THEMIS SIGNAL ANALYSIS STATISTICS RESEARCH PROGRAM

EVALUATION OF SIMULATION FROM ITS ABILITY TO SIMULATE COMPONENTS (SUBSYSTEMS) OF OVERALL SYSTEM

by

John E. Walsh

Technical Report No. 56
Department of Statistics THEMIS Contract

January 23, 1970

Research sponsored by the Office of Naval Research
Contract N00014-68-A-0515
Project NR 042-260

DEPARTMENT OF STATISTICS
Southern Methodist University

EVALUATION OF SIMULATION FROM ITS ABILITY TO SIMULATE COMPONENTS (SUBSYSTEMS) OF OVERALL SYSTEM

John E. Walsh

Southern Methodist University *

ABSTRACT

Considered is evaluation of the ability of a random-nature simulation model to represent a system. Operation of the overall system cannot be observed. However, information is available about the operation of some subsystems. These subsystems are such that their operation determines the operation of the overall system. The simulation is considered satisfactory if and only if it has a specified high probability of simulating all of these subsystems to a sufficiently close approximation (during a simulation run). Suppose that, for each subsystem, a criterion has been developed for deciding when the approximation is sufficiently close. This depends on information about the subsystem and its operation. For the evaluation, the simulation model is run a sufficient number of times. A run is considered satisfactory if and only if all the subsystems are simulated sufficiently close. These binomial data are used to investigate whether the model is satisfactory. Homogeneous and independent runs are assumed ordinarily. However, results also can be obtained that apply to heterogeneous runs. These more general results are useful when the simulation is not strictly repetitive (for example, of a man-machine nature).

* Research partially supported by ONR contract N00014-68-A-0515 and by Mobil Research and Development Corporation. Also associated with NASA Grant NGR 44-007-028 . Presented at Army Seminar on Validity of Systems Models, January 28-30, 1970.

INTRODUCTION AND DISCUSSION

Overall operation of the (single) system being simulated cannot be observed for the situation considered. Subject to this restriction, the suitability of the random-nature simulation model, for simulating the system, is to be evaluated. Available is information about operation of some subsystems. These subsystems are comprehensive in the sense that their operation determines the operation of the entire system.

This limitation on the kind of information has a disadvantage in development of the simulation model. That is, the simulation model needs to be developed so that it corresponds to the set of subsystems on which information is available. This can result in a much more detailed, and costly, simulation model than would be needed if the overall system could be observed.

The simulation model is considered satisfactory if and only if it has a high probability of simulating, in a run, all of the subsystems to an acceptable approximation. What is an acceptable approximation for a given subsystem depends on the information about that subsystem.

Ways are needed, for each subsystem, to determine when a subsystem is simulated to a reasonable approximation. This is a technical problem depending on the type of subsystem considered. Engineering design can be one important consideration, depending on the necessity for satisfying engineering specification limits. In some cases, a subsystem has a function to perform. Then this subsystem often can be judged by how well it performs that function. This simplifies consideration to how well that function of the subsystem is simulated.

Another disadvantage of the limitation on information is that the simulation model must be satisfactory for every subsystem. In some cases, the simulation could be acceptable regarding the overall system operation but not satisfactory with respect to every subsystem.

The basis for the evaluation is a sufficient number n of runs of the model. A run is labeled as satisfactory or unsatisfactory, depending on whether all the subsystems are simulated closely enough (or not). Thus, data are obtained on n binomial events. In all cases, the simulation is performed so that these events are statistically independent.

When the model is computerized, the runs are sample "values" from a fixed statistical population. Then, the binomial events are n independent trials from a binomial distribution with probability of success (a run being satisfactory) equal to some unknown value p. A (large) probability p_0 is specified such that the simulation model is considered satisfactory if and only if $p \geq p_0$. Some ways of investigating whether $p \geq p_0$ are outlined in the next section. One way is of a sequential nature.

Sometimes the simulation procedure consists of a man-machine combination. The obtaining of runs from a fixed statistical population can be very difficult under these circumstances. Moreover, identification of what is meant by "high probability of satisfactorily simulating all subsystems" experiences difficulties when the probability of a satisfactory run can change from run to run. The viewpoint adopted is that the model is satisfactory if and only if $\bar{p} \geq p_0$, where $\bar{p}$ is the arithmetic average of the probabilities of success for the n runs. Approximate investigation of $\bar{p} \geq p_0$ is considered in the next section.

The next, and final, section is concerned with statistical tests for use in deciding whether a simulation model is satisfactory on the basis of the binomial data from the runs. The null hypothesis for a test is $p = p_0$ (or $\bar{p} = p_0$). Effectively, $p = p_0$ (or $\bar{p} = p_0$) is equivalent to $p \geq p_0$ (or $\bar{p} \geq p_0$) for the tests that are considered.

OUTLINE OF TESTS

Restrictions on the size of n depend on the cost of run, the amount of time it uses, and the difficulty in determining whether a run is satisfactory (or not). The restrictions often are not very strong for computerized simulations, so that sequential tests (perhaps truncated) are a possibility. Sequential tests have advantages in specification, and determination, of Type II errors.

On the other hand, simulation runs sometimes are exceedingly expensive, time consuming, and/or difficult to classify, so that a test with minimum n is desired. Unfortunately, tests based on small amounts of data are oriented toward rejection of the hypothesis (null situation) that the model is satisfactory. That is, rejection of $p \geq p_0$ (or $\bar{p} \geq p_0$) by use of a test has a reasonably strong basis when the significance level is small. However, acceptance of $p \geq p_0$ (or $\bar{p} \geq p_0$) by a test does not have much basis. The minimum n occurs for tests where $p \geq p_0$ (or $\bar{p} \geq p_0$) is rejected unless all of the runs are satisfactory. Then, for given p_0, the value of n is determined by the value desired for the significance level.

Let p_i denote the probability of success for the i-th binomial event $(i = 1, \ldots, n)$. In the general case,

$$\bar{p} = \sum_{i=1}^{n} p_i/n.$$

All the p_i are equal to a common value p for the case of homogeneous runs.

A one-sided sequential test for investigating whether $p \geq p_0$ is given on pages 186-189 of reference 1. This is the test of $p = p(0)$ with emphasis on $p \leq p(1)$, in the notation of reference 1. The correspondence is $p(0) = p_0$, with $p(1)$ being a stated probability that is less than p_0.

Now, consider fixed sample size tests for p. These tests can be obtained directly from use of the complements of the approximate confidence intervals for the binomial parameter that are given on pages 185-186 of reference 1. The confidence intervals considered are one-sided with zero lower limit and random upper limit. Consider a test that is based on one of these confidence intervals. The test is significant if and only if the value of p_0 is outside the confidence interval. The significance level equals unity minus the value of the confidence coefficient for the interval. The alternative emphasized is $p < p_0$.

Alternatively, and perhaps easier to apply, tests can be developed directly from the binomial distribution for n independent trials with success probability equal to p. Specifically, for $p \leq p_0$, the probability that the observed number of successes does not exceed s is at most

$$\sum_{i=0}^{s} \binom{n}{i} p_0^{\,i} (1-p_0)^{n-i} \quad ,$$

which is a widely tabulated expression. Significance occurs if and only if the observed number of successes is at most s. The alternative emphasized is $p < p_0$.

Finally, consider the general situation and investigation of $\bar{p}$. The motivation for considering $\bar{p}$ is twofold. First, $\bar{p}$ seems to be reasonable consensus measure of the "probability of success." Second, approximate procedures exist that can be used to investigate $\bar{p}$. These procedures are based on the material stated for independent binomial events on pages 189-190 of reference 1 [principally for Case (II) on page 190]. Tests can be developed by setting $s_1 = 0$ and choosing s_2 so that, when $\bar{p} = p_0$, the probability that the observed number of successes is at most s_2 has probability approximately α, where α is the desired significance level for the test. Significance occurs if and only if the observed number of successes does not exceed s_2. The alternative emphasized is $\bar{p} < p_0$.

In some cases, the variation of the p_i is so small that, effectively, the binomial distribution with success probability equal to $\bar{p}$ provides a reasonable approximation to the true distribution for the number of successes. Then, the nonsequential results for investigating p are usable for investigating $\bar{p}$.

REFERENCE

Walsh, John E., HANDBOOK OF NONPARAMETRIC STATISTICS: Investigation of Randomness, Moments, Percentiles, and Distributions, D. Van Nostrand Co., Inc., 1962.

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*
SOUTHERN METHODIST UNIVERSITY

2a. REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

2b. GROUP
UNCLASSIFIED

3. REPORT TITLE
Evaluation of Simulation from its Ability to Simulate Components (Subsystems) of Overall System

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*
Technical Report

5. AUTHOR(S) *(First name, middle initial, last name)*
John E. Walsh

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
January 23, 1970	7	1

8a. CONTRACT OR GRANT NO.
N00014-68-A-0515

b. PROJECT NO.
NR 042-260

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S)
56

9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*

10. DISTRIBUTION STATEMENT
This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted for any purpose of the United States Government.

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY
Office of Naval Research

13. ABSTRACT
Considered is evaluation of the ability of a random-nature simulation model to represent a system. Operation of the overall system cannot be observed. However, information is available about the operation of some subsystems. These subsystems are such that their operation determines the operation of the overall system. The simulation is considered satisfactory if and only if it has a specified high probability of simulating all of these subsystems to a sufficiently close approximation (during a simulation run). Suppose that, for each subsystem, a criterion has been developed for deciding when the approximation is sufficiently close. This depends on information about the subsystem and its operation. For the evaluation, the simulation model is run a sufficient number of times. A run is considered satisfactory if and only if all the subsystems are simulated sufficiently close. These binomial data are used to investigate whether the model is satisfactory. Homogeneous and independent runs are assumed ordinarily. However, results also can be obtained that apply to heterogeneous runs. These more general results are useful when the simulation is not strictly repetitive (for example, of a man-machine nature).

DD FORM 1473 1 NOV 65 (PAGE 1)
S/N 0101-807-6811

UNCLASSIFIED
Security Classification

A-31408