Proceedings of the 1984
Winter Simulation Conference
S. Sheppard, U. Pooch, D. Pegden (eds.)



# ON SELECTING THE BEST OF K SYSTEMS: AN EXPOSITORY SURVEY OF INDIFFERENCE-ZONE MULTINOMIAL PROCEDURES

David Goldsman
School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332

## Abstract

We investigate the problem of selecting the 'best' one of k arbitrary systems or alternatives. Consider one observation from each of the k systems. By 'best,' we mean that system which has the highest probability of yielding the 'most desirable' of the k observations. The term 'most desirable' is defined according to some criterion of goodness determined by the experimenter. We show that this problem can be formulated as a multinomial selection problem. Hence, multinomial selection procedures are, in a sense, nonparametric procedures. An up-to-date survey of 'indifference-zone' multinomial procedures is given.

## 1. Introduction

Consider k different competing populations (or systems or alternatives, etc). A natural question to ask is: Which of these k systems is 'best?' By 'best' system, we could informally mean, e.g.

- that one of k inventory policies which maximizes profit,
- that one of k scales which is the most precise, or
- that one of k computer systems which has the greatest availability.

Thus, 'best' can take on a variety of meanings depending on the practical problem at hand.

Denote the k populations (sources of observations) as $\pi_1, \pi_2, \ldots, \pi_k$, respectively. Suppose we take independent vector-observations $(X_1, X_2, \ldots, X_k)$, where $X_i$ is from $\pi_i$, $i=1,\ldots,k$. Further, for $i=1,\ldots,k$, denote:

$$p_i = P\{X_i \text{ is the 'most desirable' of } X_1, X_2, \ldots, X_k\}.$$

The term 'most desirable' must be defined according to some criterion of goodness determined by the experimenter. Assume that nothing is known beforehand concerning the values of the $p_i$'s. Obviously, that $\pi_i$ associated with the largest of the $p_i$'s is the population which has the *highest probability of yielding the 'most desirable' observation* (of those observations from the k-vector). In this paper, our goal will be to find that $\pi_i$ associated with the largest of the $p_i$'s. We refer to that $\pi_i$ as the 'best' population.

In order to motivate this definition, consider a simple example. Let A and B be two (s,S) inventory policies. Profit is taken to be the criterion of desirability. Suppose that

$$\text{Profit from A} = 1000 \text{ w. p. } 0.001$$
$$= 0 \quad \text{ w. p. } 0.999$$

and

$$\text{Profit from B} = 0.999 \text{ w. p. } 1.$$

Clearly, $E(\text{Profit from A}) = 1 > 0.999 = E(\text{Profit from B})$; i.e., A gives the higher *average* profit. However, $P\{\text{Profit from B} > \text{Profit from A}\} = 0.999$; so B gives the higher profit *almost all of the time*. For this reason, the experimenter might justifiably consider policy B to be better than policy A.

It is therefore meaningful to consider as 'best' the policy which will most likely produce the 'most desirable' observation.

The goal of finding the 'best' population can be viewed as that of finding that cell of a k-nomial distribution with the largest underlying probability. Suppose that we take one observation from each of the k populations. Award a <u>one</u> (a 'success') to the $\pi_i$ corresponding to the 'most desirable' of these k observations (use randomization if necessary.) Award a <u>zero</u> to the remaining k-1 $\pi_i$'s. This is clearly the same as taking an observation from a multinomial distribution with cell probabilities $p_1, \ldots, p_k$.

Thus, the problem of finding the 'best' one of k arbitrary populations can be formulated as a problem of finding that one category of a k-nomial distribution with the highest underlying 'success' probability. This implies that any procedure which finds the multinomial cell associated with the largest probability is a *nonparametric* procedure. Since most real-life systems do not follow one of the 'usual' probability distributions, such nonparametric procedures are seen to be very useful. We group these nonparametric procedures under the heading of *multinomial selection procedures*. Additional motivation for the above arguments can be found in Bechhofer and Sobel (1958).

In Section 2 of this paper, we give a brief summary of the pertinent notation and terminology. In Section 3, some of the existing selection procedures are presented.

## 2. Background

We introduce notation and terminology which will be useful for investigating the problem of finding the multinomial cell which has the largest cell probability. Suppose that we take independent observations sequentially from a k-nomial distribution with cell probabilities $p_1, p_2, \ldots, p_k$, $p_i \geq 0$, $\Sigma p_i = 1$, until some stopping criterion (several of which will be given in the sequel) is met. Most of the procedures which we will study take observations (up to a limit, perhaps) until one cell has 'significantly more' successes than the other cells (whereupon the stopping criteria call for the termination of sampling).

Denote $x_{i,t}$ as the number of observations from cell i after t multinomial observations (or 'stages') have been taken, i=1,...,k; t=1,2,... Further, denote $p_{[1]} \leq p_{[2]} \leq \cdots \leq p_{[k]}$ as the ordered $p_i$'s and $x_{[1],t} \leq \cdots \leq x_{[k],t}$ as the ordered $x_{i,t}$'s. Assume that we have no a priori knowledge as to how the $p_{[i]}$'s are paired with the multinomial cells.

Our goal is to select as best that cell which is associated with $p_{[k]}$, the largest probability. If the cell corresponding to $p_{[k]}$ is actually chosen, we say that a *correct selection* (CS) has been made. Also, it is desired that the *probability of correct selection* (P{CS}) be at least $P^*$ whenever $\theta^* p_{[k-1]} \leq p_{[k]}$, where $\{P^*, \theta^*\}$ is pre-specified by the user (with $1 < \theta^* < \infty$ and $1/k < P^* < 1$). Define $\Omega_{\theta^*} = \{\underline{p} | \theta^* p_{[k-1]} \leq p_{[k]}\}$. We call $\Omega_{\theta^*}$ the *preference zone* and $\Omega^c_{\theta^*}$ the *indifference-zone*. [Multinomial procedures such as those to be considered below fall under the classification of so-called *indifference-zone selection procedures*. Another rich family of selection procedures employs the so-called *subset approach*; this approach will not be emphasized here. The reader should refer to Gupta and Panchapakesan (1979) for material concerning the indifference-zone and subset methodologies.]

We will consider the following configuration of $p_{[i]}$'s as a benchmark for comparison among procedures:

$$p_{[k]} = \theta^* p_{[i]}, \quad i=1,\ldots,k-1 \qquad \text{(SC)}$$

I.e., $p_{[i]} = (k-1+\theta^*)^{-1}$, $i=1,\ldots,k-1$; $p_{[k]} = \theta^*(k-1+\theta^*)^{-1}$. SC stands for *slippage configuration* (with slippage factor $\theta^*$). For some sampling procedures [cf: Bechhofer, Elmaghraby, and Morse (1959)], this configuration of $p_{[i]}$'s minimizes the P{CS} over $\underline{p} \in \Omega_{\theta^*}$. In this case, the SC is called the *least-favorable configuration* (LFC). Informally, the LFC can be viewed as a 'worst case' configuration (given that $\underline{p} \in \Omega_{\theta^*}$). It is not known whether the SC is the LFC for all of the multinomial procedures to be presented in the sequel. However, this is a reasonable conjecture, and we shall treat the SC as if it were, indeed, the LFC. Since we desire $P\{CS\} \geq P^*$ for all configurations $\underline{p} \in \Omega_{\theta^*}$, then (assuming the conjecture to be true) we can equivalently require that $P\{CS|\underline{p} = SC\} \geq P^*$.

Another interesting configuration is the *equal probability configuration* (EPC), where $p_i = 1/k$ for all i. Of course, the term 'correct selection' is now meaningless; but the EPC is useful as another benchmark in that we would expect such a configuration to maximize a multinomial procedure's expected sample size (i.e., the expected number of multinomial observations needed before the termination criterion is met).

Denote the sample size for a procedure $\mathcal{P}$ as $S_{\mathcal{P}}$. $E(S_{\mathcal{P}})$ is the expected sample size. Ideally, we wish to find a procedure which guarantees $P\{CS|\underline{p} = SC\} \geq P^*$ but which is also parsimonious with observations; that is, $E(S_{\mathcal{P}}|\underline{p} = SC)$ and $E(S_{\mathcal{P}}|\underline{p} = EPC)$ should be 'low.'

## 3. Some Multinomial Procedures

In this section, we concentrate on indifference-zone procedures for selecting the multinomial cell which has the largest probability. Recall that when using the indifference-zone approach, the experimenter must pre-specify two constants, $P^*$ and $\theta^*$. The procedures to be discussed below insure that:

$$P\{CS|\theta^* p_{[k-1]} \le p_{[k]}\} \ge P^*, \qquad \text{(PR)}$$

where PR stands for *probability requirement*. For all of these procedures, we establish the following conventions:

- All observations are independent multinomial observations.
- $T$ is defined to be the stage at which the procedure in question terminates sampling. $T$ may be a random variable.
- We will choose as best that cell corresponding to $x_{[k],T}$ (using randomization if necessary).

### 3.1 A single-sample procedure

The first procedure we consider is that of Bechhofer, Elmaghraby, and Morse (1959), denoted as $\mathcal{P}_{BEM}$.

Procedure $\mathcal{P}_{BEM}$:

1. Specify $k$, $P^*$, and $\theta^*$.

2. Take $N_{BEM}$ observations, where $N_{BEM} = N_{BEM}(k,P^*,\theta^*)$ is to be found in the tables of BEM (1959). $N_{BEM}$ is the number of multinomial observations which must be taken in order to satisfy the PR.//

Remarks 3.1:

1. Kesten and Morse (1959) prove that the SC is the LFC.

2. In $\mathcal{P}_{BEM}$, the number of observations we take is *fixed* at $N_{BEM}$. For this reason, the procedure is said to be a *fixed-sample* or *single-sample procedure*.

Example 3.1.1:

Suppose that $k = 3$ and that we specify $P^* = 0.75$ and $\theta^* = 3$. Use Table 1 [abstracted from BEM (1959)] in order to find $N_{BEM}$.

| $N_{BEM}$ \ $\theta^*$ | 1.1 | 1.5 | 2.0 | 3.0 |
|---|---|---|---|---|
| 1 | .355 | .429 | .500 | .600 |
| 2 | .355 | .429 | .500 | .600 |
| 3 | .362 | .464 | .563 | .696 |
| 4 | .367 | .484 | .594 | .734 |
| 5 | .370 | .496 | .617 | .769 |
| 6 | .374 | .515 | .646 | .804 |

Table 1 (for $\mathcal{P}_{BEM}$): $P\{CS|k=3,\underline{p}=LFC\}$ for selected $\theta^*$ and $N_{BEM}$

Reading down the $\theta^* = 3.0$ column, we see that $N_{BEM} = 5$ is the smallest value of $N_{BEM}$ which achieves the PR (Note that owing to the discrete nature of the multinomial distribution, $\mathcal{P}_{BEM}$ overshoots slightly the desired $P^* = 0.75$.) If we take 5 observations, the PR will be guaranteed.//

### 3.2 $\mathcal{P}_{BK}$, an improved version of $\mathcal{P}_{BEM}$

By considering the following example, it becomes apparent that $\mathcal{P}_{BEM}$ is sometimes wasteful with observations.

Example 3.2.1:

Suppose that $k = 2$, $N_{BEM} = 7$, and $\underline{x}_5 = (x_{1,5},x_{2,5}) = (4,1)$. It is obviously impossible to terminate sampling with $x_{1,T} \le x_{2,T}$; there is no chance for cell 2 to be chosen. Since cell 1 is guaranteed to be the victor regardless of the remaining two observations, we should stop sampling at $T = 5$.//

With this example in mind, we compare two procedures, the latter due to Bechhofer and Kulkarni (1984).

Procedure $\mathcal{P}_{BEM}$:

1. Specify $k$ and $N$.

2. Take $N$ observations.//

Procedure $P_{BK}$:

1. Specify k and N.

2. Take observations until either

2-A. The stage t = N or

2-B. $x_{[k],t} - x_{[k-1],t} = N-t$ (Stop sampling if the cell(s) with the second largest number of observations can only *tie* the cell corresponding to $x_{[k],t}$, even if the remaining N-t observations are taken.)//

Remarks 3.2:

1. Note that $P_{BK}$ is a *sequential* procedure.

2. It is clear that $E(S_{P_{BK}}) \le E(S_{P_{BEM}})$.

3. Bechhofer and Kulkarni show that $P\{CS|P_{BEM}\} = P\{CS|P_{BK}\}$. Thus, $P_{BK}$ *preserves the P{CS}* of the less parsimonious procedure, $P_{BEM}$; we can use the more efficient $P_{BK}$ with no loss of P{CS}.

Example 3.2.2:

Let k = 3, $P^* = 0.75$, and $\theta^* = 3$. Then $E(S_{P_{BEM}}) = N_{BEM} = 5$. It is straightforward (but tedious) to show that $E(S_{P_{BK}}) = 3.95$ in the LFC.//

3.3 A sequential procedure due to Ramey and Alam (1979)

Procedure $P_{RA}$:

1. Specify k, $P^*$, $\theta^*$.

2. Take observations until either

2-A. $x_{[k],t} = N$ or

2-B. $x_{[k],t} - x_{[k-1],t} = r$, where r and N are determined by k, $P^*$, and $\theta^*$, are to be found in tables for certain k, $P^*$, and $\theta^*$ (NB: *See Remarks below.*)//

Remarks 3.3:

1. Ramey and Alam's tables actually contain a number of errors; the user is advised to consult Bechhofer and Goldsman (1984a).

2. The number of observations which $P_{RA}$ takes is *bounded* by kN-k+1.

3. It is not known whether the SC is the LFC for all k for $P_{RA}$, but we will make the reasonable assumption that this is the case.

4. r and N are determined in such a way that the PR is satisfied and $E(S_{P_{RA}}|\underline{p} = LFC)$ is minimized over the (r,N) grid.

5. $P_{RA}$ is not directly comparable to $P_{BK}$. However, it seems that for most choices of k, $P^*$, and $\theta^*$, $P_{RA}$ requires fewer observations (on the average) than $P_{BK}$.

Example 3.3.1:

Again, let k = 3, $P^* = 0.75$, and $\theta^* = 3$. We abstract a small portion of the necessary (corrected) tables for $P_{RA}$ from Bechhofer and Goldsman (1984a).

| $P^*$ | $\theta^*$ | r | N | P{CS} | E(S) |
|---|---|---|---|---|---|
| .75 | 3.0 | 2 | 3 | .796 | 3.68 |
| .75 | 2.4 | 2 | 5 | .760 | 4.70 |
| .75 | 2.0 | 4 | 5 | .756 | 8.80 |
| .75 | 1.6 | 4 | 12 | .757 | 18.24 |

Table 2 (for $P_{RA}$):
$P\{CS|k=3,\underline{p}=LFC\}$, $E(S|.)$ for various $P^*$, $\theta^*$

We see that if r = 2 and N = 3 are chosen, a P{CS} of 0.796 will be achieved in the conjectured LFC. The overshoot of the $P\{CS|\underline{p} = LFC\}$ (0.796 vs. $P^* = 0.75$) is again due to the discrete nature of the problem. Further, *in this example*, $E(S_{P_{RA}}|\underline{p} = LFC) = 3.68 < 3.95 = E(S_{P_{BK}}|\underline{p} = LFC)$.//

3.4 An unbounded sequential procedure

Bechhofer, Kiefer, and Sobel (1968) give an *unbounded* (or *open*) sequential procedure which satisfies the PR.

Procedure $P_{BKS}$:

1. Specify k, $P^*$, $\theta^*$.

2. Take observations until

$$\sum_{i=1}^{k-1} (1/\theta^*)^{x_{[k],t} - x_{[i],t}} \le (1-P^*)/P^*.//$$

Remark 3.4: BKS show that the SC is the LFC for this procedure.

Example 3.4.1:

Let $k = 3$, $P^* = 0.75$, $\theta^* = 3$. Consulting the appropriate tables in Bechhofer and Goldsman (1984b), we immediately find that $P\{CS|\underline{p} = LFC\} = 0.842$ (.0004) and $E(S_{\mathcal{P}_{BKS}}|\underline{p} = LFC) = 4.526$ (.051). These results are Monte Carlo estimates obtained via simulation; the entries in parentheses are the accompanying standard errors. The results are fairly precise, as can be seen by the small standard errors.//

## 3.5 $\mathcal{P}_{BG}$, an improved version of $\mathcal{P}_{BKS}$

As in the above example, it turns out that $\mathcal{P}_{BKS}$ frequently yields $P\{CS|\underline{p} = LFC\} >> P^*$. This extra $P\{CS\}$ is at the cost of unnecessary observations. Bechhofer and Goldsman (1984b) give a procedure which decreases the attained $P\{CS\}$ to a level slightly greater than $P^*$, but which also saves observations.

Procedure $\mathcal{P}_{BG}$:

1. Specify $k$, $P^*$, $\theta^*$.

2. Take observations until either

2-A. $\sum_{i=1}^{k-1} (1/\theta^*)^{x_{[k],t}-x_{[i],t}} \le (1-P^*)/P^*$ or

2-B. the stage $t = N_{BG}$, where $N_{BG}$ is determined by $k$, $P^*$, $\theta^*$, and is to be found in Bechhofer and Goldsman's tables for certain values of $k$, $P^*$, $\theta^*$.//

Remarks 3.5:

1. $N_{BG}$ is chosen as the smallest upper bound on the total number of observations such that the PR is satisfied.

2. Unlike $\mathcal{P}_{BKS}$, $\mathcal{P}_{BG}$ is *bounded*.

3. It is not known whether the SC is the LFC for this procedure, but we so conjecture.

4. $\mathcal{P}_{BG}$ is not directly comparable to $\mathcal{P}_{BK}$ or to $\mathcal{P}_{RA}$. For many choices of $k$, $P^*$, $\theta^*$, it turns out that $\mathcal{P}_{BG}$ requires fewer observations (on the average) than $\mathcal{P}_{BK}$. The user should consult the relevant tables when designing an experiment.

Example 3.5.1:

Let $k = 3$, $P^* = 0.75$, $\theta^* = 3$. We abstract a small portion of the necessary tables for $\mathcal{P}_{BG}$ from Bechhofer and Goldsman (1984b):

| $P^*$ | $\theta^*$ | $N_{BG}$ | $P\{CS\}$ | $E(S)$ |
|---|---|---|---|---|
| .75 | 3.0 | 5 | .757 | 3.48 |
| .75 | 2.4 | 8 | .760 | 5.59 |
| .75 | 2.0 | 13 | .751 | 8.18 |
| .75 | 1.6 | 32 | .752 | 17.80 |

Table 3 (for $\mathcal{P}_{BG}$): $P\{CS|k=3,\underline{p}=LFC\}$, $E(S|.)$ for various $P^*$, $\theta^*$

We must choose $N_{BG} = 5$ with the resulting $P\{CS|\underline{p} = LFC\} = 0.757$ and $E(S_{\mathcal{P}_{BG}}|\underline{p} = LFC) = 3.48$.//

## 3.6 $\mathcal{P}_{BG2}$, an augmented version of $\mathcal{P}_{BG}$

We now employ the same device as was used in $\mathcal{P}_{BK}$; viz., stop sampling when the cell in second place only has a chance to tie.

Procedure $\mathcal{P}_{BG2}$:

1. Specify $k$, $P^*$, $\theta^*$.

2. Take observations until

2-A. $\sum_{i=1}^{k-1} (1/\theta^*)^{x_{[k],t}-x_{[i],t}} \le (1-P^*)/P^*$ or

2-B. $t = N_{BG2} = N_{BG}$, where $N_{BG}$ is from $\mathcal{P}_{BG}$ or

2-C. $x_{[k],t}-x_{[k-1],t} = N_{BG2}-t$.//

Remarks 3.6:

1. Clearly, $E(S_{\mathcal{P}_{BG2}}) \le E(S_{\mathcal{P}_{BG}})$.

2. By reasoning similar to that given in Bechhofer and Kulkarni (1984), $P\{CS|\mathcal{P}_{BG2}\} = P\{CS|\mathcal{P}_{BG}\}$. That is, no $P\{CS\}$ is lost between the two procedures.

3. Tables for $\mathcal{P}_{BG2}$ are currently being

prepared. See Remark 3.6.2 above for information concerning the P{CS}.

Example 3.6.1:

Again, let $k = 3$, $P^* = 0.75$, $\theta^* = 3$. Then $N_{BG2} = 5$ and $P\{CS|\underline{p} = LFC\} = 0.757$ as before. Now, $E(S_{\mathcal{P}_{BG2}}|\underline{p} = LFC) = 3.24 < 3.48 = E(S_{\mathcal{P}_{BG}}|\underline{p} = LFC)$.//

## 3.7 General remarks

We have seen procedures which follow a poset of sorts in terms of sampling efficiency. $\mathcal{P}_{BEM}$ leads to the more efficient $\mathcal{P}_{BK}$. Similarly, $\mathcal{P}_{BKS}$ leads to $\mathcal{P}_{BG}$ which, in turn, leads to $\mathcal{P}_{BG2}$. $\mathcal{P}_{RA}$ stands alone. We note that augmentations can be made to $\mathcal{P}_{RA}$, but this makes the search for the optimal combination of r and N intractable.

$$\mathcal{P}_{BEM} \longrightarrow \mathcal{P}_{BK}$$
$$\mathcal{P}_{BKS} \longrightarrow \mathcal{P}_{BG} \longrightarrow \mathcal{P}_{BG2}$$
$$\mathcal{P}_{RA}$$

In lieu of work currently in progress, we recommend use of $\mathcal{P}_{RA}$ or $\mathcal{P}_{BG2}$ when these procedures are applicable to the situation at hand.

Acknowledgement: We thank Prof. Robert E. Bechhofer of Cornell for his many comments and suggestions.

## References:

Bechhofer, R. E., Elmaghraby, S., and Morse, N. (1959). A single-sample multiple-decision procedure for selecting the multinomial event which has the highest probability. Ann. Math. Stat., 30, 102-119.

Bechhofer, R. E., and Goldsman, D. (1984a). On the Ramey-Alam sequential procedure for selecting the multinomial event which has the largest probability. School of OR&IE technical report, Cornell University, Ithaca, NY.

Bechhofer, R. E., and Goldsman, D. (1984b). Truncation of the Bechhofer-Kiefer-Sobel sequential procedure for selecting the multinomial event which has the largest probability. School of OR&IE technical report, Cornell University, Ithaca, NY.

Bechhofer, R. E., Kiefer, J., and Sobel, M. (1968). Sequential Identification and Ranking Procedures. The Univ. of Chicago Press, Chicago.

Bechhofer, R. E., and Kulkarni, R. (1984). Closed sequential procedures for selecting the multinomial events which have the highest probabilities. School of OR&IE technical report, Cornell University, Ithaca, NY.

Bechhofer, R. E., and Sobel, M. (1958). Nonparametric multiple-decision procedures for selecting that one of k populations which has the highest probability of yielding the largest observation (prelim. report). Abstract. Ann. Math. Stat., 29, p. 325.

Gupta, S. S., and Panchapakesan, S. (1979). Multiple Decision Procedures. John Wiley, New York.

Kesten, H., and Morse, N. (1959). A property of the multinomial distribution. Ann. Math. Stat., 30, 120-127.

Ramey, J. T., Jr., and Alam, K. (1979). A sequential procedure for selecting the most probable multinomial event. Biometrika, 66, 171-173.