NHPP and S-Shaped Models for Testing the Software Failure Process

Dr. Kirti Arekar
Assistant Professor
K.J. Somaiya Institute of Management Studies & Research
Vidya Nagar; Vidya Vihar; Mumbai. India.
deshmukh_k123@yahoo.com/kirtiarekar@simsr.somaiya.edu

Abstract: Non-homogeneous Poisson process plays an important role in software and hardware reliability engineering. In many realistic situations there are two or more change points in NHPP models. In the software reliability, the nature of the failure data is affected by many factors, such as testing environment, testing strategy, and resource allocation. These factors are stable through the entire process of reliability analysis. In this paper, we test the different change points according to their existence by using some test statistics.

Keywords: Change points, reliability and S-Shaped Model.

1. Introduction

NHPP models play an important role in software and hardware reliability. Musa et.al (1987), Xie (1991), Pham (1999) and Singpurwalla and Wilson (1999) among others developed the different software reliability models. The first NHPP software reliability model is proposed by Goel and Okumoto (1979), they assumed that the software failure intensity is proportional to the expected number of undetected failures. Musa and Okumoto (1984) give the log-arithmetic Poisson execution time model.

Zhao (1993), first considered the change-point problem in software reliability. He modified the Jelinski-Moranda model (1972) to estimate the location of change point. Chang (2001) and Zou (2003) uses some useful NHPP software reliability models with change point. Shyur (2003) incorporated both imperfect debugging and change-point problem into NHPP model. In the NHPP model, there is only one change-point and the unknown change point can be estimated by the maximum likelihood method or the LS method. However, in many realistic situations, the change-point is unknown. Chang (2001) shows that the two change-points oppose each other. Chen and Gupta (2001) considered a problem of multiple change points. In the present paper, we considered change-point detections. At first, we give the NHPP models with multiple change points and maximum likelihood method is used to estimate the change points and other parameters of model. To test the existence of change-point(s) the test statistics is proposed.

2. NHPP Models with Change Points

Many NHPP models are very useful to describe the software failure process. In the present paper the delayed S – Shaped model with one change point is considered, and after that delayed S-Shaped model with multiple change points are considered.

2.1 S-Shaped Model

Software failure processes are called as fault counting process. Let \( \{N(t) : t > 0\} \) is considered as the cumulative number of software failure time by t. The N (t) is called as NHPP with mean value function \( m(t) \) and failure intensity \( \lambda(t) \). Geol and Okumoto (1979) assume that the software failure intensity \( \lambda(t) \) is proportional to the expected number of undetected failure i.e.,

\[
\lambda_0(t) = \frac{d m_0(t)}{dt} = b[d - m_0(t)] \quad \ldots \ (1)
\]

Where, \( a \) is initial number of faults contained in the software and \( b \) is called as the fault detection rate, the mean value function and intensity function are,

\[
m_0(t) = a(1 - e^{-bt}) \quad \ldots \ (2)
\]

And

\[
\lambda_0(t) = ab^2 te^{-bt} \quad ; \quad \alpha > 0; \beta \in R \quad \ldots \ (3)
\]

Suppose that \( n \) software failures are obtained and software process lasted at time \( T \). Let \( 0 < t_1 < t_2 < \ldots \ldots < t_n < T \) in which failures are
observed. The log-likelihood functions of the observed data are,

$$\log L_0(a, b) = \sum_{i=1}^{n} \log \lambda_0(t_i) - m_0(T)$$

$$= \sum_{i=1}^{n} \left[ ab^2 t_i e^{-b t_i} - m_0(T) \right]$$

$$= n \log(a + 2 \log b + T - bT) - a \left(1 - e^{-bT} \right)$$

The maximum likelihood estimator of parameters $a$, $b$ are obtained by solving the following two equations,

$$\frac{n}{a} = \left(1 - e^{-bT} \right) \quad \ldots (4)$$

$$\frac{2}{b} - \sum_{i=1}^{n} \frac{(T - t_i)}{b} - \frac{nT}{(1 - e^{-bT})} = 0 \quad \ldots (5)$$

3. S-Shaped Model with Change-Points

Whenever the software testing process is going on, the nature of the failure data can be affected many factors such as testing environment, testing strategy, resources allocation and so on. In this case, it is good to use a change-point method in reliability analysis.

The fault detection rate $b$ is not a constant; it is assumed to have a change-point. Therefore, the fault detection rate at testing time $t$ can be defined as,

$$b(t) = \begin{cases} b_1 ; 0 \leq t \leq \eta \\ b_2 ; t > \eta \end{cases} \quad \ldots (6)$$

The $\eta$ is the change points $b_1$ and $b_2$ are the fault detection rates before and after the change points. If $b_1 = b_2$ the change point model is equivalent to delayed S-Shaped model. Under the assumption,

$$\lambda_1(t) = \frac{d m_1(t)}{dt} = b(t) \left[a - m_1(t)\right]$$

The mean value function and intensity function can be expressed as,

$$m_1(t) = \begin{cases} a \left[1 - e^{-b_1 t}\right] ; 0 \leq t \leq \eta \\ a \left[1 - e^{-b_1 \eta - b_2 (t-\eta)}\right] ; t > \eta \end{cases} \quad \ldots (7)$$

And

$$\lambda_1(t) = \begin{cases} a b_1^2 t e^{-b_1 t} ; 0 \leq t \leq \eta \\ a b_2^2 t e^{-b_2 (t-\eta)} ; t > \eta \end{cases} \quad \ldots (8)$$

The log-likelihood function is,

$$\log L_1(\eta, a, b_1, b_2) = \log a + ae^{-b_1 \eta - b_2 (t-\eta)} + \sum_{i=1}^{N(\eta)} \left[ \log(ab_1^2) + t_i - b_1 t_i \right]$$

$$+ \sum_{i=N(\eta)+1}^{n} \left[ \log(ab_2^2) + t_i - b_2 \eta - b_2 (t_i - \eta) \right]$$

By Nguyen et.al (1984), assumes the log-likelihood function tends to infinity as the change point $\eta$ tends to failure time $t_n$ from below. Hence, the estimate value of $\eta$ cannot be obtained by maximizing the log-likelihood function over $[0, T]$. Wang and Wang (2005) restrict the change point in the interval $[t_2, t_{n-1}]$. Therefore the estimates value of the parameters $\hat{\eta}, \hat{a}, \hat{b}_1, \hat{b}_2$ are,

$$\log L_1(\eta, a, b_1, b_2) = \max_{\eta \in [t_2, t_{n-1}]} \max_{a, b_1, b_2} \log L_1(\eta, a, b_1, b_2)$$

$$= \max_{\eta \in [t_2, t_{n-1}]} \max_{a, b_1, b_2} \log L_1(\eta, a, b_1, b_2)$$

If $\eta = t_m, 2 \leq m \leq n - 1$, then the estimates of parameter $a, b_1, b_2$ are obtained by following equations,

1. For parameter $a$,

$$\frac{n}{a} = 1 - e^{-b_1 t_m - b_2 (T-t_m)} \quad \ldots (9)$$

2. For parameter $b_1$,

$$\frac{2m}{b_1} + \sum_{i=1}^{m} (t_m - t_i) - \frac{m t_m}{1 - e^{-b_1 t_m - b_2 (T-t_m)}} = 0 \quad \ldots (10)$$

3. For parameter $b_2$,

$$\frac{2(n-m)}{b_2} + 2m(T-t_m) + \sum_{i=1}^{m} (t_m - t_i) - \frac{n(T-t_m)}{1 - e^{-b_1 t_m - b_2 (T-t_m)}} = 0 \quad \ldots (11)$$

Similarly, optimal solution of $a_m, b_{i,m}, b_{2m}$ are obtained as,
(1) For parameter \( a_m \),

\[
\frac{n}{a_m} = 1 - e^{-b(t_m - t) - (T - t_m)} \quad \ldots (12)
\]

(2) For parameter \( b_{im} \),

\[
\frac{2m}{b_{im}} + \sum_{i=1}^{m} (t_m - t_i) - \frac{n t_m}{1 - e^{-b_{im} (t_m - b_{2m})}} (T - t_m) = 0
\]

\[
\ldots (13)
\]

(3) For parameter \( b_{2m} \)

\[
\frac{2(n-m) + 2m(T-t_m)}{b_{2m}} + \sum_{i=1}^{m} (t_m - t_i) - \frac{n(T-t_m)}{1 - e^{-b_{2m} (T-t_m)}} = 0
\]

\[
\ldots (14)
\]

In addition, if \( \eta = t_m / m \leq m \leq n - 1 \), the estimate of \( a, b_1 \) and \( b_2 \) can be obtained accordingly,

\[
\frac{n}{a} = 1 - e^{-b_1 (t_m - b_{2m})} \quad \ldots (15)
\]

\[
\frac{2m-1}{b_1} + \sum_{i=1}^{m} (t_m - t_i) - \frac{n t_m}{1 - e^{-b_1 (t_m - b_{2m})}} = 0
\]

\[
\ldots (16)
\]

\[
\frac{2(n-m)+1}{b_2} + 2m(T-t_m) + \sum_{i=1}^{m} (t_m - t_i) - \frac{n(T-t_m)}{1 - e^{-b_2 (T-t_m)}} = 0
\]

\[
\ldots (17)
\]

The optimal solution is denoted by \( a_{im}, b_{im}, b_{2m} \), respectively. Then the estimates \( \hat{a}, \hat{b}_1, \hat{b}_2 \) can be obtained by comparing the values of \( \log L_1(t_m, a_m, b_{im}, b_{2m}) \): \( m = 2, \ldots, n - 1 \) and \( \log L_1(t_m, a_m, b_{im}, b_{2m}) \): \( m = 3, \ldots, n - 1 \).

Here, we present S-Shaped model with \( k \) change point, fault detection rate is given by,

\[
b(t) = [b_1, b_2, \ldots, b_{k+1}] \quad 0 \leq t \leq \eta_1; \ldots; \eta_k \leq t > \eta_k
\]

Where, \( \eta_1, \eta_2, \ldots, \eta_k \) is change point and \( b_1, b_2, \ldots, b_{k+1} \) are fault detection rates. The mean value function and intensity function are as follows:

\[
m_k(t) = \begin{cases} 
\frac{a_1 - e^{-b_1 t}}{1 - e^{-b_1 (t - \eta_1)}} & ; 0 \leq t \leq \eta_1 \\
\frac{a_2 - e^{-b_1 \eta_1 - b_2 (t - \eta_1)}}{1 - e^{-b_1 \eta_1 - b_2 (\eta_2 - \eta_1) - b_{k+1} (t - \eta_k)}} & ; \eta_1 < t \leq \eta_2 \\
\frac{a_3 - e^{-b_1 \eta_1 - b_2 \eta_2 - \ldots - b_{k+1} (t - \eta_k)}}{1 - e^{-b_1 \eta_1 - b_2 \eta_2 - \ldots - b_{k+1} (t - \eta_k)}} & ; t \geq \eta_k 
\end{cases}
\]

The log-likelihood function is,

\[
\log L_k(\eta_1, \ldots, \eta_k, a_1, b_1, \ldots, b_{k+1}) = \log a + \left( a_1 - e^{-b_1 \eta_1 - b_2 (\eta_2 - \eta_1) - b_{k+1} (t - \eta_k)} \right)
\]

\[
\ldots + \sum_{i=1}^{n} \left( \log(ab^2_{i}) + \log(t) - b_t \right)
\]

The change points are restricted in the interval \([t_1, t_n]\), the log-likelihood function is bounded and estimates of \( \hat{\eta}_1, \hat{\eta}_2, \ldots, \hat{\eta}_k, \hat{b}_1, \hat{b}_2, \ldots, \hat{b}_{k+1} \) can be maximum log likelihood function is expressed as,

\[
\log L_k(\hat{\eta}_1, \hat{\eta}_2, \ldots, \hat{\eta}_k, \hat{b}_1, \hat{b}_2, \ldots, \hat{b}_{k+1}) = \max_{\eta_1, \ldots, \eta_k \in (t_1, t_n): a_1, b_1, \ldots, b_{k+1}} \log L_k(\eta_1, \eta_2, \ldots, a_1, b_1, \ldots, b_{k+1})
\]

The estimates \( \hat{\eta}_1, \hat{\eta}_2, \ldots, \hat{\eta}_k, \hat{b}_1, \hat{b}_2, \ldots, \hat{b}_{k+1} \) can be obtained similarly.

4. Test Statistics of Single Change Point

Suppose the failure times \( t_1, t_2, \ldots, t_n \) are distributed as the order statistics in an independent and identical random sample of size \( n \) from the density,

\[
f(t; b_1, b_2, \eta) = \frac{\lambda(t)}{m(t)} I(0 < t < T)
\]

\[
= \begin{cases} 
\frac{b_2 t e^{-b_1 t}}{1 - e^{-b_1 \eta - b_2 (t - \eta)}} & ; 0 \leq t \leq \eta \\
\frac{b_2 t e^{-b_1 t}}{1 - e^{-b_1 \eta - b_2 (T - t)}} & ; \eta < t < T \\
0 & ; t \geq T
\end{cases}
\]

Where \( m(T) = \int_0^T \lambda(t) \ dt \) and \( I(\cdot) \) are the indication functions.

So we construct the likelihood ratio test statistics:
5. Steps for Testing Procedure

**STEP: 1** - For the S-Shaped delayed model the estimates of $\hat{a}, \hat{b}$ are obtained by solving equation 4 and 5.

**Step: 2** - For the S-Shaped delayed model with change-point, the estimates $\hat{\eta}, \hat{a}, \hat{b}_1, \hat{b}_2$ can be obtained by comparing the values of

$$\log L(\tau, \hat{a}, \hat{b}, \hat{\eta}) - \log L_0(\hat{a}, \hat{b})$$

as discussed earlier.

**Step: 3** - Calculate $S_n$ by equation (18).

**Step: 4** - The test at $\alpha$-level is to reject $H_0$ if

$$S_n > \chi^2_{1}(\alpha),$$

Where $\chi^2_{1}(\alpha)$ is the upper $\alpha$ point of the $\chi^2$-distribution with one degree of freedom. Otherwise we accept the hypothesis.

### 5.1 An Example

The data set shown in table 1. Are collected from the system T1 and Musa (1979). This data set includes 136 faults in the testing phase. Here we use the method proposed earlier to test the existence of change-points. The S-Shaped Delayed model fit of this data with one change-point resulted in parameter estimates of $\hat{a}_0 = 140.3$ and $\hat{b}_0 = 3.75 \times 10^{-5}$.

The S-Shaped Delayed model with one change-point resulted in parameter estimates of $\hat{a}_1 = 146.4 ; \hat{b}_1 = 1.76 \times 10^{-4} ; \hat{b}_2 = 3.12 \times 10^{-5}$, and $\hat{\tau} = 1058$ (i.e., $m = 16$). Our estimation of change-point agrees with that of Wang and Wang (2005) and Zou (2003).

Although it is clear that the estimates $\hat{b}_{11}$ and $\hat{b}_{12}$ are significantly different. If the difference between the estimates $\hat{b}_{11}$ and $\hat{b}_{12}$ is larger than the threshold value, then there is a change-point. Here we can use out test statistics to test if $\hat{\tau} = 1056$ is a change-point. The value of the test statistics is $S_n = 18.20 > \chi^2_{1}(0.05) = 3.84$, and we reject the null hypothesis i.e. there is a change-point in the testing process.

### Table 1. Software failure times data: system T1

<table>
<thead>
<tr>
<th>Software Failure times (CPUs)</th>
<th>3</th>
<th>1846</th>
<th>5324</th>
<th>1025</th>
<th>1580</th>
<th>2677</th>
<th>4229</th>
<th>5648</th>
</tr>
</thead>
<tbody>
<tr>
<td>33</td>
<td>1872</td>
<td>5389</td>
<td>1049</td>
<td>1618</td>
<td>2775</td>
<td>4229</td>
<td>5656</td>
<td></td>
</tr>
<tr>
<td>146</td>
<td>1986</td>
<td>5565</td>
<td>1062</td>
<td>1622</td>
<td>2846</td>
<td>4540</td>
<td>5702</td>
<td></td>
</tr>
<tr>
<td>227</td>
<td>2311</td>
<td>5623</td>
<td>1098</td>
<td>1635</td>
<td>2849</td>
<td>4665</td>
<td>6255</td>
<td></td>
</tr>
<tr>
<td>342</td>
<td>2366</td>
<td>6080</td>
<td>1111</td>
<td>1716</td>
<td>2936</td>
<td>4759</td>
<td>6265</td>
<td></td>
</tr>
<tr>
<td>351</td>
<td>2608</td>
<td>6380</td>
<td>1141</td>
<td>1745</td>
<td>3008</td>
<td>4829</td>
<td>6266</td>
<td></td>
</tr>
<tr>
<td>353</td>
<td>2676</td>
<td>6477</td>
<td>1144</td>
<td>1775</td>
<td>3240</td>
<td>4917</td>
<td>6373</td>
<td></td>
</tr>
<tr>
<td>556</td>
<td>3278</td>
<td>7192</td>
<td>1255</td>
<td>1856</td>
<td>3679</td>
<td>5014</td>
<td>6489</td>
<td></td>
</tr>
<tr>
<td>571</td>
<td>3288</td>
<td>7447</td>
<td>1255</td>
<td>1872</td>
<td>3764</td>
<td>5204</td>
<td>7104</td>
<td></td>
</tr>
<tr>
<td>709</td>
<td>4434</td>
<td>7644</td>
<td>1279</td>
<td>1955</td>
<td>3765</td>
<td>5248</td>
<td>7436</td>
<td></td>
</tr>
<tr>
<td>759</td>
<td>5034</td>
<td>7837</td>
<td>1312</td>
<td>2056</td>
<td>3791</td>
<td>5287</td>
<td>7540</td>
<td></td>
</tr>
<tr>
<td>836</td>
<td>5049</td>
<td>7843</td>
<td>1348</td>
<td>2101</td>
<td>3971</td>
<td>5332</td>
<td>7605</td>
<td></td>
</tr>
<tr>
<td>860</td>
<td>5085</td>
<td>7922</td>
<td>1470</td>
<td>2130</td>
<td>4058</td>
<td>5344</td>
<td>8154</td>
<td></td>
</tr>
<tr>
<td>968</td>
<td>5089</td>
<td>8738</td>
<td>1525</td>
<td>2306</td>
<td>4201</td>
<td>5443</td>
<td>8270</td>
<td></td>
</tr>
<tr>
<td>105</td>
<td>5089</td>
<td>1008</td>
<td>1526</td>
<td>2412</td>
<td>4204</td>
<td>5538</td>
<td>8456</td>
<td></td>
</tr>
<tr>
<td>172</td>
<td>5097</td>
<td>1023</td>
<td>1527</td>
<td>2591</td>
<td>4218</td>
<td>5646</td>
<td>8868</td>
<td></td>
</tr>
</tbody>
</table>

### 6. Conclusion

In software reliability the problem of change-point is considered and some NHPP software reliability model with change-point has been proposed. Practically, the change-point is unknown, and it is possible that there is more than one change-point. In this article, we construct test statistics to test the existence of change-point by using S-Shaped Model. In the testing process we find that there is existence of change-points.

In the software testing phase, sometimes the failure cannot be observed exactly and only the number of failures up to a given time is known. The data use of this testing is a grouped data. But, the limitation of my study...
that my test statistics is not used for the grouped data.
Chang (2001) suggested the testing for grouped type of
data.

References

non-homogeneous Poisson process: software
reliability with change point model, Communication in Statistics: Simulation and
Computation; 30; 623-635.

detection and estimation; Communication in
Statistics: Simulation and Computation; 30; 665-
697.

(3) Goel, A. L., Okumoto, K. (1979) : Time dependent
error detection rate model for software reliability
and other performance measures, IEEE
Transaction on Reliability;28;206-211.

(4) Jeliski, Z., Moranda, P.B. (1972) : Software
reliability research, In : Freiberger, W., ed.
Statistics computer Performance Evaluation. New

(5) Nguyen, H. T., Rogers, G.S., Walker, E. A.
(1984): Estimation in change-point hazard rate
models; Biometrika; 71;299-304.

(6) Zhao. (1993): Change-point problem in software
and hardware reliability, Communication in
Statistics: Theory and Methods; 22; 757-768.

(7) Zou, F. Z. (2003) : change-point perspective on
the software failure process; Software Testing
Verification and Reliability;13;85-93.