

Office of Naval Research

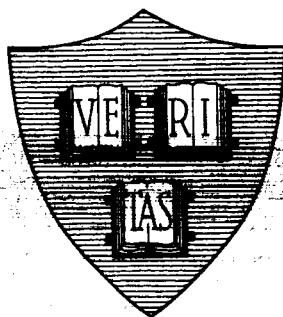
Contract Nonr-1866 (16)

NR-372-012

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Grant NGR 22-007-068

DETECTION OF CHANGES IN CHARACTERISTICS OF A GAUSS-MARKOV PROCESS



by

P. M. Newbold and Y. C. Ho

June 1967

Technical Report No. 531

"Reproduction in whole or in part is permitted by the U. S. Government. Distribution of this document is unlimited."

Division of Engineering and Applied Physics
Harvard University • Cambridge, Massachusetts

67-34543
(ACCESSION NUMBER)
43
(PAGES)
OR-87477
(NASA CR OR TMX OR AD NUMBER)

(THRU)
1
(CODE)
07
(CATEGORY)

FACILITY FORM 602

Office of Naval Research

Contract Nonr-1866(16)

NR - 372 - 012

National Aeronautics and Space Administration

Grant NGR-22-007-068

DETECTION OF CHANGES IN THE CHARACTERISTICS OF A
GAUSS-MARKOV PROCESS

By

P. M. Newbold and Y. C. Ho

Technical Report No. 531

Reproduction in whole or in part is permitted by the U. S.
Government. Distribution of this document is unlimited.

June 1967

The research reported in this document was made possible through support extended the Division of Engineering and Applied Physics, Harvard University by the U. S. Army Research Office, the U. S. Air Force Office of Scientific Research and the U. S. Office of Naval Research under the Joint Services Electronics Program by Contracts Nonr-1866(16), (07), and (32) and by the National Aeronautics and Space Administration under Grant NGR-22-007-068.

Division of Engineering and Applied Physics

/Harvard University Cambridge, Massachusetts

DETECTION OF CHANGES IN THE CHARACTERISTICS OF A GAUSS-MARKOV PROCESS

By

P. M. Newbold and Y. C. Ho

Division of Engineering and Applied Physics

Harvard University, Cambridge, Massachusetts

ABSTRACT

By an extension to the theory of sequential detection with dependent measurements, it is possible to develop a Sequential Probability Ratio Test (SPRT) to detect changes in regime in a Gauss-Markov Process rather than detecting which of two regimes exist. It is shown how a posterior form of this extended SPRT may be simplified to reduce computational complexity. The simplified SPRT's are in fact modifications of the original SPRT detecting the regime and not the change. The tests are applied to the problem of fault detection in a gyro navigational system: the results of a detailed computer simulation are given.

1. Introduction

It may often be of crucial importance to determine whether or not a fault condition has arisen in a stochastic system, at the same time as observations from the system are being used for some specific purpose. For example, if we treat a gyro navigational system as a stochastic system, it may be of vital interest to be certain that calculations based on the system output are not incorrect because of a fault condition. Although in this paper we shall only deal with faults which can be interpreted as changes in the variance of random inputs to a Gauss-Markov process, the theoretical developments outlined can equally be applied to the detection of other types of characteristic change which affect the output of the random sequence. The computational difficulty may be more or less involved depending on the nature of the change. In terms of the above example, the fault condition arising may be an increase in the variance of the drift rate of the gyro.

It should be noted that there are two distinct but related problems. The first is detecting whether the characteristics of the system have one hypothetical set of values as opposed to a second set. The second problem is the detection of a change of the characteristics from the first set of values to the second as it occurs. We shall deal with the second of these problems.

The detection of a characteristic change is a problem in statistical decision theory. The basic results used here are those of sequential analysis, originated by Wald [1]. The analysis determines from which of two probability distributions a sequence of uncorrelated samples comes, to within set error bounds. The procedure is to derive from the sequence

up till a time k a likelihood function $\theta(k)$. Its value is tested against two thresholds related to the set error bounds. If either of the thresholds are exceeded then the test is terminated by taking the appropriate decision as to which probability distribution is shown to be correct. If neither threshold is exceeded, then a $(k + 1)^{\text{th}}$ sample is taken and the same procedure repeated for the updated sequence. A detailed and very clear treatment of this Wald Sequential Probability Ratio Test (from now on abbreviated to SPRT) is given in [2].

The theory was extended to sequences of correlated samples by Bussgang and Middleton [3] and more recently a general extension has been derived by means of the state space approach by Schweppe [4]. In the present paper we shall use this approach exclusively.

Section 2 is devoted to a fairly brief exposé of Schweppe's solution, since although Schweppe only deals with the first of the two problems previously stated, the solution to the second problem is of a closely similar form. In Section 3 the complete solution to the second problem is given for the case where the characteristic change is a change in input variance. This solution retains the advantage of the SPRT in the solution of the first problem in that it has the property of being the test which on average requires the minimum number of samples of all tests to come to a decision. In Section 4 several simplified versions of the test are presented which offer considerable computational savings. An example of the test applied to a gyro navigational system is presented in Section 5. The Appendix is devoted to the derivation of performance figures for the simplified tests.

It is stressed again at this point that although the solutions are couched in terms of detecting a change in the input variance to a Gauss-Markov process, other forms of characteristic change can equally well be detected by this method.

2. The Sequential Probability Ratio Test for Correlated Samples

The aim of this section is to summarize the derivation of the SPRT detecting whether the noise input to a Gauss-Markov process has one variance or another. We do not concern ourselves with any possible change from one value to the other. The correlated samples which provide data for the test are the output from this Gauss-Markov process.

Let the process be characterized by the following equations:

$$x(k+1) = \phi x(k) + w(k) , \quad (1)$$

$$z(k) = Hx(k) + v(k) ; \quad (2)$$

where $x(k)$, $w(k)$ are vectors and the observation $z(k)$ is a scalar. $v(k)$ is a random observation error of zero mean and variance R . At any time k we are required by performing a test on the sequence of measurements $\mathcal{Z}(k) = \{z(0), z(1) \dots z(k)\}$ to find which of the following two hypotheses is true:

Hypothesis \mathcal{H}_0 : the variance of $w(k)$ is given by $Q = Q_0$

Hypothesis \mathcal{H}_1 : the variance of $w(k)$ is given by $Q = Q_1$

It is convenient but not crucial to assume that the mean of $w(k)$ is identically zero. At any time k the SPRT consists of determining the likelihood function

$$\theta_k = \log \frac{p\{\mathcal{Z}(k)/\mathcal{H}_1\}}{p\{\mathcal{Z}(k)/\mathcal{H}_0\}} , \quad (3)$$

where $p\{z(k)/H_0\}$ is the joint probability distribution of the sequence under the assumption of hypothesis H_0 , and $p\{z(k)/H_1\}$ is the corresponding distribution under the assumption of the alternate hypothesis. For independent samples we can write Eqn. (3) in the recursive logarithmic form:

$$\theta_k = \theta_{k-1} + \log \frac{p\{z(k)/H_1\}}{p\{z(k)/H_0\}} \quad (4)$$

The value of θ_k is then tested against two threshold levels A and B (where $A > B$):

If $\theta_k \leq B$ the test is terminated with the choice of H_0 .

If $\theta_k \geq A$ the test is terminated with the choice of H_1 .

If $A > \theta_k > B$ a $k+1^{\text{th}}$ sample is taken and the test is repeated.

The thresholds A and B are chosen in the following manner. We specify two error probabilities, α and β :

α is the probability of choosing H_1 when H_0 is true;

β is the probability of choosing H_0 when H_1 is true.

Then

$$A = \log \frac{1 - \beta}{\alpha} \quad \text{and} \quad B = \log \frac{\beta}{1 - \alpha} \quad (5)$$

Now since the sequence of samples is in reality correlated we can no longer reduce Eqn. (3) to Eqn. (4), and the convenient recursive form is lost. However, luckily it is possible to circumvent this problem. It is well known that the Kalman filter which gives the optimal estimate $\hat{x}(k/k)$ [given the sequence of measurements $z(k)$] of the state of a Gauss-Markov process has the property that the sequence of errors generated between the measurement $z(k)$ and its estimate $\hat{z}(k/k-1)$ is independent.

Let us present this more precisely. Since we have two hypothetical Gauss-Markov processes, differing by the value of Q , we shall have two Kalman filters. In their equations given below, the subscripts correspond to the subscript of the corresponding hypothesis.

$$\left. \begin{aligned} \hat{x}_i(k+1/k+1) &= \phi \hat{x}_i(k/k) + K_i(k+1)\{z(k+1) - H\phi \hat{x}_i(k/k)\} & (6) \\ K_i(k+1) &= P_i(k+1)H^T R^{-1} & (7) \\ M_i(k+1) &= \phi P_i(k)\phi^T + Q_i & (8) \\ P_i(k+1) &= M_i(k+1) - \frac{M_i(k+1)H^T H M_i(k+1)}{H M_i(k+1)H^T + R} & (9) \end{aligned} \right\} i = 0,1$$

Then the two sequences

$$\mathcal{E}_{H_0}(k+1) = \{\varepsilon_0(0), \varepsilon_0(1) \dots \varepsilon_0(k+1)\};$$

$$\mathcal{E}_{H_1}(k+1) = \{\varepsilon_1(0), \varepsilon_1(1) \dots \varepsilon_1(k+1)\};$$

where

$$\varepsilon_i(k) = z(k) - \hat{z}_i(k/k-1) = z(k) - H\phi \hat{x}_i(k-1/k-1); \quad i = 0,1 \quad (10)$$

are each independent random sequences.

To calculate θ_k we need to know the mean and variance of these sequences. It may easily be demonstrated that these are given by

$$E\{\varepsilon_i(k)\} = 0, \quad \text{Var}\{\varepsilon_i(k)\} \triangleq \sigma_i^2(k) = \frac{R}{1 - HK_i(k)}; \quad i = 0,1. \quad (11)$$

If we further suppose that before the start of the test the Kalman filters have been run for a sufficient number of time intervals for the variance equations, Eqns. (8)-(9), to have reached their equilibrium solution, we

need only consider two asymptotic values of $\text{Var}\{\epsilon_i(k)\}$:

$$\sigma_i^2 = R/(1 - HK_i) ; i = 0,1 \quad (12)$$

The SPRT is now carried out using the samples of ϵ_{H_0} and ϵ_{H_1} .

We have

$$\theta_{k+1} = \log \frac{p\{\mathcal{E}(k+1)/H_1\}}{p\{\mathcal{E}(k+1)/H_0\}} \quad (13)$$

which reduces to

$$\theta_{k+1} = \theta_k + \log \frac{p\{\epsilon_1(k+1)\}}{p\{\epsilon_0(k+1)\}} = \theta_k + \log \left(\frac{\sigma_0}{\sigma_1} \right) - \frac{1}{2} \left[\frac{\epsilon_1^2}{\sigma_1^2} - \frac{\epsilon_0^2}{\sigma_0^2} \right]_{k+1} \quad (14)$$

The testing algorithm is as follows:

- (a) take a new measurement $z(k+1)$;
- (b) update the estimation equations

$$\hat{x}_i(k+1/k+1) = \phi \hat{x}_i(k/k) - K_i \{z(k+1) - H\phi \hat{x}_i(k/k)\} ; \quad (15)$$

$$i = 0,1 ;$$

- (c) derive the errors $\epsilon_i(k+1)$, $i = 0,1$;
- (d) form the new likelihood function given by Eqn. (14) ;
- (e) test against the threshold levels as previously explained —

if the test does not terminate, take a new sample and repeat the operation.

The remaining point to be cleared up is what initial value to give θ_k before the start of the test. For example if we choose $\theta_0 = 0$ it would imply that before the start of the test we had no prior knowledge whether H_0 or H_1 is true. We shall see how the starting values are connected with different prior assumptions in the following sections.

3. The Exact Solution to the Problem of the Detection of a Change

In this section we solve the problem of the detection of a change in the variance of the input to the Gauss-Markov process represented by Eqns. (1)-(2). Consider a ratio test which is started at time $k = n$. At a subsequent time $n + j$ if we have made no decisions and are still testing, we are uncertain whether

- (a) no change has occurred in the system at all to date;
- (b) the change occurred before the beginning of the test;
- (c) the change occurred immediately after one of the intervals $n + i$;
 $i = 0, 1, 2, \dots, j-1$.

Thus at time $n + j$ we can list $j + 2$ exhaustive and mutually exclusive hypotheses:

- Hypothesis H_0 : no change has occurred at any instant;
- Hypothesis H_{10} : the change occurred before the test started; (16)
- Hypothesis H_{1i} : the change occurred immediately after interval
 $n + j - i$; $i = 1, 2, \dots, j$.

To each one of these hypotheses we can assign probabilities of occurrence prior to the start of the test. Let these be denoted respectively by

$$p\{H_0\} , p\{H_{10}\} , p\{H_{1i}\} ; i = 1, 2, \dots, j . \quad (17)$$

Note, however, that hypotheses H_0 , H_{1i} ; $i = 1, 2, \dots, j$ assume in each case that at time $k = n$ the process is in the pre-change state. Thus

$$p\{H_0\} = p\{H_{1i}\} ; i = 1, 2, \dots, j . \quad (18)$$

Taking into account this prior information, we can write down the expression

for a posterior probability ratio function:

$$\theta_j = \frac{\frac{1}{j} \sum_{i=1}^j p\{\mathbf{E}(n+j)/\mathcal{H}_{1i}\} p\{\mathcal{H}_0\} + p\{\mathbf{E}(n+j)/\mathcal{H}_{10}\} p\{\mathcal{H}_{10}\}}{p\{\mathbf{E}(n+j)/\mathcal{H}_0\} p\{\mathcal{H}_0\}} \quad (19)$$

Notice that this time because of the summation of the probabilities in the numerator we have not been able to use the convenient logarithmic form. Of itself this does not detract from the usefulness of the test, but makes the analytic determination of the mean and variance of θ_j very difficult.

Let us look more closely at some of the implications of this result. First of all, if we cannot assign prior probabilities of occurrence of each hypothesis, then we assume $p\{\mathcal{H}_0\} = p\{\mathcal{H}_{10}\}$ and Eqn. (19) reduces to:

$$\theta_j = \frac{1}{j} \sum_{i=1}^j \frac{p\{\mathbf{E}(n+j)/\mathcal{H}_{1i}\}}{p\{\mathbf{E}(n+j)/\mathcal{H}_0\}} + \frac{p\{\mathbf{E}(n+j)/\mathcal{H}_{10}\}}{p\{\mathbf{E}(n+j)/\mathcal{H}_0\}} \quad (20)$$

There is another, much more important implication of this result. In the previous section we saw that two Kalman filters were required, one based on the assumption of hypothesis \mathcal{H}_0 and the other on assumption of \mathcal{H}_1 . In the present case as may be expected, we again need one Kalman filter per hypothesis, that is, $j + 2$ filters.

Let us denote the Kalman filter based on \mathcal{H}_0 by \mathcal{F}_0 and the Kalman filters based on \mathcal{H}_{1i} by \mathcal{F}_{1i} ; $i = 0, 1, 2, \dots, j$. Table 3.1 below summarizes the relations between the hypotheses, filters, and values of input variance Q .

Hypothesis	Filter	Value of Q
H_0	F_0	Q_0
H_{10}	F_{10}	Q_1
H_{11} \vdots H_{1i} \vdots H_{1j}	F_{11} \vdots F_{1i} \vdots F_{1j}	Q_0 up to time $n + j - i$; Q_1 from $n + j - i + 1$ to $n + j$

Table 3.1 Kalman filters required for detecting changes.

Although there are only two basic structures for the filters, except in trivial cases the number cannot be reduced because each filter is subject to a transient at a different time interval, since the corresponding hypothesis dictates a change in Q at that time interval.

Thus, as opposed to the SPRT discussed in the previous section, the computations for this test get progressively more voluminous as time goes on. Even if the test is truncated after a fixed number of intervals and restarted, this is a great disadvantage, and has led to the development of the simplifications introduced in the following sections.

Let us now consider how to derive an iterative relation for the likelihood function in the form given in Eqn. (19). Remembering that $\mathcal{E}(k)$ still represents an independent sequence, Eqn. (19) can be written

$$\theta_j = \frac{1}{j} \sum_{i=1}^j \frac{\prod_{s=1}^j p\{\epsilon(n+s)/\mathcal{H}_{1i}\}}{\prod_{s=1}^j p\{\epsilon(n+s)/\mathcal{H}_0\}} + \frac{\prod_{s=1}^j p\{\epsilon(n+s)/\mathcal{H}_{10}\} p\{\mathcal{H}_{10}\}}{\prod_{s=1}^j p\{\epsilon(n+s)/\mathcal{H}_0\} p\{\mathcal{H}_0\}} ; \quad (21)$$

or since

$$p\{\epsilon(n+s)/\mathcal{H}_{1i}\} = p\{\epsilon(n+s)/\mathcal{H}_0\} ; \quad s = 1, 2, \dots, j-1 ,$$

$$\theta_j = \frac{1}{j} \sum_{i=1}^j \prod_{s=j-i+1}^j \frac{p\{\epsilon(n+s)/\mathcal{H}_{1i}\}}{p\{\epsilon(n+s)/\mathcal{H}_0\}} + \prod_{s=1}^j \frac{p\{\epsilon(n+s)/\mathcal{H}_{10}\} p\{\mathcal{H}_{10}\}}{p\{\epsilon(n+s)/\mathcal{H}_0\} p\{\mathcal{H}_0\}} . \quad (22)$$

As in the previous section we suppose that before the test started at time $k = n$ the Kalman filters had been run for a sufficient number of time intervals for the variance equations to have reached their equilibrium solutions. Table 3.2 gives the notation for the asymptotic values of the gains $K(k)$ and variances $\text{Var}\{\epsilon(k)\}$ for each filter.

Hypothesis	Filter	Asymptotic values at start of test as $k \rightarrow \infty$	
		Filter gain $K(k)$	Variance of error $\sigma^2(k)$
\mathcal{H}_0	\mathcal{F}_0	K_0	σ_0^2
\mathcal{H}_{10}	\mathcal{F}_{10}	K_{10}	σ_{10}^2
\mathcal{H}_{11} ⋮ \mathcal{H}_{1j}	\mathcal{F}_{11} ⋮ \mathcal{F}_{1j}	K_0	σ_0^2

Table 3.2 Asymptotic values for the Kalman filters.

Because some of the Kalman filters change their parameters during the test we need to precalculate the generated transients in the gains and variances as well as just the asymptotes. Figure 1 gives the notation used. We can

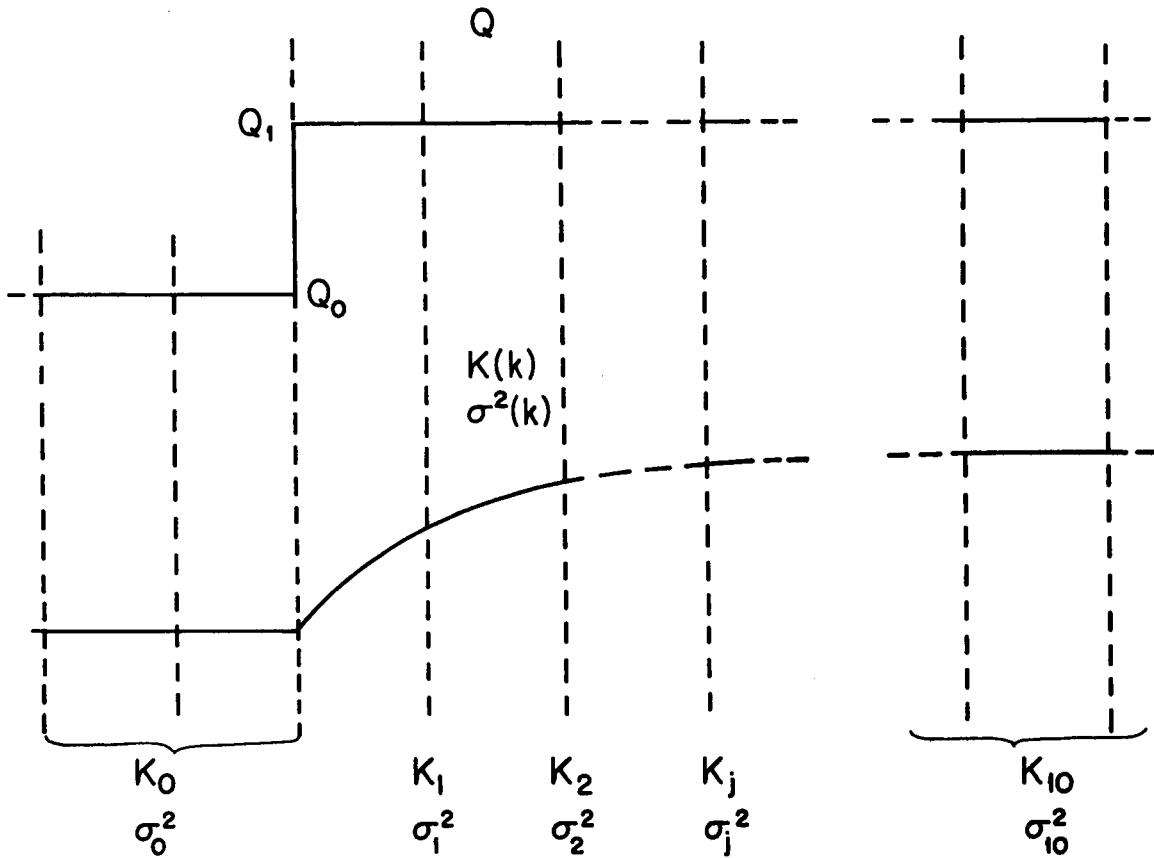


FIG. 1 NOTATION FOR FILTER GAIN $K(k)$ AND ERROR VARIANCE $\sigma^2(k)$ DURING TRANSIENT.

now write down the complete testing algorithm. At the j^{th} interval after the start of the test:

(a) take a new measurement $z(n + j)$;

(b) update the $j + 2$ estimation equations of the type of Eqn. (15) according to the logical schema of Figure 2. ;

(c) derive the $j + 2$ errors $\epsilon_0(n + j)$, $\epsilon_{1i}(n + j)$; $i = 0, 1, 2, \dots, j$;

(d) form the new likelihood function, which for $p\{\mathcal{H}_{10}\} = p\{\mathcal{H}_0\}$ may be reduced from Eqn. (22) to the following explicit recursive form:

$$\theta_j = \frac{1}{j} \sum_{i=1}^j \{\psi_i(n + j)\} + \psi_{10}(n + j) \quad ; \quad \psi_0(n + j) = 1 \quad (23)$$

$$= \frac{1}{j} \sum_{i=1}^j \left\{ \frac{\sigma_0}{\sigma_i} \exp - \frac{1}{2} \left[\frac{\epsilon_i^2}{\sigma_i^2} - \frac{\epsilon_0^2}{\sigma_0^2} \right]_{n+j} \psi_{i-1}(n + j - 1) \right\}$$

$$+ \frac{\sigma_0}{\sigma_{10}} \exp - \frac{1}{2} \left[\frac{\epsilon_{10}^2}{\sigma_{10}^2} - \frac{\epsilon_0^2}{\sigma_0^2} \right]_{n+j} \psi_{10}(n + j - 1) ;$$

(e) test as before against two threshold levels A and B . If the test does not terminate, go back to step (a) and repeat the procedure.

There is a slight difference in the derivation of the threshold values for this test. We specify the following error probabilities:

α : probability of rejecting hypothesis \mathcal{H}_0 when true;

β_i ; $i = 0, 1, 2, \dots, j$: probability of rejecting \mathcal{H}_{1i} when true;

and then the thresholds are given by

$$A = \frac{1 - \beta}{\alpha} ; B = \frac{\beta}{1 - \alpha} ; \text{ where } \beta = \frac{1}{j + 1} \sum_{i=0}^j \beta_i . \quad (24)$$

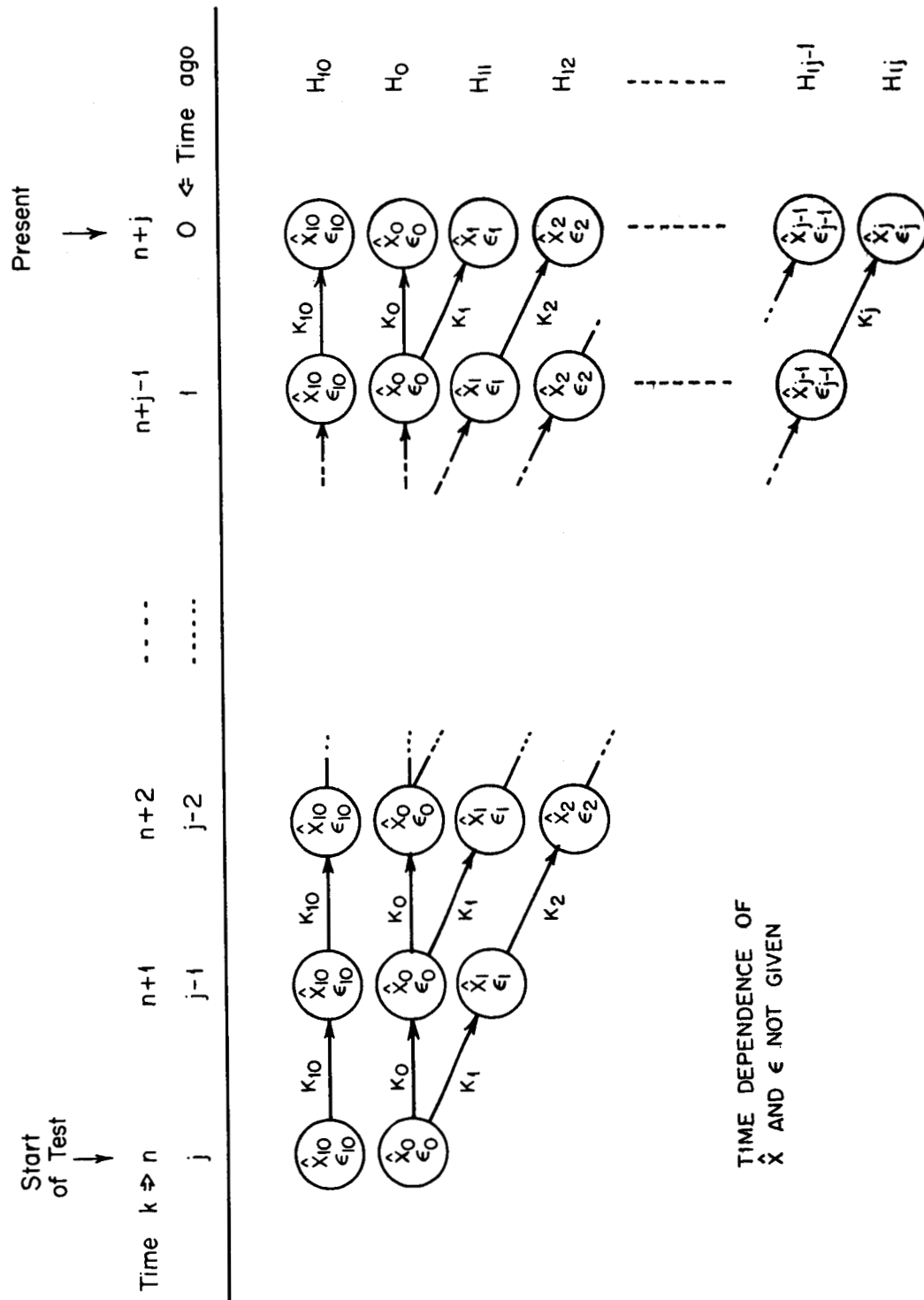


FIG. 2 LOGICAL SCHEMA FOR CALCULATION OF ESTIMATES AND ERRORS

This completes the derivation of the algorithm for testing a Gauss-Markov process for a change in a characteristic. As well as the obvious disadvantage of the increase in volume of computation as the test progresses, there is a second problem with this test. Whereas the SPRT described in Section 2 has a corresponding continuous time version, which can be obtained by taking the limit as the sampling interval decreases to zero, if this is attempted with the present SPRT we find we need an infinite number of Kalman filters. This is clearly impracticable.

In the next section, then, we remove these objections by obtaining some simpler forms of test. We should expect, although we do not prove that these tests will not perform as well as the exact test. In Section 5 experimental results show that the performance loss may be acceptable for some applications, at least.

4. Simplified Change Detection Tests

In the previous section we saw the need for simplified versions of the exact test for detecting a change in the characteristics of a process. The simplifications may be considered as arising from simplification of the expression

$$\theta_j = \frac{1}{j} \sum_{i=1}^j \frac{p\{\mathcal{E}(n+j)/\mu_{1i}\}}{p\{\mathcal{E}(n+j)/\mu_0\}} + \frac{p\{\mathcal{E}(n+j)/\mu_{10}\}p\{\mu_{10}\}}{p\{\mathcal{E}(n+j)/\mu_0\}p\{\mu_0\}} \quad (25)$$

One simplification is to assume that the process is known not to have changed before the test started. This removes the second term from Eqn.

(25) since $p\{\mu_{10}\} = 0$, leaving:

$$\theta_j = \frac{1}{j} \sum_{i=1}^j \frac{p\{\mathcal{E}(n+j)/\mu_{1i}\}}{p\{\mathcal{E}(n+j)/\mu_0\}} \quad (26)$$

This may very well be a reasonable assumption; unfortunately it does not lead to a significant reduction in computational effort because the summation term is still present in the expression.

If we therefore remove this term instead, giving

$$\theta_j = \frac{p\{\mathcal{E}(n+j)/\mu_{10}\}}{p\{\mathcal{E}(n+j)/\mu_0\}} \cdot \frac{p\{\mu_{10}\}}{p\{\mu_0\}} \quad (27)$$

what we are now saying in effect is that the change cannot ever occur during the test. This is now merely the form of the SPRT for the first problem,

which we dealt with in Section 2. The factor $\frac{p\{\mu_{10}\}}{p\{\mu_0\}}$ may be inter-

preted as the initial value θ_n . Although on the face of it this simplification is not what we want, we can in fact derive a very useful form of test from it. The idea is that to make the best use of prior information, each test should start out using the result of the previous test. So we write

$$\theta_j = \frac{p\{\mathcal{E}(n+j)/\mu_{10}\}}{p\{\mathcal{E}(n+j)/\mu_0\}} \cdot B \quad (28)$$

since from Eqn. (24), setting $\beta = \beta_0$ we have

$$p\{\mu_{10}\} = \beta ; p\{\mu_0\} = 1 - \alpha ; \text{ and thus } \theta_n = \frac{\beta}{1 - \alpha} = B .$$

Using Eqn. (28) implies that we are certain to within the error bound of

the previous test that no change has occurred. We can represent the effect of a loss of confidence during the time lapse between tests by writing

$$\theta_j = \frac{p\{\mathcal{E}(n+j)/\mathcal{H}_{10}\}}{p\{\mathcal{E}(n+j)/\mathcal{H}_0\}} \cdot \lambda B, \quad (29)$$

where λ might be a positive, monotonically decreasing function of the time lapse, depending upon any assumed failure distribution for the system.

Essentially, we are now considering again the original form of SPRT, but with the additional feature of incorporating knowledge available prior to the start of the test. (To use an analogy culled from estimation theory, the relationship between the usual SPRT and the present test as in Eqn. (28) is analogous to that between a least squares fit filter and a Kalman filter.)

We shall call this form of testing 'occasional testing'. Operationally, the present test implies that the change in the input variance occurs between tests only. All prior knowledge given or gained during the previous tests on \mathcal{H}_{10} or \mathcal{H}_0 and the failure mechanism are summarized in the initial value of θ_k at the start of a test.

There are two drawbacks to this test. If the change occurs during a test and not, as assumed, between tests, then we do not immediately know what is going to happen. If the change does occur between tests then we lose time by having to wait until the next test to detect it.

In spite of this, the test may still be useful in situations where it is not crucially important to detect a change immediately after it happens.

Each of the previous tests terminate if either of the two thresholds A or B is crossed. We can define a 'continuous test' which only

terminates when the threshold A is crossed, indicating a change. Before the change occurs, of course, the value of θ_k will move toward threshold B . The rule

$$\theta_j = \text{Max} \left\{ \frac{p\{\mathcal{E}(n+j)/\mathcal{H}_{10}\}}{p\{\mathcal{E}(n+j)/\mathcal{H}_0\}}, B \right\} \quad (30)$$

is used and testing is continued normally if B is reached. In other words, we immediately start a new test as soon as \mathcal{H}_0 is accepted. This form of test is approximately equivalent to the 'occasional test' without confidence loss, but starting a new test immediately the change occurs. Thus intuitively, the 'continuous test' will always do better than the 'occasional test', on average.

It is possible to calculate the average number of samples required to reach a decision fairly readily in the case of the simplified tests. A derivation of the average sample numbers in the 'continuous test' for the scalar case is given in the Appendix. The corresponding calculation for the 'exact test' is very difficult because the convenient logarithmic form cannot be used, and is not given here.

To summarize the last two sections, we have described three types of test by which we can detect a change in variance of the input to a process. The first test, the 'exact test', was designed to detect the change itself as it occurred. The 'occasional test' and the 'continuous test' were adaptations of the simple SPRT detecting only whether one condition or another existed, but incorporated the confidence gained from previous tests, i.e. a posterior form of SPRT. These tests could also be interpreted as simplifications of the 'exact test'.

In the following section we apply the results to the detection of faults in a gyro navigational system.

5. Application to Fault Detection in a Gyro Navigational System

We will show how the tests of the previous sections may be used to detect increases in the drift rate in a gyro system caused by incipient gyro failure. An extremely simple model of the gyro system will be used although this restriction is only imposed for the sake of clarity.

Suppose that the gyro system is installed as a navigational aid in an aircraft. There will be three independent gyros aligned along orthogonal axes; we shall restrict our considerations to one axis only. The simplest gyro drift model, discussed for example by Dushman [5], consists merely of a low-pass filter fed by white Gaussian noise. The output used for navigational calculations is the sum of the angular drift rate $\dot{e}(t)$, and the angular rate of change $\Omega(t)$ impressed on the gyro by the motion of the aircraft. The dynamic equations for this model are given by:

$$\ddot{e}(t) = F\dot{e}(t) + \xi(t) , \quad (31)$$

$$y(t) = \dot{e}(t) + \Omega(t) , \quad (32)$$

where F is a constant, and ξ is Gaussian white noise. The gyro will be considered defective if the variance of ξ exceeds a certain value. Suppose that the flight path of the aircraft is determined before takeoff: this implies that there is a predetermined 'nominal' angular rate of change $\bar{\Omega}(t)$ for the flight. If the pilot tries to fly on his predetermined flight path as closely as possible, let us suppose (rather naïvely) that the error in $\Omega(t)$ is a white Gaussian random variable:

$$\Delta\Omega(t) = \Omega(t) - \bar{\Omega}(t) ; p\{\Delta\Omega(t)\} = \mathcal{N}\{0, \sigma_{\Omega}^2\} . \quad (33)$$

We use as the measurements for the fault detection scheme samples of the

difference

$$z(t) = y(t) - \bar{\Omega}(t) = \dot{e}(t) + \Delta\Omega(t) \quad . \quad (34)$$

Although apparently now we can go straightforwardly through with deriving the equations for the fault detector, if we examine the relative magnitudes of the parameters of the gyro model we find ourselves in trouble. As typical figures we give:

Correlation time of \dot{e}	0.5 hr.	
Mean square value of \dot{e}	0.01°/hr.	(no fault)
	$\geq 0.05^\circ/\text{hr.}$	(fault)
Correlation time of $\Delta\Omega$	5 secs.	
Mean square value of $\Delta\Omega$	2°/sec.	

We see that although we can indeed treat $\Delta\Omega$ as white because it has a relatively very small correlation time, its mean square value is 6 orders of magnitude higher than that of \dot{e} . Hence it will be very difficult to detect any change at all in the variance of $\xi(t)$.

Fortunately we can take advantage of a current trend of thought which advocates a triplicated navigational system to attain higher reliability. If we assume such a system, we can say that to a large extent $\Delta\Omega$ is the same for each subsystem of the triple. We can thus remove its effects by forming the differences between the outputs of the three systems. Figure 3. shows how the three systems are interconnected.

If we assume that the values of F for each system are negligibly different, then

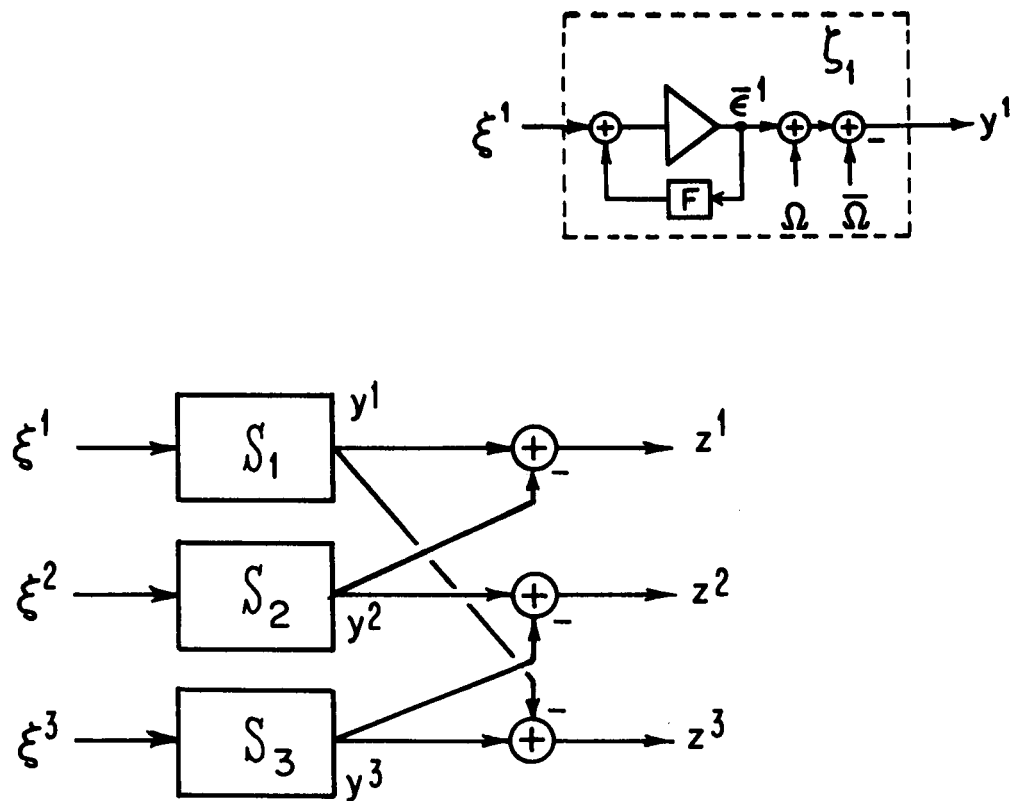


FIG. 3 INTERCONNECTIONS OF THREE GYRO SYSTEMS FOR FAULT DETECTION. (ONE AXIS ONLY)

$$\begin{aligned} \dot{x}^1(t) &= Fx^1(t) + w^1(t) \\ z^1(t) &= x^1(t) + v^1(t) \end{aligned} \quad ; \quad i = 1,2,3 \quad , \quad (35)$$

where

$$\begin{aligned} w^1 &= \xi^1 - \xi^2 \quad , \quad w^2 = \dots \text{ etc. } \quad , \\ x^1 &= \dot{e}^1 - \dot{e}^2 \quad , \quad x^2 = \dots \text{ etc. } \quad ; \end{aligned}$$

and the superscripts index the subsystems. Hopefully the differences

$$v^1 = \Delta\Omega^1 - \Delta\Omega^2 \quad ; \quad v^2 = \dots \text{ etc. } \quad ,$$

will now have a variance which is of a comparable order of magnitude to that of w . In the absence of available figures we have assumed that this is so, and treat the variance as a parameter in the ensuing tests.

Furthermore, v^1 is much more likely to be a white noise since all long term effects in $\Delta\Omega$ will be cancelled out.

By sampling the output we may use the discretized version of Eqn. (35)

$$\begin{aligned} x^1(k+1) &= \phi x^1(k) + w^1(k) \\ z^1(k) &= x^1(k) + v^1(k) \end{aligned} \quad ; \quad i = 1,2,3 \quad . \quad (36)$$

From now on we shall consider only one of the three sets of equations (36), and drop the i -superscript. Let the variances of w and v be denoted by

$$\text{Var}\{w\} = \begin{cases} Q_0 & \text{(faultfree)} \\ Q_1 & \text{(faulty)} \end{cases} \quad ; \quad (37)$$

$$\text{Var}\{v\} = R \quad ,$$

where the system is normalized so that $Q_0 = 1$. A value of $\phi = 0.8$ was used, implying a sampling period of about $3\frac{1}{2}$ minutes. There is now no obstacle to applying any of the SPRT's previously described to detecting whether a change has occurred in Q .

Two of the tests have been evaluated by digital simulation for this system:

(i) the 'continuous test';

(ii) the 'exact test' with $p\{W_{10}\} = 0$, implying that the system is known not to be faulty before the start of the test.

The 'occasional test' was not separately evaluated because the results are very similar to those of the 'continuous test'.

Taking the 'continuous test' first, two average sample numbers were evaluated by a Monte-Carlo method:

(a) the average number of samples \bar{N}_B required at the start of the test to reach a threshold (threshold B with high probability);

(b) the average number of samples \bar{N}_A required to move from threshold B to threshold A at the advent of a fault.

Figure 4. shows a typical run during the computer simulation of the test. The full curves shown in Figures 5 through 8 show the variation in \bar{N}_A and \bar{N}_B due to changes in the error probabilities, ratio of faultfree to faulty input variances, and variance of observation noise. These curves were obtained by application of the results of the Appendix. The discrete points marked are the results of the Monte-Carlo simulations: reasonable agreement is shown. Generally an increase in sample number results from:

(1) a reduction of the error probabilities α and β ; or

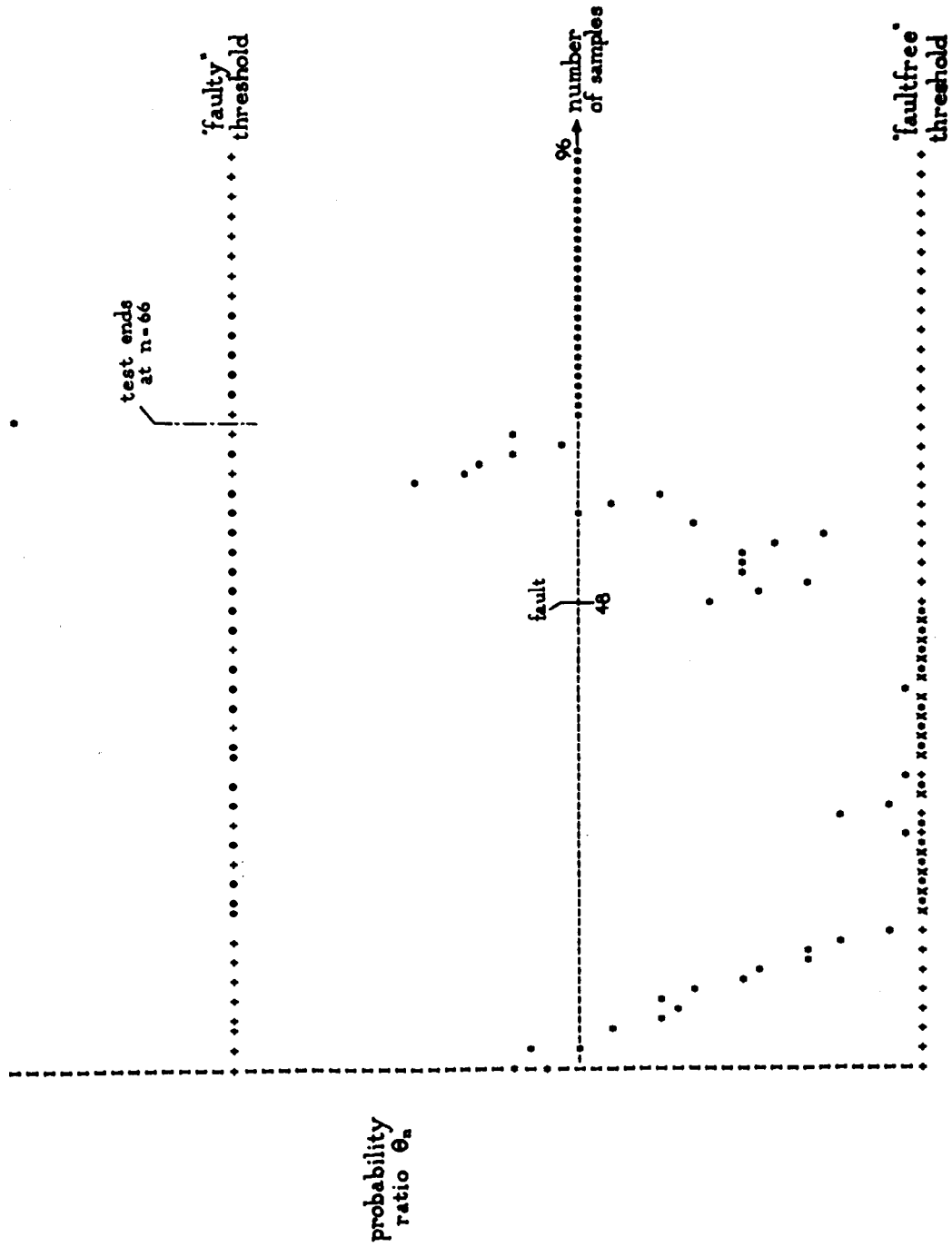


Figure 4. Typical run during computer simulation of "continuous test." Error bounds: $\alpha = \beta = 0.01$. Increase in input variance 4:1 ; $\phi = 0.8$; $R = 0.1$.

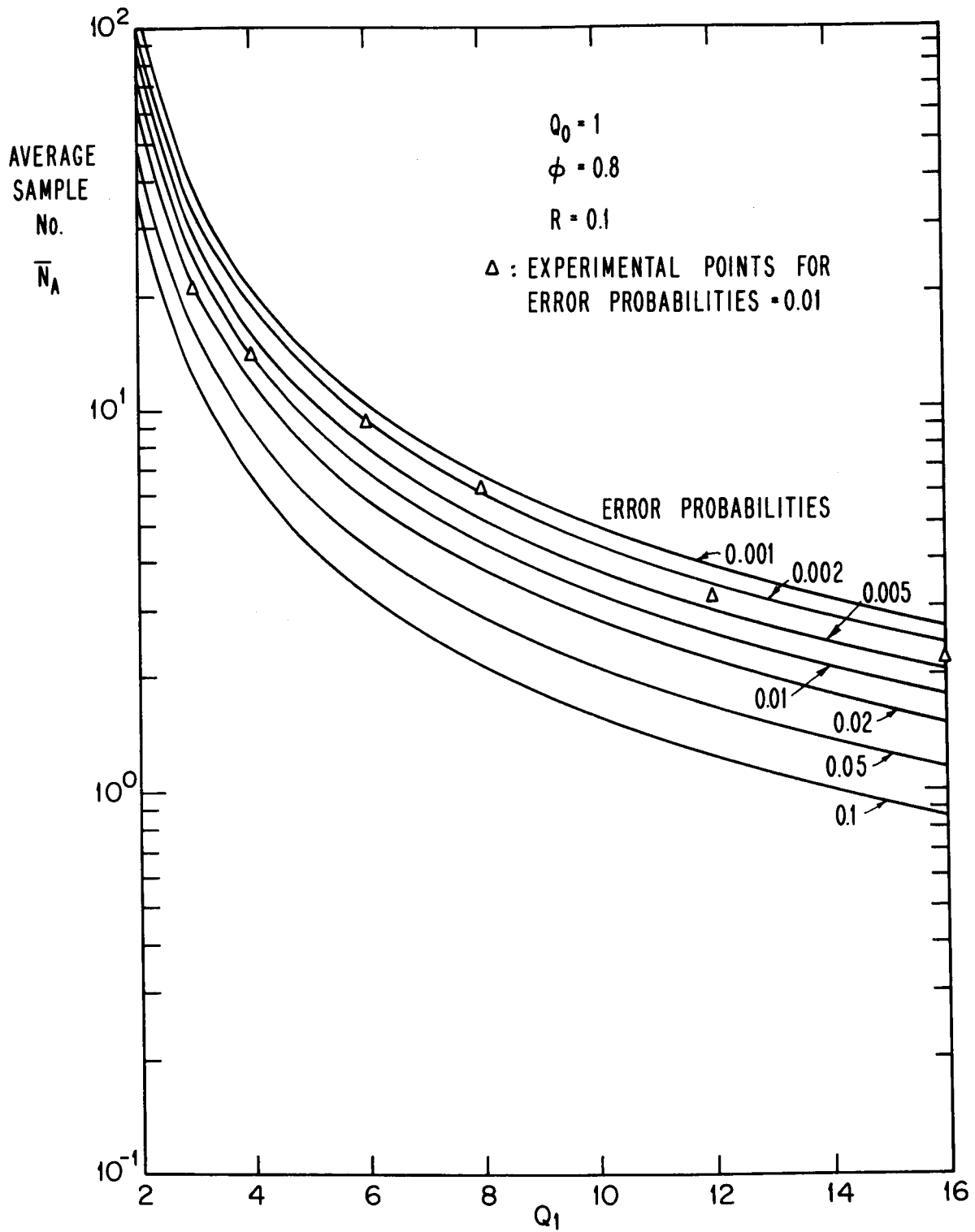


FIG. 5 AVERAGE SAMPLE NUMBER FROM LOWER TO UPPER THRESHOLD H_{10}

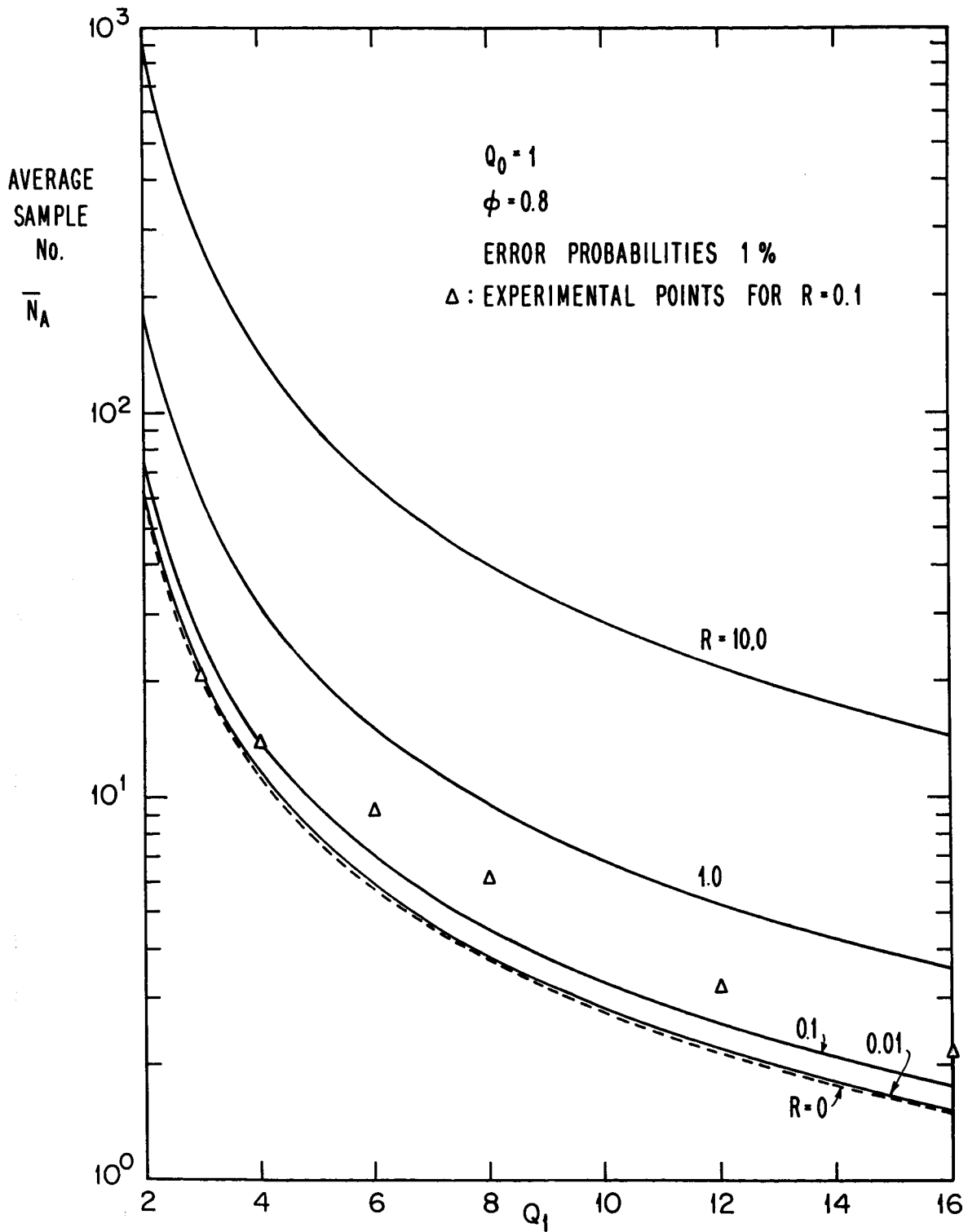


FIG. 6 AVERAGE SAMPLE NUMBER TO PASS FROM LOWER TO UPPER THRESHOLD : H_{10}

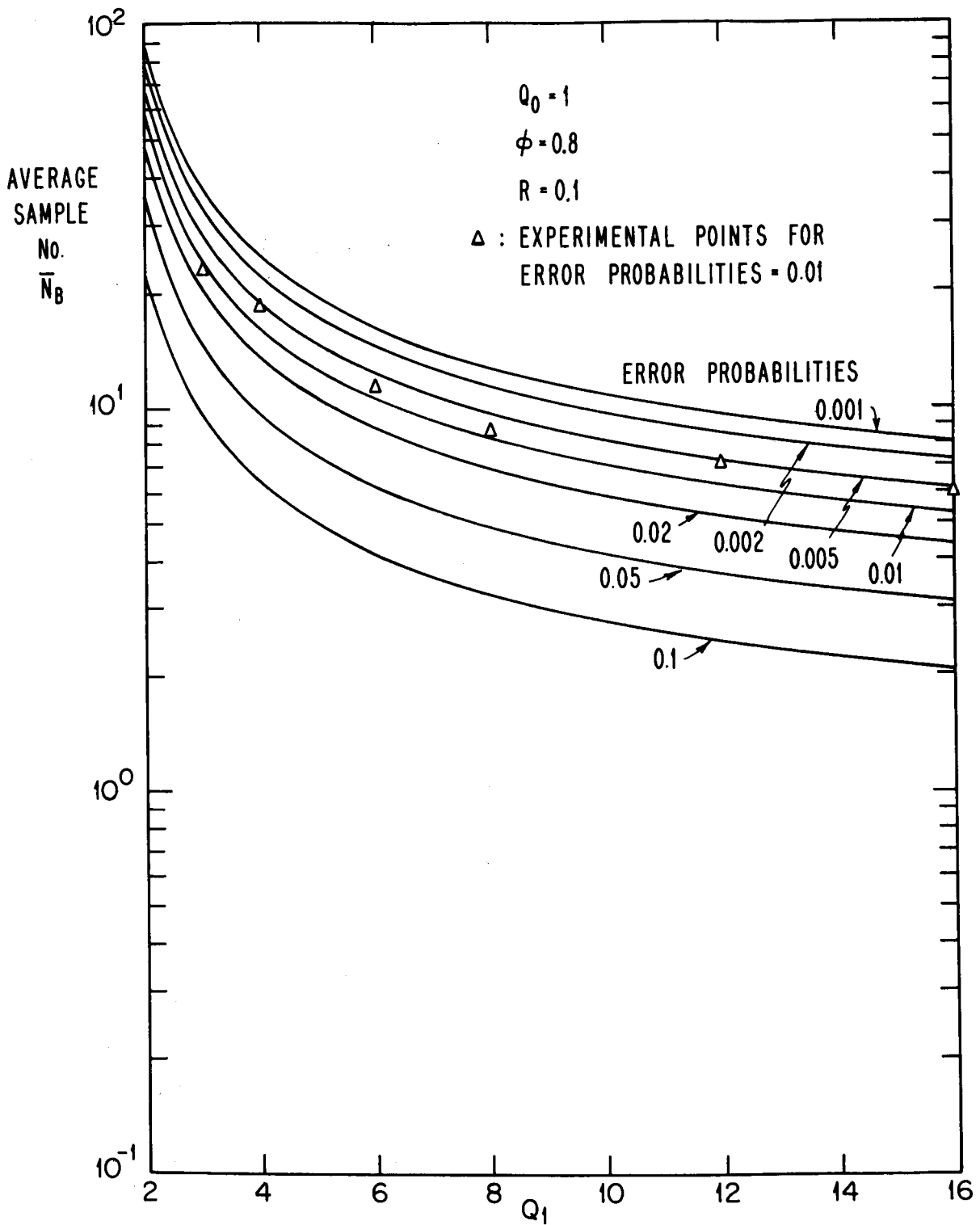


FIG. 7 AVERAGE SAMPLE NUMBER TO TERMINATION: H_0

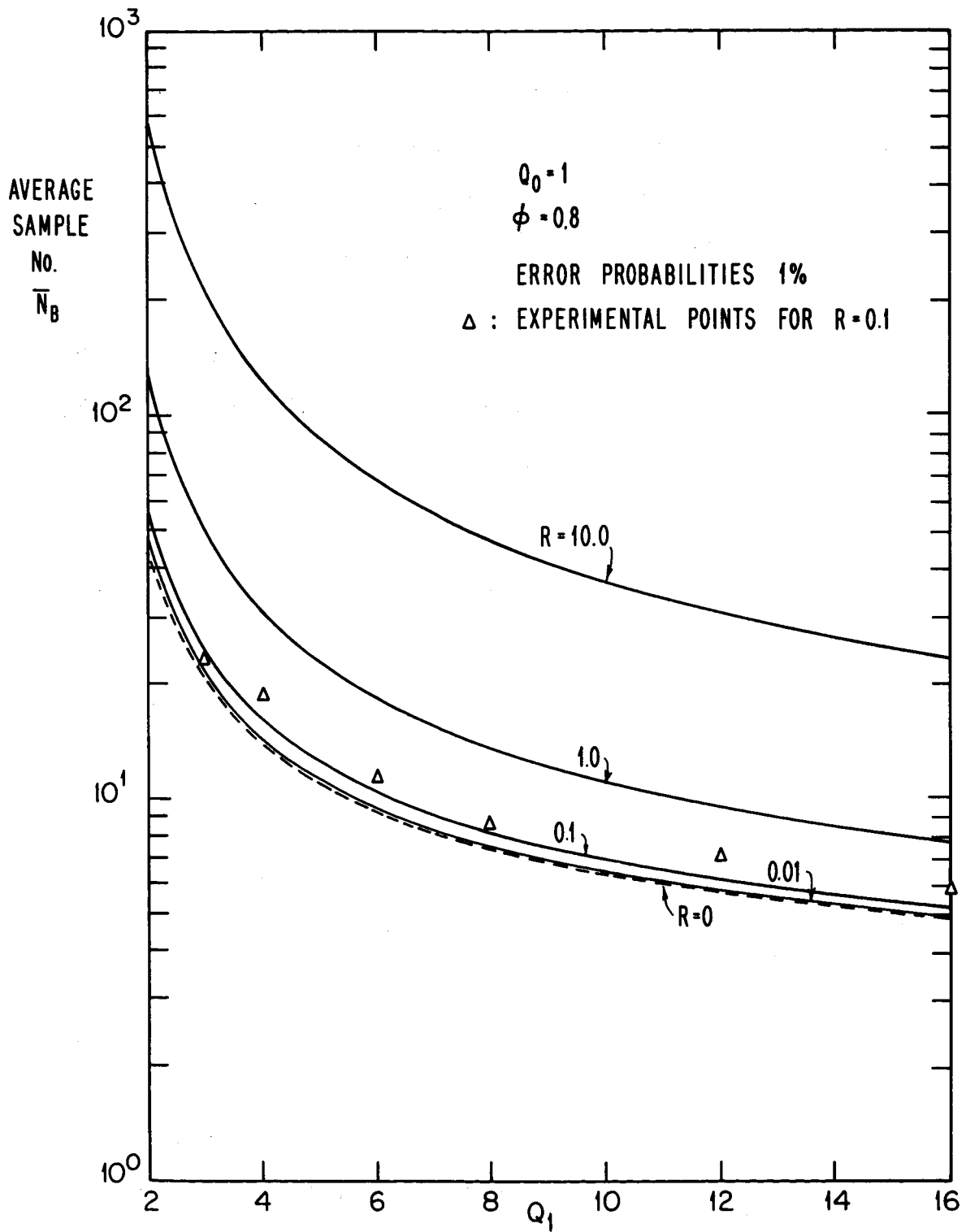


FIG. 8 AVERAGE SAMPLE NUMBER TO TERMINATION: H_0

- (ii) an increase in the variance of the observation noise R ; or
- (iii) a smaller increase in the input variance Q at the appearance of a fault.

The 'exact test' has been investigated less extensively. Figure 9. summarizes the results which have been obtained for the average sample number: it is evident that the improvement gained by the use of this test as opposed to the 'continuous test' is discernible but not very great. A curious result is that this test takes a very long time to reach the decision that no change has occurred. This can be explained by noting that in Eqn. (22), (ignoring the second term since $p\{\star_{10}\} = 0$), if \star_0 is true, on average the summands are less than unity and decrease as the number of multiplicands in the repeated product increases. The dominant

term is thus $\frac{1}{j} \cdot \frac{p\{\epsilon(n+j)/\star_{11}\}}{p\{\epsilon(n+j)/\star_0\}}$ which only goes down roughly as $\frac{1}{j}$,

where j is the number of steps since the start of the test.

Since the simplified tests consist mainly of the operations of addition and multiplication, it should soon become feasible to incorporate this type of fault detection feature as one of the duties of an on-board digital computer, especially as the computational algorithms involved may already embody Kalman filters to improve measurements from the gyro systems.

APPENDIX. Calculation of Average Sample Numbers for 'Continuous Test'

To determine the average sample numbers required for detection under each of the two hypotheses \star_{10} and \star_0 , we first need to know the average values of the increment in the probability ratio θ_k at each stage in the test. In other words, using the logarithmic form for convenience, we need to obtain expressions for

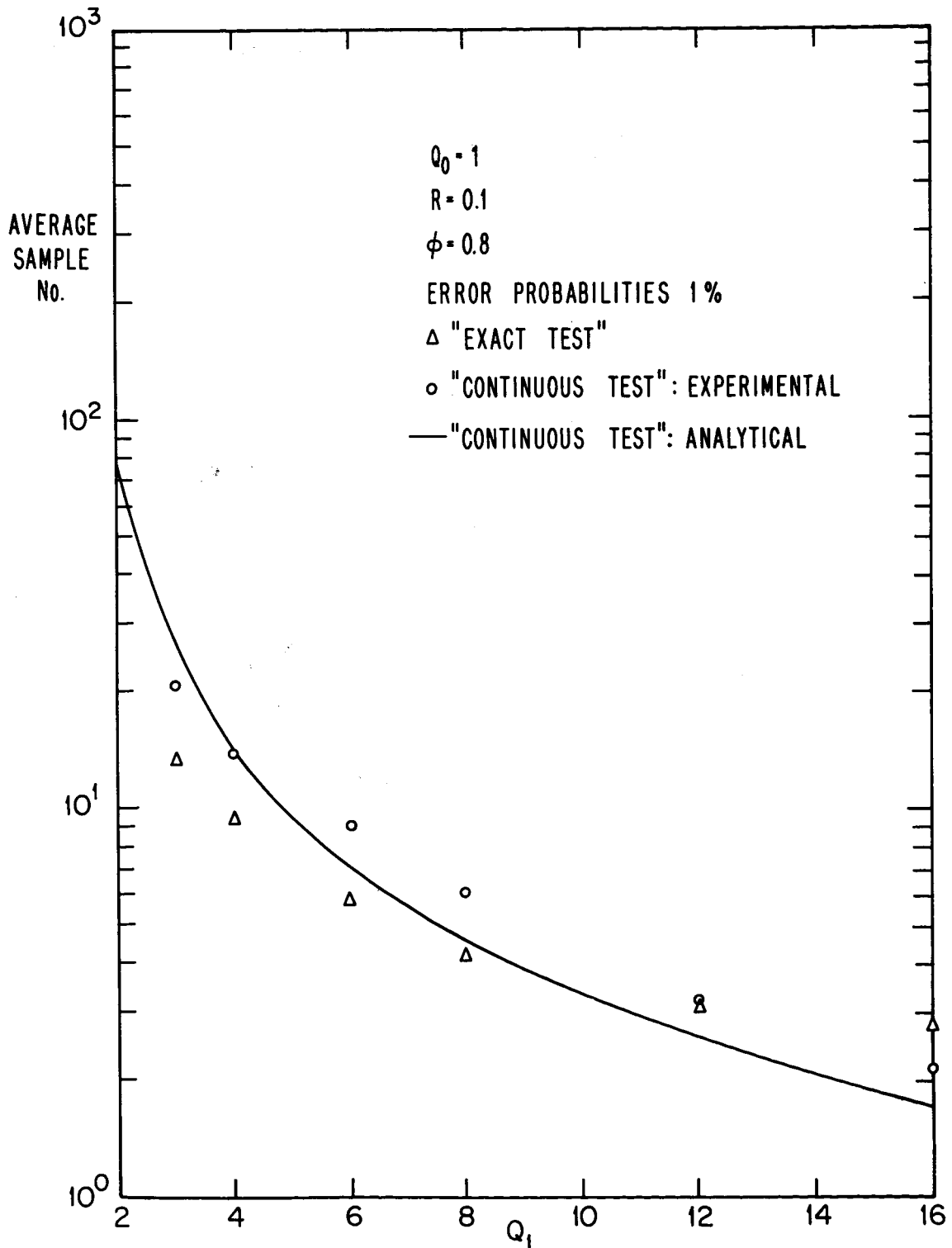


FIG. 9 AVERAGE SAMPLE NUMBER TO REACH THRESHOLD A AFTER FAULT

$$\mu_{10} = E_{\#_{10}} \left\{ \log \frac{\sigma_0}{\sigma_{10}} - \frac{1}{2} \left[\frac{\epsilon_{10}^2}{\sigma_{10}^2} - \frac{\epsilon_0^2}{\sigma_0^2} \right]_n \right\} ,$$

at time $k = n$. (38)

$$\mu_0 = E_{\#_0} \left\{ \log \frac{\sigma_0}{\sigma_{10}} - \frac{1}{2} \left[\frac{\epsilon_{10}^2}{\sigma_{10}^2} - \frac{\epsilon_0^2}{\sigma_0^2} \right]_n \right\}$$

We shall treat only the scalar case in the derivation, although it is possible to generalize.

Let the system and filter equations be given by

$$x(n+1) = \phi x(n) + w(n) ; \quad \text{Var}\{w(n)\} = q_0 \text{ or } q_1 \quad (39)$$

$$z(n) = x(n) + v(n) ; \quad \text{Var}\{v(n)\} = r \quad (40)$$

$$\hat{x}_i(n+1) = \psi_i \hat{x}_i(n) + k_i z(n+1) \quad (41)$$

$$\text{where } \psi_i = \phi \{1 - k_i\} \quad i = \begin{cases} 10 : \#_{10} \\ 0 : \#_0 \end{cases}$$

First we derive the expectations of three quantities.

$$(i) \ E\{z^2(n)\} .$$

From Eqns. (39) and (40)

$$x(n) = \phi^n x(0) + \sum_{i=0}^{n-1} \phi^i w(n-i-1) ,$$

$$E\{z^2(n)\} = E\{x^2(n)\} + r .$$

Ignoring the initial term $x(0)$

$$E\{x^2(n)\} = E\left\{ \sum_{i=0}^{n-1} \phi^i w(n-i-1) \right\}^2 = \sum_{i=0}^{n-1} \phi^{2i} q \quad ;$$

and in the limit:

$$\lim_{n \rightarrow \infty} E\{z^2(n)\} = \frac{q}{1 - \phi^2} + r \quad (42)$$

(ii) $E\{z(n) \cdot \hat{x}(n - 1)\}$.

$$z(n) = v(n) + \phi^n x(0) + \sum_{i=0}^{n-1} \phi^i w(n - i - 1) \quad .$$

and from Eqn. (41)

$$\hat{x}(n - 1) = \psi^{n-1} \hat{x}(0) + \sum_{i=0}^{n-2} \psi^i k z(n - i - 1) \quad .$$

Again ignoring the initial terms:

$$\begin{aligned} E\{z(n) \hat{x}(n - 1)\} &= E\left\{ \sum_{i=0}^{n-1} \phi^i w(n - i - 1) \cdot \sum_{i=0}^{n-2} \psi^i k z(n - i - 1) \right\} \\ &= E\left\{ \sum_{i=0}^{n-1} \phi^i w(n - i - 1) \cdot \sum_{i=0}^{n-2} \psi^i k \sum_{j=0}^{n-i-2} \phi^j w(n - i - 2 - j) \right\} \\ &= \phi k \sum_{i=0}^{n-2} \phi^{n-i-2} \sum_{j=0}^{n-i-2} \phi^j \psi^{n-i-2-j} \cdot q \quad ; \end{aligned}$$

and in the limit:

$$\lim_{n \rightarrow \infty} E\{z(n) \hat{x}(n - 1)\} = \frac{\phi k q}{1 - \phi^2} \cdot \frac{1}{1 - \phi \psi} \quad (43)$$

(iii) $E\{\hat{x}^2(n - 1)\}$.

$$E\{\hat{x}^2(n - 1)\} = E\left\{ \sum_{i=0}^{n-2} \psi^i k z(n - i - 1) \right\}^2$$

$$\begin{aligned}
E\{\hat{x}^2(n-1)\} &= k^2 E\left\{\sum_{i=0}^{n-2} \psi^i x(n-i-1)\right\}^2 + k^2 r \sum_{i=0}^{n-2} \psi^{2i} \\
&= k^2 E\left\{\sum_{i=0}^{n-2} \psi^i \sum_{j=0}^{n-i-2} \phi^j w(n-i-2-j)\right\}^2 + k^2 r \sum_{i=0}^{n-2} \psi^{2i} \\
&= qk^2 \sum_{i=0}^{n-2} \left\{\sum_{j=0}^{n-i-2} \psi^j \phi^{n-2-i-j}\right\}^2 + k^2 r \sum_{i=0}^{n-2} \psi^{2i} \quad ;
\end{aligned}$$

and in the limit

$$\lim_{n \rightarrow \infty} E\{\hat{x}^2(n-1)\} = \frac{k^2 q \{1 + \phi\psi\}}{(1 - \phi^2)(1 - \psi^2)(1 - \phi\psi)} + \frac{k^2 r}{1 - \psi^2} \quad . \quad (44)$$

We may write Eqns. (38) as

$$\mu_{10} = \log \frac{\sigma_0}{\sigma_{10}} - \frac{1}{2} \left[1 - \frac{1}{\sigma_0^2} E_{w_{10}} \{z(n) - \phi \hat{x}_0(n-1)\}^2 \right] \quad , \quad (45)$$

$$\mu_0 = \log \frac{\sigma_0}{\sigma_{10}} - \frac{1}{2} \left[\frac{1}{\sigma_{10}^2} E_{w_0} \{z(n) - \phi \hat{x}_{10}(n-1)\}^2 - 1 \right] \quad ;$$

and substituting in Eqns. (42) through (44) to obtain the asymptotic values:

$$\mu_{10} = \log \frac{\sigma_0}{\sigma_{10}} - \frac{1}{2} \{1 - v_{10}\} \quad , \quad (46)$$

$$\mu_0 = \log \frac{\sigma_0}{\sigma_{10}} - \frac{1}{2} \{v_0 - 1\} \quad ;$$

where

$$v_{10} = \frac{q_1}{1 - \phi^2} + r - \frac{2\phi^2 k_0 q_1}{(1 - \phi^2)(1 - \phi\psi_0)} + \phi^2 \left[\frac{k_0^2 q_1 (1 + \phi\psi_0)}{(1 - \phi^2)(1 - \psi_0^2)(1 - \phi\psi_0)} + \frac{k_0^2 r}{1 - \psi_0^2} \right],$$

and

$$v_0 = \frac{q_0}{1 - \phi^2} + r - \frac{2\phi^2 k_{10} q_0}{(1 - \phi^2)(1 - \phi\psi_{10})} + \phi^2 \left[\frac{k_{10}^2 q_0 (1 + \phi\psi_{10})}{(1 - \phi^2)(1 - \psi_{10}^2)(1 - \phi\psi_{10})} + \frac{k_{10}^2 r}{1 - \psi_{10}^2} \right].$$

We can now find the average sample numbers:

\bar{N}_B : the average number of samples required to reach any threshold from the start of the test, assuming μ_0 is true;

\bar{N}_A : the average number of samples to move from threshold B to threshold A when μ_{10} becomes true.

These sample numbers are given by

$$\bar{N}_B = [\alpha \log A + (1 - \alpha) \log B] / \mu_0 ; \quad (47)$$

$$\bar{N}_A = [\log A - \log B] / \mu_{10} , \quad (48)$$

where A and B are defined as in Sections 3 and 4. It should be noted that the expression for \bar{N}_A does not take any account of the fact that

θ_k may stay on the threshold B for a few samples before moving to threshold A .

The results of this Appendix are also applicable to the 'occasional test'.

References

1. Wald, A., "Sequential Analysis," John Wiley & Sons, Inc., New York, 1947.
2. Hancock, J. C., and P. A. Wintz, "Signal Detection Theory," Chaps. III and IV, McGraw Hill Inc., New York, 1966.
3. Bussgang, J. J., and D. Middleton, 'Sequential Detection of Signals in Noise,' Harvard Cruft Lab. Tech. Report No. 175, August 1955.
4. Schweppe, F. C., 'Evaluation of Likelihood Functions for Gaussian Signals,' Trans. IEEE, Vol. IT-11, No. 1, pp. 61-70, January 1965.
5. Dushman, A., 'On Gyro Drift Models and Their Evaluation,' Trans. IRE, Vol. ANE-9, No. 4., pp. 230-234, December 1962.

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)		2a. REPORT SECURITY CLASSIFICATION	
Division of Engineering and Applied Physics Harvard University Cambridge, Massachusetts		Unclassified	
		2b. GROUP	
3. REPORT TITLE			
DETECTION OF CHANGES IN THE CHARACTERISTICS OF A GAUSS-MARKOV PROCESS			
4. DESCRIPTIVE NOTES (Type of report and, inclusive dates)			
Interim technical report			
5. AUTHOR(S) (First name, middle initial, last name)			
P. M. Newbold and Y. C. Ho			
6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS	
June 1967	41	5	
8a. CONTRACT OR GRANT NO.		9a. ORIGINATOR'S REPORT NUMBER(S)	
Nonr-1866(16) & NASA Grant NGR-22-007-068		Technical Report No. 531	
b. PROJECT NO.		9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	
c.			
d.			
10. DISTRIBUTION STATEMENT			
Reproduction in whole or in part is permitted by the U. S. Government. Distribution of this document is unlimited.			
11. SUPPLEMENTARY NOTES		12. SPONSORING MILITARY ACTIVITY	
		Office of Naval Research	
13. ABSTRACT			
<p>By an extension to the theory of sequential detection with dependent measurements, it is possible to develop a Sequential Probability Ratio Test (SPRT) to detect <u>changes</u> in regime in a Gauss-Markov Process rather than detecting which of two regimes exist. It is shown how a posterior form of this extended SPRT may be simplified to reduce computational complexity. The simplified SPRT's are in fact modifications of the original SPRT detecting the regime and not the change. The tests are applied to the problem of fault detection in a gyro navigational system: the results of a detailed computer simulation are given.</p>			

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
sequential detection probability ratio test gyro fault detection						