In Classical Hypothesis testing volumes of data is to be collected and then the conclusions are drawn, which may need more time. But, Sequential Analysis of Statistical science could be adopted in order to decide upon the reliability or unreliability of the developed software very quickly. The procedure adopted for this is, Sequential Probability Ratio Test (SPRT). It is designed for continuous monitoring. The likelihood based SPRT proposed by Wald is very general and it can be used for many different probability distributions. In the present chapter we propose the performance of SPRT on 5 data sets of Time domain data as well as on 5 data sets of Interval domain data using Rayleigh and analyzed the results. The parameters are estimated using Maximum Likelihood Estimation method. The content of this chapter is published in two journals and the details are furnished below.


5.1 INTRODUCTION

Sequential Probability Ratio Test (SPRT), which is usually applied in situations, requires a decision between two simple hypothesis or a single decision point. Wald’s (1947) SPRT procedure has been used to classify the software under test into one of two categories (e.g., reliable/unreliable, pass/fail,
certified/noncertified) (Reckase, 1983). Wald's procedure is particularly relevant if the data is collected sequentially. Classical Hypothesis Testing is different from Sequential Analysis. In Classical Hypothesis testing, the number of cases tested or collected is fixed at the beginning of the experiment. In this method, the analysis is made and conclusions are drawn after collecting the complete data. However, in Sequential Analysis every case is analysed directly. The data collected up to that moment is then compared with threshold values, incorporating the new information taken from the freshly collected case. This approach makes one to draw conclusions during the data collection, and ultimate conclusion can be reached at a much earlier stage. Data collection can be terminated after few cases and decisions can be taken quickly. This leads to saving in terms of cost and human life.

In the analysis of software failure data, either TBFs or failure count in a given time interval is dealt with. If it is further assumed that the average number of recorded failures in a given time interval is directly proportional to the length of the interval and the random number of failure occurrences in the interval is explained by a Poisson process. Then it is known that the probability equation of the stochastic process representing the failure occurrences is given by a Homogeneous Poisson Process with the expression

\[ P[N(t) = n] = \frac{e^{-\lambda t} (\lambda t)^n}{n!} \]  

(5.1.1)

Stieber (1997) observes that, the application of SRGMs may be difficult and reliability predictions can be misleading, if classical testing strategies are used. However, he observes that statistical methods can be successfully applied to the failure data. He demonstrated his observation by applying the well-known sequential probability ratio test of Wald for a software failure data to detect unreliable software components and compare the reliability of different software versions. In this chapter the popular SRGM – a two parameter Weibull (i.e Rayleigh) model is considered and the principle of Stieber is adopted in detecting unreliable software in order to accept or reject the developed software. The theory proposed by Stieber is presented in Section 2 for a ready reference. Extension of this theory to the considered SRGM is presented in Section 3. Maximum Likelihood parameter estimation method is presented in Section 4. Application of the decision
rule to detect unreliable software with reference to the SRGM-Rayleigh is given in Section 5.

5.2 SEQUENTIAL TEST FOR A POISSON PROCESS

A.Wald, developed the SPRT at Columbia University in 1943. A big advantage of sequential tests is that they require fewer observations (time) on the average than fixed sample size tests. SPRTs are widely used for statistical quality control in manufacturing processes. The SPRT for Homogeneous Poisson Processes is described below.

Let \( \{N(t), t \geq 0\} \) be a homogeneous Poisson process with rate ‘\( \lambda \)’. In this case, \( N(t) \) = number of failures up to time ‘\( t \)’ and ‘\( \lambda \)’ is the failure rate (failures per unit time). If the system is put on test (for example a software system, where testing is done according to a usage profile and no faults are corrected) and that if we want to estimate its failure rate ‘\( \lambda \)’. We can not expect to estimate ‘\( \lambda \)’ precisely. But we want to reject the system with a high probability if the data suggest that the failure rate is larger than \( \lambda_1 \) and accept it with a high probability, if it is smaller than \( \lambda_0 \). As always with statistical tests, there is some risk to get the wrong answers. So we have to specify two (small) numbers ‘\( \alpha \)’ and ‘\( \beta \)’, where ‘\( \alpha \)’ is the probability of falsely rejecting the system. That is rejecting the system even if \( \lambda \leq \lambda_0 \). This is the "producer’s" risk. '\( \beta \)' is the probability of falsely accepting the system .That is accepting the system even if \( \lambda \leq \lambda_1 \). This is the “consumer’s” risk. Wald’s classical SPRT is very sensitive to the choice of relative risk required in the specification of the alternative hypothesis. With the classical SPRT, tests are performed continuously at every time point \( t > 0 \) as additional data are collected. With specified choices of \( \lambda_0 \) and \( \lambda_1 \) such that \( 0 < \lambda_0 < \lambda_1 \), the probability of finding \( N(t) \) failures in the time span \( (0,t) \) with \( \lambda_1 \), \( \lambda_0 \) as the failure rates are respectively given by

\[
P_1 = \frac{e^{-\lambda_1 t} \left[ \frac{\lambda_1}{\lambda_0} \right]^{N(t)}}{N(t)!}
\]

(5.2.1)
The ratio \( \frac{P_1}{P_0} \) at any time 't' is considered as a measure of deciding the truth towards \( \lambda_0 \) or \( \lambda_1 \), given a sequence of time instants \( t_1 < t_2 < t_3 < \ldots < t_K \) and the corresponding realizations \( N(t_1), N(t_2), \ldots, N(t_K) \) of \( N(t) \). Simplification of \( \frac{P_1}{P_0} \) gives

\[
\frac{P_1}{P_0} = \exp(\lambda_0 - \lambda_1) t + \left( \frac{\lambda_1}{\lambda_0} \right)^{N(t)}
\]

The decision rule of SPRT is to decide in favour of \( \lambda_1 \), in favour of \( \lambda_0 \) or to continue by observing the number of failures at a later time than 't' according as \( \frac{P_1}{P_0} \) is greater than or equal to a constant say A, less than or equal to a constant say B or in between the constants A and B. That is, we decide the given software product as unreliable, reliable or continue (Satyaprasad, 2007) the test process with one more observation in failure data, according to

\[
\frac{P_1}{P_0} \geq A \quad (5.2.3)
\]

\[
\frac{P_1}{P_0} \leq B \quad (5.2.4)
\]

\[
B < \frac{P_1}{P_0} < A \quad (5.2.5)
\]

The approximate values of the constants A and B are taken as

\[
A \approx \frac{1 - \beta}{\alpha}, \quad B \approx \frac{\beta}{1 - \alpha}
\]

Where \( \alpha \) and \( \beta \) are the risk probabilities as defined earlier. A good test is one that makes the \( \alpha \) and \( \beta \) errors as small as possible. The common procedure is to fix the \( \beta \) error and then choose a critical region to minimize the error or maximize the
power i.e $1 - \beta$ of the test. A simplified version of the above decision processes is to reject the system as unreliable if $N(t)$ falls for the first time above the line

$$N_u(t) = at + b_2$$

To accept the system to be reliable if $N(t)$ falls for the first time below the line

$$N_l(t) = at - b_1$$

To continue the test with one more observation on $(t, N(t))$ as the random graph of $[t, N(t)]$ is between the two linear boundaries given by equations (5.2.6) and (5.2.7) where

$$a = \frac{\lambda_1 - \lambda_0}{\log \left( \frac{\lambda_1}{\lambda_0} \right)}$$

$$b_1 = \frac{\log \left( \frac{1 - \alpha}{\beta} \right)}{\log \left( \frac{\lambda_1}{\lambda_0} \right)}$$

$$b_2 = \frac{\log \left( \frac{1 - \beta}{\alpha} \right)}{\log \left( \frac{\lambda_1}{\lambda_0} \right)}$$

The parameters $\alpha, \beta, \lambda_0$ and $\lambda_1$ can be chosen in several ways. One way suggested by Stieber is $\lambda_0 = \frac{\lambda \log(q)}{q-1}$, $\lambda_1 = q \frac{\lambda \log(q)}{q-1}$ where $q = \frac{\lambda_1}{\lambda_0}$

If $\lambda_0$ and $\lambda_1$ are chosen in this way, the slope of $N_u(t)$ and $N_l(t)$ equals $\lambda$. The other two ways of choosing $\lambda_0$ and $\lambda_1$ are from past projects (for a comparison of the projects) and from part of the data to compare the reliability of different functional areas (components).
5.3 SEQUENTIAL TEST FOR SOFTWARE RELIABILITY GROWTH MODELS

In Section 2, for the Poisson process it is known that the expected value of 
\( N(t) = \lambda t \) called the average number of failures experienced in time \( t \). This is also 
called the mean value function of the Poisson process. On the other hand if we 
consider a Poisson process with a general function (not necessarily linear) \( m(t) \) as 
its mean value function the probability equation of such a process is

\[
P[N(t) = Y] = \frac{[m(t)]^y}{y!} e^{-m(t)}, y = 0, 1, 2, \ldots
\]

Depending on the forms of \( m(t) \) various Poisson processes called NHPP are 
obtained. For our two parameter Weibull model at \( \beta = 2 \) (i.e Rayleigh), the mean 
value function is given as \( m(t) = a \left( 1 - e^{-b(t)} \right) \) where \( a > 0, b > 0 \)

It may be written as

\[
P_1 = \frac{e^{-m_1(t)} [m_1(t)]^{N(t)}}{N(t)!}
\]

\[
P_0 = \frac{e^{-m_0(t)} [m_0(t)]^{N(t)}}{N(t)!}
\]

Where, \( m_1(t), m_0(t) \) are values of the mean value function at specified sets of its 
parameters indicating reliable software and unreliable software respectively.

Let \( P_0, P_1 \) be values of the NHPP at two specifications of \( b \) say \( b_0, b_1 \), where

\( b_0 < b_1 \). It can be shown that for our model \( m(t) \) at \( b_1 \) is greater than that at \( b_0 \).

Symbolically \( m_0(t) < m_1(t) \). Then the SPRT procedure is as follows:

Accept the system to be reliable if, \( \frac{P_1}{P_0} \leq B \)

\[
\text{i.e., } \frac{e^{-m_1(t)} [m_1(t)]^{N(t)}}{e^{-m_0(t)} [m_0(t)]^{N(t)}} \leq B
\]

\[
\text{i.e., } N(t) \leq \frac{\log \left( \frac{\beta}{1-\alpha} \right) + m_1(t) - m_0(t)}{\log m_1(t) - \log m_0(t)}
\]

(5.3.1)
Decide the system to be unreliable and reject if, \( \frac{P_i}{P_o} \geq A \)

\[
\log \left( \frac{1-\beta}{\alpha} \right) + m_i(t) - m_o(t) \geq \frac{\log m_i(t) - \log m_o(t)}{\log m_i(t) - \log m_o(t)}
\]

i.e., \( N(t) \geq \frac{\log m_i(t) - \log m_o(t)}{\log m_i(t) - \log m_o(t)} \) (5.3.2)

Continue the test procedure as long as

\[
\frac{\log \left( \frac{1-\beta}{1-\alpha} \right) + m_i(t) - m_o(t)}{\log m_i(t) - \log m_o(t)} < N(t) < \frac{\log \left( \frac{1-\beta}{\alpha} \right) + m_i(t) - m_o(t)}{\log m_i(t) - \log m_o(t)}
\]

(5.3.3)

Substituting the appropriate expressions of the respective mean value function – \( m(t) \) of Rayleigh we get the respective decision rules and are given in following lines

Acceptance region:

\[
N(t) \leq \frac{\log \left( \frac{\beta}{1-\alpha} \right) + a \left( e^{-\beta t} - e^{-\alpha t} \right)}{\log \left( \frac{1-e^{-\beta t}}{1-e^{-\alpha t}} \right)}
\]

(5.3.4)

Rejection region:

\[
N(t) \geq \frac{\log \left( \frac{1-\beta}{\alpha} \right) + a \left( e^{-\beta t} - e^{-\alpha t} \right)}{\log \left( \frac{1-e^{-\beta t}}{1-e^{-\alpha t}} \right)}
\]

(5.3.5)

Continuation region:

\[
\frac{\log \left( \frac{\beta}{1-\alpha} \right) + a \left( e^{-\beta t} - e^{-\alpha t} \right)}{\log \left( \frac{1-e^{-\beta t}}{1-e^{-\alpha t}} \right)} < N(t) < \frac{\log \left( \frac{1-\beta}{\alpha} \right) + a \left( e^{-\beta t} - e^{-\alpha t} \right)}{\log \left( \frac{1-e^{-\beta t}}{1-e^{-\alpha t}} \right)}
\]

(5.3.6)

It may be noted that in the above mentioned model the decision rules are exclusively based on the strength of the sequential procedure \((\alpha, \beta)\) and the values of the respective mean value functions namely, \( m_o(t), m_i(t) \). If the mean value function is linear in ‘\( t \)’ passing through origin, that is, \( m(t) = \lambda t \) the decision rules
become decision lines as described by Stieber. In that sense equations (5.3.1), (5.3.2), (5.3.3) can be regarded as generalizations to the decision procedure of Stieber. The applications of these results for live software failure data are presented with analysis in Section 5.

5.4 SPRT ANALYSIS OF DATA SETS

In this section, the developed SPRT methodology is shown for a software failure data which is of time domain and interval domain.

5.4.1 TIME DOMAIN DATA

In this section the decision rules based on the considered mean value function for FIVE different data sets, borrowed from Pham (2006), Xie et al., (2002) and SONATA software services are evaluated. Based on the estimates of the parameter ‘b’ in each mean value function, we have chosen the specifications of $b_0 = b - \delta$, $b_1 = b + \delta$ equidistant on either side of estimate of $b$ obtained through a data set to apply SPRT such that $0 < b < 1$. Assuming the value of $\delta = 0.0025$, the choices are given in the following table. For the present model of Rayleigh, The parameters are borrowed from Chapter 2, as the same data sets are used for Time domain data.

Table 5.4.1.1: Estimates of $a$, $b$ & Specifications of $b_0$, $b_1$ for Time domain

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Estimate of ‘a’</th>
<th>Estimate of ‘b’</th>
<th>$b_0$</th>
<th>$b_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30.051590</td>
<td>0.003416</td>
<td>0.000916</td>
<td>0.005916</td>
</tr>
<tr>
<td>2</td>
<td>28.851930</td>
<td>0.011827</td>
<td>0.009327</td>
<td>0.014327</td>
</tr>
<tr>
<td>3</td>
<td>19.164356</td>
<td>0.007110</td>
<td>0.004610</td>
<td>0.009610</td>
</tr>
<tr>
<td>4</td>
<td>23.719656</td>
<td>0.004824</td>
<td>0.002324</td>
<td>0.007324</td>
</tr>
<tr>
<td>5</td>
<td>31.961497</td>
<td>0.000912</td>
<td>-0.001588</td>
<td>0.003412</td>
</tr>
</tbody>
</table>

Using the selected $b_0$, $b_1$ and subsequently the $m_h(t), m_1(t)$ for the model, we calculated the decision rules given by Equations 5.3.4 and 5.3.5, sequentially at each ‘$t$’ of the data sets taking the strength $(\alpha, \beta)$ as (0.05, 0.2). These are presented for the model in Table 5.4.1.2.
Table 5.4.1.2: SPRT analysis for 5 data sets of Time domain data

<table>
<thead>
<tr>
<th>Data Set</th>
<th>T</th>
<th>N(t)</th>
<th>Acceptance Region (≤)</th>
<th>Rejection Region (≥)</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30.02</td>
<td>1</td>
<td>-0.172918</td>
<td>0.983106</td>
<td>Rejection</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>1</td>
<td>-1.488534</td>
<td>3.528161</td>
<td></td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>2</td>
<td>-0.166213</td>
<td>4.728247</td>
<td></td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>3</td>
<td>1.577669</td>
<td>6.282285</td>
<td></td>
</tr>
<tr>
<td></td>
<td>36</td>
<td>4</td>
<td>2.245085</td>
<td>6.864429</td>
<td></td>
</tr>
<tr>
<td></td>
<td>43</td>
<td>5</td>
<td>3.358601</td>
<td>7.810780</td>
<td></td>
</tr>
<tr>
<td></td>
<td>45</td>
<td>6</td>
<td>3.652182</td>
<td>8.052846</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>7</td>
<td>4.317624</td>
<td>8.583177</td>
<td></td>
</tr>
<tr>
<td></td>
<td>58</td>
<td>8</td>
<td>5.131767</td>
<td>9.165544</td>
<td></td>
</tr>
<tr>
<td></td>
<td>63</td>
<td>9</td>
<td>5.462683</td>
<td>9.344267</td>
<td></td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>10</td>
<td>5.689586</td>
<td>9.352040</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>1</td>
<td>-0.966013</td>
<td>1.974641</td>
<td></td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>2</td>
<td>-0.726181</td>
<td>2.195982</td>
<td></td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>3</td>
<td>-0.160307</td>
<td>2.715520</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5.5</td>
<td>1</td>
<td>-0.663439</td>
<td>1.222389</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7.33</td>
<td>2</td>
<td>-0.651608</td>
<td>1.233754</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>52.5</td>
<td>1</td>
<td>-0.499423</td>
<td>2.308628</td>
<td></td>
</tr>
<tr>
<td></td>
<td>105</td>
<td>2</td>
<td>0.895501</td>
<td>3.635475</td>
<td></td>
</tr>
<tr>
<td></td>
<td>131.25</td>
<td>3</td>
<td>1.795938</td>
<td>4.486175</td>
<td></td>
</tr>
<tr>
<td></td>
<td>183.75</td>
<td>4</td>
<td>3.681803</td>
<td>6.244764</td>
<td></td>
</tr>
<tr>
<td></td>
<td>201.25</td>
<td>5</td>
<td>4.267926</td>
<td>6.781199</td>
<td></td>
</tr>
<tr>
<td></td>
<td>306.25</td>
<td>6</td>
<td>6.459276</td>
<td>8.620122</td>
<td></td>
</tr>
</tbody>
</table>

From the above table it is observed that a decision of either to accept or reject the system is reached well in advance of the last time instant of the data.

5.4.2 Interval domain data

In this section the decision rules based on the considered mean value function for Five different data sets, borrowed from Pham (2006), Zhang et al., (2002), Ohba (1984a), Misra (1983) are evaluated. Based on the estimates of the parameter ‘b’ in each mean value function, the specifications of \( b_0=b-\delta \), \( b_1=b+\delta \) equidistant on
either side of estimate of b obtained through a data set to apply SPRT such that $b_0 < b < b_1$ are chosen. Assuming the value of $\delta = 0.125$, the estimates are given in the following table.

**Table 5.4.2.1: Estimates of a, b & Specifications of b₀, b₁ for Interval domain**

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Estimate of ‘a’</th>
<th>Estimate of ‘b’</th>
<th>b₀</th>
<th>b₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>6a</td>
<td>54.765902</td>
<td>0.062527</td>
<td>-0.062473</td>
<td>0.187527</td>
</tr>
<tr>
<td>7a</td>
<td>28.129276</td>
<td>0.070566</td>
<td>-0.054434</td>
<td>0.195566</td>
</tr>
<tr>
<td>8a</td>
<td>44.88134</td>
<td>0.070928</td>
<td>-0.054072</td>
<td>0.195928</td>
</tr>
<tr>
<td>9a</td>
<td>81.026164</td>
<td>0.040014</td>
<td>-0.084986</td>
<td>0.165014</td>
</tr>
<tr>
<td>10a</td>
<td>164.246019</td>
<td>0.036012</td>
<td>-0.088988</td>
<td>0.161012</td>
</tr>
</tbody>
</table>

Using the selected $b_0$, $b_1$ and subsequently the $m_0(t)$, $m_1(t)$ for the model, the decision rules given by Equations 5.3.4, 5.3.5 are calculated, sequentially at each ‘$t$’ of the data sets taking the strength ($\alpha$, $\beta$) as (0.05, 0.2). These are presented for the model in Table 5.4.2.2.

**Table 5.4.2.2: SPRT analysis for 5 data sets of Interval domain data**

<table>
<thead>
<tr>
<th>Data Set</th>
<th>T</th>
<th>N(t)</th>
<th>Acceptance region ($\leq$)</th>
<th>Rejection Region ($\geq$)</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>6a</td>
<td>1</td>
<td>2</td>
<td>0.054640</td>
<td>2.010655</td>
<td>Accept</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3</td>
<td>2.113396</td>
<td>4.028194</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4</td>
<td>4.869323</td>
<td>6.717551</td>
<td></td>
</tr>
<tr>
<td>7a</td>
<td>1</td>
<td>1</td>
<td>-0.227468</td>
<td>1.454030</td>
<td>Accept</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>0.798963</td>
<td>2.446035</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2</td>
<td>2.165549</td>
<td>3.756856</td>
<td></td>
</tr>
<tr>
<td>8a</td>
<td>1</td>
<td>3</td>
<td>0.000412</td>
<td>1.670798</td>
<td>Reject</td>
</tr>
<tr>
<td>9a</td>
<td>1</td>
<td>6</td>
<td>0.026390</td>
<td>3.265242</td>
<td>Reject</td>
</tr>
<tr>
<td>10a</td>
<td>1</td>
<td>9</td>
<td>1.129459</td>
<td>4.753555</td>
<td>Reject</td>
</tr>
</tbody>
</table>

From the above table it is observed that a decision of either to accept or reject the system is reached much in advance of the last time instant of the data.
5.5 CONCLUSION

The table 5.4.1.2 of Time domain data as exemplified for 5 Data Sets and Table 5.4.2.2 of Interval domain data as exemplified for 5 Data Sets shows that Rayleigh model is performing well in arriving at a decision. Out of 5 Data Sets of Time domain the procedure applied on the model has given a decision of rejection for 4, acceptance for 1 and continue for none at various time instant of the data as follows. Data Set #1, #2, #3 and #4 are rejected at 1st, 10th, 3rd and 2nd instant of time respectively. Data Set #5 is accepted at 6th instant of time. Out of 5 Data Sets of Interval domain the procedure applied on the model has given a decision of rejection for 3, acceptance for 2 and continue for none at various time instants of the data as follows. Data Set #6a, #8a and #10a are rejected at 1st instant of time. Data Set #7a and #9a are accepted at 3rd instant of time. Therefore, by applying SPRT on data sets it can be concluded that we can come to an early conclusion of reliable or unreliable software.