Balancing “Test-Fix-Test” and Released System Reliability Under a Fixed Budget

Donald P. Gaver
Dept of Operations Research
Naval Postgraduate School
Monterey, CA 93943
dgaver@nps.navy.mil

Patricia A. Jacobs
Dept of Operations Research
Naval Postgraduate School
Monterey, CA 93943
pajacobs@nps.navy.mil

Ernest A. Seglie
Office of the Director,
Operational Test and
Evaluation
The Pentagon
Washington, DC
eseglie@dote.osd.mil

Situation and Problem

A new or upgraded system is designed under a proposed fixed budget $\$B$. The initial design contains an unknown number of *design faults*, or premature-failure modes, that individually may give rise to system failure if a product (individual copy) of the design is typically exercised, either during a test period or after release for actual use. If a failure occurs in actual operation this is costly, possibly dangerous, and to be avoided.

The system design is tested, and, if failures occur, or functional defects are revealed, the design faults that cause them are sought for, and (optimistically) removed (“fixed”). Testing may therefore improve the product reliability or functionality, but depletes the budget: the more testing, the better the final design may be, but the fewer product items can be manufactured and released for commercial sale or military field use because of costs of testing and “fixing.”

Mathematical models are developed and investigated to clarify the above situation for decision makers.

Initial Model

Initially there are $\boldsymbol{D}(0)$ defects in a system design. If a system copy is manufactured according to the design it contains $\boldsymbol{D}(0)-R$ ($R = 0, 1, 2, \ldots\boldsymbol{D}(0)$) defects if R have been found and successfully removed while testing. Since the probability of defect appearance (a failure) on test is taken to be $0 < p \leq 1$, independently from test to test, more than $\boldsymbol{D}(0)$ tests must typically be made. We evaluate a *run-type test stopping rule*: stop testing as soon as a success-run of r ($r = 1, 2, 3, \ldots$) consecutive successful (no failure) system copy tests have accrued; r is a decision variable that must be chosen to compromise between high probability of post-test “fielded” system design success and actual number that can be produced under the constraint of a fixed budget of $\$B(Million)$. To model the tradeoffs given cost estimates we assume a $\$c_m$/copy manufacturing cost, an individual test cost $\$c_t$, and a design defect correction cost of $\$c_f$.

Mathematical Details

Let $\boldsymbol{T}(r)$ be the random number of tests required to accrue a run of r successes (at which point stop, and accept/freeze the design for manufacture). Let $\boldsymbol{D}(r)$ be the number of defects remaining following that r-run. It can be seen that the conditional expected number of successful post-acceptance field missions is given $\boldsymbol{T}(r)$ and $\boldsymbol{D}(r)$

$$M\left(\boldsymbol{D}(r),\boldsymbol{T}(r)\right)=\frac{B-c_t\boldsymbol{T}(r)-c_m\boldsymbol{T}(r)-c_f\left(d-\boldsymbol{D}(r)\right)}{c_m}\left(Q^{\boldsymbol{D}(r)}\right) \tag{3.1}$$

where Q is the probability that a single fault remaining in the system survives a field mission ($1-Q = P$, the probability of field failure, may differ from p, the probability of test failure). It can be shown that an explicit recursion on $\boldsymbol{D}(0)$ for the joint generating function of $\boldsymbol{T}(r)$ and $\boldsymbol{D}(r)$ is (putting $\boldsymbol{D}(0) = d$)

$$g\left(z_1,z_2,d;r\right)=E\left[z_1^{\boldsymbol{D}(r)}z_2^{\boldsymbol{T}(r)}\mid \boldsymbol{D}(0)=d\right]$$

$$=z_1^d z_2^r\left[q^d\right]^r+z_2\frac{1-\left[z_2q^d\right]^r}{1-\left[z_2q^d\right]}\left[1-q^d\right]g\left(z_1,z_2,d-1;r\right) \tag{3.2}$$

from which moments can be/have been derived by differentiation, as can moments of $M(\boldsymbol{D}(r), \boldsymbol{T}(r))$; details available from the authors.

Generalization: Test-to-Test Variability of *p*.

Suppose the probability of test survival is $\tilde{q}\left(d\right)=E\left[\theta^d\right]$ where θ is a randomized version of q, now viewed as independently variable from individual test to test; likewise $\tilde{Q}\left(d\right)$ is the corresponding probability of individual field success, in general differently randomized. Explicit results corresponding to (3.2) and moments can be found (available from authors); in (3.2) it is only necessary to replace q^d by $\tilde{q}\left(d\right)$. It is convenient to let $-\ln\theta$ be random with a computable Laplace-Stieltjes transform, such as a gamma or positive stable law; if the stable law has shape parameter $\alpha\left(0<\alpha<1\right)$ then $\tilde{q}\left(d\right)=q^{d^{\alpha}}$ (normed so that $\tilde{q}\left(1\right)=q$) which can decrease very slowly with d; Feller (1966). Moments (e.g., means and variances) of numbers of field successes must generally be completed by simulation when field probability, $\boldsymbol{Q}$, is random.

Numerical Illustrations

In the results below the probability a remaining defect does not activate in the field is a constant. The unit manufacturing cost is $c_m = 2$; the cost to fix a defect is $c_f = 2$; each test costs $c_t = 1$. There is a total budget of B=100. It is assumed that the probability a defect is activated during a test is a random variable, which is independent from test to test; in particular, the expected probability none of d defects are activated during a test is $\tilde{q}_i\left(d\right)=E\left[\theta^d\right]$ where θ is a random variable having a distribution function G. In the results below, $-\ln\left(\theta\right)$ has a gamma distribution with mean μ and shape parameter β, so that

$$\tilde{q}\left(d\right)=E\left[\theta^d\right]=E\left[\exp\left\{-d\left(-\ln\left(\theta\right)\right)\right\}\right]=\left(1+\frac{\mu d}{\beta}\right)^{-\beta} \tag{4.1}$$

The parameters are chosen so that $\tilde{q}\left(1\right)=\theta$ to match the constant probability of defect activation model with one defect and results in $\mu=\beta\left(\theta^{-1/\beta}-1\right)$ with β a tuning parameter.

Table 1: Initial Number of Defects = 7; Field Success Prob. is Constant

Run Number r	E[Prob One Defect Surv. One Test] $E[q]$	Prob. One Defect Surv. a Field Mission Q_f	Gamma Shape Param. for Tests	Expected Number of Field Successes (Std. Dev.)	Expected Probability of Mission Success (Std. Dev.)	E[Number of Tests to Obtain r Succ. in a Row] (Std. Dev.)	E[Number of Syst. Purchased After Test] (Std. Dev.)
1	0.7	0.7	∞	13.1 (7.8)	0.36 (0.24)	4.5 (2.0)	39.8 (4.9)
1	0.7	0.7	3	11.1 (7.4)	0.29 (0.22)	3.8 (2.0)	41.4 (5.0)
1	0.7	0.7	1/3	6.6 (4.7)	0.15 (0.12)	2.3 (1.5)	45.3 (3.7)
3	0.7	0.7	∞	21.2 (4.9)	0.82 (0.22)	11.2 (2.1)	26.9 (3.9)
3	0.7	0.7	3	20.2 (5.4)	0.79 (0.25)	11.3 (2.4)	26.9 (4.6)
3	0.7	0.7	1/3	14.3 (7.2)	0.56 (0.35)	10.2 (4.0)	30.2 (8.2)
5	0.7	0.7	∞	19.3 (4.0)	0.93 (0.14)	14.9 (2.3)	21.0 (3.7)
5	0.7	0.7	3	18.5 (3.7)	0.93 (0.15)	15.3 (2.5)	20.4 (4.1)
5	0.7	0.7	1/3	14.4 (4.9)	0.84 (0.26)	16.6 (4.0)	18.9 (7.1)
1	0.7	0.5	∞	6.6 (7.9)	0.19 (0.24)	4.5 (2.0)	39.8 (4.9)
1	0.7	0.5	3	5.0 (7.0)	0.14 (0.21)	3.8 (2.0)	41.4 (5.0)
1	0.7	0.5	1/3	1.6 (3.4)	0.04 (0.10)	2.3 (1.5)	45.3 (3.7)
3	0.7	0.5	∞	18.4 (8.4)	0.72 (0.32)	11.2 (2.1)	26.9 (3.9)
3	0.7	0.5	3	17.1 (8.0)	0.69 (0.34)	11.3 (2.4)	26.9 (4.6)
3	0.7	0.5	1/3	10.3 (8.9)	0.44 (0.40)	10.2 (4.0)	30.2 (8.2)
5	0.7	0.5	∞	18.2 (5.2)	0.89 (0.22)	14.9 (2.3)	21.0 (3.7)
5	0.7	0.5	3	17.4 (4.9)	0.88 (0.23)	15.3 (2.5)	20.4 (4.1)
5	0.7	0.5	1/3	12.9 (6.2)	0.78 (0.34)	16.6 (4.0)	18.9 (7.1)
1	0.5	0.7	∞	21.7 (8.1)	0.65 (0.28)	6.5 (1.5)	34.9 (3.8)
1	0.5	0.7	3	18.3 (9.2)	0.54 (0.31)	5.7 (2.0)	36.9 (5.1)
1	0.5	0.7	1/3	9.4 (17.4)	0.24 (0.23)	3.2 (2.1)	43.1 (5.1)
3	0.5	0.7	∞	25.5 (2.9)	0.95 (0.12)	10.9 (1.2)	26.9 (2.1)
3	0.5	0.7	3	24.9 (3.3)	0.95 (0.13)	11.1 (1.4)	26.5 (2.4)
3	0.5	0.7	1/3	20.5 (6.4)	0.83 (0.29)	11.5 (2.9)	26.6 (5.9)
5	0.5	0.7	∞	22.7 (2.2)	0.99 (0.05)	13.4 (1.4)	23.0 (2.2)
5	0.5	0.7	3	22.2 (2.4)	0.99 (0.06)	13.7 (1.5)	22.5 (2.4)
5	0.5	0.7	1/3	19.2 (3.7)	0.97 (0.12)	15.4 (2.4)	20.0 (3.9)
1	0.5	0.5	∞	16.4(10.8)	0.50 (0.35)	6.5 (1.5)	34.9 (3.8)
1	0.5	0.5	3	12.8(11.2)	0.39 (0.36)	5.7 (2.0)	36.9 (5.1)
1	0.5	0.5	1/3	4.0 (7.3)	0.11 (0.23)	3.2 (2.1)	43.1 (5.1)
3	0.5	0.5	∞	24.6 (4.8)	0.92 (0.19)	10.9 (1.2)	26.9 (2.1)
3	0.5	0.5	3	23.9 (5.2)	0.92 (0.21)	11.1 (1.4)	26.5 (2.4)
3	0.5	0.5	1/3	18.8 (8.6)	0.77 (0.36)	11.5 (2.9)	26.6 (5.9)
5	0.5	0.5	∞	22.5 (2.7)	0.98 (0.09)	13.4 (1.4)	23.0 (2.2)
5	0.5	0.5	3	22.0 (2.9)	0.98 (0.09)	13.7 (1.5)	22.5 (2.4)
5	0.5	0.5	1/3	18.9 (4.3)	0.96 (0.16)	15.4 (2.4)	20.0 (3.9)

Discussion

The basic model studied is drastically oversimplified: for instance the design has just one stage (however, see Gaver, *et al*, (2001) for similar treatment of many sequential/series stages without budgetary restrictions). Also, cost of fix, c_f , and of test, c_t, may well have substantial variable/”random” components; these are sometimes foreseen, and deliberate attempts to avoid certain design fault tests can be made; with costly and unfortunate consequences in the field.

Overview of the Table of Measures of Testing Performance

Depending upon assumed probability of test discovery and rectification of individual faults, *and* the same for field operation, *roughly* the number purchased ranges from ~ 27 ± 4 (with about 2/3 probability) after a success-run acceptance of 3, with probability of system success in the field ranging from 0.7-0.9. If no testing is conducted, 50 copies can be purchased with success probability about 0.1. Best results clearly occur when fault survival on test is relatively small, and in the field is large. The effect of successively greater random mixing of the probability of test survival is to increase the amount of testing and to almost double the variability in required testing; if the test is comparatively forgiving (0.7 vs. 0.5) the degradation of field success probability is noticeably greater.

The lesson: the tester must pay to achieve satisfactory results, starting with a flawed design.

References

Block, H.N. and T.H. Savits (1997). Burn In. *Statistical Science*, 12/1, pp. 1-13.

Gaver, D.P. and P.A. Jacobs (1997). Testing or Fault-Finding for Reliability Growth: A Missile Destructive-Test Example. *Naval Research Logistics 44*, pp. 623-637.

Gaver, D.P., P.A. Jacobs, K.D. Glazebrook, and E.A. Seglie (2001). Probability Models for Sequential-Stage System Reliability Growth *via* Failure Mode Removal. Naval Postgraduate School Operations Research Technical Report

Feller, W. (1966). *An Introduction to Probability Theory and Its Applications*, Volume II. John Wiley & Sons, Inc., New York.