Regression Goodness-Of-Fit Test for Software Reliability Model Validation

M. Xie[a], B. Yang[a] and O. Gaudoin[b]
[a]National University of Singapore, Singapore
[b]Institut National Polytechnique de Grenoble, France

1. Introduction

Software reliability growth models based on nonhomogenerous Poisson process (NHPP) seems to be most commonly used because of their simplicity. These models are commonly used in software reliability engineering practice and most of the decision-makings such as the optimal software release time determination, the optimal testing-resource allocation, etc. are based on the results obtained from the analysis of selected model.

However a problem is the model validation and selection. If the selected model does not fit the collected software testing data relatively well, we would expect a low prediction ability of this model and the decision-makings based on the analysis of this model would be far from what is considered to be optimal decision. This article presents a simple method for model validation.

2. Model with graphical interpretation

An important class of NHPP models is the model with simple graphical interpretations [1,2]. For this type of model, the model validation can be done by a simple graphical approach. The Duane model, or the power-law process model has the following mean value function:

$$m(t) = at^b, a > 0,\ b > 0,\ t \geq 0 .$$

Taking logarithm on both sides of the above, we have that

$$\ln m(t) = \ln a + b \ln t .$$

It can be seen that if we plot the cumulative number of observed failures versus testing-time on a log-log scale, then the plotted points tends to be on a straight line.

Another model, which is shown to be better for software failure data [1], is the log-power model with the following mean value function:

$$m(t) = a \ln^b (t+1), a > 0,\ b > 0,\ t \geq 0 .$$

Taking the logarithm on both sides of this, we have that

$$\ln m(t) = \ln a + b \ln \ln(t+1) .$$

Obviously, if we plot the cumulative number of observed failures versus testing-time plus 1 on a log-double-log paper, then the plotted points should be fitted by a straight line fairly well. Other models have been presented in [3].

3. Model validation based on R^2-value

It is important to make an objective judgment with regard to its ability in fitting a data set in practice. After the plotted points are fitted with a straight line using linear regression analysis approach, the coefficient of determination, R^2, can be calculated. The R^2 value obtained can be used as a goodness-of-fit (GOF) measurement. On the other hand, it is more useful if a formal statistical GOF test based on this value can be conducted [6].

Suppose n failure times, $(T_1, \cdots, T_n)$, have been observed, then R^2 can be computed with regression analysis for specific model. If a formal goodness-of-fit test is to be established based on this R^2 value, then the test should have the following form:

> Reject the log-power model at a significance level of d if $R^2 < R_d^2$.

The above GOF test exists only if the distribution of the R^2 value is independent of model parameters, a and b. In the following section we prove that the distribution of R^2

actually does not depend on the true model parameters, thus a formal GOF test can be conducted.

Table 1. Critical values of *R*-squared test

n	10%	5%	1%
10	0.847	0.806	0.725
20	0.885	0.850	0.757
30	0.905	0.872	0.786
40	0.917	0.887	0.804
50	0.926	0.899	0.822
60	0.933	0.907	0.837
70	0.939	0.914	0.849
80	0.943	0.920	0.858
90	0.947	0.925	0.866
100	0.950	0.930	0.874

Besides the R^2 test as discussed above, there are also other tests available. In fact, it is known that any test for the power process model can be applied for the log-power model, provided that the change of variable is done, i.e., replacing T_i by $\ln(T_i + 1)$. Some well-known tests are [4-5] Kolmogorov-Smirnov test, Cramer-von Mises test and Anderson-Darling test. Extensive simulation has been carried out to compare the performance of the R^2 test, and in most cases, this tests are very comparable despite of its simplicity.

4. An example

To illustrate the simplicity of the approach, as an example, take the data in [3], the plot is made and shown in Figure 1:

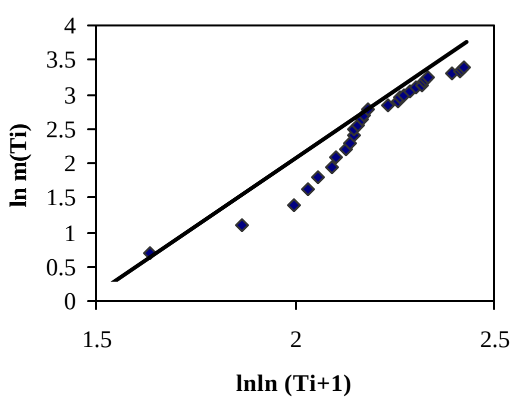

Figure 1 Plot of a data set in [3].

The calculated *R*-squared value is 0.958. Since it can be obtained from Table 1 that the critical value at significance level of 5% is 0.872, we cannot reject the LP model.

5. Discussions

The approach in this paper is interesting as it provides a statistical foundation for such a decision-making based on the simple measure, R^2. By comparing the R^2 value with the critical value, judgment can be made at a fixed confidence level. In fact, an important advantage of the logpower model is its graphical interpretation and simplicity. It is shown in [1] that this model also has closed form maximum likelihood estimates. Our study on simple test provides yet another important strength of this model.

References

[1] Xie, M. and Zhao, M. (1992) 'On the log-power software reliability model', *Proc. of the 3rd Int. Symp. on Software Reliability Engineering*, North Carolina, USA, IEEE Computer Society Press, 7-10 October, North Carolina, USA, pp.14-22.

[2] Donovan, J. and Murphy, E. (1999). Reliability growth - A new graphical model. *Quality and Reliability Engineering International*, **15**, 167-174.

[3] Xie, M. and Zhao, M. (1993) 'On some reliability growth models with graphical interpretations', *Microelectronics and Reliability*, **33**, pp. 149-167.

[4] Rigdon, S.E. (1989) 'Testing goodness-of-fit for the power law process', *Communications in Statistics: Theory and Methods*, **18**, pp. 4665-4676.

[5] Park, W.J. and Kim, Y.G. (1992) 'Goodness-of-fit tests for the power-law process', *IEEE Transactions on Reliability*, **41**, pp. 107-111.

[6] Gaudoin, O.; Yang, B. and M. Xie (2000) 'A Simple Goodness-of-fit Test for the Power-Law Process Based on the Duane Plot', *Research Report #25/2000*, Dept of Industrial Engineering, National University of Singapore.