Efficient Fully Sequential Indifference-Zone Procedures Using Properties of Multidimensional Brownian Motion Exiting a Sphere

A.B. Dieker
Columbia University
New York, NY 10027

Seong-Hee Kim
Georgia Institute of Technology
Atlanta, GA 30332-0205

July 31, 2016

Abstract

We consider a ranking and selection (R&S) problem whose goal is to select a system with the largest or smallest expected performance measure among a number of simulated systems with a pre-specified probability of correct selection. Fully sequential procedures take one observation from each survived system and eliminate inferior systems when there is clear statistical evidence that they are inferior. Most fully sequential procedures make elimination decisions based on sample performances of each possible pair of survived systems and exploit the bound crossing properties of a univariate Brownian motion. In this paper, we present new fully sequential procedures whose elimination decisions are based on sample performances of all competing systems. The analysis of the proposed procedures is based on the properties of a multidimensional Brownian motion exiting a sphere. We show that the new procedures significantly outperform a widely used fully sequential procedure while they show similar performances compared to BIZ, a recent fully-sequential procedure that uses a Bayesian approach.

Subject classification: Simulation, Ranking and Selection, Fully Sequential, Multidimensional Brownian Motion, Sphere

1. Introduction

Ranking and selection (R&S) is one of the classical and well-studied problems in the operations research literature. It aims to find the best system among a number of systems for which noisy performance information is accessible through simulation. In this paper, we assume that the best system is one with the largest or smallest expected performance, which is known as the finding-the-best problem. There are at least three approaches for the finding-the-best problem: the indifference-zone (IZ) approach, the Bayesian approach, and the optimal computing budget allocation (OCBA) approach. Kim and Nelson (2011) provide a brief review of each approach. For more information on the Bayesian approach, see Chick (2006). When there is

a fixed computing budget until a decision is made, the OCBA approach provides an efficient way to find the best system, see for example Chen and Lee (2010). This paper studies a indifference-zone (IZ) procedure, where the decision maker specifies a difference worth detecting called the IZ parameter.

Among procedures that take the IZ approach, Rinott (1978) is one of the most classical procedures. It is a two-stage procedure where a stage occurs whenever a simulation of a system resumes obtaining additional observations. Nelson et al. (2001) also propose a two-stage procedure. Rinott's procedure does not have any elimination step while procedures due to Nelson et al. (2001) eliminate systems after the first stage if there is statistical evidence that they are inferior. Therefore the latter procedure is more efficient than Rinott's procedure in terms of the number of observations needed until a decision is made. On the other hand, fully-sequential IZ procedures take one observation from competing systems and eliminate inferior systems as additional observations become available. They carry the risk of incorrectly eliminating the best system due to stochastic noise in the performance measurements. Examples of fully-sequential IZ procedures are the KN procedures from Kim and Nelson (2001), which are widely used as they are available in leading commercial simulation software. KN's parameters are chosen to control the probability of eliminating the best system. Since this probability is intractable, the procedures instead rely on a Bonferroni-type lower bound on the worst-case probability of incorrect selection, which corresponds to the best system having a mean performance that exceeds the means of the other systems by exactly the IZ parameter; this setup is known as the slippage configuration (SC). Particularly when the number of systems is large, this lower bound tends to be a poor approximation for the worst-case probability of correct selection. As discussed in Wang and Kim (2011), the result is that KN procedures tend to take many more observations than necessary to control the probability of incorrect selection, and are thus inefficient in that sense.

In seeking to circumvent the inefficiencies caused by the conservativeness of the Bonferroni bound, it has been a major challenge to devise elimination rules that are constructed from the observations of all survived systems rather than from pairwise differences between systems as in KN. The primary contribution of this paper is to develop a new family of such procedures. Our procedures build on properties of a multidimensional Brownian motion hitting a sphere rather than a univariate Brownian motion hitting a line as in KN. As far as we know, this is the first work that considers a multidimensional Brownian motion

for the derivation of R&S procedures. Naturally, our procedures make elimination decisions considering observations from all survived systems rather than pairs of systems. Frazier (2014) proposes a procedure whose elimination decisions are also based on observations from all survived systems, but the statistics in his procedure, the Bayes-inspired indifference zone procedures (BIZ), are based on Bayes probabilities. The BIZ procedures are shown to have a tight lower bound on the worst-case probability of correct selection under the slippage configuration. Experimental results show that our procedures significantly outperform KN. On the other hand, our procedures perform slightly better than BIZ under difficult scenarios (such as SC or increasing variances) but slightly worse under easier scenarios.

Preliminary work related to this work is published in the Winter Simulation Conference proceedings which include Kim and Dieker (2011) and Dieker and Kim (2012, 2014). The first two papers consider only three systems with known variances. Dieker and Kim (2014) give a procedure for a general number of systems but require known and equal variances. Moreover, the spheres that play a crucial role in the procedure all have the same radius and the procedure performs worse than KN when the means of the systems are spread out evenly. In the procedures presented in the present paper, the radii of a sphere vary as the number of survived systems decreases, outperforming KN in all scenarios; and a version of our procedure can handle unknown and unequal variances.

The paper is organized as follows. Section 2 defines our problem and introduces notation. Section 3 presents the statistics that we use for elimination decisions and explains the properties of our statistics. Section 4 proposes new fully-sequential procedures. In Section 5, we provide justifications for our procedures and approximations in order to set the parameter values of the procedures. Experimental results are presented in Section 6, followed by conclusions in Section 7.

2. Problem and Notation

This section introduces our notation and assumptions and defines the problem. We assume there are k systems ($k \geq 2$). Let X_{ij} represent the jth observation from replication (or batch) j of system i for $i = 1, \ldots, k$ and $j = 1, 2, \ldots$. Then the mean and variance of the outputs from system i are defined as $\mu_i = \mathrm{E}[X_{ij}]$ and $\sigma_i^2 = \mathrm{Var}[X_{ij}]$, respectively. We want to find the system with the largest mean μ_i.

Throughout the paper, we assume that the following assumptions hold:

Assumption 1.

$$X_{ij} \overset{iid}{\sim} \boldsymbol{N}(\mu_i, \sigma_i^2), \quad j = 1, 2, \ldots,$$

where $\overset{iid}{\sim}$ *represents 'are independent and identically distributed as' and* $\boldsymbol{N}(\mu_i, \sigma_i^2)$ *denotes the normal distribution with mean* μ_i *and variance* σ_i^2*. Moreover,* $(X_{1j}, \ldots, X_{kj})$ *and* $(X_{1j'}, \ldots, X_{kj'})$ *are independent for any* $j \neq j'$*.*

Assumption 2. $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_{k-1} \leq \mu_k - \delta$ *for* $\delta \in \mathbb{R}^+$*.*

Assumption 1 implies that observations from each system are marginally iid normally distributed and systems are simulated independently (note that this rules out common random numbers). Without loss of generality, Assumption 2 assumes that system k is the best and its mean is at least δ better than any alternative system. The parameter δ is a user-specified parameter known as the IZ parameter.

We aim to devise a method that observes systems sequentially and eliminates clearly inferior systems from further consideration. The method stops once only one system remains, and this system is declared as the best system.

Additional notation is needed for later sections:

$$\begin{aligned}
n &\equiv \text{the current number of observations or the current stage number;} \\
I &\equiv \text{set of competing systems at the } n\text{th stage;} \\
\bar{X}_i(n) &\equiv \frac{1}{n}\sum_{j=1}^{n} X_{ij}, \text{ the sample mean of system } i \text{ based on the first } n \text{ observations;} \\
\boldsymbol{X}_I(n) &\equiv |I| \times 1 \text{ vector of } \textstyle\sum_{j=1}^{n} X_{ij} \text{ for } i \in I; \\
\hat{\sigma}_i^2(n) &\equiv \text{sample variance of system } i \text{ from } X_{i1}, \ldots, X_{in} \text{ which is } \frac{1}{n-1}\sum_{j=1}^{n}(X_{ij} - \bar{X}_i(n))^2; \\
A^T &\equiv \text{the transpose of a matrix } A; \\
\delta_{|I|}^2 &\equiv \delta^2 \frac{|I|-1}{|I|}.
\end{aligned}$$

We first present statistics that our procedures use when making an elimination decision.

3. Statistics for Screening

The canonical choice for fully sequential procedures is to use $\sum_{j=1}^{n}(X_{ij} - X_{\ell j})$ for every $i \neq \ell$ as observed statistics and to eliminate a system whenever the statistics exit a so-called continuation region defined by two parallel lines such as $(-a, a)$ for a constant $a > 0$ or a function $h(n) > 0$ such as $(-h(n), h(n))$. Kim and Nelson (2014) use a triangular shaped continuation region defined by a decreasing linear function $h(n)$. Note that traditional continuation regions are defined in the two-dimensional space. Our procedures use different statistics that take a quadratic form and our continuation region is a sphere defined in a higher dimension.

Consider $x \in \mathbb{R}^s$ and $I \subset \{1, \ldots, k\}$ with $I = \{i_1, \ldots, i_s\}$. Furthermore let Γ represent the covariance matrix of $(X_{i_1 j}, X_{i_2 j}, \ldots, X_{i_s j})^T$,

$$\Gamma = \begin{bmatrix} \sigma^2_{i_1} & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2_{i_2} & 0 & \cdots & 0 \\ 0 & 0 & \sigma^2_{i_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \sigma^2_{i_s} \end{bmatrix}$$

and let V represent a $s-1$ by s matrix given by

$$V = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & -1 \\ 0 & 1 & 0 & \cdots & 0 & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & 1 & -1 \end{bmatrix}.$$

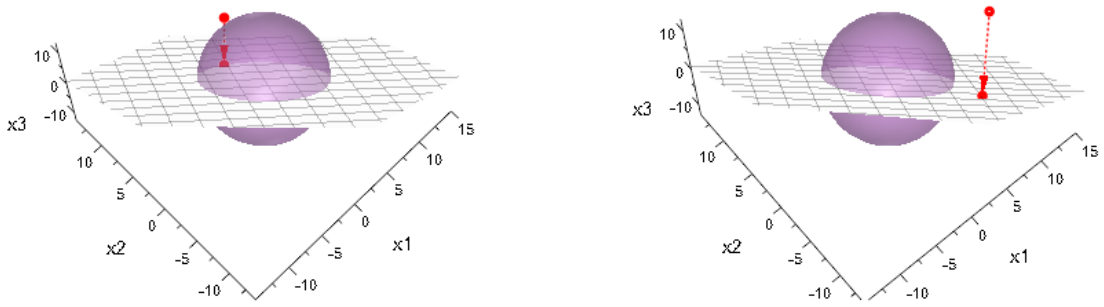
Then our statistic $\mathcal{S}_I(x)$ is defined as

$$\mathcal{S}_I(x) \equiv (Vx)^T (V\Gamma V^T)^{-1}(Vx) = \begin{bmatrix} x_{i_1} - x_{i_s} \\ \vdots \\ x_{i_{s-1}} - x_{i_s} \end{bmatrix}^T (V\Gamma V^T)^{-1} \begin{bmatrix} x_{i_1} - x_{i_s} \\ \vdots \\ x_{i_{s-1}} - x_{i_s} \end{bmatrix}$$

and our continuation region is related to this quadratic form. From the definition of $\mathcal{S}_I$ it may seem that $\mathcal{S}_I$ is complicated to calculate and that it depends on the order in which its elements are listed. The following lemma is useful in deriving a simpler form of $\mathcal{S}_I(x)$ and a corollary followed by the lemma shows that $\mathcal{S}_I(x)$ only depends on the set I, so not on the order of the elements in I. The proofs are given in the appendix.

Lemma 1. *Suppose $x \in \mathbb{R}^s$ and $I \subset \{1, \ldots, k\}$ with $I = \{i_1, \ldots, i_s\}$. If $\Pi = \Gamma V^T (V\Gamma V^T)^{-1} V$, then*

$$\mathcal{S}_I(x) = \mathcal{S}_I(\Pi x).$$



(a) Projection inside ball: no elimination

(b) Projection outside ball: elimination

Figure 1: Projected points on plane $y_1 + y_2 + y_3 = 0$ and elimination rules. A ball (with radius 7 here) is also visible.

The above lemma holds for Γ regardless of whether it has equal diagonal elements. The matrix Π is an oblique projection matrix with range $R = \{y \in \mathbb{R}^s : \sum_{i\in I} y_i = 0\}$ and null space $N = \{\alpha(1/\sigma_{i_1}^2, \cdots, 1/\sigma_{i_s}^2) : \alpha \in \mathbb{R}\}$. It becomes an orthogonal projection matrix when $\sigma_i^2 = \sigma^2$ for all $i \in I$. This lemma implies that the value of our statistic at any x equals the value of our statistic at the projected point on the plane determined by $\sum_{i\in I} y_i/\sigma_i^2 = 0$ (or $\sum_{i\in I} y_i = 0$ for equal variances). In Section 5 this lemma is used to make the elimination decision depend on the IZ parameter δ only and not on the unknown mean parameter. Using the above lemma, we next derive a simpler form of $\mathcal{S}_I(x)$ when the variances are equal.

Corollary 1. *Suppose $x \in \mathbb{R}^s$ and $I \subset \{1, \ldots, k\}$ with $I = \{i_1, \ldots, i_s\}$. If $\sigma_i^2 = \sigma^2$, then*

$$\mathcal{S}_I(x) = \frac{1}{\sigma^2}\frac{1}{|I|}\sum_{\substack{i<\ell \\ i,\ell\in I}}(x_i - x_\ell)^2 = \frac{1}{\sigma^2}\sum_{i\in I}(x_i - \bar{x})^2$$

where $\bar{x} = \frac{1}{s}\sum_{i\in I} x_i$.

This corollary shows that the order of elements listed in I does not matter when calculating $\mathcal{S}_I(x)$.

Our elimination decision rule takes a form of $\mathcal{S}_I(x) \geq r^2$ for $r \in \mathbb{R}^+$. Let x' denote the projection of x on the plane with $\sum_{i=1}^s y_i = 0$. From Lemma 1, we know that $\mathcal{S}_I(x)$ is equal to $\mathcal{S}_I(x')$. As x' is on the plane $\sum_{i=1}^s y_i = 0$, we know that $\bar{x'} = 0$. From the second equality of $\mathcal{S}_I(x)$ in Corollary 1, it is easy to see that $\mathcal{S}_I(x')$ becomes simply the squared distance between x' and the origin. Therefore our elimination decision rule implies that no elimination occurs and sampling continues when the projected point x' is inside a sphere

as in Figure 1(a); but one system (usually with the smallest value) is eliminated when the projected point x' is outside the sphere as in Figure 1(b).

Through our elimination region $\{x \in \mathbb{R}^k : \mathcal{S}_I(x_I) \geq r^2\}$, where x_I denotes the vector $(x_i)_{i\in I}$, we *simultaneously* and *automatically* verify elimination for all $2^{|I|} - 1$ nonempty subsets $J \subseteq I$. Indeed, the following lemma implies that $\{x \in \mathbb{R}^k : \mathcal{S}_J(x_J) \geq r^2\}$ is a subset of $\{x \in \mathbb{R}^k : \mathcal{S}_I(x_I) \geq r^2\}$ for every $J \subseteq I$.

Lemma 2. *Suppose $J \subseteq I \subseteq \{1, \ldots, k\}$. Then $\mathcal{S}_J(x_J) \leq \mathcal{S}_I(x_I)$ for all $x \in \mathbb{R}^k$.*

4. $\mathcal{DK}$ Procedures

In this section, we provide the descriptions of our new procedures. We first present $\mathcal{DK}_1$ for known and equal variances and extend it to unknown but equal variances, resulting in $\mathcal{DK}_2$. Then $\mathcal{DK}_3$ is presented for unknown and unequal variances.

4.1 Equal and Known Variances

The $\mathcal{DK}_1$ procedure for equal and known variances is as follows:

The $\mathcal{DK}_1$ Procedure

Setup: Select the nominal level $1-\alpha$ and the IZ parameter δ. Set $I = \{1, 2, \ldots, k\}$ and choose $\eta_{|I|}$ (which will be discussed in Section 5). Take one observation from each system. Set $n = 1$ and go to **Calculation**.

Calculation: Calculate $\mathcal{S}_I(\boldsymbol{X}_I(n))$.

Screening: If $\mathcal{S}_I(\boldsymbol{X}_I(n)) \geq \left(\frac{\sigma\cdot\eta_{|I|}}{\delta_{|I|}}\right)^2$, then eliminate the system with the smallest $\bar{X}_i(n)$ among $i \in I$. Update I by removing the eliminated system and go back to **Calculation**. Otherwise, go to **Stopping Rule**.

Stopping Rule: If $|I| = 1$, stop and declare the surviving system as the best. Otherwise, take one more observation for all $i \in I$, set $n = n + 1$, and go to **Calculation**.

Lemma 1 and the screening rule in $\mathcal{DK}_1$ imply that we have an infinite cylinder as our continuation region with a radius depending on $\eta_{|I|}$ and $\delta_{|I|}$. When a projected point x on the plane $\sum_{i\in I} y_i = 0$ is located outside the largest sphere centered at the origin inside the cylinder, elimination occurs and the screening rule is checked again with updated parameters (without obtaining additional observations). We only obtain new observations (i.e., move to the next stage) if no more elimination occurs for a given number of observations.

Lemma 2 implies that we automatically check the screening condition for all sets $J \subseteq I$ if we check it for the largest set I.

4.2 Unknown but Equal Variances

We present a straightforward variant of $\mathcal{DK}_1$ for unknown but equal variances, σ^2. As the variance parameter σ^2 is unknown, it needs to be estimated. Let $\hat{\sigma}_i^2(n)$ represent sample variance of system i. The pooled variance estimator $\hat{\sigma}_p^2(n)$ is defined as follows:

$$\hat{\sigma}_p^2(n) = \frac{1}{|I|} \sum_{i \in I} \hat{\sigma}_i^2(n).$$

Then our statistic is modified to

$$\mathcal{S}'_I(x) = \frac{1}{\hat{\sigma}_p^2(n)} \sum_{i \in I} (x_i - \bar{x})^2$$

and $\hat{\sigma}_i^2(n)$ and $\hat{\sigma}_p^2(n)$ need to be updated in the [Stopping Rule] step after additional observations are obtained. Then the $\mathcal{DK}_2$ procedure is defined below.

The $\mathcal{DK}_2$ Procedure

Setup: Select the nominal level $1-\alpha$ and the IZ parameter δ. Set $I = \{1, 2, \ldots, k\}$ and choose $\eta_{|I|}$. Take $n_0 \geq 2$ observations from each system and calculate $\hat{\sigma}_i^2(n_0)$ and $\hat{\sigma}_p^2(n_0)$. Set $n = n_0$ and go to **Calculation**.

Calculation: Calculate $\mathcal{S}'_I(\boldsymbol{X}_I(n))$.

Screening: If $\mathcal{S}'_I(\boldsymbol{X}_I(n)) \geq \left(\frac{\hat{\sigma}_p(n) \cdot \eta_{|I|}}{\delta_{|I|}} \right)^2$, then eliminate the system with the smallest $\bar{X}_i(n)$ among $i \in I$. Update I by removing the eliminated system and go back to **Calculation**. Otherwise, go to **Stopping Rule**.

Stopping Rule: If $|I| = 1$, stop and declare the surviving system as the best. Otherwise, take one more observation for all $i \in I$; set $n = n + 1$; and update $\hat{\sigma}_i^2(n)$ for $i \in I$ and $\hat{\sigma}_p^2(n)$. Then go to **Calculation**.

4.3 Unknown and Unequal Variances

This subsection extends the $\mathcal{DK}_1$ procedure to handle unknown and unequal variances, resulting in $\mathcal{DK}_3$. The main idea is to make the sampling frequency of each system proportional to the variance parameter of the system, which eventually leads to equal variances. This approach is similar to the one in Frazier (2014).

Let n_i denote the number of observations system i have received so far. In $\mathcal{DK}_1$, $n_i = n$ for any system $i \in I$ but in $\mathcal{DK}_3$, $n_i \leq n$. Also let $W_i(n) = \sum_{j=1}^{n_i} X_{ij}/n_i$ and $\boldsymbol{W}_I(n)$ represent a $|I| \times 1$ vector of $W_i(n)$ for $i \in I$.

Then

$$\mathcal{S}_I''(x) = \frac{1}{\hat{\lambda}^2} \sum_{i \in I} (x_i - \bar{x})^2$$

where

$$\hat{\lambda}^2 = \frac{\sum_{i \in I} \hat{\sigma}_i^2(n_i)}{\sum_{i \in I} n_i}.$$

We can now describe Procedure $\mathcal{DK}_3$.

The $\mathcal{DK}_3$ Procedure

Setup: Select the nominal level $1-\alpha$ and the IZ parameter δ. Also select a constant B_z. Set $I = \{1, 2, \ldots, k\}$ and choose $\eta_{|I|}$. Take n_0 observations from each system and calculate $W_i(n_0)$, $\hat{\sigma}_i^2(n_0)$ and $\hat{\lambda}^2$. Set $n = n_0$ and $n_i = n_0$ for $i \in I$, and go to **Calculation**.

Calculation: Calculate $\mathcal{S}_I''(\boldsymbol{W}_I(n))$.

Screening: If $\mathcal{S}_I''(\boldsymbol{W}_I(n)) \geq \left(\frac{\hat{\lambda} \cdot \eta_{|I|}}{\delta_{|I|}}\right)^2$, then eliminate the system with the smallest $\bar{X}_i(n)$ among $i \in I$. Update I by removing the eliminated system and go back to **Calculation**. Otherwise, go to **Stopping Rule**.

Stopping Rule: If $|I| = 1$, stop and declare the surviving system as the best. Otherwise, let $z = \arg\min \frac{n_i}{\hat{\sigma}_i^2(n_i)}$ for $i \in I$.
For each $i \in I$,

- calculate

$$\Delta_i = \left\lceil \hat{\sigma}_i^2(n_i) \cdot \frac{n_z + B_z}{\hat{\sigma}_z^2(n_z)} \right\rceil;$$

- if $\Delta_i > n_i$, then take $(\Delta_i - n_i)$ observations.

Set $n = n + 1$ and $n_i = \max(n_i, \Delta_i)$; and update $\hat{\sigma}_i^2(n_i)$ for all $i \in I$ and $\hat{\lambda}^2$. Then go to **Calculation**.

Remark: Frazier(2014) recommends $B_z = 1$.

The parameter $\eta_{|I|}$ needs to be chosen carefully so that the actual probability of correct selection is at least $1-\alpha$. In the next section, we derive some analytical results for the $\mathcal{DK}_1$ procedure and then discuss how to choose $\eta_{|I|}$.

5. Proofs and Approximations

This section presents an approximation for the probability of incorrect selection under $\mathcal{DK}_1$, which assumes known and equal variances. We use these approximations in lieu of possibly conservative bounds in order to choose the parameters $\eta_2, \ldots, \eta_k$ of $\mathcal{DK}_1$, thus bypassing a main source of inefficiencies. In the course of the presentation, we explain how we choose the parameters $\eta_2, \ldots, \eta_k$ of our procedure.

The event of incorrect selection can be partitioned according to when the best system is eliminated. If the best system is eliminated first, then we say that the level of elimination is 1. Similarly, if the second system to be eliminated is the best system, then we say that the level of elimination is 2. Thus, the possible levels of incorrect elimination are $1, \ldots, k-1$. The key building block for our approximation scheme is an approximation for the probability of incorrect selection at the first elimination level, which we discuss in Section 5.1. Other levels of incorrect elimination are studied in Section 5.2. With this, we devise a procedure for choosing the parameter $\eta_{|I|}$ for $\mathcal{DK}_1$. We then explain how $\eta_2, \ldots, \eta_k$ for $\mathcal{DK}_1$ are related to parameters for $\mathcal{DK}_2$ and $\mathcal{DK}_3$ in Section 5.3.

In the continuous analog of our problem, the discrete observation window is replaced with a continuous one. The analog of the random walk $\boldsymbol{X}_k(n)$ is $\sigma B(t)$, where $B(t)$ is a standard Brownian motion in $\mathbb{R}^k$ with drift $(\mu, \ldots, \mu, \mu+\delta) \times 1/\sigma$. Throughout this section, we study this continuous problem as a proxy for the discrete problem.

Lemma 3. *For fixed $\eta_k, \ldots, \eta_2$ and $\ell \in \{2, \ldots, k\}$, the probability of eliminination at level ℓ in $\mathcal{DK}_1$ is constant as a function of δ and σ. In particular, the probability of incorrect selection in $\mathcal{DK}_1$ does not depend on δ or σ.*

Proof. Suppose $B(\cdot)$ is a standard Brownian motion in $\mathbb{R}^N$ with $B(0) = (\sigma/\delta)x$ for some $x \in \mathbb{R}^N$ and suppose that $v \in \mathbb{R}^N$. Given an N-dimensional set S, we set

$$\tau_S = \inf\left\{t \geq 0 : \frac{\sigma}{\delta}x + B(t) - \frac{\delta}{\sigma}vt \in \frac{\sigma}{\delta}S\right\}.$$

We then have

$$\begin{aligned}
\mathbb{P}(\tau_S < \infty) &= \mathbb{P}\left(\exists t \geq 0 : \frac{\sigma}{\delta}x + B(t) - \frac{\delta}{\sigma}vt \in \frac{\sigma}{\delta}S\right) \\
&= \mathbb{P}\left(\exists t \geq 0 : \frac{\sigma}{\delta}x + B\left(\frac{\sigma^2}{\delta^2}t\right) - \frac{\sigma}{\delta}vt \in \frac{\sigma}{\delta}S\right) \\
&= \mathbb{P}\left(\exists t \geq 0 : \frac{\sigma}{\delta}x + \frac{\sigma}{\delta}B(t) - \frac{\sigma}{\delta}vt \in \frac{\sigma}{\delta}S\right) \\
&= \mathbb{P}(\exists t \geq 0 : x + B(t) - vt \in S),
\end{aligned}$$

where the first equality follows from rescaling time and the second from the Brownian scaling property. This argument extends to the hitting location, i.e., $\mathbb{P}(\tau_S < \infty, \frac{\sigma}{\delta}x + B(\tau_S) - \frac{\delta}{\sigma}v\tau_S \in \frac{\sigma}{\delta}dy)$. In particular, the hitting location scales with σ/δ.

Elimination at level ℓ amounts to successively hitting appropriate regions of sets S of the form $S = \{x : \sigma^2 \mathcal{S}_I(x_I) \geq \eta^2 k/(k-1)\}$, a set that is independent of σ in view of Corollary 1. The successive hitting locations scale with σ/δ. By the strong Markov property and the above calculation, this means that the elimination probability does not depend on δ or σ. □

5.1 Immediate (Level 1) Elimination of the Best System

Our approximation for the probability of eliminating system k first is based on an asymptotic analysis as the number of systems k goes to infinity. Our results use the commonly employed idea of (i) considering the slippage configuration (SC) where $\mu_1 = \cdots = \mu_{k-1} = \mu_k - \delta = \mu$ and (ii) replacing the (discrete) Gaussian observation sequence with a (continuous) Brownian motion.

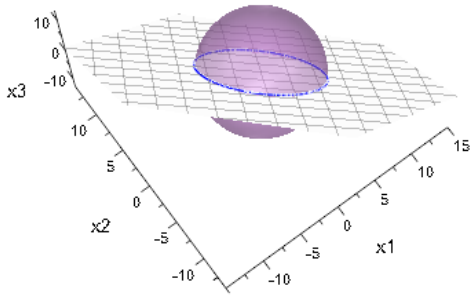
Throughout this section we use the following notation. For a given a vector $x \in \mathbb{R}^k$,

$$\mathrm{E}_k(x) = \frac{1}{k}\sum_{i=1}^{k} x_i, \quad \mathrm{Var}_k(x) = \mathrm{E}_k(x^2) - \mathrm{E}_k(x)^2,$$

where x^2 should be understood component-wise.

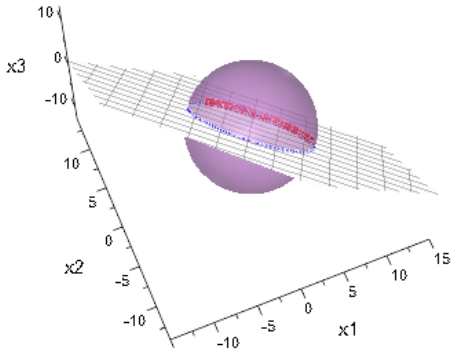
From Lemma 1 we consider $x - \bar{x}$ which corresponds to $B(t) - \mathrm{E}_k(B(t))$. Note that $B(t) - \mathrm{E}_k(B(t))$ is a standard Brownian motion with drift $(-1/k, \ldots, -1/k, (1-1/k)) \times \delta/\sigma$, which is free of the unknown mean parameter μ. Also it takes values in the hyperplane

$$H = \left\{x \in \mathbb{R}^k : \sum_{i=1}^{k} x_i = 0\right\}.$$

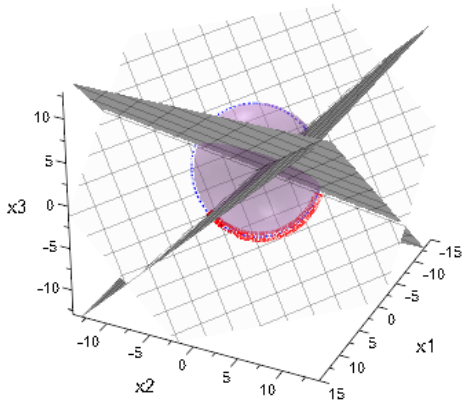


(a) Sphere C (circle here) on hyperplane H

$$C = \left\{x \in \mathbb{R}^k : \sum_{i=1}^{k} x_i = 0, \|x\| = r\right\}$$
$$E_k = \{x \in C : x_k = \min(x_1, \ldots, x_k)\}$$



(b) Region E_k (red) on hyperplane H



(c) Region E_k (red) with planes $x_1 = x_3$ and $x_2 = x_3$ on hyperplane H

Figure 2: Graphical depiction of C and E_k for $k = 3$.

Setting $r = \frac{\sigma\eta_k}{\delta_k}$, we define a k-dimensional sphere in H by

$$C = \left\{x \in \mathbb{R}^k : \sum_{i=1}^{k} x_i = 0, \|x\| = r\right\}.$$

Elimination of the best system can be formulated as $B(t) - \mathrm{E}_k(B(t))$ hitting C in the region

$$E_k = \{x \in C : x_k = \min(x_1, \ldots, x_k)\}.$$

Plane H is shown in Figure 2 when $k = 3$. The blue curve in Figure 2(a) shows C when $k = 3$ and the red curve in Figure 2(b) shows E_k, which is a part of C divided by planes $x_1 = x_3$ and $x_2 = x_3$ as shown in Figure 2(c).

We now state the main result of this section.

Lemma 4. *Let $k \geq 3$. Suppose that $Z_1, \ldots, Z_k$ are iid standard normal. The probability that the process $B(t) - \mathrm{E}_k(B(t))$ first hits C in the part E_k where the best system k gets eliminated equals*

$$\frac{\int_{-r}^{r} e^{\frac{\delta_k}{\sigma} y} d_y \mathbb{P}(Z_k = \min(Z_1, \ldots, Z_k), r(Z_k - \mathrm{E}_k(Z)) \leq y \sqrt{(k-1)\mathrm{Var}_k(Z)})}{\left(\frac{\eta_k}{2}\right)^{-\nu} \Gamma(\nu+1) I_\nu(\eta_k)}, \tag{1}$$

where $\nu = (k-3)/2$, Γ stands for the Gamma function, and I_ν for the modified Bessel function of the first kind.

Proof. Writing ζ for the drift of $B(t) - \mathrm{E}_k(B(t))$, then the hitting place of $B(t) - \mathrm{E}_k(B(t))$ on C has density f with respect to the uniform distribution u_C on C with (e.g., Rogers and Pitman (1981))

$$f(x) = \frac{e^{\langle \zeta, x \rangle}}{\int_C e^{\langle \zeta, w \rangle} u_C(dw)}, \quad x \in C.$$

This distribution is known as the von Mises distribution. The denominator can be written as (Rogers and Pitman (1981))

$$\int_C e^{\langle \zeta, w \rangle} u_C(dw) = \left(\frac{\frac{\delta_k}{\sigma} r}{2}\right)^{-\nu} \Gamma(\nu+1) I_\nu\left(\frac{\delta_k}{\sigma} r\right) = \left(\frac{\eta_k}{2}\right)^{-\nu} \Gamma(\nu+1) I_\nu(\eta_k)$$

because $\|\zeta\| = \delta \sqrt{(k-1)/k}/\sigma = \delta_k/\sigma$ and $(\delta_k/\sigma) r = \eta_k$. Note that larger values of $B_k(t) - \mathrm{E}_k(B(t))$ are more likely than smaller values when the process hits C, which should be expected because system k is the best one.

The probability of eliminating the best system in level 1 equals

$$\begin{aligned} \int_{E_k} f(x) u_C(dx) &= \mathrm{E}[1(X_k = \min(X_1, \ldots, X_k)) f(X)] \\ &= \frac{\int_{-r}^{r} e^{\frac{\delta_k}{\sigma} y} d_y \mathbb{P}(X_k = \min(X_1, \ldots, X_k), \langle \zeta, X \rangle \leq \frac{\delta_k}{\sigma} y)}{\int_C e^{\langle \zeta, w \rangle} u_C(dw)}, \end{aligned}$$

where X has a uniform distribution on C and 1 stands for the indicator function. The random vector

$$X = \frac{r(Z_1 - \mathrm{E}_k(Z), \ldots, Z_k - \mathrm{E}_k(Z))}{\sqrt{k \mathrm{Var}_k(Z)}}$$

has the uniform distribution on C by symmetry. Since $\langle \zeta, x \rangle = \frac{\delta}{\sigma} x_k$ for $x \in H$, the sought probability equals

$$\frac{\int_{-r}^{r} e^{\frac{\delta_k}{\sigma} y} d_y \mathbb{P}(Z_k = \min(Z_1, \ldots, Z_k), \frac{\delta}{\sigma} r(Z_k - \mathrm{E}_k(Z)) / \sqrt{k \mathrm{Var}_k(Z)} \leq \frac{\delta_k}{\sigma} y)}{\int_C e^{\langle \zeta, w \rangle} u_C(dw)},$$

as claimed. □

The preceding lemma yields a Monte Carlo method for calculating the probability of immediate elimination of the best system. Indeed, it states that this probability equals

$$\frac{\mathrm{E}\left[\exp\left(\eta_k \frac{Z_k - \mathrm{E}_k(Z)}{\sqrt{(k-1)\mathrm{Var}_k(Z)}}\right); Z_k = \min(Z_1, \ldots, Z_k)\right]}{\left(\frac{\eta_k}{2}\right)^{-\nu} \Gamma(\nu+1) I_\nu(\eta_k)}, \tag{2}$$

for iid standard normal $Z_1, \ldots, Z_k$. However, for large k, such a Monte Carlo method is not efficient and we instead approximate the level 1 probability (1) by replacing several of its components by asymptotic approximations. For instance, as $k \to \infty$, the random variables $\mathrm{E}_k(Z)$ and $\mathrm{Var}_k(Z)$ converge in distribution to 0 and 1, respectively, by the strong law of large numbers. The rate of convergence is relatively fast (order $1/\sqrt{k}$ by the central limit theorem). We, therefore, approximate those variables by their deterministic asymptotic approximations. The term with the minimum is slightly more complicated. Writing

$$c_k = \sqrt{2\log k} - \frac{\log\log k + \log(4\pi)}{2\sqrt{2\log k}},$$

$\min(Z_1, \ldots, Z_{k-1}) + c_{k-1}$ converges in distribution to 0. For example, see Example 3.3.29 in Embrechts, Kluppelberg and Mikosch (1997). The rate of convergence is relatively slow (order $1/\sqrt{2\log k}$), so we use an approximation based on the fact that

$$\sqrt{2\log k}(\min(Z_1, \ldots, Z_{k-1}) + c_{k-1})$$

converges in distribution to $-G$ where G is a Gumbel distributed random variable which is equal in distribution to $-\log(-\log(U))$ where U is standard uniformly distributed. Even when the central limit theorem is used for the sum instead of the law of large numbers, the minimum and sum are asymptotically independent (e.g., Chow and Teugels 1978). This motivates the approximation, for $y \in (-r, r)$,

$$\begin{aligned} & d\mathbb{P}(Z_k = \min(Z_1, \ldots, Z_k), r(Z_k - \mathrm{E}_k(Z)) \le y\sqrt{(k-1)\mathrm{Var}_k(Z)}) \\ \approx \quad & d\mathbb{P}(Z_k \le -G/\sqrt{2\log k} - c_{k-1}, rZ_k \le y\sqrt{k-1}), \end{aligned}$$

where Z_k and G are independent.

We are now ready to formulate our approximation for (1).

Lemma 5. *For fixed $a \in \mathbb{R}$, we have*

$$\begin{aligned}
&\int_{-r}^{r} e^{\frac{\delta_k}{\sigma} y} d_y \mathbb{P}(Z_k \le -a/\sqrt{2\log k} - c_{k-1}, rZ_k/\sqrt{k-1} \le y) \\
&= \exp\left(\frac{\eta_k^2}{2(k-1)}\right)\left[\Phi\left(\min\left(\max\left(-\sqrt{k-1}, \frac{-a}{\sqrt{2\log k}} - c_{k-1}\right), \sqrt{k-1}\right) - \frac{\eta_k}{\sqrt{k-1}}\right) - \Phi\left(-\sqrt{k-1} - \frac{\eta_k}{\sqrt{k-1}}\right)\right].
\end{aligned}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal random variable.

Proof. Letting Y be a centered Gaussian variable with variance $r^2/(k-1)$. For any $\kappa \in \mathbb{R}$, we then have

$$\begin{aligned}
&\int_{-r}^{r} e^{(\delta_k/\sigma)y} d_y \mathbb{P}(Z_k \le \kappa, rZ_k/\sqrt{k-1} \le y) \\
&= \int_{-r}^{r\min(\max(-1,\kappa/\sqrt{k-1}),1)} e^{(\delta_k/\sigma)y} d\mathbb{P}(Y \le y) \\
&= \int_{-r}^{r\min(\max(-1,\kappa/\sqrt{k-1}),1)} e^{(\delta_k/\sigma)y} \frac{\sqrt{k-1}}{r\sqrt{2\pi}} \exp\left(-\frac{(k-1)y^2}{2r^2}\right) dy \\
&= e^{\frac{(\delta_k/\sigma)^2 r^2}{2(k-1)}} \frac{\sqrt{k-1}}{\sqrt{2\pi} r} \int_{-r}^{r\min(\max(-1,\kappa/\sqrt{k-1}),1)} \exp\left(-\frac{\left(y - \frac{(\delta_k/\sigma)r^2}{(k-1)}\right)^2}{2r^2/(k-1)}\right) dy \\
&= e^{\frac{\eta_k^2}{2(k-1)}} \frac{\sqrt{k-1}}{\sqrt{2\pi} r} \int_{-r}^{r\min(\max(-1,\kappa/\sqrt{k-1}),1)} \exp\left(-\frac{\left(y - \frac{(\delta_k/\sigma)r^2}{(k-1)}\right)^2}{2r^2/(k-1)}\right) dy \\
&= e^{\frac{\eta_k^2}{2(k-1)}} \left[\Phi\left(\min(\max(-\sqrt{k-1}, \kappa), \sqrt{k-1}) - \frac{(\delta_k/\sigma)r}{\sqrt{k-1}}\right) - \Phi\left(-\sqrt{k-1} - \frac{(\delta_k/\sigma)r}{\sqrt{k-1}}\right)\right] \\
&= e^{\frac{\eta_k^2}{2(k-1)}} \left[\Phi\left(\min(\max(-\sqrt{k-1}, \kappa), \sqrt{k-1}) - \frac{\eta_k}{\sqrt{k-1}}\right) - \Phi\left(-\sqrt{k-1} - \frac{\eta_k}{\sqrt{k-1}}\right)\right],
\end{aligned}$$

as claimed. □

We thus approximate the probability of first eliminating the best system by

$$\frac{\exp\left(\frac{\eta_k^2}{2(k-1)}\right)\left[\mathrm{E}\Phi\left(\min\left(\max\left(-\sqrt{k-1}, \frac{-G}{\sqrt{2\log k}} - c_{k-1}\right), \sqrt{k-1}\right) - \frac{\eta_k}{\sqrt{k-1}}\right) - \Phi\left(-\sqrt{k-1} - \frac{\eta_k}{\sqrt{k-1}}\right)\right]}{(\eta_k/2)^{-\nu}\Gamma(\nu+1)I_\nu(\eta_k)}. \tag{3}$$

The expectation in (3) can be estimated through Monte Carlo by generating Gumbel random variates, which can be done fast.

5.2 Other Level Errors

For level ℓ errors for $\ell = 2, 3, \ldots, k-1$, the number of survived systems $|I|$ is $|I| = k - \ell + 1$ and it is natural to modify (3) as follows:

$$\frac{\exp\left(\frac{\eta_{|I|}^2}{2(|I|-1)}\right)\left[\mathrm{E}\Phi\left(\min\left(\max\left(-\sqrt{|I|-1}, \frac{-G}{\sqrt{2\log|I|}} - c_{|I|-1}\right), \sqrt{|I|-1}\right) - \frac{\eta_{|I|}}{\sqrt{|I|-1}}\right) - \Phi\left(-\sqrt{|I|-1} - \frac{\eta_{|I|}}{\sqrt{|I|-1}}\right)\right]}{(\eta_{|I|}/2)^{-\nu}\Gamma(\nu+1)I_\nu(\eta_{|I|})} \tag{4}$$

where $\nu = (|I|-3)/2$. In our procedure, $\eta_{|I|}$ is calculated as the solution to $(4) = \beta_\ell$ for $0 < \beta_\ell < \alpha$. We let $P_k(\ell/k, \beta_\ell)$ represent level ℓ error, the probability of incorrectly eliminating the best system at level ℓ when $\eta_{|I|}$ is calculated with target β_ℓ. Note that it does not depend on δ or σ by Lemma 3. The probability of incorrect selection (PICS) of $\mathcal{DK}_1$ is

$$\mathrm{PICS} = \sum_{\ell=1}^{k-1} P_k(\ell/k, \beta_\ell).$$

Let $\beta_0 = \alpha/(k-1)$. If $P_k(\ell/k, \beta_0)$ for $\ell = 1, \ldots, k-1$ are all approximately equal to β_0, then the overall PICS would be approximately equal to α. For large k, the analysis in Section 5.1 ensures that η_k, the solution to $(3) = \beta_0$, would result in the level 1 error approximately equal to β_0. For other level errors, we do not have control over the error probability but we propose an approximation.

For the derivation of (3), it is critical that the starting point of the corresponding Brownian motion is the origin. For levels $\ell > 1$, we start at a random point from the previous level and thus we do not necessarily have $P_k(\ell/k, \beta_0) \approx \beta_0$ if we let $\eta_{|I|}$ be the solution to $(4) = \beta_0$, unless we discard all observations from previous levels. This is not desirable because too many observations would be wasted. Instead, we seek for a heuristic way to determine $\eta_{|I|}$ under the following assumption:

Assumption 3. *For $0 < \beta_\ell < \alpha$, $\ell = 1, 2, \ldots, k-1$ and $\beta_0 = \alpha/(k-1)$,*

1. *$P_k(\ell/k, \beta_\ell) \approx \beta_\ell \cdot q_k(\ell/k)$; and*
2. *If $\beta_\ell = \beta_0$ for all ℓ, then $P_k(\ell/k, \beta_0) \approx \beta_0 \cdot g(\ell/k) \cdot v_k$ for some v_k and function g.*

Assumption 3.1 is effectively a first-order Taylor approximation under appropriate differentiability assumptions because $\lim_{\beta \downarrow 0} P_k(\ell/k, \beta) = 0$. This assumption implies that for small β_ℓ, the level ℓ error is approximately linear in β_ℓ. For example, if β_ℓ decreases in half for level ℓ, then the level ℓ error is expected to be cut in half.

We have empirical evidence for Assumption 3.2. Figure 3 shows $P_k(\ell/k,\beta_0)/P_k(1/k,\beta_0)$ for $k = 75, 150, 500$ and 1000 as a function of ℓ/k when $\alpha = 0.05$, $\delta = 0.3$ and $\sigma^2 = 1$ with 100,000 replications. (Note that the specific choice of δ and σ does not matter in view of Lemma 3.) From the figure, one can see that the shapes of graphs for various k are similar up to a multiplicative factor v_k.

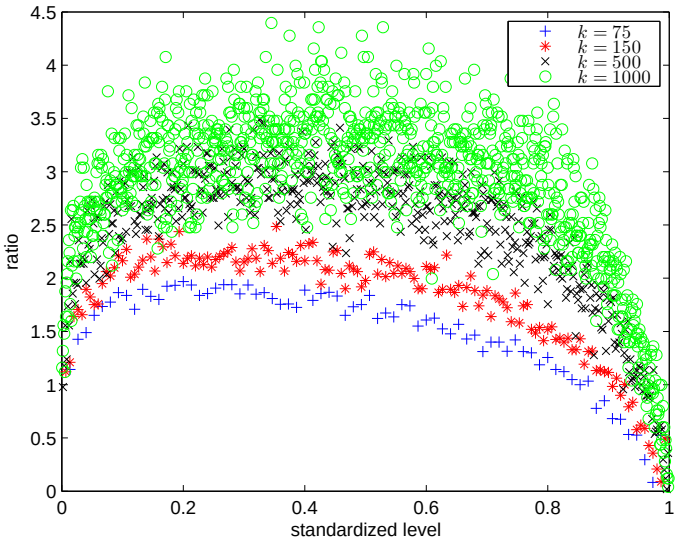


Figure 3: $P_k(\ell/k,\beta_0)/P_k(1/k,\beta_0)$ for various k when $1-\alpha = 0.95$, $\delta = 0.3$, and $\sigma^2 = 1$.

To approximate $g(w)$ for $0 < w < 1$, we use the following steps.

Step 1: Pick k_0, α, δ, and σ^2.

Step 2: Calculate $\eta_{|l|}$, the solution to $(4) = \beta_0$, for all levels.

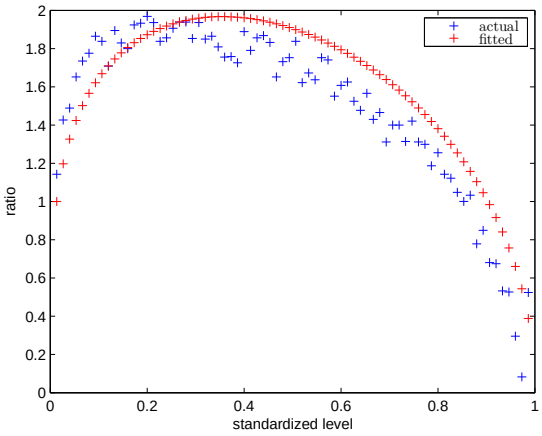
Step 3: Run experiments and approximate $g(w)/g(1/k_0)$ using points

$$\frac{\hat{P}_{k_0}(\ell/k_0,\beta_0)}{\hat{P}_{k_0}(1/k_0,\beta_0)} \quad \text{for} \quad \ell = 1,2,\ldots,k_0-1$$

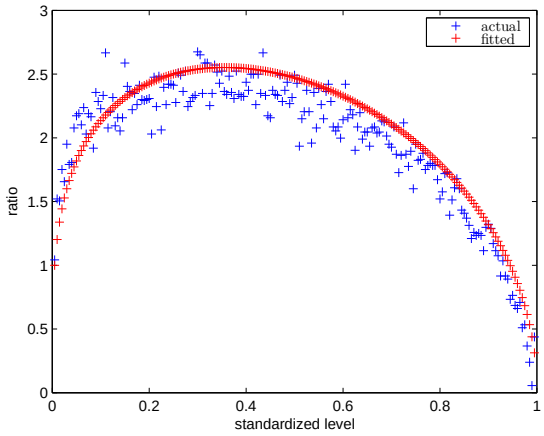
where $\hat{P}$ represents an estimated probability.

For [Step 3], one can use kernel estimation or regression to approximate $g(w)/g(1/k_0)$. We use the following beta-shaped regression for simplicity:

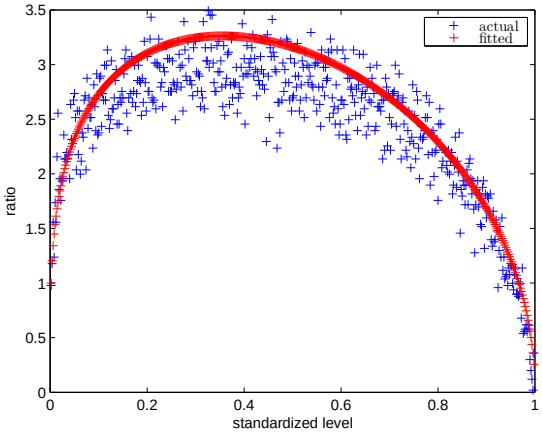
$$\frac{g(w)}{g(1/k_0)} \approx D\, w^A\, (1-w)^B \quad \text{for} \quad 0 < w < 1 \text{ and } A, B, D \in \mathbb{R}.$$



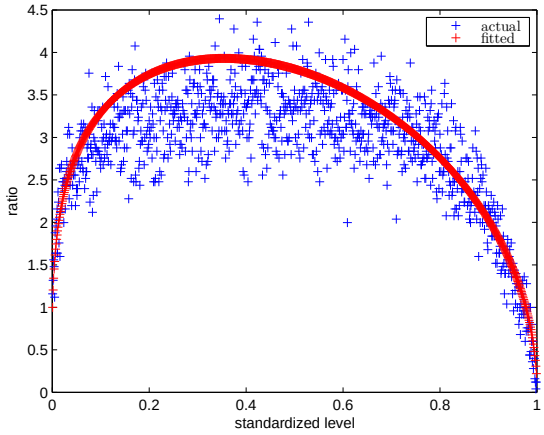
(a) $k = 75, \delta = 0.3, \alpha = 0.05$



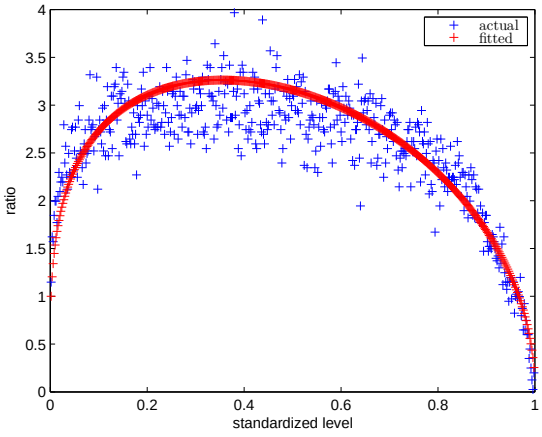
(b) $k = 200, \delta = 0.3, \alpha = 0.05$



(c) $k = 500, \delta = 0.3, \alpha = 0.05$



(d) $k = 1000, \delta = 0.3, \alpha = 0.05$



(e) $k = 500, \delta = 0.1, \alpha = 0.05$

Figure 4: $\frac{\hat{P}_k(\ell/k,\beta_0)}{\hat{P}_k(1/k,\beta_0)}$ vs. m_ℓ when $\frac{g(w)}{g(1/k_0)} \approx 1.76344\, w^{0.269518}\, (1-w)^{0.489079}$

We use $k_0 = 1000$, $\alpha = 0.05$, $\delta = 0.3$, $\sigma^2 = 1$ with 100,000 replications and obtain $A = 0.269518$, $B = 0.489079$, and $D = 1.76344$.

Once $g(w)/g(1/k_0)$ is approximated, $\eta_{|I|}$ which ensures the probability of correct selection (PCS) of $\mathcal{DK}_1$ can be calculated as follows:

Step 1: Calculate the constant m_ℓ as follows:

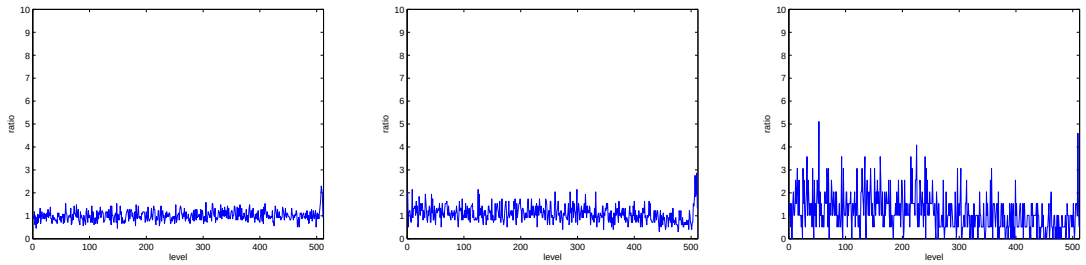
$$\begin{aligned} m_\ell &= \frac{P_k(\frac{\ell}{k}, \frac{\alpha}{k-1})}{P_k(\frac{1}{k}, \frac{\alpha}{k-1})} \approx \frac{g(\ell/k)/g(1/k_0)}{g(1/k)/g(1/k_0)} \\ &= \frac{\left(\frac{\ell}{k}\right)^A \left(1-\frac{\ell}{k}\right)^B}{\left(\frac{1}{k}\right)^A \left(1-\frac{1}{k}\right)^B} = \ell^A \left(\frac{k-\ell}{k-1}\right)^B. \end{aligned}$$

Step 2: Set $\beta_\ell = \frac{\beta_0}{m_\ell}$ and calculate $\eta_{|I|}$ from (4) $= \beta_\ell$.

As $P_k(1/k, \beta_0) \approx \beta_0$ for large k, m_ℓ is approximately the ratio between the actual level ℓ error and the target level error β_0 when $\eta_{|I|}$ is calculated from (4) $= \beta_0$. So if we know the ratio m_ℓ, then $P_k(\ell/k, \beta_0/m_\ell) \approx \beta_0$ by Assumption 3.1, which in turns implies that the overall PICS is approximately equal to α.

Figure 4 compares $\frac{\hat{P}_k(\ell/k, \beta_0)}{\hat{P}_k(1/k, \beta_0)}$ and m_ℓ for various k when a regression for a beta-shaped function is used to estimate $g(w)$ for $k_0 = 1000, \alpha = 0.05, \delta = 0.3$ and $\sigma^2 = 1$. It shows that m_ℓ is a good approximation for the actual ratio between level ℓ error and the target β_0 (or level 1 error). Figure 5 shows estimated level errors $\hat{P}_k(\ell/k, \beta_0/m_\ell)$ for $\alpha = 1, 5, 10\%$ when $k = 512$ with $\delta = 0.3$ and $\sigma^2 = 1$ and 100,000 replications. One can see that the level errors do not show a beta shape as in Figure 4. Instead the level errors fluctuate around $\beta_0 = \alpha/511$ for the three values of α, which empirically supports Assumption 3.1. Also, it shows that the function we found $g(w)$ for $\alpha = 5\%$ and $k_0 = 1000$ seems to work well for other popular choices of α, including $\alpha = 1\%$ and $\alpha = 10\%$.

The parameter $\eta_{|I|}$ requires some computation as we need to estimate the expectation in (4) by generating Gumbel random variates G, which is quick. In addition, since $\eta_{|I|}$ only depends on α and k, a table can be made for popular choices of α such as 5% and 10% and $k = 2, 3, \ldots, 10000$. Then the values of $\eta_{|I|}$ can be read from the table while running our procedure. Table A.1 in the appendix shows the values of $\eta_{|I|}$ for a few selected values of k when $\alpha = 10\%$ when one million samples of G are generated. As (4) is only accurate for large $|I|$, we sample one million standard normal samples and estimate the level error with (2)



(a) $1-\alpha = 0.90, \text{PCS} = 0.899$ (b) $1-\alpha = 0.95, \text{PCS} = 0.945$ (c) $1-\alpha = 0.99, \text{PCS} = 0.988$

Figure 5: Ratios between $\hat{P}_k(\ell/k, \beta_0/m_\ell)$ and β_0 for $\alpha = 10\%, 5\%, 1\%$ when $k = 512$, $\delta = 0.3$ and $\sigma^2 = 1$.

for $\max(1, k-9) \le \ell \le k-2$ (i.e., $|I| \le 10$). For $|I| = 2$, η_2 is set to $\eta = -\ln(2\beta_0)$, the parameter from Paulson (1964) when $k = 2$, the nominal confidence level is β_0, and its continuation region is defined by $(-a, a)$ for $a > 0$. When $k = 2$, it is not difficult to show that our elimination rule is identical to a continuation region defined by parallel lines in Paulson (1964).

5.3 Justification of Procedures for Unknown Variances

In this subsection, we discuss why $\mathcal{DK}_2$ and $\mathcal{DK}_3$ should be expected to work for unknown variances as well. For unknown variances, it is natural to replace variance parameters in $\mathcal{DK}_1$ to their estimated values. In general, it is not sufficient to replace the variance parameter with its estimated value to keep the statistical validity. It is critical to account for the variability in the estimated parameter especially when variances are estimated only once based on an initial n_0 observations. However, if variance estimators are updated on the fly in a procedure as more observations are obtained, then it can be shown that the procedure converges to the known variance case under some appropriate asymptotic regime as in Kim and Nelson (2006) and Wang and Kim (2011). We employ a variance updating scheme in $\mathcal{DK}_2$ and $\mathcal{DK}_3$ to avoid the difficulty of accounting for the variability in the estimated variance parameters.

When the decision maker believes that the variances across systems are equal (but unknown), then the natural estimator for σ^2 is the pooled variance estimator

$$\hat{\sigma}_p^2(n) = \frac{1}{|I|} \sum_{i \in I} \hat{\sigma}_i^2(n).$$

As we update $\hat{\sigma}_p^2(n)$ as more observations become available, the estimator converges to σ^2 and thus it is expected that $\mathcal{DK}_2$ works similarly to $\mathcal{DK}_1$.

When variances are unknown and unequal, we use similar arguments as in Frazier(2014). Let $n_i = \gamma\sigma_i^2 n$ for some $\gamma > 0$ and thus the number of samples obtained by stage n for system i is proportional to its variance σ_i^2. Then

$$\frac{\sum_{j=1}^{n_i} X_{ij}}{\gamma\sigma_i^2} \sim N\left(\frac{n_i}{\gamma\sigma_i^2}\mu_i, \frac{n_i}{\gamma^2\sigma_i^2}\right) = N\left(n\mu_i, \frac{n}{\gamma}\right) \approx B_{(\mu_i,1/\gamma)}(t)$$

where $B_{(\mu_i,1/\gamma)}(t)$ is a Brownian motion with drift μ_i and variance $1/\gamma$. The $\frac{\sum_{j=1}^{n_i} X_{ij}}{\gamma\sigma_i^2}$ have equal variance as long as $n_i = \gamma\sigma_i^2 n$ and thus we can apply $\mathcal{DK}_1$ to $\frac{\sum_{j=1}^{n_i} X_{ij}}{\gamma\sigma_i^2}$. Note that when $n_i = \gamma\sigma_i^2 n$,

$$n_i\lambda^2 = n_i\frac{\sum_{i\in I}\sigma_i^2}{\sum_{i\in I} n_i} = \sigma_i^2$$

where

$$\lambda^2 = \frac{\sum_{i\in I}\sigma_i^2(n_i)}{\sum_{i\in I} n_i}.$$

Then

$$\frac{\sum_{j=1}^{n_i} X_{ij}}{\gamma\sigma_i^2} = \frac{\sum_{j=1}^{n_i} X_{ij}}{\gamma n_i\lambda^2} = \frac{W_i(n)}{\gamma\lambda^2}.$$

Finally, the screening rule in $\mathcal{DK}_1$ is

$$\frac{\sum_{i\in I}\left(\frac{W_i(n)}{\gamma\lambda^2} - \frac{1}{|I|}\sum_{i\in I}\frac{W_i(n)}{\gamma\lambda^2}\right)^2}{1/\gamma} \geq \frac{1}{\gamma}\left(\frac{\eta_{|I|}}{\delta_{|I|}}\right)^2$$

which is equivalent to

$$\frac{1}{\lambda^4}\sum_{i\in I}\left(W_i(n) - \frac{1}{|I|}\sum_{i\in I} W_i(n)\right)^2 \geq \left(\frac{\eta_{|I|}}{\delta_{|I|}}\right)^2$$

or

$$\frac{1}{\lambda^2}\sum_{i\in I}\left(W_i(n) - \frac{1}{|I|}\sum_{i\in I} W_i(n)\right)^2 \geq \left(\frac{\lambda\cdot\eta_{|I|}}{\delta_{|I|}}\right)^2 \tag{5}$$

When λ^2 is replaced with its estimator $\hat{\lambda}^2$ in (5), we get the same elimination rule in the $\mathcal{DK}_3$ procedure, which is

$$\mathcal{S}_I''(\boldsymbol{W}_I(n)) \geq \left(\frac{\hat{\lambda}\cdot\eta_{|I|}}{\delta_{|I|}}\right)^2.$$

6. Experiments

In this section, we compare the performance of $\mathcal{DK}$ procedures with KN and BIZ. For unknown variances, we use the KN procedure as originally described in Kim and Nelson (2001) with $c = 1$ and $n_0 = 30$ and

Table 1: Mean and variance configurations

Configuration	Means	Variances	δ	α
SC-Equal	$\mu = [\delta, 0, \ldots, 0]$	$\sigma^2 = 100$	1	0.1
MDM-Equal	$\mu_i = -\delta i$	$\sigma^2 = 100$	1	0.1
SC-INC	$\mu = [\delta, 0, \ldots, 0]$	$\sigma_i^2 = 25\left(1 + 3\frac{i-1}{k-1}\right)^2$	1	0.1
SC-DEC	$\mu = [\delta, 0, \ldots, 0]$	$\sigma_i^2 = 25\left(1 + 3\frac{k-i}{k-1}\right)^2$	1	0.1
MDM-INC	$\mu_i = -\delta i$	$\sigma_i^2 = 25\left(1 + 3\frac{i-1}{k-1}\right)^2$	1	0.1
MDM-DEC	$\mu_i = -\delta i$	$\sigma_i^2 = 25\left(1 + 3\frac{k-i}{k-1}\right)^2$	1	0.1

Algorithm 2 of Frazier (2014) with $B_z = 1$ and $n_0 = 30$. For known variances, we use KN with $h^2 = 2\eta$ where $\eta = -\ln\left(2\frac{\alpha}{k-1}\right)$ and $n_0 = 1$, which is same as the $\mathcal{P}$ procedure in Wang and Kim (2011), and Algorithm 1 of Frazier (2014). Throughout this section, KN and BIZ refer procedures for known variances while KN-UNK and BIZ-UNK refer procedures for unknown variances.

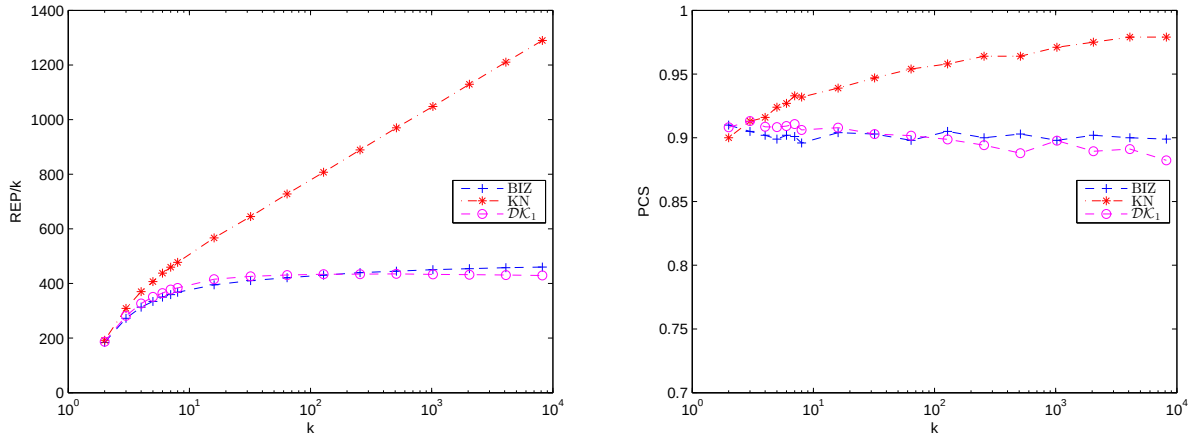
The number of systems k varies over

$$k \in \{2, 3, 4, 5, 6, 7, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096, 8192\}.$$

For the mean, we consider two mean configurations, namely slippage configuration (SC) and monotonic decreasing mean configuration (MDM); and for variances, we consider three variance configurations called Equal, INC, and DEC. Thus we have total six configurations: SC-Equal, MDM-Equal, SC-INC, SC-DEC, MDM-INC and MDM-DEC. We use same parameter settings for mean, variances, δ and α as in Frazier (2014). Table 1 gives all six mean-variance configurations and other parameter settings.

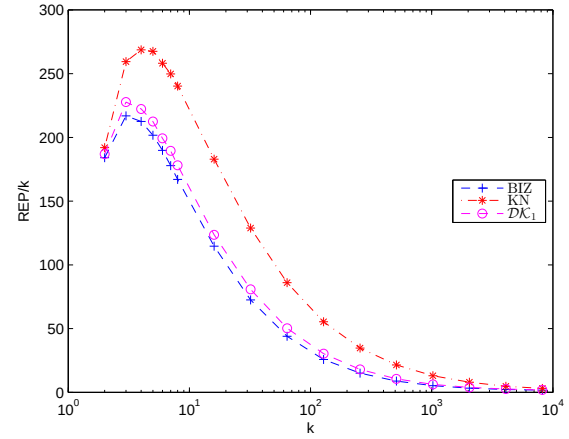
When calculating $\eta_{|I|}$ for $\mathcal{DK}$ procedures, we sample 1, 2, 4, 8, and 16 million Gumbel samples for $k \leq 1024$, $k = 2048$, $k = 4096$ and $k = 8192$, respectively, when $1 \leq \ell \leq \max(k-2, k-10)$ (i.e., $|I| \geq 11$). We also take logs to avoid numerical overflows and underflows in the denominator, since the Gamma term can be very large and the Bessel term can be very small. When $\max(1, k-9) \leq \ell \leq k-2$ (i.e., $|I| \leq 10$), we use Monte Carlo sampling through (2). When $\ell = k-1$ or only two systems are survived, we use $\eta_2 = -\ln(2\beta_0)$.

The nominal confidence level is set to $1 - \alpha = 0.9$. Estimated probability of correct selection (PCS) and an average number of observations per system until a decision is made (REP/k) are reported based on 10,000 macro replications.

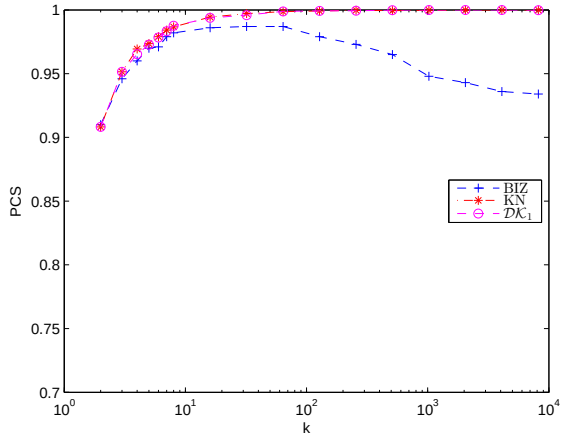


(a) SC-REP

(b) SC-PCS



(c) MDM-REP

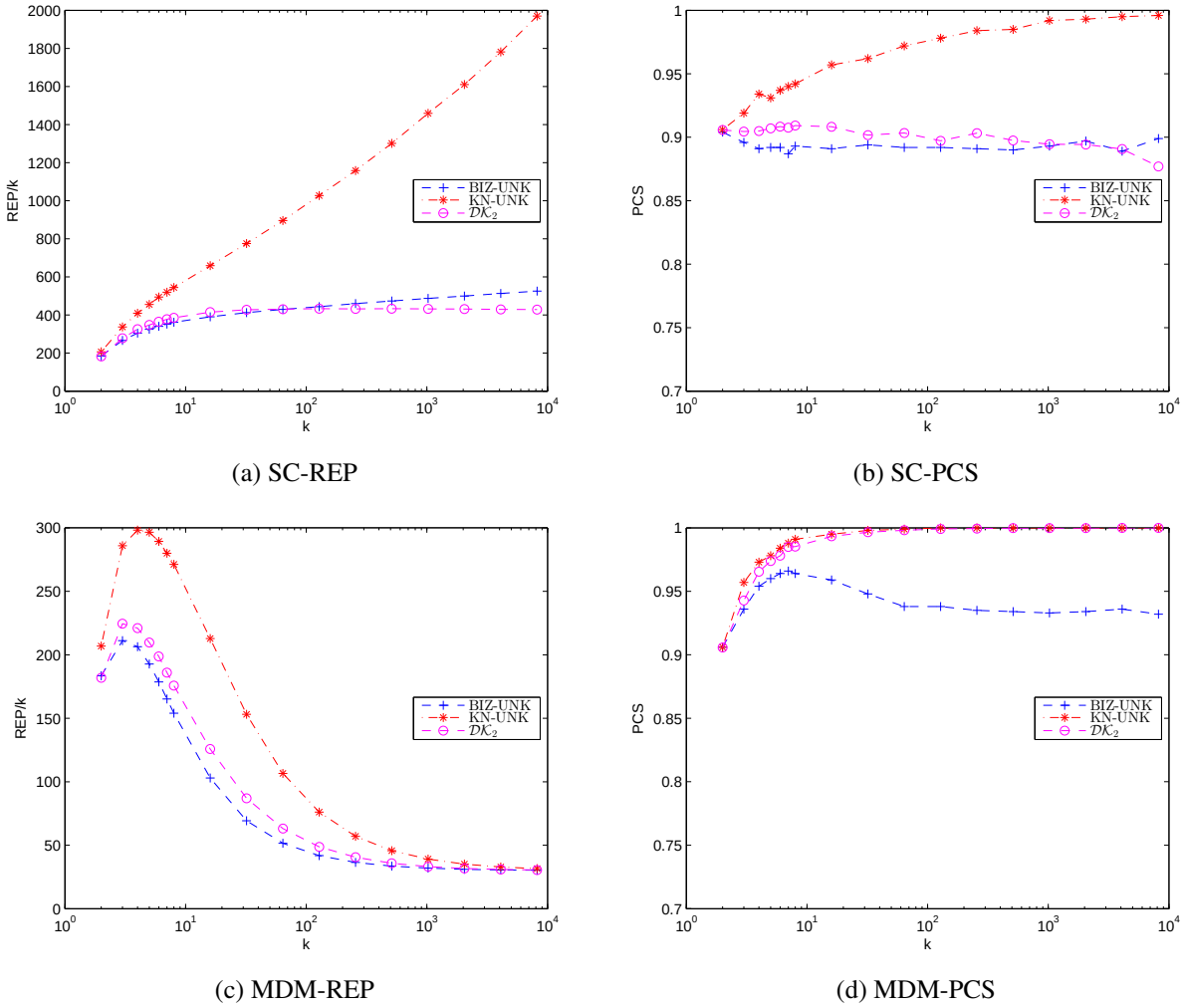


(d) MDM-PCS

Figure 6: REP/k and PCS for $\mathcal{DK}_1$ when variances are known and equal with $1-\alpha = 0.9$

6.1 $\mathcal{DK}_1$ with Known and Equal Variances

When variances are known and equal, we compare $\mathcal{DK}_1$ with KN and BIZ. Figure 6 shows REP/k and PCS under SC and MDM configurations. Procedure $\mathcal{DK}_1$ significantly outperforms KN under both SC and MDM. When k is large, $\mathcal{DK}_1$ is more than three times better than KN in terms of REP/k. On the other hand, the performances of BIZ and $\mathcal{DK}_1$ are very similar under the slippage configuration in terms of both REP/k and PCS. When k is large, $\mathcal{DK}_1$ spends a slightly fewer number of observations than BIZ but its probability of correct selection is slightly lower than BIZ. Under the monotonic decreasing mean configuration, $\mathcal{DK}_1$ achieves PCS greater than the nominal value 90% and clearly outperforms KN. However, BIZ achieves PCS close to the nominal value 90% than $\mathcal{DK}_1$ and spends slightly fewer observations than $\mathcal{DK}_1$.



(a) SC-REP

(b) SC-PCS

(c) MDM-REP

(d) MDM-PCS

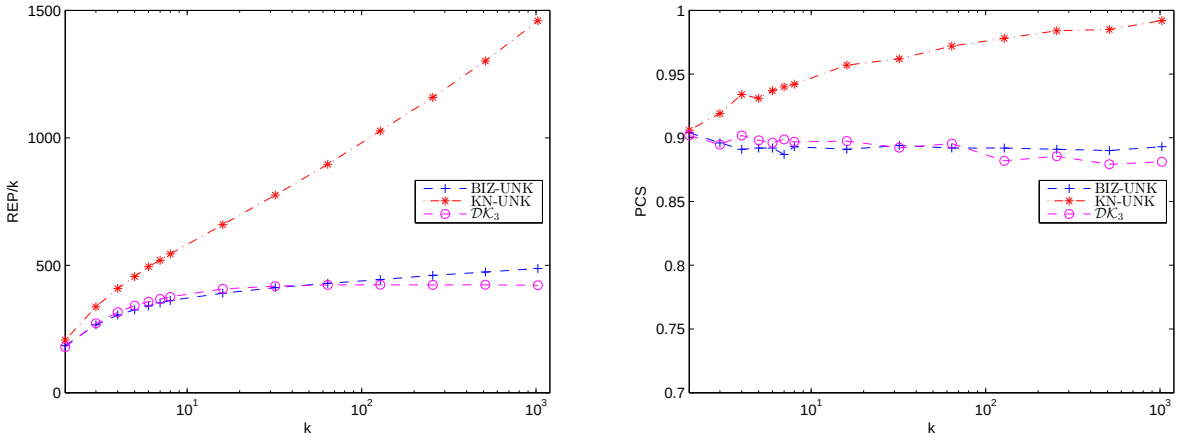
Figure 7: REP/k and PCS for $\mathcal{DK}_2$ when variances are unknown but equal with $1-\alpha = 0.9$

6.2 $\mathcal{DK}_2$ and $\mathcal{DK}_3$ with Unknown but Equal Variances

When variances are unknown but a decision maker knows that variances across systems are equal, $\mathcal{DK}_2$ or $\mathcal{DK}_3$ can be used.

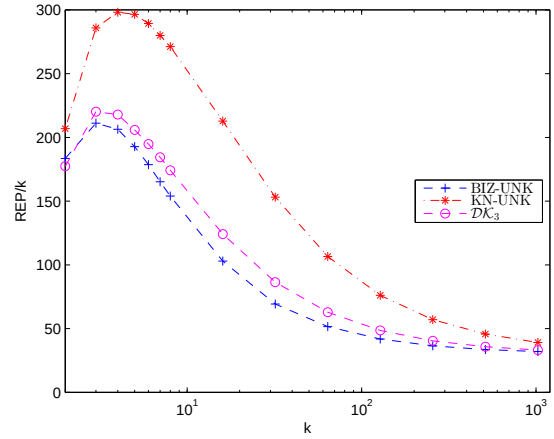
Figure 7 compares performances of $\mathcal{DK}_2$ with those of KN-UNK and BIZ-UNK. As in the case of known and equal variances, $\mathcal{DK}_2$ outperforms KN-UNK and shows similar performances as BIZ-UNK.

In reality, it is impossible to know in advance whether variances across systems are equal. In fact, equal variances across systems rarely hold. Thus we also consider $\mathcal{DK}_3$. Our experiments show that $\mathcal{DK}_3$ actually spends slightly fewer observations than $\mathcal{DK}_2$ while achieving similar PCS. Figure 8 compares $\mathcal{DK}_3$ with KN-UNK and BIZ-UNK. Graphs in Figure 8 show similar tendency as those in Figure 7.

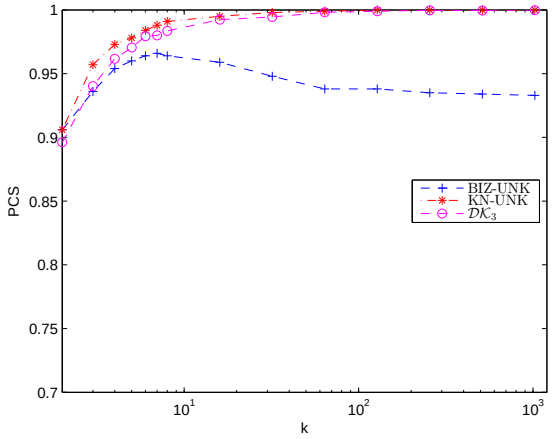


(a) SC-REP

(b) SC-PCS



(c) MDM-REP



(d) MDM-PCS

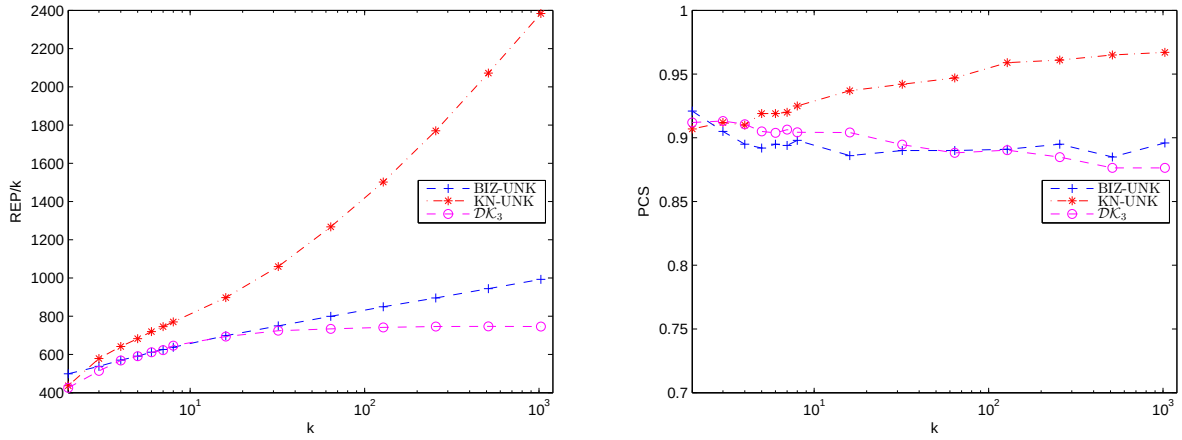
Figure 8: REP/k and PCS for $\mathcal{DK}_3$ when variances are unknown but equal with $1-\alpha = 0.9$

6.3 $\mathcal{DK}_3$ with Unknown and Unequal Variances

Finally, we consider unknown and unequal variances. Figure 9 compares the three procedures under the slippage configuration with increasing and decreasing variances while Figure 10 compares them under the MDM configuration with increasing and decreasing variances.

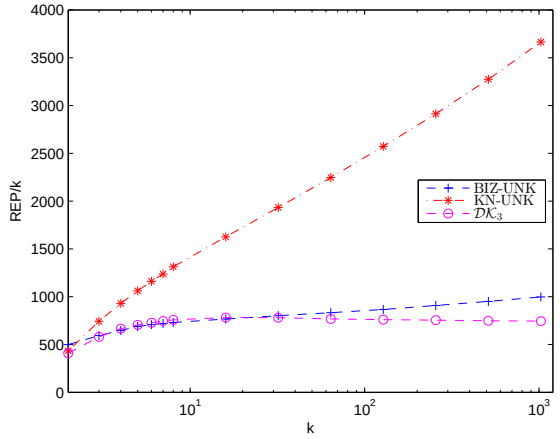
The efficiency of $\mathcal{DK}_3$ compared to KN-UNK is more obvious. When $k = 8192$, $\mathcal{DK}_3$ is four times better than KN-UNK under SC-INC and eight times better under SC-DEC in terms of REP/k while achieving PCS close to 90%. Unlike equal variances, $\mathcal{DK}_3$ spends slightly fewer observations than BIZ-UNK for small k and then outperforms it for large k under the slippage configuration.

Interestingly, under the MDM configuration with increasing variances, $\mathcal{DK}_3$ significantly outperforms

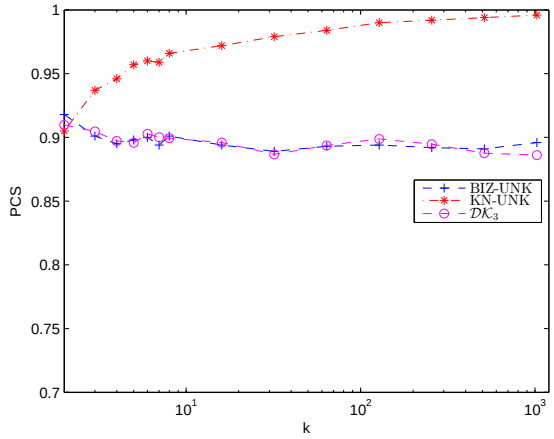


(a) SC-INC-REP

(b) SC-INC-PCS



(c) SC-DEC-REP



(d) SC-DEC-PCS

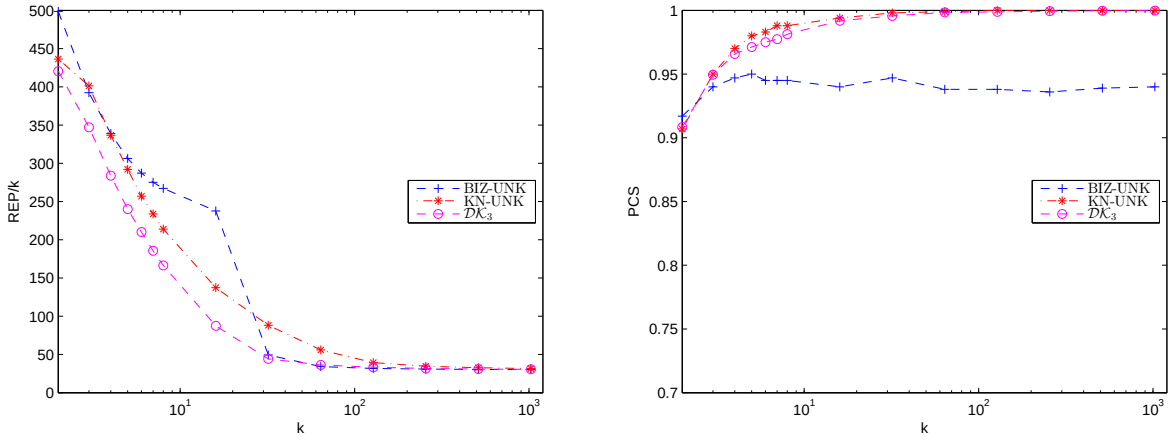
Figure 9: REP/k and PCS for $\mathcal{DK}_3$ when variances are unknown and unequal with $1-\alpha = 0.9$

both KN-UNK and BIZ-UNK, but uses more observations than BIZ-UNK under decreasing variances.

Overall, $\mathcal{DK}$ procedures achieve PCS close to the nominal value for all settings we tested and they outperform KN significantly while performing similarly to BIZ.

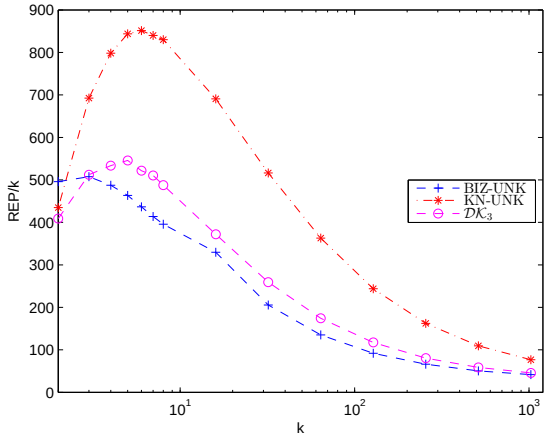
7. Conclusions

We present new fully-sequential procedures whose continuation regions are derived exploiting the properties of multidimensional Brownian motions, which is the first work in the literature. Our procedures deliver a probability of correct selection close to the nominal level. Compared to the existing state-of-art fully-sequential IZ procedure KN, the proposed procedures show a tight worst-case probability of incorrect

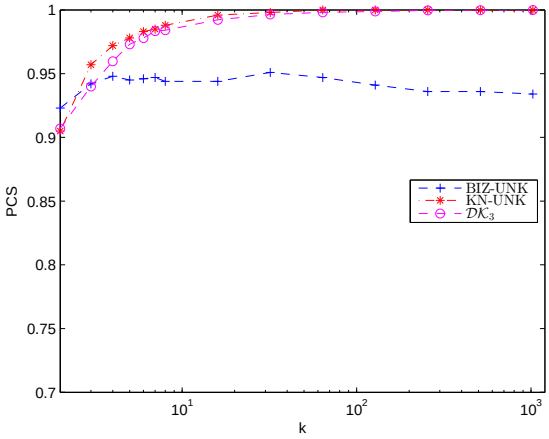


(a) MDM-INC-REP

(b) MDM-INC-PCS



(c) MDM-DEC-REP



(d) MDM-DEC-PCS

Figure 10: REP/k and PCS for $\mathcal{DK}_3$ when variances are unknown and unequal with $1-\alpha = 0.9$

selection under the slippage configuration and significant savings in the number of observations needed until a decision is made. Compared to BIZ, our procedures perform better for a large number of systems under difficult mean configurations (albeit with a slightly lower probability of correct selection than BIZ for a large number of systems), but spend slightly more (but similar) observations under easier mean configurations. There are at least two sources of slight loss in the PCS. First, in finding the functional form of $g(w)$, kernel estimation may work better than the simple beta-shaped regression line. Another source the evaluation of the analytical expression of the level error. The target β_ℓ can be really small for large k. For example, when $k = 10000$ and $\alpha = 10\%$, $\beta_{3551} \approx 1.4 \times 10^{-6}$. Thus a naive estimation of the probability based on sample average of Gumbel random variates may not be accurate and the use of variance reduction techniques might

be desirable.

Acknowledgements

This work is supported by the National Science Foundation under grant CMMI-1131047. The authors would like to thank Seunghan Lee for his insight for Lemma 1. In addition, the authors appreciate Peter Frazier for his codes and helpful comments and Barry Nelson for his helpful comments.

References

Chick, S. E. 2006. Subjective Probability and Bayesian Methodology. In *Handbooks in Operations Research and Management Science: Simulation*, edited by S. G. Henderson and B. L. Nelson. Oxford: Elsevier Science.

Chen, C.-H., and L. H. Lee. 2010. *Stochastic Simulation Optimization: An Optimal Computing Budget Allocation (System Engineering and Operations Research)*, vol 1. Singapore: World Scientific, 2010.

Chow, T. L., and J. L.Teugels. 1978. The Sum and the Maximum of I.I.D. Random Variables. In *Proceedings of the Second Prague Symposium on Asymptotic Statistics*, edited by P. Mandl and M. Huskova, 81-92. New York: North-Holland.

Dieker, A. B., and S.-H. Kim. 2012. Selecting the Best by Comparing Simulated Systems in a Group of Three When Variances are Known and Unequal. In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. M. Uhrmacher, 1-7. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Dieker, A. B., and S.-H. Kim 2014. "A Fully Sequential Procedure for Known and Equal Variances Based on Multivariate Brownian Motion". In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk, S. D. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, 3749-3760. Piscataway, New Jersey: IEEE.

Embrechts, P., C. Kluppelberg, and T. Mikosch. 1997. *Modelling Extremal Events for Insurance and Finance*. New York: Springer.

Frazier, P. 2014. A Fully Sequential Elimination Procedure for Indifference-Zone Ranking and Selection

with Tight Bounds on Probability of Correct Selection. *Operations Research* 62(4):926-942.

Nelson, B. L., J. Swann, D. Goldsman, and W. Song. 2001. "Simple procedures for selecting the best simulated system when the number of alternatives is large". *Operations Research* 49(6):950-963.

Kim, S.-H., and A. B. Dieker. 2011. Selecting the Best by Comparing Simulated Systems in a Group of Three. In *Proceedings of the 2011 Winter Simulation Conference*, edited by S. Jain, R. R. Creasey, J. Himmelspach, K. P. White, and M. Fu. 4217-4226. Piscataway, New Jersey: IEEE.

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.

Rinott, Y. 1978. "On two-stage selection procedures and related probability inequalities". *Comm. Statist.-Theory and Methods* 7(8):799-811.

Rogers, L., and J. W. Pitman. 1981. Markov Functions. *The Annals of Probability* 9:573-582.

Wang, H., and S.-H. Kim. 2011. Reducing the Conservativeness of Fully Sequential Indifference-Zone Procedures. *IEEE Transactions on Automatic Control* 58(6):1613-1619

Appendix

Proof of Lemma 1.

$$\begin{aligned}\mathcal{S}_I(\Pi x) &= (V\Pi x)^T(V\Gamma V^T)^{-1}(V\Pi x)\\ &= (V\Gamma V^T(V\Gamma V^T)^{-1}Vx)^T(V\Gamma V^T)^{-1}(V\Gamma V^T(V\Gamma V^T)^{-1}Vx)\\ &= (Vx)^T(V\Gamma V^T)^{-1}(Vx) = \mathcal{S}_I(x).\end{aligned}$$

□

Proof of Corollary 1. We first derive an explicit expression for $(V\Gamma V^T)^{-1}$. Without loss of generality, assume that $I = \{1,\dots,s\}$. Then by noting that $V\Gamma V^T$ is the covariance matrix of Vx, we get

$$Vx = \begin{bmatrix} x_1 - x_s \\ \vdots \\ x_{s-1} - x_s \end{bmatrix} \qquad \text{and} \qquad V\Gamma V^T = \begin{bmatrix} \sigma_1^2+\sigma_s^2 & \sigma_s^2 & \cdots & \cdots & \sigma_s^2 \\ \sigma_s^2 & \sigma_2^2+\sigma_s^2 & \sigma_s^2 & \cdots & \sigma_s^2 \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \sigma_s^2 \\ \sigma_s^2 & \cdots & \cdots & \sigma_s^2 & \sigma_{s-1}^2+\sigma_s^2 \end{bmatrix}.$$

For equal variances,

$$V\Gamma V^T = \sigma^2 \begin{bmatrix} 2 & 1 & \cdots & \cdots & 1 \\ 1 & 2 & 1 & \cdots & 1 \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & 1 \\ 1 & \cdots & \cdots & 1 & 2 \end{bmatrix} = \sigma^2\left(\mathrm{id}_{s-1} + \mathbf{1}_{s-1}\mathbf{1}_{s-1}^T\right)$$

where id_s is the $s\times s$ identity matrix and $\mathbf{1}_s$ is the $s\times 1$ vector of ones.

By the Sherman-Morrison formula,

$$\begin{aligned}(V\Gamma V^T)^{-1} &= \frac{1}{\sigma^2}\left(\mathrm{id}_{s-1}^{-1} - \frac{\mathrm{id}_{s-1}^{-1}\mathbf{1}_{s-1}\mathbf{1}_{s-1}^T\mathrm{id}_{s-1}^{-1}}{1+\mathbf{1}^T\mathrm{id}_{s-1}^{-1}\mathbf{1}}\right)\\ &= \frac{1}{\sigma^2}\left(\mathrm{id}_{s-1} - \frac{\mathbf{1}_{s-1}\mathbf{1}_{s-1}^T}{1+(s-1)}\right)\\ &= \frac{1}{\sigma^2}\frac{1}{s}\left(s\cdot\mathrm{id}_{s-1} - \mathbf{1}_{s-1}\mathbf{1}_{s-1}^T\right). \qquad (6)\end{aligned}$$

Then we have

$$\begin{aligned}
\mathcal{S}_I(x) &= \frac{1}{\sigma^2}\frac{1}{s}\begin{bmatrix} x_1 - x_s \\ \vdots \\ x_{s-1} - x_s \end{bmatrix}^T \left(s \cdot \mathrm{id}_{s-1} - \mathbf{1}_{s-1}\mathbf{1}_{s-1}^T\right)\begin{bmatrix} x_1 - x_s \\ \vdots \\ x_{s-1} - x_s \end{bmatrix} \\
&= \frac{1}{\sigma^2}\frac{1}{s}\left\{(s-1)\sum_{i=1}^{s-1}(x_i - x_s)^2 - 2\sum_{1\le i<\ell<s}(x_i - x_s)(x_\ell - x_s)\right\} \\
&= \frac{1}{\sigma^2}\frac{1}{s}\sum_{\substack{i<\ell \\ i,\ell\in I}}(x_i - x_\ell)^2,
\end{aligned}$$

which shows the first equality in the corollary because $|I| = s$.

Now we show the second equality of the corollary. From (6),

$$V^T(V\Gamma V^T)^{-1}V = \frac{1}{\sigma^2}\frac{1}{s}(s\cdot\mathrm{id}_s - \mathbf{1}_s\mathbf{1}_s^T) \qquad \text{and} \qquad \Pi = \Gamma V^T(V\Gamma V^T)^{-1}V = \frac{1}{s}(s\cdot\mathrm{id}_s - \mathbf{1}_s\mathbf{1}_s^T).$$

Then

$$\Pi x = \frac{1}{s}(s\cdot\mathrm{id}_s - \mathbf{1}_s\mathbf{1}_s^T)x = \begin{bmatrix} x_1 - \bar{x} \\ \vdots \\ x_s - \bar{x} \end{bmatrix}.$$

Finally,

$$\begin{aligned}
\mathcal{S}_I(\Pi x) &= (V\Pi x)^T(V\Gamma V^T)^{-1}(V\Pi x) \\
&= (\Pi x)^T[V^T(V\Gamma V^T)^{-1}V](\Pi x) \\
&= \frac{1}{\sigma^2}\frac{1}{s}\begin{bmatrix} x_1 - \bar{x} \\ \vdots \\ x_s - \bar{x} \end{bmatrix}^T (s\cdot\mathrm{id}_s - \mathbf{1}_s\mathbf{1}_s^T)\begin{bmatrix} x_1 - \bar{x} \\ \vdots \\ x_s - \bar{x} \end{bmatrix} \\
&= \frac{1}{\sigma^2}\begin{bmatrix} x_1 - \bar{x} \\ \vdots \\ x_s - \bar{x} \end{bmatrix}^T\begin{bmatrix} x_1 - \bar{x} \\ \vdots \\ x_s - \bar{x} \end{bmatrix} \\
&= \frac{1}{\sigma^2}\sum_{i=1}^{s}(x_i - \bar{x})^2.
\end{aligned}$$

□

Proof of Lemma 2. It suffices to prove the claim for $|I| = |J| + 1$. By relabeling systems if necessary, it suffices to prove the claim with $J = \{1,\ldots,s\}$ and $I = \{1,\ldots,s+1\}$. We set

$$H_{s+1} = \left\{(x_1, x_2, \ldots, x_{s+1})^T : \sum_{i=1}^{s+1} x_i = 0\right\}, \qquad Q_s = \left\{(x_1, x_2, \ldots, x_{s+1})^T : \sum_{i=1}^{s} x_i = 0, x_{s+1} = 0\right\}.$$

By the second equality of Corollary 1, it suffices to show that for $x \in \mathbb{R}^{s+1}$,

$$\mathcal{S}_I(x) \geq \frac{1}{\sigma^2} \sum_{i=1}^{s} (x_i - \bar{x}_s)^2, \tag{7}$$

where $\bar{x}_s = (x_1 + \cdots + x_s)/s$. To see that this holds, we define Ψ_s on H_{s+1} as the matrix that projects orthogonally on Q_s, i.e., $\Psi_s x = (x_1 - \bar{x}_s, \ldots, x_s - \bar{x}_s, 0)$. By Lemma 1 and (6), we have, for $x \in \mathbb{R}^{s+1}$,

$$\mathcal{S}_I(x) = \frac{1}{\sigma^2} \frac{1}{(s+1)} \begin{bmatrix} x_1 - x_{s+1} \\ \vdots \\ x_s - x_{s+1} \end{bmatrix}^T \left((s+1) \cdot \mathrm{id}_s - \mathbf{1}_s \mathbf{1}_s^T \right) \begin{bmatrix} x_1 - x_{s+1} \\ \vdots \\ x_s - x_{s+1} \end{bmatrix}$$

This representation immediately yields that

$$\mathcal{S}_I(\Psi_s x) = \frac{1}{\sigma^2} \sum_{i=1}^{s} (x_i - \bar{x}_s)^2.$$

Since projecting decreases any quadratic form, this establishes (7). □

Table A.1: $\eta_{|I|}$ when $\alpha = 10\%$

$\|I\|$	$k = 64$	$k = 32$	$k = 16$	$k = 8$	$k = 7$	$k = 6$	$k = 5$	$k = 4$	$k = 3$
64	6.09023								
63	6.55214								
62	6.79726								
61	6.97011								
60	7.08828								
59	7.18300								
58	7.25234								
57	7.31582								
56	7.35586								
55	7.38613								
54	7.41836								
53	7.44472								
52	7.45156								
51	7.46425								
50	7.47109								
49	7.47207								
48	7.47304								
47	7.46425								
46	7.45937								
45	7.44863								
44	7.43398								
43	7.42128								
42	7.39687								
41	7.38222								
40	7.35781								
39	7.33144								
38	7.30507								
37	7.27285								
36	7.24453								
35	7.20839								
34	7.17226								
33	7.14199								
32	7.09609	4.66648							
31	7.05605	5.00534							
30	7.01406	5.18796							
29	6.96816	5.29245							
28	6.91836	5.36374							
27	6.87246	5.41550							
26	6.81875	5.44577							
25	6.76308	5.46433							
24	6.70546	5.47312							
23	6.64785	5.47214							
22	6.58437	5.46140							
21	6.52089	5.44480							
20	6.45546	5.42624							
19	6.38222	5.39401							
18	6.31093	5.36569							
17	6.23378	5.32370							
16	6.15468	5.28366	3.64162						
15	6.07168	5.23386	3.89210						
14	5.98671	5.18210	4.00953						
13	5.89589	5.12448	4.06714						
12	5.80117	5.05905	4.09937						
11	5.70253	4.99558	4.11207						
10	5.60000	4.92429	4.10718						
9	5.24160	4.64987	3.95972						
8	5.05507	4.49265	3.85779	2.83287					
7	4.84121	4.32370	3.73847	2.93499	2.66279				
6	4.60000	4.11765	3.58505	2.91574	2.74253	2.47228			
5	4.32168	3.87644	3.39116	2.82603	2.69060	2.51641	2.24682		
4	3.97890	3.57201	3.14382	2.66009	2.55331	2.42129	2.24939	1.98345	
3	3.51777	3.16589	2.79554	2.39352	2.30888	2.20772	2.07880	1.90481	1.63140
2	3.22207	2.86667	2.50197	2.11705	2.03877	1.94591	1.83178	1.68365	1.47222