian perspective and tends to myopically improve some quality metric. Notable examples include the Knowledge Gradient (KG) procedure of Frazier et al. (2008), the expected value of information (EVI) procedure of Chick et al. (2010), and the approximately optimal allocation policy (AOAP) of Peng et al. (2018). Apart from the two main sub-streams, there is some recent work borrowing the idea of Thompson sampling from multi-arm bandit problems to address R&S problems (e.g., Russo (2020), Peng and Zhang (2022)). In general, fixed-budget procedures are more efficient and necessitate a smaller simulation budget than fixed-precision procedures to achieve a comparable level of decision quality. However, they typically fail to offer rigorous statistical guarantees. A recent exhaustive review on traditional R&S can be found in Hong et al. (2021).

R&S with pairwise comparisons has been explored by Groves and Branke (2019). They assume a value-based sampling outcome (defined as the score with which design i beats design j) under Gaussian noise and adapt conventional myopic R&S procedures to this setting. Further, Xiao et al. (2023) investigate this setting from an OCBA perspective, and propose heuristic algorithms to adaptively balance the derived conditions of optimal sample allocation ratios. Li et al. (2020) consider identifying the most important nodes in a random network, where the importance of nodes is determined by the stationary distribution of the transition matrix; they adapt the AOAP for this setting to actively select the pair of nodes for sampling. The closest work to our binary outcome model is that of Priekule and Meisel (2017), which develop several heuristic algorithms based on the KG procedure.

The task of identifying the top alternative through pairwise comparisons has also been investigated by the computer science community under the concept of dueling bandits (Yue and Joachims 2009). Some works assume there exists a complete ranking of alternatives, consistent with all pairwise preference probabilities, aiming to identify the so-called Condorcet winner. Examples include Jamieson and Nowak (2011), Yue et al. (2012), and Mohajer et al. (2017). When no structural assumption is made for the pairwise preference, a universal Condorcet winner may not exist. In such cases, Wu and Liu (2016) devised the double Thompson sampling procedure to seek the Copeland winner, while Heckel et al. (2019) actively learn and rank alternatives according to Borda scores. These methodologies resemble fixed-precision R&S

procedures, which aim to minimize accumulated regret or provide frequentist guarantees, with provable sample complexity bounds. A comprehensive review of dueling bandits problems and algorithms is offered by Bengs et al. (2021).

3 PROBLEM FORMULATION AND ALGORITHM

Consider k competing designs or alternatives whose performance can only be assessed through pairwise comparison. The sampling outcome $Y_{ij} \in \{0,1\}$ of any pair of designs (i,j), where $i,j \in \{1,2,\dots,k\}$ and $i \neq j$, is independent and identically distributed Bernoulli random variables with parameter $0 \leq p_{ij} \leq 1$. Here, p_{ij} reflects the probability that design i outperforms design j, satisfying the condition $p_{ij} + p_{ji} = 1$. For each design i, the average winning probability is then given by

$$p_i = \frac{1}{k-1} \sum_{j \neq i} p_{ij}. \tag{1}$$

The best design is then defined as $i^* := \arg\max_{1 \leq i \leq k} p_i$, which can be estimated by

$$\widehat{i^*} = \underset{1 \leq i \leq k}{\arg\max} \frac{1}{k-1} \sum_{j \neq i} \widehat{p}_{ij},$$

where $\widehat{p}_{ij}$ is the empirical estimator obtained from simulation experiments. Given the symmetry of sampling outcomes for (i,j) and (j,i), we can simply focus on sampling (i,j) with $i < j$ and construct the estimator $\widehat{p}_{ij} = \frac{1}{n_{ij}} \sum_{l=1}^{n_{ij}} Y_{ij}^l$ and $\widehat{p}_{ji} = 1 - \widehat{p}_{ij}$.

Remark 1 In our problem setting, we do not assume any specific structure for the winning probabilities, except that $p_{ij} + p_{ji} = 1$. Consequently, "probability cycles" can occur, and a universal Condorcet winner i^* with $p_{i^*j} > 0.5, \forall j \neq i^*$ may not exist. In such settings, alternative definitions for the best design are considered, including the Borda winner, the Copeland winner, the random-walk winner, and the von Neumann winner (see Bengs et al. (2021) for detailed definitions). In this paper, we chose to adopt the Borda score, the average winning probability defined in (1).

The challenge is to identify the top-ranked item using significantly smaller computational effort than evaluating all possible pairwise comparisons. This is because as k grows large, the total enumeration of $k(k-1)/2$ comparisons becomes computationally infeasible. We therefore consider this problem from a fixed-budget perspective and assume a sampling budget $N = O(k(k-1)/2)$. This constraint allows for limited sampling of each pair, making accurate evaluation of all pairwise comparisons impractical. As outlined in Algorithm 1, our goal is to develop a procedure to iteratively select the next pair to sample and return the estimated best design in the end. The key ingredient is the sampling policy π, which aims at efficiently allocating the limited sampling budget to maximize the quality of the returned design. Two commonly used frequentist quality metrics in the literature are the probability of correct selection (PCS) and the expected opportunity cost (EOC), defined respectively as:

$$\text{PCS} := \mathbb{P}\left[\widehat{i^*} = i^*\right] \quad \text{and} \quad \text{EOC} := \mathbb{E}\left[p_{i^*} - p_{\widehat{i^*}}\right],$$

where the randomness in $\widehat{i^*}$ may arise from both the sampling outcomes and the sampling policy.

3.1 Bayesian Framework

Several efficient R&S procedures have been developed under the Bayesian framework (e.g., KG and AOAP). Within this framework, the unknown parameters are regarded as random variables and characterized by a prior distribution. In sequential procedures, the posterior or conditional distributions for the parameters drive the sampling process. In our problem settings, we follow the Bayesian framework whereby we adopt

Algorithm 1: Framework of the sampling budget allocation procedure

Setup: Take n_0 samples for each pair of design (i,j).
Update the information set as $\mathscr{E}_{t_0}$ with $t_0 = \frac{k(k-1)n_0}{2}$.
for $t = t_0+1,\dots,N$ **do**
  Identify the next pair $(i^t, j^t) = \pi(\mathscr{E}_{t-1})$ to sample according to some policy

$$\pi : \mathscr{E} \to \{(i,j) : 1 \le i < j \le k\}.$$

  Sample (i^t, j^t) to get Y_{ij}^t and update the information set $\mathscr{E}_t$.
end
Return: $\widehat{i^*} = \arg\max_{1\le i\le k} \frac{1}{k-1}\sum_{j\neq i}\widehat{p}_{ij}$

a Beta conjugate prior. More formally, for each p_{ij} with $1 \le i < j \le k$, we assume an independent prior distribution, given by

$$p_{ij} \sim \mathrm{Beta}(\alpha_{ij}^0, \beta_{ij}^0),$$

where the uniform prior with $\alpha_{ij}^0 = \beta_{ij}^0 = 1$ is a natural and commonly used choice. Let $\mathscr{E}_t := \{(i^l, j^l, Y_{ij}^l)\}_{l=1}^t$ represent the information set that records all the sampled pairs and outcomes up to step t. Denote by n_{ij}^t the number of total samples for pair (i,j) and W_{ij}^t as the count of "wins" where design i outperforms design j. According to the Bayes rule, the posterior distribution for p_{ij} given $\mathscr{E}_t$ is:

$$p_{ij}|\mathscr{E}_t \sim \mathrm{Beta}(\alpha_{ij}^t, \beta_{ij}^t),$$

where $\alpha_{ij}^t = \alpha_{ij}^0 + W_{ij}^t$ and $\beta_{ij}^t = \beta_{ij}^0 + n_{ij}^t - W_{ij}^t$. Symmetrically, we have $p_{ji}|\mathscr{E}_t \sim \mathrm{Beta}(\alpha_{ji}^t, \beta_{ji}^t)$ with $\alpha_{ji}^t = \beta_{ij}^t$ and $\beta_{ji}^t = \alpha_{ij}^t$. The posterior expectation (mean estimator) is given by $p_{ij}^t = \alpha_{ij}^t/(\alpha_{ij}^t + \beta_{ij}^t)$ and $p_i^t = \frac{1}{k-1}\sum_{j\neq i} p_{ij}^t$. The corresponding predictive distribution for $Y_{ij}|\mathscr{E}_t$ is Bernoulli with parameter p_{ij}^t.

Although the above Bayesian formulation is natural, adapting traditional R&S procedures to this setting presents significant challenges. For example, Priekule and Meisel (2017) employ the KG framework (Frazier et al. 2008) to develop efficient algorithms for pairwise sampling. The one-step look-ahead sampling policy is defined as:

$$(i^t, j^t) = \arg\max_{(i,j)} \mathrm{KG}(i,j) := \left\{ \mathbb{E}^{t-1}\left[\max_l p_l^t \mid \mathscr{E}_{t-1}, (i,j,Y_{ij}) \right] - \max_l p_l^{t-1} \right\},$$

where $\mathbb{E}^{t-1}$ denotes the expectation conditioned on $\mathscr{E}_{t-1}$. However, this KG implementation is somewhat inefficient as the Bernoulli setting frequently leads to zero values of KG for all pairs in multiple steps of the algorithm. For the multi-step look-ahead version, we also found it difficult to characterize the optimal number of steps to use. Some other well-established traditional R&S procedures (e.g., AOAP) try to myopically improve the Bayesian posterior PCS, i.e., $\mathbb{P}\left[\cap_{j\neq i^{*,t}} p_{i^{*,t}} > p_j \mid \mathscr{E}_t\right]$ with $i^{*,t}$ being the estimated best design at step t. However, in our Bayesian formulation, the posterior characterization of p_i becomes complicated as an average of $k-1$ Beta distributions, making the posterior PCS hard to approximate. Groves and Branke (2019) get around this challenge by approximating the posterior distribution as Gaussian, enabling the application of traditional R&S techniques. However, this approximation appears to be rough for binary outcomes, as illustrated by our empirical results. Furthermore, extending the OCBA ratio (Glynn and Juneja 2004) of traditional R&S problems to this pairwise comparison setting remains challenging even for Gaussian distributed sampling outcomes. Xiao et al. (2023) propose an approximate solution, which still falls short of achieving true optimality.

3.2 Top-Two Thompson Sampling

Given the above challenges, we develop heuristics based on another Bayesian scheme for adaptive allocation of sampling effort that has recently received significant attention from various communities. Thompson sampling is an algorithm for online decision problems where actions are taken sequentially in a manner that must provide a balance between exploiting what is known to maximize immediate performance and exploring or investing to accumulate new information that may improve future performance (Russo et al. 2018). Originally developed within the context of multi-arm bandit problems, Thompson sampling balances the exploration-exploitation trade-off through a stationary randomized strategy that assigns the sampling budget to arms (alternatives) in proportion to the posterior probability of being optimal. However, when the posterior probability is concentrated around a particular value, Thompson sampling acts almost as a pure exploitation algorithm, allocating an insufficient portion of the budget for the exploration of other possibly optimal alternatives. To restore the balance, Russo (2020) introduced Top-Two Thompson sampling (TTTS), a set of Bayesian algorithms for best-arm identification with provable properties.

Peng and Zhang (2022) illustrate the relevance of Thompson sampling for R&S where a good balance between exploration and exploitation is needed for efficient allocation of the sampling budget to each alternative. To this end, we adopt TTTS to iteratively select the next pair (i, j) to sample for enhancing the accuracy of the estimated best design. To accommodate the pairwise comparison setting, the top candidate is identified in the usual manner. Then we consider two policies to designate the runner-up. The first policy, *TTTS+Random*, selects the runner-up randomly from the remaining $k-1$ alternatives. In the second policy, *TTTS+MaxVar*, the runner-up is the design that exhibits the largest posterior variance in pairwise comparison probabilities with the top candidate based on the information collected thus far. While TTTS+Random provides a cost-effective, but less targeted, approach to exploration, TTTS+MaxVar aims to refine the top design's estimate by incorporating variability considerations. Both sampling policies are heuristic; as a result, there is no guarantee that one would dominate the other. We summarize the two sampling policies in Algorithm 2 and 3. The tuning probability parameter γ controls the allocation of the sampling budget between the top design and the others, where a robust choice of 0.5 is suggested by Russo (2020) in practice. In addition, when the posterior distribution becomes highly concentrated, it may take a long time to generate $i_2 \neq i_1$. To avoid a heavy computational overhead, it is common practice to set a limit on the number of repeat loops; we adopt a cap of 100 in the numerical tests whereby the resulting computational cost is not excessive compared with existing methods. The following section illustrates the performance of our heuristic algorithms on simple test problems.

Algorithm 2: TTTS+Random

Input: Information set $\mathscr{E}_{t-1}$, a probability parameter $0 < \gamma < 1$

Sample $\widetilde{p_{ij}} \sim \text{Beta}(\alpha_{ij}^{t-1}, \beta_{ij}^{t-1})$ for all $1 \le i < j \le k$, and set $\widetilde{p_{ji}} = 1 - \widetilde{p_{ij}}$.

Compute $i_1 = \arg\max_i \widetilde{p}_i = \frac{1}{k-1} \sum_{j \neq i} \widetilde{p_{ij}}$ and generate $x \sim \text{U}(0, 1)$.

if $x \le \gamma$ **then**

  Set $i^t = i_1$.

else

  repeat

    Sample $\widetilde{p_{ij}} \sim \text{Beta}(\alpha_{ij}^{t-1}, \beta_{ij}^{t-1})$ for all $1 \le i < j \le k$, and set $\widetilde{p_{ji}} = 1 - \widetilde{p_{ij}}$.

    Compute $i_2 = \arg\max_i \widetilde{p}_i = \frac{1}{k-1} \sum_{j \neq i} \widetilde{p_{ij}}$.

  until $i_2 \neq i_1$

  Set $i^t = i_2$.

end

Randomly select $j^t \neq i^t$ and switch (i^t, j^t) if $j^t < i^t$.

Return: (i^t, j^t)

Algorithm 3: TTTS+MaxVar

Input: Information set $\mathscr{E}_{t-1}$, a probability parameter $0 < \gamma < 1$

Sample $\widetilde{p_{ij}} \sim \text{Beta}(\alpha_{ij}^{t-1}, \beta_{ij}^{t-1})$ for all $1 \le i < j \le k$, and set $\widetilde{p_{ji}} = 1 - \widetilde{p_{ij}}$.

Compute $i_1 = \arg\max_i \widetilde{p}_i = \frac{1}{k-1} \sum_{j \neq i} \widetilde{p_{ij}}$ and generate $x \sim \text{U}(0,1)$.

if $x \le \gamma$ **then**

  Set $i^t = i_1$.

else

  repeat

    Sample $\widetilde{p_{ij}} \sim \text{Beta}(\alpha_{ij}^{t-1}, \beta_{ij}^{t-1})$ for all $1 \le i < j \le k$, and set $\widetilde{p_{ji}} = 1 - \widetilde{p_{ij}}$.

    Compute $i_2 = \arg\max_i \widetilde{p}_i = \frac{1}{k-1} \sum_{j \neq i} \widetilde{p_{ij}}$.

  until $i_2 \neq i_1$

  Set $i^t = i_2$.

end

Set $j^t = \arg\max_{j \neq i^t} \text{Var}\left[p_{i^t j} \mid \mathscr{E}_{t-1}\right] = \frac{\alpha_{i^t j} \beta_{i^t j}}{(\alpha_{i^t j} + \beta_{i^t j})^2 (\alpha_{i^t j} + \beta_{i^t j} + 1)}$ and switch (i^t, j^t) if $j^t < i^t$.

Return: (i^t, j^t)

---

4 NUMERICAL ILLUSTRATION

We illustrate our heuristic algorithms with a simple synthetic problem under different numbers of competing designs. For comparison purposes, we include the following algorithms in the experiments:

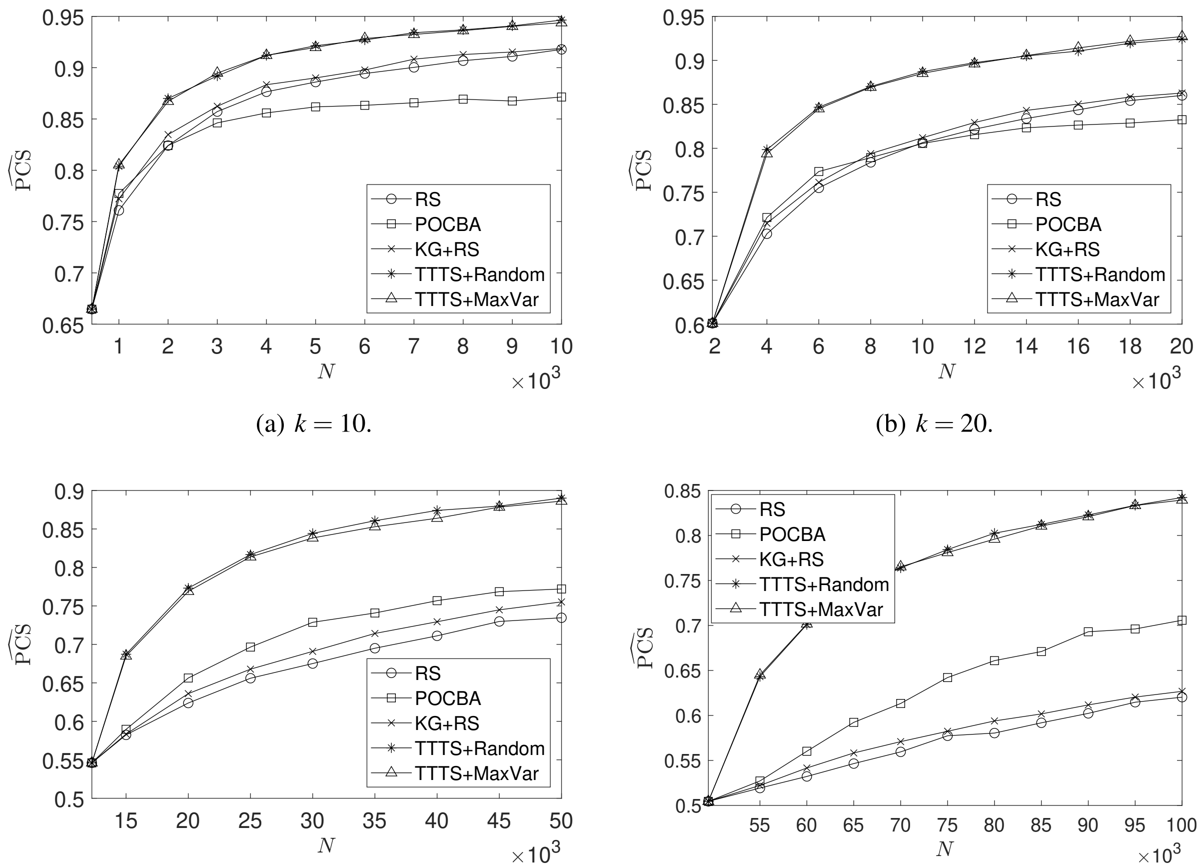
- Random Sampling (RS): In each iteration, a pair (i^t, j^t) is randomly selected for sampling. Even though it is well known to be inefficient (Jamieson and Nowak 2011), RS provides a natural and robust benchmark.
- POCBA (Groves and Branke 2019): By approximating the posterior distribution of each p_{ij} as Gaussian, this procedure looks one step ahead to improve the posterior PCS.
- KG+RS (Priekule and Meisel 2017): This algorithm uses KG to identify the pair for sampling provided that it can identify any pair with positive KG value; otherwise, it uses random sampling.

The four cases include $k = 10, 20, 50,$ and 100 competing designs, giving rise to a total number of pairs of 45, 190, 1225, and 4950, respectively. In the experiments, each pair initially receives $n_0 = 10$ samples; we set the total budget to be $N = 1000k$. The pairwise comparison probabilities p_{ij} are randomly drawn from the standard uniform distribution $\text{U}(0,1)$. We evaluate the efficiency of each algorithm through $R = 10{,}000$ independent macro replications, calculating the frequentist estimates of PCS and EOC as follows:

$$\widehat{\text{PCS}} = \frac{1}{R} \sum_{r=1}^{R} \mathbb{1}\left\{\widehat{i_r^*} = i_r^*\right\} \quad \text{and} \quad \widehat{\text{EOC}} = \frac{1}{R} \sum_{r=1}^{R} (p_{i_r^*} - p_{\widehat{i_r^*}}),$$

where i_r^* and $\widehat{i_r^*}$ denote the true and estimated best design in the r-th macro replication respectively. The results are depicted in Figure 1 and Figure 2.

From the results, we observe that the two versions of the TTTS policy yield comparable performance while consistently outperforming the other algorithms across the four cases. As the number of designs increases, the benefits of employing TTTS become more pronounced. In contrast, the KG+RS procedure exhibits only a slight advantage over RS. This observation suggests that while the KG framework theoretically makes sense, the Bernoulli setting often results in KG values of zero, thereby curtailing its potential. Notably, the performance of POCBA is disappointing, especially in the first two cases. Thus, the direct use of Gaussian approximations for Bernoulli outcomes is not robust. In summary, we find that TTTS provides an efficient framework to overcome the difficulties faced by the other procedures for selecting the best system where the performance of competing designs can only be assessed through pairwise comparisons



(a) $k = 10$. (b) $k = 20$.

(c) $k = 50$. (d) $k = 100$.

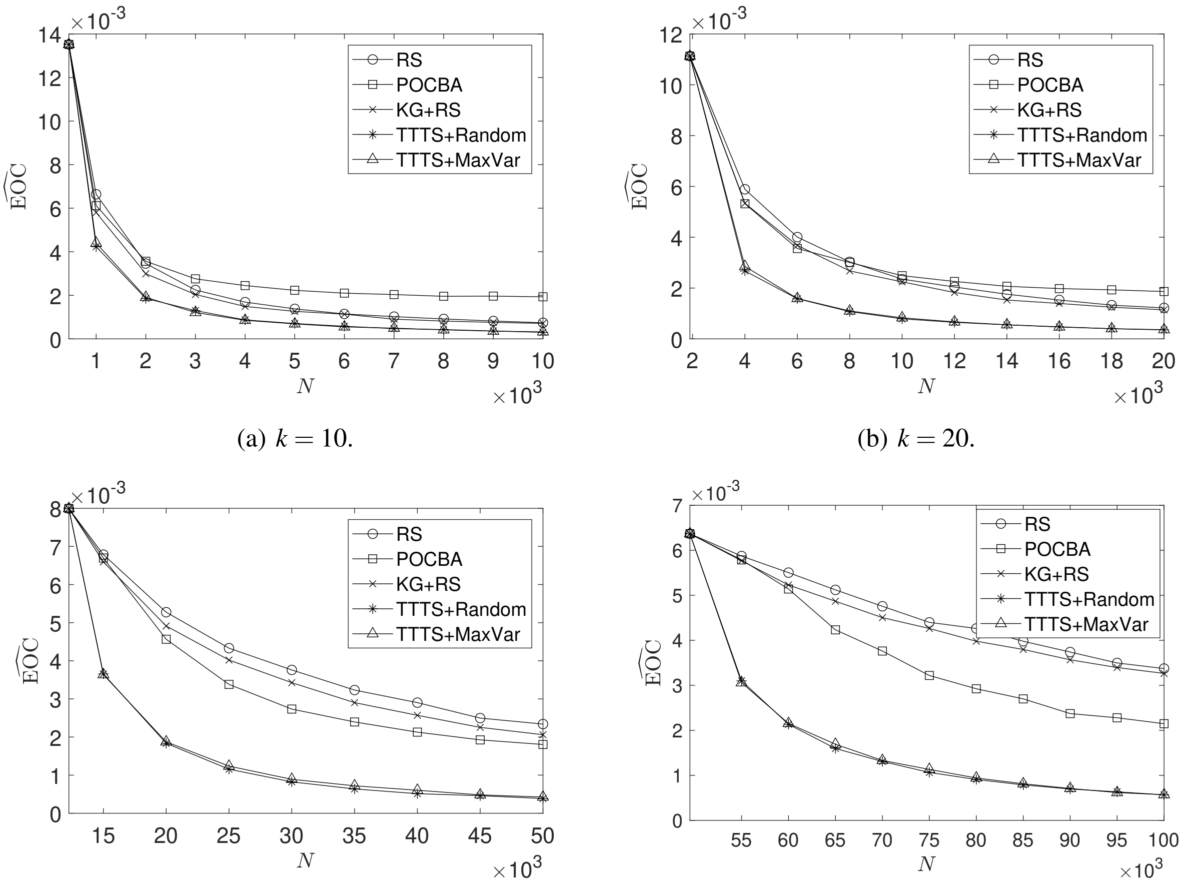
Figure 1: Comparison of the achieved $\widehat{\text{PCS}}$.

with binary outcomes. Although TTTS was originally designed to choose a single top design to sample, we show that aggregating TTTS with heuristic ways of identifying a runner-up can be both effective and robust for handling R&S with pairwise comparisons.

5 CONCLUSION

In this paper, we address the challenge of ranking and selection (R&S) where the performance of competing designs or alternatives can only be assessed by evaluating two alternatives simultaneously; the true performance of a design is defined as its Borda score, i.e., its average probability of outperforming the other competing designs. Under a fixed-budget perspective, we discuss some of the difficulties in applying popular R&S techniques to this context and propose the heuristic algorithms based on Thompson sampling to overcome those difficulties.

In our experiments, we note that Top-Two Thompson Sampling offers an effective approach to identifying the pair of designs to sample from. Shi et al. (2023) establish the posterior large deviation ratios for general adaptive sampling policies, including TTTS, for contextual R&S problems. They propose a randomized sampling policy to force sufficient exploration and prove its consistency. To generalize our observations from this limited set of experiments, our current focus is on characterizing the behavior of TTTS for pairwise comparisons and formally proving its properties.



(a) $k = 10$.

(b) $k = 20$.

(c) $k = 50$.

(d) $k = 100$.

Figure 2: Comparison of the achieved $\widehat{\mathrm{EOC}}$.

ACKNOWLEDGMENTS

This work has been partially supported by Abu Dhabi Commercial Bank through the ADCB Chair at INSEAD.

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.

Bengs, V., R. Busa-Fekete, A. El Mesaoudi-Paul, and E. Hüllermeier. 2021. “Preference-based Online Learning with Dueling Bandits: A Survey”. *Journal of Machine Learning Research* 22(7):1–108.

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(3):251–270.

Chen, Y. and I. O. Ryzhov. 2023. “Balancing Optimal Large Deviations in Sequential Selection”. *Management Science* 69(6):3457–3473.

Chick, S. E., J. Branke, and C. Schmidt. 2010. “Sequential Sampling to Myopically Maximize the Expected Value of Information”. *INFORMS Journal on Computing* 22(1):71–80.

Eriksson, B. 2013. “Learning to Top-k Search using Pairwise Comparisons”. In *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, edited by C. M. Carvalho and P. Ravikumar, 265–273. New York: Proceedings of Machine Learning Research.

Fan, W., L. J. Hong, and B. L. Nelson. 2016. “Indifference-Zone-Free Zelection of the Best”. *Operations Research* 64(6):1499–1514.

Frazier, P. I., W. B. Powell, and S. Dayanik. 2008. "A Knowledge-Gradient Policy for Sequential Information Collection". *SIAM Journal on Control and Optimization* 47(5):2410–2439.

Gao, S., W. Chen, and L. Shi. 2017. "A New Budget Allocation Framework for the Expected Opportunity Cost". *Operations Research* 65(3):787–803.

Glynn, P. and S. Juneja. 2004. "A Large Deviations Perspective on Ordinal Optimization". In *2004 Winter Simulation Conference (WSC)*, 577–585 https://doi.org/10.1109/WSC.2004.1371364.

Groves, M. and J. Branke. 2019. "Top-K Selection with Pairwise Comparisons". *European Journal of Operational Research* 274(2):615–626.

Heckel, R., N. B. Shah, K. Ramchandran, and M. J. Wainwright. 2019. "Active Ranking from Pairwise Comparisons and When Parametric Assumptions Do Not Help". *The Annals of Statistics* 47(6):3099–3126.

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.

Jamieson, K. G. and R. Nowak. 2011. "Active Ranking using Pairwise Comparisons". In *Advances in Neural Information Processing Systems 24*, edited by J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Weinberger, 2240–2248. New York: Curran Associates, Inc.

Kim, S.-H. and B. L. Nelson. 2001. "A Fully Sequential Procedure for Indifference-Zone Selection in Simulation". *ACM Transactions on Modeling and Computer Simulation* 11(3):251–273.

Li, H., X. Xu, Y. Peng, and C.-H. Chen. 2020. "Efficient Learning for Selecting Important Nodes in Random Network". *IEEE Transactions on Automatic Control* 66(3):1321–1328.

Lu, T. and C. Boutilier. 2011. "Robust Approximation and Incremental Elicitation in Voting Protocols". In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*, edited by T. Walsh, 287–293. Cambridge, MA: American Association for Artificial Intelligence Press.

Mohajer, S., C. Suh, and A. Elmahdy. 2017. "Active Learning for Top-*K* Rank Aggregation from Noisy Comparisons". In *Proceedings of the 34th International Conference on Machine Learning*, edited by D. Precup and Y. W. Teh, 2488–2497. New York: Proceedings of Machine Learning Research.

Paulson, E. 1964. "A Sequential Procedure for Selecting the Population with the Largest Mean From K Normal Populations". *The Annals of Mathematical Statistics* 35(1):174–180.

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

Peng, Y. and G. Zhang. 2022. "Thompson Sampling Meets Ranking and Selection". In *2022 Winter Simulation Conference (WSC)*, 3075–3086 https://doi.org/10.1109/WSC57314.2022.10015392.

Priekule, L. and S. Meisel. 2017. "A Bayesian Ranking and Selection Problem with Pairwise Comparisons". In *2017 Winter Simulation Conference (WSC)*, 2149–2160 https://doi.org/10.1109/WSC.2017.8247947.

Rinott, Y. 1978. "On Two-Stage Selection Procedures and Related Probability-Inequalities". *Communications in Statistics – Theory and Methods* 7(8):799–811.

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

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

Shi, X., Y. Peng, and G. Zhang. 2023. "Top-Two Thompson Sampling for Selecting Context-Dependent Best Designs". In *2023 Winter Simulation Conference (WSC)*, 3400–3411 https://doi.org/10.1109/WSC60868.2023.10407595.

Shin, D., M. Broadie, and A. Zeevi. 2018. "Tractable Sampling Strategies for Ordinal Optimization". *Operations Research* 66(6):1693–1712.

Wu, H. and X. Liu. 2016. "Double Thompson Sampling for Dueling Bandits". In *Advances in Neural Information Processing Systems 29*, edited by D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, 649–657. New York: Curran Associates, Inc.

Xiao, H., Y. Zhang, G. Kou, S. Zhang and J. Branke. 2023. "Ranking and Selection with Pairwise Comparisons". *Naval Research Logistics* 70(3):284–302.

Yue, Y., J. Broder, R. Kleinberg, and T. Joachims. 2012. "The K-Armed Dueling Bandits Problem". *Journal of Computing Systems Science* 78(5):1538–1556.

Yue, Y. and T. Joachims. 2009. "Interactively Optimizing Information Retrieval Systems as a Dueling Bandits Problem". In *Proceedings of the 26th Annual International Conference on Machine Learning*, edited by A. P. Danyluk, L. Bottou, and M. L. Littman, 1201–1208. New York: Association for Computing Machinery.

Zhong, Y., S. Liu, J. Luo, and L. J. Hong. 2022. "Speeding Up Paulson's Procedure for Large-Scale Problems using Parallel Computing". *INFORMS Journal on Computing* 34(1):586–606.

AUTHOR BIOGRAPHIES

DONGYANG LI is a doctoral student in the Department of Industrial Systems Engineering and Management at the National University of Singapore. His email address is dongyang_li@u.nus.edu and his website is https://cde.nus.edu.sg/c4ngp/staff/li-dongyang/.

ENVER YÜCESAN is a Professor in the Technology and Operations Management Area at INSEAD, Fontainebleau, France. His email address is enver.yucesan@insead.edu and his website is https://www.insead.edu/faculty/enver-yucesan.

CHUN-HUNG CHEN is a Professor in the Department of Systems Engineering and Operations Research at George Mason University. His email address is cchen9@gmu.edu and his website is https://mason.gmu.edu/~cchen9/.