Proceedings of the 1988 Winter Simulation Conference
M. Abrams, P. Haigh, and J. Comfort (eds.)

Indifference-zone selection procedures for choosing the best airspace configuration

Douglas Gray
Decision Technologies
American Airlines, Inc.
Dallas, Tx 75261

David Goldsman
School of ISyE
Georgia Tech
Atlanta, GA 30332

ABSTRACT

We present a real-world application of a ranking and selection procedure for selecting the best of a number of competing systems. We concentrate on an indifference-zone normal means procedure. Our example involves the selection of an airspace configuration which minimizes airspace route delays.

1. INTRODUCTION

This paper illustrates a real-world application of a statistical procedure for selecting the best of a number of competing alternatives. The problem at hand concerns the selection of an "optimal" airspace configuration for a major European international and domestic airport.

During the past five years, the Stockholm airport experienced a significant increase in air traffic; this increase is expected to continue at the rate of 11% per year. The local Aviation Authority in charge of air traffic control was concerned that a third runway would be necessary to handle the anticipated volume jump. To operate the runway in the most efficient and safe manner, the terminal airspace would have to be restructured, and new air traffic control procedures would have to be developed.

Our work involved simulating a number of proposed alternative airspace configurations (AAC's) for the airport. An airspace configuration is characterized by a collection of airspace routes for arrivals and departures which traverse an airport's terminal airspace. The goal was to select the best of the candidate AAC's on the basis of various criteria of goodness. We wished to determine the most "efficient" airspace route structure. We hoped to differentiate among the candidate AAC's by simulating each under various levels and patterns of air traffic. We also hoped to identify the limitations and points of congestion for each of the AAC's.

We carried out the simulations using SIMMOD, a fast, discrete-event airport/airspace simulation model. Our SIMMOD simulation model directly evaluates each AAC's capacity to handle the anticipated future traffic increases. Event files were generated for the current level of traffic, as well as for 50%, 75%, and 100% increases in traffic.

The performance measures that were used to differentiate among the AAC's are as follows:

- Overall average route travel time
- Overall average route delay
- Individual average route travel time
- Individual average route delay
- Individual maximum route travel time
- Individual maximum route delay

Although the procedure presented herein can be applied to any of the above measures, the results in this article will focus on the measurement and estimation of individual average route delay. Airspace delays, and hence travel times, are of a stochastic nature due to varying route traffic levels and air traffic control procedures (e.g., aircraft separation, airspeed control); this stochastic nature is obviously what makes the problem of finding the best AAC difficult.

The remainder of the paper concerns an *indifference-zone* (IZ) procedure for selecting the "best" among a number of competing alternatives. In Section 2, we discuss the *normal means* IZ selection problem. A procedure for selecting the best normal population is given in Section 3, and an illustrative example is presented in Section 4.

2. THE NORMAL MEANS PROBLEM

In order to find the best AAC, and to more clearly accentuate the differences among competing AAC's, it is desirable to rank the alternatives based upon their performances with respect to a given quantitative measure of interest. Since the goal is to find that AAC which has the smallest average airspace route delay, it stands to reason that the desired AAC is that which yields the smallest simulated average route delay. Unfortunately, since the simulated processes are of a random nature, a ranking of sample means alone is not sufficient for the purpose of finding the best AAC. The variance of the associated random variables from each system must be taken into account when determining the sample size required to adequately assess system performance.

We would like to be assured that the *correct selection* (CS) of the best AAC from a group of k competing systems will be made with at least a certain high probability, say P^*, where, to avoid trivialities, $1/k < P^* < 1$. The higher we specify the desired $P\{CS\}$, the greater the number of sampling observations will be required.

Another consideration in selecting the best system concerns the difference between the performance measures from the best and second best systems. Suppose we denote the individual expected route delay arising from AAC i by μ_i, $i = 1,\ldots,k$, and the associated ordered μ_i's by $\mu_{[1]} \le \mu_{[2]} \le \cdots \le \mu_{[k]}$. (The μ_i's, $\mu_{[i]}$'s, and their pairings are completely unknown.) Since we prefer individual average route delays to be as small as possible, the mean difference between the two best systems in the ongoing example is $\mu_{[2]} - \mu_{[1]}$. The smaller this quantity is, the greater the amount of sampling that will be required to differentiate between the two systems. Naturally, if $\mu_{[2]} - \mu_{[1]}$ is extremely small, say less than $\delta^* > 0$, then for all practical purposes, it would not matter which of the two corresponding systems we choose as best;

any loss incurred would be negligible, and we would be "indifferent" as to which of the two was chosen.

Procedures which attempt to select the best of k competing systems on the basis of a quantitative performance measure subject to the considerations described above are known as *indifference-zone* (IZ) selection procedures, and were first proposed by Bechhofer (1954). A particularly useful procedure for finding that one of k normal populations which has the smallest mean is given by Dudewicz and Dalal (D-D) (1975). We note that the D-D procedure requires that the observations taken within a particular system be independent and identically distributed (i.i.d.), luxuries rarely present in the simulation environment. It will be necessary to alter the D-D procedure in certain obvious ways so that it can be applied in our simulation setting. To this end, Iglehart (1977), Law and Kelton (L-K) (1982), Koenig and Law (1982), Sullivan and Wilson (1984), and others present a number of variations on the D-D theme.

We implemented the L-K procedure to select the best of a group of AAC's. The quantitative measure by which they were to be evaluated was the average delay incurred by aircraft on a certain critical airspace route. Specifically, we implemented L-K's variation of the D-D two-stage IZ procedure for selecting the system with the smallest mean. The procedure asks the user to specify the aforementioned constants P^* and δ^*, and guarantees the *probability requirement*

$$P\{CS\} \geq P^* \text{ whenever } \mu_{[2]} - \mu_{[1]} \geq \delta^*. \qquad (1)$$

We remark that the L-K procedure can be generalized to select good subsets from among the k alternative systems.

3. A NORMAL MEANS PROCEDURE

Let the random variable of interest be denoted as

$$D_{ij} \equiv \sum_{m=1}^{n_{ij}} D_{ijm}/n_{ij} ,$$

where D_{ijm} is the airspace delay incurred on the critical route by the mth aircraft in the jth iteration (i.e., "day") of the simulation for the ith AAC, and n_{ij} is the number of aircraft observed to have flown the critical route during that iteration, $i = 1,\ldots,k$, $j = 1,2,\ldots$ (n_{ij} is a random variable.) Thus, D_{ij} is the sample average airspace delay incurred by flights travelling the critical route in iteration j of AAC i. We also let

$$\mu_{ij} \equiv E[D_{ij}].$$

The L-K procedure requires that for fixed i, $D_{i1}, D_{i2},\ldots$ are i.i.d. normal random variables. The system being modelled, i.e., the airspace configuration and its activities, is assumed to have the properties of a terminating simulation. That is, the system "reboots" itself at midnight each day, and there are no carryover effects from one day to the next. This assumption is valid since the airport in question operated from about 6:00 a.m. to 10:00 p.m. and experienced negligible traffic between 10:00 p.m. and 6:00 a.m. At the start of each "day," we independently re-seed the active random number generators, and we initialize the independent runs of a particular AAC under conditions sampled from a given distribution.

Thus, $D_{i1}, D_{i2}, \ldots$ are i.i.d. with mean $\mu_i \equiv \mu_{ij}$ and variance σ_i^2, say. Of course, the σ_i^2's are not necessarily all equal; this is what makes the problem difficult. The normality assumption is reasonable for a number of reasons. Since the D_{ij}'s are sample means of supposedly identically distributed (but not necessarily independent) D_{ijm}'s, a central limit theorem allows us to assume normality of the D_{ij}'s for sufficiently large corresponding n_{ij}'s. In our application, the n_{ij}'s were typically ≥ 90. *A fortiori*, the D-D procedure's performance is claimed to be somewhat robust against departures from the normality assumption.

Our problem is to find the AAC corresponding to the smallest expected airspace route delay, $\mu_{[1]}$, subject to the constraints given in (1). The procedure consists of two *stages* of sampling. The first stage takes an initial sample of n_0 independent days of each AAC's daily average critical route delays. The purpose of the first stage is to estimate the σ_i^2's so that efficient sampling can be carried out in the second stage. The choice of the first stage sample size is up to the user. We note that a very small n_0 might result in poor estimates of the σ_i^2's, and the subsequent second stage of sampling might then require a larger than necessary sample size. On the other hand, if n_0 is "too large," then waste occurs in the first stage.

We define the first stage sample means and variances as follows:

$$\bar{D}_i^{(1)} \equiv \sum_{j=1}^{n_0} D_{ij}/n_0 , \quad i = 1,\ldots,k$$

and

$$S_i^2 \equiv \sum_{j=1}^{n_0} \frac{[D_{ij} - \bar{D}_i^{(1)}]^2}{n_0 - 1} , \quad i = 1,\ldots,k.$$

We use S_i^2 as an estimator for σ_i^2. The total sample size required from system i (first stage plus second stage) is

$$N_i \equiv \max \{n_0+1, \| hS_i^2/(\delta^*)^2 \| \} ,$$

where $\|z\|$ is the smallest integer that is greater than or equal to z. The constant h depends on k, P*, and n_0; it is the unique solution to

$$\int_{-\infty}^{\infty} [F(x+h)]^{k-1} dF(x) = P^* ,$$

where $F(\cdot)$ is the cumulative distribution function of the t-distribution with $n_0 - 1$ degrees of freedom. Tables of h-values can be found in L-K or D-D.

The second stage of the procedure consists of taking $N_i - n_0$ additonal replications (days) from system i, i = 1,...,k, and obtaining the second stage sample means

$$\bar{D}_i^{(2)} \equiv \sum_{j=n_0+1}^{N_i} D_{ij}/(N_i - n_0) .$$

We then define weights

$$W_{i1} = \frac{n_0}{N_i} \left[1 + \left\{ 1 - \frac{N_i}{n_0} \left[\frac{1 - (N_i - n_0)(\delta^*)^2}{h^2 S_i^2} \right] \right\}^{\frac{1}{2}} \right]$$

and

$$W_{i2} = 1 - W_{i1}, \quad i = 1,\ldots,k.$$

The formidable form of the W's is necessary to guarantee the probability requirement (1). Finally, we define the weighted sample means

$$\tilde{D}_i \equiv W_{i1}\bar{D}_i^{(1)} + W_{i2}\bar{D}_i^{(2)} .$$

The system with the smallest weighted sample mean is the one selected as the best of the AAC's with respect to average daily critical route delays.

4. AN EXAMPLE

We illustrate the use of the procedure with a simple example. We wanted to select that one of k AAC's which minimizes certain expected delays. Management had narrowed down the problem to that of selecting the best of k = 3 AAC's. Choice 1 had traffic fly over the top of the airport, while the other two choices required more direct routes to the airport. The parameters for the example were as follows: $P^* = 0.90$, $\delta^* = 0.365$ minutes (after some initial empirical investigation), and $n_0 = 20$ days. Computer time for running the necessary simulations was regarded as expensive; so it was desirable to be parsimonious with observations. From the tables in L-K, we have h = 2.342. After the first stage of sampling, we calculated

$$\bar{D}_1^{(1)} = 4.35, \quad S_1^2 = 1.39, \quad N_1 = 58$$

$$\bar{D}_2^{(1)} = 2.15, \quad S_2^2 = 0.79, \quad N_2 = 33$$

$$\bar{D}_3^{(1)} = 1.78, \quad S_3^2 = 0.64, \quad N_3 = 27$$

Thus, in the second stage of sampling, we had to run 38, 13, and 7 additional days worth of simulations for AAC's 1, 2, and 3, respectively. After the second stage of sampling concluded, we had the following weighted averages.

$$\tilde{D}_1 = 4.75, \quad \tilde{D}_2 = 2.12, \quad \tilde{D}_3 = 1.81.$$

Since AAC 3 corresponded to the smallest sample mean, we selected that as best. Note, however, that AAC 2 came in a "close second" (i.e., within δ^* of AAC 3).

We conducted analogous simulation runs for other values of P^*, δ^*, and n_0, and found that AAC 3 was almost always the winner.

5. CONCLUSIONS

We have seen that the IZ normal means procedure due to D-D and modified by L-K is a viable and applicable procedure for the evaluation of alternative or competing simulated and real-world systems. The procedure considered here is useful in situations where estimation of the measure of interest is made difficult by the highly variable stochastic nature of the systems under study; in such cases, it is not feasible to select the best system solely on the basis of point estimates for the mean. The procedure discussed in this paper is easy to understand and simple to use. Generalizations of this procedure (e.g., select the m best of k competing systems) are also readily available.

REFERENCES

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

Dudewicz, E.J. and S.R. Dalal (1975). "Allocation of Observations in Ranking and Selection with Unequal Variances." *Sankhya-B*, 28-78.

Iglehart, D.L. (1977). "Simulating Stable Stochastic Systems, VII: Selecting the Best System." *TIMS Studies Management Science*, 37-50.

Koenig, L.W. and A.M. Law (1982). "A Procedure for Selecting a Subset of Size m Containing the l Best of k Normal Populations, with Applications to Simulation." Tech. Report #82-9, Dept. of Management Information Systems, Univ. of Arizona, Tucson, AZ.

Law, A.M. and W.D. Kelton (1982). *Simulation Modeling and Analysis*. McGraw-Hill, New York.

Sullivan, D.W. and J.R. Wilson (1984). "Restricted Subset Selection for Normal Populations with Unknown and Unequal Variances." *Proc. Winter Simul. Conf.*, 266-274.

DOUGLAS GRAY is a Consultant with American Airlines Decision Technologies. He holds a B.S. in Mathematical Sciences from Loyola College of Maryland, and an M.S. in Operations Research from Georgia Tech. His research interests include large-scale simulation analysis, and ranking and selection.

Decision Technologies
American Airlines, Inc.
MD-2B56, Box 619616
Dallas/Fort Worth Airport, TX 75261-9616
(817)355-1485

DAVE GOLDSMAN is an Assistant Professor in the School of ISyE at Georgia Tech. He holds a Ph.D. from the School of OR&IE at Cornell. His research interests include simulation output analysis, and ranking and selection.

School of ISyE
Georgia Tech
Atlanta, GA 30332
(404)894-2365
dgoldsma@gtri01.bitnet