Performance Analysis of Sequential Tests in Process Control

Submitted for Consideration to the Nicholson Student Paper Competition

by

Kristin K. Hoyer†
Department of Industrial Engineering/Management Science
Northwestern University
Evanston, Illinois, 60208

† The entrant would like to thank her advisor Prof. Maria Rieders for all of her assistance, time, and patience with respect to this paper and the research presented herein.

DISCLAIMER

1 Introduction

In recent years, a great deal of emphasis has been placed on quality control of industrial processes. In particular, numerous statistical techniques exist which are designed to continually check an industrial process for machine or component failure, thereby determining if the process is *under control*, or if it is becoming *out of control*. In this study, we consider a very powerful class of quality control techniques known as sequential tests.

Sequential tests classify a set of observations in a manner similar to statistical hypothesis tests, but are characterized by a random sample size. Perhaps the best known sequential test is Wald's sequential probability ratio test [27]. The sequential probability ratio test (SPRT) is a log likelihood ratio based test for simple or composite hypotheses. After taking each observation, the SPRT decides whether to accept the null hypothesis, reject the null hypothesis, or continue sampling. For the purpose of quality control, an SPRT can be conducted repeatedly over time as incoming observations are received.

The SPRT is only one of several process control techniques which can be formulated as a sum of random increments inside an interval of the real line. Once the sum leaves the interval of interest, the test is terminated and a decision is made which characterizes the observations. In this study, we focus attention on the SPRT; however, the analysis presented here applies equally well to other control techniques based on sums of random variables (such as cumulative sum control charts).

One important question is how effectively a given sequential test will monitor the process under study. The effectiveness of a sequential test can be summarized by two performance measures: (1) how long it will take for the test to make an incorrect decision given the process is either under control or out of control and (2) how quickly the test will detect failure of a process which is becoming gradually (or

rapidly) out of control Both performance measures correspond to passage times of a sum of random increments to some threshold. These times are, in general, quite difficult to compute. Research to date includes limited results with respect to these performance measures. Namely, several techniques exist which allow us to approximate the distribution and moments of (1), however, virtually no results exist which allow us to study (2).

The significance of obtaining these performance measures is emphasized for safety-critical or mission-critical applications. Such systems often use a high degree of built-in redundancy to increase reliability, resulting in numerous correlated processes to be monitored. At Argonne National Laboratory (ANL), expert systems are under development for process control of safety-critical or mission-critical systems, (see [11]-[13]). These expert systems utilize numerous sequential tests operating both simultaneously and in sequence for the purpose of detecting and diagnosing component failures of highly redundant systems. The expert systems were originally developed to monitor ANL's EBR-II nuclear reactor coolant pumps but have been generalized to include other systems. Current applications of these expert systems include the nuclear reactor industry, the airline industry, and surgical equipment among others. The research developed here is motivated by the critical nature of these applications, and is currently being used by the author and the scientists at ANL as a foundation for design and analysis of the expert systems for process control.

In this research, we develop a methodology for obtaining performance measures of sequential tests for process control. This methodology is developed in two stages. First, we formulate a bivariate Markov model in discrete time with continuous state space. The first parameter of this model tracks the value of the test statistic over time and the second parameter monitors the state of the process under study. Called a Markov additive process, the model allows us to study sequential tests under non-i.i.d. process behavior.

In particular, we can model the behavior of sequential tests for under control processes, gracefully failing processes, suddenly failing processes, and out of control processes. Second, using stochastic order relations, we develop a simple computational technique for obtaining passage times of our Markov additive model. This technique involves discretizing the state space of our model in a particular manner and studying the resulting Markov chain. We prove that bounds on the first passage time distribution can be obtained using this method and that these bounds become exact as the number of discrete states increases. Finally, we use this technique to compute the performance measures of interest for sequential tests in process control.

It should be noted that the method developed here for computing first passage times applies to discrete time Markov additive processes in general and has utility in a variety of contexts. Markov additive processes (sums of random variables defined on a Markov chain) have been used to study various applications, including queuing models, dam models, shock models, and insurance risk models. In many cases, passage times of these models are of interest. With the limited exception of problems having a discrete state space, no well-established computational techniques exist for obtaining passage times of Markov additive processes. Therefore, utility of the technique developed here extends well beyond the area of process control.

In the following section, we review the sequential probability ratio test and state some of its properties. In Section 3, we formulate a Markov additive model which allows us to study the sequential probability ratio test under various types of process behavior. Section 4 develops the theoretical results and the methodological approach that allow us to bound the first passage time distributions of our model. Section 5 illustrates our techniques through numerical examples.

2 The Sequential Probability Ratio Test

Sequential tests are concerned with characterizing data from some random experiment in a manner similar to hypothesis testing. A sequential test differs from a standard hypothesis test primarily through a random sample number which depends on the outcome of the experiment. A sequential test may be described as follows. After collecting each observation, a rule is given for taking one of three actions. The possible actions are

(1) terminate the test and accept the null hypothesis
(2) terminate the test and reject the null hypothesis
(3) continue sampling

For a more complete development of the SPRT, see Ghosh and Sen [10], Chapter 3. For simplicity, assume X_1, X_2,... are independent, identically distributed (i.i.d.) random variables with absolutely continuous distribution. The following development and analysis presented in this section is valid under more general conditions and can be done in analogy. Consider the following hypotheses:

H_0: X_1, X_2,... have density $f_0(\ )$
H_1: X_1, X_2,... have density $f_1(\ )$

The SPRT for the above hypothesis proceeds as follows: select constants a and b with $0<b<a<\infty$. Denoting the ith observation by x_i, let $\lambda_n = \prod_{i=1}^{n} f_1(x_i)/f_0(x_i)$ represent the likelihood ratio at stage n. According to the value of λ_n take the following action:

If	$\lambda_n \le b$	terminate and accept H_0
If	$b < \lambda_n < a$	continue sampling
If	$\lambda_n \ge a$	terminate and reject H_{01}

The SPRT is implemented by taking the logarithm of the likelihood ratio λ_n. Let $Z_i = \log[\ f_1(x_i)/f_0(x_i)]$, and $S_n = \log \lambda_n = \Sigma_{1\le i\le n} Z_i$. Define $A = \log(a)$ and $B = \log(b)$. Then the SPRT is equivalent to the following:

If	$S_n \leq B$	terminate and accept H_0
If	$B < S_n < A$	continue sampling
If	$S_n \geq A$	terminate and reject H_0

In practice, the values A>0 and B<0 are determined by the desired error probabilities of the test and are further mentioned below. Clearly, one SPRT is equivalent to a random walk $S_n = \Sigma_{1 \leq j \leq n} Z_i$ inside an interval with $S_0 = 0$ and two boundaries B and A. Define $N = \min \{n>0: S_n \leq B \text{ or } S_n \geq A\}$. Then N is a stopping time which determines when S_n leaves the interval (B,A) thus terminating the test. N is called the *sample number* of the test. Several methods exist for approximating the random variable N (see e.g. Wald [27], Martinsek [17], Berk [3], and Siegmund [26]), however, these results are valid only under the null or alternative hypothesis. The method we develop here can be applied to approximate the sample number under more general conditions.

Let $\alpha = P\{$ reject $H_0 \mid H_0$ true$\}$ be the type I error probability. i.e. the false alarm probability, and $\beta = P\{$ accept $H_0 \mid H_1$ true$\}$ be the type II error probability for the SPRT, i.e. the missed alarm probability. A relationship between a,b, α, and β exists which allows one to specify error probabilities for the SPRT and from that determine the constants a and b. In particular, to ensure that $\alpha \leq \alpha_0$ and $\beta \leq \beta_0$ for some prespecified values α_0 and β_0, Wald's [27] boundaries a and b can be used where

$$a = \frac{1-\beta_0}{\alpha_0} \tag{2.1a}$$

$$b = \frac{\beta_0}{1-\alpha_0} \tag{2.1b}$$

We now describe the use of the SPRT in quality control. Suppose observations from an incoming process

are i.i.d. Gaussian random variables each with mean μ_0 and variance σ^2. Furthermore, suppose that when the mean of an incoming process falls outside the interval $[\mu_0 - \delta, \mu_0 + \delta]$, it is considered to be out of control. Since the SPRT is a one-sided test, it is necessary to simultaneously run two tests. This two-sided SPRT for process control is given by the following:

I. H_0: $X_1, X_2, \ldots$ are normally distributed with mean μ_0, and variance σ^2
H_1: $X_1, X_2, \ldots$ are normally distributed with mean $\mu_1 = \mu_0 - \delta$, and variance σ^2

II. H_0: $X_1, X_2, \ldots$ are normally distributed with mean μ_0, and variance σ^2
H_1: $X_1, X_2, \ldots$ are normally distributed with mean $\mu_1 = \mu_0 + \delta$, and variance σ^2

In testing for the mean of a Gaussian population as in tests I and II, the log likelihood ratio Z_i is easily shown to be the following:

$$Z_i = \frac{\mu_1 - \mu_0}{\sigma^2} X_i + \frac{\mu_0^2 - \mu_1^2}{2\sigma^2} \tag{2.2}$$

Let Z^I_i (Z^{II}_i) be the log likelihood ratio increment for test I (II). Similarly, let S^I_n (S^{II}_n) be the corresponding SPRT test statistic for tests I (II). Then Z^I_i (Z^{II}_i) can be obtained by using $\mu_1 = \mu_0 - \delta$ ($\mu_1 = \mu_0 + \delta$) in expression (2.2).

The log likelihood ratios Z^I_i and Z^{II}_i then serve as the respective increments for the SPRT test statistics in tests I and II. When a test decides H_0, its corresponding SPRT sum S^k_n for k=I,II is set to 0 and a new test begins. If either SPRT decides H_1, the testing terminates and the process is determined out of control.

The key issue addressed in this research is that of determining how well a sequential test will monitor the process under study. One primary performance measure of interest for sequential tests is the amount of

time needed for the test to exceed the threshold A. This time until the first H_1 decision is referred to in the control literature as the *run length*. For an SPRT used in process control, the run length corresponds to the first time a test decides H_1. The run length of the SPRT is a critical performance measure for examining the behavior of the test under various types of process behavior. The run length under i.i.d. observations has been addressed for both one- and two-sided cumulative sum (CUSUM) sequential tests (see Böhm and Hackl [6], Reynolds [23], Brooks and Evans [7], Woodall [29], [30], and Waldmann [28]). However, of primary interest is the behavior of the sequential test when the process under study consists of non-i.i.d. observations. Specifically, we would like to know the run length of the SPRT under various types of process failure, i.e. when the mean of the process is changing over time. No such results exist for the SPRT, and only a few exist for cumulative sum tests. Existing results are limited to rough approximations (see Bagshaw and Johnson [1]).

3 The Sequential Probability Ratio Test as a Markov Additive Process

In this section, we propose a model which allows us to study the one-sided SPRT under various types of process behavior. When the incoming observations are i.i.d., the increments of the SPRT are also i.i.d. and the SPRT test statistic $S_n = \Sigma_{1 \le i \le n} Z_i$ forms a Markov process. Suppose the observations are no longer i.i.d. but the increments $Z_i; i=1,2,...$ are allowed to vary in distribution based on the progress of a finite Markov chain $\{D_n, n \ge 0\}$. In other words, we let $D_n; n=1,2,...$ represent the mean of the incoming process X_n at time n and consider the bivariate Markov process $\{(S_n, D_n); n=0,1,2,...\}$ The resulting bivariate Markov process is called a Markov additive process. A formal definition is included in the following subsection.

3.1 Definition of a Markov Additive Process

In general, a Markov additive process (MAP) is an additive function S_n of a Markov process D_n so that the pair (S_n, D_n) forms a bivariate Markov process. More practically, an MAP consists of some additive random variable S_n which depends incrementally on the progress of a finite Markov chain. The following definition, contributed to Keilson and Rao [15], formally defines a Markov additive process in discrete time.

Definition 1: A two dimensional Markov process (S_n,D_n) for n=0,1,2,... is said to be a *Markov additive process in discrete time* if it satisfies the following criteria:

(i) $\{D_n, n \geq 0\}$ is a finite Markov chain on state space $\{0,1,...R\}$ governed by a transition probability matrix $\underline{B} = \{b_{rs}\}$

(ii) S_n is a sum $S_n = S_0 + \Sigma_{i \leq n} Z_i$ dependent on the chain in such a way that if the ith transition takes the chain $\{D_n, n \geq 0\}$ from state r to state s, the increment Z_i has c.d.f. $P\{Z_i \leq t\} = \chi_{rs}(t)$.

MAPs have been used to model a wide variety of situations. Examples include queuing, dam, shock and insurance risk models (see e.g. [14], [15]). The following section develops a Markov additive model for the one-sided SPRT which allows the process under study to vary in some characteristic manner.

3.2 Model Formulation

As noted in Section 2, the sequential probability ratio test is simply a sum $S_n = \Sigma_{1 \leq k \leq n} Z_k$ with $S_0 = 0$ inside an interval (B,A), where Z_k is a linear function of the incoming process X_k. Suppose $\{D_n, n \geq 0\}$ is a Markov chain which represents the mean of incoming process X_n so that the X_n are normally distributed with variance σ^2 and mean specified by D_n for n=1,2,... For the test of the mean of i.i.d. Gaussian

observations proposed in Section 2, $Z_n = (\mu_1 - \mu_0)X_n/\sigma^2 + (\mu_0^2 - \mu_1^2)/(2\sigma^2)$. Therefore, if X_n is gaussian with mean ν and variance σ^2, Z_n is gaussian with mean $(\mu_1 - \mu_0)\nu/\sigma^2 + (\mu_0^2 - \mu_1^2)/(2\sigma^2)$ and variance $(\mu_1 - \mu_0)^2/\sigma^2$. In this case, it is easy to see that the pair (S_n,D_n) is a Markov Additive Process in discrete time with Z_n having continuous state space. The transition function of the bivariate Markov process (S_n,D_n) is

$$P\{S_{n+1} \le t, D_{n+1} = s \mid S_n = w, D_n = r\} = b_{r,s}\Phi\left(\frac{\sigma}{\sqrt{\mu_0 - \mu_1}}\left[t - w - \frac{(\mu_0 - \mu_1)\nu_s}{\sigma^2} + \frac{\mu_1^2 - \mu_0^2}{2\sigma^2}\right]\right) \qquad (3.1)$$

where ν_s is the process mean corresponding to $D_{n+1} = s$, and $\Phi(.)$ represents the cumulative distribution function for a standard normal random variable.

Modeling the SPRT in this way allows us to study a variety of process behaviors. For example, by setting R=0 so that the state space of D_n is $\Omega=\{0\}$, we can model the case where $X_1,X_2,...$ are i.i.d. By letting $D_n \in \Omega = \{0,1,...R\}$ be a pure birth process with corresponding means $\{\nu_0,\nu_1,...\nu_R\}$ and ν_R an absorbing state, we can model a process whose mean is increasing up to some critical value. For R=1 and $\nu_0 < \nu_1$ we can model a rapid change in the mean of an incoming process. For large R and $\nu_0 < \nu_1 < ... < \nu_R$ we can model a gradual increase in the mean of a process. A variety of other types of process behavior can be modeled in this way. However, from a practical point of view and as verified through discussions with the engineers at ANL, it will suffice to study process failure as described by the three types mentioned here. The transition diagrams of D_n for each of these three examples are shown in Figure 1.

3.3 Performance Measures

Under the three types of process behavior shown in Figure 1, we would like to know how quickly the SPRT for process control will decide H_1, thus determining the process to be out of control. Therefore, we would like obtain first passage times of the process (S_n,D_n) to the absorbing boundary $\{S_n \ge A\}$. In case

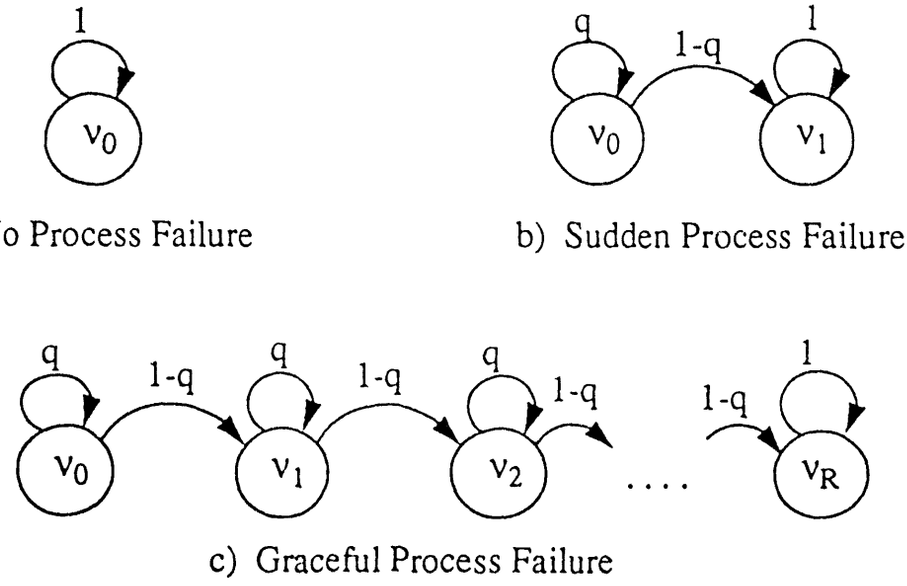


Figure 1. Transition Diagram of $\{D_n: n \geq 0\}$ for Three Types of Process Behaviors

a, the trivial case where the incoming observations are i.i.d., the passage time of S_n to A will enable us to compute the frequency of false alarms, i.e. how often the SPRT will decide the process is out of control given that it is actually under control. In cases b and c, first passage times to the boundary $\{S_n \geq A\}$ yield information about how quickly the SPRT will detect a change in the process mean. The run length is given by the passage time T_A where

$$T_A = \min\{n: S_n \geq A\} \tag{3.2}$$

However, direct computation of passage times has, in general, proven to be a difficult problem. The majority of research involving passage times of Markov Additive processes has been mostly of theoretical interest. For example, Miller [18] studies convexity properties of boundary

problems for MAP's. Keilson and Wishart [16] derive a central limit theorem for such processes. Çinlar [8],[9] derives several properties of MAP's such as a strong Markov property. Ney and Nummelin [19],[20] study the large deviations properties of such processes. Sadowsky [25] derives a data dependent extension of Wald's identity for MAP's. No methods exist, however, specifically for computing transient performance measures of Markov additive processes in general. Only when Z_i has a discrete state space can such measures be easily obtained. In this limited case, passage times can be computed by considering the resulting bivariate Markov Chain and using well-known methods. In the following section, we develop a method for computing passage times of the more general problem.

4 Passage Times for Markov Additive Processes

In this section, we consider discretizing the state space of Z_i for i=1,2,... so that we can use well-established methods to compute first passage time distributions of the resulting Markov chain. Brook and Evans [7] proposed discretizing the state space as a method for approximating the run length distribution in a one-sided CUSUM test and Woodall [30] extended this technique to a two-sided test. However, this method provides no means to assess the accuracy of the resulting approximation. The following development proposes a method for discretizing the state space of a Markov Additive Process so that bounds on first passage times can be obtained. The next subsection presents the notation that will be used throughout the analysis. Subsection 4.2 defines the concepts needed in the development and states previously proved results. Subsection 4.3 develops the theory for obtaining first passage time bounds. In Subsection 4.4, we present a method for discretizing the state space and in Subsection 4.5, we prove convergence results which demonstrate asymptotic exactness of the method.

4.1 Notation

Throughout the analysis, we will let the superscripts U and L be used to denote random variables which provide upper and lower bounds, respectively, on the quantity of interest (e.g. Z^U_i, and Z^L_i will denote upper and lower bounds for Z_i). The following notation will be used throughout the entire analysis:

D_n; n=0,1,2,... a finite Markov chain on $\Omega = \{0,1,2,...R\}$, with $D_0 = d_0$

$Z_i \in \mathbf{R}$; i=0,1,2,... a sequence of absolutely continuous random variables with the property that if the Markov chain's ith transition is from state r to state s, then the c.d.f. if Z_i is $\Psi_{rs}(\ )$

$Z^U_i \in \mathbf{R}$; i=0,1,2,... a sequence of random variables with the property that if the Markov chain's ith transition is from state r to state s, then the c.d.f. if Z^U_i is $\chi_{rs}(\ )$ where $1-\chi_{rs}(t) \geq 1-\Psi_{rs}(t)$ for all $t \in \mathbf{R}$, and all $r,s \in \Omega$

$Z^L_i \in \mathbf{R}$; i=0,1,2,... a sequence of random variables with the property that if the Markov chain's ith transition is from state r to state s, then the c.d.f. if Z^L_i is $\Gamma_{rs}(\ )$ where $1-\Gamma_{rs}(t) \leq 1-\Psi_{rs}(t)$ for all $t \in \mathbf{R}$, and all $r,s \in \Omega$

$S_n = \sum_{i \leq n} Z_i$

$U_n = \sum_{i \leq n} Z^U_i$

$L_n = \sum_{i \leq n} Z^L_i$

$G \subseteq \Omega$ a subset of the state space of the Markov chain D_n

$T^L_{xA} = \min\{n : L_n \geq A\}$, given $(L_0 = x, D_0 = d_0)$

$T^S_{xA} = \min\{n : S_n \geq A\}$, given $(S_0 = x, D_0 = d_0)$

$T^U_{xA} = \min\{n : U_n \geq A\}$, given $(U_0 = x, D_0 = d_0)$

$T^L_{xB} = \min\{n : L_n \leq B\}$, given $(L_0 = x, D_0 = d_0)$

$T^S_{xB} = \min\{n : S_n \leq B\}$, given $(S_0 = x, D_0 = d_0)$

$T^U_{xB} = \min\{n : U_n \leq B\}$, given $(U_0 = x, D_0 = d_0)$

4.2 Definitions and Existing Results

The bounds on first passage time distribution to be developed later are based on the concept of stochastic order relations. In this subsection, we review the definition and basic properties of stochastic ordering of random variables. The proofs can be found in Ross [24].

Definition 2: Given random variables X, and Y defined on $\mathbf{R}$, X is said to be *stochastically larger than* Y if $P\{X > t\} \geq P\{Y > t\}$ for all $t \in \mathbf{R}$, and the relation is denoted by $X \geq_{st} Y$.

Lemma 1: If $X \geq_{st} Y$, then $E[X] \geq E[Y]$.

Lemma 2: $X \geq_{st} Y$ iff $E[f(X)] \geq E[f(Y)]$ for all nondecreasing functions f.

Lemma 3: If F and G are distribution functions with $1\text{-}F(a) \geq 1\text{-}G(a)$ for all a, then there exists random variables X and Y having c.d.f. F and G respectively so that $P\{ X \geq Y\} = 1$.

4.3 Development of Bounds for First Passage Times

From Section 4.1 and the definition of an MAP in Section 3, it is easy to verify that (S_n,D_n), (L_n,D_n), and (U_n,D_n) are all Markov additive processes in discrete time which have the same underlying Markov chain $\{D_n, n\geq 0\}$. The following lemma establishes an order relation between the three processes.

Lemma 4: For (S_n,D_n), (L_n,D_n), and (U_n,D_n) as defined in Section 4.1.

$$Z^U_i \geq_{st} Z_i \geq_{st} Z^L_i \tag{4.1}$$

Proof: From the construction of the processes $\{L_n, n\geq 0\}$, and $\{U_n, n\geq 0\}$, it follows that if the ith transition of the Markov chain takes the chain from state r to state s, then $1-\chi_{rs}(t)\geq 1-\varphi_{rs}(t) \geq 1-\Gamma_{rs}(t)$ for all t. The result follows immediately since this is true for any $r,s\in\Omega$. Q.E.D.

The next lemma demonstrates conditional independence of the increments Z_i of a Markov Additive process given the progress of the underlying Markov chain $\{D_n, n\geq 0\}$

Lemma 5: For a Markov Additive Process as defined in Section 4.1, Z_{n+1} is conditionally independent

$$P\{Z_{n+1}\leq t \mid Z_n, D_n, D_{n+1}\} = P\{Z_{n+1}\leq t \mid D_n, D_{n+1}\} \tag{4.2}$$

of Z_n given D_n and D_{n+1}. In other words,

Proof: The proof follows from definition 4.1, part 2. Q.E.D.

Lemma 6: Given Z^U_i, Z_i, and Z^L_i as defined in Section 4.1 with $Z^U_i \geq_{st} Z_i \geq_{st} Z^L_i$ for i=1,2,...n, then for any componentwise nondecreasing function f( ), $f(Z^U_1,Z^U_2,...Z^U_n) \geq_{st} f(Z_1,Z_2,...,Z_n) \geq_{st} f(Z^L_1,Z^L_2,...Z^L_n)$.

Proof: Lemma 3 implies that for $Z^U_i \geq_{st} Z_i \geq_{st} Z^L_i$ we can define a $U^*_i =^d Z^U_i$ and a $L^*_i =^d Z^L_i$ with the property that $P\{U^*_i \geq Z_i \geq L^*_i\} = 1$. Suppose we are given $D_0 = d_0$, $D_1 = d_1$, ..., $D_n = d_n$. Then the sequences U^*_i, Z_i, and L^*_i i=1,2,...n are sequences of independent random variables. Using this and the assumption that f is a componentwise nondecreasing function, we obtain

$$\{f(L^*_1,L^*_2,...L^*_n)>t \mid D_0=d_0,D_1=d_1,...,D_n=d_n\} \Rightarrow \{f(Z_1,Z_2,...Z_n) > t \mid D_0=d_0,D_1=d_1,...,D_n=d_n\}$$
$$\Rightarrow \{f(U^*_1,U^*_2,...U^*_n) > t \mid D_0=d_0,D_1=d_1,...,D_n=d_n\}$$

almost surely. Therefore,

$$P\{f(L^*_1,L^*_2,...L^*_n)>t|D_0=d_0,D_1=d_1,...,D_n=d_n\} \le P\{f(Z_1,Z_2,...Z_n) > t|D_0=d_0,D_1=d_1,...,D_n=d_n\}$$

$$\le P\{f(U^*_1,U^*_2,...U^*_n) > t| D_0 = d_0,D_1=d_1,...,D_n=d_n\}.$$

Using Lemma 3 and unconditioning gives

$$\begin{aligned}
P[f(Z_1^L,Z_2^L,..,Z_n^L)>t] &= \sum_{d_0,...,d_n} P[f(Z_1^L,Z_2^L,..Z_n^L>t|d_0,..d_n]P[d_0,..d_n] \\
&= \sum_{d_0,...,d_n} P[f(L_1^*,L_2^*,..L_n^*)>t|d_0,..d_n]P[d_0,..d_n] \\
&\le \sum_{d_0,...,d_n} P[f(Z_1,..Z_n)>t|d_0,..d_n]P[d_0,..d_n] \\
&\le \sum_{d_0,...,d_n} P[f(U_1^*,U_2^*,..U_n^*)>t|d_0,..d_n]P[d_0,..d_n] \\
&= \sum_{d_0,...,d_n} p[f(Z_1^U,Z_2^U,..Z_n^U)>t|d_0,..d_n]P[d_0,..d_n] \\
&= P[f(Z_1^U,Z_2^U,..Z_n^U)>t)
\end{aligned}$$

Q.E.D.

Lemma 6 gives upper and lower bounds on the first passage time distribution to either A>0 or B<0 for Markov Additive processes.

Theorem 1: Given Z^U_i, Z_i, and Z^L_i as defined above with $Z^U_i \ge_{st} Z_i \ge_{st} Z^L_i$ for i=1,2,...n, and the first passage times T^L_{xA}, T^S_{xA}, T^U_{xA}, T^L_{xB}, T^S_{xB}, T^U_{xB} with $x \in (B,A)$, then

$$T^U_{xA} \le_{st} T^S_{xA} \le_{st} T^L_{xA}, \quad (4.3a)$$

$$\text{and } T^U_{xB} \ge_{st} T^S_{xB} \ge_{st} T^L_{xB} \quad (4.3b)$$

Proof: The proof is divided into two cases.

Case 1 First passage time to A>0. We note that, the first passage time

$$T^S_{xA} = \min\{n: S_n \ge A | S_0 = 0, D_0 = d_0\} \Rightarrow \{T^S_{xA} > n\} = \{\max_{1\le k\le n} S_k < A | S_0 = 0, D_0 = d_0\}.$$

Let $f(x_1,x_2,..,x_n) = \max_{1\le k\le n}\left\{\sum_{0\le i\le k} x_i\right\}$. Clearly f ( ) is a componentwise nondecreasing function.

The first passage time $\{T^U_{xA}>n\}=\{f(Z^U_1,Z^U_2,...Z^U_n)<A\}$,

$\{T^S_{xA}>n\}=\{f(Z_1,Z_2,...Z_n)<A\}$, and

$\{T^L_{xA}>n\} = \{f(Z^L_1,Z^L_2,...Z^L_n)<A\}$.

Lemma 4 implies that

$$P\{f(Z^U_1,Z^U_2,...Z^U_n) < A\} \ge P\{f(Z_1,Z_2,...Z_n) <A\}\ge P\{f(Z^L_1,Z^L_2,...Z^L_n) < A\},$$

so we have the relation $P\{T^U_{xA}> n\}\le P\{T^S_{xA}> n\} \le P\{T^L_{xA}> n\}$.

Case 2 First passage time to B<0. We note that, the first passage time

$$T^S_{xB} = \min\{n: S_n\le B|\ S_0 = 0, D_0 = d_0\} \Rightarrow \{T^S_{xA} > n\} = \{\min_{1\le k\le n} S_k > B|\ S_0 = 0, D_0 = d_0\}.$$

Let $f(x_1,x_2,..,x_n) = \min_{1\le k\le n}\left\{\sum_{0\le i\le k} x_i\right\}$ Again f( ) is a componentwise nondecreasing function. Then the first passage time

$$\{T^U_{xB}>n\}=\{f(Z^U_1,Z^U_2,...Z^U_n) > B\},$$

$$\{T^S_{xB}>n\}=\{f(Z_1,Z_2,...Z_n) > B\}, \text{ and}$$

$\{T^L_{xB}>n\} = \{f(Z^L_1,Z^L_2,...Z^L_n)>B\}$.

So, Lemma 4 implies that

$$P\{f(Z^U_1,Z^U_2,...Z^U_n)>B\}\le P\{f(Z_1,Z_2,...Z_n)>B\}\le P\{f(Z^L_1,Z^L_2,...Z^L_n)>B\},$$

so we have the relation $P\{T^U_{xB}> n\}\ge P\{T^S_{xB}> n\} \ge P\{T^L_{xB}> n\}$. Q.E.D.

4.4 Method for Bounding First Passage Times

Theorems 1 and 2 yield a method for bounding the first passage distribution of the sum S_n which has absolutely continuous increments Z_i, i=1,2,..., and is defined on the Markov Chain $\{D_n, n\ge 0\}$. By

constructing discrete-valued random variables Z^U_i and Z^L_i so that $Z^U_i \geq_{st} Z_i \geq_{st} Z^L_i$ we can use the well-established finite state Markov chain methods to find the first passage time distributions of the corresponding MAP's (U_n,D_n), and (L_n,D_n) which will provide the desired bounds. An illustration of the order relationship between Z^U_i, Z_i, and Z^L_i is given in Figure 2.

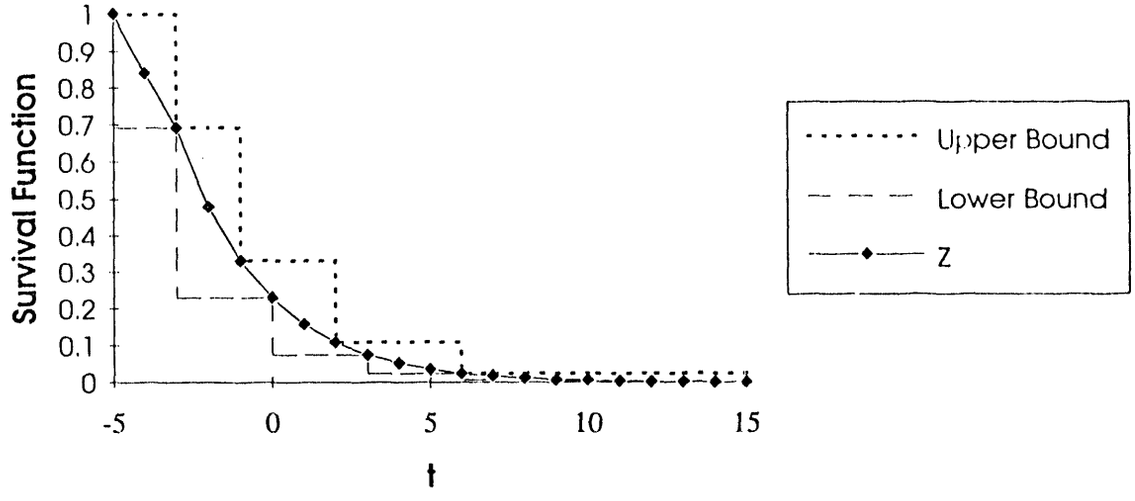


Figure 2. Relationship between the Survival Functions of the Upper Bound Z^U, the Lower Bound Z^L, and Z

Constructing the discrete random variables Z^U_i and Z^L_i is straightforward. To this end, let Z_i be a real-valued uniformly bounded continuous random variable with $|Z_i| \leq L$. Consider, the first passage time T^S_{xA} where $E[T^S_{xA}] < \infty$. We partition the interval [-L,L] into M disjoint intervals E_k, k=1,2,...M so that $\bigcup_k E_k$ = [-L,L], and define probabilities $p_k = P\{Z_i \in E_k\}$, k=1,2...M. The values p_k;k=1,2,...M become the probabilities of being in each discrete state and since Z_i is continuous, this partitioning of the state space is the same for both discrete random variables Z^U_i and Z^L_i. The difference between the two discrete random variables is the values assigned to each of the M discrete states. In particular, for $Z^L_i \leq_{st} Z \leq_{st} Z^U_i$,

divide the interval [-L, L] into M intervals each of length w = 2L/M, so that

$$E_k = (a_k, b_k] = (-L+w(k-1), -L+wk] \text{ for } k=1,2,...M$$

and let $Z^L_i = a_k$ with probability p_k and $Z^U_i = b_k$ with probability p_k. This discretization procedure is illustrated in Figure 3.

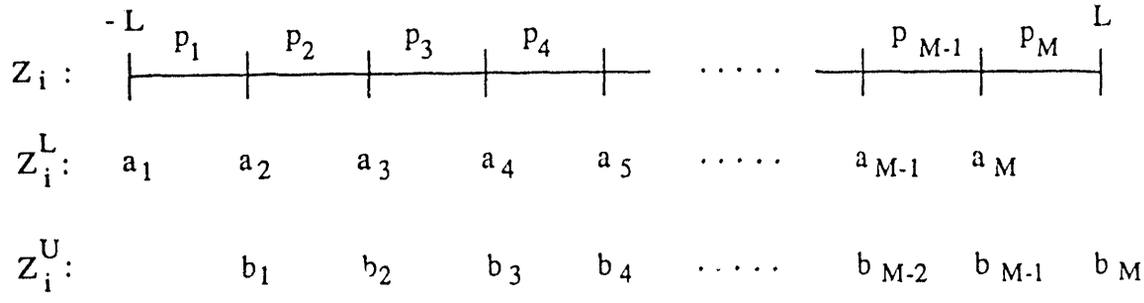


Figure 3. Method for Discretizing the State Space of Z_i to Bound First Passage Times

Note that by defining another discrete random variable whose values lie in the middle of each interval E_k, k=1,2,...M the resulting process would yield a first passage time distribution (mean) in between that for the process Z^U_i and Z^L_i and may yield a good approximation for Z_i.

This technique can be used as easily for the Markov Additive case as for the case where the Z_i are i.i.d. Assuming Z_i is uniformly bounded for any transition of the Markov chain (i.e $\Gamma_{rs}(-L_{rs}) = 0$ and $\Gamma_{rs}(L_{rs}) = 1$ for some L_{rs}, and all r,s $\in \Omega$), then let $L = \max_{r,s} \{L_{rs}\}$ and discretize the state space of each c.d.f. as in Figure 3.

Computation of first passage time distributions for the discrete-valued Markov chains is straightforward.

The first passage time to a recurrent subset of states can be computed by making the subset absorbing, and utilizing the following method. Given a finite Markov chain with transition probability matrix $\underline{P}$ and initial distribution $\underline{\pi}_0$, suppose the state space consists of a set T of transient states, so that $\underline{P}$ has the following structure:

$$\underline{P} = \begin{bmatrix} \underline{P_1} & \underline{0} \\ \underline{R} & \underline{Q} \end{bmatrix} \tag{4.4}$$

where P_1 is an r x r matrix corresponding to the subset of recurrent states. Let $\underline{M} = (\underline{I} - \underline{Q})^{-1}$ denote the fundamental matrix, and let T denote the set of transient states. Then for $i,j \in T$, $N_{ij}; i,j \in T$ is the number of times, starting from i, that state j is visited before the process leaves T.

It is well known that $\underline{M} = [\, E[N_{ij}] \,]$ for $i,j \in T$ (see, e.g. Bhat [6]). Let $\underline{e}$ denote a column vector of all 1's. Then the expected first passage time to the subset T^C after starting in some state in T according to the initial distribution $[\underline{\pi}_{0:i}]_{i \in T}$ is given by $\underline{\pi}_0^T \underline{M}\, \underline{e}$ the expected number of visits to all states in T before leaving the transient set. The distribution of the passage time to T^C, denoted here by N, is given by

$$P\{N>n\} = \underline{\pi}_0^T \underline{Q}^n \underline{e} \tag{4.5}$$

Therefore, to compute bounds on the distribution of the run length for an SPRT, we use the Markov chains (L_n, D_n) and (U_n, D_n) constructed by discretizing the state space of Z_i as in Figure 3, set $T = \{i : i < A\}$ and use equation (4.7).

The approach taken here is an extension of the one developed for the computing the run length of a one-sided CUSUM test by Brook and Evans [7], and of a two-sided CUSUM test by Woodall [30]. Since the CUSUM test is a random walk in the interval [0,A), Brook and Evans [7] proposed dividing this interval

into M subintervals of equal length, and placing a discrete state in the center of each interval and assigning it a probability associated with being in that interval. Analysis of the corresponding discrete Markov chain gives their approximation to the run length distribution of the CUSUM procedure. Their approach provides no assessment of the accuracy of the approximation. The method developed here modifies their technique so that actual bounds on run length distribution can be obtained, and generalizes their approach to non-i.i.d. process behavior. Furthermore, in the following section we prove that the bounds obtained by the method developed here approach the exact distribution asymptotically.

4.5 Convergence Results

The above method requires that the number of discrete states M in the approximation be specified. It is therefore of useful to know how tight the bounds on the first passage distribution are for a given M. Let $\Rightarrow^d$ denote convergence in distribution. The following Theorem proves that the bounds become exact as the number of discrete states approaches ∞.

Theorem 3: For $T^U_{xA} = \min\{n: U_n \geq A\}$ and $T^L_{xA} = \min\{n: L_n \geq A\}$, $T^U_{xA} \Rightarrow^d T^S_{xA}$ and $T^L_{xA} \Rightarrow^d T^S_{xA}$ as the number of discrete states $M \rightarrow \infty$.

Proof: The result is proven for T^U_{xA} using a standard coupling argument. Suppose we have M discrete states. Let I_C denote the indicator function on a set C. Define a process

$$Z^{\#}_i = \sum_k (wk - L)\, I_{\{w(k-1)-L < Z_i \leq wk-L\}}$$

with $Z^{\#}_0 = x$, $U^{\#}_n = \sum_i Z^{\#}_i$, and $T_{UA}{}^{\#} = \min\{n: U^{\#}_n \geq A\}$. Then clearly

$$P\{Z^{\#}_i = wk-L\} = P\{Z^U_i = wk-L\} \text{ for all } k=1,2,\ldots M$$

and so $T^{\#}_{UA} =^d T^U_{xA}$. Now suppose $w(j-1)-L < Z_i \leq wj-L$ for some $j \in \{1,2,\ldots M\}$. Then we have

$$|Z_i - Z^{\#}_i| < w = 2L/M.$$

Since Z_i is uniformly bounded by $L < \infty$, given $\varepsilon>0$ we can find $M_0 = \lceil 2L/\varepsilon \rceil$ so that for all $M > M_0$, $|Z_i - Z^{\#}_i| < w \le \varepsilon$. Therefore, $Z^{\#}_i \to Z_i$ as $M\to\infty$ pointwise (almost surely). Therefore, since T^S_{xA} is finite a.s., $T^{\#}_{UA} \to T^S_{xA}$ a.s. Finally since $T^{\#}_{UA} =^d T^S_{xA}$, then $T^U_{xA} \Rightarrow T^S_{xA}$ as $M\to\infty$. The proof of T^L_{xA} is identical and uses an analogous coupled process. Q.E.D.

Theorem 3 ensures that as the number of discrete state in the approximation becomes very large, the passage time distribution will converge to the exact distribution. The following result demonstrates a monotonic relationship between the bounds as they converge.

Theorem 4: Let T^U_{xA} (M) and T^L_{xA}(M) denote the upper and lower bound passage times when M discrete states are used in the approximation. Consider the subsequence $n(M) = 2^M$ for M=1,2,.... Then T^L_{xA} (n(M)) $\ge T^L_{xA}$ (n(M+1)) , T^U_{xA} (n(M)) $\le T^U_{xA}$ (n(M+1)), T^U_{xB} (n(M)) $\ge T^U_{xB}$ (n(M+1)) , and T^L_{xB} (n(M)) $\le T^L_{xB}$ (n(M+1) a.s. for all M=1,2,...

Proof: The result will be proven for T^U_{xA} (n(M)) . The proof for the other three cases is analogous. Let $w_{n(M)}$ be the interval width when M discrete states are used. As before, define a process

$$Z^*_i(n(M)) = \sum_{k \le n(M)} (-L+wk) I_{\{-L+w(k-1) < Zi \le -L+wk\}}.$$

Then $P\{Z^*_i(n(M)) = -L+wk\} = P\{Z^U_i(n(M)) = -L+wk\}$ for all k=1,2,...n(M). Now suppose

$$-L+w_{n(M)}(j-1) < Z_i \le -L+w_{n(M)}j \text{ for some } j \in \{1,2,...n(M)\}.$$

Then $Z^*_i(n(M)) = -L+w_{n(M)}j$. Looking at n(M+1), n(M+!) = 2n(M), and $w_{n(M)} = 2w_{n(M+1)}$ so each interval is divided in half. Based on the value of Z_i, there are two possibilities for $Z^*_i(n(M+1))$.

(1) If Z_i lies in the upper half of the interval $(-L+w_{n(M)}(j-1),-L+w_{n(M)}j]$. Specifically, $L+2w_{n(M+1)}(j-1/2) < Z_i \le -L+2w_{n(M+1)}j$, then $Z^*_i(n(M+1)) = -L+2w_{n(M+1)}j = Z^*_i(n(M))$, or

(2) If Z_i lies in the lower half of the interval $(-L+w_{n(M)}(j-1)\ ,-L+w_{n(M)}j]$, or we have $L+2w_{n(M+1)}(j-1) < Z_i \leq -L+2w_{n(M+1)}(j-1/2)$, then

$$Z^*_i(n(M+1)) = -L+2w_{n(M+1)}(j-1/2) < Z^*_i(n(M)).$$

So $Z^*_i(n(M+1)) \leq Z^*_i(n(M))$ a.s. Therefore, $Z^U_i(n(M+1)) \leq_{st} Z^U_i(n(M))$ and applying Theorem 1 gives the desired result. Q.E.D.

5 Numerical Results

We can now obtain bounds on the run length of the SPRT for various types of process behavior. In particular, we are interested in the run length of the SPRT under (1) no process failure, (2) rapid process failure, and (3) graceful process failure. The first performance measure gives us the time to a false decision by the SPRT, while the second and third tell us how quickly the test will detect a failing process.

Figures 4 and 5 plot upper and lower bounds on the run length for the one sided SPRT with $\mu_0 = 0$, $\mu_1 = 1, \sigma^2 = 0.5$, and $\alpha_0 = \beta_0 = 0.01$. Figure 4 plots the run length distribution under no process failure with $M = 50$. The bounds on the run length distribution mimic one another in shape and remain quite close together yielding accurate information about the actual run length distribution of an SPRT.

Figure 5 demonstrates the mean run length as a function of discrete state space size M. Figure 5a plots mean run length under no process failure. The bounds are compared against a Monte Carlo simulation estimate based on the importance sampling method of Siegmund[26]. Figures 5b and 5c plot the expected run length under rapid process failure and graceful process failure, respectively. The bounds are compared against straightforward Monte Carlo simulation. The models for rapid process failure and graceful process failure were formulated as in Figure 3, where q was selected so that the mean time until the mean reaches

its critical level ν_R would correspond to a rate of failure commonly found in the nuclear reactor applications at ANL.

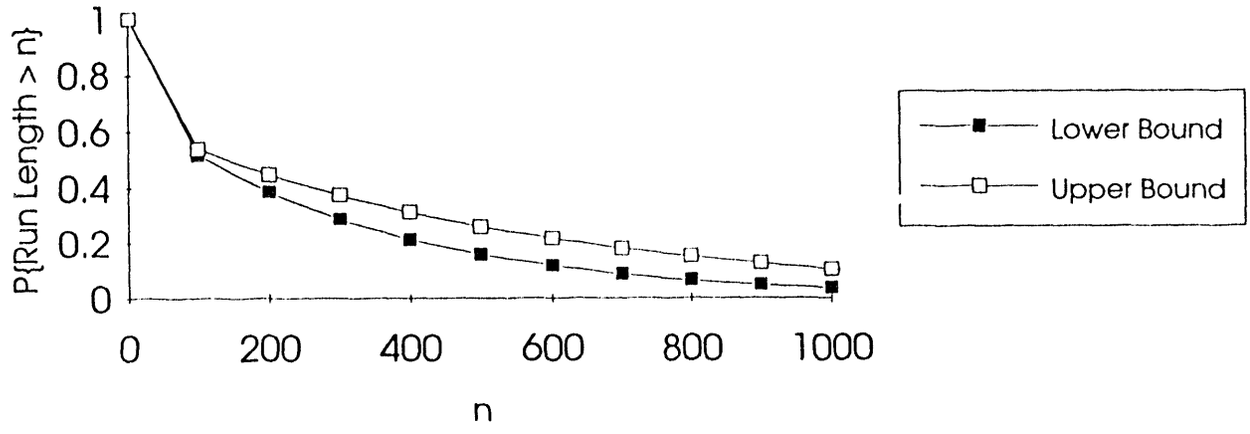
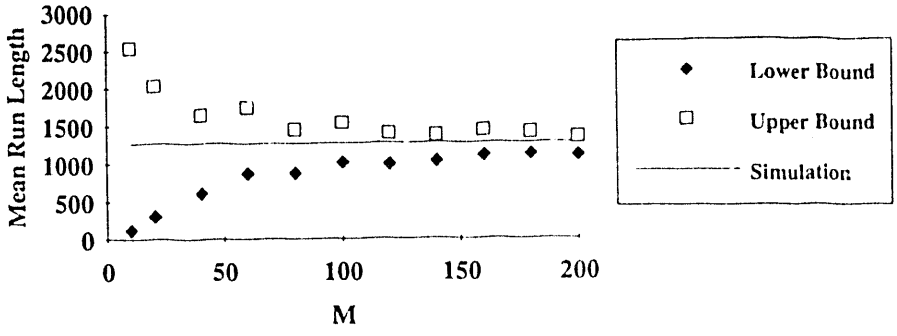


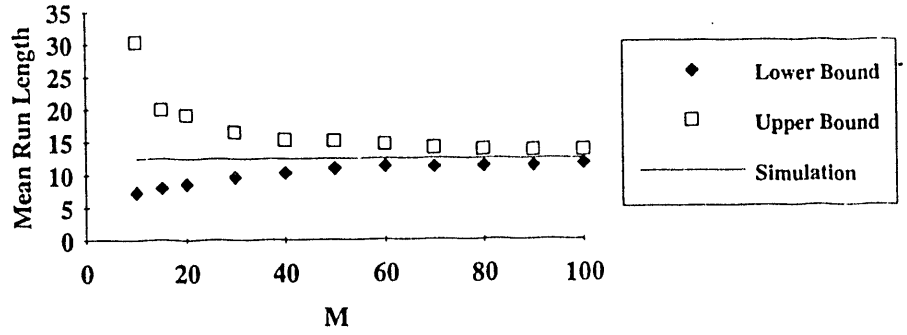
Figure 4. Run Length of the SPRT Under No Process Failure

Figure 5 demonstrates that the more quickly a process fails, the shorter the run length of an SPRT. For example, the mean run length of an SPRT under no process failure is 1424. This value decreases to 25 for a gracefully failing process and to 12 for a rapidly failing process. This is to be expected since an H_1 decision by the SPRT becomes more likely as the mean of an incoming process increases.

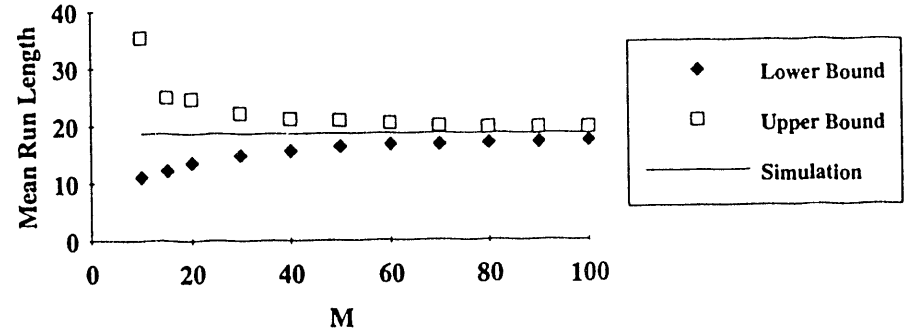
Figure 5 also indicates convergence of the bounds to the exact value as M becomes very large. The numerical results also suggest that a longer mean run length for the original problem will yield slower convergence of the bounds to the exact solution. For example, under no process failure, $M \geq 120$ is required to reach within 15% of the simulated mean run length for the upper and lower bounds. For gracefully failing and rapidly failing processes the bounds become within 15% of the simulated value for $M \geq 50$, and $M \geq 40$, respectively.



a) Run Length Under No Process Failure



b) Run Length Under Rapid Process Failure



c) Run Length Under Graceful Process Failure

Figure 5. Bounds on the Expected Run Length of the SPRT under Three Process Behaviors as a Function of Discrete State Space Size

6 Conclusion

In this work, we develop a useful technique for computing performance measures of sequential tests in process control and illustrate these methods through numerical examples. Development of this technique

has been motivated largely by the following: (1) a need to assess the performance of process control techniques for safety-critical or mission-critical applications, and (2) the desire to characterize these performance metrics in a variety of scenarios. The methods developed here serve as a basis for the design and performance analysis of expert systems for process control. These unique expert systems utilize numerous sequential tests running both simultaneously and in sequence for the purpose of detecting and diagnosing failures of process signals from highly redundant systems.

In the process of developing this methodology, we extend the current research in several ways. First, we formulate a Markov additive model for the SPRT which allows us to study the behavior of sequential tests under various non-i.i.d. process behavior, including under control processes, slowly failing processes, rapidly failing processes, and out of control processes. Using this innovative model, the performance measures of interest can be viewed as passage times of a Markov additive process.

Second, using stochastic order relations, we develop a technique for computing bounds on first passage time distributions of Markov additive processes. The technique is based on discretizing the state space of the original process in a particular manner, and can be used to obtain performance measures of interest not only in process control, but also in a variety of other contexts. In particular, passage times of Markov additive processes are of interest in many applications, and no established techniques exist for obtaining such performance measures in general. The method developed here provides a straightforward technique for obtaining not only moments, but the distribution of these performance measures.

Finally, the techniques developed here for passage times of the sequential probability ratio test can be extended to include other performance measures, and other applications. For example, the approach developed here may be used to compute the sample number and operating characteristic (the probability

a test decides H_1) of sequential tests under non-i.i.d. process behavior. Furthermore, sequential tests for correlated processes, such as Markov sequences or autoregressive processes, have been the focus of recent research. The methodology developed here can be generalized for the purpose of analyzing such tests. These issues are subjects of current and future research.

References

1. Bagshaw, M. and Johnson, R. "The Effect of Serial Correlation on the Performance of CUSUM Tests II", *Technometrics* (1975), v. 17, n. 1, pp 80.

2. Bajic, D. "New Simple Method for Solving First Passage Time Problem", *Electronic Letters* (1991), v 27, n 16, pp 1419-1421.

3. Berk, R. "Some Asymptotic Aspects of Sequential Analysis", *Ann. Stat.* (1973), v. 1, pp 126-138.

4. Bhat, U. Elements of Applied Stochastic Processes, John Wiley & Sons, Inc, New York, 1972.

5. Bhate, D. "Approximation to the Distribution of Sample Size for Sequential Tests, I. Tests for Simple Hypotheses", *Biometrika* (1955). v 42, pp 130-144.

6. Böhm, W. and Hackl, P. "Improved Bounds for the Average Run Length of Control Charts Based on Finite Weighted Sums", *Ann. Stat.* (1990), v. 18, n. 4, pp 1895-1899.

7. Brook, D. and Evans, D. "An Approach to the Probability Distribution of CUSUM Run Length", *Biometrika* (1972), v. 59, n. 3, pp 539-549.

8. Çinlar, E. "Markov Additive Processes. I", *Z. Wahrscheinlichkeitstheorie verw. Geb.* (1972), v. 24, pp 85-93.

9. Çinlar, E. "Markov Additive Processes. II", *Z. Wahrscheinlichkeitstheorie verw. Geb.* (1972), v. 24, pp 95-121.

10. Ghosh, B. and Sen, P, Handbook of Sequential Analysis, Marcel Dekker, Inc, 1991.

11. Gross, K. and Humenik, K., "Sequential Probability Ratio Test for Nuclear Plant Component Surveillance", *Nucl. Tech.* (1991), v. 93,131-137.

12. Gross, K, Singer, R, Humenik, K, and Walsh, M, "Expert System for Reactor Coolant Pump Surveillance", *International Fast Reactor Safety Meeting Conference Proceedings* (1990).

13. Humenik, K. and Gross, K., "Sequential Probability Ratio Tests for Reactor Signal Validation and Sensor Surveillance Applications", *Nucl. Sci and Eng.* (1990), v. 105, pp 383-390.

14. Kaspi, H. "Storage Processes with Markov Additive Input and Output", *Math. Oper. Res.* (1984), v. 9, n. 3, pp 424-440.

15. Keilson, J. and Rao, S.A Process with Chain Dependent Growth Rate'' *J. Appl Prob* (1970), v 7, p 699-711.

16. Keilson, J. and Wishart, D. "A Central Limit Theorem for Processes Defined on a Finite Markov

Chain", *Proc Camb Phil Soc* (1965), v. 60, pp 547-567.

17. Martinsek, A., "A Note on the Variance and Higher Central Moments of the Stopping Time of an SPRT", *J. Amer. Statist. Assoc.*(1981), v 76, pp 701-703.

18. Miller, H. "Absorption Probabilities for Sums of Random Variables Defined on a Finite Markov Chain", *Proc. Camb. Phil Soc* (1960)

19. Ney, P., and Nummelin, E. "Markov Additive Processes I. Eigenvalue Properties and Limit Theorems", *Ann Prob* (1987), v. 15, n. 2, pp 561-592.

20. Ney, P., and Nummelin, E. "Markov Additive Processes II. Large Deviations", *Ann Prob* (1987), v. 15, n. 2, pp 593-609.

21. Page, E. "Continuous Inspection Schemes", *Biometrika*, 41, 100-114.

22. Rego, V."Naive Asymptotics for Hitting Time Bounds in Markov Chains", *Acta Informatica* (1992), v 29, pp 579-594.

23. Reynolds, M. "Approximations to the Average Run Length in Cumulative Sum Control Charts", *Technometrics* (1975), v. 17, n. 1, pp 65-71.

24. Ross, Introduction to Probability Models, Wiley, 1989.

25. Sadowsky, J, "A Dependent Data Extension of Wald's Identity and Its Application to Sequential Test Performance Computation", *IEEE Trans. Info. Theory* (1989), v 35, n 4, pp 834-842.

26. Siegmund, D, "Importance Sampling in the Monte Carlo Simulation of Sequential Tests", *Ann. Stat.* (1976), v 4, n 4, pp 673-684.

27. Wald, A. Sequential Analysis, Wiley, 1947.

28. Waldmann, K. "Bounds on the Distribution of One-Sided and Two-Sided CUSUM Quality Control Schemes", *Technometrics* (1986), v. 28, n. 1, pp 61-67.

29. Woodall, W. "The Distribution of the Run Length of One-Sided Procedures for Continuous Random Variables", *Technometrics* (1983), v. 25, n. 3, pp 295-301.

30. Woodall, W. "On the Markov Chain Approach to the Two-Sided CUSUM Procedure", *Technometrics* (1984), v. 26, n. 1, pp 41-46.

DATE
FILMED
5/31/94
END