Proceedings of the 2012 Winter Simulation Conference
C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. M. Uhrmacher, eds.

RANKING AND SELECTION MEETS ROBUST OPTIMIZATION

Ilya O. Ryzhov

Robert H. Smith School of Business
University of Maryland
College Park, MD 20742, USA

Boris Defourny
Warren B. Powell

Operations Research & Financial Engineering
Princeton University
Princeton, NJ 08544, USA

ABSTRACT

The objective of ranking and selection is to efficiently allocate an information budget among a set of design alternatives with unknown values in order to maximize the decision-maker's chances of discovering the best alternative. The field of robust optimization, however, considers risk-averse decision makers who may accept a suboptimal alternative in order to minimize the risk of a worst-case outcome. We bring these two fields together by defining a Bayesian ranking and selection problem with a robust implementation decision. We propose a new simulation allocation procedure that is risk-neutral with respect to simulation outcomes, but risk-averse with respect to the implementation decision. We discuss the properties of the procedure and present numerical examples illustrating the difference between the risk-averse problem and the more typical risk-neutral problem from the literature.

1 INTRODUCTION

Consider a decision-maker who must choose one out of finitely many expensive alternatives. The alternatives could represent different factory layouts, investment strategies, supply contracts, or funding decisions. One example is the problem of hedging electricity contracts. The decision-maker must choose a number of forward contracts for delivering electricity to meet future demand. Buying too few contracts means having to make up the difference on the spot market; buying too many means having to sell back the excess at the market price. The goal is to select the buying strategy with the lowest average cost.

In all of these cases, we suppose that the number of alternatives is relatively small, limited to a few of the most promising choices. However, the cost of implementing a choice is substantial, and the decision-maker may use stochastic simulation to estimate the values of various alternatives before committing to a final decision. The output of the simulation provides new information, allowing the decision-maker to make a more informed implementation decision. The goal of the ranking and selection (R&S) problem is to allocate the simulation budget efficiently in order to maximize the quality (variously defined) of the final decision; see e.g. Bechhofer et al. (1995) for an introduction.

At the same time, the learning process itself is uncertain. Even with simulation, the decision-maker's beliefs about the problem can be misleading. For example, the decision-maker may severely underestimate the variance of the simulation output (i.e. the performance of a particular alternative). It may also be that we place too much or too little confidence in our initial beliefs about the performance values. Then, the value of the alternative implemented by the decision-maker may, in fact, be much worse than estimated. When the cost of error (choosing a poor investment strategy or signing a contract with an unreliable supplier) is particularly great, a risk-averse decision-maker may prefer an alternative that may be suboptimal, but performs as well as possible in a worst-case situation. The present paper examines R&S from the risk-averse point of view.

We adopt the Bayesian approach to R&S (see Chick 2006 for a survey), which assumes that we begin with prior beliefs about the alternatives, and uses various probability distributions to model our

978-1-4673-4781-5/12/$31.00 ©2012 IEEE

uncertainty about the prior. In this paper, we use a standard model from DeGroot (1970), where the unknown performance values of the alternatives are assumed to be independent, and the variance of the simulation output is assumed to be known. This model has been used, e.g., by Gupta and Miescke (1994) and Gupta and Miescke (1996) to derive "value of information" procedures, which sequentially simulate alternatives that are believed to have the highest potential for improving the implementation decision. There are numerous extensions of this approach to other learning models: for example, Chick et al. (2010) considers a Bayesian model with independent priors and unknown sampling variance. However, all of this literature has assumed the decision-maker to be risk-neutral (explicitly stated in Chick and Frazier 2009 and Chick and Gans 2009, but implicitly assumed elsewhere), and the resulting learning procedures have sought to maximize the expected value of the alternative selected for implementation.

Another school of thought uses frequentist statistics to model learning, and seeks to maximize the probability that the implementation decision matches the true best alternative. See Hong and Nelson (2009) for an accessible introduction to this approach. The indifference-zone method (Kim and Nelson 2001; Kim and Nelson 2006) guarantees a certain probability of correct selection (PCS), given that the true value of the best alternative is sufficiently greater than the second-best. In practice, these methods are somewhat conservative (Wang and Kim 2011), and may even "over-deliver" (Branke et al. 2007), producing a better PCS than we require. This is in line with our desire for a more conservative or risk-averse approach, though it is still assumed that the decision-maker is maximizing the value of the implementation decision.

We introduce risk-averse implementation decisions into R&S by making a connection to the field of robust optimization (Ben-Tal et al. 2009). This methodology uses a maxi-min objective, optimizing our performance for worst-case realizations of the unknown values. The implementation decision thus maximizes with respect to a fixed sample path chosen in an adversarial way. Robust optimization has been applied in a wide variety of settings (see Ben-Tal and Nemirovski 2002 or Bertsimas et al. 2007 for examples), but never, to our knowledge, in ranking and selection. The philosophy of robust optimization is to convert stochastic problems into deterministic ones on worst-case sample paths, while the main principle of R&S is that unknown values can only be learned through stochastic simulation. Nonetheless, we can use robust optimization concepts that allow risk-averse decisions without the need for a specific utility function (a subject discussed e.g. by Waeber et al. 2010).

In this paper, we propose a new approach to R&S that integrates a robust implementation decision into a Bayesian learning framework. We then derive a value of information procedure that allocates simulations based on expected improvement in the robust decision, suggesting a decision-maker who is risk-averse with respect to implementation, but risk-neutral with respect to simulation. In the basic model with independent priors and independent replications, this policy is shown to generalize the knowledge gradient (KG) formula of Frazier et al. (2008), retaining the KG policy's theoretical property of global convergence. We present numerical comparisons with the risk-neutral KG method and highlight situations where our "robust KG" policy adds value. We have found that the robust approach is particularly useful when 1) the simulation output has high variance, 2) the simulation budget is relatively small, and 3) the learning model is improperly specified.

2 ROBUST RANKING AND SELECTION

We illustrate the concept of robust R&S on a Bayesian learning model with independent priors and independent observations with known variance. Suppose that there are M alternatives, and that a simulation of alternative $x \in \{1,...,M\}$ produces an observation $W_x \sim \mathcal{N}\left(\mu_x, \sigma_\varepsilon^2\right)$. The variance σ_ε^2 is known to the decision-maker (and can be made to depend on x), while the mean μ_x is unknown. We begin with a Bayesian prior $\mu_x \sim \mathcal{N}\left(\theta_x^0, \left(\sigma_x^0\right)^2\right)$ for each x, assuming that μ_x and μ_y are independent for $x \neq y$.

We consider adaptive sequential policies for simulation allocation, in which the decision-maker simulates one alternative at a time, and the nth simulation decision can depend on the previous $n-1$ decisions and the outcomes of the corresponding simulations. Let $\mathscr{F}^n$ be the sigma-algebra generated by the first n

simulation decisions $x^0, x^1, ..., x^{n-1}$ and by the observations $W^1_{x^0}, W^2_{x^1}, ..., W^n_{x^{n-1}}$. We will use the notation P^n and $\mathbb{E}^n$ to represent conditional probabilities and expectations given $\mathscr{F}^n$.

The conditional distribution of μ_x given $\mathscr{F}^n$ is still normal (DeGroot 1970) with mean θ^n_x and variance $(\sigma^n_x)^2$. The posterior parameters are updated recursively via the equations

$$\theta^{n+1}_x = \begin{cases} \frac{(\sigma^n_x)^{-2}\theta^n_x + \sigma_\varepsilon^{-2} W^{n+1}_x}{(\sigma^n_x)^{-2} + \sigma_\varepsilon^{-2}} & \text{if } x^n = x \\ \theta^n_x & \text{if } x^n \neq x, \end{cases} \tag{1}$$

and

$$\left(\sigma^{n+1}_x\right)^2 = \begin{cases} \left((\sigma^n_x)^{-2} + \sigma_\varepsilon^{-2}\right)^{-1} & \text{if } x^n = x \\ (\sigma^n_x)^2 & \text{if } x^n \neq x. \end{cases} \tag{2}$$

From (1), it can also be shown (Powell and Ryzhov 2012) that the conditional distribution of θ^{n+1}_x given $\mathscr{F}^n$ is normal with mean θ^n_x and variance

$$(\tilde{\sigma}^n_x)^2 = (\sigma^n_x)^2 - \left(\sigma^{n+1}_x\right)^2. \tag{3}$$

Due to the independence assumptions on μ_x and μ_y, an observation of x only provides information about x, and not about any other alternative. Because the posterior is normal, our beliefs are completely characterized by the parameters $\theta^n = (\theta^n_1, ..., \theta^n_M)$ and $\sigma^n = (\sigma^n_1, ..., \sigma^n_M)$. Our end goal will be to design a policy π that chooses an alternative $X^{\pi,n}(\theta, \sigma)$ for the $(n+1)$st simulation based on the most recent available information.

However, before designing such a policy, we first need to specify the objective function that the policy would seek to optimize. In the classical R&S problem considered by Gupta and Miescke (1996) and Frazier et al. (2008), the optimal policy satisfies the objective

$$\sup_\pi \mathbb{E}^\pi \left(\max_x \mathbb{E}^{\pi,N} \mu_x\right) = \sup_\pi \mathbb{E}^\pi \left(\max_x \theta^N_x\right), \tag{4}$$

where N is the total number of simulations available to us. This objective function characterizes a risk-neutral decision-maker (Chick and Gans 2009). At time N, the decision-maker valuates alternative x in terms of its posterior mean. The implementation decision at time N is thus

$$X^{RN,N}\left(\theta^N, \sigma^N\right) = \arg\max_x \theta^N_x, \tag{5}$$

and the learning policy in (4) is chosen to maximize the expected value of this implementation decision.

We propose to replace (5) by the *robust implementation decision*

$$X^{RA,N}\left(\theta^N, \sigma^N\right) = \arg\max_x \left(\min_{\mu \in \mathscr{E}^N} \mu_x\right), \tag{6}$$

where the set $\mathscr{E}^N \subseteq \mathbb{R}^M$ has the property that

$$P^N\left(\mu \in \mathscr{E}^N\right) \geq 1 - \varepsilon \tag{7}$$

for some risk-tolerance parameter ε, specified in advance by the decision-maker. Instead of using the posterior mean to valuate x, we assume that μ_x will be chosen in an adversarial way, and we implement the best alternative for this worst-case scenario. The following result shows that (6) can be transformed into an intuitive generalization of (5). The proof interprets $\mathscr{E}^N$ as an ellipsoid in $\mathbb{R}^M$ and uses Lagrangian duality to rewrite the maxi-min objective.

Proposition 1 The robust implementation decision in (6) can be reformulated as

$$X^{RA,N}\left(\theta^N,\sigma^N\right)=\arg\max_x\left(\theta_x^N-\alpha\sigma_x^N\right), \tag{8}$$

where $\alpha=\sqrt{F^{-1}_{\chi^2_M}(1-\varepsilon)}$ and $F_{\chi^2_M}$ is the cdf of the chi-squared distribution with M degrees of freedom.

The set $\mathscr{E}^N$ is not precisely specified by (7). There are many sets that have this property, and we could potentially use any of them. The set used to derive Proposition 1 is technically convenient, but it also has the property that no point outside $\mathscr{E}^N$ has a density greater than or equal to the density of any point in $\mathscr{E}^N$. In other words, points are included in $\mathscr{E}^N$ in decreasing order of density.

The robust approach to simulation selection thus penalizes our beliefs about an alternative by a factor that grows with our uncertainty about the beliefs. Since σ_x^n decreases every time we measure x, we become less likely to implement an alternative that we have not simulated, or that we have simulated infrequently. Our next step is to replace (4) by the objective

$$\sup_\pi \mathbb{E}^\pi\left(\max_x \theta_x^N-\alpha\sigma_x^N\right), \tag{9}$$

so that we maximize the expected value of the robust implementation decision.

This step merits additional discussion. The meaning of (9) is that the decision-maker is risk-averse with respect to the implementation decision, but risk-neutral with respect to the observations collected from the learning policy, as indicated by the $\mathbb{E}^\pi$ operator. We argue that this is the most appropriate formulation for ranking and selection. A poor implementation decision incurs significant economic costs due to an inefficient factory layout or a contract with an unreliable supplier. While a simulation could also have economic costs (as in Chick and Gans 2009), they are much smaller in magnitude. The purpose of simulation is to experiment with different options before committing to an implementation, so it is reasonable to suppose that the decision-maker will have a higher risk tolerance for simulation decisions.

We support this argument with the following theoretical analysis. Consider a situation where there is one alternative with unknown mean μ and prior parameters θ^0, σ^0, and a second alternative whose mean is known to be zero. Simulating the second alternative thus provides no information, since we already know its value perfectly. Now define

$$V^n\left(\theta^n,\sigma^n\right)=\begin{cases}\max\left\{0,\theta^N-\alpha\sigma^N\right\} & n=N\\ \max\left\{V^{n+1}\left(\theta^n,\sigma^n\right),\min_{W^{n+1}\in\mathscr{D}^n}V^{n+1}\left(\theta^{n+1},\sigma^{n+1}\right)\right\} & n<N,\end{cases} \tag{10}$$

where $\mathscr{D}^n$ satisfies $P^n\left(W^{n+1}\in\mathscr{D}^n\right)\geq 1-\rho$, with ρ being a second risk-tolerance parameter. Define a policy π^* that measures the unknown alternative if

$$\min_{W^{n+1}\in\mathscr{D}^n}V^{n+1}\left(\theta^{n+1},\sigma^{n+1}\right)>V^{n+1}\left(\theta^n,\sigma^n\right), \tag{11}$$

and measures the known alternative otherwise. The policy π^* is robust with respect to both measurement and implementation decisions, and looks out to the end of time horizon, analogous to the optimal policy in the dynamic programming formulation of the risk-neutral R&S problem in Frazier et al. (2008). However, for some choices of ε and ρ, this policy can be shown to conduct no exploration.

Theorem 1 Suppose that $\rho=\frac{\varepsilon}{2}$. Then, for any N, the policy π^* will never measure the unknown alternative.

Proof: We show by induction that

$$V^n\left(\theta^n,\sigma^n\right)=\max\left\{0,\theta^n-\alpha\sigma^n\right\}. \tag{12}$$

This holds for $n=N$ by definition. Suppose now that (12) holds for time $n+1$. Then, (10) becomes

$$V^n\left(\theta^n,\sigma^n\right)=\max\left\{0,\theta^n-\alpha\sigma^n,\min_{W^{n+1}\in\mathscr{D}^n}\left(\max\left\{0,\theta^{n+1}-\alpha\sigma^{n+1}\right\}\right)\right\}.$$

It suffices to prove that

$$\min_{W^{n+1}\in\mathscr{D}^n}\left(\max\left\{0,\theta^{n+1}-\alpha\sigma^{n+1}\right\}\right)\leq\max\left\{0,\theta^n-\alpha\sigma^n\right\}.$$

Observe that

$$\min_{W^{n+1}\in\mathscr{D}^n}\left(\max\left\{0,\theta^{n+1}-\alpha\sigma^{n+1}\right\}\right)=\max\left\{0,\min_{W^{n+1}\in\mathscr{D}^n}\theta^{n+1}-\alpha\sigma^{n+1}\right\}.$$

Recalling (3), we can bound the probability that $\theta^{n+1}-\alpha\sigma^{n+1}<\theta^n-\alpha\sigma^n$ by calculating

$$\begin{aligned}P^n\left(\theta^n+\tilde{\sigma}^nZ-\alpha\sigma^{n+1}<\theta^n-\alpha\sigma^n\right)&=\Phi\left(-\alpha\frac{\sigma^n-\sigma^{n+1}}{\tilde{\sigma}^n}\right)\\&=\Phi\left(-\alpha\sqrt{\frac{\sigma^n-\sigma^{n+1}}{\sigma^n+\sigma^{n+1}}}\right)\\&\geq\Phi(-\alpha)\\&=\frac{\varepsilon}{2}.\end{aligned}$$

The last line comes from the fact that $F_{\chi^2_1}\left(\alpha^2\right)=1-\varepsilon$, and if we have a single unknown alternative,

$$F_{\chi^2_1}\left(\alpha^2\right)=P\left(Z^2<\alpha^2\right)=P\left(-\alpha<Z<\alpha\right)=1-2\Phi(-\alpha),$$

where Z is standard normal. It follows that $\Phi(-\alpha)=\frac{\varepsilon}{2}$.

By the definition of ρ, the set $\mathscr{D}^n$ contains some W^{n+1} for which $\theta^{n+1}-\alpha\sigma^{n+1}<\theta^n-\alpha\sigma^n$, whence

$$\max\left\{0,\min_{W^{n+1}\in\mathscr{D}^n}\theta^{n+1}-\alpha\sigma^{n+1}\right\}<\max\left\{0,\theta^n-\alpha\sigma^n\right\},$$

as required. It follows that

$$\min_{W^{n+1}\in\mathscr{D}^n}V^{n+1}\left(\theta^{n+1},\sigma^{n+1}\right)<V^{n+1}\left(\theta^n,\sigma^n\right),$$

so by (11), the policy π^* will never measure the unknown alternative. □

Theorem 1 suggests that risk-aversion to both measurement and implementation is "too conservative," in that it might cause us to never explore an alternative even with an optimal policy. Even in an infinite horizon ($N\rightarrow\infty$), we will never learn the true value of this alternative. In the ranking and selection literature, this is generally viewed as undesirable behaviour for a learning policy. Furthermore, if $\theta^0-\alpha\sigma^0<0$, there is no chance that we will ever implement the unknown alternative. We thus claim that (9) is the proper objective function for robust R&S, and we now proceed to propose a learning policy for this setting.

3 ALLOCATING THE SIMULATION BUDGET

We apply the value of information approach (surveyed in Chick 2006) to the robust objective function in (9). Essentially, we derive the policy that is optimal for a budget of $N=1$. This policy can be defined as

$$X^{RKG,n}\left(\theta^n,\sigma^n\right)=\arg\max_x\mathbb{E}^n\left[\left(\max_y\theta^{n+1}_y-\alpha\sigma^{n+1}_y\right)-\left(\max_y\theta^n_y-\alpha\sigma^n_y\right)\mid x^n=x\right]. \tag{13}$$

The name RKG means Robust Knowledge Gradient, to emphasize the connection to the related knowledge gradient policy of Frazier et al. (2008). The right-hand side of (13) calculates the expected improvement

in the value of the robust implementation decision due to a single measurement of x. This can be viewed as the marginal value of information, since we only look ahead to the outcome of the next measurement. We then choose the alternative with the highest marginal value. One major advantage of this methodology is that (13) can be computed in closed form.

For a fixed x, we calculate

$$\mathbb{E}^n\left[\max_y \theta_y^{n+1} - \alpha\sigma_y^{n+1} \,|\, x^n = x\right] = \mathbb{E}\max\left\{\left(\max_{y\neq x}\theta_y^n - \alpha\sigma_y^n\right), \left(\theta_x^n - \alpha\sigma_x^{n+1}\right) + \tilde{\sigma}_x^n Z\right\}.$$

This follows from the independence of μ_x and μ_y for $y \neq x$, as well as from the conditional distribution of θ_x^{n+1} given $\mathscr{F}^n$, whose variance was given in (3). Recall from (2) that the posterior variance does not depend on the outcome of the simulation, so σ_x^{n+1} is known given $\mathscr{F}_x^n$ and given $x^n = x$.

Using a computational result by Clark (1961), we calculate

$$\begin{aligned}
&\mathbb{E}\max\left\{\left(\max_{y\neq x}\theta_y^n - \alpha\sigma_y^n\right), \left(\theta_x^n - \alpha\sigma_x^{n+1}\right) + \tilde{\sigma}_x^n Z\right\} \\
= \;& \max\left\{\left(\max_{y\neq x}\theta_y^n - \alpha\sigma_y^n\right), \theta_x^n - \alpha\sigma_x^{n+1}\right\} + \tilde{\sigma}_x^n f\left(-\left|\frac{\left(\theta_x^n - \alpha\sigma_x^{n+1}\right) - \left(\max_{y\neq x}\theta_y^n - \alpha\sigma_y^n\right)}{\tilde{\sigma}_x^n}\right|\right)
\end{aligned}$$

where $f(z) = z\Phi(z) + \phi(z)$ and ϕ is the standard normal pdf. We can now rewrite (13) as

$$X^{RKG,n}(\theta^n, \sigma^n) = \arg\max_x \nu_x^{RKG,n}$$

where

$$\nu_x^{RKG,n} = \tilde{\sigma}_x^n f\left(-\left|\frac{\left(\theta_x^n - \alpha\sigma_x^{n+1}\right) - \left(\max_{y\neq x}\theta_y^n - \alpha\sigma_y^n\right)}{\tilde{\sigma}_x^n}\right|\right) + \tilde{\nu}_x^n, \tag{14}$$

$$\tilde{\nu}_x^n = \max\left\{\left(\max_{y\neq x}\theta_y^n - \alpha\sigma_y^n\right), \theta_x^n - \alpha\sigma_x^{n+1}\right\} - \left(\max_y \theta_y^n - \alpha\sigma_y^n\right). \tag{15}$$

In the special case where $\alpha = 0$, we have $\tilde{\nu}_x^n = 0$ and (14) becomes

$$\nu_x^{KG,n} = \tilde{\sigma}_x^n f\left(-\left|\frac{\theta_x^n - \max_{y\neq x}\theta_y^n}{\tilde{\sigma}_x^n}\right|\right),$$

which is precisely the knowledge gradient policy of Frazier et al. (2008), designed for the risk-neutral objective (4). This is consistent with our formulation in (9) of robust R&S as a generalization of the risk-neutral problem with an extra penalty for variance. In the robust setting, the value of information $\nu_x^{RKG,n}$ consists of two components. The first term on the right-hand side of (14) represents the potential of the random observation to improve our implementation decision, just as in the risk-neutral setting. The second term, defined in (15), represents the benefits obtained by reducing the variance in our beliefs, which also improves our implementation decision by shrinking the uncertainty ellipsoid $\mathscr{E}^N$.

We now summarize our asymptotic analysis of the RKG policy, with the full derivations given in Defourny et al. (2012). Like its risk-neutral counterpart, the robust value of information is bounded. It follows that, as $N \to \infty$, $\nu_x^{RKG,n}$ has an almost sure limit.

Proposition 2 For all n, $\nu_x^{RKG,n} \leq \left(\alpha + \frac{1}{\sqrt{2\pi}}\right) \cdot \left(\max_x \sigma_x^0\right)$ almost surely.

Using the similar structure of $\nu_x^{RKG,n}$ and $\nu_x^{KG,n}$, it can be shown that $\nu_x^{RKG,n} \to 0$ almost surely if x is measured infinitely often. It can also be shown that $\nu_x^{RKG,n} = 0$ if and only if $\sigma_x^n = 0$. Combining these properties, we arrive at the following convergence result. It is worth noting that the proof only requires our simulation outcomes to be unbiased with respect to the true values. We do not actually need the Bayesian modeling assumptions on μ in order to obtain convergence.

Theorem 2 The event that the RKG policy measures every x infinitely often occurs w.p. 1.

By the strong law of large numbers, it follows that $\theta_x^n \to \mu_x$ a.s. under the robust KG policy. Although RKG is more conservative than its risk-neutral counterpart, it still conducts a sufficient amount of exploration, in contrast with the result of Theorem 1, where even the optimal policy for risk-averse measurement decisions could produce inconsistent results. Theorem 2 provides additional evidence that the right way to model the robust R&S problem is by assuming risk-averse implementation decisions and risk-neutral measurement decisions.

4 NUMERICAL EXAMPLES

We now discuss several numerical examples illustrating situations where the robust policy helps to reduce the risk of making a poor implementation decision. These experiments were conducted on simulated data. Each example considers a problem with $M = 50$ alternatives. The prior means θ_x^0 were set to zero for all x, and the prior variances $\left(\sigma_x^0\right)^2$ were chosen from a uniform distribution on $[50, 450]$. In each experiment, we ran 10^4 simulations, each consisting of some fixed number N of measurements. At the beginning of each simulation, we generated a true value $\mu_x \sim \mathcal{N}\left(\theta_x^0, \left(\sigma_x^0\right)^2\right)$. The simulations thus covered a wide variety of configurations of true values.

Our objective in these experiments was to compare performance in risk-neutral and robust settings to obtain insights into exactly how robust R&S differs from the classical version. In order to control for the effects of the policy as much as possible, we compared RKG to its risk-neutral counterpart (labeled simply KG) from Frazier et al. (2008). We used two criteria

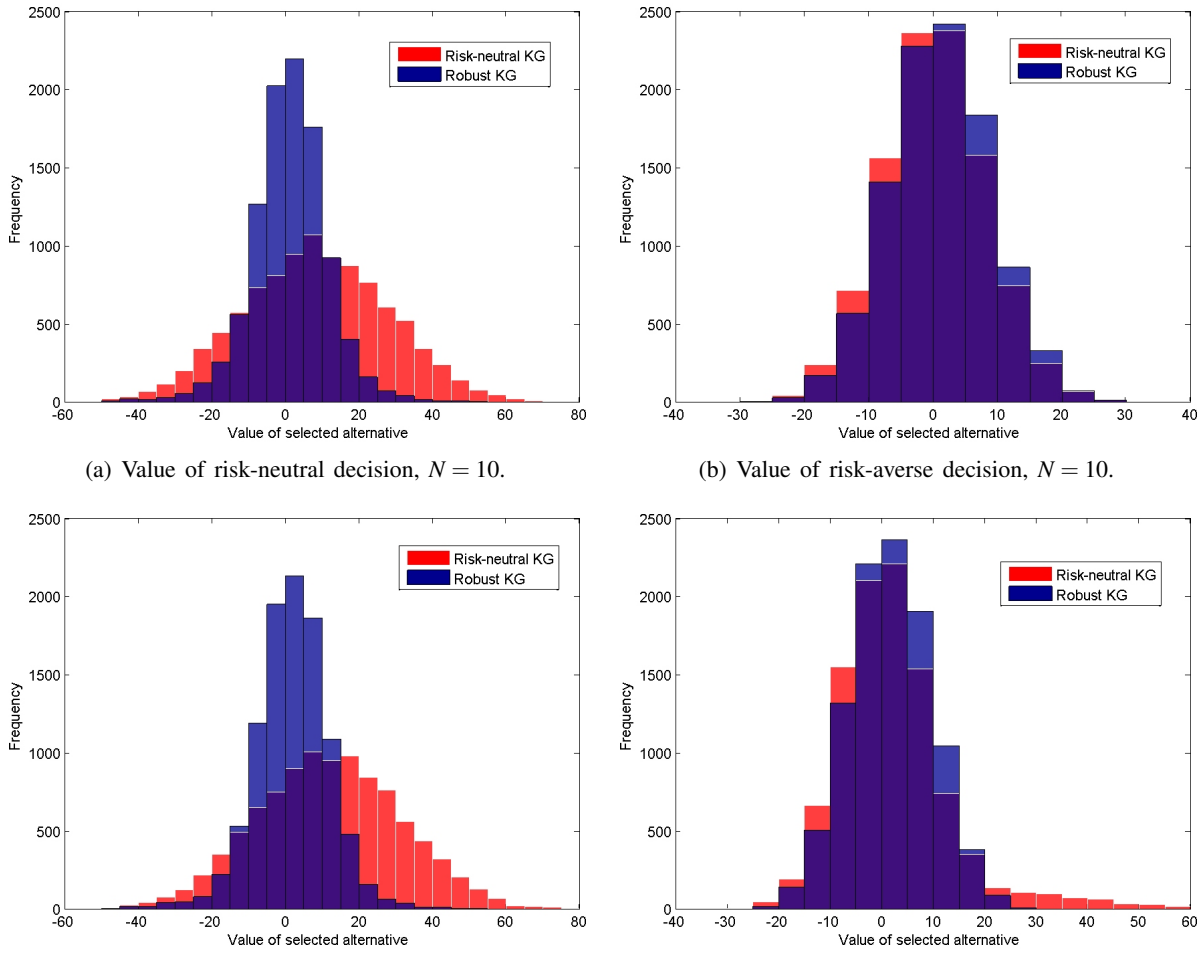
$$C^{RN,N} = \mu_{X^{RN,N}(\theta^N,\sigma^N)}, \qquad C^{RA,N} = \mu_{X^{RA,N}(\theta^N,\sigma^N)}.$$

Four experiments were conducted in total, all with the same value of α. The first three considered different time horizons with N equal to 10, 20, and 50, but were otherwise set up in the same way, with $\sigma_\varepsilon^2 = 10^4$. The fourth experiment used the value $\sigma_\varepsilon^2 = 10^2$ to update the beliefs in (1) and (2), but the actual observations W_x^{n+1} were sampled from a different distribution $\mathcal{N}\left(\mu_x, 10^4\right)$ with much higher measurement noise. This experiment models a situation where the decision-maker is over-confident about the accuracy of the simulation model, and uses a value of σ_ε^2 that is much lower than the actual sampling noise. We refer to this as an "improper" model.

Table 1 summarizes the empirical means and standard errors of these values over 10^4 different samples of μ from the prior. In the first three experiments, we see several consistent trends. For both policies, the robust implementation decision is much more conservative (achieves smaller performance values) than risk-neutral implementation, but these values also exhibit much smaller variance. Switching to RKG achieves a similar effect: the policy leads us to lower performance values, but achieves those results much

Table 1: Means and standard errors of the values of risk-neutral and risk-averse implementation decisions under different learning policies.

		Risk-neutral implementation		Risk-averse implementation	
Problem	Policy	Mean value	Std. error	Mean value	Std. error
$N = 10$	KG	8.2818	402.6150	0.2694	69.3759
	RKG	1.2686	113.4012	1.1422	62.8785
$N = 20$	KG	11.8275	392.5398	2.3608	136.5311
	RKG	1.8784	109.6800	1.7464	62.6704
$N = 50$	KG	17.7332	342.1730	12.1738	322.0734
	RKG	2.6673	108.3321	2.3402	61.5915
Improper	KG	11.5074	321.8716	11.5616	321.6481
	RKG	15.9245	374.3321	16.3508	360.8486



(a) Value of risk-neutral decision, $N = 10$.

(b) Value of risk-averse decision, $N = 10$.

(c) Value of risk-neutral decision, $N = 20$.

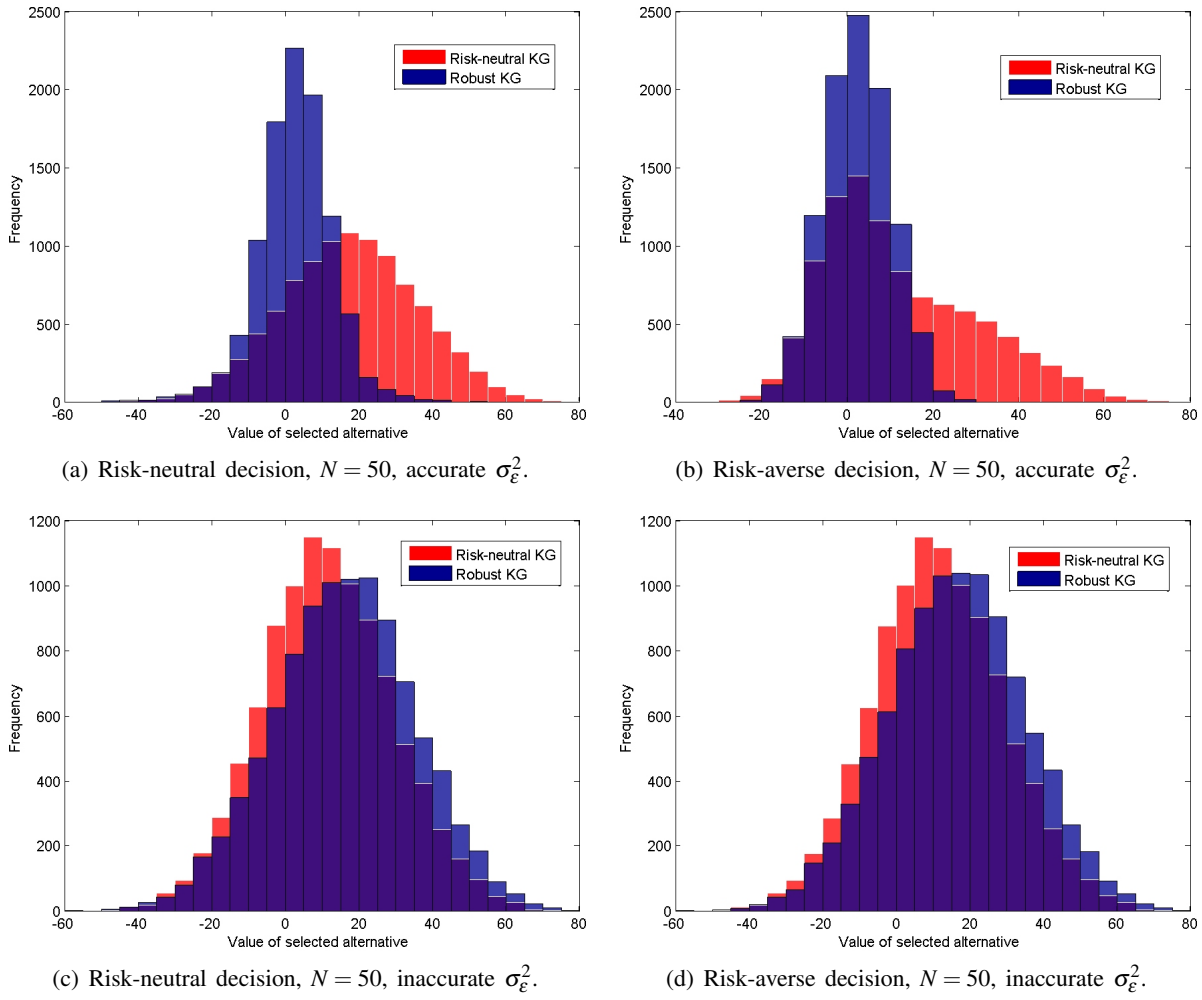
(d) Value of risk-averse decision, $N = 20$.

Figure 1: Empirical distributions of the values of the risk-neutral and risk-averse implementation decisions over 10^4 simulations.

more consistently. Under risk-neutral implementation decisions, RKG achieves a threefold reduction in variance over KG. For risk-averse implementation decisions, the variance grows for KG, but stays nearly constant for RKG.

The results for the improper model are quite different. The variance for both policies is similar in magnitude (actually slightly greater for RKG). However, RKG now outperforms KG on average, as well. Recall that this is a model where the decision-maker specified a value of σ_ε^2 that was too low, thus placing too much confidence in new information. In this setting, a risk-averse implementation decision actually produces better results on average, enhanced even further by the use of RKG. This appears to suggest that robust R&S may be a useful model in a situation where it is difficult to obtain accurate estimates of σ_ε^2, perhaps due to lack of prior experience with the problem. In the literature, an alternate approach to this problem is to put a prior on the sampling variance or to estimate it through a separate procedure; we note, however, that for small measurement budgets, such estimates will be subject to considerable uncertainty.

However, the table does not give a complete picture. Figures 1 and 2 give histograms of the observed values of $C^{RN,N}$ and $C^{RA,N}$ over 10^4 sample paths. For the risk-neutral criterion, we see a classic trade-off between risk and return: the distribution for RKG shows less spread, but the peak itself is smaller than the peak for the risk-neutral policy. Higher values of the risk-tolerance parameter α will decrease the variance



(a) Risk-neutral decision, $N = 50$, accurate σ_ε^2.

(b) Risk-averse decision, $N = 50$, accurate σ_ε^2.

(c) Risk-neutral decision, $N = 50$, inaccurate σ_ε^2.

(d) Risk-averse decision, $N = 50$, inaccurate σ_ε^2.

Figure 2: Empirical distributions of the values of the risk-neutral and risk-averse implementation decisions over 10^4 simulations.

even more, but the peak will move even further to the left; lower values of α will lead to greater similarity between the two policies.

For the given value of α, we see that RKG reduces the negative tails in Figures 1(a) and 1(c). However, in Figure 2(a), this effect is virtually gone, and the reduction in variance comes mostly at the expense of the positive tails. The insight here is that, with a properly specified model, the given sampling budget $N = 50$ is large enough to find a good alternative with sufficiently high accuracy. At that point, worst-case scenarios become so unlikely that RKG adds little value.

At the same time, when the model is improperly specified, $N = 50$ is much more restrictive. Thus, in Figure 2(c), RKG actually begins to outperform KG on average, with a much larger positive tail. Our conclusion is that risk-aversion can become very valuable when the sampling budget is relatively small, or when the prior information is inaccurate.

Finally, if we elect to use the robust implementation decision, RKG appears to be a better policy overall than KG, reducing the positive tails in Figures 1(b) and 1(d) and moving the entire histogram to the right. Even for $N = 50$, RKG still produces a slight reduction in the negative tails, although for this budget the effect is insignificant. However, when we switch to the improper model, Figure 2(d) shows that RKG yields the same benefits as in Figure 2(c).

While these experiments are based on a particular set of simulated data, they do suggest that robust R&S primarily adds value when the given simulation budget can be used to help us make a better decision, but is too small to conclusively identify the best alternative. The meaning of "too small" is problem-dependent and may be due to high measurement noise or an improperly specified model. Overall, we recommend robust R&S as a tool for small-sample simulation where a single replication is very expensive or very noisy, as well as for problems where information has to be collected from time-consuming experiments in the field.

5 CONCLUSION

We have proposed a framework for ranking and selection in which the decision-maker is risk-averse with respect to the outcome of the final selection decision. Simulations are allocated according to a Bayesian value of information policy that is risk-neutral with respect to measurement decisions. This setup allows us to make more conservative decisions, while still maintaining a sufficient degree of exploration. Experimental results suggest that this approach adds the most value when observations are noisy, the simulation budget is small, and the learning model is incorrectly specified.

This paper has not discussed how the concept of robust R&S may be carried over to more complex learning models. However, our preliminary work has shown that computationally tractable robust policies can also be derived for problems with correlated beliefs (where a single measurement provides information about multiple alternatives) as well as for global optimization problems with continuous decision spaces. We believe that the conjunction of robust optimization and ranking and selection offers a new way to think about hedging risk in simulation optimization.

REFERENCES

Bechhofer, R. E., T. J. Santner, and D. M. Goldsman. 1995. *Design and Analysis of Experiments for Statistical Selection, Screening and Multiple Comparisons*. New York: J.Wiley & Sons.

Ben-Tal, A., L. El Ghaoui, and A. Nemirovski. 2009. *Robust Optimization*. Princeton University Press.

Ben-Tal, A., and A. Nemirovski. 2002. "Robust optimization – methodology and applications". *Mathematical Programming* 92 (3): 453–480.

Bertsimas, D., D. B. Brown, and C. Caramanis. 2007. "Theory and Applications of Robust Optimization". *SIAM Review* 53 (3): 464–501.

Branke, J., S. Chick, and C. Schmidt. 2007. "Selecting a Selection Procedure". *Management Science* 53 (12): 1916–1932.

Chick, S. E. 2006. "Subjective Probability and Bayesian Methodology". In *Handbooks of Operations Research and Management Science, vol. 13: Simulation*, edited by S. Henderson and B. Nelson, 225–258. North-Holland Publishing, Amsterdam.

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.

Chick, S. E., and P. I. Frazier. 2009, December. "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.

Chick, S. E., and N. Gans. 2009. "Economic analysis of simulation selection problems". *Management Science* 55 (3): 421–437.

Clark, C. E. 1961. "The greatest of a finite set of random variables". *Operations Research* 9 (2): 145–162.

Defourny, B., I. O. Ryzhov, and W. B. Powell. 2012. "The robust approach to simulation selection". *Working paper, Princeton University*.

DeGroot, M. H. 1970. *Optimal Statistical Decisions*. John Wiley and Sons.

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.

Gupta, S., and K. Miescke. 1994. "Bayesian look ahead one stage sampling allocations for selecting the largest normal mean". *Statistical Papers* 35:169–177.

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 (2): 229–244.

Hong, L. J., and B. L. Nelson. 2009, December. "A Brief Introduction To Optimization Via Simulation". In *Proceedings of the 2009 Winter Simulation Conference*, edited by M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin, and R. G. Ingalls, 75–85. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, 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.

Kim, S.-H., and B. L. Nelson. 2006. "Selecting the best system". In *Handbooks of Operations Research and Management Science, vol. 13: Simulation*, edited by S. G. Henderson and B. L. Nelson, 501–534. North-Holland Publishing, Amsterdam.

Powell, W. B., and I. O. Ryzhov. 2012. *Optimal Learning*. John Wiley and Sons.

Waeber, R., P. I. Frazier, and S. G. Henderson. 2010, December. "Performance Measures for Ranking and Selection Procedures". In *Proceedings of the 2010 Winter Simulation Conference*, edited by B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan, 1235–1245. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Wang, H., and S.-H. Kim. 2011. "On the Conservativeness of Fully Sequential Indifference-Zone Procedures". *Submitted for publication*.

AUTHOR BIOGRAPHIES

ILYA O. RYZHOV is an Assistant Professor in the Robert H. Smith School of Business at the University of Maryland. He received a Ph.D. in Operations Research and Financial Engineering from Princeton University in 2011. His research deals with the interface between optimal learning and the broader area of stochastic optimization, with applications in disaster relief, energy, and operations management. His work has appeared in *Operations Research*, and he is a co-author of the book *Optimal Learning*, published in 2012 by John Wiley & Sons. His email address is iryzhov@rhsmith.umd.edu.

BORIS DEFOURNY is an Associate Professional Specialist in the Department of Operations Research and Financial Engineering at Princeton University. He received an Electrical Engineering degree in 2005, and a Ph.D. in Applied Sciences in 2010, from the University of Liege, Belgium. His research is in sequential decision-making under uncertainty, using techniques from stochastic programming, robust optimization, and machine learning. He works on stochastic optimization problems for electric power systems. His email address is defourny@princeton.edu.

WARREN B. POWELL is a Professor in the Department of Operations Research and Financial Engineering at Princeton University, and director of CASTLE Laboratory (http://www.castlelab.princeton.edu) and the Princeton Laboratory for Energy Systems Analysis (http://energysystems.princeton.edu). He has coauthored over 150 refereed publications in stochastic optimization, stochastic resource allocation and related applications. He is the author of the book *Approximate Dynamic Programming: Solving the curses of dimensionality* and a co-author of *Optimal Learning*, published by John Wiley & Sons. Currently, he is involved in applications in energy, transportation, finance and homeland security. His email address is powell@princeton.edu.