EARLY CHARACTERIZATION OF THE DEFECT REMOVAL PROCESS

Yashwant K. Malaiya*
Computer Science Department
Colorado State University
Fort Collins CO 80523
(303) 491-7031
malaiya@cs.colostate.edu

Abstract

In the early phase of testing, the defect detection rate often shows an erratic behavior. The delayed S-shaped model is robust in this situation and is now widely used by the Japanese. Recent investigations suggests that the logarithmic model has much better overall predictive capability; however, it is often unstable at the beginning. Here a technique is presented that makes the logarithmic model much more robust in the early phases.

1 INTRODUCTION

Managing the reliability of software is always a critical part of software development. Using a reliability model a manager can project the number of bugs that need to removed and the resources (time and personnel) that are needed to obtain the target reliability level. The capability to make good projections sufficiently early can allow a manager to allocate the resources efficiently in order to meet the release date. Unlike many other statistical applications, the test data recorded contains a large amount of noise. This often means that the fit of a model, as measured by traditional statistical measures (like R^2) [Camp87]is less perfect than desired. The application of a Software Reliability Model (SRM) is specially problematic in the early phases of testing. Not only is there considerable noise, often the software does not show clear signs of a reliability growth for a while. The reliability growth, as reflected by a declining fault detection rate (or failure intensity λ), often is not apparent until the software is relatively stable and defect clusters have been removed. Most SRMs like the exponential and the logarithmic models assume a reliability growth [Musa87], [MS90]. They tend to break down during the early phases, yielding negative values for parameters which are expected to be positive. If used for making projections, they will generate values which are clearly impossible. The delayed S-shaped model, proposed by Yamada et al. [YO85], assumes that the curve for λ against time (Figure 1) initially rises and then passes through a slope of 0 at the peak. After that λ declines indicating reliability growth. It thus fits better in situations where there is no reliability growth in the beginning. It exhibits a stable behavior at the beginning, thus making it suitable for early planning. It has become specially popular in Japan.

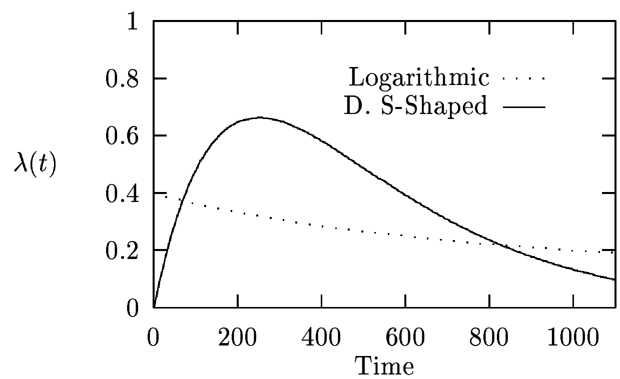

Figure 1: The failure intensity for the two models

We have recently evaluated predictability of some of the popular SRMs using data sets from several different sources, including Japan. Using a predictability measure, termed Average Error (AE), we found that the logarithmic models work best in a majority of cases [MKV91]. In other cases also it performs satisfactorily. The delayed S-shaped model performed poorly for

*This work was partly supported by an SDIO/IST project monitored by ONR

most data sets. Other researchers appear to have obtained similar results. This is in spite of the fact that the delayed S-shaped model often provides a better fit. These results suggest that the logarithmic model generally describes the overall defect detection process more accurately. As discussed in the next section, the logarithmic model assumes a decline in fault detection rate λ. When the reliability growth is not occurring, the application of the logarithmic model often yields negative value for one of the parameters. Thus no projections can be made until λ starts declining. Here an approach is presented which will handle these "bad" data points.

2 The Major Models: Formulation

Table 1 below summarizes the basic features of the three major two-parameter models. They are frequently stated in terms of two parameters, β_0 and β_1, as shown. For all of them $\beta 1$ is the time scale-factor. The maximum value of λ, denoted below by λ_m, is $\beta_0\beta_1$, for all models. For the exponential and the delayed S-shaped models, β_0 is the total number of faults that will be eventually detected. Since the logarithmic model is an infinite-faults model, β_0 has no such significance and simply serves as a fault-count scale-factor.

One good way to visually view data is to look at m-plots (Figure 2). A plot of $1/\lambda$ (i.e. MTTF) against time will be linear if the behavior is logarithmic as shown in the last column. The exponential and the delayed S-shaped models will have an exponential trend, the latter has a minimum at t=$1/\beta_1$.

In some situations, it is useful to use $\lambda_m(\beta_0\beta_1)$ as a parameter instead of β_0. Our analysis with actual data from many projects shows that λ_m is somewhat more robust compared with β_0. Table 2 illustrates this for the parameters of the logarithmic model.. In addition λ_m is meaningful for all the three models, whereas β_0 lacks a physical meaning for the logarithmic model. Recasting the models in terms of parameters λ_m and β_1, we have these expressions for $\lambda(t)$:

$$Exponential: \lambda(t) = \lambda_m e^{-\beta_1 t} \quad (1)$$

$$Logarithmic: \lambda(t) = \frac{\lambda_m}{1+\beta_1 t} \quad (2)$$

$$Delayed\ S-Shaped: \lambda(t) = \lambda_m \beta_1 t e^{-\beta_1 t} \quad (3)$$

Model	$\lambda(t)$	Maximum λ	$1/\lambda$ Trend
Expo.	$\beta_0\beta_1 e^{-\beta_1 t}$	$\lambda(0) = \beta_0\beta_1$	Exponen.
Log.	$\beta_0\beta_1/(1+\beta_1 t)$	$\lambda(0) = \beta_0\beta_1$	Linear
S-sh.	$\beta_0\beta_1^2 t e^{-\beta_1 t}$	$\lambda(1/\beta_1) = \beta_0\beta_1$	Exponen.

Table 1: Major 2-parameter models

Time	β_0	β_1	λ_m
571	-12.63	-3.48	43.94
968	50.50	1.29	65.09
1986	8.68	25.11	217.95
3098	12.50	7.74	96.68
5049	11.61	9.65	112.09
5324	36.87	1.03	37.80
6380	43.38	0.83	35.88
7644	44.73	0.79	35.48
10089	29.11	1.61	46.93
10982	46.76	0.71	33.21
12559	50.89	0.63	31.81
13486	72.12	0.38	27.49
15806	57.62	0.53	30.32
17458	64.28	0.44	28.59
19556	63.81	0.45	28.70
24127	42.00	1.03	43.33
24493	67.82	0.39	26.71
36799	23.98	-3.13	-75.05
40580	31.58	3.73	117.76
42296	44.91	0.75	33.46
49171	41.33	0.96	39.79
52875	49.16	0.57	28.14
56463	57.26	0.40	22.83
62651	56.64	0.41	23.14
71043	52.25	0.50	26.28
82702	45.43	0.82	37.35
Median	45.43	0.75	35.48

Table 2: Estimated parameters for data set P1 [MKV91]

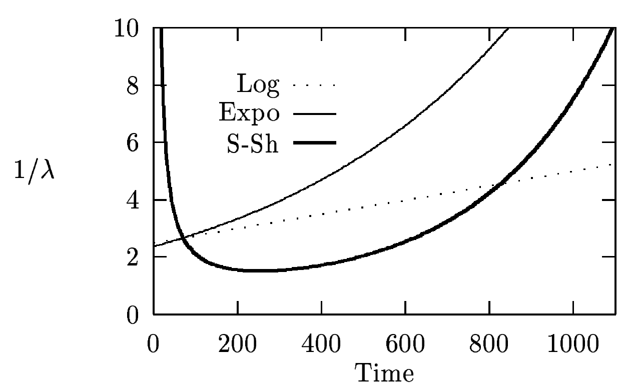

Figure 2: The m-plots for different models

3 Problems in Parameter Estimation

The three expressions above can be used directly to estimate parameters using non-linear regression. Fortunately, in all three cases it is possible to linearize the expression using suitable coordinate transformation. This allows use of linear regression which is computationally much more efficient. The equation (2) may be restated as:

$$1/\lambda = \frac{1}{\lambda_m} + \frac{\beta_1}{\lambda_m} t \tag{4}$$

Since by definition, both β_1 and λ_m are required to be positive, both the slope and intercept in an m-plot are expected to be positive. As discussed above, the fault detection rate often tends to be erratic at the beginning of testing. If only a few initial data points are available, the best least square fit using equation (4) may yield a line with a negative slope. This can render application of the logarithmic model invalid for some of the initial data points. For example if one uses the first four points in Figure 3, the slope of the trend would be negative. One possible approach can be to look for a least square fit with the constraint that the slope must be positive. This can, however, be computationally complex. In addition, this slope can vary widely because only a few points may be available for fitting.

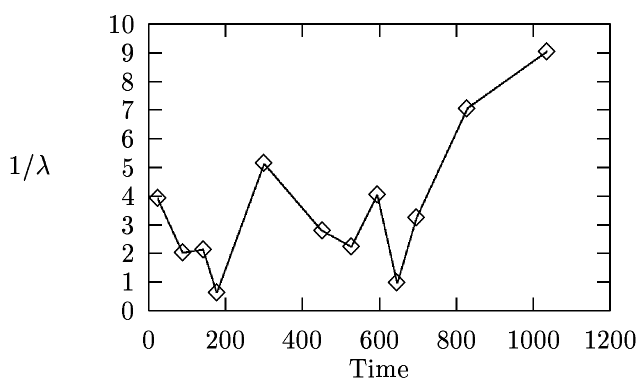

Figure 3: Points on the m-plot for data set 3.2 [MKV91]

Here, a new approach is suggested to handle the problem in parameter estimation. It is based on the view that for a small period in the early part of testing, the reliability growth is small and the variations in $1/\lambda$ are merely noise superimposed on the constant $1/\lambda_m$. A possible approach is to assume that β_1 is equal to some previously assumed low value. Thus, the parameter estimation should use these rules:

1. If the use of equation (4) results in negative slope, the following correction should be made:
   (a) The constant (intercept) is replaced by the average value.
   (b) A predetermined value β_{1L} is used for β_1. It must be a low estimate so that it is not optimistic in the long term. Such a correction generally would be needed only during the early phases of testing.
2. No correction is needed if the parameter values are positive.

The value of β_{1L} does not have to be estimated accurately because it can be shown that for low values of t the behavior is relatively insensitive with respect to β_1. Consider the expression for λ and μ,

$$\lambda = \frac{\lambda_m}{1+\beta_1 t} \tag{5}$$

$$= \lambda_m (1+\beta_1 t)^{-1} \tag{6}$$

$$= \lambda_m (1 - \beta_1 t + \ldots)\ for\ \beta_1 t < 1 \tag{7}$$

$$\approx \lambda_m\ for\ \beta_1 t << 1 \tag{8}$$

$$\mu = \frac{\lambda_m}{\beta_1} ln(1+\beta_1 t) \tag{9}$$

$$= \frac{\lambda_m}{\beta_1}(\beta_1 t - \frac{\beta_1^2 t^2}{2} + \ldots)\ for\ \beta_1 t < 1 \tag{10}$$

$$\approx \lambda_m t\ for\ \beta_1 t << 1 \tag{11}$$

One can estimate β_{1L} a-priori using the approach below. When a satisfactory value of β_1 is found using test data, the a-priori value is disregarded.

Table 3 illustrates the use of the above rules. The value for β_{1L} was arbitrarily chosen to be equal to 4.10^{-5}.

Data Set		Additional Faults				
Time	Faults	Actual	Log	Expo	D S-sh	Log adj
0	0	328				
58.8	15	313				
117.6	44	284	Failure	1792515	166	313
164.6	66	262	Failure	13382	102	325
188.2	103	225	Failure	835816	1969	383
410.4	146	182	140	226	48	140
491.3	175	153	155	170	53	155
560.4	206	122	178	182	70	178
629.5	223	105	120	118	54	120
661.1	255	73	154	183	94	142
729.4	276	52	122	136	76	122
926.6	304	24	47	55	33	47
1143.6	328	0	0	0	0	0

Table 3: Comparisons with corrected logarithmic model for data set 3.2 [MKV91]

4 Estimation of β_{1L}

While at this time accurate estimation of β_1 is not possible, an approximate β_{1L} for corrective use can be found. For the exponential model, the parameters can be empirically estimated using [MSKS90]:

$$\beta_0^E = D \cdot S \tag{12}$$
$$\beta_1^E = K \frac{r}{S \cdot Q} \tag{13}$$

where S is the number of source instructions, r is the object-instruction execution rate (for the computer being used) and Q is average number of object instructions per source instruction (often in the vicinity of 4). D is the defect density, which varies between 20 to 3 per KLOC at the beginning of the system test phase. The fault exposure ratio K has been found to vary from 1.4x10-7 to 10.6x10-7. It has been suggested [MSKS90] that the logarithmic model parameters β_0, β_1 may be related with the exponential model parameters β_0^E, β_1^E in the following way.

$$\beta_0 = \frac{1}{a}\beta_0^E \tag{14}$$
$$\beta_1 = a\beta_1^E \tag{15}$$

where a is in the vicinity of 5.

Here our objective is to estimate λ_m and β_1. In the beginning of testing, λ_m is estimated simply by taking the average of the values. The value of β_1 is then estimated by using:

$$\beta_1 = \frac{\lambda_m}{\beta_0} = \frac{a\lambda_m}{\beta_0^E} = \frac{a\lambda_m}{DS} \tag{16}$$

The size S of the code is known. The initial defect density D may be estimated using past experience for the same design team. For example, if we assume that the initial defect density is less than 20/KLOC, and if we take a=5, a low estimate for β_1 would be:

$$\beta_{1L} = 0.25\frac{\lambda_m}{S} \tag{17}$$

An estimate of β_1 using equation (18) uses both static and dynamic data. It can also be estimated using only static information using equations (15) and (16). The choice may depend on whether D or K can be estimated more accurately. Satisfactory methods for empirically estimating D have been proposed [TK85] whereas no methods for estimating K exist at this time.

5 Concluding Remarks

In the beginning of testing, sometimes the failure intensity varies erratically. In contrast with the delayed S-shaped model, here the view taken is that the noise essentially masks the trend in the beginning. Thus initially the dynamic data should not be used to evaluate the trend. It is suggested that until a definite trend is apparent, an a-priori value of β_1 should be used.

References

[Camp87] S.K. Campbell, Applied Business Statistics, Harper & Row, New York, 1987, pp. 622-684.

[Musa87] J.D. Musa, A. Iannino and K. Okumoto, Software Reliability: Measurement, Prediction, Application, McGraw-Hill, New York, 1987.

[MS90] Y.K. Malaiya, P.K. Srimani, Editors, Software Reliability Models, Theoretical Developments, Evaluation and Applications, IEEE Computer Society Press, 1990.

[YO85] S. Yamada and S. Osaki, "Software Reliability Growth Modeling: Models and Applications," IEEE Trans. Software Engineering, Dec. 1985, pp. 1431-1437.

[MKV91] Y.K. Malaiya, N. Karunanithi and P. Verma, "Predictability of Software Reliability Models," Technical Report, 1991, 18 pages.

[MSKS90] Y.K. Malaiya, S. Sur, N. Karunanithi and Y.C. Sun, "Implementation Considerations for Software Reliability," Proc. 8th Annual IEEE Software Reliability Symp., June 1990, pp. 6.21-6.30.

[TK85] M. Takahashi and Y. Kamayachi, "An Empirical Study of a Model for Program Error Prediction," Proc. 8th Int. IEEE Conf. on Software Engineering, August 1985, pp. 330Œ-336.