Proceedings of the 2017 Winter Simulation Conference
W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, eds.

A BAYESIAN RANKING AND SELECTION PROBLEM WITH PAIRWISE COMPARISONS

Laura Priekule
Stephan Meisel

School of Business and Economics
University of Muenster
Leonardo-Campus 3
48149 Muenster, GERMANY

ABSTRACT

We consider a ranking and selection problem where sampling of two alternatives at once is required for learning about the true performances of the individual alternatives. The true performance of an alternative is defined as its average probability of outperforming the other alternatives. We derive and numerically compare four different solution approaches. Two Knowledge Gradient sampling policies are compared with a pure exploration policy and with a knockout tournament. The knockout tournament serves as a natural benchmarking approach with respect to pairwise comparisons, and determines the sampling budget provided to the other approaches. Our numerical results show that the Knowledge Gradient policies outperform both knockout tournament and pure exploration, and that they lead to significant improvements already at a very small number of pairwise comparisons. In particular we find that a nonstationary Knowledge Gradient policy is the best of the considered approaches for ranking and selection with pairwise comparisons.

1 INTRODUCTION

In this work, we consider a Bayesian ranking and selection (R&S) problem, where each sample observation involves a pair of alternatives. We distinguish between the two general cases of (a) R&S with pairwise comparisons, and (b) R&S with pairwise sampling. In the former case, observing two alternatives at the same time is essential for being able to gain information about the alternatives' performances. In the latter case observing pairs is not essential, i.e., two alternatives are observed although it is possible to gain information by individual observation. Pairwise sampling can be useful, e.g., for reducing the overall time required for an experiment, or for eliciting learning advantages due to correlations between alternatives.

Pairwise comparisons naturally occur in settings where two opposing alternatives are competing against each other, e.g., in military training and sports. However, there are far more cases where relying on pairwise comparisons is appropriate. Qualitative pairwise comparisons provide a straightforward assessment approach whenever the performances of alternatives must be judged with respect to a (possibly large) number of criteria, as well as whenever the performances can only be judged subjectively. As an example, consider virtual prototyping (Wang 2003) for the purpose of getting customer feedback on a number of design alternatives for a future product. Quantitative pairwise comparisons (where we observe that one alternative performs a certain number of times better than the other) occur for example in settings where a pair of alternatives is sampled in parallel under similar conditions that cannot be restored in the future. Examples include A/B testing (Siroker and Koomen 2013) of alternative designs with human-in-the-loop simulations (Narayanan and Kidambi 2011).

The approach proposed in this work could be applied to R&S with quantitative comparisons and with pairwise sampling, provided that sample observations are interpreted as qualitative performance observations. However, the approach is tailored to R&S with qualitative pairwise comparisons, i.e., to R&S problems

where two alternatives must necessarily be observed at once in order to gain information, and where one can merely observe which of two alternatives performs better. The fact that precisely one pair out of the set of alternatives must be observed distinguishes our problem from multinomial selection problems (see, e.g., Vieira Jr, Kienitz, and Belderrain (2010), and Tollefson et al. (2014)) with more than two alternatives.

Our approach differs from prior works on Bayesian R&S problems mainly in two interdependent respects: integration of pairwise comparisons, and the chosen performance model. We propose sampling policies for deciding which pair of alternatives to compare next. Note that in Bayesian learning (see, e.g., de Groot (1970)) sampling policies are typically designed for deciding which single alternative to evaluate next. We define the true performance of an alternative in terms of its unknown winning probabilities, i.e. in terms of the probabilities of outperforming each of the other alternatives in a pairwise comparison. We assign a Bernoulli distribution with a beta prior for the unknown winning probability to each pair of alternatives. Note that, in contrast, previous works on Bayesian R&S typically model the average true performances of the alternatives as means of normal distributions with normal priors.

By introducing pairwise comparisons in the context of Bayesian R&S, our work is positioned between the method of paired comparisons (see, e.g., David (1988) for an introductory overview) and classic Bayesian R&S (see, e.g., Powell and Ryzhov (2012), Ch.4). The method of paired comparisons (originating from psychometrics) aims either at obtaining a ranking of all alternatives, or at the general study of preferences (Kendall and Smith 1940) based on (a subset of) all possible pairwise comparisons between alternatives. In contrast to R&S, the method does not primarily aim at finding the best alternative with an efficient sampling policy.

In this work we adapt two Bayesian R&S sampling policies to pairwise comparisons, that were originally designed for choosing single alternatives. In particular, we propose a variation of the Knowledge Gradient (KG) policy introduced by Frazier, Powell, and Dayanik (2008) as an extension of the work of Gupta and Miescke (1996), and we propose a variation of the KG* policy introduced by Chick and Frazier (2009) as well as by Frazier and Powell (2010). KG policies are based on the value of information, i.e., sampling decisions are made such that the expected value of the gained information is maximized. Most works on KG policies assume normally distributed sample observations and normal beliefs about the unknown performances of alternatives, which leads to R&S problems with a normal-normal conjugate prior model.

The normal-normal model can be adapted to pairwise sampling (referred to as case (b) earlier). A KG policy allowing for both pairwise sampling and sampling of single alternatives is proposed by Frazier, Xie, and Chick (2011) and extended by Xie, Frazier, and Chick (2016). In these works, the value of information regarding two alternatives is calculated under the assumption that a quantitative pairwise comparison is made. However, the actual observation, is given in terms of pairwise sampling and results in a sample from a multivariate normal distribution. More precisely, if the two chosen alternatives have a negative sampling correlation, they are sampled independently, whereas in case of positive correlation the alternatives are sampled using common random numbers.

In our Bayesian R&S problem, however, each sample observation results from a qualitative pairwise comparison. As a consequence, it is clearly inappropriate to assume a normal sampling distribution and a normal-normal conjugate prior. Instead we assume that a qualitative pairwise comparison results in a sample from a Bernoulli distribution. This leads to the question of how to define the interrelationship between the true performance of an alternative and its winning probabilities. Two modeling approaches exist:

The first approach (widely adapted in the context of the method of paired comparisons) describes the true performances of the alternatives as unknown parameters, i.e., as normal expectations, and defines the winning probabilities as a function of these parameters. Popular examples for this approach are the Thurstone-Mosteller model (Thurstone 1927, Mosteller 1951) and the Bradley-Terry model (Bradley and Terry 1952). In a Bayesian context, this modeling approach implies a departure from the classic conjugate prior model and raises the need for approximate Bayesian analysis.

In the second modeling approach (adapted in this work, and referred to as the multibinomial model), the winning probabilities are the unknown parameters and the alternatives' performances must be defined with these probabilities. In our work, an alternative's performance is given in terms of its average probability of outperforming another alternative. The beliefs are given in terms of beta distributions, resulting in a Bernoulli-beta conjugate prior model with the standard recursive updating equations.

To the best of our knowledge, this work is the first to consider pairwise comparisons with the multibinomial model in the context of Bayesian R&S. A number of related works exist outside of the scope of R&S. These works propose sequential procedures adapting (variations of) the Bradley-Terry model for competitive sports or online gaming. However, their aim is to develop player rating systems, rather than efficient identification of the best alternative, i.e., the focus lies on learning the unknown skills of players without any sampling decisions to be made. Glickman (1999) is the first to propose a Bayesian approach to this problem by establishing a solid statistical foundation for the well-known Elo chess rating system (Elo 1978). Motivated by multiplayer online gaming, Glickmann's approach has later been extended to settings with performance observations resulting from games with more than two players or teams of players (Herbrich, Minka, and Graepel (2006), Dangauthier et al. (2007)).

Ryzhov, Awais, and Powell (2011) introduce a variation of the R&S problem with pairwise comparisons, referred to as the match-making problem. Motivated by the work of Herbrich, Minka, and Graepel (2006), the problem is modeled from the point of view of one fixed player. At each sampling stage an appropriate opponent has to be selected in order to guarantee a satisfying game for the fixed player. More precisely, the objective is to maximize the expected probabilities of a draw across all stages, and pairwise comparisons are made with one fixed alternative and another alternative which has to be selected. The proposed sampling policy also is a KG policy. However, besides the departures from our approach regarding the objective and the model, at each sampling stage only one alternative has to be selected instead of two.

In contrast to the works discussed above we propose a ranking and selection problem with pairwise comparisons. The performance of an alternative is modeled as the the average probability of outperforming another alternative in a pairwise comparison. We solve the problem with variations of two KG sampling policies, as well as with a pure exploration policy. As a natural benchmark for these three Bayesian policies we consider a straightforward knockout tournament approach (Vu, Altman, and Shoham 2009).

The remainder of this paper is structured as follows. In Section 2 we provide a formal problem statement. In Section 3 we introduce the sampling policies that we apply to the problem and that we compare numerically. Section 4 first describes our experimental setup including the knockout tournament approach used as benchmark, and then discusses numerical results. Section 5 concludes the paper.

2 PROBLEM FORMULATION

Recall that in our R&S problem pairwise comparisons are mandatory for gaining qualitative information about the alternatives' performances. While many applications in military training and sports naturally enforce the need for pairwise comparisons, there are other applications where the need for pairwise comparisons is introduced deliberately. As an example, consider virtual prototyping (Wang 2003) for getting customer feedback on product design alternatives. In this application, adopting the method of pairwise comparisons from psychometrics (e.g., by presenting a pair of alternative designs to each customer that logs in to the company website) can serve the purpose of receiving accurate feedback about the customers' preferences. Selecting the pair to be presented to the customer that logs in next is key for identifying the customers' preferred design as quickly as possible.

Against this background we now provide a formal statement of the considered R&S problem. We introduce the definition of the true performance of an alternative in terms of its winning probabilities (Section 2.1), we formulate the Bayesian updating equations for our beliefs about the alternatives' performances after each observation (Section 2.2), and we formulate the criterion an optimal sampling policy must meet (Section 2.3).

2.1 Performance of an Alternative

Suppose there are M distinct alternatives. For each alternative $i \in M$ we define a vector $p_i = (p_{i1}, p_{i2}, \ldots, p_{iM})$ of unknown probabilities, such that the winning of alternative i over j follows the Bernoulli distribution with a success probability p_{ij}. For the sake of notational simplicity and without loss of generality, for all $i \in M$ we may set $p_{ii} = 0.5$. As the events 'i wins over j' and 'j wins over i' are mutually exclusive, the equation $p_{ij} + p_{ji} = 1$ is always satisfied for all $i, j \in M$.

For each alternative we define the average win probability

$$\bar{p}_i = \frac{1}{M} \sum_{j=1}^{M} p_{ij}.$$

We make the assumption that the best alternative i^* is defined as the alternative with the largest average win probability, i.e.,

$$i^* \in \arg\max_{i \in M} \bar{p}_i. \tag{1}$$

Note that given an R&S problem with pairwise comparisons the choice of performance measure depends on the specific application under consideration. In the virtual prototyping example discussed earlier, looking for an alternative with a high average win probability may be interpreted as looking for an alternative that represents the mainstream customer preference.

As a consequence of the unknown success probabilities, the true average win probabilities are also unknown, and it is not possible to obtain the optimal value of (1). Instead we learn and improve our estimates of the unknown winning probabilities by making N sequential observations of individual alternative pairs. At each time $0 \leq n < N$, we choose an alternative pair $(i, j)^n$ for a pairwise comparison and observe a binary value $W_{ij}^{n+1} \sim \mathrm{Br}(p_{ij})$, which equals 1 if i wins over j, and 0 otherwise. Due to the obvious symmetry of pairs, without loss of generality alternative pairs considered for sampling are given by the set $\widetilde{M} = \{(i, j) \mid i \in M, i < j \leq M\}$. Thus, the number of possible alternative pairs to sample from amounts to $|\widetilde{M}| = (M^2 - M)/2$. After the sequential sampling procedure has been completed, we will make the implementation decision by choosing the alternative that we believe is the best.

2.2 Bayesian Updating of Beliefs

At each stage n of the sequential sampling procedure, our beliefs are given in terms of probability distributions. In particular, at any time $0 \leq n \leq N$ our beliefs about the winning probabilities p_{ij} are beta distributed with parameters α_{ij}^n and β_{ij}^n,

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

and the symmetry yields $p_{ji} = (1 - p_{ij})$.

Without loss of generality, for all $i \in M$ we set $p_{ii} \sim \mathrm{Beta}(1,1)$. We write $S^n := (\alpha^n, \beta^n)$ for the state of knowledge at time n, where $\alpha^n = (\alpha_{ij})_{M \times M}$ and $\beta^n = (\beta_{ij})_{M \times M}$. If at time n we choose to sample alternative pair $(i, j)^n$, our state of knowledge evolves according to the Bayesian updating equations, which, for all pairs (i, j) satisfying $j > i$, are given by

$$\alpha_{ij}^{n+1} = \begin{cases} \alpha_{ij}^n + W_{ij}^{n+1} & \text{if } (i,j)^n = (i,j) \\ \alpha_{ij}^n & \text{otherwise,} \end{cases}$$

$$\beta_{ij}^{n+1} = \begin{cases} \beta_{ij}^n + (1 - W_{ij}^{n+1}) & \text{if } (i,j)^n = (i,j) \\ \beta_{ij}^n & \text{otherwise.} \end{cases}$$

Our beliefs regarding the remaining pairs are updated using the symmetry, and the values α_{ii}^n and β_{ii}^n equal constantly one.

2.3 Criterion for an Optimal Policy

We write $\mathbb{E}^n$ to indicate expectations taken with respect to distributions induced by our beliefs at time n. With this notation our expected average win probability of an alternative i at time n is given by

$$\mathbb{E}^n[\bar{p}_i] = \frac{1}{M}\sum_{j=1}^{M}\mathbb{E}^n[p_{ij}] = \frac{1}{M}\sum_{j=1}^{M}\frac{\alpha_{ij}^n}{\alpha_{ij}^n+\beta_{ij}^n}.$$

Consequently, the implementation decision at time N is given by

$$\max_{i\in M}\mathbb{E}^N[\bar{p}_i] = \max_{i\in M}\frac{1}{M}\sum_{j=1}^{M}\frac{\alpha_{ij}^N}{\alpha_{ij}^N+\beta_{ij}^N}.$$

And, letting Π be the set of all possible selection policies $\pi = ((i,j)^0,\ldots,(i,j)^{N-1})$, an optimal policy must satisfy

$$\sup_{\pi\in\Pi}\mathbb{E}^{\pi}\left[\max_{i\in M}\frac{1}{M}\sum_{j=1}^{M}\frac{\alpha_{ij}^N}{\alpha_{ij}^N+\beta_{ij}^N}\right], \tag{2}$$

where $\mathbb{E}^{\pi}$ denotes the expectation induced by a policy π. Since calculating the exact distribution of $\max_{i\in M}\mathbb{E}^N[\bar{p}_i]$ given a policy π in (2) is computationally infeasible, we rely on sampling policies that only represent approximations to the optimal policy solving (2).

3 SAMPLING POLICIES

In this section we briefly describe the three sampling policies used in the computational experiments of Section 4.

3.1 Pure Exploration Policy

At each step n the *pure exploration* policy selects a pair of alternatives $(i,j)^n = (i,j)$ with probability $1/\widetilde{M}$. The policy does neither take into account the current beliefs nor the information collected from previous pairwise comparisons for selecting a pair. The two policies introduced in Sections 3.2 and 3.3 partly rely on pure exploration. However, they also take into account the current state of knowledge and consider the expected information value of making a pairwise comparison between any two alternatives for making a sampling decision.

3.2 Knowledge Gradient Policy

The KG policy (Frazier, Powell, and Dayanik 2008) selects at each step n the pair of alternatives that maximizes the so called *KG factor*. The KG factor of an alternative pair (i,j) looks one sampling step into the future and captures the expected value of making a pairwise comparison between i and j. For an alternative pair (i,j) the KG factor is given by

$$\nu_{ij}^{KG,n} = \mathbb{E}^n\left[\max_x\frac{1}{M}\sum_{y=1}^{M}\frac{\alpha_{xy}^{n+1}}{\alpha_{xy}^{n+1}+\beta_{xy}^{n+1}} - \max_x\frac{1}{M}\sum_{y=1}^{M}\frac{\alpha_{xy}^{n}}{\alpha_{xy}^{n}+\beta_{xy}^{n}}\right], \tag{3}$$

and the KG decision rule for a given knowledge state is

$$X^{KG}(S^n) = \arg\max_{(i,j)\in\widetilde{M}}\nu_{ij}^{KG,n}.$$

At time n, the predictive distribution of the parameters $(\alpha_{ij}^{n+1}, \beta_{ij}^{n+1})$ is discrete, where the corresponding probabilities are given by

$$\mathbb{P}(\alpha_{ij}^{n+1} = \alpha_{ij}^n + 1,\ \beta_{ij}^{n+1} = \beta_{ij}^n) = \frac{\alpha_{ij}^n}{\alpha_{ij}^n + \beta_{ij}^n},$$
$$\mathbb{P}(\alpha_{ij}^{n+1} = \alpha_{ij}^n,\ \beta_{ij}^{n+1} = \beta_{ij}^n + 1) = \frac{\beta_{ij}^n}{\alpha_{ij}^n + \beta_{ij}^n}.$$

And, setting $C_{ij}^n = \max_{x \neq i,j} \bar{p}_x^n$, the expectation in (3) can be reformulated as

$$\begin{aligned} {\nu_{ij}}^{KG,n} = & \\ & \frac{\alpha_{ij}^n}{\alpha_{ij}^n + \beta_{ij}^n} \max\left\{ C_{ij}^n,\ \bar{p}_i^n + \frac{\beta_{ij}^n}{(\alpha_{ij}^n + \beta_{ij}^n)(\alpha_{ij}^n + \beta_{ij}^n + 1)},\ \bar{p}_j^n - \frac{\beta_{ij}^n}{(\alpha_{ij}^n + \beta_{ij}^n)(\alpha_{ij}^n + \beta_{ij}^n + 1)} \right\} \\ & + \frac{\beta_{ij}^n}{\alpha_{ij}^n + \beta_{ij}^n} \max\left\{ C_{ij}^n,\ \bar{p}_i^n - \frac{\alpha_{ij}^n}{(\alpha_{ij}^n + \beta_{ij}^n)(\alpha_{ij}^n + \beta_{ij}^n + 1)},\ \bar{p}_j^n + \frac{\alpha_{ij}^n}{(\alpha_{ij}^n + \beta_{ij}^n)(\alpha_{ij}^n + \beta_{ij}^n + 1)} \right\} - \max_x \bar{p}_x^n. \end{aligned} \tag{4}$$

A case study of (4) reveals, that the KG factor depends on the absolute values of the pairwise differences between $\bar{p}_i^n$, $\bar{p}_j^n$ and C_{ij}^n. In order for the KG factor not to be equal to zero, the absolute values of these differences should not exceed thresholds that are dominated by the values of $\alpha_{ij}^n/(\alpha_{ij}^n + \beta_{ij}^n)(\alpha_{ij}^n + \beta_{ij}^n + 1)$ and $\beta_{ij}^n/(\alpha_{ij}^n + \beta_{ij}^n)(\alpha_{ij}^n + \beta_{ij}^n + 1)$. Obviously, with growing alphas and betas these thresholds get very small quickly, and once all factors equal zero, there is no value of sampling according to the KG factor.

As a possible circumvention of this issue, we propose a modified KG policy, which, for a given sampling budget N, makes decisions like the original KG policy as long as the maximal KG factor is greater than zero, and chooses an alternative for a pairwise comparison randomly otherwise. Thus, it is a hybrid between KG and the pure exploration policy described in the previous subsection.

3.3 KG* Policy

In contrast to the KG policy, the KG* policy (Frazier and Powell 2010) looks several sampling steps ahead into the future and considers a number of subsequent observations for each alternative pair. Let $\nu^{KG,n}(n_{ij})$ denote the KG factor if we were to perform n_{ij} pairwise comparisons with an alternative pair (i, j). In this case, we have the following probabilities describing the discrete predictive distribution of $(\alpha_{ij}^{n+1}, \beta_{ij}^{n+1})$,

$$\mathbb{P}(\alpha_{ij}^{n+1} = \alpha_{ij}^n + k,\ \beta_{ij}^{n+1} = \beta_{ij}^n + n_{ij} - k) = \binom{n_{ij}}{k} \left(\frac{\alpha_{ij}^n}{\alpha_{ij}^n + \beta_{ij}^n}\right)^k \left(\frac{\beta_{ij}^n}{\alpha_{ij}^n + \beta_{ij}^n}\right)^{n_{ij}-k} \qquad k = 0, \ldots, n_{ij}.$$

And the KG factor as in (3) can be reformulated as

$$\begin{aligned} \nu^{KG,n}(n_{ij}) = & \sum_{k=0}^{n_{ij}} \binom{n_{ij}}{k} \left(\frac{\alpha_{ij}^n}{\alpha_{ij}^n + \beta_{ij}^n}\right)^k \left(\frac{\beta_{ij}^n}{\alpha_{ij}^n + \beta_{ij}^n}\right)^{n_{ij}-k} \cdot \\ & \cdot \max\left\{ C_{ij}^n,\ \bar{p}_i^n + \frac{\alpha_{ij}^n(k - n_{ij}) + k\beta_{ij}^n}{(\alpha_{ij}^n + \beta_{ij}^n)(\alpha_{ij}^n + \beta_{ij}^n + n_{ij})},\ \bar{p}_j^n + \frac{\alpha_{ij}^n(n_{ij} - k) + k\beta_{ij}^n}{(\alpha_{ij}^n + \beta_{ij}^n)(\alpha_{ij}^n + \beta_{ij}^n + n_{ij})} \right\} - \max_x \bar{p}_x^n \end{aligned}$$

The KG* decision rule for a given state S^n and fixed sampling budget N is then given by

$$X^{KG^*,n}(S^n) = \arg\max_{(i,j) \in \widetilde{M}} \max_{0 < n_{ij} \leq N-n} \frac{\nu_{ij}^{KG}(n_{ij})}{n_{ij}}.$$

As in case of the Knowledge Gradient policy, we also modify the KG* policy such that an alternative pair is selected by pure exploration if all KG factors are equal to zero. Note that the KG* policy is not stationary, and that it depends not only on the current state, but also on the number of sample observations left.

4 COMPUTATIONAL EXPERIMENTS

In Section 4.1 we introduce the problem instances and the performance measures used in our computational experiments. Moreover, we explain the knockout tournament approach that serves as a benchmarking approach for the sampling policies, and that determines the sampling budget that the policies are provided with. In Section 4.2 we compare the sampling policies and the benchmarking approach numerically.

4.1 Experimental Setup

In order to compare the performances of the pure exploration policy and the Knowledge Gradient policies, we do experiments with 5, 10 and 20 individual alternatives, i.e., with $M \in \{5, 10, 20\}$. The corresponding number of alternative pairs $\widetilde{M}$ are 10, 105 and 190, respectively. For each $M \in \{5, 10, 20\}$ we randomly generate 25 problem instances. In particular, 5 different initial beliefs are generated, where the entries of the initial prior parameters α^0 and β^0 (except α^0_{ii} and β^0_{ii}) are random samples from a uniform discrete distribution on the set $\{1, 2, 3, 4, 5\}$. The purpose of using such a set, is to emulate the state of having initial priors that give a rather small amount of information with the precision not being too high. Subsequently, 5 different truths with $p_{ij} \sim \text{Beta}(\alpha^0_{ij}, \beta^0_{ij})$ are sampled per initial belief. This represents a situation in which our prior beliefs provide a reasonably good idea about the true winning probabilities.

We consider the winner of a knockout tournament (Schwenk 2000) as benchmark for the sampling policies on each problem instance. Similar to a typical tennis tournament, a knockout tournament (KOT) consists of stages in which the alternatives compete pairwise against each other with the losers being eliminated and the winners progressing to the next stage until only one alternative is left. Since the initial number of alternative pairs not always is a perfect power of two, one randomly selected alternative advances to the next stage, if there is a stage with an uneven number of alternatives ("wildcard"). We choose KOT as benchmark for the sampling policies, because it represents a straightforward method that directly applies to problems where pairwise comparisons are mandatory. The following two KOT versions are considered:

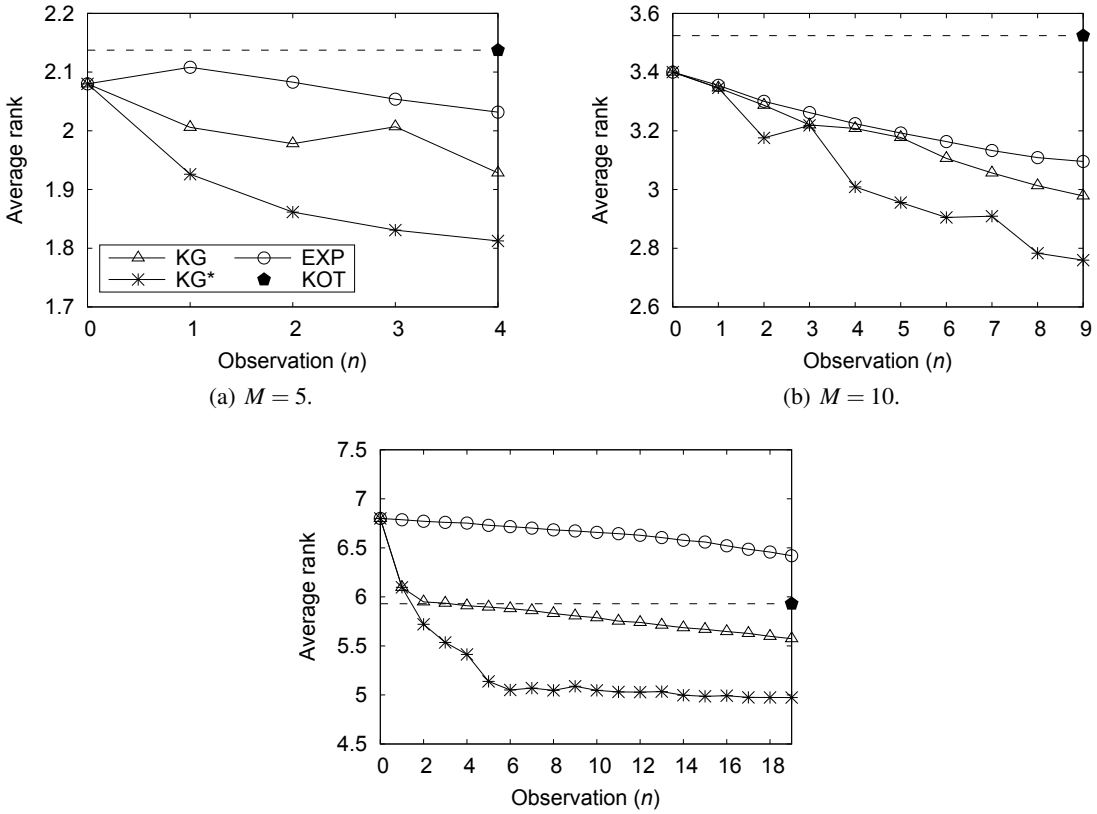
We first apply a basic knockout tournament by setting the alternative pairs randomly in each stage without taking into account the possibility of a non-random seeding. (Note that the tournament designs introduced in the literature require at least partly known success probabilities or even certain monotonicity of the success probabilities (Horen and Riezman 1985, Vu, Altman, and Shoham 2009).) We then extend the the basic knockout tournament approach by allowing for a fixed number n_{KOT} of comparisons per pair in every stage. If the results (in terms of wins of the alternatives) of n_{KOT} comparisons of a pair are tied, one of the two alternatives is chosen randomly for passing to the next stage. With this extension the number of pairwise comparisons needed for a knockout tournament to yield a winner is then determined by the values of M and n_{KOT}.

In our experiments with M alternatives, the sampling budget N the policies are provided with is set equal to the number of pairwise comparisons required by a knockout tournament with n_{KOT}. We are then able to compare the performances of the sampling policies with a straightforward procedure that yields a winner without requiring any kind of prior information regarding the alternatives.

Many authors in the Bayesian sampling literature use as a standard performance measure for sampling policies the opportunity cost of the believed best alternative with respect to the true value (in our problem given by $\max_{i \in M} \bar{p}_i - \bar{p}_{\arg\max_{i \in M} \mathbb{E}^N[\bar{p}_i]}$). However, since we use the KOT approach as benchmark, and since in KOT the assumption is common that the alternatives can only be ranked qualitatively from best to worst, we prefer to evaluate in terms of the true ranks of both the tournament winners and the believed best alternatives of the policies. To each alternative i, we assign its true rank r_i in the following way: The

alternative with the highest true average win probability has rank one, the alternative with the second best value has rank two, etc.. In order to estimate the expected true rank of the believed best alternative at each step of the sampling process we average the results of the sampling policies and of KOT over 10^3 different sample paths for each problem instance, and over the 25 instances.

4.2 Results



(a) $M = 5$. (b) $M = 10$.

(c) $M = 20$.

Figure 1: Evolution of the true rank of the believed best alternative by the hybrid KG, KG* and pure exploration policies during the sampling process given the true rank of the winner of a standard knockout tournament with $n_{KOT} = 1$ as a benchmark.

Figure 1 illustrates the evolution of the average true rank of the believed best alternative during the sampling processes of each of the policies of Section 3, with 5 alternatives (Figure 1a), 10 alternatives (Figure 1b) and 20 alternatives (Figure 1c). We compare the policies with the standard knockout tournament, given in terms of $n_{KOT} = 1$. Thus, the benchmark is given by the average true rank of the winner of a standard knockout tournament and the number of observations N amounts to 4, 9 and 19 concerning the cases of 5, 10 and 20 alternatives, respectively. Note that the resulting budget N for each M covers only a small fraction of the number of alternative pairs taken into consideration (approx. 24%, 8 % and 10%). The value of the benchmark produced by the knockout tournament and the time of its realization are marked by the black pentagon. Dashed lines highlight the benchmark throughout the policies' sampling processes.

We observe that on average the initially best believed alternative lies around the lower bound of the top third of the alternatives and, in the cases $M = 5$ and $M = 10$, already trumps the average winner of the knockout tournament, as do all the following best believed alternatives. In the 20 alternatives case,

the initially believed best alternative is ranked lower than the tournament winner by approximately 0.9. However, with the Knowledge Gradient policies, the best believed alternative overtakes the tournament winner in the course of a few observations, in particular, at four observations with KG and already at two observations with KG*. By contrast, the pure exploration policy never reaches the benchmark, as it selects alternative pairs randomly without a specific selection criterion. This competitive disadvantage to both of the Knowledge Gradient policies grows with the number of alternatives if the budget of observations is proportionally small. None of the sampling policies comes close to the best alternative within the given budgets of observations. However, the KG* policy outperforms all other policies, and in case of $M = 20$ the rank improves on average by almost 2 over the sampling process, despite the low budget of observations.

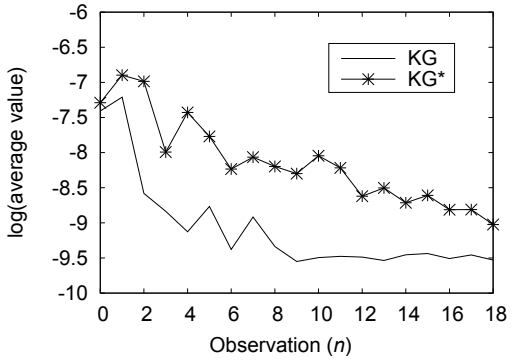
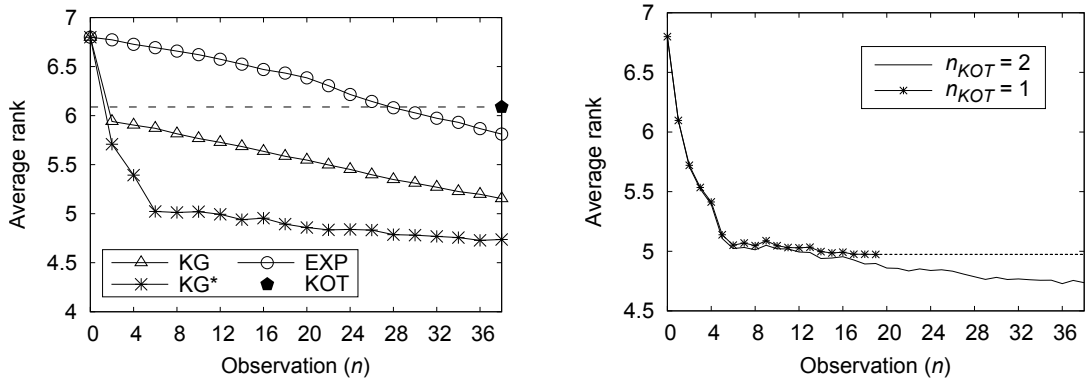


Figure 2: Comparison of the KG and KG* values for a setting with $M = 20$ and $n_{KOT} = 1$.

Comparing the two Knowledge Gradient policies, we observe that in general the KG* policy performs clearly better than the KG policy. This is due to the fact that KG* explores the alternative pairs more efficiently than KG. In Figure 2, showing $M = 20$ and $n_{KOT} = 1$, we illustrate the natural logarithm of the average KG and KG* values at each measurement step. The figure illustrates that the average KG* values are always higher than the corresponding KG values. This obviously results from the fact that the observations we make are binary, and, as discussed in Section 3.2, the KG value quickly drops to zero under certain conditions. As a consequence of our modifications to the KG policy, it behaves like pure exploration if the KG value is zero, and therefore doesn't take into account any value of information at this point. Here the KG* policy has an advantage, since it looks several steps ahead into the future and calculates the maximal average value of information, which mostly still results in very low values, that are, however, greater than zero.



(a) Learning curves for $M = 20$ and $n_{KOT} = 2$.

(b) KG* learning curves for $n_{KOT} = 1$ and $n_{KOT} = 2$.

Figure 3: Evolution of the true rank of the best believed alternative.

As pointed out earlier, letting $n_{KOT} = 1$ results in a very small budget of observations compared with the number of possible pairs of alternatives. As a consequence the sampling policies' task to find the truly best alternative is particularly hard, unless the initial priors precisely match the truth. One would expect the policies to find lower ranked best believed alternatives, as the budget of observations grows at fixed M. Figure 3a shows the same comparison as Figure 1c, but with a higher observation budget of $N = 38$ due to $n_{KOT} = 2$. Compared with the number of possible pairs, the budget is still low, but at least it now amounts to 20% of the number of possible pairs. The figure shows that with the higher budget the average true rank of the knockout tournament winner is even slightly lower than before. Our experiments show that the knockout tournament with multiple comparisons yields a higher ranked winner only in settings with wide ranged true average win probabilities $\bar{p}_i$. However, if these probabilities have rather similar values, a knockout tournament with slightly higher n_{KOT} does not necessarily yield a better result on average.

In contrast, the sampling policies benefit from the larger budget and the true rank of the final best believed alternatives increases as expected. With both the KG policy and pure exploration, the knowledge state at time n does not depend on the actual budget of observations. As a consequence, the policies' average learning curves in Figure 3a appear as mere extensions of their learning curves illustrated in Figure 1c.

The case is different for the KG* policy, since given a higher budget N this policy is able to look ahead into a more distant future. As shown in Figure 3b, the higher sampling budget leads to more efficient learning in early stages of the sampling process. Here, the low budget ($n_{KOT} = 1$) learning curve ends at $n = 19$ and the line extends the curve at the constant value of the achieved average rank. Figure 3b illustrates that the high budget ($n_{KOT} = 2$) learning curve is located below the low budget curve, and, thus, indicates that the learning efficiency grows as the budget of observations increases.

5 CONCLUSIONS

We have considered a ranking and selection problem that requires pairwise comparisons of alternatives for being able to gain information about the alternatives' performances. We adapt the multibinomial model and assume that the unknown true performance of an alternative is given in terms of its average probability of outperforming the other alternatives in a pairwise comparison. We numerically compare three sampling policies for the ranking and selection problem: a variation of the Knowledge Gradient Policy, a variation of the KG* policy, as well as a pure exploration policy.

We consider the true rank of the final believed best alternative as performance measure for a policy. As a natural benchmark for the sampling policies we use the true rank of the winner of a knockout tournament, where the policies' sampling budgets of observations is set equal to the number of pairwise comparisons required for the tournament to determine a winner. Our numerical results show that KG* is the best of the considered approaches, yielding the best believed alternative with the highest rank on average. The pure exploration policy, on contrary, shows slow learning performance. The Knowledge Gradient policy performs significantly better than pure exploration. However, it shows a disadvantage compared to the KG* policy due to its myopic nature in combination with binary observations.

Many avenues for future work exist. The next steps will include thorough analyses of the impacts of different types of prior-truth relationships and of an increasing sampling budget. Moreover, more advanced policies that exploit correlations between pairs of alternatives will be developed.

REFERENCES

Bradley, R. A., and M. E. Terry. 1952. "Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons". *Biometrika* 39 (3/4): 324–345.

Chick, S. E., and P. I. Frazier. 2009. "The Conjunction of the Knowledge Gradient and the Economic Approach to Simulation Selection". In *Proceedings of the 2009 Winter Simulation Conference*, edited

by M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin, and R. G. Ingalls, 528–539. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Dangauthier, P., R. Herbrich, T. Minka, and T. Graepel. 2007. "TrueSkill Through Time: Revisiting the History of Chess.". In *Advances in Neural Information Processing Systems*, Volume 20, 337–344: MIT-Press.

David, H. A. 1988. *The Method of Paired Comparisons*. 2nd ed. Lubrecht & Cramer, Limited.

de Groot, M. H. 1970. *Optimal Statistical Decisions*. New York: McGraw-Hill.

Elo, A. E. 1978. *The Rating of Chessplayers, Past and Present*. Arco Pub.

Frazier, P. I., and W. B. Powell. 2010. "Paradoxes in Learning and the Marginal Value of Information". *Decision Analysis* 7:378–403.

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

Frazier, P. I., J. Xie, and S. E. Chick. 2011. "Value of Information Methods for Pairwise Sampling with Correlations". In *Proceedings of the 2011 Winter Simulation Conference*, edited by S. Jain, R. R. Creasey, J. Himmelspach, K. P. White, and M. Fu, 3974–3986. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Glickman, M. E. 1999. "Parameter Estimation in Large Dynamic Paired Comparison Experiments". *Applied Statistics* 48:377–394.

Gupta, S., and K. Miescke. 1996. "Bayesian Look Ahead One-Stage Sampling Allocations for Selection of the Best Population". *Journal of Statistical Planning and Inference* 54:229–244.

Herbrich, R., T. Minka, and T. Graepel. 2006. "Trueskill: A Bayesian Skill Rating System". In *Advances in Neural Information Processing Systems*, Volume 19, 569–576: MIT-Press.

Horen, J., and R. Riezman. 1985. "Comparing Draws for Single Elimination Tournaments". *Operations Research* 33 (2): 249–262.

Kendall, M. G., and B. B. Smith. 1940. "On the Method of Paired Comparisons". *Biometrika* 31 (3/4): 324–345.

Mosteller, F. 1951. "Remarks on the Method of Paired Comparisons: I. The Least Squares Solution Assuming Equal Standard Deviations and Equal Correlations". *Psychometrika* 16 (1): 3–9.

Narayanan, S., and P. Kidambi. 2011. "Interactive Simulations: History, Features, and Trends". In *Human-in-the-Loop Simulations*, edited by L. Rothrock and S. Narayanan, 1–13: Springer.

Powell, W. B., and I. Ryzhov. 2012. *Optimal Learning*. Hoboken, New Jersey: Wiley.

Ryzhov, I. O., T. Awais, and W. B. Powell. 2011. "May the Best Man Win: Simulation Optimization for Match-Making in E-Sports". In *Proceedings of the 2011 Winter Simulation Conference*, edited by S. Jain, R. R. Creasey, J. Himmelspach, K. P. White, and M. Fu, 4234–4245. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Schwenk, A. J. 2000. "What is the Correct Way to Seed a Knockout Tournament?". *The American Mathematical Monthly* 107 (2): 140–150.

Siroker, D., and P. Koomen. 2013. *A/B Testing: The Most Powerful Way to Turn Clicks Into Customers*. Hoboken, New Jersey: Wiley.

Thurstone, L. L. 1927. "The Method of Paired Comparisons for Social Values.". *The Journal of Abnormal and Social Psychology* 21 (4): 384–400.

Tollefson, E., D. Goldsman, A. Kleywegt, and C. Tovey. 2014. "Optimal Selection of the Most Probable Multinomial Alternative". *Sequential Analysis* 33:491–508.

Vieira Jr, H., K. H. Kienitz, and M. C. N. Belderrain. 2010. "Multinomial Selection Problem: A Study of BEM and AVC Algorithms". *Communications in Statistics - Simulation and Computation* 39:971–980.

Vu, T., A. Altman, and Y. Shoham. 2009. "On the Complexity of Schedule Control Problems for Knockout Tournaments". In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*, 225–232. International Foundation for Autonomous Agents and Multiagent Systems.

Wang, G. G. 2003. "Definition and Review of Virtual Prototyping". *Journal of Computing and Information Science in Engineering* 2:232–236.

Xie, J., P. I. Frazier, and S. E. Chick. 2016. "Bayesian Optimization via Simulation with Pairwise Sampling and Correlated Prior Beliefs". *Operations Research* 64 (2): 542–559.

AUTHOR BIOGRAPHIES

LAURA PRIEKULE is a Ph.D. student in the School of Business and Economics at the University of Muenster, Germany. She holds a German Diploma (M.S. equivalent) in Mathematics from University of Dresden, Germany. Her research interests lie in stochastic optimization and in particular sequential decision making under uncertainty. Her email address is laura.priekule@uni-muenster.de.

STEPHAN MEISEL is an Assistant Professor in the School of Business and Economics at the University of Muenster, Germany. He holds a Ph.D. in Operations Research from University of Braunschweig, Germany. His research interests lie in stochastic optimization and in particular sequential decision making under uncertainty with applications in energy and transportation. His email address is stephan.meisel@uni-muenster.de.