TUTORIAL ON INDIFFERENCE-ZONE NORMAL MEANS RANKING AND SELECTION PROCEDURES

David Goldsman
School of ISyE
Georgia Institute of Technology
Atlanta, GA 30332

ABSTRACT

This tutorial concerns the ranking and selection problem of choosing that one of k normal populations (with common known variance) which has the largest mean. We discuss a number of procedures which guarantee a specified probability of selecting the correct population. We concentrate on so-called indifference-zone procedures.

1. INTRODUCTION

Ranking and selection procedures are a collection of methods which enable an experimenter to answer such statistical questions as:

o Which of k normal populations has the largest mean?

o Which of k Bernoulli populations has the largest "success" probability?

o Which is the most probable cell of a multinomial distribution?

The purpose of this tutorial is to introduce the reader to basic ranking and selection notation and techniques. We concentrate on the problem of choosing that one of k normal populations (with common known variance) which has the largest mean. This problem serves as a springboard to the other questions mentioned previously. For more comprehensive treatments on ranking and selection, see Gibbons, Olkin, and Sobel (1977) and Gupta and Panchapakesan (1979).

The organization of this paper is as follows. In Section 2, we discuss the so-called indifference-zone approach in ranking and selection. Section 3 introduces and compares a number of procedures for finding the normal population with the largest mean. An overview is given in Section 4.

2. THE INDIFFERENCE-ZONE APPROACH

Suppose $\pi_1, \ldots, \pi_k$ are k normal populations such that π_i has mean μ_i and *common known variance* σ^2, $i = 1, \ldots, k$. We wish to ascertain which of the k π_i's has the largest mean; we shall hereinafter refer to that π_i as the "best" population. Let $\mu_{[1]} \leq \mu_{[2]} \leq \cdots \leq \mu_{[k]}$ denote the ordered but unknown μ_i's. We do not know the values of the $\mu_{[i]}$'s nor do we know how they are paired with the π_i's. So the goal of finding the best π_i can be restated as that of finding the population corresponding to $\mu_{[k]}$.

A typical procedure for choosing the best population usually requires the experimenter to take a certain number of observations in a prescribed way from each π_i; the selection is then made using statistics calculated from these observations. Since the observations are realizations of random variables, it is possible that the experimenter will not choose the best π_i. However, if the best population is indeed chosen, we say that the experimenter has made a *correct selection* (CS).

In order to implement the procedures which will be discussed in this tutorial, the experimenter must supply two constants, δ^* and P^*,

prior to the start of experimentation. $\delta^* > 0$ is specified with the following thought in mind: If $\mu_{[k]} - \mu_{[k-1]}$ is very small, it stands to reason that an experimenter would regard the populations corresponding to $\mu_{[k]}$ and $\mu_{[k-1]}$ as *practically* the same. So δ^* is chosen as the smallest value of $\mu_{[k]} - \mu_{[k-1]}$ that the experimenter views as "worth detecting." In other words, if $\mu_{[k]} - \mu_{[k-1]}$ were to be $\geq \delta^*$, the experimenter would prefer to choose the π_i associated with $\mu_{[k]}$; if $\mu_{[k]} - \mu_{[k-1]}$ were to be $< \delta^*$, the experimenter would be indifferent about choosing either the population corresponding to $\mu_{[k]}$ or that corresponding to $\mu_{[k-1]}$. Hence, $\Omega_{\delta^*} \equiv \{\mu \mid \mu_{[k]} - \mu_{[k-1]} \geq \delta^*\}$ is called the *preference-zone*, and its complement is the *indifference-zone*.

The procedures we will examine in the next section guarantee the *indifference-zone probability requirement* introduced by Bechhofer (1954):

$$P\{CS\} \geq P^* \text{ whenever } \mu \in \Omega_{\delta^*} \qquad (PR)$$

Of course, we can force $P\{CS\} = 1/k$ simply by choosing as best one of the k π_i's at random. To avoid this triviality, we require $1/k < P^* < 1$.

3. INDIFFERENCE-ZONE PROCEDURES

In this section, we discuss a number of indifference-zone procedures [i.e., procedures which satisfy (PR)] for finding that one of k normal populations with common known variance which has the largest mean.

3.1 A Single-Stage Procedure

If the size of the sample to be taken from each π_i is a constant determined prior to the start of experimentation, we say that the procedure is a *single-stage* procedure. We first examine a single-stage procedure due to Bechhofer (1954).

Procedure P_B:

1. Specify $\{k, \sigma^2, \delta^*, P^*\}$, with $\delta^* > 0$ and $1/k < P^* < 1$.

2. Calculate $n_B \leftarrow \| (c_{k,P^*}\sigma/\delta^*)^2 + 1 \|$, where c_{k,P^*} is a tabled constant [see, e.g., Bechhofer (1954)], and $\| \cdot \|$ is the greatest integer ("floor") function.

3. Take n_B independent observations from each of the k populations.

4. Let X_{ij} denote the jth observation from π_i, $i = 1, \ldots, k$ and $j = 1, \ldots, n_B$. Calculate the k sample means:

$$\bar{X}_i \leftarrow \sum_{j=1}^{n_B} X_{ij}/n_B, \quad i = 1, \ldots, k.$$

5. Select as best that population corresponding to the largest $\bar{X}_i$.

As a simple example, suppose that $k = 3$, $P^* = 0.75$, and $\delta^* = 0.2$. From the tables in Bechhofer (1954), we have $c_{3,0.75} = 1.4338$. Further suppose that $\sigma^2 = 1$. Then $n_B = \| 52.39 \| = 52$. So we must take 52 independent observations from each of the three populations.

3.2 An Open Sequential Procedure

If a procedure is not a single-stage procedure, we shall say that it is *multi-stage* (or *sequential*). A sequential procedure is *open* if, prior to the start of sampling, the experimenter can not place an upper bound on the number of observations to be taken from each population. Otherwise, a sequential procedure is *closed*. The next procedure is an open sequential procedure from Bechhofer, Kiefer, and Sobel (1968).

Procedure P_{BKS}:

1. Specify $\{k, \sigma^2, \delta^*, P^*\}$, with $\delta^* > 0$ and $1/k < P^* < 1$.

2. Set the *stage counter* $n \leftarrow 0$.

3. Set $n \leftarrow n + 1$. Take one observation X_{in} from each π_i, $i = 1, \ldots, k$.

4. For all i, calculate

$$Y_{in} \leftarrow \sum_{j=1}^{n} X_{ij} .$$

Further, let $Y_{max} \leftarrow \max_{1 \leq i \leq k} Y_{in}$.

5. Calculate

$$Z_n \leftarrow \sum_{i=1}^k \exp\{-\delta^*(Y_{max} - Y_{in})/\sigma^2\} \ .$$

6. If $Z_n \leq 1/P^*$, stop sampling, and select as best that Π_i corresponding to Y_{max}. Otherwise, go to Step 3.

Let N_{BKS} denote the random number of stages required for procedure P_{BKS} to terminate (i.e., the value of the stage counter n at procedure termination). Clearly, the distribution of N_{BKS} depends on the underlying configuration of the population means $\mu_1, \ldots, \mu_k$. Two configurations are of special interest:

If $\mu_{[1]} = \mu_{[k-1]} = \mu_{[k]} - \delta^*$, we say that μ is in the *least favorable configuration* (LFC). The LFC is a "worst case" configuration for μ_i's in the preference-zone.

If $\mu_{[1]} = \mu_{[k]}$, we say that μ is in the *equal means configuration* (EMC). The EMC is a "worst case" configuration over the entire space of μ's.

Denote the expected number of stages for Procedure P_{BKS} when μ is in the LFC (EMC) by $E[N_{BKS}|\mu=LFC]$ ($E[N_{BKS}|\mu=EMC]$). It can be shown that the LFC maximizes $E[N_{BKS}]$ over all configurations in the preference-zone. That is, $E[N_{BKS}|\mu\in\Omega_{\delta^*}] \leq E[N_{BKS}|\mu=LFC]$. Consideration of the EMC is also important since $E[N_{BKS}|\mu] \leq E[N_{BKS}|\mu=EMC]$ for all μ. If (unknown to the experimenter) μ is in the EMC, procedure P_{BKS} might be quite costly to run until natural termination.

Continuing our simple example, again suppose that $k = 3$, $\sigma^2 = 1$, $P^* = 0.75$, and $\delta^* = 0.2$. Bechhofer and Goldsman (1986) show that, for this example, Procedure P_{BKS} yields $P\{CS|\mu=LFC\} \approx 0.777$, $E[N_{BKS}|\mu=LFC] \approx 38.89$, and $E[N_{BKS}|\mu=EMC] \approx 47.98$. Thus, for this example, these expected values are *smaller* than the 52 observations required by the single-stage Procedure P_B. (The number of observations required by P_B is determined before the start of experimentation, and is independent of μ once sampling commences.)

An unappealing feature of P_{BKS} is the fact that it is open. Bechhofer and Goldsman (1986) note that P_{BKS} always yields $P\{CS|\mu=LFC\}$ which is larger than the desired P^*. (In the previous example, $P\{CS|\mu=LFC\} \approx 0.777 > 0.75 = P^*$.) They study a closed version of P_{BKS} which uses *sampling truncation*, that is, the automatic termination of sampling after a certain predetermined number of stages have been conducted. Compared to P_{BKS}, truncation results in less probability "overprotection", as well as slightly smaller expected numbers of stages.

3.3 A Closed Sequential Elimination Procedure

We next discuss an interesting class of sequential procedures due to Paulson (1964). As sequential sampling proceeds, populations which appear to be inferior are permanently *eliminated* from further consideration.

Procedure P_P:

1. Specify $\{k, \sigma^2, \delta^*, P^*\}$ with $\delta^* > 0$, $1/k < P^* < 1$, and $0 < \lambda \leq \frac{1}{2}\delta^*$. [Paulson suggests taking $\lambda = \delta^*/4$; indeed, empirical calculations in Ramberg (1966) show that such a choice often results in considerable savings of observations over other values of λ. However, it can be shown analytically that as P^* approaches 1, the choice $\lambda = \frac{1}{2}\delta^*$ is, in a certain sense, optimal.]

2. Calculate the quantities

$$a_\lambda \leftarrow \frac{\sigma^2}{\delta^* - \lambda} \ln(\frac{k-1}{1-P^*})$$

and

$$W_\lambda \leftarrow \| a_\lambda/\lambda \| \ ,$$

3. Initialize the stage counter $n \leftarrow 0$ and the set of contending (non-eliminated) populations $I_0 \leftarrow \{1, 2, \ldots, k\}$.

4. Set $n \leftarrow n + 1$. Take one observation X_{in} from each Π_i such that $i \in I_{n-1}$.

5. For all $i \in I_{n-1}$, calculate

$$Y_{in} \leftarrow \sum_{j=1}^n X_{ij} \ .$$

Further, let $Y_{max} \leftarrow \max_{i \in I_{n-1}} Y_{in}$.

6. Determine the new set of contending populations:

$$I_n \leftarrow \{i \in I_{n-1} | Y_{max} - Y_{in} < a_\lambda - n\lambda\} .$$

7. If the cardinality $|I_n| = 1$ _or_ if $n = W_\lambda + 1$, then terminate sampling, and select as best that π_i corrsponding to Y_{max}. Otherwise ($|I_n| > 1$ and $n \leq W_\lambda$), go to Step 4.

Let N_P denote the random number of stages required for procedure P_P to terminate sampling. We ran a Monte Carlo experiment for our simple example with $k = 3$, $\sigma^2 = 1$, $P^* = 0.75$, $\delta^* = 0.2$, and $\lambda = \frac{1}{2}\delta^*$. The experiment revealed that $P\{CS | \mu = LFC\} \approx 0.89$, $E[N_P | \mu = LFC] \approx 79.60$, and $E[N_P | \mu = EMC] \approx 100.57$. The tremendous probability overshoot ($P\{CS | \mu = LFC\} \gg P^*$) is purchased at the expense of the expected values, both of which are much greater than the corresponding values from Procedures P_B and P_{BKS}. However, as alluded to previously, P_P performs more parsimoniously (for this example) when $\lambda = \delta^*/4$. Also, it turns out that the performance of P_P improves when P^* is closer to unity. Further, improvements in the Paulson procedure due to Fabian (1974) and Hartmann (1986) greatly reduce the expected values to competitive levels.

An advantage of an elimination-type procedure such as P_P is the fact that once a population is eliminated, no additional observations are ever taken from that π_i. Thus, whereas Procedures P_B and P_{BKS} require kn_B and kN_{BKS} total *observations*, respectively (i.e., k observations per stage), Procedure P_P requires *less than* kN_P total observations.

3.4 A Two-Stage Elimination Procedure

It can sometimes be costly (in terms of time and money) to take observations sequentially. Rather, the experimenter might prefer to conduct the entire selection procedure in one or two stages, both of which take a predetermined number of observations from each population. Tamhane and Bechhofer (1977,1979) give a class of two-stage procedures for finding the best population. We will concentrate here on the two-stage procedure which they denote as $P_2(C_1)$; for simplicity, we

shall call this Procedure P_{TB}. Procedure P_{TB} is an elimination procedure since, after the first stage, P_{TB} eliminates from further consideration populations which seem to be inferior.

<u>Procedure P_{TB}</u>:

1. Specify $\{k, \sigma^2, \delta^*, P^*\}$ with $\delta^* > 0$ and $1/k < P^* < 1$.

2. Find constants c_1, c_2, and d, tabled in Tamhane and Bechhofer (1979), pp. 348–349.

3. Calculate the constants

$$n_1 \leftarrow \| (c_1\sigma/\delta^*)^2 + 1 \|,$$

$$n_2 \leftarrow \| (c_2\sigma/\delta^*)^2 + 1 \|, \text{ and}$$

$$h \leftarrow d\delta^*/c_1 .$$

n_1 (n_2) is the sample size to be taken from each (contending) population in the first (second) stage of experimentation. h is a constant which will be used in the elimination process.

4. First stage of sampling: Take n_1 independent observations from each of the k π_i's.

5. Calculate the k first stage sample means:

$$\bar{X}_i^{(1)} \leftarrow \sum_{j=1}^{n_1} X_{ij}/n_1, \quad i = 1,\ldots,k.$$

Further, let $\bar{X}_{max}^{(1)} \leftarrow max_{1 \leq i \leq k} \bar{X}_i^{(1)}$.

6. Determine which populations will enter the second stage of sampling:

$$I \leftarrow \{ i | \bar{X}_i^{(1)} \geq \bar{X}_{max}^{(1)} - h \} .$$

7. Second stage of sampling: If $|I| = 1$, stop sampling and select as best the single contending π_i. Otherwise, take n_2 additional observations from all π_i's such that $i \in I$.

8. Calculate

$$\bar{X}_i \leftarrow \sum_{j=1}^{n_1+n_2} X_{ij}/(n_1+n_2), \quad i \in I.$$

Further, let $\bar{X}_{max} \leftarrow max_{i \in I}\bar{X}_i$.

9. Select as best that Π_i corresponding to $\bar{X}_{max}$.

Continuing our example, let $k = 3$, $\sigma^2 = 1$, $P^* = 0.75$, and $\delta^* = 0.2$. Referring to the Tamhane and Bechhofer tables, we find that $c_1 = 0.9986$, $c_2 = 0.9485$, and $d = 3.989$. Simple calculations give $n_1 = 25$ and $n_2 = 23$. So the *maximum possible* number of observations taken from a given population is 48.

Procedure P_{TB} possesses a number of appealing properties: Let T_{TB} denote the (random) total number of observations required by P_{TB}. It can be shown that P_{TB} is *uniformly* better than the single-stage Procedure P_B; T_{TB} is always less than kn_B. Procedure P_{TB} is (approximately) minimax in the sense that $\sup_\mu E[T_{TB}|\mu]$ is (approximately) minimized subject to (PR), n_1 and n_2 nonnegative integers, and $h \geq 0$.

4. OVERVIEW AND EXTENSIONS

The procedures outlined in Section 3 serve to give the reader the flavor of various simple indifference-zone procedures. Comparison among the above procedures is the subject of current research. We also point out that the LFC and EMC are not the only configurations of interest. In a real-life application, the experimenter might very well be conservative in his choice of δ^*; that is, he may actually expect $\mu_{[k]} - \mu_{[k-1]} > \delta^*$. Therefore, another interesting research topic concerns the extent to which such a "favorable" configuration of μ_i's improves selection procedure performance.

Ranking and selection techniques exist to find the best normal population when less information about the variances is available. For instance, we might only assume that the k populations have common (but unknown) variance. Or, we might assume that the variances are completely unknown. Generalizing further, we might be interested in selecting that component of a k-variate normal population which has the largest mean.

As mentioned previously, ranking and selection methods extend beyond normal populations. For instance, it might be desired to select the best of a number of drug treatments in a clinical trial; such a selection amounts to choosing that Bernoulli population with the largest "success" parameter. Perhaps the experimenter is interested in choosing the most popular variety of a certain product; it is possible to interpret this problem as that of determining the most probable multinomial cell.

The indifference-zone approach is not the only methodology available to practitioners. A large class of so-called *subset selection* procedures [introduced by Gupta (1956)] is available.

Also of interest is the fact that many ranking and selection procedures can directly be extended for use in the computer simulation environment. This is a bit surprising since most ranking and selection procedures require independent and identically distributed observations within each population – a situation which is rarely present in simulations. See Goldsman (1983,1985) for relevant surveys.

REFERENCES

Bechhofer, R.E. (1954). "A Single-Sample Multiple Decision Procedure for Ranking Means of Normal Populations with Known Variances." *Ann. Math. Stat.*, 25, 16–39.

Bechhofer, R.E. and D. Goldsman (1986). Technical Report (in preparation), School of OR&IE, Cornell Univ., Ithaca, NY.

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

Fabian, V. (1974). "Note on Anderson's Sequential Procedures with Triangular Boundary." *Annals of Statistics*, 2, 170–175.

Gibbons, J.D., I. Olkin, and M. Sobel (1977). *Selecting and Ordering Populations*. John Wiley, New York.

Goldsman, D. (1983). "Ranking and Selection in Simulation." *Proceedings of the 1983 Winter Simulation Conference*, 386–393.

Goldsman, D. (1985). "Ranking and Selection Procedures Using Standardized Time Series." *Proceedings of the 1985 Winter Simulation Conference*, 120–124.

Gupta, S.S. (1956). *On a Decision Rule for a Problem in Ranking Means*. Ph.D. Dissertation (Mimeo Ser. No. 150). Inst. of Statist., Univ. of N. Carolina, Chapel Hill.

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

Hartmann, M. (1986). "An Improvement on Paulson's Sequential Procedure for Selecting the Largest Normal Mean." Technical Report, School of OR&IE, Cornell Univ., Ithaca, NY.

Paulson, E. (1964). "A Sequential Procedure for Selecting the Population with the Largest Mean from k Normal Populations." *Ann. Math. Stat.*, 35, 174–180.

Ramberg, J.S. (1966). *A Comparison of the Performance Characteristics of Two Sequential Procedures for Ranking Means of Normal Populations*. Master's Thesis, School of OR&IE, Cornell Univ., Ithaca, NY.

Tamhane, A.C. and R.E. Bechhofer (1977). "A Two-Stage Minimax Procedure with Screening for Selecting the Largest Normal Mean." *Comm. Stat. – Theor. Meth.*, A6, 1003–1033.

Tamhane, A.C. and R.E. Bechhofer (1979). "A Two-Stage Minimax Procedure with Screening for Selecting the Largest Normal Mean (II): An Improved PCS Lower Bound and Associated Tables." *Comm. Stat. – Theor. Meth.*, A8, 337–358.

AUTHOR'S BIOGRAPHY

DAVE GOLDSMAN is an Assistant Professor of Industrial and Systems Engineering at Georgia Tech. He holds degrees from Syracuse and Cornell, where he spent last summer as a Visiting Scientist. His research interests include simulation output analysis, and ranking and selection.

Dave Goldsman
School of ISyE
Georgia Tech
Atlanta, GA 30332
(404) 894-2365