Sequential Probability Ratio Test for Long-Term Radiation Monitoring

Kenneth D. Jarman, L. Eric Smith, and Deborah K. Carlson

Abstract – Among the possible decision-making algorithms for sequentially-acquired radiation sensor data is the Sequential Probability Ratio Test (SPRT). The suitability of the SPRT for long-term monitoring applications is discussed, and the decision-making performance of the SPRT is compared to that of the commonly used single-interval test (SIT). The analysis spans a wide range of signal and background count rates so that results are applicable to sensors of all sizes operating in different ambient conditions, with a spectrum of alarm thresholds. It is demonstrated that, for these simulated long-term monitoring scenarios, decisions to issue an alarm when the measured count rate equals the threshold count rate are made 3 to 5 times faster using the SPRT than with the SIT. The ability of the SPRT to provide an “all-clear” indication and the need for SPRT truncation strategies to limit decision times when the measured count rate falls between background and the specified threshold are also discussed. Under an early termination scenario, it is shown that a truncated SPRT retains a higher probability of detection.

I. INTRODUCTION

A PROTOTYPE instrumented container system for detecting illicit nuclear and radiological materials concealed in containerized cargo is being developed by Pacific Northwest National Laboratory (PNNL), Richland, WA. At the core of the prototype system are low-cost, low-power radiation sensors mounted on the walls of intermodal shipping containers. Decision-making algorithms that process the raw data from radiation sensors during oceanic transit are an integral part of the project.

These decision-making algorithms should issue an alarm when a certain threshold is exceeded and should provide an “all-clear” indication when the radiation signature of the container matches the expected background signature. Sensitivity is measured in terms of the estimated time to decision, that is, the measurement time needed to decide whether sensor data are consistent with a specified threshold or the expected background, while meeting the statistical confidence requirements of the measurement scenario (i.e., the false alarm probability and the detection probability).

Setting decision thresholds in long-term monitoring scenarios, such as instrumented containers, can be formulated in terms of a statistical hypothesis test on accumulated sensor counts. A standard approach is to accumulate counts over a fixed period of time and compare the resulting total count to a single critical level (the single-interval test or SIT).

However, sensor counts in long-term monitoring are often transmitted at set intervals (e.g., every few hours during a five-day oceanic transit), suggesting the application of sequential statistical analysis methods with the goal of timely decision making. The Sequential Probability Ratio Test (SPRT) gives a straightforward method for setting alarm and background thresholds against which counts can be sequentially compared until the test is conclusive. The SPRT is known to minimize the average sample size (i.e., average decision times) among all sequential tests with given error probabilities under rather general conditions [1], [2]. The properties of the SPRT make it appropriate for many radiation detection problems in which alarm decisions must be made. For example, the SPRT has been shown to improve decision times for the relatively short-duration portal monitoring of vehicles [3], [4] and personnel and packages [5], [6]. Properties of the SPRT have also been exploited for detection of “off-normal” operation in nuclear reactor surveillance [7], [8] and many other sensor-based applications.

II. METHODS

The decision making that accompanies the long-term monitoring of radiation fields can be generalized into three primary scenarios: 1) measured count rate is consistent with, or less than, the nominal background count rate resulting in an “all-clear” decision, 2) measured count rate is consistent with, or greater than, a specified threshold count rate resulting in “alarm” decision, and 3) measured count rate is greater than the nominal background count rate, but less than the threshold alarm rate, so that decisions cannot be made at the desired level of statistical confidence in the allowable measurement period.

Each of the three scenarios can be couched in statistical terms as well, given an observed sensor total count rate with a
known probability distribution but unknown parameter \( \lambda \), a nominal average background count rate \( \lambda_0 \), and nominal average signal plus background count rate \( \lambda_s \). The first scenario is the null hypothesis \( H_0 \) that the observed count is due to background only and is tested against the second scenario, or alternative hypothesis \( H_1 \), that the count includes threshold signal. More precisely, it is assumed that counts in each case have the same distribution but different mean count rates. Letting the random variable \( X \) (or random process \( X_t \) for sequential analysis) represent a total sensor count with mean count rate \( \lambda \), the hypotheses for the first two scenarios are \( H_0 : \lambda = \lambda_0 \); \( H_1 : \lambda = \lambda_s \). Imposing maximum probabilities of false positive and false negative decisions, one derives a threshold on \( X \) (or thresholds on log likelihood ratios related to \( X_t \)) for deciding whether to reject the null hypothesis. In the third scenario, where \( \lambda_0 < \lambda < \lambda_1 \), a decision to reject or fail to reject \( H_1 \) can still be made, but at reduced statistical confidence levels. In these cases, the test is terminated and a probability of detection is calculated based on the acquired data to that point.

Most of the analysis presented in this paper focuses on the first two idealized scenarios wherein an “all-clear” or an “alarm” is issued, but considerable discussion is devoted to truncation strategies for the third scenario. Before proceeding, a few definitions are in order:

- \( \lambda_0 \) = nominal mean background count rate, in counts per second (cps)
- \( \lambda_1 \) = threshold alarm count rate, typically pre-specified based on operational constraints for false alarms and detection probability, for a typical \( \lambda_0 \)
- \( \tau = (\lambda_1 - \lambda_0) / \lambda_0 \), the ratio of the mean net signal count rate to mean background count rate
- \( \lambda \) = true mean count rate of observed process (unknown prior to observation)
- \( \alpha_0 \) = target false alarm probability (probability of raising an alarm when \( \lambda = \lambda_0 \))
- \( \beta_0 \) = target false negative probability (probability of failing to raise alarm when \( \lambda = \lambda_1 \))
- \( \alpha \) = true false alarm probability, commonly referred to as FAP
- \( \beta \) = true false negative probability, where the probability of detection, often referred to as \( P_D \), is \( 1 - \beta \)
- \( \text{TTD} = \text{time to detection} = \text{observation period for SIT} \) and \( E[\lambda | H_1] \), the average total observation time when \( \lambda = \lambda_1 \), for SPRT
- \( \text{TTC} = \text{time to clear} = E[\lambda | H_0] \), the average total observation time when \( \lambda = \lambda_0 \), for SPRT

The probabilities \( \alpha_0, \beta_0, \alpha, \) and \( \beta \) are also commonly referred to as error rates. Note that \( \lambda_1 \) is the minimum mean count rate at which a target detection probability of at least \( 1 - \beta \) is to be achieved (\( P_D = 1 - \beta \)). For true mean count rate \( \lambda < \lambda_1 \), actual detection probability is less than \( P_0 \). For the instrumented container monitoring application, very low error rates are appropriate, and \( \alpha_0 - \beta_0 - 10^{-4} \) is specified in the results presented here.

The analysis also makes several assumptions important to appreciate when translating the simulated results of this paper into actual application. First, systematic uncertainties such as detector drift and temperature dependencies are not assessed. Only statistical uncertainty is considered in this analysis, and perfect knowledge of the mean background count rate is assumed. Second, instrument malfunctions and methods to detect those malfunctions from the data stream are not considered. Finally, for the SPRT, observation intervals are allowed to be arbitrarily small. In practice, the observation intervals of both SPRT and SIT will, of course, be constrained to a practical minimum.

### A. Single Interval Test

For a fixed sample or “single interval test” (SIT), counts are collected over a single time \( T \) and the total number of counts is compared to a single critical level to decide whether to raise an alarm. Given a probability distribution for counts, one inverts the cumulative distribution function to obtain a critical count level, \( L_C \), and \( T \).

Intense sources will lead to short decision times and low total counts, for which the Poisson distribution is appropriate. Weaker sources will lead to longer decision times and higher total counts, in which case the Gaussian distribution can be a good approximation to the Poisson distribution. In this latter case a simple analytical result is obtained that illuminates this process. Let background counts have mean \( N_0 = \lambda_0 T \), and let background plus signal counts have mean \( N_1 = \lambda_1 T \). The critical level is given by \( L_C = N_0 + A_{\sigma_0} \sigma_0 \), where \( \sigma_0 = \sqrt{N_0} \) and \( A_{\sigma_0} \) is a constant that depends only on \( \alpha_0 \) (e.g., for \( \alpha_0 = 0.05 \), \( A_{\sigma_0} \approx 1.645 \)). The total count at which \( P_D \) is achieved is given by \( N_1 = L_C + A_{\sigma_1} \sigma_1 \), where \( \sigma_1 = \sqrt{N_1} \) and \( A_{\sigma_1} \) is a constant that depends on \( \beta_0 \). Combining equations above, one can solve for \( T = \text{TTD} \):

\[
\text{TTD} = \left( \frac{A_{\sigma_0} \sqrt{N_0} + A_{\sigma_1} \sqrt{T}}{\lambda_1 - \lambda_0} \right)^2.
\]

In this case TTD is the observation time required by the test for any value of \( \lambda \). However, as noted previously, for \( \lambda < \lambda_1 \), the true detection probability will be less than \( 1 - \beta \).
B. Sequential Probability Ratio Test

The SPRT is an application of hypothesis testing in which a random process is observed sequentially, and at each observation a log likelihood ratio is compared to two thresholds that are determined by the prescribed error rates. If a threshold is reached or crossed, a decision regarding the null or alternative hypotheses is reached (alarm or “all-clear”) [1,2]. In this framework the decision time is a random variable and depends on the true count rate $\lambda$. A crucial difference between this sequential approach and the SIT is that the latter is inconclusive when an alarm is not raised, whereas with the SPRT one can continue observing until the test concludes either $H_0$ or $H_1$. Relationships for the average time to make a decision under $H_0$ and $H_1$ using the SPRT algorithm are derived below.

Within each time interval of length $\Delta T$ the sensor count is modeled as a Poisson random variable, where the average background and background plus signal counts per time interval are given by $\lambda_0 \Delta T$ and $\lambda_1 \Delta T$, respectively. Let $X_n$ be the total sensor count at step $n$ with unknown count rate $\lambda$, and let $x_i$ represent the sensor count within the $i$th interval, so that $X_n = \sum_{i=1}^{n} x_i$. Let $Z_i$ be the log likelihood ratio

$$Z_i = \log \frac{f(x_i)}{f_0(x_i)} = \log \left( \frac{\lambda_i}{\lambda_0} \right) x_i - (\lambda_i - \lambda_0) \Delta T \quad (2)$$

where $f_j$ is the probability density of $x_i$ under $H_j$, $j = 0,1$. Let $a$ and $b$ be given (as yet unknown) thresholds with $b < 0 < a$. At each observation $n = 1,2,\ldots$ the log likelihood ratio for total counts, $I_n = \sum_{i=1}^{n} Z_i$, is compared to $a$ and $b$. At step $n$, if $I_n \leq b$ then the test is terminated and $H_0$ is decided (background only). If $I_n \geq a$ then the test is terminated and $H_1$ is decided (signal present). Otherwise, the test continues. Let $N$ be the number of observations needed to reach a decision. $T = N \Delta T$ is referred to as the stopping time. Assuming negligible overshoot of the thresholds $a$ and $b$ by $I_n$, approximate formulas for several properties of the SPRT were obtained by Wald [1,2]. In particular, the thresholds are given by

$$a = \log \frac{1-\beta_0}{\alpha_0} \quad \text{and} \quad b = \log \frac{\beta_0}{1-\alpha_0} \quad (3)$$

and it is of note that $a$ and $b$ are independent of the distribution of the $x_i$. For these thresholds, the actual error rates $\alpha$ and $\beta$ are not identical to the desired rates $\alpha_0$ and $\beta_0$; however, it is easily shown that $\alpha + \beta \leq \alpha_0 + \beta_0$, $\alpha \leq \alpha_0/(1-\beta_0)$, and $\beta \leq \beta_0/(1-\alpha_0)$. Thus, for the very low error rates required here, the actual error rates cannot be much greater than, and as it turns out are often smaller than, the desired rates.

The average stopping time under $H_1$, or time to detection, may be approximated by

$$\text{TTD} = E[N \mid H_1] \Delta T \approx \frac{(1-\beta_0)a + \beta_0 b}{\log(\lambda_1/\lambda_0)\lambda_1 - (\lambda_1 - \lambda_0)} \quad (4)$$

This is the time to detection value that is compared to that obtained from the SIT method (1).

Similarly, the average stopping time under $H_0$, or time to clear, may be approximated by

$$\text{TTC} = E[N \mid H_0] \Delta T \approx \frac{\alpha_0 a + (1-\alpha_0) b}{\log(\lambda_1/\lambda_0)\lambda_0 - (\lambda_1 - \lambda_0)} \quad (5)$$

Note that the overshoot of the thresholds $a$ and $b$ cannot be ignored for Poisson processes. However, by comparison to exact results for sequential analysis on continuous-time Poisson processes (see [9,10]) we found that, in this case, the neglect of overshoot does not lead to significant approximation errors.

III. Results

A. Average Time to Decision: SPRT versus SIT

Results for $\tau$ in the range $10^{-3}$ to $10^3$ and, initially, for a single value of $\lambda_0 = 10$ cps are considered. The resulting analytical estimates for average decision times using the SIT (1) and the SPRT (4 and 5) are compared in Fig. 1. In this figure Gaussian counts are assumed for background mean counts greater than 80 (corresponding to $\tau < 0.1$), otherwise Poisson counts are assumed.
Fig. 1. Comparison of average TTD and TTC using SPRT to TTD using SIT for $\lambda_0 = 10$ cps. Target error rates are $\alpha_0 = \beta_0 = 10^{-4}$.

From Fig. 1 and Table I, it is evident that SPRT TTD is estimated to be 3 to 5 times faster than the SIT TTD, which would be extremely important if SIT detection times are longer than allowable measurement times (e.g., longer than oceanic transit).

**Table I.** Average simulated time to decision for SPRT and SIT methods.

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>Wald</th>
<th>Sim.</th>
<th>% of SIT TTD</th>
<th>% &lt; SIT TTD</th>
<th>Wald</th>
<th>Sim.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001</td>
<td>1.84E+06</td>
<td>1.98E+06</td>
<td>36</td>
<td>99.6</td>
<td>1.84E+06</td>
<td>1.98E+06</td>
</tr>
<tr>
<td>0.01</td>
<td>1.85E+04</td>
<td>1.98E+04</td>
<td>36</td>
<td>99.6</td>
<td>1.85E+04</td>
<td>1.98E+04</td>
</tr>
<tr>
<td>0.1</td>
<td>1.90E+02</td>
<td>2.05E+02</td>
<td>35</td>
<td>99.6</td>
<td>1.96E+02</td>
<td>2.11E+02</td>
</tr>
<tr>
<td>1</td>
<td>2.38E+00</td>
<td>2.60E+00</td>
<td>31</td>
<td>99.8</td>
<td>3.00E+00</td>
<td>3.15E+00</td>
</tr>
<tr>
<td>10</td>
<td>5.62E-02</td>
<td>6.43E-02</td>
<td>25</td>
<td>99.9</td>
<td>1.21E-01</td>
<td>1.23E-01</td>
</tr>
<tr>
<td>100</td>
<td>2.52E-03</td>
<td>3.22E-03</td>
<td>20</td>
<td>99.9</td>
<td>9.65E-03</td>
<td>9.75E-03</td>
</tr>
<tr>
<td>1000</td>
<td>1.56E-04</td>
<td>2.09E-04</td>
<td>18</td>
<td>99.9</td>
<td>9.27E-04</td>
<td>9.32E-04</td>
</tr>
</tbody>
</table>

Additionally, Fig. 1 also illustrates the fact that the SIT method, as it is commonly implemented, has no ability to issue an “all-clear” when the measured count rate is consistent with the expected background count rate. If the measured count rate is less than the critical level $L_C$, the test result is inconclusive.

By contrast, the SPRT method does provide a rigorous estimate of TTC, a capability that may be a decided advantage in applications such as instrumented containers where clearing containers may be as important as identifying suspect ones.

The results of Fig. 1 and Table I are based on a single value of $\lambda_0$, but it is clear that TTD and TTC are functions of $\lambda_0$ and not simply a function of the ratio $\tau$. Average TTD and TTC for SPRT and SIT TTD can be rigorously shown to depend on $\lambda_0$ only through dependence on the ratio $\tau$ and a linear scaling of $1/\lambda_0$. The ratio of any two of these decision time estimates depends on $\lambda_0$ only through $\tau$. Therefore, the qualitative behavior of Fig. 1 remains the same for all simulated values of $\tau$ and $\lambda_0$, as shown in Fig. 2 where $\lambda_0$ is varied over five decades from 0.1 cps to 1000 cps to cover a range of sensor types and ambient background conditions. For example, background count rates in small neutron radiation sensors may be less than 1 cps while background rates in large plastic scintillator gamma-ray detectors often exceed 1000 cps. These findings are consistent with well-known results from the general SPRT theory and the previous applications to radiation detection, nuclear safeguards, and nuclear reactor surveillance cited earlier.

**B. Time to Decision Distributions: SPRT versus SIT**

The distribution of SPRT decision times is obtained by simulation, which is easily carried out using MATLAB [11]. This allows a more empirical estimate of average times to decision as well as quantification of the fraction of SPRT decisions taking less time than the SIT TTD, as shown in Table I. For the range of $\tau$ considered here, the simulated SPRT outperforms the SIT better than 99% of the time when $\lambda = \lambda_0$ or $\lambda = \lambda_1$. The average TTD and TTC estimated from simulations (100,000 simulated tests for each value of $\tau$) are very close to the analytical approximations. The reduction in decision time shown in this table is based on simulation mean rather than the Wald approximation.

**C. Truncated SPRT and Early Termination**

Although the SPRT clearly gives much faster results than the SIT for the idealized scenarios when $\lambda = \lambda_0$ or $\lambda = \lambda_1$, Table I shows that on rare occasions, the SPRT decision time can exceed the SIT time and be excessively long, due to the random nature in the SPRT. Additionally, a documented criticism of the SPRT is that decision times can be very long in the third scenario described earlier—when the measured count rate is between the expected background and some target alarm threshold ($\lambda_0 < \lambda < \lambda_1$) [1]. Both situations are likely to occur in realistic long-term monitoring and must be addressed via truncation strategies [6]. That is, in cases where the SPRT has not reached a decision by a given time, the decision is forced.

An example of how the SPRT truncation might be put to practice helps to shed light on the $\lambda_0 < \lambda < \lambda_1$ region where both the SPRT and SIT decision times could be intolerably long. For instance, the measurement time required to meet the statistical confidence criteria may be estimated at 100 days, yet the container ship voyage is only 10 days. In that case, both tests must be terminated “early,” and the question is which of the decision algorithms provides a higher-confidence decision.
When the SIT is not allowed to run to the full time required to meet the original confidence criteria, those criteria cannot be completely met. In particular, the probability of detection will decrease for all $\lambda$. For example, the prescribed probability of detection $1 - \beta_0$ at $\lambda_0$ cannot be achieved; the probability $1 - \beta_0$ is achieved now for some mean count rate threshold greater than $\lambda_0$. This threshold is calculated, given the new observation time, in a similar manner as TTD is calculated given $\lambda_0$ (Section 2.1).

On the other hand, when the SPRT cannot run to completion, the test must be truncated. A natural truncation strategy is to stop the sequential test at a specified step $N_0$ and accept $H_0$ when $I_{N_0} \leq 0$ and reject $H_0$ when $I_{N_0} > 0$ if the SPRT has not yet concluded. Similar to the early termination of the SIT, this has the effect of reducing the probability of detection.

The results for an example “early termination” are shown in Fig. 3. In this example, both tests are stopped at $\frac{1}{2}$ the pre-specified SIT TTD, and the probability of alarm is compared for $\lambda > \lambda_0$. The latter probability can be obtained for the SIT directly from the distribution of counts given the new observation time and original error rates. The same probability for SPRT was estimated using simulations, using the truncation strategy described above. Fig. 3 demonstrates that the SPRT retains a higher probability of detection than the SIT under these early termination conditions. Further, the SPRT never takes longer than the SIT, and on average takes far less time, to make a decision.

![Fig. 3](image)

**Fig. 3.** Probability of detection when truncating at observation time less than SIT TTD. Truncation time here is $\frac{1}{2}$ the SIT TTD, $\lambda_0 = 10$ cps, and target error rates are $\alpha_0 = \beta_0 = 10^{-4}$.

### IV. Summary

The periodic transmission of data from radiation sensors in long-term monitoring applications opens the door to decision-making algorithms that can leverage sequential data collection into shorter decision-making times when compared to decision times from the commonly used single interval tests. Analytical estimates of average time to alarm, when the measured count rate is consistent with a specified alarm threshold, are 3 to 5 times faster using the SPRT method when compared to the SIT method, for signal to background ratios ranging from $10^{-3}$ to $10^3$. This superiority of the SPRT holds over the wide range of background count rates and alarm thresholds for most conceivable long-term radiation monitoring scenarios. Importantly, the SPRT framework incorporates the ability to issue an “all-clear” decision when the measured count rate is consistent with the expected background count rate whereas SIT test is inconclusive under these conditions. A disadvantage of the SPRT method is that decision times for certain values of measured count rate between expected background and alarm threshold can be somewhat longer than the SIT values. Truncation strategies can be exercised to mitigate this problem, and to make decisions when measurements must be terminated before either the SPRT or SIT tests can meet the prescribed error rates. In these situations, the truncated SPRT methods retain a higher probability of detection, with decision times that are less than or equal to the SIT method.

### V. Acknowledgment

The authors would like to thank K.H. Jarman, K.C. Gross, and P. Huang for helpful suggestions, and D.N. Anderson for providing seminal ideas and discussion.

### VI. References
