A Bivariate Software Reliability Model with Change-Point and Its Applications

Shinji Inoue, Shigeru Yamada

Department of Social Management Engineering, Graduate School of Engineering, Tottori University
E-mail: {ino,yamada}@sse.tottori-u.ac.jp

Received February 28, 2011; Revised March 20, 2011; accepted March 23, 2011

Abstract

Testing-time when a change of a stochastic characteristic of the software failure-occurrence time or software failure-occurrence time-interval is observed is called change-point. It is said that effect of the change-point on the software reliability growth process influences on accuracy for software reliability assessment based on a software reliability growth model (SRGM). We propose an SRGM with the effect of the change-point based on a bivariate SRGM, in which the software reliability growth process is assumed to depend on the testing-time and testing-effort factors simultaneously, for accurate software reliability assessment. And we discuss an optimal software release problem for deriving optimal testing-effort expenditures based on our model. Further, we show numerical examples of software reliability assessment based on our bivariate SRGM and estimation of optimal testing-effort expenditures by using actual data.

Keywords: Software Reliability, Software Reliability Growth Factor, Change-Point, Bivariate Software Reliability Growth Model, Optimal Testing-Effort Expenditing Problem

1. Introduction

We are required to conduct quantitative software quality/reliability assessment in terms of software quality assurance in a testing phase. And it is very important to measure software quality/reliability of the final software product with accuracy in the testing-phase. A software reliability growth model (abbreviated as SRGM) [1-4] is known as one of the useful mathematical tool for quantitative assessment of software reliability. This mathematical model enables us to describe a software reliability growth process observed in the actual testing-phase by treating the software failure-occurrence or the software fault-detection phenomenon as random variables.

Needless to say, it is preferable that the SRGMs are developed under feasible modeling assumptions, which reflect actual software failure-occurrence phenomena. Most of SRGMs proposed so far have been developed under the following assumptions: (1) the software reliability growth process depends only on the testing-time essentially, (2) the stochastic characteristics for the software failure-occurrence or the software fault-detection phenomenon does not change throughout the testing-phase. In an actual testing-phase, it is not necessarily that the common assumptions mentioned above is always appropriate. That means, it is natural to consider the software reliability growth process observed in the actual testing-phase depends on not only the testing-time but also the other software reliability growth factors, such as the testing-coverage, the testing-effort expenditures, the number of executed test-cases [5-7]. And the stochastic characteristics for the software failure-occurrence or the software fault-detection phenomenon changes due to changing the fault-target, the difference of the fault-density for each module, and so forth [8-14]. Especially, testing-time when the stochastic characteristic for the software failure-occurrence or the software fault-detection phenomenon notably changes is called change-point [8].

This paper discusses a bivariate SRGM with the effect of a change of the software reliability growth factors for overcoming the problem mentioned above. Our bivariate SRGM enables us to describe a software reliability growth process depending on the testing-time and the testing-effort factors, and also enables us to consider the effect of the change of the software reliability growth factors at change-point. Further, we discuss an optimal release problem for deriving optimal testing-effort expenditures based on our bivariate SRGM. Finally, we show numerical examples for two-dimensional software
reliability analysis and estimation of an optimal testing-effort expenditures based on our bivariate model by using actual software fault data.

2. Two-Dimensional Software Reliability Growth Modeling

A bivariate SRGM in which the number of detectable faults in a software system is assumed to be finite can be developed by the following modeling assumptions [5, 15, 16]:

(A1) Whenever a software failure is observed, the fault is detected immediately, and no new faults are introduced in the fault-detection procedures.

(A2) Each software failure occurs at independently and identically distributed random times with the bivariate probability distribution function \( F(s,u) = \Pr\{S \leq s, U \leq u\} \), where \( S \) and \( U \) are the random variables representing the testing-time and cumulative testing-effort expenditure, respectively. And \( \Pr\{A\} \) represents the joint density function and the probability of event \( A \), respectively.

(A3) The initial number of faults in the software system, \( N_0(>0) \), is a random variable, and is finite.

Figure 1 shows the stochastic quantities for the two-dimensional software failure-occurrence or the software fault-detection phenomenon. Now we define the two-dimensional stochastic process \( \{N(s,u), s \geq 0, u \geq 0\} \) [17] representing the number of faults detected in the two-dimensional space \([0,s] \times [0,u]\). Then, we have a bivariate probability mass function that \( m \) faults are detected in the two-dimensional space \([0,s] \times [0,u]\) as

\[
\Pr\{N(s,u) = m\} = \sum_{n=m}^{\infty} \binom{n}{m} \left[ F(s,u) \right]^m \left[ 1 - F(s,u) \right]^{n-m} \Pr\{N_0 = n\}
\]

\( (m = 0, 1, 2, \ldots) \)  

(1)

from the modeling assumptions. From Equation (1), we can say that the stochastic behavior of the software failure-occurrence or the software fault-detection phenomenon can be characterized by assuming the probability mass function for the initial number of faults in the software system, \( N_0 \). In this paper, we assume that the initial fault content follows a Poisson distribution with parameter \( \omega \). Then we have:

\[
\Pr\{N(s,u) = m\} = \frac{(\omega F(s,u))^m}{m!} \exp\left(-\omega F(s,u)\right)
\]

(2)

from Equation (1). Equation (2) is essentially equivalent to a two-dimensional nonhomogeneous Poisson process (abbreviated as a two-dimensional NHPP) [5, 6] with mean value function \( E[N(s,u)] = \lambda(s,u) = \omega F(s,u) \), where \( E[\cdot] \) denotes the expectation.

3. Two-Dimensional Change-Point Modeling

We discuss a bivariate software reliability growth modeling with effect of change-point on the software reliability
growth process. In our research, we extend the assumption (A2) as that the stochastic characteristics of the software failure-occurrence phenomenon is changed at change-point. And the change-point occurs just only one time throughout testing-phase for the sake of simplification of this discussion.

Let us denote the change-point for the testing-time and the testing-effort factors by \( \tau = \{ r_s, r_t \} \) \((0 < r_s < r_t, 0 < u < u)\), where \( u \) represents the testing-effort expenditure expended up to the testing termination-time \( s \). To start with, we define stochastic quantities for our bivariate software reliability growth modeling with effect of the change-point as shown in Figure 2. And we assume the following relationship between the software failure-occurrence phenomenon and those after the change-point:

\[
\begin{align*}
M_1 &= \alpha_s (X_i), \\
C_1 &= \alpha_s (Z_i), \\
K_1 &= \alpha_u (Y_i), \\
D_1 &= \alpha_u (W_i),
\end{align*}
\]

where \( \alpha_i(\cdot) \) and \( \alpha_s(\cdot) \) represent testing-environmental functions [18] for the testing-time and testing-effort factors, respectively. The testing-environmental function characterizes the relationship of the software failure-occurrence phenomenon before and after the change-point. Concretely, we assume that the relationship can be formulated as:

\[
R = \alpha Q 
\]  

(4)

in this paper. In Equation (4), \( R \) and \( Q \) represent column vectors, \( R = (m \times k)^T \) \((m > 0, k > 0)\) and \( Q = (x y)^T \) \((x > 0, y > 0)\), where the superscript \( T \) represents the transposed matrix and \( m, k, x, \) and \( y \) are the observed data for the random variables \( M, K, X, \) and \( Y \) in Figure 2, respectively. And \( \alpha \) is the constant vector, \( \alpha = (\alpha_s, \alpha_u) \), which represents the relative magnitude of the effect of change-point on the software reliability growth process for each factor. Equation (4) is one of the examples for the testing-environmental function. However, we can get to know the effect of the change-point on the software reliability growth process simply by assuming Equation (4) as the testing-environmental function.

Now we suppose that \( n(>0) \) faults have been detected up to change-point and their fault data from the test-beginning have been observed as \( (x_0, y_0), (x_1, y_1), \cdots (x_n, y_n) \), where \( x_0 = 0, y_0 = 0, 0 < x_1 < x_2 \cdots < x_n \leq r_n, \) and \( 0 < y_1 < y_2 \cdots < y_n \leq u_n \). Let us suppose that the number of faults detected in the testing-territory before change-point, \( \{ 0, s \} \times \{ 0, u \} (s \leq r_s, u \leq r_u) \), follows the two-dimensional NHPP in Equation (5) with mean value function \( \Lambda_A(s, u) \). Then, the bivariate probability distribution function for the \( (M_1, K_1) \) can be derived by its cofunction. The cofunction is derived as the following conditional probability:

\[
Pr[M_1 > s, K_1 > u] = Pr[X_{n+1} > \tau_s - x_s + s / \alpha_s, Y_{n+1} > \tau_u - y_u + u / \alpha_u]
\]

(5)

The denominator in Equation (5) can be derived as:

\[
Pr[X_{n+1} > \tau_s - x_s, Y_{n+1} > \tau_u - y_u]
\]

(6)

based on the property of the two-dimensional NHPP. And also, we have the numerator as:

\[
\exp[-\Lambda_B(\tau_s, \tau_u) + \Lambda_B(x_s, \tau_u) + \Lambda_B(y_s, \tau_u) - \Lambda_B(x_s, y_u)]
\]

(7)

From Equations (6) and (7), Equation (5) can be written as:

\[
Pr[M_1 > s, K_1 > u] = \exp[-\Lambda_B(\tau_s + s / \alpha_s, \tau_u + u / \alpha_u)] + \Lambda_B(\tau_s + s / \alpha_s, \tau_u + u / \alpha_u) - \Lambda_B(\tau_s + s / \alpha_s, \tau_u)
\]

(8)

From Equation (8), the expected number of faults detected after the change-point, \( \Lambda_A(s, u) \), is derived as:

\[
\Lambda_A(s, u) = -\log[Pr[M_1 > s - r_s, K_1 > u - r_u]]
\]

(9)

\[
\Lambda_A(s, u) = \Lambda_B(\tau_s, \tau_u) + \Lambda_A(s, u)
\]

(10)

Equation (10) imply that our bivariate change-point model for software reliability assessment can be devel-
opied by assuming the two-dimensional mean value function before change-point, \( \Lambda_0(s, u) \).

4. Software Reliability Assessment Measure

Software reliability assessment measures are derived by stochastic properties of the SRGM, and play an important role in quantitative software reliability assessment based on the SRGM. An operational software reliability [5] is defined as the probability that a software failure does not occur in the time-interval \((s_e, s_e + \eta)\) \( (s_e \geq 0, \eta \geq 0) \) given that the testing has been going up to testing-time \( s_e \) and the testing-effort has been expended up to \( u_e \) by testing-time \( s_e \). From the basic notion of this measure and the stochastic properties in Equation (1), we can generally formulate the operational software reliability as:

\[
R(\eta \mid s_e, u_e) = \sum_k \Pr\{N(s_e + \eta, u_e) = k \mid N(s_e, u_e) = k\} \\
\quad \times \Pr\{N(s_e, u_e) = k\} \\
\quad \times \left[ \frac{n!}{k!} \left( 1 - F(s_e + \eta, u_e) \right)^k \right] \\
\quad \times \Pr\{N_0 = n\}
\]

(11)

Assuming that \( N_0 \) follows a Poisson distribution with parameter \( \omega \), we can derive the operational software reliability function as:

\[
R(\eta \mid s_e, u_e) = \exp\left[-\{\Lambda(s_e + \eta, u_e) - \Lambda(s_e, u_e)\}\right]
\]

(12)

by using Equation (11).

5. Parameter Estimation

Parameters of our bivariate model with change-point for software reliability assessment can be estimated by the method of maximum-likelihood. Suppose that we have observed \( K \) data pairs \((s_k, u_k, y_k)\) \( (k = 0, 1, 2, \cdots K) \) with respect to the number of fault \( y_k \), which have been detected in the testing-territory \([0, t_s] \times [0, t_u] \times [t_s, t_u] \times [t_u, u_e] \). The logarithmic likelihood function, \( \ln L(\theta, \alpha \mid \tau) \), for the two-dimensional stochastic process \( \{N(s, u), s \geq 0, u \geq 0\} \) given the change-point \( \tau \) can be derived as:

\[
\ln L(\theta, \alpha \mid \tau) = \sum_{k=1}^K (y_k - y_{k-1}) \cdot \ln \left[ \Lambda(s_k, u_k; \theta, \alpha \mid \tau) \right] \\
\quad - \Lambda(s_{k-1}, u_{k-1}; \theta, \alpha \mid \tau) \\
\quad - \sum_{k=1}^K \ln \left[ \left( y_k - y_{k-1} \right)^\tau \right]
\]

(13)

where \( \theta \) represents the set of the parameter in \( \Lambda(s, u) \). Then, we obtain the following equation:

\[
\frac{\partial \ln L(\theta, \alpha \mid \tau)}{\partial \theta} = \frac{\partial \ln L(\theta, \alpha \mid \tau)}{\partial \alpha} = 0
\]

(14)

The maximum-likelihood estimates are obtained by solving Equation (14) numerically.

6. Optimal Software Release Problem

If debugging cost before change-point and its after change-point a different each other, a trade-off problem between the effect of the change-point on the software reliability growth process and the related cost arises in a conventional optimal software release problem [19]. This paper discusses an optimal problem for estimating optimal shipping time and change-point of a software system based on our model proposed in this paper.

Now, we define cost parameter for formulating the expected total software cost with change-point as follows:

\( c_1 \): debugging cost for one fault before the change-point in the testing-phase \((c_1 > 0)\),

\( c_2 \): debugging cost for one fault after the change-point in the testing-phase \((c_2 > 0)\),

\( c_3 \): debugging cost for one fault in the operational phase \((c_1 < c_3, c_2 < c_3)\)

\( c_4 \): testing cost per arbitrary testing-time and testing-effort \((c_4 > 0)\).

Letting \( S, U, \beta_s, \) and \( \beta_u \) be the termination time of testing, the total testing-effort expended up to the termination time of the testing, the time duration from \( t_s \) to \( S \), and the testing-effort expended from \( t_u \) to \( U \), we have the expected total software cost, \( C(S, U, \beta_s, \beta_u) \), as:

\[
C(S, U, \beta_s, \beta_u) = c_1 \Lambda_1(S - \beta_s, U - \beta_u) \\
+ c_2 \left[ \Lambda_2(S, U) - \Lambda_1(S - \beta_s, U - \beta_u) \right] \\
+ c_3 \{ \omega - \Lambda_2(S, U) \} + c_4 SU
\]

(15)

Generally, \( S^*, U^*, \beta_s^*, \) and \( \beta_u^* \), which are the cost-optimal release time, the optimal testing-effort expenditures, the optimal time duration from \( t_s \) to \( S^* \), and the optimal testing-effort expended from \( t_u \) to \( U^* \), are derived by minimizing the expected total software cost, \( C(S, U, \beta_s, \beta_u) \). Therefore, \( S^*, U^*, \beta_s^*, \) and \( \beta_u^* \), are obtained by solving the following equations simultaneously:

\[
\frac{\partial C(S, U, \beta_s, \beta_u)}{\partial S} = \frac{\partial C(S, U, \beta_s, \beta_u)}{\partial U} = \frac{\partial C(S, U, \beta_s, \beta_u)}{\partial \beta_s} = \frac{\partial C(S, U, \beta_s, \beta_u)}{\partial \beta_u} = 0
\]

(16)

We should note that Equation (16) is a necessary condition for the optimal solutions. Equations (15) and (16) can be simplified to an optimal testing-effort expending problem for estimating optimal testing-effort expendi-
Figure 3. Estimated two-dimensional mean value function with effect of change-point, \( \tau_s = 6, \tau_u = 12.89 \)

Figure 4. Estimated operational software reliability, \( \hat{R}(\eta \mid 19, 47.65) \)

7. Numerical Examples

We show numerical examples of our bivariate model and its application to an optimal software release problem. In this paper, we use actual data consisted of 19 data pairs: \((s_k, u_k, y_k)(k = 0, 1, 2, \ldots, 19), t_0 = 19(\text{weeks}), s_0 = 47.65(\text{CPU hours}), y_{19} = 328)\) [20]. And we assume that the software failure-occurrence time distribution before change-point follows the following bivariate probability distribution function [21]:

\[
F(s, u) = \begin{cases} 
(1 - e^{-\omega s})(1 - e^{-bu})(1 + ze^{-au-bu}) & (a > 0, b > 0, -1 \leq z \leq 1) 
\end{cases}
\]

And we set \(a_s = a_u = a\) for simplification and we also set \(\tau = \{\tau_s, \tau_u\} = \{6, 12.89\}\) by following the actual behavior of the software failure-occurrence phenomenon. Then, we estimate the parameters as \(\hat{\omega}, \hat{a}, \hat{b}, \hat{z},\) and \(\hat{\alpha}\), which are the estimates of \(\omega, a, b, z,\) and \(\alpha\) by the method of maximum likelihood discussed in Section 5.

Figure 3 shows the two-dimensional behavior of the estimated mean value function with effect of the change-point, where \(\tau_s = 6, \tau_u = 12.89\). In Figure 3, the dotted line represents the actual behavior of the cumulative number of detected faults and the curved surface the estimated behavior. We see that the expected number of faults is estimated to be zero outside the software failure-occurrence territory, which has been explained in Figure 2. This is one of the feature for our two-dimensional SRGM with the effect of the change-point. Further, Figure 4 shows the estimated operational software reliability, \(\hat{R}(\eta \mid 19, 47.65)\). From Figure 4, we can estimate the operational software reliability at 0.3 weeks after the termination time of the testing, \(\hat{R}(0.3 \mid 19, 47.65)\), to be about 0.036.

Then, we show numerical examples for an optimal release problem for deriving optimal testing-effort expenditure based on our bivariate change-point model. This problem is one of the simplified problem for our optimal software release problem. In this problem, \(h_2(S, U)\) in Equation (19) can be derived as:

\[
h_2(S, U) = \frac{\partial A_2(S, U)}{\partial U} = \frac{-c_2 - c_3}{c_3 - c_1} \frac{s - h_2(S, U)}{h_2(S, U)} = 0
\]

where \(h_2(S, U) = \frac{\partial A_2(S, U)}{\partial U} \).
Table 1. Sensitivity analysis of the optimal testing-effort expenditures.

<table>
<thead>
<tr>
<th>$c_3$</th>
<th>$c_4$</th>
<th>$U^*$</th>
<th>$C(S, U^*, \beta_1, \beta_2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>5</td>
<td>3.2501</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>3</td>
<td>10.856</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>1</td>
<td>28.840</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>1</td>
<td>43.318</td>
</tr>
<tr>
<td>1</td>
<td>20</td>
<td>1</td>
<td>57.253</td>
</tr>
<tr>
<td>1</td>
<td>40</td>
<td>1</td>
<td>70.975</td>
</tr>
<tr>
<td>1</td>
<td>80</td>
<td>1</td>
<td>84.606</td>
</tr>
</tbody>
</table>

Figure 5. Time-dependent behavior of the estimated expected total software cost and $\hat{h}_2(S, U)$. ($\tau_s = 6$, $\tau_u = 12.89$, $c_5 = 1$, $c_3 = 5$, and $c_4 = 1$).

$U^* = 28.84$ (CPU hours) and $C(S, U^*, \beta_1, \beta_2) = 2925.33$.

8. Concluding Remarks

This paper discussed a bivariate software reliability growth modeling with the effect of the change-point and an optimal software release problem based on our model. Further, we showed numerical examples of software reliability analysis and an optimal testing-effort expending problem based on our model by using actual data. Our bivariate SRGM with the effect of change-point is expected to contribute high accuracy assessment of software reliability in a testing-phase. In the further studies, we have to investigate the effectiveness and validity of our model by using a lot of data sets collected from actual software development projects, and have to give a discussion on sufficient conditions for the optimal software release problem.

9. Acknowledgement

This work was supported in part by the Grant-in-Aid for Scientific Research (C), Grant No. 22510150, from the Ministry of Education, Sports, Science, and Technology of Japan.

10. References


