Optimal Discrete Stopping Times For Reliability Growth Tests

John Quigley

Department of Management Science

University of Strathclyde

40 George Street

Glasgow UK

G1 1QE

(tel):   +44 141 548 3152

(fax):   +44 141 552 6686

e-mail: j.quigley@strath.ac.uk

Abstract

Often, the duration of a reliability growth development test is specified in advance and the decision to terminate or continue testing is conducted at discrete time intervals. These features are normally not captured by reliability growth models. This paper adapts a standard reliability growth model to determine the optimal time for which to plan to terminate testing. The underlying stochastic process is developed from an Order Statistic argument with Bayesian inference used to estimate the number of faults within the design and classical inference procedures used to assess the rate of fault detection. Inference procedures within this framework are explored where it is shown the Maximum Likelihood Estimators possess a small bias and converges to the Minimum Variance Unbiased Estimator after few tests for designs with moderate number of faults. It is shown that the Likelihood function can be bimodal when there is conflict between the observed rate of fault detection and the prior distribution describing the number of faults in the design. An illustrative example is provided.

Key Words: Discrete Reliability Growth, NHPP, Order Statistics, Cost Model

1.   Introduction

The aim of a reliability growth program is to improve the reliability of an item through the identification of design weaknesses and subsequent re-design. Typically, this is achieved through Test Analyze And Fix (TAAF) programs, whereby an item is placed on test for a period of time to expose weaknesses. Through such testing, faults are exposed, corrective actions implemented and the reliability of the item is improved[1,2]. The data generated from such a program will be periodically analyzed to determine whether the test should continue or terminate.

Reliability growth tests are costly in terms of time and resources required, which results in much scepticism as to their benefits in comparison to their costs. Despite this, due to the risk adverse approach to reliability, growth testing continues to be widely used (see Hobbs[2]). Many reliability growth models exist to support decision-making, see Jewell[3], Xie[4], Ansell et al[5] for a review and critical appraisal. These models aim to provide information about the effectiveness of testing for improving reliability by measuring changes in the observed rate of failure. Inference procedures for these models that follow a strictly classical statistical paradigm require more data than is often available, while Bayesian approaches often require prior distributions on parameters that are too abstract to be meaningful to practitioners[6].

The Modified IBM Model[7] (IEC 61164) overcomes the aforementioned criticisms through combining engineering judgement about the inherent concerns with the design to be tested with observed failures on test. It makes use of a register of potential design weaknesses to formulate a prior distribution describing the number of faults within the design. Processes used to capture such beliefs include FMEA[8], HAZOP[9] and elicitation of engineering judgment[6]. The model is a continuous time reliability growth model, where the rate of fault detection is estimated classically. The model was developed with the primary aim of supporting inference on failure

mode detection. Notwithstanding, inference regarding the number of faults within a design will support in-service operational performance predictions.

In this paper, we are concerned with the situation where test reviews are conducted periodically, say weekly or monthly, to decide whether testing should continue or be terminated. Moreover the test data used to support the decision is a record of the number of faults detected between each review. There exist a number of discrete reliability growth models that could describe this situation, however, many of the classical models have been criticized for not having rigorous statistical inference procedures and the Bayesian models have been criticized for not making use of specific knowledge of faults[10].

We consider discretizing the Modified IBM model. This results in a model that is equivalent to the Generalized Binomial Software Reliability Growth Model when the number of initial faults within the design are described with a Poisson prior distribution as developed by Dohi et al[11]. In their paper, the authors evaluated the estimation properties of the Maximum Likelihood Estimators (MLE) of both the mean of the prior distribution and the probability of fault detection parameter. Their evaluation was based on an empirical study and showed that the MLE was not always the best procedure and that the quality of the estimator did not always improve with more data. We consider the mean of the prior to be provided by a calibrated expert and show through a simulation study that the MLE converges quickly to the Minimum Variance Unbiased Estimator (MVUE) for situations were there is moderate number of faults. Moreover we show that that Likelihood function can be temporarily bimodal, which can explain the deterioration of the quality of the MLE observed in the work by Dohi et al[11].

Section two describes how the modified IBM model was adapted from a continuous reliability growth model to a discrete one. Moreover, the likelihood function is constructed where point and interval estimate procedures are derived.

Section three describes an investigation into the quality of the Maximum Likelihood Estimate (MLE) for this model for bias and variability. Section four embeds the model with a cost model to assess the optimal number of tests for which to plan. Section five is an illustrative example of using the model to plan for an optimal number of tests and how to update this in light of data generated from the tests.

2. Description of Discrete Modified IBM RG Model Assumptions

2.1 Discussion of Modeling Assumptions

It is assumed that an item starts the development program test with an unknown number of faults, say N. The accumulated time until faults can be detected is assumed independent and identically distributed. Moreover, it is assumed that the time to realize a particular fault is exponentially distributed. These modeling assumptions are consistent with earlier work such as (Rosner[12], Jelenski and Moranda[13], Goel and Okumoto[14]).

At fixed intervals, known apriori, the item on test is removed and investigated for faults. We will refer to each interval as a test and as such we consider the development program to consist of a sequence of identical tests. Exact time of fault realization is not recorded, only that it was realized within the interval between the inspections. Similar testing regimes are considered within Crow[15] and Robinson and Dietrich[16,17], with different fault realization assumptions. We denote the conditional probability that a fault is detected during a particular test given it existed within the design at the start of the test with p. Relating this to the assumption that the time to

realize a fault is Exponentially distributed with hazard rate, say μ, and assuming the length of a test is t then $p = 1 - e^{-\mu t}$.

Corrective actions are made and re-designs are assumed not to introduce other faults. Much information is often gained through observing the manner in which a fault has been realized. While it is optimistic to assume that the engineer has learned enough about the item to remove the fault without introducing additional faults, if it were felt that this assumption was unduly optimistic then this is easily dealt with through re-visiting the elicitation processes for the prior distribution, as discussed in Walls and Quigley[6].

Finally, the prior distribution describing the number of faults that exist within the design is assumed to be a Poisson distribution with mean λ. This is consistent with most reliability growth models (e.g. (Cozzolino[18], Crow[15], Jewell[3], Robinson and Dietrich[16,17], Ebrahimi[19], Calabrai et al[20]) as mathematically this is equivalent to assuming a Non-Homogeneous Poisson Process (NHPP).

Quigley and Walls[21] showed this modeling framework to be more appropriate than the Power Law framework proposed under IEC 1164[22], for situations where prior knowledge of the number of faults is available.

2.2   MLE for Probability of Detecting a Fault

We consider a situation where an item has been exposed to a series of j tests. The item possesses an unknown number of faults, denoted by *N*. We denote the number of faults detected on the ith test with n_i. The number of defects prior to test, i.e. *N*, is an unknown constant whose epistemic uncertainty is measured with a Poisson distribution with mean λ. The realizations of faults on a particular test assuming they exist are independently and identically distributed with probability *p*. Finally, it is assumed that the corrective actions do not introduce any new faults and completely

5

remove the fault that provoked the action. The likelihood function for this situation is presented in Eq. (1).

$$L\left(p\mid N, \underset{\sim}{n}\right)=\left(\begin{array}{c} N \\ \left(N-\sum_{i=1}^{j} n_i\right), n_1,\ldots,n_j \end{array}\right)\left(1-p\right)^{\left(N-\sum_{i=1}^{j} n_i\right) j}\prod_{i=1}^{j} p^{n_i}\left(1-p\right)^{n_i(i-1)}$$

$where:$
$$\underset{\sim}{n}=\left(n_1, n_2,\ldots,n_j\right)$$

(1)

Eq. (1) is the product of the likelihood of the faults that have not been exposed, i.e. $N-\sum_{i=1}^{j} n_i$, avoiding exposure on j consecutive tests and the likelihood of the faults that were exposed being realized during specific tests. This likelihood is conditional on the initial number of faults with the design, i.e. N. The likelihood in Eq. (2) is obtained through evaluating the expectation of Eq. (1) with respect to N; the distribution for N is a Poisson distribution conditioned on being at least equal to the number of faults detected on the j completed tests.

We introduce the following notation for simplicity.

$$n_T=\sum_{i=1}^{j} n_i$$
$$n_e=\sum_{i=1}^{j} i\times n_i$$

The rationale behind the notation is to denote the total number of faults detected, i.e. n_T, and the total exposure to test conditions experienced by those faults that have been realized, i.e. n_e.

$$L\left(p\middle|\underset{\sim}{n}\right)=E\left[\frac{N(0)!}{\left(N(0)-n_T\right)!\prod_{i=1}^{j}n_i!}(1-p)^{(N(0)-n_T)j}\prod_{i=1}^{j}p^{n_i}(1-p)^{n_i(i-1)}\right]$$

$$=\frac{p^{n_T}(1-p)^{n_T}}{(1-p)^{(j+1)n_T}\prod_{i=1}^{j}n_i!}\sum_{n=n_T}^{\infty}\frac{n!}{(n-n_T)!}(1-p)^{jn}\frac{\dfrac{\lambda^n e^{-\lambda}}{n!}}{1-\sum_{k=0}^{n_T}\dfrac{\lambda^k e^{-\lambda}}{k!}}$$

Re-expressing *(n-n_T)* with *m* we obtain the following.

$$L\left(p\middle|\underset{\sim}{n}\right)=\frac{\lambda^{n_T}e^{-\lambda}p^{n_T}(1-p)^{jn_T+n_e}}{\left(1-\sum_{k=0}^{n_T}\dfrac{\lambda^k e^{-\lambda}}{k!}\right)(1-p)^{(j+1)n_T}\prod_{i=1}^{j}n_i!}\sum_{m=0}^{\infty}\frac{\left[\lambda(1-p)^j\right]^m}{m!}$$

$$=\frac{\lambda^{n_T}e^{-\lambda\left[1-(1-p)^j\right]}p^{n_T}(1-p)^{n_e-n_T}}{\left(1-\sum_{k=0}^{n_T}\dfrac{\lambda^k e^{-\lambda}}{k!}\right)\prod_{i=1}^{j}n_i!}$$

(2)

The Score function Eq. (3) is derived through differentiating the logarithm of Eq. (2) with respect to *p*.

$$S\left(p\middle|\underset{\sim}{n}\right)=-\lambda j(1-p)^{j-1}+\frac{n_T}{p}-\frac{n_e-n_T}{1-p}$$

(3)

Solving Eq. (3) for *p* provides the Maximum Likelihood Estimator (MLE) $\hat{p}$.

A closed form solution does not exist for $\hat{p}$ however, through re-arranging the Score function we can obtain the following equation.

$$\hat{p} = \frac{n_T}{j\lambda\left(1-\hat{p}\right)^j + n_e} \qquad (4)$$

The expression for the MLE in Eq. (4) has an intuitive explanation. The numerator is the total number of faults detected in j tests and the numerator is a measure of the expected exposure of faults to test conditions experienced by the end of the jth test, where n_e is the exposure from those faults that have been detected and $j\lambda\left(1-\hat{p}\right)^j$ is the expected contribution from those faults not yet exposed. Note that the limit as the number of tests conducted approaches infinity is the ratio of the number of faults detected to the exposure of faults to tests.

2.3   Cramer Rao Lower Bound

The Information function for this model is obtained from taking the expectation with respect to the data of the negative of the derivative of the Score function.

$$I\left(p|\underset{\sim}{n}\right) = \begin{cases} E_{\underset{\sim}{n}}\left[\dfrac{n_1}{p^2}\right] & , j = 1 \\[4ex] E_{\underset{\sim}{n}}\left[\dfrac{n_T}{p^2} + \dfrac{n_e - n_T}{\left(1-p\right)^2} - \lambda j\left(j-1\right)\left(1-p\right)^{j-2}\right] & , j \geq 2 \end{cases} \qquad (5)$$

We note the following two expectations that will be used to calculate the expectation of Eq. (5).

$$E\left[n_T\right] = \lambda\left[1 - \left(1-p\right)^j\right]$$

$$E\left[n_e\right] = \lambda\left[1 - \left(j+1\right)\left(1-p\right)^j + \left(1-p\right)\frac{1 - \left(1-p\right)^j}{p}\right] \qquad (6)$$

The expectation of Eq. (5) with respect to $\underset{\sim}{n}$ is:

$$I(p) = \begin{cases} \dfrac{\lambda}{p}, & j=1 \\ \lambda \left[\dfrac{1}{(1-p)p^2} - \left(j^2 + \dfrac{1-p}{p^2} \right)(1-p)^{j-2} \right], & j \geq 2 \end{cases} \tag{7}$$

Therefore, by the Cramer Rao Lower Bound[23] the variance of an unbiased estimator

for p must have variance at least as large as the inverse of Eq. (7).

2.4 *Bimodality of Likelihood*

The MLE for q can be solved through the following fixed-point iteration.

$$g(p) = \frac{n_T}{j\lambda(1-p)^j + n_e} \tag{8}$$

The MLE of p will be the convergence of the following successive iteration:

$$p_q = g(p_{q-1})$$

This fixed-point iteration will converge to a unique solution if the following condition

is met (Burden and Faires[24]). As

$$p \in [0,1]$$
$$g(p) \in C\left[\frac{n_T}{n_e}, \frac{n_T}{j\lambda + n_e} \right] \tag{9}$$

For the iteration to be guaranteed to converge to a unique solution we require:

$$\left| \frac{dg(p)}{dp} \right| = \frac{\lambda j^2 (1-p)^{j-1} n_T}{\left[n_e + \lambda j \ (1-p)^j \right]^2} \leq 1 \tag{10}$$

This implies:

$$n_e^2 + \lambda (1-p)^{j-1} j \left(2n_e \ (1-p) - jn_T \right) + \lambda^2 \left[j \ (1-p)^j \right]^2 \geq 0 \tag{11}$$

Investigating the roots of this quadratic equation with respect to λ, the root of Eq. (11)

with respect to λ is:

$$\lambda_{root} = \frac{-\left(2n_e(1-p) - jn_T\right) \pm \sqrt{jn_T\left(jn_T - 4n_e(1-p)\right)}}{2j(1-p)^{j+1}} \qquad (12)$$

A necessary condition for a real solution to the radical in this equation is:

$$jn_T \geq 4n_e \qquad (13)$$

As λ must be a positive number and both the constant and quadratic term of Eq. (11) are positive it follows that if there are no real roots to Eq. (11) then there will always be a unique solution. Re-arranging Eq. (13) we require d to be less than 0 to ensure convergence to a unique solution.

$$d = jn_T - 4n_e \qquad (14)$$

Therefore, convergence to a unique solution must occur during the first four tests. The expectation of Eq. (14) is obtained through the following.

$$E[D] = j\lambda\left[1 - (1-p)^j\right] - 4\lambda\left[1 - (j+1)(1-p)^j + (1-p)\frac{1 - (1-p)^j}{p}\right]$$

$$= \lambda\left\{\left(j - \frac{4}{p}\right) + \left(3j + \frac{4}{p}\right)(1-p)^j\right\}$$

Figure 1 illustrates the locus of points for p and j where the expectation of Eq. (14) is less than 0. We see that there is an inverse relation between the number of tests and the efficacy of the testing, if p is large then is requires few tests to leave the region where a unique solution is guaranteed. It is worth noting that for the region where a unique solution is not guaranteed does not necessarily mean the Likelihood function is bimodal. Moreover, as the number of tests, i.e. j, approaches infinity then condition Eq. (11) is met, so the condition of bimodality will be transient if it exists at all.

INSERT FIGURE 1

2.5 *Confidence Intervals for p*

We present three methods for obtaining confidence intervals for p. Firstly, conditioning on the total number of faults that have been detected only. This

approach makes no use of the observed rate of occurrence and as such will be conservative but the simplest to construct. Secondly, confidence intervals will be constructed through parametric Bootstrapping. This approach will be the most accurate but requires re-sampling from a Poisson process. Thirdly, the intervals will be constructed through the likelihood ratio statistic. This approach works for moderate size data sets and will require the solution to a non-linear implicit function.

Method 1   Conservative

We are presuming that λ is provided by expert assessment and as such the only parameter to be estimated with the data is p. We have assumed that faults are realized according to a Poisson process. After j consecutive tests the total number of faults detected, i.e. $\sum_{i=1}^{j} n_i$, has a Poisson distribution with mean $\lambda\left(1-\left(1-p\right)^{j}\right)$.

Using the formula for a 100(1-α)% Confidence Interval for the mean of a Poisson distribution (see Johnson et al[25]) we have the following.

$$\lambda\left(1-\left(1-p_L\right)^{j}\right)=0.5\chi^2_{2n_t,\frac{\alpha}{2}}$$

$$p_L=1-\sqrt[j]{\max\left(1-\frac{0.5\chi^2_{2n_t,\frac{\alpha}{2}}}{\lambda},1\right)}$$

$$\lambda\left(1-\left(1-p_U\right)^{j}\right)=0.5\chi^2_{2n_t,\frac{\alpha}{2}}$$

$$p_U=1-\sqrt[j]{\max\left(1-\frac{0.5\chi^2_{2(n_t+1),1-\frac{\alpha}{2}}}{\lambda},1\right)} \qquad (15)$$

Method 2   Parametric Bootstrap

The principle underpinning a parametric Bootstrap Confidence[26] interval is to estimate the parameters of the stochastic model using the data available and then use the model to simulate data sets, each of which is used to re-estimate the model parameters. The variability observed in the re-estimated parameters is used to assess the level of confidence in the original MLE's.

We are assuming that the time until a fault is realized is Exponentially distributed. However, we are modeling the test in which the faults are realized and as such this has a Geometric distribution. Moreover, we are assuming that the realization of faults occurs independently of each other. Furthermore, we are assuming that the number of faults within the design is a Poisson random variable with mean λ. Therefore, the number of faults detected within any test, say for example the ith test, has a Poisson distribution with mean $\lambda (1-p)^{i-1} p$.

The following algorithm can be used for constructing parametric Bootstrap confidence intervals for this model when there have been j tests conducted.

1   Estimate $\hat{p}$ from Eq. (4)

2   For k from 1 to runs do

3   For i from 1 to j do

4   Simulate N_{ki} from Poisson distribution with mean $\lambda \left(1-\hat{p} \right)^{i-1} \hat{p}$

5   Next i

6   Using N_{ki} (for i from 1 to j) estimate $\hat{p}_k$ from Eq. (4)

7   Next j

The lower and upper values of $\hat{p}_k$ (for k from 1 to runs) will provide the parametric Bootstrap values. The number of runs required will vary depending on the variability within the model which is greater for larger values of λ.

Method 3   Likelihood Ratio

Likelihood ratio confidence are useful for medium size sample sizes (see Lawless[27]) and are derived from the fact that -2 times the natural logarithm of the relative likelihood function converges quickly to a χ^2 distribution as described in Eq. (16).

$$-2\left[\lambda\left[\left(1-p\right)^j-\left(1-\hat{p}\right)^j\right]+n_T\ln\left(\frac{p}{\hat{p}}\right)+\left(n_e-n_T\right)\ln\left(\frac{1-p}{1-\hat{p}}\right)\right]\sim\chi^2_{(1)} \qquad (16)$$

This expression can be solved for required confidence level to determine appropriate interval for *p*.

2.6   Bayesian Updating of Number of Faults Remaining Undetected

The prior distribution on the number of faults updates to a Poisson posterior distribution as the following demonstrates.

$$\pi\left(N(0)=d\big|\underset{\sim}{n}\right)=\frac{L\left(p\big|N(0),\underset{\sim}{n}\right)\pi\left(N(0)=d\right)}{\sum\limits_{\forall d}L\left(p\big|\underset{\sim}{n}\right)\pi\left(N(0)=d\right)}$$

$$=\frac{\dfrac{n!}{\left(n-n_T\right)!\prod\limits_{i=1}^{j}n_i!}\left(1-p\right)^{(n-n_T)j}\left[\prod\limits_{i=1}^{j}p^{n_i}\left(1-p\right)^{n_i(i-1)}\right]\dfrac{\lambda^n e^{-\lambda}}{n!}}{\sum\limits_{n=n_T}^{\infty}\dfrac{n!}{\left(n-n_T\right)!\prod\limits_{i=1}^{j}n_i!}\left(1-p\right)^{(n-n_T)j}\left[\prod\limits_{i=1}^{j}p^{n_i}\left(1-p\right)^{n_i(i-1)}\right]\dfrac{\lambda^n e^{-\lambda}}{d!}}$$

Changing the notation where *R* denotes the remaining number of faults we have the following.

$$\pi\left(R=r\middle|\underset{\sim}{n}\right)=\frac{\dfrac{(r+n_T)!}{r!\prod\limits_{i=1}^{j}n_i!}(1-p)^{rj}\,p^{n_T}\,(1-p)^{n_e-n_T}\dfrac{\lambda^{r+n_T}e^{-\lambda}}{(r+n_T)!}}{\sum\limits_{r=0}^{\infty}\dfrac{(r+n_T)!}{r!\prod\limits_{i=1}^{j}n_i!}(1-p)^{rj}\,p^{n_T}\,(1-p)^{n_e-n_T}\dfrac{\lambda^{r+n_T}e^{-\lambda}}{(r+n_T)!}}$$

$$=\frac{\left[\lambda(1-p)^j\right]^r e^{-\lambda(1-p)^j}}{r!}$$

(17)

Which is Poisson with mean $\lambda(1-p)^j$.

3 Investigation into the Quality of $\hat{p}$

The investigation to assess the quality of the MLE was conducted through a simulation exercise using Maple version 8 software[28]. The simulation study consisted of 1000 runs for each parameter combination. For each run the number of faults were simulated from a Poisson distribution and for each fault a time of detection was simulated from the geometric distribution, where time is measured discretely to indicate which test in the sequence exposed the fault. The parameters used within the simulation were the expected number of faults prior to starting any test, i.e. λ, and the probability that a fault will be exposed on a particular test, i.e. p. The exercise considered only the first 4 sequential tests, i.e. j=1..4, p ranging from 0.1 to 0.9 and λ ranging from 5 to 20. The choice of the range of parameter values and tests numbers was to focus the assessment on small sample sizes.

  The MLE for p was evaluated after each test. The estimation procedure was evaluated for bias and accuracy. These are described in the following.

3.1 Bias

We have chosen to measure bias with relative error Eq. (18).

$$relative\,error = \frac{\displaystyle\sum_{i=1}^{1000}\frac{\hat{p_i}-p}{p}}{1000} \qquad (18)$$

We see from Figure 2 that the absolute relative error decreases as λ increases across all simulations. In addition the relative error is increases in absolute value for small values of p while for larger values is approaches 0. Therefore, relative error is greatest in situations where there are few expected faults and efficacy of testing is poor. Overall, the MLE performs well with respect to error, as the relative error is worse for small value of p then absolute error is insignificant.

INSERT FIGURE 2

3.3   Accuracy

In order to assess the accuracy of the MLE the root of the mean square error (RMSE) was evaluated against the Cramer Rao Lower Bound (CRLB).

$$ratio = \sqrt{\frac{\displaystyle\sum_{i=1}^{1000}\left(\hat{p_i}-p\right)^2}{1000}}{\frac{1}{I(p)}} \qquad (19)$$

Figure 3 illustrates the ratio for various parameter values of p and λ for the first four tests. It can be seen that as λ increases the ratio approaches 1 and for small p the ratio increases as the number of tests increase; this is consistent with the results from the bias study. Overall the estimator performs well. For certain parameter combinations the RMSE outperformed the CRLB; this is possible due to the bias in the MLE, as the CRLB is a bound for unbiased estimators. Through using the expert judgment to assess the number of faults likely to be in the design we are creating the possibility for inserting bias into the estimation procedure; although this is typically

compensated for by the use of both the data and the expert assessment in the estimator of p.

<div align="center">INSERT FIGURE 3</div>

4.   Cost Model

We consider a situation where a prototype has been constructed and we seek to assess the appropriate number of consecutive tests for which to plan. We assume we have conducted an elicitation exercise and as such have a Poisson distribution describing the number of faults within the design.   Moreover, we assume each test cost C payable at the start of each test and the discount rate for money is δ for the period of one test. Therefore the present value at project time 0 of conducting J consecutive tests is:

$$C\sum_{i=0}^{J-1} e^{-\delta i} = C\frac{1-e^{-\delta J}}{1-e^{-\delta}} \tag{20}$$

The expected number of faults remaining undetected following J consecutive tests is $\lambda(1-p)^J$. If we assume that the operation time to realize a fault is independent and identically distributed whose distribution has Laplace Transform $L_T(z)$ and that the penalty associated with realizing a fault in operation is P then the present value at project time 0 of the costs incurred by the faults realized in operation ca be expressed as the following.

$$e^{-\delta J}\left[P\sum_{i=1}^{N-\sum_{k=1}^{J} n_k} e^{-\delta T_i}\right] \tag{21}$$

We evaluate the expected value of Eq. (21) with respect to T_i and then with respect to N.

$$E_{T_i}\left[e^{-\delta J}\left[P\sum_{i=1}^{N-\sum_{k=1}^{J}n_k}e^{-\delta T_i}\right]\right]=e^{-\delta J}\left[P\sum_{i=1}^{N-\sum_{k=1}^{J}n_k}E_{T_i}\left[e^{-\delta T_i}\right]\right]$$

$$=e^{-\delta J}P\sum_{i=1}^{N-\sum_{k=1}^{J}n_k}L_T(\delta) \tag{22}$$

$$=e^{-\delta J}P\left(N-\sum_{k=1}^{J}n_k\right)L_T(\delta)$$

Taking the expected value of this function with respect to the number of faults that would exist within the design following J consecutive tests we obtain the following.

$$P\lambda L_T(\delta)\left[(1-p)e^{-\delta}\right]^J \tag{23}$$

Combining this with the expected test cost we have the following expression for Expected Total Costs, TC.

$$TC=C\frac{1-e^{-\delta J}}{1-e^{-\delta}}+P\lambda L_T(\delta)\left[(1-p)e^{-\delta}\right]^J \tag{24}$$

We seek the value of J that will minimize TC. We achieve this through differentiating with respect to J to obtain Eq. (25) and solving for its root to obtain Eq. (26). A second derivative test will show that this is a minimum and finally we will show that the integer solution for J will be the integer immediately above or below Eq. (26).

$$\frac{dTC}{dJ}=C\frac{\delta e^{-\delta J}}{1-e^{-\delta}}+P\lambda L_T(\delta)\left(\ln(1-p)-\delta\right)\left[(1-p)e^{-\delta}\right]^J \tag{25}$$

We denote the root of Eq. (26) with J^*.

$$J^*=\frac{\ln\left[\dfrac{C\delta}{\left(1-e^{-\delta}\right)P\lambda L_T(\delta)\left(\delta-\ln(1-p)\right)}\right]}{\ln(1-p)}$$

$$=\frac{\ln\left(\dfrac{C}{P}\right)+\ln\left[\dfrac{\delta}{\left(1-e^{-\delta}\right)\lambda L_T(\delta)\left(\delta-\ln(1-p)\right)}\right]}{\ln(1-p)} \tag{26}$$

Second derivative of *TC* is presented in Eq. (27), which we will show to be positive for all real value of *J*.

$$\frac{d^2TC}{dJ^2} = -C\frac{\delta^2 e^{-\delta J}}{1-e^{-\delta}} + P\lambda L_T(\delta)\big(\ln(1-p)-\delta\big)^2\big[(1-p)e^{-\delta}\big]^J \qquad (27)$$

The following argument demonstrates that the second derivative Eq. (27) is negative for all real values of *J*. This is achieved through the deriving the value for *J* where Eq. (27) becomes positive and showing that it is an imaginary number.

$$-C\frac{\delta^2 e^{-\delta J}}{1-e^{-\delta}} + P\lambda L_T(\delta)\big(\ln(1-p)-\delta\big)^2\big[(1-p)e^{-\delta}\big]^J > 0$$

$$-C\frac{\delta^2}{1-e^{-\delta}} + P\lambda L_T(\delta)\big(\ln(1-p)-\delta\big)^2\big[(1-p)\big]^J > 0$$

$$J > \frac{\ln\left[\dfrac{C\delta^2}{P\lambda L_T(\delta)\big(\ln(1-p)-\delta\big)^2\big(1-e^{-\delta}\big)}\right]}{\ln(1-p)}$$

$$J > -J^* + \frac{2\ln\left(\dfrac{C}{P}\right) + \ln\left(\dfrac{\delta}{\ln(1-p)-\delta}\right)}{\ln(1-p)}$$

$$J > -J^* - \frac{i\pi + \ln\left(1-\dfrac{\ln(1-p)}{\delta}\right)}{\ln(1-p)}$$

where $i = \sqrt{-1}$

This implies that J^* is a minimum for *TC*.

Because *TC* is a continuous function and $\dfrac{dTC}{dJ}$ is a monotonically increasing function of *J* then the integer solution must be the integer which is immediately greater than J^* or less than J^*.

5 Illustrative Example

Consider an item to be placed on test. We will consider the analysis supported by the modeling at three different points in time; namely pre-test, after the third test and after 10 tests.

Pre-test Analysis

The prior distribution measuring the number of faults within the item is Poisson distributed with a mean of 10. Each test costs $10000 and the penalty for realizing a fault in service is $100 000. The rate of interest is 5% and a test requires a month to complete. Past experience has shown that the probability of fault being realized on a test is approximately 0.25. Table 1 is a list of the values of the parameters used with the cost model Eq. (24) to obtain the optimal number of tests to conduct Eq. (26).

INSERT TABLE 1

Figure 4 illustrates the expected cost associated with conducting a number of different tests, and as can be confirmed with Eq. (26) the optimal number of tests to plan for at time 0 is 14, which an expected cost of $168 811. It is worth noting the rate of change of the expected costs changes much more dramatically for tests less than the optimal than greater than the optimal. Therefore, over estimating the number of tests may not incur excessive costs.

INSERT FIGURE 4

Analysis after Three Tests

We assume that testing commenced and the project manager had planned for 14 tests. Therefore we expected at project time 0 that testing will be complete in 14 months. Table 2 contains the results of the first three tests, which were available after 3 months.

Using Eq. (4) the MLE of p is 0.3, which is slightly higher than the original estimate. Substituting this into Eq. (17) results in an expected number of faults remaining in the design of 3.4. Using these updated estimates with the cost model Eq. (24) we obtain a new estimate of the optimal number of further test to conduct, namely 9 more tests, which will result in an expected further costs of $113 753. Therefore, we have lowered the expected number of tests after the data was available from the first three tests.

In order to assess the level of confidence in these estimates we consider 95% confidence intervals for p. Table 3a provides a comparison of the lower bounds for p from the three methods while Table 3b provides a comparison using the upper bounds for p from the three methods. Due to the asymmetric intervals about the confidence level and the finite support for the parameter p, there is little different in Table 3b compared with Table 3b. Comparing the figures in Table 3a, we see that while there is a large variation in the upper bound of recommended tests for which to plan, e.g. the conservative method is almost twice as many tests as the Relative Likelihood; all methods suggest testing should continue.

Analysis after Ten Tests

Seven more tests were conducted resulting in a total of ten tests. There were two more faults detected after tests 3. These occurred on test 4 and 7. The updated MLE of p based on the data from all ten tests is 0.36 and the expected number of faults remaining undetected is 0.12. The cost optimal number of tests is 0 based on these updated figures and the expected further costs assuming incurred assuming no further testing is conducted is $22 315. Table 4 is a summary of the results obtained from using the lower bound estimate of p from the three methods. The samples size is

perhaps too small for the Likelihood Ratio Statistic, as the confidence interval appears optimistic. It is worth noting the sensitivity of the cost optimal number of tests in the lower bound, as a drop from a p of 0.142 to 0.063 results in an increase in expected costs by a factor of 2.36.

INSERT TABLE 4

6. Conclusions

Research to-date on reliability growth modeling has focused primarily on modeling the rate at which improvements are made to support predictions of required test time to achieve target levels of reliability. This implicitly assumes the decision-maker believes there is a need for testing and can identify a target level of reliability. Therefore, the decisions to test or not are typically made through ad hoc assessments, informally trading off test time today for reliability in the future. This paper has attempted to formally account for the key variables that are relevant for such a decision when considering a potential sequence of reliability growth tests.

A model was presented for supporting inference during a reliability growth program where data are discrete. The model is a discrete development of earlier models. The modeling assumptions are consistent with many reliability growth models. The data necessary to operationalize the model are collected during typical design and development programs and therefore require little extra effort to collect.

One shortcoming of the model is the assumption that faults are removed with certainty once they are discovered. While this is a naïve assumption, a model that explicitly considered imperfect repairs would rely on further assumptions. The model in its current form can address this shortcoming through re-assessing the number of engineering concerns following major re-designs.

A second shortcoming of the model is with the requirement of explicitly measuring the penalty of realizing a fault in operation. We have simply measured this

with one cost, however in practice this will vary depending on the type of fault that is revealed. This is being considered for further developments of the modeling.

References

1. Blischke, W and Murthy, D (2000) *Reliability: Modeling, Prediction and Optimization*, Wiley.

2. Hobbs GK (2000) *Accelerated Reliability Engineering – HALT and HASS,* Wiley

3. Jewell, W (1984) "A General Framework for Learning Curve Reliability Growth Models", *Operations Research*, **32**, pp 547-558.

4. Xie, M (1993) Software Reliability Models - a Selected Annotated Bibliography, *Software Testing, Verification and Reliability*, **Vol. 3,** pp. 3-28.

5. Ansell JI and Philips, MJ (1994) Practical Methods for Reliability Data, Oxford University Press.

6. Walls L and Quigley J (2001) "Building Prior Distributions to Support Bayesian Reliability Growth Modelling Using Expert Judgement", *Reliability Engineering and System Safety,* 74, pp 117-128.

7. Walls L, Quigley J and Kraisch M (2005) 'Comparison of Two Models for Managing Reliability Growth During Product Development', *IMA Journal of Mathematics Applied in Business and Industry,* Vol 16 pp. 12-22.

8. Gilchrist, W (1993) Modelling Failure Mode and Effects Analysis, *International Journal of Quality and Reliability Management*, **10**, 16-23.

9. Kletz T.A. (1997) "Hazop – Past and Future", *Reliability Engineering and System Safety*, **55**, 263-6.

10. Fries A and Sen A (1996) A Survey of Discrete Reliability-Growth Models, *IEEE Transactions on Reliability*, **45**, pp 582-604

11. Dohi T, Yasui K and Osaki S (2003) Software Reliability Assessment Models Based on Cumulative Bernoulli Trial Processes, *Mathematical and Computer Modelling*, **38**, pp 1177-1184.

12. Rosner, N (1961) System Analysis - Non-Linear Estimation Techniques, Proc. Seventh National Symposium on Reliability and Quality Control, Institute of Radio Engineers .

13. Jelinski, Z and Moranda, P (1972) Software reliability research. *Statistical Computer Performance Evaluation* (ed. W. Freiberger), 485-502, (Academic Press, New York).

14. Goel AL and Okumoto, K (1979) Time Dependent Error Detection Rate Model For Software Reliability and Other Performance Measures, *IEEE Transactions on Reliability,* **28**, 206-211.

15. Crow, LH (1974) Reliability Analysis of Complex Repairable Systems Reliability and Biometry**,** Ed F Proschan and RJ Serfling, SIAM.

16. Robinson, D and Deitrich, D (1987), A New Nonparametric Growth Model**,** *IEEE Trans. Rel.*, **R-36**, pp. 411-418.

17. Robinson D and Dietrich, D (1989) A Non-parametric-Bayes Reliability Growth Model, *IEEE Trans Rel.* **R-38,** 591-598.

18. Cozzolino, JM (1968) Probabilistic Models of Decreasing Failure Rate Processes, *Naval Research Logistics Quarterly*, **15**, 361-374.

19. Ebrahimi, N (1996) How to Model Reliability Growth When Times of Design Modifications are Known, *IEEE Transactions on Reliability,* **45**, 45-58.

20. Calabrai, R Guida, M and Pulcini, G (1996) A Reliability-Growth Model in a Bayes-Decision Framework, *IEEE Transactions on Reliability,* **45**, 505-510.

21. Quigley, J.L. and Walls L.A. (1999) "Measuring the effectiveness of reliability growth testing", *Quality Reliability Engineering International,***15,** 87-93

22. IEC 1164 (1995) *Reliability growth - statistical tests and estimation methods*

23. Arnold, S (1990) Mathematical Statistics, Prentice Hall

24. Burden, R and Faires J (1989) *Numerical Analysis* PWS Kent

25. Johnson N, Kotz S and Kemp A (1993) Univariate Discrete Distributions, Wiley.

26. Efron B and Tibshirani R, (1998) *An Introduction to the Bootstrap*, Chapman and Hall/CRC.

27. Lawless, J (2003) Statistical Models and Methods for Lifetime Data, Second Edition, Wiley.

28. http://www.maplesoft.on.ca/

Table 1

Parameter	Value
P	$1 000 000
δ	0.0041
C	$10 000
P	0.25
$L_T(\delta)$	0.19
λ	10
J^*	14
Expected Minimum Costs	$168 811

Table 2 Data from first three tests

Test	Number of Faults Exposed
1	3
2	2
3	2

Table 3a   Lower Confidence Interval for p after three tests

	Conservative	Relative Likelihood	Parametric Bootstrap
J*	46	24	34
λ	8.7	7.0	8.0
Costs	$577 530	$301 931	$426 196

Table 3b   Upper Confidence Interval for p

	Conservative	Relative Likelihood	Parametric Bootstrap
J*	0	2	1
λ	0	0.4	0.1
Costs	$0	$27 817	$12 940

Table 4:   Results from Lower Bound of p after ten tests

	Conservative	Relative Likelihood	Parametric Bootstrap
p	0.057	0.142	0.063
J^*	33	12	31
λ	5.6	2.2	5.4
Costs	$444 192	$180 650	$425 986
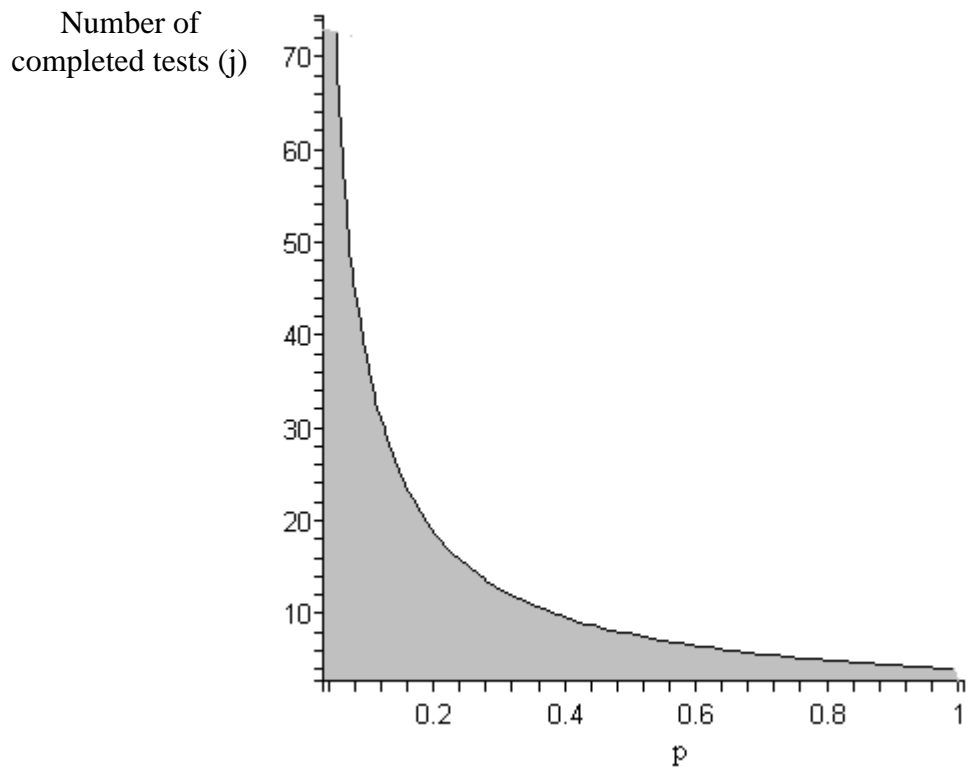
Figure 1: Region of guaranteed unique MLE
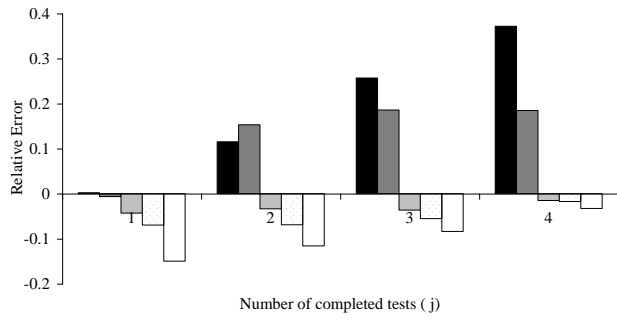
a) $\lambda = 5$
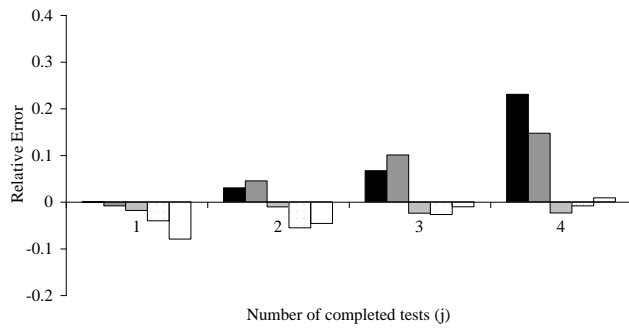
b) $\lambda = 10$
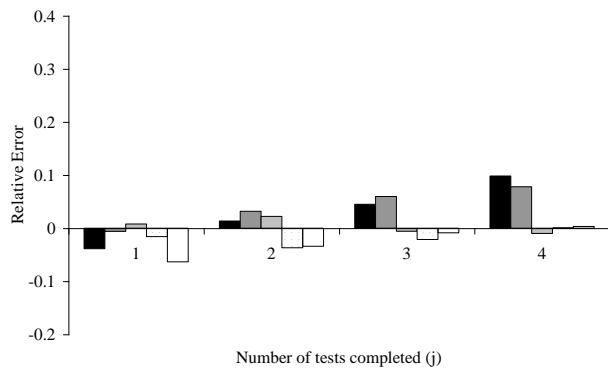
c) $\lambda = 15$
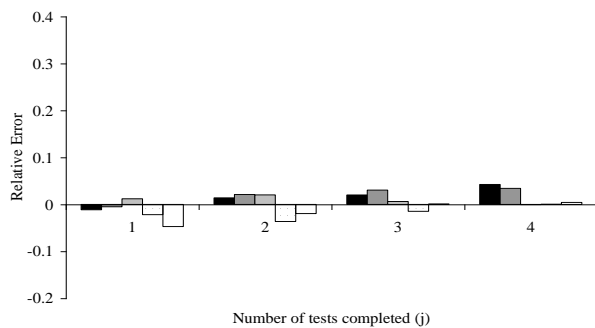
d) $\lambda = 20$

■ p=0.1 ▨ p=0.25 ▨ p=0.5 ☐ p=0.75 ☐ p=0.9

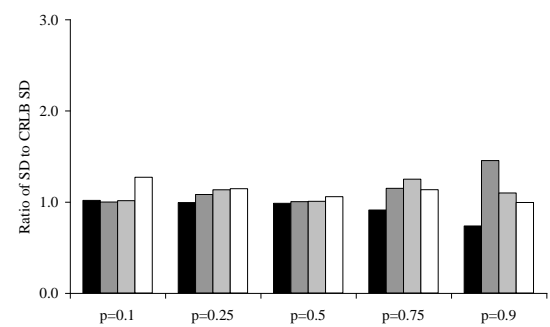Figure 2   Relative Error for Tests 1 to 4
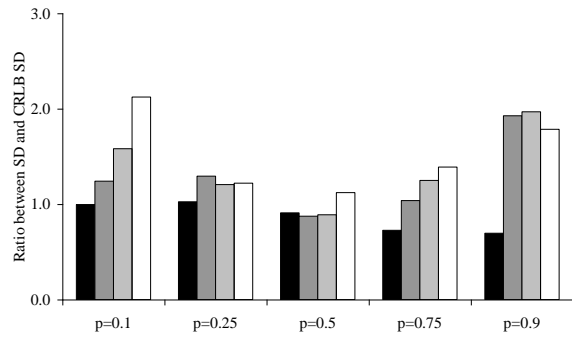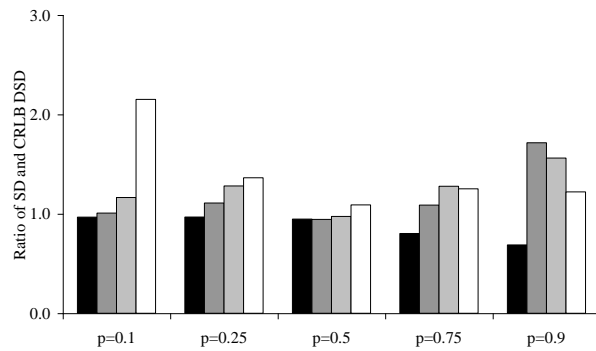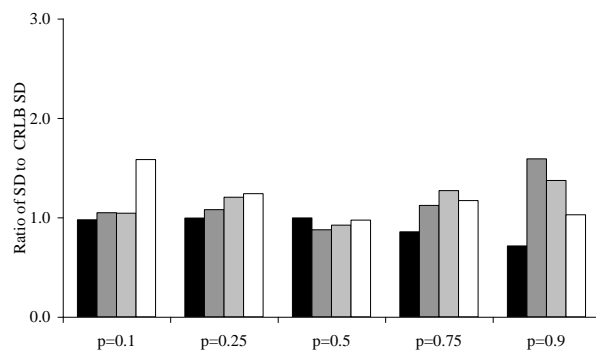
Figure 3   Ratios of Standard Deviations for p from 0.1 to 0.9
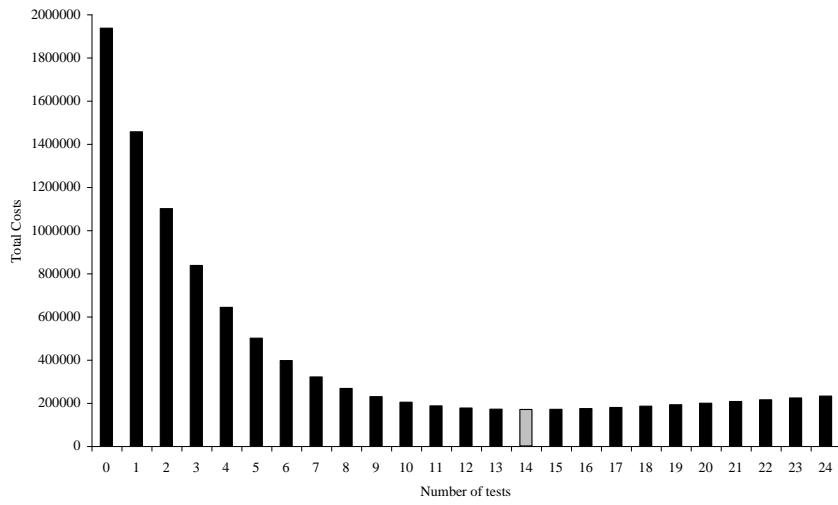
Figure 4   Expected total costs for various numbers of planned tests

BIOGRAPHY

John Quigley, *PhD, CStat*

Department of Management Science

University of Strathclyde

Glasgow G1 1QE, SCOTLAND

Email: j.quigley@strath.ac.uk

John Quigley earned a BMath in Actuarial Science from the University of Waterloo, Canada and a PhD from the Department of Management Science, University of Strathclyde, Scotland. Currently, he is a Senior Lecturer with research interests in applied probability modeling, statistical inference and reliability growth modeling. He is also a Member of the Safety and Reliability Society, a Chartered Statistician and an Associate of the Society of Actuaries.