Distribution Based Change-Point Problem with Two Types of Imperfect Debugging in Software Reliability

P. K. Kapur¹  Sameer Anand²  V. B. Singh³

¹ Department of Operational Research, University of Delhi, India, pkkapur1@gmail.com
² S. S. College of Business Studies, University of Delhi, India, sanand_or@yahoo.com
³ Delhi College of Arts & Commerce, University of Delhi, India, singh_vb@rediffmail.com

ABSTRACT
Software testing is an important phase of software development life cycle. It controls the quality of software product. Due to the complexity of software system and incomplete understanding of software, the testing team may not be able to remove/correct the fault perfectly on observation/detection of a failure and the original fault may remain resulting in a phenomenon known as imperfect debugging, or get replaced by another fault causing fault generation. In case of imperfect debugging, the fault content of the software remains same while in case of fault generation, the fault content increases as the testing progresses and removal/correction results in introduction of new faults while removing/correcting old ones. During software testing fault detection/correction rate may not be same throughout the whole testing process, but it may change at any time moment. In the literature various software reliability models have been proposed incorporating change-point concept. In this paper we propose a distribution based change-point problem with two types of imperfect debugging in software reliability. The models developed have been validated and verified using real data sets. Estimated Parameters and comparison criteria results have also been presented.

KEYWORDS
Non-homogenous Poisson process, software reliability growth model, hazard rate, imperfect debugging.

NOTATION

\( m(t) \): the mean value function or the expected number of faults detected or removed by time \( t \).
\( a(t) \): total fault content of software dependent on time.
\( p \): the probability of fault removal on a failure (i.e., the probability of perfect debugging).
\( a \): the rate at which the faults/errors may be introduced during the debugging process.
\( b \): fault removal/correction rate.
\( k(t) \): intensity function for NHPP models or fault detection rate per unit time.
\( F(t) \): distribution functions for fault removal/correction times.
\( f(t) \): density functions for fault removal/correction times.
\( z(t) \): hazard rate function.
\( \beta \): learning parameter in logistic function.

INTRODUCTION
Computer software is embedded in systems of all kinds: transportation, medical, telecommunications, military, industrial processes, entertainment, office products...the list is almost endless. Software is virtually inescapable in a modern world. And as we move into the twenty-first century, it will become the driver for new advances in everything from elementary education to genetic engineering. Software development consists of different phases: requirement analysis, design, coding, testing, implementation and maintenance called SDLC. Research has been conducted in software reliability engineering over the past three decades and many software reliability growth models (SRGM) have been proposed. The Software Reliability Growth Model (SRGM) is the tool, which can be used to evaluate the software quantitatively, develop test status, schedule status and monitor the changes in reliability performance.

Research has been conducted in software reliability engineering over the past three decades and many software reliability growth models (SRGM) have been proposed. The pioneering attempt in non-homogenous Poisson process based on SRGM was made by Goel and Okumoto (G-O) [1]. The model describes the failure observation phenomenon by an exponential curve. There are also SRGM that describe either S-shaped curves or a mixture of exponential and S-shaped curves (flexible). Some of the important contributions of these type of models are due to Yamada et al. [27], Ohba [18], Bittanti et al. [3], Kapur and Garg [16], Kapur et al. [17], Pham [24] etc.

In most of the models discussed above it is assumed that whenever an attempt is made to remove a fault, it is removed with certainty i.e. a case of perfect debugging. But the debugging activity is not always perfect because of number of factors like tester’s skill/expertise etc. In practical software development scenario, the number of failures observed/detected may not be necessarily same as the number of errors removed/corrected. Kapur and Garg [16] have discussed in their error removal phenomenon model that as testing grows and testing team gains experience, additional numbers of faults are removed without them causing any failure.
The testing team, however, may not be able to remove/correct fault perfectly on observation/detection of a failure and the original fault may remain leading to a phenomenon known as imperfect debugging, or replaced by another fault resulting in fault generation. In case of imperfect debugging the fault content of the software is not changed, but because of incomplete understanding of the software, the original detected fault is not removed perfectly. But in case of fault generation, the total fault content increases as the testing progresses because new faults are introduced in the system while removing the old original faults.

Model due to Obha and Chou [18] is an fault generation model applied on G-O model and has been also named as Imperfect debugging model. Kapur and Garg [22] introduced the imperfect debugging in G-O model. They assumed that the FDR per remaining faults is reduced due to imperfect debugging. Thus the number of failures observed/detected by time infinity is more than the initial fault content. Although these two models describe the imperfect debugging phenomenon yet the software reliability growth curve of these models is always exponential. Moreover, they assume that the probability of imperfect debugging is independent of the testing time. Thus, they ignore the role of the learning process during the testing phase by not accounting for the experience gained with the progress of software testing. Pham [24] developed an SRGM for multiple failure types incorporating fault generation. Zhang et al. [26] proposed a testing efficiency model which includes both imperfect debugging and fault generation, modeling it on the number of failures experienced/observed/detected, however both imperfect debugging and fault generation are actually seen during fault removal/correction. Recently, Kapur et al. [12] proposed a flexible SRGM with imperfect debugging and fault generation using a logistic function for fault detection rate which reflects the efficiency of the testing/removal team.

We execute the program in specific environment and improve its quality by detecting and correcting faults. Many SRGM assume that, during the fault detection process, each failure caused by a fault occurs independently and randomly in time according to the same distribution Musa et al. [21]. But the failure distribution can be affected by many factors such as running environment, testing strategy, defect density and resource allocation. On the other hand, in practice, if we want to detect more faults for a short period of time, we may introduce new techniques or tools that are not yet used, or bring in consultants to make a radical software risk analysis. In addition, there are newly proposed automated testing tools for increasing test coverage and can be used to replace traditional manual software testing regularly. The benefits to software developers testers include increased software quality, reduced testing costs, improved release time to market, repeatable test steps, and improved testing productivity. These technologies can make software testing and correction easier, detect more bugs, save more time, and reduce much expense. Altogether, we wish that the consultants, new automated test tools or techniques could greatly help us in detecting additional faults that are difficult to find during regular testing and usage, in identifying and correcting faults most cost effectively and in assisting clients to improve their software development process. Thus, the fault detection rate may not be smooth and can be changed at some time moment \( \tau \) called change-point. Many researchers have incorporated change point in software reliability growth modeling. Firstly Zhao [28] incorporated change-point in software and hardware reliability. Huang et al. [7] used change-point in software reliability growth modeling with testing effort functions. The imperfect debugging with change-point has been introduced in software reliability growth modeling by Shyur [25]. Kapur et al. [9,14] introduced various testing effort functions and testing effort control with change-point in software reliability growth modeling. Goswami et al.[6] and Kapur et al.[15] proposed a software reliability growth model for errors of different severity using change-point. The multiple change-points in software reliability growth modeling for fielded software has been proposed by Kapur et al. [11]. Later on SRGM based on stochastic differential equations incorporating change-point concept has been proposed by Kapur et al. [10].

**BASIC ASSUMPTION**

The NHPP models are based on the assumption that the software system is subject to failures at random times caused by manifestation of remaining faults in the system. Hence NHPP are used to describe the failure phenomenon during the testing phase. The counting process \( \{N(t), t \geq 0\} \) of an NHPP process is given as follows.

\[
Pr\{N(t) = k\} = \frac{(m(t))^k}{k!} e^{-m(t)}, \quad k = 0,1,2,\ldots
\]  

and

\[
m(t) = \int_0^t \lambda(x) dx
\]

The intensity function \( \lambda(x) \) (or the mean value function \( m(t) \)) is the basic building block of all the NHPP models existing in the software reliability engineering literature. The proposed models are based upon the following basic
assumptions:
1. Failure fault removal phenomenon is modeled by NHPP.
2. Software is subject to failures during execution caused by faults remaining in the software.
3. Failure rate is equally affected by all the faults remaining in the software.
4. When a software failure occurs, an instantaneous repair effort starts and the following may occur:
   (a) Fault content is reduced by one with probability \( p \)
   (b) Fault content remains unchanged with probability \( 1-p \).
5. During the fault removal process, whether the fault is removed successfully or not, new faults are generated with a constant probability \( \alpha \).
6. Fault detection/removal rate may change at any time moment.

Assumption 4 and 5 captures the effect of imperfect debugging and fault generation respectively.

MODEL DEVELOPMENT

In this section, we formulate distribution based software reliability growth models incorporating change-point and two types of imperfect debugging. Since the faults in the software systems are detected and eliminated during the testing phase, the number of faults remaining in the software system gradually decreases as the testing procedure goes on. Thus under the common assumptions for software reliability growth modeling, we consider the following linear differential equation.

\[
\frac{d m(t)}{dt} = b(t)(a - m(t))
\]  (3)

Where \( b(t) \) is a fault detection rate per remaining faults at testing time \( t \). Here we consider the fault detection rate as hazard rate \( z(t) \), initial fault is not the constant but the function of time and incorporating the imperfect debugging. So the above equation can be written as

\[
\frac{d m(t)}{dt} = z(t) p(a(t) - m(t))
\]  (4)

We assume that faults can be introduced during the debugging phase with a constant fault introduction rate \( \alpha \). Therefore, the fault content rate function, \( a(t) \), is a linear function of the expected number of faults detected by time \( t \). That is

\[
a(t) = a + \alpha m(t),
\]

above equation becomes

\[
\frac{d m(t)}{dt} = z(t) p\left(a + \alpha m(t) - m(t)\right)
\]  (4)

In the proposed model we assume that the hazard rate \( z(t) = \frac{f(t)}{1 - F(t)} \), the fault introduction rate \( \alpha \) and probability of perfect debugging \( p \), may be changed at some time moment \( \tau \) called change point.

After incorporating change-point, we get the following form of fault detection/removal, probability of perfect debugging and fault generation rate.

\[
z(t) = \begin{cases} 
\frac{f_1(t)}{1 - F_1(t)} & \text{for } t \leq \tau \\
\frac{f_2(t)}{1 - F_2(t)} & \text{for } t > \tau
\end{cases}
\]  (5)

Where \( F_1, f_1, F_2, f_2 \) are the distributions, density functions before and after change point respectively.

The equation (5) can be rewritten as

Probability of perfect debugging rate will be

\[
p = \begin{cases} 
p_1 & \text{for } t \leq \tau \\
p_2 & \text{for } t > \tau
\end{cases}
\]  (6)

and fault content rate

\[
a(t) = \begin{cases} 
a + \alpha m(t) & \text{for } t \leq \tau \\
\alpha \left(m(t) - m(t)\right) & \text{for } t > \tau
\end{cases}
\]  (7)

The Equation (5) can be rewritten as

\[
z(t) = \frac{f_1(t)}{1 - F_1(t)}U(t - \tau) + \frac{f_2(t)}{1 - F_2(t)}U(t - \tau)
\]

Using unit step function give by

\[
U(x) = \begin{cases} 
0 & \text{if } x < 0 \\
1 & \text{if } x > 1
\end{cases}
\]

Similarly equation (6) & (7) can also be rewritten as

\[
p = pU(t - \tau) + pU(t - \tau)
\]

& \( a(t) = a + \alpha m(t)U(t - \tau) + \alpha \left(m(t) - m(t)\right)U(t - \tau) \)

Now using equation (5), (6) and (7), the equation (4) can be rewritten as,

\[
\frac{d m(t)}{dt} = \begin{cases} 
\frac{f_1(t)}{1 - F_1(t)}p_1(a + \alpha m(t) - m(t)) & \text{for } t \leq \tau \\
\frac{f_2(t)}{1 - F_2(t)}p_1(a + \alpha m(t) + \alpha \left(m(t) - m(t)\right) - m(t)) & \text{for } t > \tau
\end{cases}
\]  (8)
After solving the above equations, we get the following solutions

\[
m(t) = \begin{cases} 
\frac{a}{(1-\alpha_1)} \left[1-(1-F_1(t))^{\beta_1}\right] & \text{for } t \leq \tau \\
\frac{a}{(1-\alpha_2)} \left\{ \left[1-(1-F_1(t))^{\beta_1}\right] + \left(\frac{a_1-a_2}{1-a_2}\right)m(\tau) \right\} & \text{for } t > \tau 
\end{cases}
\] (9)

\[
m(t) = \begin{cases} 
\frac{a}{(1-\alpha_1)} \left[1-(1+\beta_1)(\exp(-b_1)t)\right]^{\beta_1} & \text{for } t \leq \tau \\
\frac{a}{(1-\alpha_2)} \left\{ \left[1-(1+\beta_1)(\exp(-b_1)t)\right]^{\beta_1} + \left(\frac{a_1-a_2}{1-a_2}\right)m(\tau) \right\} & \text{for } t > \tau 
\end{cases}
\] (15)

**SRGM-1**

The following exponential distribution function is used to model SRGM-1:

Let \( T \sim \exp(b_1) \) for \( t \leq \tau \)

i.e. \( F_1(t) = 1 - \exp(-b_1t) \) for \( t \leq \tau \) (10)

and

Let \( T \sim \exp(b_2) \) for \( t > \tau \)

\[
F_2(t) = 1 - \exp(-b_2t) \quad \text{for } t > \tau
\] (11)

Substituting the value of \( F_1(t) \) and \( F_2(t) \) from Equation (10) and (11) into Equation (9), we get:

\[
m(t) = \begin{cases} 
\frac{a}{(1-\alpha_1)} \left[1-\exp\left(-b_1p_1(1-\alpha_1)t\right)\right] & \text{for } t \leq \tau \\
\frac{a}{(1-\alpha_2)} \left[1-\exp\left(-b_1p_1(1-\alpha_1)t - b_2p_2(1-\alpha_2)(t-\tau)\right)\right] + \left(\frac{a_1-a_2}{1-a_2}\right)m(\tau) & \text{for } t > \tau
\end{cases}
\] (12)

The above model can be reduced to the model given by Shyur [25] if we consider the perfect debugging and no fault generation.

**SRGM-2**

Let \( F(t) \) be a two-stage Erlangian distribution function i.e. ,

\( T \sim \text{Erlang-2}(b_1) \) for \( t \leq \tau \)

i.e. \( F_1(t) = 1 - (1+b_1t)\exp(-b_1t) \) for \( t \leq \tau \)

(13)

And

\( T \sim \text{Erlang-2}(b_2) \) for \( t > \tau \)

i.e. \( F_2(t) = 1 - (1+b_2t)\exp(-b_2t) \) for \( t > \tau \) (14)

Substituting the value of \( F_1(t) \) and \( F_2(t) \) from Equation (13) and (14) into Equation (9), we get:

\[
m(t) = \begin{cases} 
\frac{a}{(1-\alpha_1)} \left[\left(\frac{1+\beta_1}{1+\beta_1}\right)\exp\left(-b_1p_1(1-\alpha_1)t\right)\right] & \text{for } 0 \leq t \leq \tau \\
\frac{a}{(1-\alpha_2)} \left\{ \left[\left(\frac{1+\beta_1}{1+\beta_1}\right)\exp\left(-b_1p_1(1-\alpha_1)t\right)\right] + \left(\frac{a_1-a_2}{1-a_2}\right)m(\tau) \right\} & \text{for } t > \tau
\end{cases}
\] (13)

\[
m(t) = \begin{cases} 
\frac{a}{(1-\alpha_1)} \left[\left(\frac{1+\beta_1}{1+\beta_1}\right)\exp\left(-b_1p_1(1-\alpha_1)t\right)\right] & \text{for } 0 \leq t \leq \tau \\
\frac{a}{(1-\alpha_2)} \left\{ \left[\left(\frac{1+\beta_1}{1+\beta_1}\right)\exp\left(-b_1p_1(1-\alpha_1)t\right)\right] + \left(\frac{a_1-a_2}{1-a_2}\right)m(\tau) \right\} & \text{for } t > \tau
\end{cases}
\] (18)

For further simplifying the estimation procedure we may assume \( \alpha_1 = \alpha_2 = \alpha \) and \( p_1 = p_2 = p \).
Model Validation, Comparison Criteria and Data Analyses

Model Validation
To illustrate the estimation procedure and application of the SRGM (existing as well as proposed) we have carried out the data analysis of real software data set. The parameters of the models have been estimated using statistical package SPSS and the change-point of the data sets have been judged by using change-point analyzer.

Data set 1 (DS-1)
The first data set (DS-1) had been collected during 35 months of testing a radar system of size 124 KLOC and 1301 faults were detected during testing. This data is cited from Brooks and Motley [4]. The change-point for this data set is 17th month.

Data set 2 (DS-2)
The second data set (DS-2) had been collected during 19 weeks of testing a real time command and control system and 328 faults were detected during testing. This data is cited from Ohba [19]. The change-point for this data set is 6th week.

Comparison Criteria for SRGM
The performance of SRGM are judged by their ability to fit the past software fault data (goodness of fit) and predicting the future behavior of the fault.

Goodness of Fit criteria
The term goodness of fit is used in two different contexts. In one context, it denotes the question if a sample of data came from a population with a specific distribution. In another context, it denotes the question of “How good does a mathematical model (for example a linear regression model) fit to the data”?

The Mean Square - Error (MSE):
The model under comparison is used to simulate the fault data, the difference between the expected values, \( \hat{m}(t_i) \) and the observed data \( y_i \) is measured by MSE as follows.

\[
MSE = \frac{1}{k} \sum_{i=1}^{k} (\hat{m}(t_i) - y_i)^2
\]

where \( k \) is the number of observations. The lower MSE indicates less fitting error, thus better goodness of fit [17].

Coefficient of Multiple Determination (R2):
We define this coefficient as the ratio of the sum of squares resulting from the trend model to that from constant model subtracted from 1.

\[
R^2 = 1 - \frac{\text{residual SS}}{\text{corrected SS}}
\]

R2 measures the percentage of the total variation about the mean accounted for the fitted curve. It ranges in value from 0 to 1. Small values indicate that the model does not fit the data well. The larger \( R^2 \), the better the model explains the variation in the data [17].

Bias
The difference between the observation and prediction of number of failures at any instant of time \( i \) is known as PEi (prediction error).

The average of PEs is known as bias. Lower the value of Bias better is the goodness of fit [8].

Variation
The standard deviation of prediction error is known as variation.

\[
Variation = \sqrt{\frac{1}{N-1} \sum (PE_i - \text{Bias})^2}
\]

Lower the value of Variation better is the goodness of fit [8].

Root Mean Square Prediction Error
It is a measure of closeness with which a model predicts the observation.

\[
\text{RMSPE} = \sqrt{(\text{Bias}^2 + \text{Variation}^2)}
\]

Lower the value of Root Mean Square Prediction Error better is the goodness of fit [8].

Data Analyses

For DS-1
The parameter estimation and comparison criteria results for DS-1 of all the models under consideration can be viewed through Table I(a) and I(b). It is clear from the table that the value of \( R^2 \) for SRGM-3 is higher and value of MSE is lower in comparison with other models and provides better goodness of fit for DS-1.

For DS-2
The parameter estimation and comparison criteria results for DS-2 of all the models under consideration can be viewed through Table II(a) and II(b). It is clear from the table that the value of \( R^2 \) for SRGM-3 is higher and value of MSE is lower in comparison with other models and provides better goodness of fit for DS-2.
<table>
<thead>
<tr>
<th>SRGM-2</th>
<th>.990</th>
<th>104.906</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRGM-3</td>
<td>.992</td>
<td>83.111</td>
</tr>
</tbody>
</table>

**GOODNESS OF FIT CURVES**

For DS-1

![Goodness of fit curves](image1)

For DS-2

![Goodness of fit curves](image2)

**CONCLUSION**

In this paper we have developed a distribution based change-point problem with two types of imperfect debugging in software reliability. With this approach, we can derive existing models and propose new model. All these models have been validated and verified using real data sets. Parameter estimates, comparison results and goodness of fit curves have also been presented.

**FUTURE SCOPE**

In future, we will try to develop more models in the same line by using Erlang normal, weibull and gamma distribution functions. Models can be extended for multiple change-points problem.

**REFERENCE**

Distribution Based Change-Point Problem with Two Types of Imperfect Debugging in Software Reliability


