Optimal and Adaptable Reliability Test Planning Using Conditional Methods

Mark A. Powell
P.O. Box 51194
Idaho Falls, ID 83405-1194

Abstract. Developing optimal cost verification strategies for performance requirements using the analysis or test methods can be difficult, if not impossible. Developing such a strategy that is also adaptable to unexpected test results is even more difficult. The author investigates a reliability requirement verification example from the automobile industry for cost optimization and adaptability to unexpected failures using classical methods. Verification strategies for this example are then developed using conditional methods. These conditional verification strategies demonstrate significant cost efficiencies, and provide easy adaptability to unexpected failures. These strategies further provide understandable decision criteria for test success or failure. Using conditional methods for verification of performance requirements provides opportunities for cost savings, adaptability to unexpected test results data, and clear decisions for product acceptance.

Introduction

Planning the verification of any performance requirement using data essentially consists of an inversion of statistical hypothesis testing. Using classical procedures, the null hypothesis chosen for the test is that *the product fails to meet the requirement*. This null is selected because in a classical hypothesis test, one can only find evidence to support rejection of the null, not to support acceptance of either the null or alternate hypotheses. If sufficient evidence exists to reject the null, then there exists sufficient evidence that the requirement is satisfied. This is a subtle yet important concept that is often overlooked in test planning. If the null that *the product meets the requirement* is selected, failure to find evidence to support rejection of the null is a much weaker position for verification. Failure to find evidence that the requirement is not satisfied is not evidence that it is satisfied.

Typically, a decision maker will accept that a product's performance requirement is satisfied if the test results provide sufficient *confidence* that the requirement is indeed satisfied. In a decision-making context, the term *confidence* refers to the assurance or probability that the requirement is satisfied given the test results. When using classical procedures for verification, the term *confidence* has an entirely different meaning, and the significance of this difference is discussed later in this report.

Conditional methods may be used for verification of a performance requirement instead of classical procedures. With conditional methods, much of the sensitivity to null hypothesis selection disappears. Further, conditional methods naturally operate in the decision-making context, and provide quantitative measures of the *confidence* that the requirement is satisfied given the test results.

A reliability requirement verification example was selected for this report because most products and systems are designed to meet important and high reliability levels. Products designed to high reliability requirements can be very expensive to test, both in dollars and time, even with accelerated life test processes. The difficulty is inherent in the fact that a product meeting a high level reliability requirement should not fail. Of course, the very lack of a failure in reliability testing does wonders for the *confidence* of the decision maker who authorized funds for the test, but complicates terribly the statistical inference problem using classical procedures.

Using classical procedures, a rejection region for the test statistic is selected based on the reliability requirement and the level of *confidence* desired by the decision maker that the requirement is satisfied. Constraint criteria follow for the test results data that would cause the test statistic to just fall into this rejection region. These constraint criteria for the test results data then become the criteria for test success.

Costing for such a reliability verification test, subject to random events, is never simple. The cost function for reliability testing is always a function of numbers of units to be tested and duration on test, the primary constraint criteria for test result data. There indeed may be other parameters, but these two are the primary driving factors for overall cost.

Finding the minimal cost test plan for reliability where failures are not expected to be observed can be quite complex. Dodson (Dodson 1996) proposes an elegant graphical method for cost optimization for an automobile industry reliability example. This method is very direct and very simple, yet is based on classical procedures for development of the verification strategy. Dodson's optimization method further provides a superior communication vehicle with management with whom the decision lays for test approval and funding.

In this report, the author applies conditional methods to this baseline automobile industry example to simplify both reliability verification test planning, and to find an optimal cost test strategy easily adaptable to unexpected test results.

Baseline Automobile Industry Example

This report intends to demonstrate how conditional methods may be used in planning a minimal cost reliability verification strategy. Dodson's automobile industry example is used as a baseline for comparison.

In this baseline example, failures for a new vehicle are modeled using the Weibull distribution (Weibull 1951) with only scale and shape parameters θ and β respectively, with the location parameter set to zero to allow failures on test start. The shape parameter, β, is believed strongly by the manufacturer from past experience to have a value of 3. (The assumption of a value for β is sometimes referred to as an *informal Bayesian* process (Abernethy 2000). This is truly a misnomer, but there is no reason to not proceed using this assumption.) Thus, the resulting failure model is the one parameter Weibull density function in equation 1.

$$f(x \mid \theta) = \left(\frac{3}{\theta}\right)\left(\frac{x}{\theta}\right)^2 e^{-\left(\frac{x}{\theta}\right)^3} \tag{1}$$

The reliability requirement for this new vehicle is 95% at 100,000 miles, $R(100{,}000 \text{ miles}|\theta) \geq 0.95$. This requirement in turn translates to a minimum requirement on the scale parameter θ, the solution for which is presented in equation 2.

$$\theta \geq \frac{100{,}000 \; miles}{\left[-\ln(Min(R(100{,}000 \; miles \mid \theta))\right]^{\frac{1}{3}}} = 269{,}141 \; miles \tag{2}$$

The cost function for the test in this example is a combination of costs for labor, development time, machine time, and costs for the prototype vehicles to be tested. Where *T* is the duration of the test and *n* is the number of prototypes to be tested, equation 3 provides this cost function assuming none of the prototypes fail before *T*.

$$C(T,n) = T*(C_l + C_d + C_m) + n*C_p \quad (3)$$

where C_l is labor cost per unit time,
C_d is development time cost per unit time,
C_m is machine time per unit time per machine, and
C_p is prototype vehicle cost per unit.

In this example, C_m is $0.50, C_l is $0.45, and C_d is $0.50, all per simulated mile. C_p is $20,000 per prototype. The manufacturer requires that the reliability requirement of 95% at 100,000 miles be met with 90% *confidence*.

For this baseline example, it is assumed that all *n* vehicles survive to duration *T*, there are no failures during the test. Dodson proceeds with a classical procedure for selecting values of *n* and *T* that satisfy the test requirement, at the decision maker's desired level of *confidence*, and plots the resulting test costs neatly as a function of *n*, the number of prototypes needed for the test. This plot is reproduced in Figure 1.

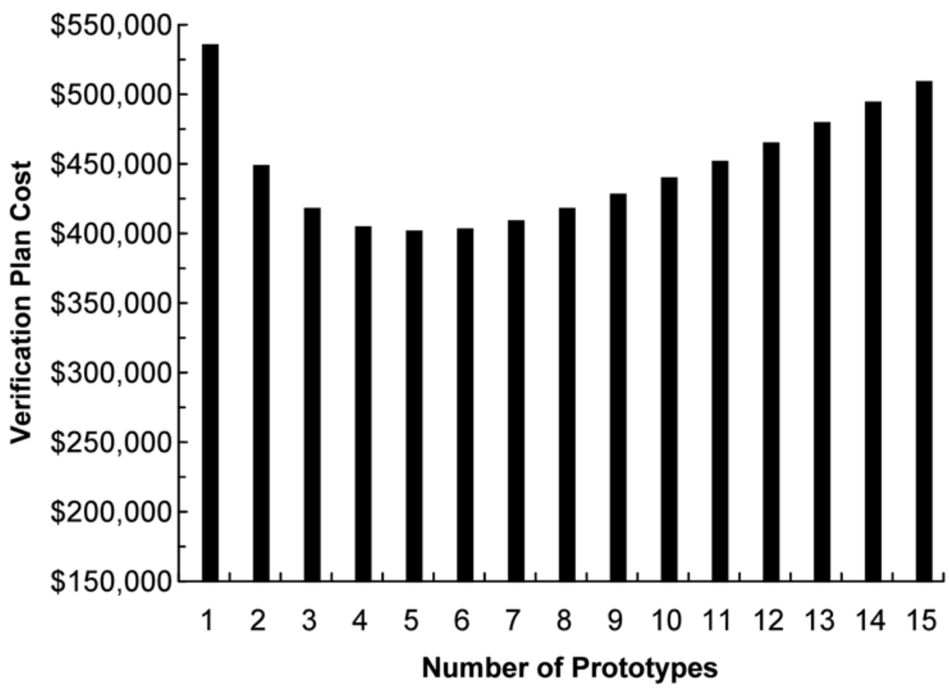

Figure1: The minimum cost test plan for the Baseline Example is evident.

As is clear in Figure 1, the minimum cost test plan will use five prototypes. The duration on test T will be 207,840 simulated miles. This optimal test plan will cost the manufacturer $401,368, provided *none* of the five prototypes fail. Dodson recommends that should a prototype unexpectedly fail, it be replaced and the duration on test T be recalculated. Dodson avers that in these cases, the test plan should be either still minimum cost, or very nearly so.

Methods

Conditional inferential methods are particularly useful for reliability analysis where no failures are observed (Powell, Shepard, 2002). They can be applied equally as well to reliability test planning where no failures are expected.

With conditional methods, the probability density for the baseline example is developed for the unknown scale parameter θ of the Weibull distribution given the test results data. Once this density is formulated, any statistical estimate may be computed. To develop the density for θ given the data, Bayes' Law (Jeffreys 1961) is employed.

$$f(\theta \mid data) \propto f(data \mid \theta)\, f(\theta) \tag{4}$$

In equation 4, $f(data|\theta)$ is the *likelihood*. This is the same likelihood function used to calculate the maximum likelihood estimate of θ. $f(\theta)$ is the *prior* density for θ. The *prior* density is selected as a model of the decision maker's uncertainty for, knowledge or ignorance of, θ before the test results data can be obtained. $f(\theta|data)$, the density of θ given the data, is called the *posterior* density. The proportionality in equation 4 is insignificant – the proportionality constant may be calculated by integrating the *posterior* density over all values of θ.

For this example, the decision maker should have a priori knowledge or belief that θ will be close to if not exceed 269,141 miles. After all, these prototypes were specifically designed to meet or exceed the reliability requirement, which is indeed satisfied if $\theta \geq 269{,}141$ miles. The prototype design engineers can be expected to usually voice strong opinions that the reliability requirement is satisfied as well. However, if the *prior* density is modeled assuming no a priori knowledge of the distribution of θ, independence from the decision maker's and design engineers' beliefs is assured, eliminating any bias for success. This is called using a *noninfomative* or *ignorance prior* (Berger 1985). Because θ is a scale parameter, a Jeffrey's prior (Sivia 1996) suitably models a priori ignorance as in equation 5.

$$f(\theta) = \frac{1}{\theta} \tag{5}$$

Now, given that there will exist as data N_f failures and $n\text{-}N_f$ survivors to simulated mileage T, the *posterior* density for θ is formulated using the Weibull distribution with f_i being the i^{th} failure time.

$$f(\theta \mid data) \propto \left(\prod_{i=1}^{N_f} \left(\frac{3}{\theta}\right)\left(\frac{f_i}{\theta}\right)^2 e^{-\left(\frac{f_i}{\theta}\right)^3} \right) * \left(\prod_{j=1}^{n-N_f} e^{-\left(\frac{T}{\theta}\right)^3} \right) * \left(\frac{1}{\theta}\right)$$
$$\propto \left(\theta^{-(3N_f+1)}\right)\left(\prod_{i=1}^{N_f} f_i^2 e^{-\left(\frac{f_i}{\theta}\right)^3} \right) e^{-(n-N_f)\left(\frac{T}{\theta}\right)^3} \quad (6)$$

In equation 6, the first term to the right of the first proportion sign is the *likelihood* for the failure data, the second is the *likelihood* for the survivor data, and the remaining term is the Jeffrey's *prior* for θ. Expecting no failures, as in the baseline example where there are as data *n* survivors to simulated mileage *T*, the *posterior* for θ simplifies even further in equation 7.

$$f(\theta \mid n\, survivors\, to\, T) \propto \frac{1}{\theta} e^{-n\left(\frac{T}{\theta}\right)^3} \quad (7)$$

Unfortunately, these *posterior* densities given the data in equations 6 and 7 are not particularly conducive to analytical integration, limiting the ability to calculate the statistics and probabilities needed to select the test criteria. Further, since they are not recognizable as known probability density models, ordinary Monte Carlo integration techniques may not be used. The remedy to this dilemma is to use Monte Carlo integration techniques via Markov Chain Monte Carlo (MCMC) methods. MCMC methods allow full range sampling of any arbitrary distribution given a formulation for the density (Gilks, et al 1996). With sufficient MCMC sampling of the *posterior* densities in equations 6 and 7, very accurate approximations can be computed for almost any statistic of interest.

For example, if there are *N* samples of θ given *n* survivors to simulated mileage *T*, it is very simple to compute the probability that the reliability requirement of 95% at 100,000 miles is satisfied. Recall from equation 2 that when $\theta \geq 269{,}141$ miles, the reliability requirement is satisfied. To calculate this probability, simply count the number of samples of $\theta \geq 269{,}141$ miles and divide by *N*.

As mentioned in the introduction, to develop the reliability requirement test plan, it is necessary to determine the constraints on the test results data that would provide the decision maker with the desired level of *confidence* that the reliability requirement is satisfied. In a decision analysis context, the term *confidence* defines the probability that the predicted consequence will be observed, in this case that the reliability requirement is satisfied. Hence, considering a decision analysis perspective for this test plan, it is necessary to find the combinations of *n* survivors to simulated mileage *T* that will produce a 90% probability that the reliability requirement is satisfied. Another way to say this is that it is necessary to find *n* and *T* such that P($\theta \geq 269{,}141$ miles) ≥ 90%.

As an aside, it is important to mention that the term *confidence* in classical statistics has an entirely different interpretation from that usually in a decision maker's mind. Using classical procedures, once a parameter estimate has been developed from a given set of data, a *confidence* interval about the estimate for a *significance* level α may be computed from the data as well. The classical interpretation of this interval is as follows. If this test could be repeated a very large number of times, calculating from the new test results data obtained each time a new parameter estimate and *confidence* interval, the true value of the parameter being estimated will lie inside this interval, in the long run, a fraction 1-α of the time. There exists no means to determine if the true value of the parameter is inside the one interval obtained from the one set of test results data from a single execution of the test. This is quite different from the interpretation of *confidence* used by decision makers and in decision analysis, and as discussed later, much more conservative than what a decision maker might really require.

Now back to the test plan problem. As described in the introduction, the task now is to find constraints on the test results data, the number of prototypes to be used n and the test duration T, such that the reliability requirement will be satisfied, i.e., that P(θ≥269,141 miles) ≥ 90%. Note in the *posterior* density in equation 7 that n and T are related. It is reasonable to define a new variable x to make things a bit easier.

$$x = n^{\frac{1}{3}}T \tag{8}$$

Now in equation 9, the *posterior* density becomes simpler.

$$f(\theta \mid x) \propto \frac{1}{\theta} e^{-\left(\frac{x}{\theta}\right)^3} \tag{9}$$

By sampling this *posterior* density using MCMC methods for different values of x, it is easy compute the P(θ≥269,141 miles) as a function of values of x. The value of x that just provides P(θ≥269,141 miles) = 90% is the test criterion. The relationship in equation 8 is then used to find values of T for integer values of n (numbers of prototypes to be used in the test), input these into the cost function of equation 3, and graph as Dodson did to find the optimal cost test plan.

Results

Using the conditional inferential procedure just described assuming no prototypes fail during the test, the needed test criterion was x=135,053.7. Figure 2 is the Dodson type chart for cost using this test criterion showing both Dodson's test plans and the conditional test plans.

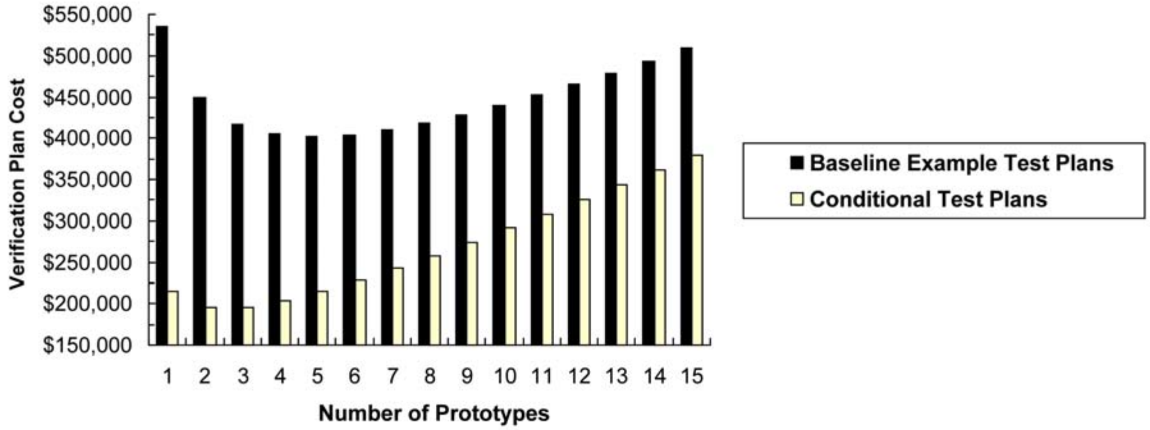

Figure 2: Conditional test plans demonstrate appreciable cost savings over Dodson's classical test plans.

Figure 2 yields some interesting results. First, conditional test plans using up to 15 prototypes are all less expensive than the baseline minimum cost test plan using five prototypes. Second, the minimum cost conditional test plan requires only two prototypes and costs less than half the minimum cost baseline test plan. The test duration for the minimum cost conditional test plan is just 107,192.2 simulated miles.

Revisiting the Baseline Example: A few comments about the baseline minimum cost test plan and the classical inferential method to produce it are in order. First, from a general testing perspective, it is always desirable to have a high probability of test success if the requirement is indeed satisfied. It would not be surprising to hear that most project managers expect the test engineers to develop a test that has a full 100% probability of success if the requirement is satisfied in the design. Of course, some test engineers avoid discussing this statistic just for this reason. Most decision-making managers are not professional statisticians. In this baseline example, the probability that all five prototypes survive until 207,840 simulated miles (the probability of passing this test given that the reliability requirement was indeed satisfied) is simple to calculate.

$$\begin{aligned}
&P\left(n\,vehicles\,survive\,to\,T \mid R(100{,}000)=0.95\right)\\
&\quad = P\left(5\,vehicles\,survive\,to\,207{,}840\,miles \mid R(100{,}000)=0.95\right)\\
&\quad = \prod_{i=1}^{5} P\left(vehicle_i\,survives\,to\,207{,}840\,miles \mid R(100{,}000)=0.95\right) \qquad (10)\\
&\quad = \prod_{i=1}^{5} R\left(207{,}840\,miles \mid \theta = 269{,}141\,miles\right)\\
&\quad = R\left(207{,}840\,miles \mid \theta = 269{,}141\,miles\right)^5 = 0.63096^5 = 0.1
\end{aligned}$$

The baseline minimum cost test plan has only a least upper bound 10% chance of success when the requirement is satisfied! This begs the question, how much more would a test plan cost that provides a much higher probability of success given the reliability requirement is met? An interesting observation on Figure 1 is that the probability of the test succeeding given the reliability requirement is satisfied always has that least upper bound of 10%, independent of the number of prototypes tested, and subsequently independent of the cost of the test. In each case, there is as much as a 90% probability that at least one of the five prototypes will fail before 207,840 simulated miles. This test plan seems destined for failure, even if the requirement is satisfied.

Consider also the possibility that the design engineers are given this test plan along with the reliability requirement to satisfy in their design. Quality assurance specifications often accompany performance specifications when provided to design engineering. Now, design engineers are typically proud of their work, and want their design to pass all tests. Noting the low probability of passing the test if the design just meets the reliability requirement, the design engineers can improve these odds by overdesigning for an even higher reliability, specifically to pass this test. Such overdesign costs money in design and in eventual production, and will be a waste of resources that cuts into net profits down the line.

A second comment considers what should be done if this rather high probability case that at least one prototype fails before 207,840 simulated miles occurs. The procedure Dodson recommends should a prototype fail is to replace it with another prototype and continue testing. Of course, with a failure, the duration to which the remaining prototypes and the new one must survive must be recalculated. This is done by using the standard classical formula for the lower confidence limit for θ. Equation 2 determined what this numerical value must be, and the desired *confidence* level of 90% yields a *significance* level $\alpha = 0.1$. Equation 11 expresses the standard classical formula for recalculation with the single failure and a replacement.

$$\theta_{L,0.1} = 269{,}141\,miles = \left(\frac{2\sum_{i=1}^{5} t_i^{\,3}}{\chi^2_{0.1,2r+2}} \right)^{\frac{1}{3}} = \left(\frac{2\left(t_f^3 + 4*T^3 + (T - t_f)\right)^3}{\chi^2_{0.1,4}} \right)^{\frac{1}{3}} \tag{11}$$

where t_i is the simulated mileage at failure or censored mileages (censored mileages are survivors at duration T (4)),
r is the number of failures (1),
L indicates the lower *confidence* limit is to be used, and
t_f is the simulated mileage at the one failure.

By solving equation 11 for T as a function of the one failure at t_f, some rather interesting results are obtained. Figures 3 and 4 are plots of test duration T and test cost C as functions of the failure mileage t_f. The peak test duration T and peak test cost C both occur when the failure mileage t_f = 67,316.9, and these peaks are 134,634 simulated miles and $315,213 respectively. The total test duration shrinks when there is an unexpected failure! And the test costs even less with a failure! Even though the cost is increased by the replacement prototype, the overall cost reduces because the time on test reduces dramatically.

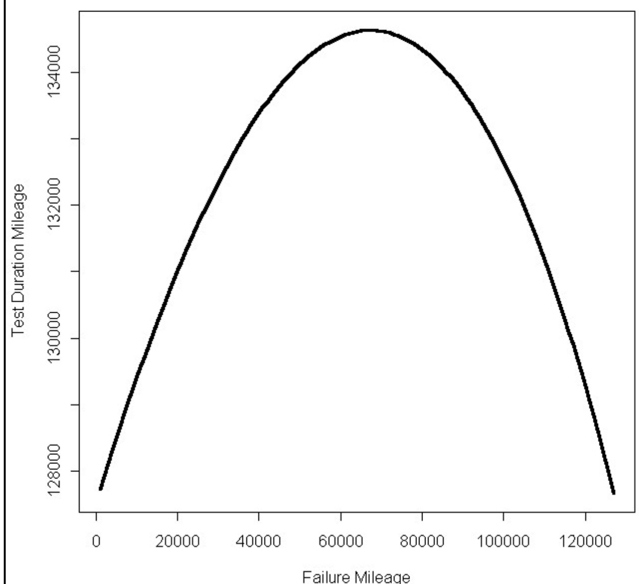

Figure 3: Test duration is a simple function of failure mileage.

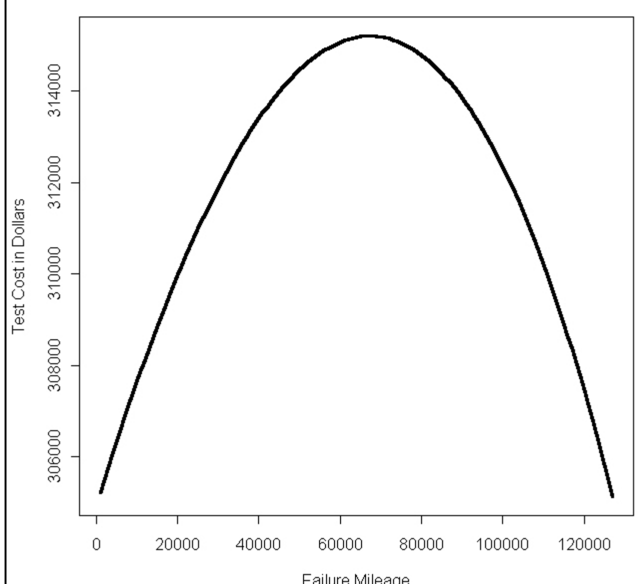

Figure 4: Test Cost is also a simple function of failure mileage.

The maximum failure mileage for a single failed prototype is 127,534 simulated miles. At this point, 90% *confidence* that the requirement of 95% reliability at 100,000 miles is achieved if none of the other four prototypes have failed. Should the only failure occur right at this point where the requirement is met, the additional prototype is not needed, and the test cost is only $285,213, much less than if no failures occurred at all.

The single failure of one of the five prototypes highlights a pathology in formulation of the minimum cost test plan developed for the baseline example where no failures are expected to occur. Obviously, the test will pass if four of the five prototypes survive to 134,634 simulated miles, regardless of when the single failure occurs. Further, the cost is minimized if the one failure occurs at t_f = 127,534 simulated miles. The test is over then, provided of course that the test engineer breaks one of the prototypes at this mileage instead of waiting for all five to survive to 207,840 simulated miles.

The probability that the test will be passed given that the reliability requirement is satisfied and that only one prototype fails is also a function of the failure mileage. Figure 5 demonstrates that this probability has improved dramatically over the least upper bound 10% chance for the test plan with no failures. If the failing prototype is failed on test insertion, at zero simulated miles, this is an increase to a little over 58%. At the minimum cost failure mileage of t_f = 127,534 simulated miles, the probability of test success given the reliability requirement is satisfied is over 65%. Big improvements, but not enough to satisfy most decision makers.

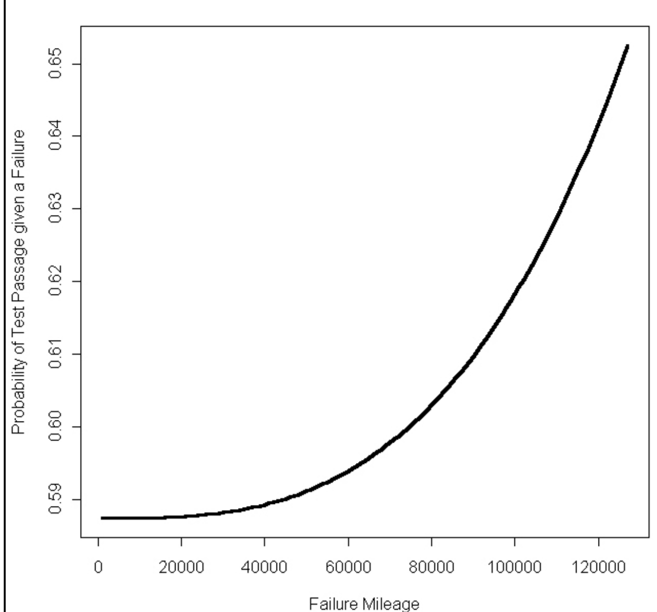

Figure 5: The probability of test success given the reliability requirement is satisfied increases with failure mileage.

The reduction in test duration and test cost, and increase in probability of passing the test when a single failure occurs are completely counter-intuitive. It just seems natural that a failure of one of the prototypes before the end of the test should produce the opposite effects.

These scenarios get much more complicated if more than one prototype fails before the end of the test. This greatly complicates any attempt at finding any minimum cost test strategy using classical methods. Dodson chose not to address this complication at all.

A third comment concerns the decision analysis for this test plan. Recall that this test plan was developed for a 90% *confidence* (in the classical sense) that the reliability requirement would be satisfied with test success. What *confidence* level in the decision analysis context, i.e., the probability that the reliability requirement is satisfied with test success, does this test plan produce? Using the five prototypes and test duration of 207,840 simulated miles as data with conditional methods, it is observed that the probability that the requirement is satisfied is much higher than needed, P($\theta \geq$269,141 simulated miles) = 99.92%. This is not a bad result if the decision maker is willing to pay more than double for the extra 9.92% probability that the test demonstrates the desired reliability, with a very high risk of test failure given the requirement is actually satisfied. Of course, the decision maker accustomed to classical testing methods may actually prefer such a high probability of requirement satisfaction on test success, and set the classical *confidence* level requirement heuristically allowing for the difference in interpretations.

With no failures, using classical procedures, it is very difficult to determine the proper combinations of n prototypes and test duration T that provide the required level of *confidence* that the requirement is met, 90%. As discussed in the introduction to this report, developing a test plan for a performance requirement using data inverts the normal hypothesis testing process. The critical region is established for rejection of the null hypothesis (that *the requirement is not satisfied*), from which one can determine the constraints for the data such that the test statistic falls in the rejection region. In this case, it is expected that no data, no failures will be observed. Classical procedures require quite a few assumptions to address this inversion of the normal hypothesis testing process with no data. These assumptions lead to overconservative test plans from a decision analysis perspective, and a less than optimal cost overall testing strategy.

Revisiting the Minimum Cost Conditional Test Plan: The probability that test success for the minimum cost conditional test plan, assuming no failures and that the reliability requirement is satisfied, can be computed using equation 10 with appropriate values of n and T. Doing so, it is observered that there is at least an 88.13% probability of test success provided the reliability requirement is satisfied. This is obviously an improvement over the least upper bound 10% probability of test success for the baseline classical test plan, but still not as high as most test engineers or decision-making managers would normally prefer. At this point, it is also obvious that the probability that the test will cost $195,428.70 is at least 88.13%, with no more than an 11.87% probability that it will cost more (one of the prototypes fails).

Suppose that one of the two prototypes fails before reaching a simulated mileage of 107,192.2. There exists an 11.87% probability of this happening during the test. Using the replacement, recalculation, and continue strategy proposed by Dodson, it is easy using conditional methods to compute a new test duration T and new costs. Intuitively, it would be expected that the test duration and the test cost to increase since a failure was observed. In test planning, it is not possible to know when this failure t_f will occur, but it must occur at or before 107,192.2 simulated miles, else the test is successful. t_f then becomes another random variable in the test planning problem. The joint *posterior* density for θ and t_f is formulated in equation 12.

$$f(\theta, t_f \mid 4\, survivors\, to\, T, 1\, survivor\, to\, T - t_f) \propto \theta^{-4} t_f^{\,2} e^{-\left[\left(\frac{t_f}{\theta}\right)^3 + \left(\frac{T - t_f}{\theta}\right)^3 + \left(\frac{T}{\theta}\right)^3\right]} \tag{12}$$

The probability that the reliability requirement is satisfied is still solely dependent on θ, the marginal distribution of which is obtained by integrating out the other random variable t_f. By sampling the joint *posterior* in equation 12 using MCMC methods, the marginal distribution for θ is simply the collection of θ samples. The value of T is then found for which

P($\theta \geq$269,141) $\geq$ 0.9. Table 1 presents a synopsis of the information that can be provided to the decision maker from conditional test planning.

	Zero Failures		One Failure		Zero or One Failure	
Number of Prototypes	Total Test Cost	P(Pass Test)	Total Test Cost	Delta P(Pass Test)	Expected Total Cost	Gross P(Pass Test)
1	$215,827.90	0.8813	$713,241.24	0.0548386	$244,966.16	0.9361386
2	$195,428.70	0.8813	$517,393.80	0.080711	$222,441.00	0.962011
3	$195,779.50	0.8813	$465,758.00	0.0840301	$219,280.60	0.9653301
4	$203,363.80	0.8813	$441,570.70	0.0917905	$225,833.58	0.9730905
5	$214,520.80	0.8813	$431,058.35	0.0993902	$236,466.27	0.9806902

Table 1: Decision information from Conditional Test Planning for zero or one failure cases.

As can be seen in Table 1, the definition of *optimum cost verification strategy* is a little more complicated. There are more factors to be considered than just minimum cost assuming no failures. It is obvious that the cost given one failure is decreasing as the number of prototypes goes up, so perhaps using more prototypes would be more optimal overall. The probability of the test passing allowing a single failure given the reliability requirement is satisfied is increasing with increased numbers of prototypes. There should be a minimum cost allowing a single failure at higher numbers of prototypes (extending the table vertically), but the cost for the most likely case (no failures) goes up and the expected cost goes up. The expected cost given no more than one failure is minimum for the test plan using three prototypes.

Test plan cost for zero failures, allowing a failure, expected cost, and probability of passing the test given the requirement is satisfied, all have to be balanced with the decision maker's risk tolerance to find the truly optimal cost test strategy. A wise decision maker balances these factors with the consequences of the design failing to satisfy the reliability requirement. There may be other factors in this balancing act as well. This decision can get quite complicated, especially if more than one failure is allowed, but conditional methods can provide all the information the decision maker needs to identify the truly optimal verification strategy.

Limitations Using Conditional Methods: Software packages implementing conditional inferential methods to support decision making are not exactly abundant. The primary reason is that the Markov Chains used to sample the *posteriors* for calculating statistics require some manual tuning. This tuning can get quite tricky in cases where no data are available, and only survivors can be used. The Markov Chains can track out of stationarity and censoring of the chain requires more manual effort.

Another minor limitation is that in the search for the test duration *T* that satisfy the reliability requirement to the required *confidence* level (decision analysis sense), *T* is seldom found exactly. The method used in this report was to find values of *T* that closely bracketed (within a few percent to either side) the 90% *confidence* level, fit with linear regression and solve for the 90% *confidence* value of *T*. Using linear regression introduces some error, but its level is usually in the third or fourth significant digit. Decision-making in this case is effectively quantized at percentiles, deciles, or higher, so this error is too small to have any impact.

These limitations are not severe. The conditional analysis presented in this report took less than a day, and considerably more time was invested in investigating the impacts of a single failure on the baseline example minimum cost test plan using classical procedures.

Conclusions

Imagine if you will: the Test Engineer (TE) who developed the test plan using classical procedures in Figure 1 is going to present it to the Project Manager (PM) for the soon to be announced Zeus 5000 SUV.

TE: "As you can see in Figure 1, the minimum cost test plan for verifying the reliability requirement for the Zeus 5000 to the 90% confidence level you requested requires five prototypes and will cost $401,368."

PM: "That's a pretty expensive test. Given that the designers met the reliability requirement, what is the probability that the Zeus 5000 will pass this test?"

TE: "At least 10%."

PM: "You mean to tell me that over $400K of my budget is going to a reliability test with a 90% chance of failing, even if we meet the reliability requirement? What is this test going to cost me if one of the five prototypes fails in testing?"

TE: "Interesting you should ask that question. If one of the prototypes does fail, depending on when it fails, the test cost reduces to as little as $285,213 and to no more than $315,213. Also, if we get a failure, the probability that we pass the test given that the reliability requirement is met goes up to at least 58%."

PM: "What?!! This doesn't make any sense! If we have a failure, the test will cost less and have a better chance of passing? You better go back and redo your test plan such that it makes some sense. I can't take this forward to my management!"

Alternate final PM response: "Really?!! Okay, I'll tell you what we'll do. Give me your Figure 1 and I will forward the $400K test estimate to my management. You have one of the test folks break one of the prototypes early in the test. That way I'll have higher probabilities of the test succeeding, the requirement being met, and being able to come in under budget. We win all around!"

Cynicism aside: except for the alternate final response, the conversation above is not all that far-fetched. Consider this conversation if the design engineer had used conditional methods in designing the test plan.

TE: "As you can see by the white bars in Figure 2, the minimum cost test plan for verifying the reliability requirement for the Zeus 5000 to the 90% confidence level you requested requires two prototypes and will cost $195,428.70. We do recommend however that we use three prototypes, it only costs $350 more."

PM: "Okay, that's a little expensive though. Given that the designers met the reliability requirement, what is the probability that we will pass this test?"

TE: "Well we can save some money if you can come down on the 90% confidence. But in this case, the probability of passing the test is a little over 88%."

PM: "No, I won't come down any, I am just barely comfortable at that confidence level. How much will the cost increase if one of the prototypes fails?"

TE: "With three prototypes, the test cost can go up to $465K. Now I know that sounds high, but the expected cost if we allow one failure is only $219K, and we have a 96.5% probability of passing the test, given the reliability requirement is satisfied. This data is contained in Table 1. As you can see, the $219K expected cost is the minimum expected cost as well."

PM: "Okay, let me summarize this to see if I have it straight. Assuming outright that the designers met the reliability requirement: 1) we have an 88% chance this test will pass and cost a little under $200K, and about a 12% chance that one of the three prototypes will fail and it will cost $465K if we continue testing; 2) with the failure and the $465K cost, we have a 96.5% chance that the test will pass; and, 3) the test plan with three prototypes has an expected cost of about $219K, which is the minimum expected cost."

TE: "That sums it up pretty well."

PM: "All right, the design engineers probably overdesigned a little for this requirement, so I am reasonably comfortable with taking this plan using three prototypes allowing one failure up to corporate. Start the detailed planning."

As should be obvious by now, the second conversation went much better because the test plan was developed considering the decision maker's thought processes, not classical interpretations of *confidence*. Conditional methods naturally provide these considerations.

Using conditional methods with Dodson's graphical optimization method provides a reliability test plan for this example that was much less expensive, had a higher probability for success, behaved much more intuitively, and was easily adaptable if the unexpected failure occurred. Further, additional costs that made sense were quantified if that unexpected failure did occur.

Conditional methods such as those presented in this report can be used for verification planning for any performance requirement, and will provide information upon which a decision maker can properly select the test plan that works best.

References

Abernethy, Robert B., *The New Weibull Handbook, Fourth Edition*. Robert B. Abernethy, North Palm Beach, Florida, 2000.

Berger, James O., *Statistical Decision Theory and Bayesian Analysis, Second Edition*. Springer-Verlag, New York, 1980.

Dodson, Bryan, *A Technique for the Optimization of Reliability Verification Tests*. Proceedings from the 49th Annual Quality Congress, Cincinnati, Ohio, 1995.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., *Markov Chain Monte Carlo in Practice*. Chapman & Hall, Boca Raton, Florida, 1996

Jeffreys, Harold, *Theory of Probability*. Oxford University Press, Oxford, 1961.

Powell, M. A., Sheppard, E. B., Applications of Conditional Inferential Methods for Operational Cost Savings for US Coast Guard C130 Aircraft Maintenance. Proceedings from the 12th Annual International Symposium, International Council on Systems Engineering, Las Vegas, 2002.

Sivia, D. S., *Data Analysis, A Bayesian Tutorial*. Oxford University Press, Oxford, 1996.

Biography

Mark A. Powell. Mr. Powell is currently adjunct faculty in the Mechanical Engineering department of the University of Idaho at Idaho Falls after a career of over 30 years in systems engineering and project management. He remains affiliated with the University of Texas at Austin where he is performing research applying conditional inferential techniques for manned spaceflight missions. Mr. Powell maintains an active consulting practice for engineering and management, and has had customers worldwide. His most recent consultancy is supporting government oversight on the $22B Future Combat System for the Army and the Defense Advanced Research Projects Agency. Mr. Powell was recently appointed as a delegate to the Organization of International Standards to represent the interests of INCOSE in the harmonization of systems and software engineering standards. Mr. Powell further serves

INCOSE in the Education and Research Technical Committee and Technical Visions Workshops.