RANKING AND SELECTION WITH UNKNOWN CORRELATION STRUCTURES

Ilya O. Ryzhov

Huashuai Qu

Michael C. Fu

Department of Mathematics
University of Maryland
College Park, MD 20742, USA

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

ABSTRACT

We create the first computationally tractable Bayesian statistical model for learning unknown correlations among estimated alternatives in fully sequential ranking and selection. Although correlations allow us to extract more information from each individual simulation, the correlation structure is itself unknown, and we face the additional challenge of simultaneously learning the unknown values and unknown correlations from simulation. We derive a Bayesian procedure that allocates simulations based on the value of information, thus exploiting the correlation structure and anticipating future changes to our beliefs about the correlations. We test the model and algorithm in a simulation study motivated by the problem of optimal wind farm placement, and obtain encouraging empirical results.

1 INTRODUCTION

Ranking and selection (R&S) deals with the problem of selecting the best of K design alternatives. The mean performance values of these alternatives are unknown, but can be estimated through sequential simulation. Classical ranking and selection create independent estimates of each mean value, and use these estimates to allocate simulation budget between alternatives. However, in many applications, the performance values are naturally correlated, in a way that is itself difficult to estimate.

As a motivating example, consider the problem of choosing a location for a new wind farm. We have a number of candidate locations in a certain region, and we have access to historical data (Cosgrove et al. 2003) about the wind speeds at these different locations. Wind speeds are known to be volatile, however, and energy output also depends on a variety of other factors such as pressure gradient, frictional forces, and Coriolis effect, so we would like to perform an additional investigation of several locations before we commit to one of them. Suppose that we can estimate the energy output at a given location using a simulator that models the physics of the problem with stochastic wind speeds. The output at different locations will be highly correlated. However, the correlation between two locations is not a simple function of physical distance, but also depends on other factors such as wind currents and the topography of the region.

A new and difficult problem arises. Correlation should work to our advantage: simulating one location should also provide information about similar locations. However, we do not know exactly which locations are "similar." Our goal is to find the location with the highest mean performance value, exactly as in the classic ranking and selection framework, but the task is complicated by the presence of unknown correlations that have to be learned together with the means. The problem of unknown correlations has received almost no attention in the literature to date. Previous work on correlated R&S has dealt almost exclusively with known correlation structures or with correlations arising from common random numbers.

Classical R&S uses frequentist statistics to make decisions; see Bechhofer et al. (1995) for an overview of such methods. Such procedures often have the objective (known as the indifference-zone criterion) of ensuring that the best alternative is selected with a certain probability. Numerous surveys and tutorials on

indifference-zone selection are available, including Kim and Nelson (2006), Hong and Nelson (2009) and Kim (2013).

A different approach to the same problem uses Bayesian statistics to express the decision-maker's belief about the unknown values. The procedures then make decisions by trading off the estimated value of an alternative against the decision-maker's uncertainty about this estimate. This school of thought, known variously as "value of information" or "knowledge gradient" procedures, goes back to Gupta and Miescke (1996). Chick (2006) surveys the Bayesian approach; see Chick et al. (2010) for an example of more recent work. The method of optimal computing budget allocation (see Chen et al. 2008, or the monograph by Chen and Lee 2010) can also be included in this category, as it allows for Bayesian beliefs. Powell and Ryzhov (2012) provides an up-to-date survey of Bayesian learning techniques.

Correlations have been addressed to some extent in the literature. Most such studies assume that correlations are induced by common random numbers inside the simulator; see e.g., Yang and Nelson (1991) or Nelson and Matejcik (1995). Recent work in the Bayesian setting (Fu et al. 2007; Frazier et al. 2009) allows the decision-maker to have correlated beliefs, so that a single simulation provides information about multiple alternatives. In applications, correlated beliefs significantly improve performance and enable us to solve very large problems (even problems with continuous decision spaces, as in Scott et al. 2010). Yet, there has been almost no work on Bayesian R&S with *unknown* correlation structures. To our knowledge, Chick and Inoue (2001) is the only work of this kind in the simulation literature.

This lack of attention may be due to the difficulty of creating a computationally tractable Bayesian statistical model for learning unknown correlations from observations of individual alternatives. The normal-Wishart distribution is a standard conjugate learning model for situations where each alternative can be sampled simultaneously; however, in our setting, we observe only one alternative at a time. In statistics, this is known as a "missing-data" problem, since we are trying to estimate a multivariate distribution from scalar observations. Statisticians have made a number of attempts to derive conjugate prior distributions for learning models with missing data; examples include Kadane and Trader (1988), Dominici et al. (2000), or a recent work by Triantafyllopoulos (2008). However, most of these models present serious computational difficulties in an R&S setting. Chick and Inoue (2001) uses a tractable prior, but limits decisions to a very particular structure (first screening out some alternatives, then allocating simulations equally among the rest).

We present what we believe to be the first work to address this problem from an R&S standpoint. We create a new Bayesian model for simultaneously learning unknown means and unknown correlation structures on the basis of scalar observations. Although our prior is not exactly conjugate, we create an optimal approximation of conjugacy by minimizing the Kullback-Leibler (KL) divergence between the true posterior and the normal-Wishart distribution (a technique known as density projection). We then derive a new value of information procedure for this new problem setting that requires no more computational effort than an existing procedure for known correlation structures. This paper summarizes our work on this problem and presents simulation experiments motivated by the wind farm placement problem. The numerical results demonstrate that the new procedure is very promising.

This paper is organized as follows. Section 2 describes the problem of R&S with unknown correlation structures and presents our Bayesian learning model. Section 3 introduces the KG policy for choosing simulations and summarizes our study of its properties. Section 4 describes our experiments. Finally, we conclude the paper in Section 5.

2 LEARNING UNKNOWN CORRELATION STRUCTURES

We summarize the model and give a sketch of the derivation; the full details for all the results presented in this section can be found in Qu et al. (2012).

Let $\{1, 2, \cdots, K\}$ be a set of alternatives. Let $\hat{\mathbf{Y}}$ be a multivariate normal random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. We assume that $\boldsymbol{\Sigma}$ is invertible and define $\mathbf{R} = \boldsymbol{\Sigma}^{-1}$ to be the *precision matrix* of $\hat{\mathbf{Y}}$. The vector $\hat{\mathbf{Y}}$ describes the behavior of K alternatives, all observed concurrently. Both $\boldsymbol{\mu}$ and $\mathbf{R}$ are

unknown to us, and our goal is to discover the alternative x with the largest underlying mean μ_x. Let $\hat{y}_x$ be a scalar random variable following the marginal distribution of the xth component of $\hat{\mathbf{Y}}$. We view $\hat{y}_x$ as the output of a stochastic simulation of the behavior of the xth alternative. The sampling distribution given $\boldsymbol{\mu}$ and $\mathbf{R}$ is univariate normal with probability density function

$$p(\hat{y}_x|\boldsymbol{\mu},\mathbf{R}) \propto \left|e_x'\mathbf{R}^{-1}e_x\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\hat{y}_x - \mu_x)'(e_x'\mathbf{R}^{-1}e_x)^{-1}(\hat{y}_x - \mu_x)\right\}, \tag{1}$$

where $e_x = (0,\ldots,1,\ldots,0)$ is a $K \times 1$ vector, with 1 at the xth component, and 0 at others. Here the prime denotes transpose. We write (1) in terms of the kernel of the probability density, omitting a normalizing constant.

Correlations between alternatives are incorporated into the non-diagonal precision matrix $\mathbf{R}$. Therefore, it is logical to suppose that a single observation $\hat{y}_x$ should also provide some information about other alternatives $x' \neq x$ that are correlated with x. Unfortunately, the exact correlation structure is unknown, so we do not know how much extra information is provided.

2.1 Learning Model

We adopt the Bayesian view, where the unknown $\boldsymbol{\mu}$ and $\mathbf{R}$ are viewed as a random vector and a random matrix respectively. Their distribution represents the decision-maker's prior beliefs about the unknown quantities. We assume that

$$\boldsymbol{\mu}|\mathbf{R} \sim \mathcal{N}_K(\boldsymbol{\theta}, q\mathbf{R}), \quad \mathbf{R} \sim \mathcal{W}_K(b, \mathbf{B}).$$

This notation means that $\mathbf{R}$ is a random matrix following a Wishart distribution, whereas $\boldsymbol{\mu}$ is multivariate normal conditional on $\mathbf{R}$. The Wishart distribution (see e.g., Gupta and Nagar 2000 for a detailed discussion) is parametrized by a scalar b and a $K \times K$ matrix $\mathbf{B}$, with the density

$$p(\mathbf{R}) = \frac{1}{Z(b,\mathbf{B})}|\mathbf{R}|^{\frac{b-K-1}{2}} \exp\left\{-\frac{1}{2}\operatorname{tr}(\mathbf{B}\mathbf{R})\right\},$$

where

$$Z(b,\mathbf{B}) = \pi^{\frac{K(K-1)}{4}}\left|\frac{\mathbf{B}}{2}\right|^{-\frac{b}{2}} \prod_{i=1}^{K}\Gamma\left(\frac{b+1-i}{2}\right)$$

is a normalization constant. Therefore the joint prior distribution of $\boldsymbol{\mu}$ and $\mathbf{R}$ is

$$p(\boldsymbol{\mu},\mathbf{R}) = \frac{1}{Z(b,\mathbf{B})}|\mathbf{R}|^{\frac{b-K-1}{2}} \exp\left\{-\frac{1}{2}\operatorname{tr}(\mathbf{B}\mathbf{R})\right\}\left(\frac{q}{2\pi}\right)^{\frac{K}{2}}|\mathbf{R}|^{\frac{1}{2}} \exp\left\{-\frac{q}{2}(\boldsymbol{\mu} - \boldsymbol{\theta})'\mathbf{R}(\boldsymbol{\mu} - \boldsymbol{\theta})\right\},$$

where the vector $\boldsymbol{\theta}$ and scalar q are additional parameters. If we could observe the entire vector $\hat{\mathbf{Y}} \sim \mathcal{N}_K(\boldsymbol{\mu}, \mathbf{R})$, standard results from Bayesian analysis (DeGroot 2004) tell us that the posterior density

$$p(\boldsymbol{\mu},\mathbf{R}|\hat{\mathbf{Y}}) = \frac{p(\hat{\mathbf{Y}}|\boldsymbol{\mu},\mathbf{R})p(\boldsymbol{\mu},\mathbf{R})}{\iint p(\hat{\mathbf{Y}}|\boldsymbol{\mu},\mathbf{R})p(\boldsymbol{\mu},\mathbf{R})d\boldsymbol{\mu}d\mathbf{R}}$$

would also be a normal-Wishart distribution with parameters

$$q^* = q + 1, \tag{2}$$

$$b^* = b + 1, \tag{3}$$

$$\boldsymbol{\theta}^* = \frac{q\boldsymbol{\theta} + 1}{q + 1}, \tag{4}$$

$$\mathbf{B}^* = \mathbf{B} + \frac{q}{q+1}(\boldsymbol{\theta} - \hat{\mathbf{Y}})(\boldsymbol{\theta} - \hat{\mathbf{Y}})'. \tag{5}$$

This property is known as *conjugacy* of the normal-Wishart prior and the normal sampling distribution.

Now suppose we sample from alternative x only (as is common in fully sequential R&S), collecting observation $\hat{y}_x \sim \mathcal{N}(\mu_x, e_x' \mathbf{R}^{-1} e_x)$. Applying Bayes' rule, the joint posterior distribution of $\boldsymbol{\mu}$ and $\mathbf{R}$, given the observation $\hat{y}_x$, can be written as

$$p(\boldsymbol{\mu}, \mathbf{R}|\hat{y}_x) \propto \frac{1}{Z(b, \mathbf{B})} |\mathbf{R}|^{\frac{b-K-1}{2}} \exp\left\{-\frac{1}{2} \mathrm{tr}(\mathbf{BR})\right\} \left(\frac{q}{2\pi}\right)^{\frac{K}{2}}$$

$$\cdot |\mathbf{R}|^{\frac{1}{2}} \exp\left\{-\frac{q}{2}(\boldsymbol{\mu} - \boldsymbol{\theta})' \mathbf{R}(\boldsymbol{\mu} - \boldsymbol{\theta})\right\} (2\pi \mathbf{R}^{-1})_{xx}^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{R}^{-1})_{xx}^{-1}(\hat{y}_x - \mu_x)^2\right\}. \qquad (6)$$

Factoring the distribution in (6) into the conditional distribution of $\boldsymbol{\mu}$ given $\mathbf{R}$ and the marginal distribution of $\mathbf{R}$, we observe that the conditional distribution of $\boldsymbol{\mu}$ given $\mathbf{R}$ is a multivariate normal distribution, but the marginal distribution of $\mathbf{R}$ is no longer a Wishart distribution.

The distribution in (6) presents computational difficulties. If we can no longer observe the entire vector $\hat{\mathbf{Y}}$, the conjugacy property is lost. The non-conjugacy of our prior distribution makes it difficult to develop sequential simulation procedures, which rely on our ability to compactly represent our uncertainty with a small number of parameters. Instead of working with (6) directly, we force conjugacy with the density projection technique. We replace the posterior in (6) by a normal-Wishart distribution whose parameters are chosen to resemble the true posterior as closely as possible, in the sense of minimizing the Kullback-Leibler (KL) divergence.

Let $\xi(\boldsymbol{\mu}, \mathbf{R})$ be a distribution from the normal-Wishart family such that

$$\boldsymbol{\mu}|\mathbf{R} \sim \mathcal{N}_k(\boldsymbol{\theta}^*, q^* \mathbf{R}), \quad \mathbf{R} \sim \mathcal{W}_K(b^*, \mathbf{B}^*).$$

Define $\mathcal{D}(\xi|p)$ to be the Kullback-Leibler (KL) divergence between $p(\boldsymbol{\mu}, \mathbf{R}|\hat{y}_x)$ and $\xi(\boldsymbol{\mu}, \mathbf{R})$, which is given by

$$\mathcal{D}(\xi|p) = \mathbb{E}_\xi \left(\log \frac{\xi(\boldsymbol{\mu}, \mathbf{R})}{p(\boldsymbol{\mu}, \mathbf{R}|y_x)}\right), \qquad (7)$$

where $\mathbb{E}_\xi[\cdot]$ is the expectation with respect to $\xi(\boldsymbol{\mu}, \mathbf{R})$. This quantity is used to measure the "distance" between the distributions ξ and p. Lower KL divergence suggests that there is more similarity between the two distributions. We wish to find

$$(q^*, b^*, \boldsymbol{\theta}^*, \mathbf{B}^*) = \mathrm{argmin}\, \mathbb{E}_\xi [\mathcal{D}(\xi|p)],$$

the set of parameters that project (according to KL divergence) the normal-Wishart distribution on the true posterior in (6).

Proposition 1 The KL divergence $\mathcal{D}(\xi|p)$ defined in (7) can be expressed as

$$\mathcal{D}(\xi|p) = \frac{b^* - b}{2} \left(-\log|\mathbf{B}^*| + \sum_{i=1}^{K} \psi\left(\frac{b^* - i + 1}{2}\right)\right) - \frac{b^* K}{2} + \frac{b^*}{2} \mathrm{tr}\left(\mathbf{B}(\mathbf{B}^*)^{-1}\right)$$

$$+ \log \frac{Z(b, \mathbf{B})}{Z(b^*, \mathbf{B}^*)} + \frac{1}{2}\left[K \log \frac{q^*}{q} + K \frac{q}{q^*} + q(\boldsymbol{\theta} - \boldsymbol{\theta}^*)' b^* (\mathbf{B}^*)^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}^*)\right]$$

$$+ \frac{1}{2} \log \mathbf{B}_{xx}^* - \frac{1}{2} \psi\left(\frac{b^* - K + 1}{2}\right) + \frac{1}{2q^*} + \frac{1}{2}(\hat{y}_x - \theta_x^*)^2 \frac{b^* - K + 1}{\mathbf{B}_{xx}^*} + C,$$

where $\psi(x) = d\ln\Gamma(x)/dx$ is the digamma function and C is some constant.

Using Proposition 1, we summarize the results in the following theorem.

Theorem 2 The normal-Wishart distribution that minimizes the KL divergence $\mathscr{D}(\xi|p)$ defined in (7) has the parameters

$$q^* = q + \frac{1}{K}, \tag{8}$$

$$b^* = b + \Delta b, \tag{9}$$

$$\boldsymbol{\theta}^* = \left(qb^*(\mathbf{B}^*)^{-1} + \frac{b^* - K + 1}{e_x'\mathbf{B}^* e_x} e_x e_x' \right)^{-1} \left(qb^*(\mathbf{B}^*)^{-1}\boldsymbol{\theta} + \frac{b^* - K + 1}{e_x'\mathbf{B}^* e_x} \hat{y}_x e_x \right), \tag{10}$$

$$\mathbf{B}^* = \mathbf{B} + \frac{\Delta b}{b}\mathbf{B} + \left(\frac{b^*}{b+1} \right) \left(\frac{q}{\frac{qb^*}{b^* - K + 1} + 1} \frac{(\hat{y}_x - \theta_x)^2}{B_{xx}^2} - \frac{1}{bB_{xx}} \right) \mathbf{B} e_x e_x' \mathbf{B}. \tag{11}$$

Using the Sherman-Morrison-Woodbury formula (see e.g., Golub and Van Loan 1996), we can rewrite (10) without inverse matrices as

$$\boldsymbol{\theta}^* = \boldsymbol{\theta} + \frac{\hat{y}_x - \theta_x}{\frac{qb^*}{b^* - K + 1} B_{xx} + B_{xx}} \mathbf{B} e_x. \tag{12}$$

The most crucial aspect of (12) is that a single scalar observation is now used to update the entire posterior mean vector through the matrix $\mathbf{B}$. Similar behavior occurs in the Kalman filter-like update used by Frazier et al. (2009) in the case of known correlation structures. In that setting, the update incorporates both the variance of the current belief and the known variance of the observations. However, when the correlation structure is unknown, the matrix $\mathbf{B}$ is used to estimate both variances.

The parameter q in the prior distribution is not a prior sample size; it is intended to be a reflection of prior precision relative to the sample size that is tunable by the researcher or practitioner to reflect prior confidence. It increases by $1/K$, analogously to the result for complete observations in (2), where a single vector sample increases q by 1. When we only sample from one alternative, the confidence parameter should be increased by $1/K$.

The parameter b is also increased by 1 when we have complete observations. However, when we sample from only one alternative, the increment Δb actually depends on $(q, b, \hat{y}_x, \theta_x, B_{xx})$. The quantity Δb does not have a closed-form expression, but can easily be obtained numerically via a bisection procedure on the interval $[0, 1]$. However, we have observed in our experiments that the optimal values of Δb are smaller than $1/K$ and approach $1/K$ asymptotically over time.

2.2 Predictive Distribution

The predictive distribution of $\hat{y}_x$ represents the decision-maker's beliefs about the next observation given the prior distribution on the unknown parameters. In Section 3, we introduce a policy that uses the predictive distribution to look ahead to the outcome of a simulation decision. In preparation for this discussion, we now summarize the predictive distribution for the normal-Wishart model.

We provide the definition of the multivariate t distribution for completeness (Kotz and Nadarajah 2004).

Definition 1 A p-dimensional random vector $\mathbf{X} = (X_1, \ldots, X_p)$ is said to have the p-variate t distribution with degrees of freedom ν, mean vector $\boldsymbol{\mu}$, and correlation matrix $\mathbf{R}$ if its joint pdf is given by

$$f(\mathbf{x}) = \frac{\Gamma((\nu + p)/2)}{(\pi\nu)^{p/2}\Gamma(\nu/2)|\mathbf{R}|^{1/2}} \left[1 + \frac{1}{\nu}(\mathbf{x} - \boldsymbol{\mu})'\mathbf{R}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \right]^{(\nu+p)/2}.$$

The predictive distribution of $\hat{y}_x$ follows from the following two lemmas.

Lemma 3 The predictive distribution of a complete observation $\hat{\mathbf{Y}}$ is a multivariate t distribution with $b - K + 1$ degrees of freedom, mean vector $\boldsymbol{\theta}$ and correlation matrix $(q+1)\mathbf{B}/q(b - K + 1)$.

Lemma 4 The predictive distribution of $\hat{y}_x$ is

$$\hat{y}_x \sim t\left(b-K+1, \theta_x, \frac{q(b-K+1)}{(q+1)B_{xx}}\right), \tag{13}$$

which is a univariate Student's t-distribution with $b-K+1$ degrees of freedom, mean θ_x and variance $(q+1)B_{xx}/q(b-K-1)$.

Using the predictive distribution found in (13), we can derive another expression for the updating equation of $\boldsymbol{\theta}$ in (12). Define

$$T = \frac{\hat{y}_x - \theta_x}{\left(\frac{q+1}{q(b-K+1)}B_{xx}\right)^{1/2}},$$

$$\tilde{s}(\mathbf{B},x) = \frac{\left(\frac{q+1}{q(b-K+1)}\right)^{1/2}}{\left(\frac{q(b+\Delta b)}{b+\Delta b-K+1}+1\right)B_{xx}^{1/2}}\mathbf{B}e_x.$$

Then, (12) can be rewritten as

$$\boldsymbol{\theta}^* = \boldsymbol{\theta} + \tilde{s}(\mathbf{B},x)T, \tag{14}$$

where T has a Student's t-distribution with $b-K+1$ degrees of freedom, mean 0 and scale parameter 1.

3   THE VALUE OF INFORMATION

Value of information procedures allocate the simulation budget by evaluating the potential of new observations to improve the current estimate of the best value (see Chick 2006 for a survey). The information potential is defined in terms of the expected difference in the estimated objective value before and after the next observation occurs. We do not know exactly how an observation of alternative x will change our beliefs about the best alternative, but we can compute an expectation over the predictive distribution in (14). In this way, we can "look ahead" to the random outcome of the next observation, attempting to anticipate the results before we see them. If we sample from alternative x in the nth stage and collect observation $\hat{y}_x^{n+1}$, the value of information is defined as

$$\mathcal{V}(x) = \mathbb{E}^n\left[\max_i \theta_i^{n+1} \mid x^n = x\right] - \max_i \theta_i^n,$$

where $\mathbb{E}^n$ is the conditional expectation taken with respect to $\mathcal{F}^n$, the σ- algebra generated by $x^0, \hat{y}^1, x^1, \hat{y}^2, \ldots, x^{n-1}, \hat{y}^n$, i.e., all decisions and observations before stage n.

Using (14), the expected value of information can be rewritten as

$$\mathcal{V}(\boldsymbol{\theta}^n, \tilde{s}(\mathbf{B}^n,x), m) = \mathbb{E}^n\left[\max_i \theta_i^n + \tilde{s}(\mathbf{B}^n,x^n)T_m \mid x^n = x\right] - \max_i \theta_i^n, \tag{15}$$

where $\mathcal{V}$ is defined by $\mathcal{V}(\mathbf{a},\mathbf{b},m) = \mathbb{E}[\max_i a_i + b_i T_m] - \max_i a_i$, $\mathbf{a}$ and $\mathbf{b}$ are $K \times 1$ vectors. The random variable T_m follows a univariate Student's t-distribution with degrees of freedom m, mean 0 and variance 1.

We introduce a fully sequential policy called the *Projected Learning of Unknown Correlations with Knowledge gradients* (PLUCK). The PLUCK policy chooses its measurement decision by

$$X^{PLUCK,n} \in \underset{x}{\operatorname{argmax}} \mathcal{V}(\boldsymbol{\theta}^n, \tilde{s}(\mathbf{B}^n,x), m).$$

We now show how the value of information can be computed exactly.

3.1 Computation of the Value of Information

Define a function $h : \mathbb{R} \mapsto \{1, 2, \ldots, K\}$ as

$$h(t) := \max(\operatorname*{argmax}_{i} a_i + b_i t).$$

The function h tells us which alternative will appear to be the best given $T_m = t$. Instead of calculating $\mathscr{V}(\mathbf{a}, \mathbf{b}, m)$ directly, we first rewrite $a_{h(T_m)} + b_{h(T_m)} T_m$ as a telescoping sum,

$$a_{h(0)} + b_{h(0)} T_m + \left[\sum_{i=h(0)}^{h(T_m)-1} (a_{i+1} - a_i) + (b_{i+1} - b_i) T_m \right] + \left[\sum_{i=h(T_m)}^{h(0)-1} (a_i - a_{i+1}) + (b_i - b_{i+1}) T_m \right].$$

Using standard techniques (see e.g., Frazier et al. 2009), we can find a non-decreasing sequence $\{c_i\}_{i=0}^{K}$ such that $h(z) = i$ if and only if $z \in [c_{i-1}, c_i)$. It follows that $\mathscr{V}(\mathbf{a}, \mathbf{b}, m)$ can be written as

$$\mathscr{V}(\mathbf{a}, \mathbf{b}, m) = \sum_{i=1}^{K-1} (b_{i+1} - b_i) \mathbb{E}[(-|c_i| + T_m)^+].$$

To compute this quantity, we need an expression for the tail expectation of a univariate Student's t-distribution. Let $g_m(\cdot)$ and $G_m(\cdot)$ be the pdf and cdf, respectively, of a standard Student's t-distribution with m degrees of freedom. We proceed with the computation by rewriting $\mathbb{E}\left[(T_m - |c_i|)^+\right]$ as

$$\mathbb{E}\left[(T_m - |c_i|)^+\right] = \mathbb{E}(T_m \cdot \mathbf{1}_{\{T_m > |c_i|\}}) - |c_i|(1 - G_m(|c_i|)).$$

It can be shown that

$$\mathbb{E}(T_m \cdot \mathbf{1}_{\{T_m > |c_i|\}}) = \frac{m + c_i^2}{m - 1} g_m(|c_i|).$$

Therefore we can obtain a closed-form expression for the value of information as follows:

$$\mathscr{V}(\mathbf{a}, \mathbf{b}, m) = \sum_{i=1}^{K-1} (b_{i+1} - b_i) \left(\frac{m + c_i^2}{m - 1} g_m(|c_i|) - |c_i|(1 - G_m(|c_i|)) \right). \tag{16}$$

3.2 Monotonicity of the Value of Information

The value of information calculated in (16) depends on the degrees of freedom m of the Student's t-distribution. As the PLUCK policy collects information, m will increase. The following result shows that, all else being equal, information is less valuable when m is larger. Since $m = b - K + 1$, the degrees of freedom are directly related to the total amount of information we have collected thus far, represented by b. In other words, the same information is less valuable when we have already accumulated many other observations.

Theorem 5 For fixed $K \times 1$ vectors $\mathbf{a}$ and $\mathbf{b}$, the value of information $\mathscr{V}(\mathbf{a}, \mathbf{b}, m)$ is a decreasing function in the degrees of freedom parameter m.

We point out an interesting consequence of Theorem 5. If we take $m \to \infty$, the random variable T_m converges weakly to a standard normal distribution. This is precisely what occurs when the correlation structure is known: the value of information in (15) will be calculated using a normal random variable (see Powell and Ryzhov 2012). It follows that, given the same set of beliefs about the means and covariances, information will be more valuable when the correlation structure is unknown than when the latter is known. To put it simply, we learn more from a single observation when we are learning both means and covariances.

Table 1: Final opportunity cost and standard errors for the experiments.

Experiment	Performance Measure	Policies			
		PLUCK	CKG	Greedy	LL
1	Opportunity Cost	0.0851	0.2192	0.1950	0.2657
	Standard Errors	0.0036	0.0050	0.0049	0.0056
2	Opportunity Cost	0.0031	0.1037	0.1434	0.2424
	Standard Errors	0.0026	0.0049	0.0061	0.0060
3	Opportunity Cost	0.0773	0.1430	0.2716	0.2402
	Standard Errors	0.0027	0.0051	0.0053	0.0022
4	Opportunity Cost	0.0917	0.1254	0.1784	0.3501
	Standard Errors	0.0038	0.0040	0.0039	0.0061

4 APPLICATION TO WIND FARM PLACEMENT

Within the context of the wind farm placement problem described earlier, we compare the performance of the PLUCK policy with other learning policies. We use hourly wind speed data (Cosgrove et al. 2003) across the United States with latitude and longitude resolution of 0.125 degrees. The data consist of two components: the zonal component u (in the west-east direction) and the meridional component v (in the north-south direction). Assuming for the purpose of this example that all the wind turbines can be placed in the right direction, we focus on the magnitude of the wind speed, which is defined as $\sqrt{u^2 + v^2}$. The objective is to select the location with the highest wind speed over a set of 64 locations.

4.1 Experiment Settings

We considered data from four regions across the United States: California, Texas, Iowa and Washington. All regions have high wind capacity installed or high percentage of wind power generation in recent years. For each of the four regions, we selected 64 different locations sitting on an 8×8 grid (area of the grid is about 3500 square miles to 4500 square miles) within the region as alternatives for wind farm placement.

We used 1800 days of data to estimate the mean and covariance matrix of a multivariate normal distribution. We then chose these fitted parameters to represent the "true" underlying sampling distribution. In our experiments, individual observations were sampled by simulating from multivariate normal distributions with the "true" parameter values. However, the policies were not allowed to see the "truth" when making decisions. We considered four policies in the comparison: the PLUCK policy, the correlated KG (CKG) policy in Frazier et al. (2009), a Greedy policy and the LL policy with linear loss in Chick and Inoue (2001). We briefly explain the distinctions between these policies below.

Both PLUCK and CKG are designed to sample sequentially, with CKG assuming a known covariance structure. The greedy policy is also sequential, choosing the alternative with the highest value of $\boldsymbol{\theta}^n$ at time step n and using our approximate normal-Wishart model to update the beliefs. By contrast, the LL policy of Chick and Inoue (2001) first screens out a subset of alternatives, then allocates the simulation budget equally among the rest, and measures them in batch. This structure allows a conjugate normal-Wishart prior to be used, with the drawback that the policy often samples alternatives that do not provide a lot of useful information.

4.2 Experimental Results

It is critical to collect information efficiently when the decision-maker's prior beliefs are inaccurate or misleading. To show that the PLUCK policy is particularly effective in such a situation, we used a small number of data points to create a prior for which the location that appeared to be the best was quite different from the true best location. We discuss this issue further below; for now, we note that each policy was given a budget of $N = 100$ measurements to correct this initial error. Table 1 give the final opportunity
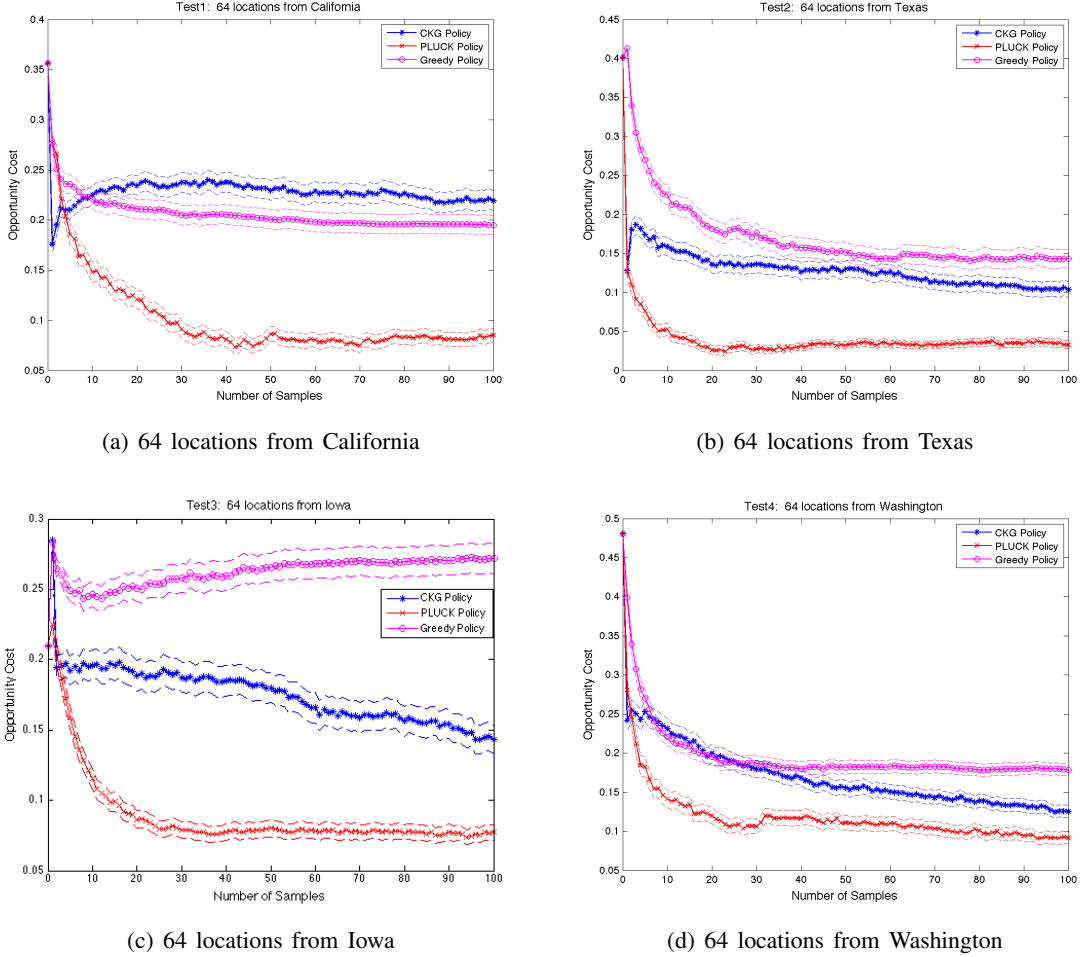
(a) 64 locations from California

(b) 64 locations from Texas

(c) 64 locations from Iowa

(d) 64 locations from Washington

Figure 1: Averaged opportunity cost vs. the number of samples.

cost, which is defined as

$$C^\pi = \max_x \mu_x - \mu_{(\operatorname{argmax}_x \boldsymbol{\theta}_x^N)}$$

after *N* measurements for each policy π, averaged over 1000 sample paths. Lower opportunity cost suggests that a policy selects an alternative closer to the best. The PLUCK policy outperforms all other competing policies by a statistically significant amount, while the correlated KG policy has smaller final opportunity cost than the other two policies. The LL policy performs poorly in all four experiments, possibly because the simulation budget is quite small relative to the number of alternatives.

Figure 1 shows how this performance measure changes as the number of samples *N* varies from 1 to 100. The bands indicate the mean performance measures plus or minus two standard errors. The LL policy is omitted from these pictures because it allocates simulations in batch rather than sequentially, by dividing them uniformly across any alternatives that were not screened out. The opportunity cost for PLUCK tends to decrease over time. For CKG, there is a general tendency toward improvement; however, there is also a sharp degradation in performance for every short time horizons. We conjecture that this behavior arises because CKG assumes a known covariance structure. If the prior beliefs about the correlations are inaccurate, this misdirects the way in which CKG incorporates new information into the posterior. A small amount of information can thus make CKG produce even worse performance than what could be obtained with just the prior.
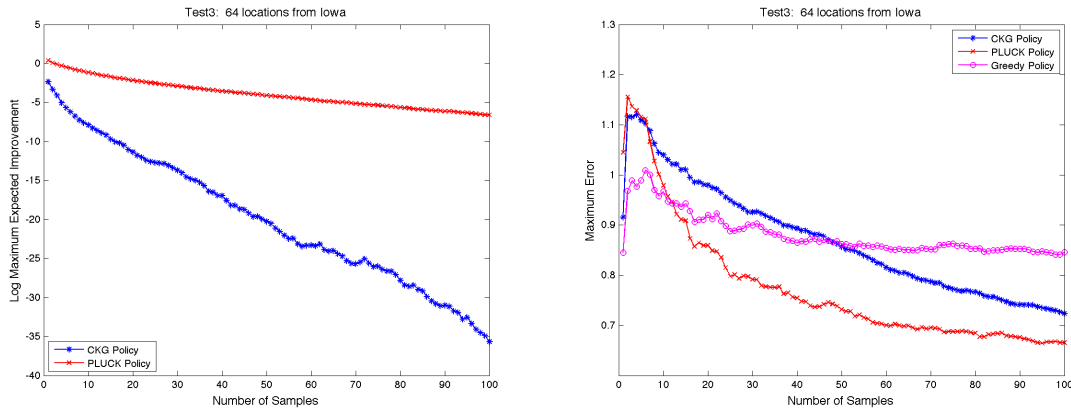
(a) Expected Improvement from Experiment 3   (b) Maximum Error from Experiment 3

Figure 2: Value of information and Maximum Error.

We also considered a different set of experiments in which results were averaged across multiple priors constructed by small-sample bootstrapping from the wind speed data. Overall, we found that PLUCK still outperformed the competition, with the caveat that all policies were more heavily affected by the initial degradation in performance (the early iterations needed to get a handle on the true correlation structure).

We also make two interesting observations from the experimental results. Figure 2(a) shows the logarithm of the value of information as computed by both PLUCK and CKG (for a particular experiment),
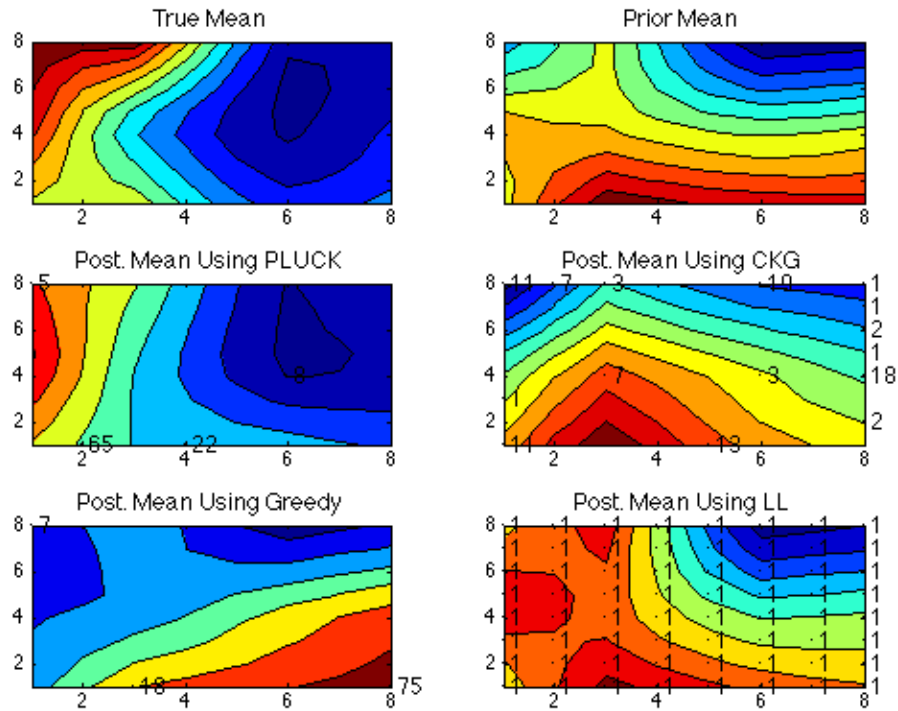

Figure 3: Contour map of different policies.

while Figure 2(b) shows the maximum absolute difference between the posterior and true means for various policies. Figure 2(a) shows that the value of information is much higher when we consider unknown correlations, as suggested by Theorem 5. Figure 2(b) shows that the PLUCK policy does a better overall job of estimating the true values.

Figure 3 gives contour maps of the true means, prior means and posterior means after 100 measurements with four different policies. The number of times that each alternative is measured is also shown on the contour maps (zeros are omitted). Red colors indicate higher values. We can see that the true best alternative is in the upper-left corner, whereas the prior misdirects us toward bottom-center. After 100 measurements, the PLUCK policy captures the general trend of the true values, whereas CKG is still stuck on beliefs that resemble the prior. The LL policy comes closer to the true values, but its batch structure ends up allocating many samples to alternatives that do not provide a lot of useful information.

5 CONCLUSION

We have presented the first computationally tractable statistical learning model for fully sequential ranking and selection with unknown correlation structures. We have also derived a value of information procedure that anticipates new information about both the true values and the true correlations when allocating simulations. Previous work in this area has required known correlation structures, an assumption that is likely to be violated in many applications. We believe that our work offers a useful way to tackle problems with difficult correlation structures, and opens up new problem classes for Bayesian optimal learning.

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.

Chen, C.-H., D. He, M. C. Fu, and L. H. Lee. 2008. "Efficient simulation budget allocation for selecting an optimal subset". *INFORMS Journal on Computing* 20 (4): 579–595.

Chen, C.-H., and L. H. Lee. 2010. *Stochastic Simulation Optimization: An Optimal Computing Budget Allocation*. World Scientific.

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 K. Inoue. 2001. "New Procedures to Select the Best Simulated System Using Common Random Numbers". *Management Science* 47 (8): 1133–1149.

Cosgrove, B. A., D. Lohmann, K. E. Mitchell, P. R. Houser, E. F. Wood, J. C. Schaake, A. Robock, C. Marshall, J. Sheffield, Q. Duan, L. Luo, R. W. Higgins, R. T. Pinker, J. D. Tarpley, and J. Meng. 2003. "Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project". *Journal of Geophysical Research* 108 (D22): 8842–8853.

DeGroot, M. H. 2004. *Optimal Statistical Decisions*. WCL ed. New York: Wiley-Interscience.

Dominici, F., G. Parmigiani, and M. Clyde. 2000. "Conjugate analysis of multivariate normal data with incomplete observations". *Canadian Journal of Statistics* 28 (3): 533–550.

Frazier, P. I., W. B. Powell, and S. Dayanik. 2009. "The knowledge-gradient policy for correlated normal rewards". *INFORMS Journal on Computing* 21 (4): 599–613.

Fu, M. C., J.-Q. Hu, C.-H. Chen, and X. Xiong. 2007. "Simulation Allocation for Determining the Best Design in the Presence of Correlated Sampling". *INFORMS Journal on Computing* 19 (1): 101–111.

Golub, G. H., and C. F. Van Loan. 1996. *Matrix Computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press.

Gupta, A. K., and D. K. Nagar. 2000. *Matrix variate distributions*. Chapman & Hall.

Gupta, S. S., and K. J. 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.

Kadane, J., and R. Trader. 1988. "A Bayesian treatment of multivariate normal data with observations missing at random". *Statistical Decision Theory and Related Topics* 4 (1): 225–234.

Kim, S.-H. forthcoming 2013. "Statistical Ranking and Selection". In *Encyclopedia of Operations Research and Management Science*, edited by S. I. Gass and M. C. Fu. Springer.

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. Henderson and B. Nelson, 501–534. North-Holland Publishing, Amsterdam.

Kotz, S., and S. Nadarajah. 2004. *Multivariate t Distributions and Their Applications*. New York: Cambridge University Press.

Nelson, B., and F. Matejcik. 1995. "Using Common Random Numbers for Indifference-Zone Selection and Multiple Comparisons in Simulation". *Management Science* 41 (12): 1935–1945.

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

Qu, H., I. O. Ryzhov, and M. C. Fu. 2012. "Simulation selection with unknown correlation structures". *Working paper, University of Maryland*.

Scott, W. R., W. B. Powell, and H. P. Simão. 2010, December. "Calibrating simulation models using the knowledge gradient with continuous parameters". In *Proceedings of the 2010 Winter Simulation Conference*, edited by B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan, 1099–1109. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Triantafyllopoulos, K. 2008. "Missing observation analysis for matrix-variate time series data". *Statistics and Probability Letters* 78 (16): 2647–2653.

Yang, W., and B. Nelson. 1991. "Using Common Random Numbers and Control Variates in Multiple-Comparison Procedures". *Operations Research* 39 (4): 583–591.

AUTHOR BIOGRAPHIES

HUASHUAI QU is a Ph.D. candidate in Applied Mathematics, Statistics, and Scientific Computation at the University of Maryland. His interests lie in the broad areas of optimal learning and simulation optimization. His email address is huashuai@math.umd.edu.

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.

MICHAEL C. FU is Ralph J. Tyser Professor of Management Science in the Robert H. Smith School of Business, with a joint appointment in the Institute for Systems Research and affiliate faculty appointment in the Department of Electrical and Computer Engineering, all at the University of Maryland. His research interests include simulation optimization and applied probability, with applications in supply chain management and financial engineering. He has published four books: *Conditional Monte Carlo: Gradient Estimation and Optimization Applications* (1998 INFORMS Simulation Society Outstanding Publication Award); *Simulation-based Algorithms for Markov Decision Processes*; *Perspectives in Operations Research*; and *Advances in Mathematical Finance*. His email address is mfu@umd.edu.