# Efficient Multinomial Selection in Simulation

**J. O. Miller,[1] Barry L. Nelson,[2] Charles H. Reilly[3]**

[1] *Department of Operational Sciences, Air Force Institute of Technology, Wright-Patterson AFB, Ohio 45433-7765, USA*

[2] *Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois 60208-3119, USA*

[3] *Department of Industrial Engineering and Management Systems, University of Central Florida, Orlando, Florida 32816-2450, USA*

Received March 1997; revised February 1998; accepted 22 February 1998

**Abstract:** Consider a simulation experiment consisting of $v$ independent vector replications across $k$ systems, where in any given replication one system is selected as the best performer (i.e., it wins). Each system has an unknown constant probability of winning in any replication and the numbers of wins for the individual systems follow a multinomial distribution. The classical multinomial selection procedure of Bechhofer, Elmaghraby, and Morse (Procedure BEM) prescribes a minimum number of replications, denoted as $v^*$, so that the probability of correctly selecting the true best system (PCS) meets or exceeds a prespecified probability. Assuming that larger is better, Procedure BEM selects as best the system having the largest value of the performance measure in more replications than any other system. We use these same $v^*$ replications across $k$ systems to form $(v^*)^k$ pseudoreplications that contain one observation from each system, and develop Procedure AVC (All Vector Comparisons) to achieve a higher PCS than with Procedure BEM. For specific small-sample cases and via a large-sample approximation we show that the PCS with Procedure AVC exceeds the PCS with Procedure BEM. We also show that with Procedure AVC we achieve a given PCS with a smaller $v$ than the $v^*$ required with Procedure BEM. © 1998 John Wiley & Sons, Inc. Naval Research Logistics 45: 459–482, 1998

**Keywords:** multinomial; ranking and selection; simulation

## 1. INTRODUCTION

Suppose we have $k \geq 2$ independent populations, denoted $\pi_1, \pi_2, \ldots, \pi_k$. In a simulation context each population is a simulated system. We consider the problem of selecting the best of the $k$ systems based on simulated results for all of the systems.

Let $X_{ji}$ represent the $i$th replication from system $j$ of some performance measure. Each system ($\pi_j, j = 1, 2, \ldots, k$) has an unknown constant probability ($p_j, j = 1, 2, \ldots, k$) of having the largest value of the performance measure. We define the best system as the

*Correspondence to:* J. O. Miller
Contract grant sponsor: National Science Foundation; contract grant number: DMI-9622065

 CCC 0894-069X/98/050459-24

system most likely to have the largest performance measure in any comparison across all systems. Such a comparison corresponds to a multinomial trial, where one and only one system can win in any given trial. Our objective is to find the system that is most likely to be the best performer in a single trial among the systems, as opposed to identifying the best average performer over the long run, with a minimum amount of data. This is known as the multinomial selection problem (MSP).

Bechhofer and Sobel [3] introduced the use of multinomial selection procedures to find the system most likely to produce the largest observation on a given vector-trial. Goldsman [5] first suggested the more general use of this type of procedure to find the system most likely to produce the "most desirable" observation on a given vector-trial, where "most desirable" can be almost any criterion of goodness. A classical solution procedure for the MSP, Procedure BEM (Bechhofer, Elmaghraby, and Morse [1]), prescribes a minimum number $v^*$ of independent vector replications across all systems so that the probability of correctly selecting the true best system (PCS) meets or exceeds a prespecified probability. Assuming that larger is better, BEM selects as best the system having the largest value of the performance measure in more replications than any other.

MSP applications include selecting the best of a set of tactical or strategic military actions; finding the design that performs best in a one-time catastrophic event, such as an earthquake; selecting the production schedule most likely to result in completing all jobs on time; selecting the investment portfolio most likely to provide the largest return; or selecting the computer system with the highest probability of completing a series of tasks without failure. Each of these applications involves the comparison of quantitative measures of performance among competing systems as opposed to comparing qualitative measures. For the type of MSP considered in this study, we require a quantitative measure of system performance such that each system in each trial can be compared with the performance of other systems across any or all of the remaining trials.

Let $\mathbf{X}_i = (X_{1i}, X_{2i}, \ldots, X_{ki})$ represent the $i$th replication across all $k$ systems. Let $Y_{ji} = 1$ if $X_{ji} > X_{\ell i}$, for $\ell = 1, 2, \ldots, k$, but $\ell \neq j$; and let $Y_{ji} = 0$ otherwise. In other words, $Y_{ji} = 1$ if $X_{ji}$ is the largest observation in $\mathbf{X}_i$. In case of a tie for the largest value, we randomly select one of the tied populations as the best.

Suppose that there are $v$ independent replications across all systems, and let $Y_j = \sum_{i=1}^{v} Y_{ji}$ represent the number of times system $j$ wins out of these $v$ replications. Let $p_j = \Pr\{X_{ji} > X_{\ell i}, \forall \ell \neq j\}$ where $0 < p_j < 1$ and $\sum_{j=1}^{k} p_j = 1$. Then $\sum_{j=1}^{k} Y_j = v$ and the $k$-variate discrete random variable $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_k)$ follows a multinomial distribution with success probabilities $\mathbf{p} = (p_1, p_2, \ldots, p_k)$. Therefore, the probability mass function for $\mathbf{Y}$ with parameters $v$ and $\mathbf{p}$ is

$$\Pr\{Y_1 = y_1, Y_2 = y_2, \ldots, Y_k = y_k\} = \frac{v!}{\prod_{j=1}^{k} y_j!} \prod_{j=1}^{k} p_j^{y_j}.$$

Due to convention and convenience when comparing simulated system responses, the responses are typically grouped by replication, corresponding to a trial in a physical experiment. Grouping independent system responses in this fashion is arbitrary, and, since our simulated responses are quantitative, we can compare any observation from one system with any observation from each of the remaining systems. This means that a single observation from system 1 can be grouped in a vector comparison with any one of the $v$ observations from system 2, and with any one of the $v$ observations from system 3, and so on, up to and

including any one of the $v$ observations from system $k$. A total of $v^k$ vector comparisons (trials) can be formed with $v$ independent observations from the $k$ systems. We incorporate this setup in a new MSP procedure, which we call AVC, for All Vector Comparisons. By performing only the $v$ vector comparisons where the observations for each system are grouped by replication, as is done with BEM, we disregard the information available from the remaining $v^k - v$ comparisons.

Our results suggest a number of advantages of AVC over BEM. For specific small-sample cases, we show that AVC has a larger PCS than BEM for a fixed $v$. We show this analytically for small values of $v$ and $k$, and also present simulation results for up to $k = 10$ systems and $v = 50$ vector replications. Looking at these results from a slightly different perspective, we also demonstrate achievement of a desired PCS with a smaller value of $v$ when using AVC as compared to BEM. The first perspective emphasizes a more efficient use of the available data to increase PCS. The second view points towards a more efficient way to design a simulation experiment using the smallest value of $v$ required to achieve a desired PCS.

Unlike BEM, the PCS for AVC depends on the distributions of the simulation outputs, not just on $p_1, p_2, \ldots, p_k$. However, we also show that the dependence is weak. This fact, along with the difficulty of analytically evaluating the PCS of AVC for even small $k$ and $v$, leads us to a large-sample approximation (LSA) for the PCS using AVC. As $v \to \infty$, any distributional differences in PCS with AVC disappear. Therefore, our LSA is distribution-independent, and we use this fact to estimate the PCS with AVC. Our LSA demonstrates that asymptotically the PCS with AVC is larger than the PCS with BEM. Additionally, this LSA shows that AVC can provide better discrimination between the systems at the same level of confidence and with the same data.

The paper is organized as follows: We first provide a brief review of the MSP and the classical approach to solving it. Then we describe our new procedure, AVC, and present analytical results covering a variety of specific population distributions for the performance measures. Our LSA is then presented by recasting PCS in terms of a point estimation problem for the multinomial success probabilities, $p_j$, $j = 1, 2, \ldots, k$. Empirical results follow for specific distributions and include simulations designed to test the robustness of our LSA.

## 2. BACKGROUND

Bechhofer, Elmaghraby, and Morse [1] describe a single-stage procedure for selecting the multinomial event (population or system) which has the largest success probability. BEM requires the specification of $P^*$ (where $1/k < P^* < 1$), a minimum probability of correctly identifying the population with the largest success probability (i.e., the best population), and $\theta^*$ (where $1 < \theta^* < \infty$), the minimum ratio of the largest success probability to the second largest success probability that we want to be able to detect. The procedure, as adapted to simulation, consists of the following steps:

PROCEDURE 1 BEM:

1. For given $k$ and $\theta^*$, find the minimum value of $v$, denoted $v^*$, that guarantees that the PCS is at least $P^*$.
2. Generate $v^*$ independent replications for each population.

3. Compute $Y_j = \sum_{i=1}^{v^*} Y_{ji}$, for $j = 1, 2, \ldots, k$.
4. Let $Y_{(1)} \leq Y_{(2)} \leq \cdots \leq Y_{(k)}$ be the ranked sample counts from step 3. Select the population associated with the largest count, $Y_{(k)}$, as the best population. In case of a tie for the largest count, randomly select one of the tied populations as the best.

To determine the appropriate $v^*$ in step 1, let $p_{[1]} \leq p_{[2]} \leq \cdots \leq p_{[k]}$ denote the ranked success probabilities for the $k$ populations. Since only values of the ratio $\theta = p_{[k]}/p_{[k-1]}$ greater than or equal to $\theta^*$ are of interest, we are indifferent between the best and the next-best population for values of $\theta < \theta^*$. A procedure of this type is referred to as an *indifference-zone* procedure. Select $v^*$ as the minimum number of independent vector observations required to achieve a PCS greater than or equal to $P^*$ whenever $\theta \geq \theta^*$.

We define the least favorable configuration (LFC) of $[\mathbf{p}] = (p_{[1]}, p_{[2]}, \ldots, p_{[k]})$ as the configuration where PCS is a minimum over all configurations with $\theta \geq \theta^*$ [4]. If we obtain a PCS $\geq P^*$ with our selected $v^*$ under the LFC, then a PCS of at least $P^*$ can be guaranteed for *any* configuration of $[\mathbf{p}]$ with $\theta \geq \theta^*$. Keston and Morse [7] prove that the LFC for BEM is given by

$$p_{[1]} = p_{[2]} = \cdots = p_{[k-1]} = \frac{1}{\theta^* + k - 1},$$

$$p_{[k]} = \frac{\theta^*}{\theta^* + k - 1}. \tag{1}$$

Although we only need to consider the LFC for designing sampling plans, the PCS can be calculated for any $[\mathbf{p}]$ with $p_{[k]} > p_{[k-1]}$ as follows.

Let $\pi_{[j]}$ be the population associated with $p_{[j]}$ and let $y_{[j]}$ represent the number of wins for $\pi_{[j]}$. Thus, the subscripts for the populations and the associated numbers of wins are based on the ranking of the $p_j$s. We refer to the PCS using BEM for a fixed $k$ and $v$ as $\text{PCS}^{\text{bem}}$. For any fixed $k$ and $v$, $\text{PCS}^{\text{bem}}$ can be expressed as

$$\text{PCS}^{\text{bem}}([\mathbf{p}]) = \sum_{\mathbf{y}} \frac{1}{t(\mathbf{y})} \frac{v!}{\prod_{j=1}^{k} y_{[j]}!} \prod_{j=1}^{k} p_{[j]}^{y_{[j]}},$$

where the summation is over all vectors $\mathbf{y} = (y_{[1]}, y_{[2]}, \ldots, y_{[k]})$ such that $\sum_{j=1}^{k} y_j = v$, $y_{[k]} \geq y_{[j]}$ $(j = 1, 2, \ldots, k-1)$, and $t(\mathbf{y})$ is a function of $y_{[1]}, y_{[2]}, \ldots, y_{[k]}$ representing the number of populations tied for the most wins [1].

## 3. ALL VECTOR COMPARISONS (AVC)

We propose a method to provide a PCS greater than or equal to $\text{PCS}^{\text{bem}}$ (in at least some cases) using the same replications $\mathbf{X}_i$, $i = 1, 2, \ldots, v^*$. We use the BEM parameters $k$, $P^*$, and $\theta^*$, and we execute the first step of BEM to find a value of $v^*$. However, rather than comparing the $i$th replication for each system with the $i$th replications of the other systems, consider instead a total of $(v^*)^k$ pseudoreplications formed by associating each $X_{ji}$ $(j = 1, 2, \ldots, k;\ i = 1, 2, \ldots, v^*)$, with all possible combinations of the remaining $X_{\ell h}$ $(\ell = 1, 2, \ldots, k;\ \ell \neq j;\ h = 1, 2, \ldots, v^*)$. Each such pseudoreplication contains one

observation from each population. Note that the $(v^*)^k$ pseudoreplications include the $v^*$ independent replications from which the pseudoreplications are formed.

Define

$$Z_j = \sum_{a_1=1}^{v^*} \sum_{a_2=1}^{v^*} \cdots \sum_{a_k=1}^{v^*} \prod_{\ell \neq 1; \ell \neq j}^{k} \phi(X_{ja_j} - X_{\ell a_\ell}) \tag{2}$$

for $j = 1, 2, \ldots, k$ with

$$\phi(c) = \begin{cases} 1, & c > 0, \\ 0, & c < 0, \\ \text{randomly assign 0 or 1}, & c = 0. \end{cases}$$

Thus, $Z_j$ represents the number of times out of $(v^*)^k$ pseudoreplications that population $\pi_j$ wins (ties broken randomly) and $\Sigma_{j=1}^{k} Z_j = (v^*)^k$.

Our new procedure consists of the following steps:

PROCEDURE 2 AVC:

1. Given values for $k$, $P^*$, and $\theta^*$, use step 1 of Procedure BEM to determine a value for $v^*$.
2. Generate $v^*$ independent replications for each population and construct the additional $(v^*)^k - v^*$ pseudoreplications possible with one value from each of the populations.
3. Compute $Z_j$ using Eq. (2).
4. Let $Z_{(1)} \leq Z_{(2)} \leq \cdots \leq Z_{(k)}$ be the ranked sample counts from step 3. Select the population associated with the largest count, $Z_{(k)}$, as the best population. In case of a tie for the largest count, randomly select one of the tied populations as the best.

Suppose we modify step 1 of Procedure AVC to use the minimum $v$ where $\text{PCS}^{\text{avc}} \geq P^*$. We demonstrate later that a smaller number of replications are required with AVC relative to BEM to achieve $P^*$. We provide such values of $v$ in this paper.

$\text{PCS}^{\text{avc}}$ can be expressed as

$$\text{PCS}^{\text{avc}}([\mathbf{p}]) = \sum_{\mathbf{z}} \frac{1}{t(\mathbf{z})} \Pr\{Z_{[1]} = z_{[1]}, \ldots, Z_{[k]} = z_{[k]}\},$$

where the summation is over all vectors $\mathbf{z} = (z_{[1]}, z_{[2]}, \ldots, z_{[k]})$ such that $\Sigma_{j=1}^{k} z_j = v^k$, $z_{[k]} \geq z_{[j]}$, $j = 1, 2, \ldots, k-1$, and $t(\mathbf{z})$ is a function of $z_{[1]}, z_{[2]}, \ldots, z_{[k]}$ representing the number of populations tied for the most wins. Each $z_{[j]}$ represents the number of times that $\pi_{[j]}$ wins out of the $v^k$ pseudoreplications. Unfortunately, $\mathbf{Z}$ does not follow a multinomial distribution, so that we must refer to the distributions of the original observations, $X_{ji}$, to calculate $\text{PCS}^{\text{avc}}$. Analytical and simulation results using a number of different population distributions show that $\text{PCS}^{\text{avc}}$ depends weakly on the underlying distributions of the $X_{ji}$.

## 4. ANALYTICAL RESULTS

The following discussion illustrates a number of important properties of the AVC method. First, we demonstrate the improvement possible with AVC for specific cases. We also show a weak dependence in the AVC results on the underlying population distributions for the $X_{ji}$. Lastly, we demonstrate the difficulty in obtaining analytical results for even a small number of populations and observations, and thus provide motivation for our large sample approximation of $\mathrm{PCS}^{\mathrm{avc}}$ which is distribution independent.

Initially, we restrict our attention to continuous distributions for the $X_{ji}$, which eliminates the possibility of ties among the observations. We let $\pi_{[k]}$ be the best population and assume all the remaining populations, $\pi_{[1]}, \pi_{[2]}, \ldots, \pi_{[k-1]}$, are identically distributed. This setup gives us the LFC for BEM when $p_{[k]}/p_{[k-1]} = \theta^*$. We also assume that all population distributions belong to the same parametric family. We calculate $\mathrm{PCS}^{\mathrm{avc}}$ by conditioning on the joint density of all the order statistics for the $v$ independent replications from $\pi_{[k]}$.

Consider a set of $v$ vector replications across all populations. Combine all the observations from all populations and rank them from smallest to largest. Refer to each observation by its rank and consider permutations of these ranks. For any such permutation we can determine the value of $Z_{[k]}$ and calculate the probability of obtaining that arrangement of ranks. We refer to such an arrangement as a *rank order.* Recall that $Z_{[k]}$ represents the number of times the best population, $\pi_{[k]}$, wins out of the $v^k$ pseudoreplications. For illustrative purposes, let $X$ represent an observation from $\pi_{[k]}$ and let $O$ represent an observation from any of the remaining inferior populations.

As an example, suppose $k = 3$, $v = 2$. Then

$$\Pr\{Z_{[3]} = 8\} = \Pr\{O_{(1)} < O_{(2)} < O_{(3)} < O_{(4)} < X_{(1)} < X_{(2)}\}, \tag{3}$$

$$\Pr\{Z_{[3]} = 6\} = 4 \Pr\{O_{(1)} < O_{(2)} < O_{(3)} < X_{(1)} < O_{(4)} < X_{(2)}\}. \tag{4}$$

These probability expressions, (3) and (4), do not identify which observation from which inferior population each $O$ represents. However, in evaluating these expressions we must consider all permutations of the $O$s with respect to observation number and population, and account for each unique combination of adjacent $O$'s. For probability statement (3), there is only one combination of adjacent $O$'s from the 4! permutations of the $O$'s that is less than both $X$s. In the rank order for probability statement (4), since any one of the $O$'s can be associated with $O_{(4)}$, we have four distinct combinations (in terms of which set of $O$'s are adjacent) that result in this one rank order. This is why the coefficient "4" appears on the right-hand side of Eq. (4). In general this coefficient is $\binom{n}{r}$, where $n = v(k-1)$ is the total number of observations from the inferior populations and $r$ is the largest number of these observations that are adjacent. Similar arguments can be used to derive expressions for possible values of $Z_{[k]}$ for integers $k, v \geq 2$. For this example, there is only a single rank order that results in each value of $Z_{[k]}$. As $k$ or $v$ get even moderately large, there will be many rank orders that result in the same value for $Z_{[k]}$. In addition, the rank order of all the data may not uniquely determine the value of $Z_{[k]}$. Therefore, the calculation of the probability of each value of $Z_{[k]}$ becomes extremely tedious with increasing $k$ or $v$.

Restricting our attention to $k = 2$ populations, it is interesting to note that the vector comparisons with AVC are analogous to the comparisons that form the Wilcoxon rank-sum statistic [11]. Let $W$ equal the sum of the ranks of the observations from the best

**Table 1.** Analytical expressions for improvement of AVC over BEM.

| $k$ | $v$ | Distribution | $\Delta\text{PCS} = \text{PCS}^{\text{avc}} - \text{PCS}^{\text{bem}}$ |
|---|---|---|---|
| 2 | 2 | Exponential[a] | $\dfrac{\lambda\mu(\mu - \lambda)}{(2\lambda + \mu)(\lambda + 2\mu)(\lambda + \mu)} > 0$ |
| 2 | 2 | Continuous uniform[b] | $\dfrac{A(B - A)}{6B} > 0$ |
| 2 | 2 | Bernoulli[c] | $\frac{1}{4}(1 + 2p_x p_o - p_x - p_o)(p_x - p_o) > 0$ |
| 2 | 3 | Exponential | $\dfrac{\lambda^2\mu^2(16\lambda^2 + 37\lambda\mu + 16\mu^2)(\mu - \lambda)}{(3\lambda + 2\mu)(2\lambda + 3\mu)(2\lambda + \mu)(\lambda + 2\mu)(\lambda + \mu)^3} > 0$ |
| 2 | 3 | Continuous uniform | $\dfrac{3A^2(B - A)}{20B^3} > 0$ |
| 3 | 2 | Exponential | $\dfrac{2\lambda\mu^2(4\lambda^3 + 41\lambda^2\mu + 84\lambda\mu^2 + 41\mu^3)(\mu - \lambda)}{(2\lambda + 3\mu)(2\lambda + \mu)(\lambda + 4\mu)(\lambda + 3\mu)(\lambda + 2\mu)(\lambda + \mu)^2} > 0$ |
| 3 | 2 | Continuous uniform | $\dfrac{7A(B - A)}{15B^2} > 0$ |

[a] $X \sim \exp(\lambda)$, $O \sim \exp(\mu)$ with $0 < \lambda < \mu$.
[b] $X \sim U(0, B)$, $O \sim U(0, A)$ with $0 < A < B$.
[c] $X \sim \text{Ber}(p_x)$, $O \sim \text{Ber}(p_o)$ with $p_x > p_o$.

population. Then $W$ is the Wilcoxon rank-sum statistic, and our $Z_{[2]}$ is the Mann–Whitney U-statistic. Therefore, $W$ can be expressed as a function of our $Z_{[2]}$ as

$$W = Z_{[2]} + \frac{v}{2}(v + 1).$$

In terms of $W$, AVC always makes a correct selection for $W > \text{E}[W]$ (incorrect selection for $W < \text{E}[W]$), where $\text{E}[W]$ is the expected value of $W$ under the assumption that the two populations are identical in distribution. For our discussion, $\text{E}[W] = (v/2)(2v + 1)$, and $W > \text{E}[W]$ is equivalent to $Z_{[2]} > v^2/2$ (i.e., the best population wins in more than half of the pseudoreplications).

If we specify a particular distribution family for our populations, then we can derive formulas to compare $\text{PCS}^{\text{avc}}$ with $\text{PCS}^{\text{bem}}$ for very small $k$ and $v$. Table 1 presents results for exponential, continuous uniform, and Bernoulli distributions.

Each expression in the $\Delta$PCS column of Table 1 includes a positive term involving the difference of the respective parameters: $(\mu - \lambda)$, $(B - A)$, or $(p_x - p_o)$. These terms all illustrate an improvement in PCS with AVC when $X$ is the best population. When substituting values for the parameters of the distributions shown to achieve a common value of $\theta$, we notice a weak distributional dependence in the $\text{PCS}^{\text{avc}}$ values. The magnitudes of these differences are examined in detail in [10] and are also discussed in Section 6. The weak distributional dependence of $\text{PCS}^{\text{avc}}$, along with the difficulty of computing $\text{PCS}^{\text{avc}}$ for small $k$ and $v$, motivates the following large-sample approximation (LSA).

## 5. LARGE-SAMPLE APPROXIMATION

The results presented so far for small $k$ and $v$ show that $\text{PCS}^{\text{avc}}$ is weakly distribution dependent. By redefining $\text{PCS}^{\text{bem}}$ and $\text{PCS}^{\text{avc}}$ in terms of point estimators for each of the

individual system success probabilities, we arrive at distribution-free results as the sample size goes to infinity.

### 5.1. Preliminaries

Using our previous notation, we have

$$p_j = \Pr\{X_{ji} > X_{\ell i},\ \forall \ell \neq j\}.$$

Let the distribution of $X_{ji}$ depend upon the sample size, $X_{ji} \sim F_j^{(v)}$. We construct the $F_j^{(v)}$ such that the $F_j^{(v)}$ converge to a common distribution, $F$, for all $j$ as $v$ approaches infinity, but for finite $v$

$$\Pr\{X_{ji} > X_{\ell i},\ \forall \ell \neq j \,|\, \text{sample size } v\} = p_j(v) = \begin{cases} \dfrac{1}{k} + \dfrac{(k-1)\delta}{\sqrt{v}}, & j = 1, \\ \dfrac{1}{k} - \dfrac{\delta}{\sqrt{v}}, & j \neq 1. \end{cases} \tag{5}$$

We can choose $\delta > 0$ so that $(p_1(v), p_2(v), \ldots, p_k(v))$ is a LFC for any "reasonable" finite value of $v$; the configuration (5) is, for all practical purposes, a completely general LFC.

Under (5) Population 1 is the best. Define

$Y_j(v)$ = number of wins for system $j$ under BEM with sample size $v$,

$Z_j(v)$ = number of wins for system $j$ under AVC with sample size $v$,

which gives us point estimators

$$\hat{p}_j(v) = \frac{Y_j(v)}{v},$$

$$\bar{p}_j(v) = \frac{Z_j(v)}{v^k}.$$

Thus, our BEM estimators are denoted by $\hat{p}_j$ and our AVC estimators by $\bar{p}_j$. Notice that (ignoring the asymptotically vanishing probability of a tie) $\text{PCS}^{\text{bem}} = \Pr\{\hat{p}_1 > \hat{p}_j,\ \forall j \neq 1\}$ and $\text{PCS}^{\text{avc}} = \Pr\{\bar{p}_1 > \bar{p}_j,\ \forall j \neq 1\}$. Our approach is based on the fact that standardized versions of $\hat{p}$ and $\bar{p}$ are asymptotically multivariate normal (MVN). However, when the distributions are fixed with respect to $v$, then as the sample size increases, both $\text{PCS}^{\text{bem}}$ and $\text{PCS}^{\text{avc}}$ approach 1, masking the differences between the two procedures. To eliminate this effect and isolate the improvement with AVC, we simultaneously let the ratio, $p_1(v)/p_j(v)$ $(j \neq 1)$, approach 1 at the canonical rate of $1/\sqrt{v}$, as shown in (5).

### 5.2. BEM Estimators

Consider the asymptotic behavior of $PCS^{bem}$ as the number of vectors, $v$, goes to infinity. Our approach is structured around a result of Lehmann [9] which we state below as a lemma.

LEMMA 1 (ASYMPTOTIC DISTRIBUTION OF STANDARDIZED BEM ESTIMATORS): Let $\mathbf{Y}(v) = (Y_1(v), Y_2(v), \ldots, Y_k(v))$ be distributed as a multinomial random variable with parameters $v$ and $\mathbf{p}(v) = (p_1(v), p_2(v), \ldots, p_k(v))$, with $p_j(v)$ defined as in Eq. (5). Let

$$A_j(v) = \frac{Y_j(v) - v/k}{\sqrt{v}}, \qquad j = 1, 2, \ldots, k.$$

Then as $v \to \infty$

$$\begin{pmatrix} A_2(v) \\ \vdots \\ A_k(v) \end{pmatrix} \overset{D}{\Rightarrow} \text{MVN}\left[ \begin{pmatrix} -\delta \\ \vdots \\ -\delta \end{pmatrix}, \begin{pmatrix} \frac{1}{k}(1 - \frac{1}{k}) & -(\frac{1}{k^2}) & \cdots & -(\frac{1}{k^2}) \\ -(\frac{1}{k^2}) & \frac{1}{k}(1 - \frac{1}{k}) & \cdots & -(\frac{1}{k^2}) \\ \vdots & \vdots & \ddots & \vdots \\ -(\frac{1}{k^2}) & -(\frac{1}{k^2}) & \cdots & \frac{1}{k}(1 - \frac{1}{k}) \end{pmatrix} \right].$$

PROOF: See Lehmann [9].

Lemma 1 is critical to proving the following theorem which we later use to equate asymptotic $PCS^{bem}$ with asymptotic $PCS^{avc}$.

THEOREM 1 (ASYMPTOTIC $PCS^{bem}$): Let $\mathbf{Y}(v) = (Y_1(v), Y_2(v), \ldots, Y_k(v))$ be distributed as a multinomial random variable with parameters $v$ and $\mathbf{p}(v) = (p_1(v), p_2(v), \ldots, p_k(v))$, with $p_j(v)$ defined as in Eq. (5). Then

$$\overrightarrow{PCS}^{bem} \equiv \lim_{v \to \infty} \Pr\{Y_1(v) > Y_j(v)\} = \Pr\left\{ \max_{\ell = 2, \ldots, k} Q_\ell < \frac{k\delta}{\sqrt{2/k}} \right\}, \tag{6}$$

where

$$\begin{pmatrix} Q_2 \\ \vdots \\ Q_k \end{pmatrix} \sim \text{MVN}\left[ \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 1/2 & \cdots & 1/2 \\ 1/2 & 1 & \cdots & 1/2 \\ \vdots & \vdots & \ddots & \vdots \\ 1/2 & 1/2 & \cdots & 1 \end{pmatrix} \right]. \tag{7}$$

PROOF: See the Appendix.

### 5.3. AVC Estimators

Consider the asymptotic behavior of $\text{PCS}^{\text{avc}}$ as the number of vectors, $v$, goes to infinity. Our approach is structured around a result of Lehmann [8] and Randles and Wolfe [11]. We state this result below as Lemma 2, where some of the notation has been simplified to match the context of our problem.

We first show that our AVC estimator is a $k$-sample U-statistic for the parameter $\mathbf{p} = (p_1, p_2, \ldots, p_k)$, the system success probabilities. From [11], we say $\mathbf{p}$ is estimable of degree $(1, 1, \ldots, 1)$ for distributions $(F_1, F_2, \ldots, F_k)$ of $(X_{1i}, X_{2i}, \ldots, X_{ki})$ in some family of distributions $\mathcal{F}$, if $(X_{1i}, X_{2i}, \ldots, X_{ki})$ is the smallest sample size (one observation from each system) for which there exists an unbiased estimator of $\mathbf{p}$ for every $(F_1, F_2, \ldots, F_k) \in \mathcal{F}$. Formally stated

$$E_{(F_1,\ldots,F_k)}[h^{(j)}(X_{11}, \ldots, X_{k1})] = p_j$$

for $j = 1, 2, \ldots, k$, for a $k$-sample symmetric kernel $h^{(j)}(\cdot)$. In our case this kernel is

$$h^{(j)} = \prod_{\ell=1;\ell\neq j}^{k} \phi(X_{ji} - X_{li}). \tag{8}$$

Thus, $h^{(j)} = 1$ if the observation from the $j$th system is the largest in any vector comparison across all systems. A $k$-sample U-statistic is the average value of such a kernel over all vectors of observations with one observation from each system, which is precisely (2) divided by $v^k$. Therefore, $\bar{p}_j$, $j = 1, 2, \ldots, k$, are each $k$-sample U-statistics.

LEMMA 2 (ASYMPTOTIC DISTRIBUTION OF AVC ESTIMATORS): Let $\bar{p}_1, \bar{p}_2, \ldots, \bar{p}_k$ be $k$-sample U-statistics, with $\bar{p}_j$ corresponding to a parameter $p_j$ of degree $(1, 1, \ldots, 1)$ and symmetric kernel $h^{(j)}(\cdot)$, for $j = 1, 2, \ldots, k$. Let $N = kv$, where $v$ is the sample size from each of $k$ populations. Then the joint limiting distribution of

$$\begin{pmatrix} \sqrt{N}(\bar{p}_1 - p_1) \\ \vdots \\ \sqrt{N}(\bar{p}_k - p_k) \end{pmatrix} \overset{\mathcal{D}}{\Rightarrow} \text{MVN}\left[ \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \Sigma = \|\sigma^{(a,b)}\| \right]$$

as $v \to \infty$, where

$$\sigma^{(a,b)} = \sum_{j=1}^{k} \frac{1}{\lambda_j} \xi_j^{(a,b)} \tag{9}$$

for $\lambda_j = \lim_{N\to\infty} v/N = 1/k$. The quantities $\xi_j^{(a,b)}$ are given by

$$\xi_j^{(a,b)} = \text{Cov}[H_{j1}^{(a)}, H_{j2}^{(b)}] = E[H_{j1}^{(a)} H_{j2}^{(b)}] - p_a p_b,$$

where

$$H_{j1}^{(a)} = h^{(a)}(X_{1\alpha_1}, X_{2\alpha_2}, \ldots, X_{k\alpha_k}),$$

$$H_{j2}^{(b)} = h^{(b)}(X_{1\beta_1}, X_{2\beta_2}, \ldots, X_{k\beta_k}),$$

and the sets $(\alpha_1, \alpha_2, \ldots, \alpha_j, \ldots, \alpha_k)$ and $(\beta_1, \beta_2, \ldots, \beta_j, \ldots, \beta_k)$ have only the $j$th element in common, the elements in each set represent positive integers, and $a, b \in \{1, \ldots, k\}$. See the Appendix for more details on calculating the covariance terms.

PROOF: See Lehmann [8].

Proceeding in much the same manner as we did in moving from Lemma 1 to Theorem 1, we can define $\text{PCS}^{\text{avc}}$ as a probability statement involving a function of $k$ and the maximum of $(k-1)$ random variables created by subtracting one of the standardized $\bar{p}_j$s from each of the remaining standardized $\bar{p}_i s$, $i \neq j$. These random variables are the $Q_2$, $Q_3, \ldots, Q_k$ in the following theorem.

THEOREM 2 (ASYMPTOTIC $\text{PCS}^{\text{avc}}$): Let $\bar{p}_1, \bar{p}_2, \ldots, \bar{p}_k$ be $k$-sample U-statistics, with $\bar{p}_j$ corresponding to a parameter $p_j$ of degree $(1, 1, \ldots, 1)$ and symmetric kernel $h^{(j)}(\cdot)$, for $j = 1, 2, \ldots, k$. Let $N = kv$ and $\mathbf{p}(v) = (p_1(v), p_2(v), \ldots, p_k(v))$, with $p_j(v)$ defined as in Eq. (5). Then under our model with $F_j^{(v)} \to F$ as $v \to \infty$

$$\vec{\text{PCS}}^{\text{avc}} \equiv \lim_{v \to \infty} \Pr\{\bar{p}_1(v) > \bar{p}_j(v)\} = \Pr\left\{ \max_{\ell=2,\ldots,k} Q_\ell < \frac{k\delta}{\sqrt{2/(2k-1)}} \right\}, \tag{10}$$

where $(Q_2, \ldots, Q_k)$ has the same MVN distribution as that given by (7).

PROOF: See the Appendix.

### 5.4. Combining BEM and AVC Results

Recall $\vec{\text{PCS}}^{\text{avc}}$ and $\vec{\text{PCS}}^{\text{bem}}$ represent the asymptotic PCS for AVC and BEM, respectively, under the setup described in Subsections 5.3 and 5.2. Combining the results from Eqs. (6) and (10), we have

$$\vec{\text{PCS}}^{\text{avc}} \geq \vec{\text{PCS}}^{\text{bem}}$$

since

$$\frac{k\delta^{\text{avc}}}{\sqrt{2/(2k-1)}} \geq \frac{k\delta^{\text{bem}}}{\sqrt{2/k}}$$

for $k \geq 2$ with $\delta^{\text{avc}} = \delta^{\text{bem}}$. Thus we have equal asymptotic PCS, that is,

$$\vec{\text{PCS}}^{\text{avc}} = \vec{\text{PCS}}^{\text{bem}} \tag{11}$$

if and only if

$$\frac{k\delta^{\text{avc}}}{\sqrt{2/(2k-1)}} = \frac{k\delta^{\text{bem}}}{\sqrt{2/k}}.$$

Solving for $\delta^{\text{avc}}$ we have

$$\delta^{\text{avc}} = \delta^{\text{bem}}\sqrt{k/(2k-1)}. \tag{12}$$

We use $\delta^{\text{bem}}$ and $\theta^{\text{bem}}$ ($\delta^{\text{avc}}$ and $\theta^{\text{avc}}$) to represent the difference $p_{[k]} - p_{[k-1]}$ or the ratio $p_{[k]}/p_{[k-1]}$, respectively, associated with BEM (AVC) calculations. The relationship between $\delta^{\text{avc}}$ and $\delta^{\text{bem}}$ is used to define a relationship between $\theta^{\text{avc}}$ and $\theta^{\text{bem}}$ which also guarantees (11). This allows us to use BEM calculations to approximate AVC results.

Consider the following illustration. We have a problem with a specified $(\theta^*, k, P^*)$, where we want to find the minimum sample size required with AVC. To approximate the required sample size, $v$, for AVC we will set

$$v^{\text{avc}}(\theta^{\text{avc}} = \theta^*, k, P^*) = v^{\text{bem}}(\theta^{\text{bem}}, k, P^*), \tag{13}$$

where $v^{\text{avc}}$ and $v^{\text{bem}}$ denote the $v$ required for AVC or BEM, respectively, and $\theta^{\text{bem}}$ is such that (11) holds.

We can make this approximation since

$$\text{PCS}^{\text{bem}}(v^{\text{bem}}(\theta^{\text{bem}}, k, P^*)) \geq P^*,$$

and from (11) we have

$$\vec{\text{PCS}}^{\text{avc}}(\theta^{\text{avc}}, k) = \vec{\text{PCS}}^{\text{bem}}(\theta^{\text{bem}}, k),$$

which leads us to

$$\text{PCS}^{\text{avc}}(v^{\text{bem}}(\theta^{\text{bem}}, k, P^*)) \approx P^*.$$

Using Algorithm 2.1 of [10], our LSA for $\theta^{\text{avc}}$ defines the asymptotically equivalent $\theta^{\text{bem}}$ to be

$$\theta^{\text{bem}} = \frac{1 + (k-1)\left(\dfrac{\theta^{\text{avc}} - 1}{\theta^{\text{avc}} + k - 1}\right)\sqrt{\dfrac{2k-1}{k}}}{1 - \left(\dfrac{\theta^{\text{avc}} - 1}{\theta^{\text{avc}} + k - 1}\right)\sqrt{\dfrac{2k-1}{k}}}. \tag{14}$$

It will always be the case that $\theta^{\text{bem}} > \theta^{\text{avc}}$ when $\theta^{\text{avc}} = \theta^*$. Thus approximation (14) allows use of standard BEM calculations or tables for estimating the number of vectors required

**Table 2.** Asymptotically equivalent $\theta$ values using LSA.

| $k$ | $\theta^*$ | $\theta^{avc}$ ($\theta^{bem} = \theta^*$) | $\theta^{bem}$ ($\theta^{avc} = \theta^*$) |
|---|---|---|---|
| 2 | 1.2 | 1.1604 | 1.2506 |
| | 2.0 | 1.7479 | 2.3798 |
| 3 | 1.2 | 1.1526 | 1.2633 |
| | 2.0 | 1.7205 | 2.4297 |
| 4 | 1.2 | 1.1494 | 1.2689 |
| | 2.0 | 1.7124 | 2.4390 |
| 5 | 1.2 | 1.1476 | 1.2720 |
| | 2.0 | 1.7092 | 2.4400 |
| 10 | 1.2 | 1.1443 | 1.2778 |
| | 2.0 | 1.7061 | 2.4326 |

to achieve $P^*$ using AVC. We illustrate the use of this approximation in constructing Table 9.

We can modify the approximation in (14) to estimate an equivalent $\theta^{avc}$ when $\theta^{bem} = \theta^*$. It will always be the case that $\theta^{avc} < \theta^{bem}$ when $\theta^{bem} = \theta^*$. This form of the approximation has little practical use. However, it does reflect another benefit of AVC in terms of a smaller $\theta$, indicating the ability of AVC to discriminate smaller differences between the best and the next best system with the same value of $v$ as BEM for $\theta^*$. This advantage becomes important in a case where we need to detect as small a difference as possible with a fixed number of vector replications.

We provide conversions using both of these approximations for some common values of $\theta^*$ in Table 2 and present simulation results testing the robustness of the approximations in Section 7.

## 6. EMPIRICAL RESULTS

In order to allow easy comparison with available BEM results, we select population distributions for our simulations that allow us to control the value of $\theta^*$. These distributions are the exponential, continuous uniform, and the Bernoulli presented in Section 4. In addition, to consider a less peaked continuous distribution without the restricted range of the continuous uniform, we look at a set of gamma distributions with a shape parameter of 3.

As in our analytical results, we consider population distributions that belong to the same parametric family. We arbitrarily designate $\pi_1$ as the best population and the remaining populations are identically distributed. Let $X_j$ represent a random observation from $\pi_j$, $j = 1, 2, \ldots, k$. We have

$$\Pr\{\text{Best Population Wins}\} = \Pr\{X_1 > \max(X_2, \ldots, X_k)\}.$$

We then define

$$\theta = \frac{\Pr\{X_1 > \max(X_2, \ldots, X_k)\}}{(1 - \Pr\{X_1 > \max(X_2, \ldots, X_k)\})/(k-1)}.$$

By setting $\theta = \theta^*$, we can then fix one or more parameters for one of the distributions and

**Table 3.** Computed parameter values for given distribution in LFC.

| $k$ | $\theta$ | Exponential[a] | Gamma[b] | Continuous uniform[c] | Bernoulli[d] |
|---|---|---|---|---|---|
| 2 | 1.2 | 1.2000 | 1.1021 | 1.1000 | 0.4091 |
| 2 | 2.0 | 2.0000 | 1.4442 | 1.5000 | 0.1667 |
| 3 | 1.2 | 1.1589 | 1.0855 | 1.0667 | 0.4208 |
| 3 | 2.0 | 1.7808 | 1.3751 | 1.3333 | 0.2192 |
| 4 | 1.2 | 1.1368 | 1.0760 | 1.0500 | 0.4250 |
| 4 | 2.0 | 1.6632 | 1.3340 | 1.2500 | 0.2426 |
| 5 | 1.2 | 1.1227 | 1.0698 | 1.0400 | 0.4264 |
| 5 | 2.0 | 1.5885 | 1.3061 | 1.2000 | 0.2545 |

[a] Values are for $\mu$ with $\lambda = 1$.
[b] Values are for $\pi_1$'s scale parameter, with scale parameter of 1 for all remaining populations and shape parameter of 3 for all populations.
[c] Values are for $B$ with $A = 1$.
[d] Values are for $p_o$ with $p_x = 0.5$.

solve for the remaining parameter to carry out our simulations at a given $\theta^*$. Table 3 lists parameters for $\theta = 1.2$ and $\theta = 2.0$ with $k = 2, 3, 4, 5$ for each of our four distributions.

Our simulation consists of the following steps.

1. Model all systems using the same distribution family, with system 1 arbitrarily the best, and all remaining systems identically distributed such that $\theta = \theta^*$. Initialize $\text{SUM}^{\text{bem}}$ and $\text{SUM}^{\text{avc}}$ to 0 and set $v = 2$.
2. Generate a set of $v$ random vector replications, where each replication contains one observation for each of the $k$ systems.
3. For BEM, group the observations across systems by vector replication and count up the number of wins for each system. These are our $Y_j, j = 1, 2, \ldots, k$.
4. For AVC, form the $v^k$ pseudoreplications from the $v$ vector replications and count the number of wins for each system. These are our $Z_j, j = 1, 2, \ldots, k$.
5. If $Y_1$ (BEM count associated with the best system) is larger than $Y_j, j = 2, 3, \ldots, k$, increase $\text{SUM}^{\text{bem}}$ by 1. If $Y_1$ ties for the largest count with $t$ other systems, $t = 1, 2, \ldots, k - 1$, increase $\text{SUM}^{\text{bem}}$ by $1/(t + 1)$. If $Y_1 < Y_j$, for any $j, j = 2, 3, \ldots, k$, do not increase $\text{SUM}^{\text{bem}}$.
6. If $Z_1$ (AVC count associated with the best system) is larger than $Z_j, j = 2, 3, \ldots, k$, increase $\text{SUM}^{\text{avc}}$ by 1. If $Z_1$ ties for the largest count with $t$ other systems, $t = 1, 2, \ldots, k - 1$, increase $\text{SUM}^{\text{avc}}$ by $1/(t + 1)$. If $Z_1 < Z_j$, for any $j, j = 2, 3, \ldots, k$, do not increase $\text{SUM}^{\text{avc}}$.
7. Repeat steps 2–6 for $M$ macroreplications. Compute $\text{PCS}^{\text{bem}} = \text{SUM}^{\text{bem}}/M$ and $\text{PCS}^{\text{avc}} = \text{SUM}^{\text{avc}}/M$.
8. Increase $v$ and repeat steps 2–7.

Taking parameter values from Table 3, we estimated $\text{PCS}^{\text{bem}}$ and $\text{PCS}^{\text{avc}}$ using the simulation described above for $k = 2, 3, 4, 5$ populations out to $v = 50$ vectors for each of the three continuous distributions at $\theta = 1.2$ and 2.0. Due to limited computer time, Bernoulli distributions were only simulated for $k = 2$ and 3 populations at $\theta = 1.2$. All simulation results are for $M = 100{,}000$ macroreplications using a separate random number stream for each population, but common random numbers across distributions. Standard errors for the PCS values are on the order of 0.0015. More complete results are available in [10].

**Table 4.** PCS results for $k = 2$ populations with $\theta = 1.2$.

| | | $PCS^{avc}$ | | | |
|---|---|---|---|---|---|
| $v$ | $PCS^{bem}$ | Exponential | Uniform | Gamma | Bernoulli |
| 2 | 0.5430 | 0.5532 | 0.5579 | 0.5532 | 0.5565 |
| 6 | 0.5835 | 0.5950 | 0.5990 | 0.5940 | 0.6148 |
| 10 | 0.6099 | 0.6232 | 0.6261 | 0.6238 | 0.6513 |
| 14 | 0.6290 | 0.6463 | 0.6470 | 0.6460 | 0.6781 |
| 18 | 0.6462 | 0.6638 | 0.6653 | 0.6662 | 0.7034 |
| 22 | 0.6624 | 0.6812 | 0.6822 | 0.6813 | 0.7223 |
| 26 | 0.6757 | 0.6950 | 0.6976 | 0.6957 | 0.7407 |
| 30 | 0.6882 | 0.7078 | 0.7104 | 0.7088 | 0.7576 |
| 34 | 0.6987 | 0.7206 | 0.7228 | 0.7200 | 0.7716 |
| 38 | 0.7094 | 0.7324 | 0.7346 | 0.7314 | 0.7853 |
| 42 | 0.7198 | 0.7427 | 0.7460 | 0.7429 | 0.7955 |
| 46 | 0.7279 | 0.7528 | 0.7549 | 0.7529 | 0.8073 |
| 50 | 0.7362 | 0.7614 | 0.7638 | 0.7612 | 0.8179 |

Table 4 lists results for each of our distributions out to $v = 50$ vectors for $k = 2$ populations at $\theta = 1.2$. The $PCS^{bem}$ column is from simulations using exponential populations. The difference in the $PCS^{avc}$ values among the continuous distributions is generally found in the third decimal place. However, we see a more significant difference between the Bernoulli $PCS^{avc}$ and any of the continuous $PCS^{avc}$ values. Figure 1 demonstrates the distributional dependence of $PCS^{avc}$ for exponential and Bernoulli populations. We also notice significant improvement in Table 4 with $PCS^{avc}$ over $PCS^{bem}$ for all of the distributions. Figure 2 illustrates the improvement with $PCS^{avc}$ over $PCS^{bem}$ for $k = 2$–5 exponential populations. Looking closely at Figure 2, the spread between $PCS^{avc}$ and $PCS^{bem}$ appears to be increasing slightly as $k$ increases. This is most readily apparent when comparing the $k = 2$ results to the $k = 3$ results. It is also apparent from both Figures 1 and 2 that the spread between the $PCS^{avc}$ and $PCS^{bem}$ values widens as $v$ increases. However, we know that as $v$ approaches infinity both $PCS^{avc}$ and $PCS^{bem}$ approach 1, so that this spread will eventually disappear.

These results clearly show an improvement in PCS with AVC for all values of $k$ and $v$ considered, and also illustrate the weak dependence of $PCS^{avc}$ on the underlying population distributions.

## 7. ROBUSTNESS OF LSA

To check the accuracy of our LSA, we performed a simulation study. The study covered a set of values for $P^*$ (0.75, 0.90, and 0.95) and $\theta^*$ (1.2 and 2.0) with exponential, continuous uniform and gamma distributions for $k = 2, 3, 4, 5$ populations and Bernoulli distributions for $k = 2, 3$. For the exponential and continuous uniform distributions, results are also presented for $k = 10$ at $\theta^* = 2.0$. Results are not available for all distributions at $\theta^* = 1.2$ for $k > 2$. This is due to the significant amount of computing time required to obtain these results because of the much larger number of vector replications required than for $\theta^* = 2.0$. We have included all results available for $\theta^* = 1.2$ in Tables 5 and 7. Results for $\theta^* = 2.0$ are shown in Tables 6 and 8.

We first consider the approximation in (13) and perform the following steps.

1. Select a $k$ and $\theta^*$ and set $\theta^{avc} = \theta^*$. This indicates that we are interested in calculating AVC results at $\theta^*$.

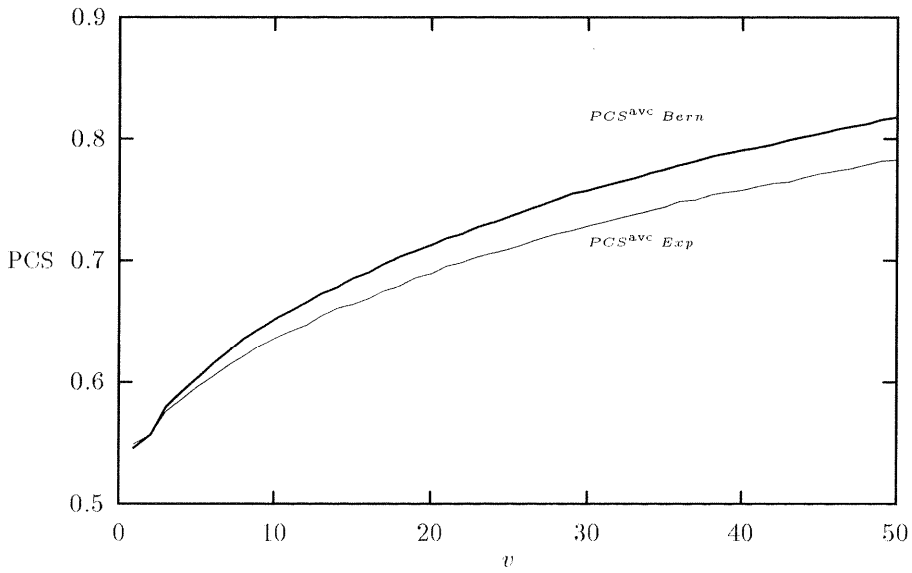


**Figure 1.** $PCS^{avc}$ for exponential and Bernoulli populations: $k = 2$, $\theta = 1.2$.

2. Solve for $\theta^{bem}$ using (14).
3. The calculated value of $\theta^{bem}$ will not be in a standard BEM table. Use FORTRAN code developed by Goldsman [6] to find $v^{bem}(\theta^{bem}, k, P^*)$ for $P^* = 0.75, 0.90, 0.95$. Denote these values as $v^{bem}(.75)$, $v^{bem}(.90)$, and $v^{bem}(.95)$, respectively.

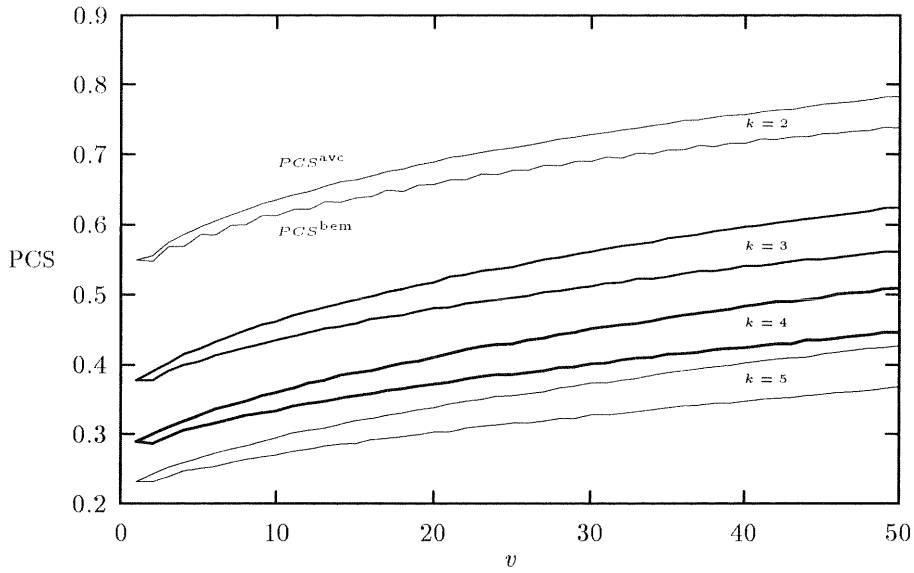


**Figure 2.** PCS for exponential populations, $\theta = 1.2$.

**Table 5.** PCS achieved for $\theta^{avc} = \theta^* = 1.2$ using LSA.

| $k$ | $\theta^{bem}$ | Distribution | $P^* = 0.75$ | $P^* = 0.90$ | $P^* = 0.95$ |
|---|---|---|---|---|---|
| 2 | 1.2506 | Exponential | 0.7502 (0.0014) | 0.9012 (0.0009) | 0.9510 (0.0007) |
| | | Gamma | 0.7481 (0.0014) | 0.8994 (0.0010) | 0.9493 (0.0007) |
| | | Uniform | 0.7499 (0.0014) | 0.9010 (0.0009) | 0.9506 (0.0007) |
| | | Bernoulli | 0.7856 (0.0013) | 0.9318 (0.0008) | 0.9723 (0.0005) |
| 3 | 1.2633 | Exponential | 0.7470 (0.0014) | 0.8997 (0.0009) | 0.9490 (0.0007) |
| | | Gamma | 0.7446 (0.0014) | 0.8994 (0.0010) | 0.9482 (0.0007) |
| | | Uniform | 0.7474 (0.0014) | 0.8989 (0.0010) | 0.9481 (0.0007) |
| 4 | 1.2689 | Exponential | 0.7469 (0.0014) | 0.8964 (0.0010) | 0.9485 (0.0007) |
| | | Gamma | 0.7462 (0.0014) | 0.8984 (0.0010) | 0.9489 (0.0007) |
| | | Uniform | 0.7469 (0.0014) | 0.8948 (0.0010) | 0.9467 (0.0007) |

4. Perform simulation runs to estimate $\text{PCS}^{avc}$ values at $k$ and $\theta^{avc} = \theta^*$ when using the $v^{bem}$ values from step 3. We are looking for the following:

$$\text{PCS}^{avc}(v^{bem}(.75); (\theta^{avc}, k)) \approx 0.75,$$

$$\text{PCS}^{avc}(v^{bem}(.90); (\theta^{avc}, k)) \approx 0.90,$$

$$\text{PCS}^{avc}(v^{bem}(.95); (\theta^{avc}, k)) \approx 0.95.$$

Values estimated in step 4 above are reported in Tables 5 and 6. If our LSA is good, all $\text{PCS}^{avc}$ values in Tables 5 and 6 should be close to the $P^*$ listed at the top of the column in which they appear. The table values include the estimated $\text{PCS}^{avc}$ value and the associated standard error in parentheses. All simulation runs use the models described in Section 6 and are based on $M = 100{,}000$ macroreplications for the three values of $v$ found in step 3 above.

To illustrate how this approximation works for a numerical example, say we want to find

**Table 6.** PCS achieved for $\theta^{avc} = \theta^* = 2.0$ using LSA.

| $k$ | $\theta^{bem}$ | Distribution | $P^* = 0.75$ | $P^* = 0.90$ | $P^* = 0.95$ |
|---|---|---|---|---|---|
| 2 | 2.3798 | Exponential | 0.7614 (0.0013) | 0.8914 (0.0010) | 0.9450 (0.0007) |
| | | Gamma | 0.7608 (0.0013) | 0.8924 (0.0010) | 0.9456 (0.0007) |
| | | Uniform | 0.7625 (0.0013) | 0.8889 (0.0010) | 0.9426 (0.0007) |
| | | Bernoulli | 0.7673 (0.0013) | 0.9380 (0.0008) | 0.9773 (0.0005) |
| 3 | 2.4297 | Exponential | 0.7390 (0.0014) | 0.8909 (0.0010) | 0.9426 (0.0007) |
| | | Gamma | 0.7380 (0.0014) | 0.8910 (0.0010) | 0.9440 (0.0007) |
| | | Uniform | 0.7391 (0.0014) | 0.8873 (0.0010) | 0.9391 (0.0007) |
| | | Bernoulli | 0.7833 (0.0013) | 0.9279 (0.0008) | 0.9687 (0.0007) |
| 4 | 2.4390 | Exponential | 0.7461 (0.0014) | 0.8911 (0.0010) | 0.9464 (0.0007) |
| | | Gamma | 0.7457 (0.0014) | 0.8929 (0.0010) | 0.9391 (0.0008) |
| | | Uniform | 0.7446 (0.0014) | 0.8858 (0.0010) | 0.9409 (0.0007) |
| 5 | 2.4400 | Exponential | 0.7485 (0.0014) | 0.8910 (0.0010) | 0.9443 (0.0007) |
| | | Gamma | 0.7473 (0.0014) | 0.8940 (0.0010) | 0.9454 (0.0007) |
| | | Uniform | 0.7454 (0.0014) | 0.8824 (0.0010) | 0.9372 (0.0007) |

**Table 7.** PCS achieved for $\theta^{bem} = \theta^* = 1.2$ using LSA.

| $k$ | $\theta^{avc}$ | Distribution | $P^* = 0.75$ | $P^* = 0.90$ | $P^* = 0.95$ |
|---|---|---|---|---|---|
| 2 | 1.1604 | Exponential | 0.7486 (0.0014) | 0.8998 (0.0009) | 0.9500 (0.0007) |
| | | Gamma | 0.7493 (0.0014) | 0.9008 (0.0009) | 0.9495 (0.0007) |
| | | Uniform | 0.7505 (0.0013) | 0.9011 (0.0009) | 0.9500 (0.0007) |
| 3 | 1.1526 | Exponential | 0.7514 (0.0014) | 0.8991 (0.0010) | 0.9500 (0.0007) |
| | | Gamma | 0.7502 (0.0014) | 0.8997 (0.0010) | 0.9500 (0.0007) |
| | | Uniform | 0.7510 (0.0013) | 0.8983 (0.0009) | 0.9493 (0.0007) |
| 4 | 1.1494 | Exponential | 0.7504 (0.0014) | 0.8998 (0.0010) | 0.9501 (0.0007) |
| | | Gamma | 0.7520 (0.0014) | 0.9002 (0.0009) | 0.9509 (0.0007) |
| | | Uniform | 0.7499 (0.0014) | 0.8975 (0.0010) | 0.9489 (0.0007) |

$v^{avc}$ ($\theta^{avc} = 1.2$, $k = 3$, $P^* = 0.90$). Using (14), we obtain $\theta^{bem} = 1.2633$, and using FORTRAN code developed by Goldsman [6], we find $v^{bem}$ ($\theta^{bem} = 1.2633$, $k = 3$, $P^* = 0.90$) = 264. To show how good an approximation this provides for our specified $v^{avc}$, we simulate $M = 100{,}000$ macroreplications each containing 264 vector replications using $\theta^{avc} = 1.2$ with exponential populations, and obtain $\widehat{PCS}^{avc} = 0.8997$ with a standard error of 0.0009. So here our LSA is very good. These results are included in Table 5, where we see that the results for all the distributions achieve the desired $P^*$ to the second decimal place in almost all cases. In fact, we note that the Bernoulli results are significantly larger than $P^*$ in many cases. The cases where we see more of a departure from $P^*$ are for smaller values of $P^*$ where $v$ is typically less than 30. For a few cases, the estimated PCS values fall more than two standard errors below $P^*$, but these differences are practically insignificant.

We also notice that $v^{bem}$ ($\theta^{bem} = 1.2$, $k = 3$, $P^* = 0.90$) = 437 [2]. Comparing this with our approximate $v^{avc}$ ($\theta^{avc} = 1.2$, $k = 3$, $P^* = 0.90$) = 264, we see a nearly 40% reduction in the number of replications required with AVC.

**Table 8.** PCS achieved for $\theta^{bem} = \theta^* = 2.0$ using LSA.

| $k$ | $\theta^{avc}$ | Distribution | $P^* = 0.75$ | $P^* = 0.90$ | $P^* = 0.95$ |
|---|---|---|---|---|---|
| 2 | 1.7479 | Exponential | 0.7692 (0.0013) | 0.9021 (0.0009) | 0.9471 (0.0007) |
| | | Gamma | 0.7702 (0.0013) | 0.9037 (0.0009) | 0.9477 (0.0007) |
| | | Uniform | 0.7703 (0.0013) | 0.9013 (0.0009) | 0.9453 (0.0007) |
| | | Bernoulli | 0.7950 (0.0013) | 0.9373 (0.0008) | 0.9726 (0.0005) |
| 3 | 1.7205 | Exponential | 0.7471 (0.0014) | 0.8971 (0.0010) | 0.9460 (0.0007) |
| | | Gamma | 0.7456 (0.0013) | 0.8975 (0.0009) | 0.9464 (0.0007) |
| | | Uniform | 0.7459 (0.0014) | 0.8933 (0.0010) | 0.9432 (0.0007) |
| | | Bernoulli | 0.7911 (0.0013) | 0.9332 (0.0008) | 0.9701 (0.0005) |
| 4 | 1.7124 | Exponential | 0.7419 (0.0014) | 0.8944 (0.0010) | 0.9461 (0.0007) |
| | | Gamma | 0.7418 (0.0014) | 0.8956 (0.0010) | 0.9464 (0.0007) |
| | | Uniform | 0.7403 (0.0014) | 0.8888 (0.0010) | 0.9414 (0.0007) |
| 5 | 1.7092 | Exponential | 0.7458 (0.0014) | 0.8950 (0.0010) | 0.9474 (0.0007) |
| | | Gamma | 0.7452 (0.0014) | 0.8976 (0.0010) | 0.9478 (0.0007) |
| | | Uniform | 0.7412 (0.0014) | 0.8881 (0.0010) | 0.9410 (0.0007) |
| 10 | 1.7061 | Exponential | 0.7576 (0.0014) | 0.9024 (0.0009) | 0.9516 (0.0007) |
| | | Uniform | 0.7476 (0.0014) | 0.8891 (0.0009) | 0.9411 (0.0007) |

**Table 9.** Minimum number of vectors to achieve $P^*$ for AVC (BEM).

| $k$ | $\theta^*$ | $P^* = 0.75$ | $P^* = 0.90$ | $P^* = 0.95$ |
|---|---|---|---|---|
| 2 | 1.01 | 12171 (18371) | [a] | [a] |
| | 1.05 | 509 (765) | 1839 (2759) | 3027 (4545) |
| | 1.10 | 133 (201) | 483 (723) | 793 (1191) |
| | 1.20 | 37 (55) | 133 (199) | 217 (327) |
| | 2.00 | 3 (5) | 9 (15) | 15 (23) |
| 3 | 1.05 | 1544 (2565) | 3741 (6211) | 5526 (9165) |
| | 1.10 | 401 (666) | 972 (1615) | 1436 (2385) |
| | 1.20 | 108 (181) | 264 (437) | 388 (645) |
| | 2.00 | 7 (12) | 17 (29) | 25 (42) |
| 4 | 1.20 | 187 (326) | 398 (692) | 565 (979) |
| | 2.00 | 12 (20) | 25 (43) | 36 (61) |
| 5 | 1.20 | 271 (486) | 541 (964) | 748 (1331) |
| | 2.00 | 17 (29) | 33 (58) | 46 (81) |

[a] Numbers not available due to large computation time.

To be complete we consider modifying the approximation in (14) by setting $\theta^{\text{bem}} = \theta^*$ and solving for $\theta^{\text{avc}}$. We then perform a similar set of simulation runs to obtain estimated $\text{PCS}^{\text{avc}}$ values at $k$ and $\theta^{\text{avc}}$ using the appropriate values of $v^{\text{bem}}$. Results of these runs are reported in Tables 7 and 8. As with the previous form of our LSA, we are looking for estimated $\text{PCS}^{\text{avc}}$ values that are close to the $P^*$ listed at the top of the column in which they appear. These results show that the LSA is good in this direction as well.

The benefit from this form of the LSA is reflected by $\theta^{\text{avc}} < \theta^*$. This indicates that AVC can provide better discrimination between the systems at the same level of confidence and with the same data.

## 8. CONCLUSIONS

When trying to pick the best system out of $k$ systems, there are many instances when this selection should be based on one-time performance rather than long-run average performance. Multinomial selection procedures provide a framework for defining such a problem, and Procedure BEM is the classical approach for solving it. Procedure AVC is an alternative approach designed to obtain a higher PCS by performing all possible comparisons across all systems for a given set of system performance data. Construction of Procedure AVC closely follows that of BEM, allowing researchers to easily move from a standard approach to our new approach.

From the simulation design point of view, AVC can also be used to our advantage by allowing us to plan a smaller number of replications to achieve a desired PCS, $P^*$. Table 9 presents comparisons of the minimum number of independent replications needed to achieve a given $P^*$ for AVC and BEM. The AVC values are obtained using our LSA in (14) with $\theta^{\text{avc}} = \theta^*$ to find $\theta^{\text{bem}}$ and then running an exact code for $\text{PCS}^{\text{bem}}$ provided by Goldsman [6] at $\theta = \theta^{\text{bem}}$. Values for BEM are taken from [2] or obtained using Goldsman's code [6] for unavailable values of $\theta^{\text{bem}} = \theta^*$. As $k$ increases, we see a more dramatic reduction in the number of vector observations needed with AVC to achieve the same $P^*$. The reduction in $v$ goes from roughly 34% at $k = 2$ up to 44% at $k = 5$. So the advantages of AVC over BEM appear greater for more challenging MSPs.

## APPENDIX: PROOFS

PROOF OF THEOREM 1: From Lemma 1 we have the MVN distribution of $(A_2(v), \ldots, A_k(v))$. Since $A_i(v) - A_1(v) = A_i(v) + \sum_{j=2}^{k} A_j(v)$ $(i \neq 1)$, we take the difference of MVN random variables and obtain

$$\begin{pmatrix} A_2(v) - A_1(v) \\ \vdots \\ A_k(v) - A_1(v) \end{pmatrix} \overset{v}{\Rightarrow} \text{MVN}\left[ \begin{pmatrix} -k\delta \\ \vdots \\ -k\delta \end{pmatrix}, 2/k \begin{pmatrix} 1 & 1/2 & \cdots & 1/2 \\ 1/2 & 1 & \cdots & 1/2 \\ \vdots & \vdots & \ddots & \vdots \\ 1/2 & 1/2 & \cdots & 1 \end{pmatrix} \right]. \tag{15}$$

Assuming population 1 is the best, in terms of $\overrightarrow{\text{PCS}}^{\text{bem}}$, we can state

$$\overrightarrow{\text{PCS}}^{\text{bem}} = \lim_{v\to\infty} \Pr\{Y_1(v) > Y_j(v), \forall j \neq 1\} = \lim_{v\to\infty} \Pr\left\{ \frac{Y_1(v) - v/k}{\sqrt{v}} > \frac{Y_j(v) - v/k}{\sqrt{v}}, \forall j \neq 1 \right\}$$

$$= \lim_{v\to\infty} \Pr\{A_j(v) - A_1(v) < 0, \forall j \neq 1\} = \Pr\{W_j < 0, j = 2, 3, \ldots, k\}, \tag{16}$$

where $(W_2, W_3, \ldots, W_k) \sim (15)$. If we add $k\delta$ to each $W_j$ to obtain a random vector with a mean of zero, then from (16), we have

$$\overrightarrow{\text{PCS}}^{\text{bem}} = \Pr\left\{ \frac{W_j + k\delta}{\sqrt{2/k}} < \frac{k\delta}{\sqrt{2/k}}, j = 2, \ldots, k \right\} = \Pr\left\{ Q_j < \frac{k\delta}{\sqrt{2/k}}, j = 2, \ldots, k \right\},$$

where $(Q_2, \ldots, Q_k)$ has the same MVN distribution as that given by (7). □

PROOF OF THEOREM 2: In the following proof, we first develop the asymptotic covariance matrix for a vector of standardized AVC point estimators using individual U-statistic covariance terms. Much of the proof involves defining these covariance terms, of which there are four distinct cases. We then use the resulting covariance matrix to derive the required expression for the asymptotic $\text{PCS}^{\text{avc}}$.

Define $\Sigma(N)$, the covariance matrix computed for $(F_2^{(v)}, F_3^{(v)}, \ldots, F_k^{(v)})$, as

$$\Sigma(N) = \text{Var}\begin{bmatrix} \sqrt{N}\bar{p}_2(N) \\ \vdots \\ \sqrt{N}\bar{p}_k(N) \end{bmatrix}.$$

Let $p_j(N) = E_N[\bar{p}_j(N)]$, the expected value at sample size $N = kv$ [here $p_j(N)$ and $\bar{p}_j(N)$ are equivalent to $p_j(v)$ and $\bar{p}_j(v)$, respectively, with $v$ replaced with $kv$]. Lehmann [8] shows that Lemma 2 holds even if the distributions of the data depend upon the sample size provided

$$\Sigma(N) \to \Sigma,$$

where $\Sigma$ is nonsingular as $N \to \infty$. We assume

$$\text{Var}\begin{pmatrix} \sqrt{N}(\bar{p}_2(N) - p_2(N)) \\ \vdots \\ \sqrt{N}(\bar{p}_k(N) - p_k(N)) \end{pmatrix} \to \Sigma$$

for any reasonable set of $F_i^{(v)}$. We define $\Sigma$ as $\|\sigma^{(a,b)}\|$ in (9).

To proceed, we need to consider the covariance terms defined in Lemma 2:

$$\begin{aligned}\xi_i^{(a,b)} &= \mathrm{E}[H_{i1}^{(a)} H_{i2}^{(b)}] - \mathrm{E}[H_{i1}^{(a)}]\mathrm{E}[H_{i2}^{(b)}] \\ &= \mathrm{E}[H_{i1}^{(a)} H_{i2}^{(b)}] - 1/k^2\end{aligned}$$

since

$$H_{i1}^{(a)} = \begin{cases} 1, & \text{if } X_{a\alpha_a} > X_{\ell\alpha_\ell} \quad \forall \ell \neq a, \\ 0, & \text{otherwise,} \end{cases}$$

$$H_{i2}^{(b)} = \begin{cases} 1, & \text{if } X_{b\beta_b} > X_{\ell\beta_\ell} \quad \forall \ell \neq b, \\ 0, & \text{otherwise.} \end{cases}$$

Using our kernel from (8), we can express

$$\mathrm{E}[H_{a1}^{(a)} H_{a2}^{(a)}] = \mathrm{E}\left[ \prod_{i=1, i\neq a}^{k} \phi(X_{a\alpha_a} - X_{i\alpha_i}) \times \prod_{\ell=1, \ell\neq a}^{k} \phi(X_{a\alpha_a} - X_{\ell\beta_\ell}) \right].$$

We can then write this as a probability statement combining the two indices to come up with

$$\mathrm{E}[H_{a1}^{(a)} H_{a2}^{(a)}] = \Pr\{ X_{a\alpha_a} > \max_{\ell \neq \mathrm{a}} \{ X_{\ell\alpha_\ell}, X_{\ell\beta_\ell} \} \}.$$

This particular case is easy to illustrate, and we extend this development for other expected value terms in less detail.

There are a number of different cases we need to consider for the covariance terms, $\xi_i^{(a,b)}$. In this notation, $a$ represents the population with the largest value in $H_{i1}^{(a)}$; $b$ represents the population with the largest value in $H_{i2}^{(b)}$; and $i$ represents the one population that has the same observation in both vectors (pseudoreplications). We can enumerate the different cases for the covariance terms based on the values of $a$, $b$, and $i$. We have the following four cases:

1. $a = b = i$; $\xi_a^{(a,a)}$.
2. $a = b \neq i$; $\xi_i^{(a,a)}$.
3. $a \neq b \neq i$; $\xi_i^{(a,b)}$.
4. $a \neq b$, $a = i$ or $b = i$; $\xi_a^{(a,b)}$ or $\xi_b^{(a,b)}$.

Asymptotically $X_j \sim F \ \forall j$ which allows us to construct distribution-free expressions for $\xi_i^{(a,b)}$ since each $X_{ij}$ has the same probability of being the largest value in a single vector. For case 1 we have

$$\xi_a^{(a,a)} = \Pr\{ X_{a\alpha_a} > \max_{\ell \neq \mathrm{a}} \{ X_{\ell\alpha_\ell}, X_{\beta_\ell} \} \} - \frac{1}{k^2} = \frac{1}{2k-1} - \frac{1}{k^2}. \tag{17}$$

This follows since we have $2(k-1)+1$ independent and identically distributed random variables and we want the probability that a particular one is the largest. In our context this means we want the probability that a single observation from population $a$ is a winner in two separate vectors containing no other common observations.

Case 2 is more difficult to approach. We have

$$\xi_i^{(a,a)} = \Pr\left\{ \begin{matrix} X_{a\alpha_a} > X_{i\alpha_i}, X_{a\alpha_a} > \max_{\ell\neq a,i}\{X_{\ell\alpha_\ell}\}; \\ X_{a\beta_a} > X_{i\alpha_i}, X_{a\beta_a} > \max_{\ell\neq a,i}\{X_{\ell\beta_\ell}\} \end{matrix} \right\} - \frac{1}{k^2},$$

where the common observation in each of the two vectors is not the largest, and both observations from population $a$ are the largest in their respective vectors. The ordering of the random variables from this pair of vectors must look like the following:

$$X_{a\alpha_a} > \{h \text{ of the } X_{\ell\alpha_\ell}\} > X_{a\beta_a} > \{(k-2)-h \text{ of the } X_{\ell\alpha_\ell},\ (k-2) \text{ of the } X_{\ell\beta_\ell}, \text{ and } X_{i\alpha_i}\}$$

or interchanging $X_{a\alpha_a}$ and $X_{a\beta_a}$

$$X_{a\beta_a} > \{h \text{ of the } X_{\ell\beta_\ell}\} > X_{a\alpha_a} > \{(k-2)-h \text{ of the } X_{\ell\beta_\ell},\ (k-2) \text{ of the } X_{\ell\alpha_\ell}, \text{ and } X_{i\alpha_i}\},$$

where $h = 0, 1, \ldots, k-2$. For each subset of size $h$ there are

$$h!((k-2)-h+(k-2)+1)$$

equally likely orderings, and there are

$$\binom{k-2}{h}$$

ways to select an $h$. With a total of $(2k-1)!$ possible orderings, we then have

$$\xi_i^{(a,a)} = \frac{2\sum_{h=0}^{k-2}\binom{k-2}{h}h!((k-2)-h+(k-2)+1)!}{(2k-1)!} - \frac{1}{k^2} = \frac{2}{k(2k-1)} - \frac{1}{k^2}. \tag{18}$$

We proceed in a similar fashion for case 3

$$\xi_i^{(a,b)} = \Pr\left\{\begin{matrix} X_{a\alpha_a} > X_{i\alpha_i},\ X_{a\alpha_a} > \max_{\ell\neq a,i}\{X_{\ell\alpha_\ell}\}; \\ X_{b\beta_b} > X_{i\alpha_i},\ X_{b\beta_b} > \max_{\ell\neq b,i}\{X_{\ell\beta_\ell}\} \end{matrix}\right\} - \frac{1}{k^2} = \xi_i^{(a,a)} = \frac{2}{k(2k-1)} - \frac{1}{k^2} \tag{19}$$

since all random variables are identically distributed and there is no distinction between $X_{b\beta_b}$ and $X_{a\beta_a}$.

For case 4 we have

$$\xi_a^{(a,b)} = \Pr\{X_{a\alpha_a} > \max_{\ell\neq a}\{X_{\ell\alpha_\ell}\};\ X_{b\beta_b} > X_{a\alpha_a},\ X_{b\beta_b} > \max_{\ell\neq a,b}\{X_{\ell\beta_\ell}\}\} - \frac{1}{k^2}.$$

As we did for $\xi_a^{(a,a)}$ previously, we need to identify all possible arrangements of the random variables from two vectors that meet the above conditions. The following orderings work:

$$X_{b\beta_b} > \{h \text{ of the } X_{\ell\beta_\ell}\} > X_{a\alpha_a} > \{(k-2)-h \text{ of the } X_{\ell\beta_\ell},\ (k-1) \text{ of the } X_{\ell\alpha_\ell}\}.$$

Then proceeding as we did for $\xi_a^{(a,a)}$ we obtain for $a \neq b$

$$\xi_a^{(a,b)} = \xi_b^{(a,b)} = \frac{1}{k(2k-1)} - \frac{1}{k^2}. \tag{20}$$

Given covariance expressions from (17) and (18), we find the diagonal terms of $\sigma^{(a,a)}$ from (9) as

$$\sigma^{(a,a)} = k\sum_{i=1}^{k} \xi_i^{(a,a)} = k[\xi_a^{(a,a)} + (k-1)\xi_i^{(a,a)}] = \frac{k-1}{2k-1}. \tag{21}$$

With the covariance expressions from (19) and (20) we can find the off-diagonal terms of $\sigma^{(a,b)}$ from (9) as

$$\sigma^{(a,b)} = k\left[\xi_a^{(a,b)} + \xi_b^{(a,b)} + \sum_{\ell \neq a,b} \xi_\ell^{(a,b)}\right] = \frac{-1}{2k-1}. \tag{22}$$

Combining the terms in (22) with those in (21), we let $\eta = 1/(k-1)$, and we have

$$\boldsymbol{\Sigma} = \left(\frac{k-1}{2k-1}\right)\begin{pmatrix} 1 & -\eta & \cdots & -\eta \\ -\eta & 1 & \cdots & -\eta \\ \vdots & \vdots & \ddots & \vdots \\ -\eta & -\eta & \cdots & 1 \end{pmatrix} \tag{23}$$

From [8] we know that

$$\begin{pmatrix} \sqrt{N}(\bar{p}_2(N) - (1/k - \delta/\sqrt{v})) \\ \vdots \\ \sqrt{N}(\bar{p}_k(N) - (1/k - \delta/\sqrt{v})) \end{pmatrix} \Rightarrow \text{MVN}\left[\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \boldsymbol{\Sigma}\right]. \tag{24}$$

Assuming population 1 is the best, in terms of $\vec{\text{PCS}}^{\text{avc}}$ we have

$$\vec{\text{PCS}}^{\text{avc}} = \lim_{N\to\infty} \Pr\{\bar{p}_1(N) > \bar{p}_j(N), \forall j \neq 1\} = \lim_{N\to\infty} \Pr\{\sqrt{N}(\bar{p}_j(N) - 1/k) - \sqrt{N}(\bar{p}_1(N) - 1/k) < 0, \forall j \neq 1\} = \Pr\{\bar{W}_j < 0, \forall j \neq 1\},$$

where

$$\begin{pmatrix} \bar{W}_2 \\ \vdots \\ \bar{W}_k \end{pmatrix} \sim \text{MVN}\left[\begin{pmatrix} -k\sqrt{k}\delta \\ \vdots \\ -k\sqrt{k}\delta \end{pmatrix}, \boldsymbol{\Xi}\right].$$

Using our variance and covariance terms from (23), the diagonal terms of $\boldsymbol{\Xi}$ are

$$\sigma^{(a,a)} + \sigma^{(b,b)} - 2\sigma^{(a,b)} = \frac{2k}{2k-1},$$

and the off-diagonal terms are

$$\sigma^{(a,b)} - \sigma^{(a,1)} - \sigma^{(1,b)} + \sigma^{(1,1)} = \frac{k}{2k-1}.$$

Combining these terms we have

$$\Xi = \frac{2k}{2k-1}\begin{pmatrix} 1 & 1/2 & \cdots & 1/2 \\ 1/2 & 1 & \cdots & 1/2 \\ \vdots & \vdots & \ddots & \vdots \\ 1/2 & 1/2 & \cdots & 1 \end{pmatrix}.$$

Then

$$\vec{\mathrm{PCS}}^{\mathrm{ave}} = \Pr\left\{\frac{\bar{W}_j + k\sqrt{k\delta}}{\sqrt{2k/(2k-1)}} < \frac{k\sqrt{k\delta}}{\sqrt{2k/(2k-1)}}, j = 2, \ldots, k\right\} = \Pr\left\{Q_j < \frac{k\delta}{\sqrt{2/(2k-1)}}, j = 2, \ldots, k\right\},$$

where $(Q_2, \ldots, Q_k)$ has the same MVN distribution as that given by (7). □

## ACKNOWLEDGMENTS

This work was partially supported by NSF Grant No. DMI-9622065. The authors acknowledge the helpful discussions of this work with Dave Goldsman of Georgia Tech and helpful comments from the referee.

## REFERENCES

[1] Bechhofer, R.E., Elmaghraby, S., and Morse, N., "A Single-Sample Multiple-Decision Procedure for Selecting the Multinomial Event which Has the Highest Probability," *Annals of Mathematical Statistics,* March, **30,** 102–119 (1959).

[2] Bechhofer, R.E., Santner, T., and Goldsman, D., *Design and Analysis of Experiments for Statistical Selection, Screening and Multiple Comparisons,* Wiley, New York, **29,** 1995.

[3] Bechhofer, R.E., and Sobel, M., "Non-Parametric Multiple-Decision Procedures for Selecting That One of $k$ Populations Which Has the Highest Probability of Yielding the Largest Observation," preliminary report, Abstract, *Annals of Mathematical Statistics,* March, 325 (1958).

[4] Gibbons, J.D., Olkin, I., and Sobel, M., *Selecting and Ordering Populations: A New Statistical Methodology,* Wiley, New York, 1977.

[5] Goldsman, D., "A Multinomial Ranking and Selection Procedure: Simulation and Applications," in S. Shepard, U.W. Pooch, and C.D. Pegden, Eds., *Proceedings of the 1984 Winter Simulation Conference,* The Institute of Electrical and Electronics Engineers, Piscataway, NJ, 1984, pp. 259–264.

[6] Goldsman, D., personal communication, 1995.

[7] Kesten, H., and Morse, N., "A Property of the Multinomial Distribution," *Annals of Mathematical Statistics,* March, **30,** 120–127 (1959).

[8] Lehmann, E.L., "Robust Estimation in Analysis of Variance," *Annals of Mathematical Statistics,* September, **34,** 957–966 (1963).

[9] Lehmann, E.L., *Testing Statistical Hypotheses,* 2nd ed., Wiley, New York, 1986.

[10] Miller, J.O., "Efficient Multinomial Selection in Simulation," Ph.D. thesis, Department of Industrial, Welding, and Systems Engineering, The Ohio State University, 1997 (unpublished).

[11] Randles, H.R., and Wolfe, D.A., *Introduction to the Theory of Nonparametric Statistics,* Wiley, New York, 1979.