Multinomial Selection Problem: A Study of BEM and AVC Algorithms

HÉLCIO VIEIRA, Jr., KARL HEINZ KIENITZ, AND MISCHEL CARMEN NEYRA BELDERRAIN

Technological Institute of Aeronautics, São José dos Campos, Brazil

The two well-known and widely used multinomial selection procedures Bechhofer, Elmaghraby, and Morse (BEM) and all vector comparison (AVC) are critically compared in applications related to simulation optimization problems.

Two configurations of population probability distributions in which the best system has the greatest probability \( p_i \) of yielding the largest value of the performance measure and has or does not have the largest expected performance measure were studied.

The numbers achieved by our simulations clearly show that none of the studied procedures outperform the other in all situations. The user must take into consideration the complexity of the simulations and the performance measure probability distribution properties when deciding which procedure to employ.

An important discovery was that the AVC does not work in populations in which the best system has the greatest probability \( p_i \) of yielding the largest value of the performance measure but does not have the largest expected performance measure.

Keywords AVC; BEM; BG; KN; Multinomial selection problem; Simulation optimization.

Mathematics Subject Classification 62P30; 60A72.

1. Introduction

Procedures that are used to solve the multinomial selection problem (MSP) belong to a wide class of algorithms dedicated to perform the optimization of simulated systems. The problem of optimizing simulated systems is called simulation optimization (SO) and has received great attention in the literature in recent decades.

Given a system that needs to be optimized whose performance measure can only be evaluated through simulation, the purpose of SO is to find the values of the controlled parameters (decision variables) that optimize the performance measure. This article focuses on multinomial selection (MS) procedures applied to SO problems.

Received September 23, 2009; Accepted February 10, 2010

Address correspondence to Hélcio Vieira, Jr., Technological Institute of Aeronautics, Praca Marechal Eduardo Gomes, 50/12.228-900, São José dos Campos, SP, Brazil; E-mail: junior_hv@yahoo.com.br
MS procedures are concerned with the process of selecting the best of \( k \) systems, where best is defined as “the system most likely to have the largest performance measure in any comparison across all systems” (Miller et al., 1998, p. 460). Stated in other words, MS procedures select as best the system with greatest probability \( p_i \), where \( p_i (i = 1, 2, \ldots, k) \) and \( \sum_{i=1}^{k} p_i = 1 \) denotes the single trial probability of system \( i \) yielding the largest value of the performance measure.

The interest on MSP is due to, at least, these reasons:

- “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” (Miller et al., 1998, p. 477). When this situation occurs, the great number of procedures for selection of the best (selecting the system with the largest expected performance measure) cannot be used.
- “Since most real-life systems do not follow the ‘usual’ probability distributions, such multinomial selection procedures are seen to be very useful” (Chen, 1988b, p. 440). Even if the purpose is to select the system with the largest expected performance (selection of the best procedures), if the probability distribution of the systems outcomes is unknown or does not follow the “usual” ones, MS procedures can be the (only) solution.

Two MS procedures that are well known and widely used are the BEM (Bechhofer et al., 1959) and all vector comparisons (AVC; Miller et al., 1998).

The purpose of this article is to present a study of these two procedures (BEM and AVC) showing their strengths and weaknesses.

The article is organized as follows. The next section introduces the BEM and AVC algorithms. Section 3 provides an application oriented study of the performance of these two procedures and Sec. 4 presents our conclusions.

2. BEM and AVC Algorithms

2.1. BEM Algorithm

This algorithm was proposed by Bechhofer et al. (1959) to solve the MSP.

It is required that the decision maker defines the following values:

- The smallest value \( \theta^* \) of the rate \( \theta_{ij} = \frac{p_i}{p_j} \) that is worth detecting.
- The smallest acceptable value \( P^* \) of the probability of correct selection (PCS) of choosing the population with the greatest \( p_i \) when \( \theta_{ij} \geq \theta^* \), \( \forall i, j \).

In other words, the decision maker desires that:

\[
\Pr\{\text{correct decision} \mid \theta_{ij} \geq \theta^*\} \geq P^* \tag{1}
\]

The BEM procedure is basically a power test; i.e., it calculates the minimum number \( N_{\text{BEM}} \) of independent vector replications across all systems that should be collected.

Let

\[
a_1 = 2 \arcsin \left( \frac{\theta^*}{\theta^* + k - 1} \right) - 2 \arcsin \left( \frac{1}{\theta^* + k - 1} \right) \tag{2}
\]

\[
b_1 = 2 + 2 \sqrt{\frac{\theta^*}{(k - 1)(\theta^* + k - 2)}} \tag{3}
\]
\[ N_{\text{BEM}} = \left\lceil \frac{A^2 b_i}{2a_i^2} \right\rceil \]  

(4)

where \( \theta^* \) is the rate \( \theta_{ij} = \frac{p_i}{p_j} \) that is worth detecting, \( k \) is the number of systems under evaluation, \( \lceil x \rceil \) denotes the smallest integer equal to or greater than \( x \) and \( \Lambda \) is the smallest value that makes the equality (5) (true Bechhofer, 1954, p. 22).

\[
\int_{-\infty}^{\infty} [F(y + \Lambda)]^{k-1} [1 - F(y)]^{k-1} f(y) dy = P^*
\]

(5)

where \( F(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-x^2/2} dx \) and \( f(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \).

Let

\[
V_{ij} = \begin{cases} 
1, & \text{if } X_{ij} > X_{mj}, \\
0, & \text{if } X_{ij} < X_{mj},
\end{cases} \quad \forall m \neq i
\]

(6)

where \( X_{ij} \) is the \( j \)th performance measurement of population \( i \). If there are ties \( (X_{ij} = X_{mj}) \), randomly choose one of the tied systems as winner \( (V_{ij} = 1) \) and make \( V_{ij} = 0 \) for all the others.

BEM guarantees with probability \( P^* \) that the population with greatest mean (7) will be the one with greatest \( p_i \).

\[
\overline{V}_b = \frac{1}{N_{\text{BEM}}} \sum_{j=1}^{N_{\text{BEM}}} V_{bj}.
\]

(7)

### 2.2. **AVC Algorithm**

The AVC procedure was proposed by Miller et al. (1998) as an improvement to the method BEM.

“The key insight is that the formation of vector observations by replication number—\((X_{1j}, X_{2j}, \ldots, X_{kj})\)—is arbitrary; any vector formed with one output from each system has the same distribution. Thus, \( N \) replications from each system can form \( N^k \) vector observations” (Kim and Nelson, 2006, p. 528).

As stated by Kim and Nelson (2006), the AVC algorithm is almost identical to BEM, except that the performance measurement matrix \( X = [X_{ij}] \) is not formed by the \( N_{\text{BEM}} \) line vectors anymore. It will be augmented to a new matrix \( X_{\text{AVC}} \) through inclusion of all the possible \( (N_{\text{BEM}})^k \) combinations of the original \( N_{\text{BEM}} \) independent observations from the \( k \) systems.

In this fashion, Miller et al. (1998) claimed that for the same number of measurements \( N_{\text{BEM}} = N_{\text{AVC}} \), AVC will have a greater PCS than BEM. On the other hand, for a fixed PCS, AVC will require \( N_{\text{AVC}} < N_{\text{BEM}} \).

In order to obtain \( N_{\text{AVC}} \), the same equations of procedure BEM (Eqs. (2)–(5)) with \( \theta^* \) replaced by \( \theta^*_{\text{AVC}} \) are used:

\[
\theta^*_{\text{AVC}} = \frac{\frac{1}{k} \left( k - 1 \right)^{\frac{\theta^* - 1}{\theta^* + k - 1}} \sqrt{\frac{2k - 1}{k}}}{1 - \left( \frac{\theta^* - 1}{\theta^* + k - 1} \right)^{\frac{2k - 1}{k}}}
\]

(8)
where \( \delta^* \) is smallest value of the rate \( \theta_{ij} = p_i/p_j \) that is worth detecting and \( k \) is the number of systems under evaluation.

After the obtainment of the \( N_{AVC} \) independent vector replications across all systems, a “total of \( (N_{AVC})^k \) pseudoreplications formed by associating each \( X_{ij} \) \( (j = 1, 2, \ldots, k; \ j = 1, 2, \ldots, N_{AVC}) \), with all possible combinations of the remaining \( X_{ih} \) \( (l = 1, 2, \ldots, k; \ l \neq j; \ h = 1, 2, \ldots, N_{AVC}) \)” (Miller et al., 1998, p. 462) will be constructed. These pseudoreplications are the ones that will be used in Eqs. (2)–(5).

3. Numerical Study

This section provides BEM and AVC performance analysis for two qualitatively different configurations of population probability distributions:

- Configuration A: a family of distributions in which the best system has the greatest probability \( p_i \) of yielding the largest value of the performance measure and has the largest expected performance measure.
- Configuration B: a family of distributions in which the best system has the greatest probability \( p_i \) of yielding the largest value of the performance measure but does not have the largest expected performance measure.

3.1. Configuration A

This family of distributions includes the exponential, continuous uniform, gamma, and Bernoulli distributions. They have been studied by Miller et al. (1998) by comparing the performance of their proposal (the AVC algorithm) with the BEM algorithm.

We replicated the simulations done by Miller et al. (1998) and reached values quite close to the ones they achieved. Table 1 reproduces the results of their work.

As stated by the authors of AVC, “the reduction in \( N \) goes from roughly 34% at \( k = 2 \) to 44% at \( k = 5 \). So the advantages of AVC over BEM appear greater for more challenging MSPs” (p. 477). We deepen this study taking into consideration the computational time required to perform both algorithms.

For \( k = 2 \), \( \theta^* = 1.010, P^* = 0.750 \) and considering the time to perform one vector comparison to be \( 10^{-4} \) s: BEM takes 1.837 s (18,371 comparisons) to select the best performer and AVC takes 14,813.324 s/3.914 h (12, 171; \( 128 \), 133, 241 comparisons) to do the same. If the time to simulate one replication of the system is greater than 2.388 s, AVC will outperform BEM (considering computational time). Because one of the hypotheses of simulation optimization is that it is expensive to generate each simulation replication (Fu, 2006, p. 577), AVC seems to be a good choice to solve the MSP.

Another more realistic example should be \( k = 8 \), \( \theta^* = 1.200, P^* = 0.900 \) and considering the same time to perform one vector comparison (\( 10^{-4} \) s): BEM would take 0.185 s (1,852 comparisons) to select the best performer and AVC would take 1.091 \( \cdot 10^{20} \) s/3.031 \( \cdot 10^{15} \) h (1, 011 \( \cdot 10^{8} \) comparisons) to do the same. In this case, AVC would outperform BEM only if the time to simulate one replication of the system would be greater than 3.605 \( \cdot 10^{13} \) h.

The numbers shown above indicate that AVC will outperform BEM in simulations in which the number of systems is low. If the number of systems is medium/large, it will outperform BEM only if the complexity of the simulation (time to perform one replication) is very large.
Both BEM and AVC are open procedures; i.e., it is known in advance the exact number of measurements the procedure will take. We decided to compare their performances with a closed sequential procedure—a procedure in which it is not known in advance the exact number of measurements needed to stop it. At first, we tried to perform the comparison with the procedure “R” proposed by Chen (1988a). Unfortunately, the available tables for R implementation with a $\theta^* = 1.200$ display a maximum PCS of 69%, which is a much smaller PCS than the smallest one achieved by AVC in Table 2 (76%), not allowing a fair comparison. Another closed sequential procedure that is well known and respected is the BG (Bechhofer and Goldsman, 1985). Bechhofer et al. (1995) stated that “procedure BG is superior, over
a broad range of practical $(t, \Theta^*, P^*)$-values, to all procedures that have thus far been proposed for the multinomial selection problem.” For this reason, the closed procedure used for comparison with BEM and AVC was the BG.

Because the best in this configuration also has the largest mean, we decided to perform an additional comparison among BEM and AVC performances with a selection of the best algorithm called KN (Kim and Nelson, 2001). Remember that selection of the best algorithms aims to find the system with greatest expected performance measure.

In its two initial lines, Table 2 presents the results of the last line of Table 4 of Miller et al. (1998) and in the third and fourth lines the results of our simulations with procedures BG and KN. The probability distributions of the performance measure data studied were the exponential, continuous uniform, gamma, and Bernoulli. The achieved PCS is the percentage of times the algorithm found the true best in $n$ independent macroreplications. The number of simulations for BEM and AVC is fixed ($N \cdot k = 50 \cdot 2 = 100$). It was used as desired PCS for BG and KN the PCS achieved by AVC in Miller et al.’s (1998) studies. For example, the desired PCS used for BG and KN was 76.14% (with exponential distribution). The number of simulations for BG and KN was the average number of simulations needed to achieve the desired PCS.

The KN algorithm was the best performer in almost all cases (except with uniform distributions in which the BG performed better than the KN).

Observe that although the KN procedure was designed to deal with normally independently identically distributed (iid) populations (and the data analyzed do not have this distribution), it had a PCS greater than the desired one and also had a lower expected number of simulations than both BEM and AVC. This result confirms what Dudewicz (1971) found: the MS procedures always require much larger sample size than selection of the best procedures.

This analysis suggests that when the best has the largest $p_i$ and the largest mean, it is better to use the closed sequential algorithm called BG or the selection of the best procedure called KN because they provided a greater PCS with a lower number of simulations than both BEM and AVC did.

### 3.2. Configuration B

A situation that does not allow the use of procedures for selecting the best occurs when the best system has the greatest probability $p_i$ of yielding the largest value of the performance measure but does not have the largest expected performance measure. Because all the studies performed by Miller et al. (1998) when comparing AVC with BEM were on systems with distribution of probability described in the Configuration A subsection, we decided to perform additional studies in this different configuration. The probability distribution given in (9) and (10) was used for our study.

\[
\begin{align*}
\Pr[X_{i,j} = 10] &= \hat{p} \\
\Pr[X_{i,j} = 2] &= \frac{1 - \hat{p}}{2} \\
\Pr[X_{i,j} = 0] &= \frac{1 - \hat{p}}{2}
\end{align*}
\]
Table 3
PCS (mean time in seconds to perform one macroreplication) achieved for several different designs of Configuration B

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th><em><em>$P^</em> = 0.75$</em>*</th>
<th></th>
<th></th>
<th></th>
<th><em><em>$P^</em> = 0.90$</em>*</th>
<th></th>
<th></th>
<th></th>
<th><em><em>$P^</em> = 0.95$</em>*</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\theta^*$</td>
<td>AVC</td>
<td>BEM</td>
<td>BG</td>
<td>AVC</td>
<td>BEM</td>
<td>BG</td>
<td>AVC</td>
<td>BEM</td>
<td>BG</td>
<td>AVC</td>
</tr>
<tr>
<td>2</td>
<td>1,000</td>
<td>1.200</td>
<td>0.723</td>
<td>0.787</td>
<td>0.779</td>
<td>0.870</td>
<td>0.907</td>
<td>0.928</td>
<td>0.927</td>
<td>0.958</td>
<td>0.950</td>
<td>(6.133 · 10^{-2})</td>
</tr>
<tr>
<td></td>
<td>2,000</td>
<td></td>
<td>(3.240 · 10^{-2})</td>
<td>(7.300 · 10^{-3})</td>
<td>(1.400 · 10^{-3})</td>
<td>(2.178 · 10^{-1})</td>
<td>(3.220 · 10^{-2})</td>
<td>(5.000 · 10^{-3})</td>
<td>(6.133 · 10^{-2})</td>
<td>(5.440 · 10^{-2})</td>
<td>(7.800 · 10^{-2})</td>
<td>(9.946)</td>
</tr>
<tr>
<td>3</td>
<td>200</td>
<td>1.200</td>
<td>0.100</td>
<td>0.780</td>
<td>0.800</td>
<td>0.020</td>
<td>0.920</td>
<td>0.900</td>
<td>N/A</td>
<td>0.945</td>
<td>0.950</td>
<td>(6.000 · 10^{-3})</td>
</tr>
<tr>
<td></td>
<td>2,000</td>
<td></td>
<td>(2.717 · 10^{-4})</td>
<td>(5.642 · 10^{-4})</td>
<td>(1.187 · 10^{-4})</td>
<td>(2.400 · 10^{-3})</td>
<td>(2.000 · 10^{-3})</td>
<td>(3.643 · 10^{-4})</td>
<td>(6.000 · 10^{-3})</td>
<td>(3.000 · 10^{-3})</td>
<td>(5.483 · 10^{-4})</td>
<td>(9.715)</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>1.200</td>
<td>0.625</td>
<td>0.825</td>
<td>0.765</td>
<td>0.670</td>
<td>0.915</td>
<td>0.910</td>
<td>N/A</td>
<td>0.945</td>
<td>0.950</td>
<td>(8.770 · 10^{-2})</td>
</tr>
<tr>
<td></td>
<td>2,000</td>
<td></td>
<td>(1.370 · 10^{-2})</td>
<td>(2.300 · 10^{-3})</td>
<td>(4.097 · 10^{-4})</td>
<td>(1.538 · 10^{-1})</td>
<td>(4.700 · 10^{-3})</td>
<td>(9.643 · 10^{-4})</td>
<td>(5.032 · 10^{-1})</td>
<td>(7.200 · 10^{-3})</td>
<td>(1.100 · 10^{-3})</td>
<td>(9.715)</td>
</tr>
<tr>
<td>5</td>
<td>100</td>
<td>1.200</td>
<td>0.470</td>
<td>0.820</td>
<td>0.755</td>
<td>0.380</td>
<td>0.910</td>
<td>0.905</td>
<td>N/A</td>
<td>0.945</td>
<td>0.950</td>
<td>(2.407 · 10^{-1})</td>
</tr>
<tr>
<td></td>
<td>2,000</td>
<td></td>
<td>(5.953 · 10^{-3})</td>
<td>(3.700 · 10^{-3})</td>
<td>(8.126 · 10^{-4})</td>
<td>(1.344 · 10^{1})</td>
<td>(7.300 · 10^{-3})</td>
<td>(1.600 · 10^{-3})</td>
<td>(1.553 · 10^{-1})</td>
<td>(5.020 · 10^{-2})</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>200</td>
<td></td>
<td>(7.980 · 10^{-2})</td>
<td>(3.200 · 10^{-3})</td>
<td>(6.410 · 10^{-3})</td>
<td>(1.010 · 10^{-2})</td>
<td>(2.500 · 10^{-3})</td>
<td>(1.330 · 10^{-2})</td>
<td>(3.400 · 10^{-3})</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

N/A indicated numbers not available due to computer lack of memory; $P^*$ is the desired PCS; bold numbers denote cells that did not achieve $P^*$. 
\[
\begin{align*}
\Pr \{X_{i,j} = 8 \} &= \bar{p} \\
\Pr \{X_{i,j} = 7 \} &= 1 - \bar{p}
\end{align*}
\]  

(10)

where \( \bar{p} = \theta^* \bar{p} \), \( \bar{p} = \frac{1}{k + \theta^* - 1} \), \( k \) is the number of systems and \( \theta^* \) is the smallest rate \( \theta_{ij} = \frac{p_i}{p_j} \); the decision maker wants to detect. The values of \( \bar{p} \) and \( \bar{p} \) calculated in this fashion produce a situation known as least favorable configuration (LFC)—the configuration where PCS is a minimum over all configurations with \( \theta \geq \theta^* \) (Gibbons et al., 1977).

The practical use of formulae (9) and (10) can be exemplified as follows. Let \( k = 3 \) and \( \theta^* = 1.200 \), then the probability \( p_i \) of achieving the largest value of the performance measure and the expected value of the random variables will be

\[ p_1 = 0.375, \quad p_2 = p_3 = 0.313 \quad \text{and} \quad E[X_1] = 4.375, \quad E[X_2] = E[X_3] = 7.313 \]

System 1 has the largest \( p_i \) but does not have the largest mean.

The results of our simulations are described in Table 3, where the results are the average time in seconds to perform one macroreplication and the average PCS achieved on \( n \) macroreplications for a number of systems \( k = 2, 3, 4, \) and \( 5 \) with \( \theta^* = 1.200 \) and \( 2.000 \). The number of macroreplications used was inversely proportional to the number of systems analyzed because the time necessary to complete each AVC macroreplication increases exponentially with \( k \). The algorithms were implemented in MATLAB on a computer with AMD Turion X2 Dual-Core 2.1 GHz CPU and 4 GB of RAM.

Table 3 Analysis of shows that:

- AVC was not able to achieve the required PCS at \( \theta^* = 1.200 \) and was able to achieve it at \( \theta^* = 2.000 \) only with \( k = 2 \).
- The performance of the AVC algorithm decays with the number of systems \( k \). For example, the sequence of achieved PCS for a desired PCS \( P^* = 0.750 \), \( \theta^* = 2.000 \), and \( k = 2, 3, 4, \) and \( 5 \) was, respectively, 0.749, 0.625, 0.470, and 0.230.
- The implementation of AVC in real-world problems can be difficult. For example, for \( P^* = 0.950 \), \( \theta^* = 1.200 \), and \( k = 3 \), the number of comparisons to be performed is \( 390^3 = 59,319,000 \), the size of the matrix that stores the vector \( V_{AVC} \) of comparisons is \( 59,319,000 \cdot 3 = 177,957,000 \), and considering each cell of the mentioned matrix stores an integer (8 KB), it would be necessary to have 1.357 GB of memory available to run the algorithm. This fact can be noticed in Table 3 by the percentage (37.500%) of AVC simulations that was not completed due to computer lack of memory.
- BEM and BG procedures performed well in all situations.
- The BG procedure had the smallest mean time to complete one macroreplication in all different designs of configuration B.

4. Conclusions

A study of the procedures BEM and AVC was presented showing their strengths and weaknesses. Section 2 presented a literature review of BEM and AVC algorithms and Sec. 3 provided a study of these two procedures.
Our study was divided in two parts:

- The first was dedicated to problems in which the best system has the greatest probability $p_i$ of yielding the largest value of the performance measure and has the largest expected performance measure: our results showed that AVC will outperform BEM in simulations in which the number of systems is low. If the number of systems is medium/large, it will outperform BEM only if the complexity of the simulation (time to perform one replication) is very large. Furthermore, there are indications that both BG and KN algorithms outperform both AVC and BEM because they achieved greater PCS with a lower number of simulations.

- The second part was dedicated to problems in which the best system has the greatest probability $p_i$ of yielding the largest value of the performance measure but does not have the largest expected performance measure. Our results showed that
  - AVC was not able to achieve the required PCS at $\theta^* = 1.200$ and was able to achieve it at $\theta^* = 2.000$ only with $k = 2$.
  - The performance of the AVC algorithm decays with the number of systems $k$.
  - The implementation of AVC in real-world problems can be difficult.
  - BEM and BG procedures performed well in all situations.
  - BG was the fastest tested algorithm.

The numbers achieved by our simulations clearly show that none of the studied procedures outperform the others in all situations. The user must take into consideration the complexity of the simulations and the performance measure probability distribution proprieties when deciding which procedure to employ.

An important discover was that the AVC algorithm does not work in the Configuration B class of problems.

Future efforts should focus on finding out why AVC does not work in populations in which the best system has the greatest probability $p_i$ of yielding the largest value of the performance measure but does not have the largest expected performance measure.

References


