

## COMPARISON OF LIMIT STANDARDS USING A SEQUENTIAL PROBABILITY RATIO TEST

Roy R. Creasey, Jr.

College of Business and Economics  
Longwood University  
Farmville, VA 23909, U.S.A.

K. Preston White, Jr.

Dept. of Systems and Information Engineering  
University of Virginia  
Charlottesville, VA 22904, U.S.A.

### ABSTRACT

We continue our research into the comparison, via simulation experiments, of a stochastic system to a limit standard. A limit standard is defined as a maximum and/or minimum standard. We have found that evaluation methods using proportions provide a statistically valid comparison of this family of standards. In this paper, we present the use of Wald's sequential probability ratio test (SPRT) as a comparison method. We define a fully sequential analysis procedure, as well as initial test results.

### 1 INTRODUCTION

One problem the simulation analyst consistently encounters is the proper comparison of the simulation output to a standard. A "standard" is an identified benchmark or source to which the output of a simulation is compared (Nelson and Goldsman 2001; Law and Kelton 2000).

Standards can take on a variety of forms. The most common form found in simulation output analysis is that of expected performance. (For example, Nelson and Goldsman 2001; Law and Kelton 2000; Kim and Nelson 2001; Kim and Nelson 2006). This form, however, cannot be applied to other forms of standards; in particular, when the standard is a maximum and/or a minimum. This form of standard has been defined as a limit standard (Creasey 2005).

There are many examples of limit standards in manufacturing, business, engineering and military environments. For example, many federal road construction contracts specify a maximum vehicle stopping time of thirty minutes.

Our research finds that the use of proportional statistics in the analysis of limit standards provides a consistent and valid comparison. For a system of  $n$  observations, we are interested in only one characteristic of each observation – whether or not the observation exceeds the standard. We can therefore treat this characteristic as a binomial distribu-

tion, irregardless of the underlying distribution of each observation. (Conover 1980)

This is the principle applied to acceptance sampling. The distribution of the measurement and the severity of the defect (how much the defect was out of tolerance) is not considered. The lot (or system) is accepted or rejected based on the sample size and the number of defectives observed in the sample, which in turn estimates the proportion of observations meeting the standard.

In this paper, we propose a method derived from sequential probability ratio tests (SPRT) (Wald, 1947). SPRT has been utilized in the quality control discipline to determine, sequentially, the acceptance or rejection of lots (Montgomery 2001).

Specifically, SPRT considers the ratio of two joint probability distribution functions (p.d.f); one for the null hypothesis and one for the alternate hypothesis. If the parameter of interest is the proportion of the system exceeding the standard,  $p$ , the hypotheses are

$$\begin{aligned} H_0: p &= p_1 \\ H_1: p &= p_2, \end{aligned}$$

where  $p_1$  is the customer-desired proportion exceeding the standard at a probability of acceptance  $P_a = 1 - a$ , and  $p_2$  is the customer-desired proportion exceeding the standard at a probability of acceptance  $P_a = \beta$ , and  $p_1 < p_2$ .

We are, therefore, interested in monitoring the change in the ratio as subsequent observations are incorporated. Eventually, one joint p.d.f. will prove to be superior, or closer to the joint p.d.f of the system under test. The general equation used by Wald is

$$A < \frac{\prod_{i=1}^m f(x_i, p_1)}{\prod_{i=1}^m f(x_i, p_2)} < B, \quad (1)$$

where  $m$  is the number of observations incorporated and  $A$  and  $B$  are as defined in Section 3. (Wald, 1947) If the ratio is less than or equal to  $A$ , the null hypothesis is accepted, while the alternate hypothesis is accepted if the ratio is greater than or equal to  $B$ .

This paper is organized as follows. In Section 2, we define the problem, the system under investigation and the assumptions taken. From these definitions, we provide a procedure in Section 3. Section 4 provides results of this procedure applied to an experiment. Finally, conclusions and areas of future research are outlined in Section 5.

### 2 SYSTEM DESCRIPTION

Let  $\omega_i$  denote the  $i$ th system under investigation, where  $i = 0, 1$ . Let  $\omega_0$  denote the standard system. The other system ( $\omega_1$ ) will be referred to as the simulated system, or system under test. Define  $X_j$  as the  $j$ th independent and identically distributed (i.i.d.) observation of interest from system  $\omega_1$ , for  $j = 1, 2, \dots, n$ . For the standard system, there is only one standard of interest, which is the limit standard. The limit standard is, therefore, defined as a *known standard*, meaning it is a discrete value with no variation. (Nelson and Goldsman 2001) Let  $T$  denote this limit standard, which can be a maximum ( $T_{\max}$ ), a minimum ( $T_{\min}$ ) or both.

Define  $d_m$  as the number of  $m$  observations that exceed the standard. In other words,

$$d_m = \sum_{j=1}^m b_j, \tag{2}$$

where  $b_j = 1$  when  $X_j > T_{\max}$  or  $X_j < T_{\min}$   
 $b_j = 0$  otherwise.

### 3 PROCEDURE

Using the description and assumptions stated in Section 2, we describe a procedure that will take a decision regarding the hypotheses  $H_0$  and  $H_1$  with the optimal (smallest) number of observations. This procedure is guaranteed to eventually reach a decision, regardless of the initial selection of  $p_1, p_2, a$ , or  $\beta$ , per Wald (1947).

**Assignment:** Assign values for  $p_1, p_2, a, \beta$  and  $T$  (as  $T_{\max}$  and/or  $T_{\min}$ ). From these assignments, determine

$$A = \frac{(1 - \beta)}{a},$$

$$B = \frac{(1 - a)}{\beta},$$

$$C = \frac{(1 - p_2)}{(1 - p_1)},$$

$$D = \frac{p_2}{p_1}.$$

Determine an initial set of observations  $n_0$ , using

$$n_0 = \left\lceil \text{abs} \left( \frac{\log \frac{1}{B}}{\log \frac{1}{C}} \right) \right\rceil, \tag{3}$$

where “abs” means “absolute value” and  $\lceil \cdot \rceil$  indicates rounding up to the next integer. Finally, set  $d_m = 0$  for  $m = 0$ .

**Initialization:** Take  $n_0$  observations from system  $\omega_1$ .

**Analysis (initial observations):** Determine  $d_{n_0}$  using Equation (2).

**Stopping Rule (initial observations):** If  $d_{n_0} = 0$ , then stop and conclude that the system meets the standard with confidence level of at least  $1 - a$ . If  $d_{n_0} > 0$ , continue with subsequent analysis below.

**Analysis (subsequent observations):** Take one additional observation from the system. Determine  $d_m$  using Equation (2) and the acceptance and rejection numbers,  $a_m$  and  $r_m$ , as follows.

$$a_m = \left\lfloor \frac{\log \frac{1}{B} + m \log \frac{1}{C}}{\log D - \log C} \right\rfloor,$$

$$r_m = \left\lfloor \frac{\log A + m \log \frac{1}{C}}{\log D - \log C} \right\rfloor,$$

where  $\lfloor \cdot \rfloor$  indicates rounding to the next lower integer.

**Stopping Rule (subsequent observations):** If  $d_m \leq a_m$ , then stop and conclude that the system meets the standard with confidence level of at least  $1 - a$ . If  $d_m \geq r_m$ , then stop and conclude that the system does not meet the standard. If  $r_m < d_m < a_m$ , then return to **Analysis (subsequent observations)**.

## 4 EXPERIMENT

We developed a model, using Arena®, of vehicle flow and delay within a road construction zone. (Kelton, et al. 2004) The goal of this effort is to determine a traffic control policy that meets the federal maximum vehicle delay of 30 minutes. In our experiment, we established a 16-vehicle delay policy, meaning traffic in a direction would be stopped until 16 vehicles were detained. At that point, traffic in the opposite direction was stopped, and after the vehicles already in motion left the construction zone, the 16 (or, at this point, 17 or 18) vehicles were allowed to proceed. The observation of interest was the time a vehicle was stopped.

Although federal road construction contracts impose a 30 minute maximum delay requirement, the contracts do not include the parameters necessary to analyze the requirement using a model. Therefore, for this experiment, we defined  $p_1 = 0.05$  and  $\alpha = 0.05$ , establishing a requirement that, if 95 percent of the vehicles in a replication meet the standard, that replication would be accepted with  $P_a = 0.95$ . Similarly, we defined  $p_2 = 0.06$  and  $\beta = 0.06$ , establishing a requirement that, if 94 percent of the vehicles in a replication meet the standard, that replication would be accepted with  $P_a = 0.06$ .

We conducted 200 replications. Each replication ran for 1000 minutes, resulting in an average of 1136 observations per replication. Using the parameters established above, we applied the procedure established in Section 3. From Equation (3), the initial (minimum) number of observations required in a replication was determined to be  $n_0 = 262$ .

The sequential results from a replication are shown in Figure 1. (Note: The “N/A” zone is the area where the number of failed observations are more than the number of observations, which is not possible.) Of the 200 replications, 193 initiated in either the “Indifference” or “Accept  $H_0$ ” zones, and entered and/or remained in the “Accept  $H_0$ ” zone. The average number of observations before a decision was taken was 371, with a range of 262 (using the procedure in Section 3) and 444. The seven failed replications had a large number of failed observations early in the replication, resulting in a transition to the “Accept  $H_1$ ” zone. Each of these failed replications eventually returned to the “Indifference” zone.

## 5 CONCLUSIONS

We draw two conclusions from this research. First, we conclude the model meets the standard, using the policy and parameters specified in Section 4.

Second, we conclude the use of SPRT is a statistically valid method to analyze limit standards. Subsequent research shall restate this conclusion as a theorem.

During our analysis, we discovered a small anomaly with the procedure outlined in Section 3. As shown in Figure 1, soon after the replication entered the “Accept  $H_0$ ” zone, another failed observation was detected, resulting in a return to the “Indifference” zone. We certainly do not want to accept the results of this replication until we are confident subsequent observations will not cause a return to the “Indifference” zone. In no case did a replication return to the “Indifference” zone when the number of observations in the “Accept  $H_0$ ” zone was 20 or more. Future research will modify the Procedure in Section 3 to minimize this “straddling” anomaly.

This topic has provided us with a wealth of future endeavors. As stated in Section 2, we assumed the observations are i.i.d. It can be argued the observations from the above experiment are not i.i.d. Therefore, if we wish to relax this assumption, then we must consider the use of a generalized SPRT (GSPRT). Also, Wald provided several theorems and proofs with the SPRT method. For example, Wald proved that the use of SPRT results in the optimal, or smallest, number of observations necessary to take a decision. We wish to take his theorems and assess applicability to simulation output analysis.

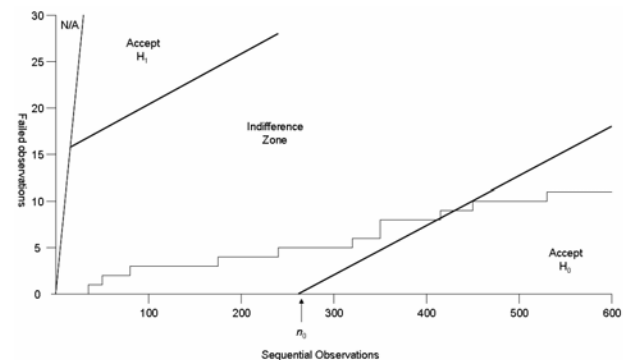


Figure 1. Analysis of a Replication Using SPRT Approach.

## REFERENCES

- Conover, W.J. (1980). *Practical Nonparametric Statistics*, 2<sup>nd</sup> ed., Wiley, New York.
- Creasey, R., K. White, M. Marks, and B. Waller (2005). “Determination of the ‘Best’ System That Meets a Limit Standard,” *Proceedings of the 2005 Winter Simulation Conference*, IEEE, Piscataway, NJ.
- Kelton, W., R. Sadowski and D. Sturrock (2004). *Simulation with Arena®*, 3<sup>rd</sup> ed., McGraw-Hill, Boston.
- Kim, S. and B. Nelson (2001). “A Fully Sequential Procedure for Indifference-Zone Selection in Simulation.” *ACM TOMCS*, 11 (3), 251-273.
- Kim, S. and B. Nelson (2006). “On the Asymptotic Validity of Fully Sequential Selection Procedures for Steady-State Simulation.” *Operations Research*, 54 (3).

- Law, A. and W. Kelton (2000). *Simulation Modeling and Analysis*, 3<sup>rd</sup> ed. McGraw-Hill, Boston.
- Miller, J., B. Nelson and C. Reilly (2002). "Estimating the Probability That a Simulated System Will Be the Best." *Naval Research Logistics*, **49** (4), 341-358.
- Montgomery, D.C. (2001). *Introduction to Statistical Quality Control*, 4<sup>th</sup> ed., Wiley, New York.
- Nelson, B. and D. Goldsman (2001). "Comparisons with a Standard in Simulation Experiments." *Management Science*: v. 47, no. 3, 449-463.
- Wald, A. (1947). *Sequential Analysis*, Wiley, New York.

#### AUTHOR BIOGRAPHIES

**ROY R. CREASEY, JR.** is a Lecturer in Management at Longwood University, and a Ph.D. student in Systems and Information Engineering at the University of Virginia. He received the B.S. from the Virginia Military Institute and the M.E. from the University of Virginia. He has seventeen years prior experience as a systems and test engineer

in the defense industry. He is a student member of INFORMS and IEEE/SMC. His e-mail is <creaseyrr@longwood.edu>.

**K. PRESTON WHITE, JR.** is a Professor of Systems and Information Engineering at the University of Virginia. He received the B.S.E., M.S., and Ph.D. degrees from Duke University. He has held faculty appointments at Polytechnic University and Carnegie-Mellon University and served as Distinguished Visiting Professor at Newport News Shipbuilding and at SEMATECH. He is U.S. Editor for *International Abstracts in Operations Research* and Associate Editor for *International Journal of Intelligent Automation* and *IEEE Transactions on Electronics Packaging Manufacturing Technology*. He is a member of INFORMS, SCS, and INCOSE and a senior member of IEEE and IIE. He sits on the Advisory Board of VMASC and represents IEEE/SMC on the Winter Simulation Conference Board. His e-mail is <kpwhite@virginia.edu>.