Proceedings of the 1996 Winter Simulation Conference
ed. J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain

GETTING MORE FROM THE DATA IN A MULTINOMIAL SELECTION PROBLEM

J. O. Miller

Department of Industrial, Welding
& Systems Engineering
The Ohio State University
Columbus, OH 43210, U.S.A.

Barry L. Nelson

Department of Industrial Engineering
& Management Sciences
Northwestern University
Evanston, IL 60208, U.S.A.

Charles H. Reilly

Department of Industrial Engineering & Management Systems
University of Central Florida
Orlando, FL 32816, U.S.A.

ABSTRACT

We consider the problem of determining which of k simulated systems is most likely to be the best performer based on some objective performance measure. The standard experiment is to generate v independent vector observations (replications) across the k systems. A classical multinomial selection procedure, BEM (Bechhofer, Elmaghraby, and Morse), prescribes a minimum number of replications so that the probability of correctly selecting the true best system 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. We propose using these same v replications across k systems to form v^k pseudoreplications (no longer independent) that contain one observation from each system, and again select as best the system having the largest value of the performance measure in more pseudoreplications than any other. We expect that this new procedure, AVC (all vector comparisons), dominates BEM in the sense that AVC will never require more independent replications than BEM to meet a prespecified probability of correct selection. We present analytical and simulation results to show how AVC fares versus BEM for different underlying distribution families, different numbers of populations and various values of v. We also present results for the closely related problem of estimating the probability that a specific system is the best.

1 INTRODUCTION AND MOTIVATION

Motivating Example: As tactical war planning analysts, we are directed to provide the Joint Task Force Commander with the best plan to cripple the enemy's command and control. "Best" means achieving the highest level of cumulative damage expectancy (CDE) against a selected set of targets given current intelligence estimates of enemy defense capabilities and available friendly forces. Our team prepares four independent attack plans and we simulate v replications across all four plans. For each replication we compare the CDE between each of the four plans. Since the chosen plan can only be executed a single time, we select as the best plan the one that has the largest CDE in most of the replications.

We consider the general problem of selecting the best of $k \geq 2$ independent populations, $\pi_1, \pi_2, \ldots, \pi_k$, where in our context "populations" is taken to mean simulated systems. This is known as the multinomial selection problem (MSP). Let $\mathbf{X}_i = (X_{1i}, X_{2i}, \ldots, X_{ki})$ represent a vector of independent observations of some common performance measure across all populations on the i^{th} replication. For each i, the best population is the population with the largest X_{ji}. The goal is to find the population that is most likely to be the best performer among the populations, as opposed to identifying the best average performer over the long run. Applications include selecting the best of a set of tactical or strategic military actions, as presented above. In the areas of marketing research or opinion surveys, we might determine the most popular brand, flavor, etc., or the most favored candidate or position on a political issue. An example in the area of structural engineering is finding the design that performs best in a one-time catastrophic event, such as an earthquake. The goal in any MSP is to achieve a prespecified probability of correctly selecting the best population with a minimum amount of data.

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 populations, and let $Y_j = \sum_{i=1}^{v} Y_{ji}$ represent the number of times population j wins out of these v replications. So $\sum_{j=1}^{k} Y_j = v$ and the k-variate discrete random variable $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_k)$ follows a multinomial distribution with $p_j = \Pr\{\pi_j \text{ wins}\}, j = 1, 2, \ldots, k$, where $0 < p_j < 1, j = 1, 2, \ldots, k$, and $\sum_{j=1}^{k} p_j = 1$. The probability mass function for $\mathbf{Y}$ with parameter v and success probabilities $\mathbf{p} = (p_1, p_2, \ldots, p_k)$ is

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

The goal of MSP is to find the population π_i associated with the largest p_i.

The paper is organized as follows: We first provide a brief review of MSP and the classical approach to solving it. Then we describe our new procedure that uses the same data and increases the probabilitly of correctly selecting the best population. Some analytical results are presented next, along with empirical results for a number of specific distributions for the X_{ji}. Finally, we describe the closely related problem of estimating p_j, the probability that population j will be the best.

2 BACKGROUND

Bechhofer, Elmaghraby and Morse (1959) describe a single-stage procedure for selecting the multinomial event (population) 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 θ^* (where $1 < \theta^* < \infty$), the ratio of the largest success probability to the second largest success probability. The procedure consists of the following steps:

Procedure BEM

1. For given k and θ^*, find the minimum value of v that guarantees that the probability of selecting the best population 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 θ^* 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 approach.* Select v as the minimum number of independent vector observations required to achieve a probability of correct selection (PCS) greater than or equal to P^* whenever $\theta \geq \theta^*$.

If we obtain a PCS $\geq P^*$ with our selected v under the least favorable configuration (LFC) of $\mathbf{p} = (p_{[1]}, p_{[2]}, \ldots, p_{[k]})$, a PCS of at least P^* can be guaranteed for any configuration of $\mathbf{p}$ with $\theta \geq \theta^*$. Keston and Morse (1959) prove that the LFC for BEM is given by:

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

However, the PCS can be calculated for any configuration of $\mathbf{p}$ with $p_{[k]} > p_{[k-1]}$. Let $\pi_{[j]}$ be the population associated with $p_{[j]}$ and let $Y_{[j]}$ represent the number of wins for $\pi_{[j]}$. So the subscripts for the populations and the associated number 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 PCS$^{\text{bem}}$. For any fixed k and v, PCS$^{\text{bem}}$can be expressed as

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

where the summation is over all vectors $\mathbf{Y} = (Y_{[1]}, \ldots, Y_{[k]})$ such that $Y_{[k]} \geq Y_{[j]}$ $(j = 1, 2, \ldots, k-1)$, and t is the number of populations tied for the most wins.

3 NEW METHOD

We propose a method to provide a PCS greater than or equal to PCS$^{\text{bem}}$ using the same replications $\mathbf{X}_i, i = 1, 2, \ldots, v$. In other words, an improvement in PCS for "free." We use the BEM parameters k, P^*, and θ^*, and we execute the first step of BEM to find a value of v. Then a total of v^k pseudoreplications are 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

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

for $j = 1, 2, \ldots, k$, where $\phi(X_{ji}, X_{\ell i})$ is an indicator function that takes the value 1 only if X_{ji} is larger than $X_{\ell i}$. Thus, Z_j represents the number of times out of v^k pseudoreplications that population π_j wins. This new procedure consists of the following steps:

Procedure AVC

1. Given values for k, P^*, and θ^*, 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 equation (1).
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.

We conjecture that the PCS with AVC, referred to as PCS^{avc}, is greater than or equal to PCS^{bem}. PCS^{avc} can be expressed

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

where the summation is over all vectors $\mathbf{Z} = (Z_{[1]}, \ldots, Z_{[k]})$ such that $Z_{[k]} \geq Z_{[j]}$, $j = 1, 2, \ldots, k-1$, and t is the number of populations tied for the most wins. Since $\mathbf{Z}$ does not follow a multinomial distribution, we must go back to the distributions of the original observations, X_{ji}, $j = 1, 2, \ldots, k$; $i = 1, 2, \ldots, v$ to calculate PCS^{avc}. Analytical and simulation results using a number of different population distributions show that PCS^{avc} depends on the underlying distributions of the $X_{ji}\,s$.

4 ANALYTICAL RESULTS

Initially, we restrict our attention to continuous distributions for the $X_{ji}\,s$ which eliminates the possibility of ties among observations. We can calculate PCS^{avc} by conditioning on the joint density of all the order statistics for the v independent replications from $\pi_{[k]}$. Let $\pi_{[k]}$ follow some distribution and let X represent an observation from $\pi_{[k]}$. Let all the remaining populations be identically distributed and let O represent an observation from any of these populations. This set up gives us the LFC for BEM.

Consider permutations of the ranks of the observations from all populations. 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*.

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

$$\begin{aligned}\Pr\{Z_{[3]} = 8\} &= \Pr\{O_{(1)} < O_{(2)} < O_{(3)} \\ &\quad < O_{(4)} < X_{(1)} < X_{(2)}\} \quad (2)\\ \Pr\{Z_{[3]} = 6\} &= 4\Pr\{O_{(1)} < O_{(2)} < O_{(3)} \\ &\quad < X_{(1)} < O_{(4)} < X_{(2)}\} \quad (3)\end{aligned}$$

Probability statement (2) uniquely covers all 4! permutations of the O's that are less than both X's. However, in probability statement (3), since all the O's are not adjacent, we need to consider how many of the 4! permutations of the O's result in a unique combination of adjacent O's. This is why the coefficient '4' appears on the right-hand side of equation (3). 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 multiple rank orders that result in the same value for $Z_{[k]}$ and the calculation of the probability of each value of $Z_{[k]}$ becomes extremely tedious.

If we pick a particular distribution family for X and O, then we can obtain experimental results with simulation as well as derive formulas to compare PCS^{avc} with PCS^{bem}. First, consider $X \sim \exp(\lambda)$ and $O \sim \exp(\mu)$ and let $\lambda < \mu$, where λ and μ are exponential rates. For $k = 2$, $v = 2$, we have $p_{[2]} = \Pr\{X > O\} = \mu/(\lambda + \mu)$ and $p_{[1]} = \Pr\{X < O\} = \lambda/(\lambda + \mu)$. To calculate PCS^{bem}, we need to consider vectors $\mathbf{Y} = (Y_{[1]}, Y_{[2]})$ such that $Y_{[2]} \geq Y_{[1]}$.

With $v = 2$, the only possible $\mathbf{Y}'s$ with $Y_{[2]} \geq Y_{[1]}$ are $(0,2)$ and $(1,1)$. This gives us

$$\begin{aligned} PCS^{bem} &= \Pr\{Y_{[2]} = 2\} + \frac{1}{2}\Pr\{Y_{[2]} = 1\} \quad (4) \\ &= p_{[2]}^2 + \frac{1}{2}\, 2\, p_{[1]}p_{[2]} \\ &= \frac{\mu}{\lambda + \mu} \end{aligned}$$

Similarly, to calculate PCS^{avc}, we need to consider vectors $\mathbf{Z} = (Z_{[1]}, Z_{[2]})$ such that $Z_{[2]} \geq Z_{[1]}$. With $v^k = 4$, the only $\mathbf{Z}'s$ with $Z_{[2]} \geq Z_{[1]}$ are: $(0,4)$, $(1,3)$, $(2,2)$. So

$$\begin{aligned} PCS^{avc} &= \Pr\{Z_{[2]} = 4\} + \Pr\{Z_{[2]} = 3\} \\ &+ \frac{1}{2}\Pr\{Z_{[2]} = 2\}. \quad (5) \end{aligned}$$

When $X \sim \exp(\lambda)$, the joint distribution of $(X_{[1]}, X_{[2]})$ is $f_{12}(a,b) = 2\lambda^2 e^{-\lambda(a+b)}$. The probabilities for the values of $Z_{[2]}$ in Equation (5) and PCS^{avc} can then be found as follows:

$$\begin{aligned} \Pr\{Z_{[2]} = 4\} &= \int_0^\infty \int_0^b (1 - e^{-\mu a})^2 \times 2\lambda^2 e^{-\lambda(a+b)} da\, db \\ &= \frac{\mu^2}{(2\lambda + \mu)(\lambda + \mu)} \\ \Pr\{Z_{[2]} = 3\} &= 2\int_0^\infty \int_0^b (1 - e^{-\mu a})(e^{-\mu a} - e^{-\mu b}) \times 2\lambda^2 e^{-\lambda(a+b)} da\, db \\ &= \frac{2\lambda^2}{(2\lambda + \mu)(\lambda + \mu)^2} \\ \Pr\{Z_{[2]} = 2\} &= \int_0^\infty \int_0^b (e^{-\mu a} - e^{-\mu b})^2 \times 2\lambda^2 e^{-\lambda(a+b)} da\, db \; + \\ &\quad 2\int_0^\infty \int_0^b (1 - e^{-\mu a}) e^{-\mu b} \times 2\lambda^2 e^{-\lambda(a+b)} da\, db \\ &= \frac{2\lambda\mu(4\lambda\mu + \mu^2 + \lambda^2)}{(2\lambda + \mu)(\lambda + 2\mu)(\lambda + \mu)^2} \\ PCS^{avc} &= \frac{\mu(\lambda^2 + 6\lambda\mu + 2\mu^2)}{(2\lambda + \mu)(\lambda + 2\mu)(\lambda + \mu)} \end{aligned}$$

Given expressions for both PCS^{bem} and PCS^{avc}, we can find the improvement in PCS with AVC from

$$\begin{aligned} \Delta PCS &= PCS^{avc} - PCS^{bem} \\ &= \frac{\lambda\mu(\mu - \lambda)}{(2\lambda + \mu)(\lambda + 2\mu)(\lambda + \mu)} > 0 \quad (6) \end{aligned}$$

The $(\mu - \lambda) > 0$ term in equation (6) shows that when X is the best population, AVC always shows an improvement in PCS over BEM. Similar analytical results have been obtained for exponential populations with $k = 2, v = 3$ and $k = 3, v = 2$. Additionally, results for continuous uniform populations ($k = 2, v = 2,3;\ k = 3, v = 2$) show an increase in PCS with AVC, but the increase is different than it is for the exponential populations.

As an illustration of how AVC compares to BEM for discrete distributions, let $X \sim \text{Bern}(p_x)$ and $O \sim \text{Bern}(p_o)$ with $p_x > p_o$. For $k = 2$, $v = 2$, we have $p_{[2]} = \Pr\{X > O\} = 1/2(p_x + 1 - p_o)$ and $p_{[1]} = \Pr\{X < O\} = 1/2(p_o + 1 - p_x)$. From equation (4) we obtain

$$PCS^{bem} = \frac{1}{2}(p_x + 1 - p_o)$$

PCS^{avc} can be calculated as we did for our exponential example by finding the probabilities for the following values of $Z_{[2]}$ and using Equation (5).

$$\begin{aligned} \Pr\{Z_{[2]} = 4\} &= \frac{9}{16}p_x^2 - \frac{5}{8}p_x^2 p_o + \frac{1}{8}p_x^2 p_o^2 \\ &+ \frac{3}{8}p_x - \frac{3}{4}p_x p_o + \frac{3}{8}p_x p_o^2 \\ &- \frac{1}{8}p_o + \frac{1}{16}p_o^2 + \frac{1}{16} \\ \Pr\{Z_{[2]} = 3\} &= -\frac{3}{4}p_x^2 + \frac{3}{2}p_x^2 p_o - \frac{1}{2}p_x^2 p_o^2 \\ &+ \frac{1}{2}p_x - \frac{1}{2}p_x p_o^2 - \frac{1}{2}p_o \\ &+ \frac{1}{4}p_o^2 + \frac{1}{4} \\ \Pr\{Z_{[2]} = 2\} &= -\frac{1}{8}p_x^2 - \frac{3}{4}p_x^2 p_o + \frac{3}{4}p_x^2 p_o^2 \\ &- \frac{1}{4}p_x + \frac{3}{2}p_x p_o - \frac{3}{4}p_x p_o^2 \\ &- \frac{1}{4}p_o - \frac{1}{8}p_o^2 + \frac{3}{8} \\ PCS^{avc} &= \frac{1}{4}(p_o^2 - p_x^2) + \frac{1}{2}p_x p_o(p_x - p_o) \\ &+ \frac{3}{4}(p_x - p_o) \end{aligned}$$

Then calculating the difference in PCS

$$\Delta PCS = \frac{1}{4}(p_x - p_o)(1 + 2p_x p_o - p_x - p_o) > 0 \quad (7)$$

We again see a term, $(p_x - p_o) > 0$ in Equation (7), which shows an improvement in PCS with AVC when X is the better population. Figure 1 illustrates the difference in the improvement with AVC for exponential and Bernoulli populations. Notice that the

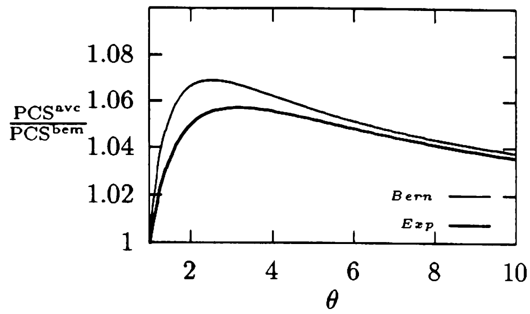

Figure 1: Exponential and Bernoulli Populations: $k = 2$, $v = 2$, $p_x = \frac{7}{8}$

maximum improvement with AVC occurs at a θ value between 2 and 3.

We have determined that the increase in PCS with AVC can be attributed to a subset of the total possible rank orders. Most rank orders result in the same conclusion with either BEM or AVC. For illustration consider $k = 2$. For any rank order with $Z_{[2]} > v^2/2$, AVC always makes the correct selection, however; for values of $Z_{[2]}$ close to $v^2/2$, BEM can make an incorrect selection. We refer to such rank orders as a *gain*. Similarly, for rank orders with $Z_{[2]} < v^2/2$, AVC always makes an incorrect selection; however, for such values of $Z_{[2]}$ close to $v^2/2$, BEM can make a correct selection. We refer to such a rank order as a *loss*.

Given a particular rank order, the probability of a gain (loss) is simply the fraction of the time AVC is correct and BEM is incorrect (BEM is correct and AVC is incorrect). The contribution to the overall PCS is then the probability of obtaining a particular rank order times the probability of the gain or loss with that rank order. Summing this contribution over all possible rank orders with a gain or loss provides an alternative way to quantify the difference in PCS between BEM and AVC. Calculations using this approach with exponential and continous uniform populations ($k = 2, v = 2, 3$; $k = 3, v = 2$) match results for finding the gain in PCS by taking $\text{PCS}^{\text{avc}} - \text{PCS}^{\text{bem}}$. Our conjecture is that we can systematically identify all rank orders with a gain or a loss and then group these gains and losses in some fashion to show the sum of the probabilties of the gains exceeds the sum of the probabilties of the losses.

5 EMPIRICAL RESULTS

Due to the difficulty in calculating analytical results for even small k and v, analytical results were only presented for a single k and v over a range of values for θ. The simulation results presented in this section, can easily show the relationship between PCS^{bem} and PCS^{avc} over a range of vectors and number of populations for a fixed θ. Mirroring the analytical results already presented, Figures 2 - 4 illustrate the improvement in PCS with AVC for up to $v = 50$ vectors for exponential and Bernoulli populations. All results are for 100,000 replications using a separate random number stream for each population and common random numbers for each set of distributions. The standard errors for PCS^{bem} and PCS^{avc} are approximately 0.0015. All runs were done using the LFC for BEM with $\theta = 1.2$. This was accomplished by fixing the parameter for one distribution and varying the other with increasing k to maintain a constant θ. For the exponential populations $\lambda = 1$ and for the Bernoulli populations $p_x = 1/2$.

Figure 2 clearly illustrates the improvement in PCS^{avc} over PCS^{bem} for 2 to 5 exponential populations. Likewise, Figure 3 shows similiar results for 2 and 3 Bernoulli populations. Figure 4 demonstrates the dependence of PCS^{avc} on the underlying population distributions.

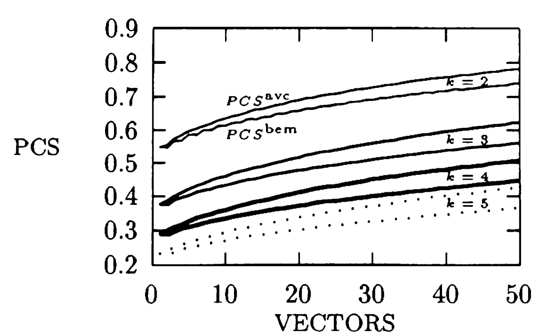

Figure 2: Exponential Populations, $\theta = 1.2$

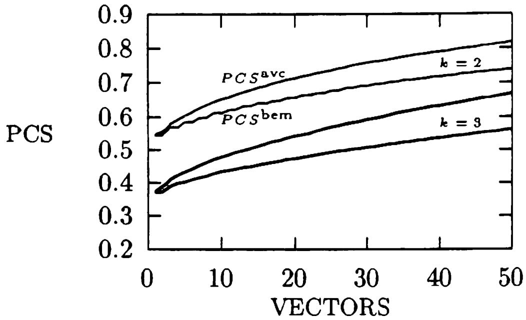

Figure 3: Bernoulli Populations, $\theta = 1.2$

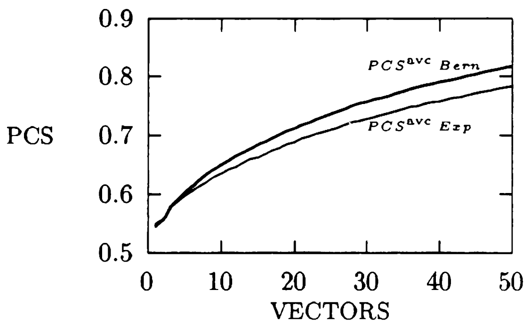

Figure 4: Exponential and Bernoulli Populations: $k = 2$, $\theta = 1.2$

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

6 POINT ESTIMATION

Although the focus of this paper is selecting the best population, a closely related problem is *estimation* of $p_j = \Pr\{\pi_j \text{ wins}\}$. If v vector observations are taken, then the natural point estimator associated with the BEM method is

$$\widehat{p}_j = \frac{Y_j}{v}$$

the fraction of wins out of v replications achieved by population j. It is well known that $\text{E}[\widehat{p}_j] = p_j$ and $\text{Var}[\widehat{p}_j] = p_j(1-p_j)/v$.

The natural point estimator associated with the AVC method is

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

the fraction of wins out of v^k psuedoreplications achieved by population j, where the pseudoreplications are obtained by taking all combinations of k observations with one from each population. Clearly, $\text{E}[\bar{p}_j] = p_j$. Unfortunately, the variance of $\bar{p}_j$ is quite complex. However, the following analysis shows that we should anticipate a variance reduction relative to $\text{Var}[\widehat{p}_j]$.

To simplify the development from here on, suppose that we are trying to estimate p_1. Let

$$h(X_{1i}, X_{2i}, \ldots, X_{ki}) = \prod_{\ell=2}^{k} \phi(X_{1i}, X_{\ell i})$$

an indicator function that will take the value 1 only if X_{1i} is largest among $(X_{1i}, X_{2i}, \ldots, X_{ki})$. Then equation (1) can be rewritten as

$$Z_1 = \sum_{a_1=1}^{v} \sum_{a_2=1}^{v} \cdots \sum_{a_k=1}^{v} h(X_{1a_1}, X_{2a_2}, \ldots, X_{ka_k}).$$

When written in this way we can verify that $\bar{p}_1 = Z_1/v^k$ is a *k-sample U-statistic* with kernel h of order $(1, 1, \ldots, 1)$ (Randles and Wolfe 1979, pp. 104–109). We can show that

$$\text{Var}[\widehat{p}_1] = \text{Var}[\bar{p}_1] + \text{E}\left[\text{Var}(\widehat{p}_1 | \widehat{F}_1, \widehat{F}_2, \ldots, \widehat{F}_k)\right]$$

where $\widehat{F}_j$ is the empirical cdf of the X_{ji}. This shows that $\text{Var}[\bar{p}_1] \leq \text{Var}[\widehat{p}_1]$. But how much variance reduction should we expect?

Since $\bar{p}_1$ is a U-statistic, its asymptotic distribution (as v goes to infinity) is known. Specifically,

$$\sqrt{kv}(\bar{p}_1 - p_1) \Longrightarrow \text{N}(0, \sigma^2_{\text{avc}})$$

where σ^2_{avc} depends on the distribution of the X_{ji}, but is much less complicated than the finite-sample variance.

To compare $\bar{p}_1$ to $\widehat{p}_1$, we note that

$$\sqrt{kv}(\widehat{p}_1 - p_1) \Longrightarrow \text{N}(0, \sigma^2_{\text{bem}})$$

where $\sigma^2_{\text{bem}} = kp_1(1-p_1)$. Therefore, a ratio of $\sigma^2_{\text{avc}}/\sigma^2_{\text{bem}} < 1$ implies that $\bar{p}_1$ is asymptotically superior to $\widehat{p}_1$. Table 1 shows one such calculation for exponential populations with $\lambda_1 = 1.0$, $\lambda_2 = \cdots = \lambda_k = 0.99$ (we have changed the problem slightly in that a smaller performance measure is considered better, allowing us to exploit the fact that the minimum of exponentially distributed random variables is also exponentially distributed).

The table shows a substantial variance reduction for $\bar{p}_1$, increasing to a limit of 50% as the number of populations to compare gets large (the limit of 1/2 holds for exponential distributions with any parameters). Thus, when there are more populations, making the comparison problem harder, the payoff is largest.

Table 1: Ratio of the Asymptotic Variances of Estimators for p_1

k	$\sigma^2_{\mathrm{avc}}/\sigma^2_{\mathrm{bem}}$
2	0.667
3	0.600
4	0.572
5	0.556
10	0.527
20	0.513
100	0.503
∞	0.500

7 CONCLUDING REMARKS

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.

Given fixed values of k, P^*, and θ^*, we conjecture that $\mathrm{PCS}^{\mathrm{avc}} \geq \mathrm{PCS}^{\mathrm{bem}}$. An interesting question is how many fewer replications are needed for an AVC-like procedure to perform just as well as BEM. Table 2 presents some preliminary comparisons of the minimum number of independent replications needed to achieve a given P^* for AVC and BEM. Values for BEM are taken from Bechhofer, Santner, and Goldsman (1995). The AVC values are from simulations (10,000 replications) using exponential populations under the LFC for BEM. As P^* increases and the difference between the best population and the other populations decreases, we see a more dramatic reduction in the number of vector observations needed with AVC to achieve the same P^*.

Table 2: Minimum v with AVC and BEM

		$P^* = .90$		$P^* = .95$	
k	θ^*	avc	bem	avc	bem
2	2.0	12	15	19	23
	1.2	116	199	128	327
3	2.0	21	29	33	42

ACKNOWLEDGMENTS

Dr. Nelson's work was partially supported by National Science Foundation Grant No. DMI-9622065.

REFERENCES

Bechhofer, R. E., S. Elmaghraby, and N. Morse. 1959. A single-sample multiple-decision procedure for selecting the multinomial event which has the highest probability. *Annals of Mathematical Statistics* 30:102–119.

Bechhofer, R. E., T. Santner, and D. Goldsman. 1995. *Design and analysis of experiments for statistical selection, screening and multiple comparisions.* New York: John Wiley & Sons, Inc.

Kesten, H., and N. Morse. 1959. A property of the multinomial distribution. *Annals of Mathematical Statistics* 30:120–127.

Randles, R. H., and D. A. Wolfe. 1979. *Introduction to the theory of nonparametric statistics.* New York: John Wiley.

AUTHOR BIOGRAPHIES

J.O. MILLER is a Major in the United States Air Force and a Ph.D. student in the Department of Industrial, Welding and Systems Engineering at The Ohio State University. His research interests include simulation, stochastic processes, and ranking and selection.

BARRY L. NELSON is an associate professor in the Department of Industrial Engineering and Management Sciences at Northwestern University. He is interested in the design and analysis of computer simulation experiments, particularly statistical efficiency, multivariate output analysis and input modeling. He is the simulation area editor for *Operations Research* and will be Program Chair for the 1997 Winter Simulation Conference.

CHARLES H. REILLY is a Professor and Chair in the Department of Industrial Engineering and Management Systems at the University of Central Florida. His principal research interests are generation methods for synthetic optimization problems and the relationships between problem generation methods and the outcomes of computational experiments. Dr. Reilly is a member of IIE, INFORMS, and ASEE; he is the Immediate Past Director of the Operations Research Division of IIE.