Deriving a frequentist conservative confidence bound for probability of failure per demand for systems with different operational and test profiles

Peter Bishop, Andrey Povyakalo

City University, London

Abstract

Reliability testing is typically used in demand-based systems (such as protection systems) to derive a confidence bound for a specific operational profile. To be realistic, the number of tests for each class of demand should be proportional to the demand frequency of the class. In practice however, the actual operational profile may differ from that used during testing. This paper provides a means for estimating the confidence bound when the test profile differs from the profile used in actual operation. Based on this analysis the paper examines what bound can be claimed for different types of profile uncertainty and options for dealing with this uncertainty. We also show that the same conservative bound estimation equations can be applied to cases where different measures of software test coverage and operational profile are used.

Keywords: Statistical testing, Confidence bounds, Operational profile, Software reliability

1. Introduction

Nuclear protection systems are designed to protect against a range of safety-related plant incidents (known as postulated initiating events or PIE). A PIE can affect one or more plant parameters (such as temperature, pressure and neutron flux). These plant parameters are monitored by the protection system and the reactor is tripped if the plant parameters go outside the safe operational envelope.

In the UK, a probabilistic safety assessment (PSA) is required to justify the safety of nuclear plant. As part of this process, the performance of the protection system must be quantified in terms of probability of failure on demand, *pfd*, where the demand can be any of the PIE events. There are accepted means for estimating the *pfd* arising from hardware failures, but we also need to include an estimate for the *pfd* of the software if the protection system is computer-based. Statistical reliability testing [1, 2] is one means of estimating the software *pfd* of a demand-based system to some confidence bound, and it is recommended in functional safety standards such as IEC 61508 [3]. For example, reliability

testing was performed as part of the independent confidence building programme required by the UK Office for Nuclear Regulation (ONR) for the computer-based primary protection system (PPS) at Sizewell B nuclear power station [4]. The PPS was subjected to 5000 simulated demands to support a *pfd* claim of 10^{-3}. Reliability testing is also planned for new nuclear power stations to be installed in the UK [5].
The confidence bound derived from statistical reliability testing is based on a number of modelling assumptions. The stated assumptions in IEC 61508 [3] for the low demand rate case are:

1. The test data distribution is equal to the distribution of demands during on-line operation.
2. Test runs are statistically independent from each other, with respect to the cause of a failure.
3. An adequate mechanism exists to detect any failures which may occur.
4. Number of test cases $N > 100$.
5. No failure occurs during the N test cases.

The second assumption can be met in the protection system context as the protection system is normally reset after a reactor trip (so the software always starts from the same initial state).

The third assumption is requires a perfect "oracle" that determines if a failure has occurred. The required response is relatively easy to determine for PIE events in a nuclear plant since each simulated PIE is expected to result in a reactor trip.

The last two assumptions will also be met in a nuclear protection context as many thousands of tests are needed for the required *pfd* and the software has to be corrected and retested from scratch if a failure is observed.

To satisfy assumption 1, the number of tests for each class of demand (i.e. for each PIE) should be proportional to the demand frequency of that class during operation, so the confidence bound estimate cannot be used if the test and operational profiles differ.

This paper presents a means for estimating the confidence bound when the test profile differs from the profile used in actual operation. Based on this analysis, the paper examines what bound can be claimed for different types of profile uncertainty and the options for dealing with this uncertainty.

We also show that the same conservative bound equations can be applied in contexts where the software reliability bound and input profile are characterised in different ways.

2. Problem Statement

If a system is subjected to N test demands without failure [1], we can follow the approach suggested by Neyman and Pearson [6], Neyman [7], Clopper and Pearson [8] as it is presented by Wang [9] and identify an upper confidence bound, q, on the probability of failure on demand Q to a confidence $1 - \alpha$ as

the largest value such that the hypothesis "$H0 : Q = q$" is not rejected against the alternative "$H1 : Q < q$" at the significance level α.

Thus, q must satisfy the following equation:

$$(1 - q)^N = \alpha \tag{1}$$

However, it is often the case that the system handles different *classes* of demand, e.g. a protection system that protects against different PIE events. These demand classes are assumed to be disjoint, i.e. only a single demand can occur at any point in time.

Testing over a series of classes can be characterised by a test plan vector:

$$\mathbf{n} = \{n_1, n_2, \ldots, n_m\} \tag{2}$$

where m is a number of demand classes, n_i is the number of tests for demand class i, and the total number of tests is:

$$N = \sum_{i=1}^{m} n_i \tag{3}$$

The distribution of tests over the demand classes can be characterised by a test distribution profile vector:

$$\hat{\mathbf{p}} = \{\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_m\} \tag{4}$$

where $\hat{p}_i = n_i/N, i = 1 \ldots m$

When this multiple demand class system is used in operation it will be subject to an operational profile:

$$\mathbf{p} = \{p_1, p_2, \ldots, p_m\} \tag{5}$$

Ideally the operational and test profile distributions will match so that $\mathbf{p} = \hat{\mathbf{p}}$. However, in practice the operational profile $\mathbf{p}$ will vary if the system is used in different environments or there is uncertainty in the likelihood of different external events. So we need some means to determine a bound q_s to some confidence $(1 - \alpha)$ for a different operational profile $\mathbf{p}$ given a prior set of tests $\mathbf{n}$.

3. Problem Formulation

For some (unknown) vector of demand class *pfd*s

$$\mathbf{q} = \{q_1, q_2, \ldots, q_m\} \tag{6}$$

the likelihood of observing no failures with test plan $\mathbf{n}$ is:

$$P(\mathbf{q}, \mathbf{n}) = \prod_{i=1}^{m} (1 - q_i)^{n_i}, (0 \leq q_i \leq 1) \tag{7}$$

The $(1-\alpha)$ confidence area for all possible *pfd* vectors, $\mathbf{q}'$, is

$$D(\mathbf{n}, \alpha) = \{\mathbf{q}' : P(\mathbf{q}', \mathbf{n}) \geq \alpha\} \tag{8}$$

For an arbitrary vector of demand class *pfd*s $\mathbf{q}$ and operational profile $\mathbf{p}$, the system *pfd*, Q_S, is simply the weighted average of the vector of $\mathbf{q}$ values, i.e.

$$Q_S(\mathbf{q}, \mathbf{p}) = \mathbf{q} \cdot \mathbf{p} = \sum_{i=1}^{m} q_i p_i \tag{9}$$

The confidence area (8) constrains the set of permissable $\mathbf{q}$ vectors and induces a confidence interval for Q_S with the upper bound:

$$q_s = \max_{\mathbf{q} \in \mathbf{D}(\mathbf{n}, \alpha)} Q_s(\mathbf{q}, \mathbf{p}) \tag{10}$$

We therefore need a method for solving (10) for an arbitrary demand profile $\mathbf{p}$.

It is straightforward to solve (10) numerically for any profile $\mathbf{p}$ and test vector $\mathbf{n}$. However a numerical analysis does not permit any general conclusions to be drawn about the impact of changes in the operational profile $\mathbf{p}$.

With an analytic derivation of the confidence bound, we can model the impact of a mismatch between the test profile and the actual demand profile and identify general strategies for designing test profiles that reduce the sensitivity of the bound to uncertainties in the operational profile.

The next section describes the approach we developed to derive an analytic solution for the confidence bound.

4. Solution Approach

In Appendix A we use Lagrangian multipliers to identify the stationary points that represent the potential solutions to (10) but the solution space is complex. There are $2^m - 1$ stationary points and the optimal point depends on the specific values used in $\mathbf{p}$ and $\mathbf{n}$. As a result, there is no simple analytic solution that can be applied to all operational profiles. So we developed an alternative approach for obtaining an analytic solution by deriving a conservative approximation for (10) that makes the problem easier to solve.

In this reformulation, the likelihood (7) is approximated as:

$$\tilde{P}(\mathbf{q}, \mathbf{n}) = \prod_{i=1}^{m} \exp\left(-q_i n_i\right), \quad 0 \leq q_i \leq 1 \tag{11}$$

It is a standard result [10] that

$$\exp\left(-q_i n_i\right) \geq (1 - q_i)^{n_i} \tag{12}$$

Thus

$$P(\mathbf{q}, \mathbf{n}) \geq \alpha \tag{13}$$

implies,

$$\tilde{P}(\mathbf{q}, \mathbf{n}) \geq \alpha \tag{14}$$

Therefore, the approximated confidence area

$$\tilde{D}(\mathbf{n}, \alpha) = \{\mathbf{q}' : \tilde{P}(\mathbf{q}', \mathbf{n}) \geq \alpha\} \tag{15}$$

is a superset of the exact confidence area, i.e.

$$\tilde{D}(\mathbf{n}, \alpha) \supseteq D(\mathbf{n}, \alpha) \tag{16}$$

As a result, the approximate solution will always be conservative relative to the exact solution, i.e. for a given α, $\mathbf{n}$, $\mathbf{p}$

$$\tilde{q_s} \geq q_s \tag{17}$$

where

$$\tilde{q_s} = \max_{\mathbf{q} \in \tilde{\mathbf{D}}(\mathbf{n}, \alpha)} Q_s(\mathbf{q}, \mathbf{p}) \tag{18}$$

The log of the approximated likelihood (11) is

$$\ln \tilde{P}(\mathbf{q}, \mathbf{n}) = \sum_{i=1}^{m} -q_i n_i, \quad 0 \leq q_i \leq 1 \tag{19}$$

So the log of constraint (14) can be rearranged to become:

$$\sum_{i=1}^{m} \tilde{q}_i n_i \leq \ln \frac{1}{\alpha}, \quad 0 \leq \tilde{q}_i \leq 1 \tag{20}$$

If we define

$$\Delta\tilde{q}_i = \tilde{q}_i p_i$$

then constraint (20) can be reformulated as

$$\sum_{i=1}^{m} \Delta\tilde{q}_i \frac{n_i}{p_i} \leq \ln \frac{1}{\alpha}, \quad 0 \leq \Delta\tilde{q}_i \leq p_i \tag{21}$$

and equation (9) becomes:

$$Q_S(\mathbf{q}, \mathbf{p}) = \sum_{i=1}^{m} \Delta\tilde{q}_i \tag{22}$$

In this reformulation, the goal is to choose a set of $\Delta\tilde{q}_i$ values that maximise (22) subject to constraint (21).

Constraint (21) is a simple linear constraint for $\Delta\tilde{q}_i$. To maximise (22) we need to assign $\Delta\tilde{q}$ values to the demand class that makes the *smallest* contribution to reaching the upper limit of $\ln 1/\alpha$. So the procedure for maximising $\sum_{i=1}^{m} \Delta\tilde{q}_i$ is:

- Order the demand classes i in terms of increasing n_i/p_i value.
- Assign $\Delta\tilde{q}_i$ up to the limit of p_i for each demand class in turn until the confidence bound, $\ln(1/\alpha)$, is reached.
- In general, the last class with $\Delta\tilde{q}_i > 0$ can only be "filled" partially, i.e. $\Delta\tilde{q}_i < p_i$.
- The $\Delta\tilde{q}_i$ values for the remaining demand classes are set to zero.

This assignment strategy is illustrated in Figure 1.

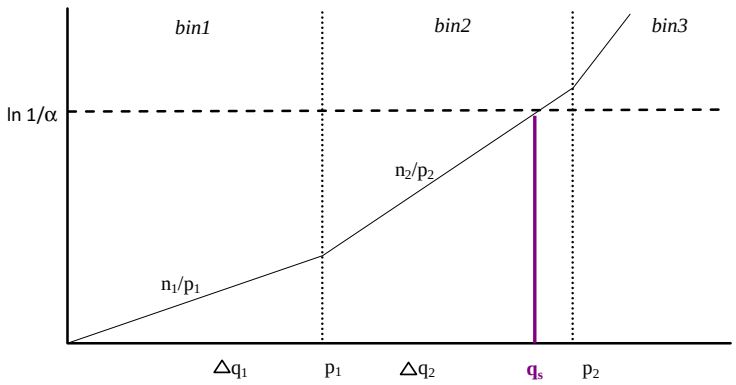

Figure 1: Illustration of bin-filling

It is a variant of the "bin-filling" strategy used in [11] where there is a worst case demand "bin" and this bin should be filled first.

There is an exception to this filling rule when several bins have an identical effect i.e. when the n_i/p_i values for the bins are the same. For equivalent bins, it does not matter how the $\Delta\tilde{q}_i$ values are distributed amongst the bins provided the limit p_i is respected for each bin.

If we assume that only a single bin k needs to be filled where:

$$\frac{n_k}{p_k} = \min_{i=1\ldots m}\left(\frac{n_i}{p_i}\right) \tag{23}$$

then equation (22) reduces to $Q_S(\mathbf{q},\mathbf{p}) = \Delta\tilde{q}_k$ and hence constraint (21) reduces to

$$Q_S(\mathbf{q},\mathbf{p})\frac{n_k}{p_k} \leq \ln\frac{1}{\alpha} \tag{24}$$

So the confidence bound, $\tilde{q}$ is:

$$Q_S(\mathbf{q},\mathbf{p}) \leq \tilde{q} = \frac{p_k}{n_k}\ln\frac{1}{\alpha} \tag{25}$$

or equivalently:

$$\tilde{q} = \max_{i=1\ldots m}\left(\frac{p_i}{n_i}\right)\ln\frac{1}{\alpha} \tag{26}$$

This upper bound still applies if more than one bin needs to be filled, but $\tilde{q}$ would be unattainable within the approximated confidence area (15) and hence be too pessimistic. This is illustrated in Figure 2

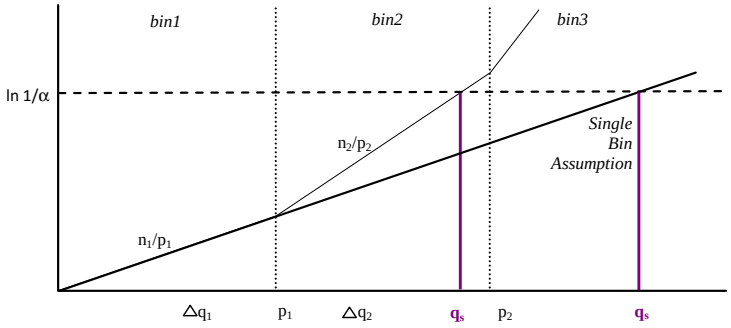

Figure 2: Conservatism of the single bin assumption

In practice however, the single bin assumption can be fulfilled when $\Delta\tilde{q}_k = p_k$ implies:

$$\Delta\tilde{q}_k \frac{n_k}{p_k} = \tilde{q}_k p_k \frac{n_k}{p_k} = \tilde{q}_k n_k \geq \ln \frac{1}{\alpha} \tag{27}$$

Since bin k will be part-filled if equation (27) is satisfied when $\tilde{q_k} \leq 1$, it follows that the single bin criterion can be met if:

$$n_k \geq \ln \frac{1}{\alpha} \tag{28}$$

For example, for 99% confidence, $\ln 1/\alpha = 4.6$, so a minimum of 5 demands on that bin would ensure that the "single bin" constraint (28) is met.

5. Approximation Accuracy

The analytic solution, $\tilde{q}_s$, is a conservative approximation to the true confidence bound q.

Appendix A.3 makes a comparison between a Lagrangian solution for the exact non-linear programming model in equations (7 – 10) and the approximate log-linear solution given in equation (26). The Lagrangian analysis derives a set of stationary points that represent potential solutions. For the equivalent of the single bin case where, for some $\mathbf{p}, \mathbf{n}$ there is a worst case bin k that determines the result, where:

$$q_s = p_k(1 - \alpha^{1/n_k}) \approx \frac{p_k}{n_k} \ln \frac{1}{\alpha}$$

For other cases where the worst case Lagrange solution point is multi-dimensional, any single-dimensional point bound under-estimates the true bound q_s, so for all $\mathbf{p}, \mathbf{n}$

$$q_s \geq p_k(1 - \alpha^{1/n_k}) = q_s^* \tag{29}$$

where bin k is the demand class with the minimum gradient value $\frac{n_k}{p_k}$. So the error relative to the exact bound q_s is constrained by

$$\frac{\tilde{q}_s - q_s}{q_s} \leq \frac{\tilde{q}_s - q_s^*}{q_s^*} = \frac{\ln 1/\alpha}{1 - \alpha^{1/n_k}} - 1 \tag{30}$$

It is clear from [30] that the approximation error could be high if the n_k value is small. However the bound estimation error need not be large if the test strategy is designed to avoid having a worst case bin k where n_k is small. The design of test strategies is discussed in more detail in Section 7. A less conservative bound on the approximation error can be found in Appendix A.4.

6. Modelling the Impact of a Profile Mismatch

With an analytic approximation to the confidence bound we can examine the impact of a mismatch between the test profile and the actual demand profile.

As noted earlier, if there is an exact match between the test profile $\hat{\mathbf{p}}$ and the demand profile $\mathbf{p}$ then we obtain the standard statistical test result

$$\tilde{q}_s \leq \frac{\ln 1/\alpha}{N} \tag{31}$$

For a mismatched profile $\mathbf{p} \neq \hat{\mathbf{p}}$

$$\tilde{q'}_s = \left(\frac{p_k}{\hat{p}_k}\right) \frac{\ln 1/\alpha}{N} \tag{32}$$

where bin $i = k$ maximises the ratio $p_i/\hat{p}_i$. As a result, we can define a scale factor S that represents the scale-up in the confidence limit $\tilde{q}_s$ for a mismatched demand profile, where:

$$S = \frac{\tilde{q}'_s}{\tilde{q}_s} = \left(\frac{p_k}{\hat{p}_k}\right) \tag{33}$$

This scale factor S remains the same regardless of the choice of confidence value, $1 - \alpha$.

Equation 32 can be interpreted as reducing the number of "relevant" test demands to $N' = N/S$ which are distributed over classes $i = \{1 \ldots m\}$ according to the new profile, where:

$$n'_i = \frac{N}{S} p_i$$

For the demand class k that determines S

$$n'_k = N p_k \frac{\hat{p}_k}{p_k} = n_k$$

This means that all n_k test demands for class k are included in the bound calculation for the new profile. For other demand classes $n'_i < n_i$, so effectively some test demands for these classes are "discarded" if they do not fit into the new

profile, resulting in a reduced number of tests N' that are considered "relevant" when deriving the revised confidence bound.

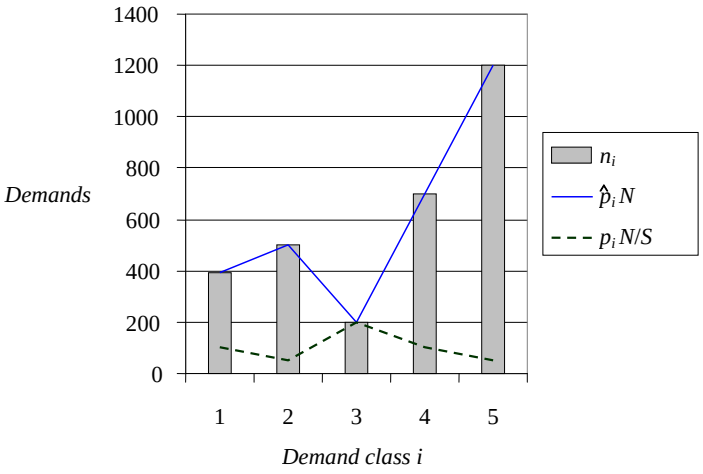

Figure 3: "Relevant" demands for different profiles

Figure 3 provides a graphical illustration of the variation of the number of "relevant" demands with the demand profile. The solid line shows the relevant demands when the profile matches the distribution of test demands. The dashed line shows the number of relevant demands when the operational profile does not match the test distribution. In this case, all the test demands for the worst case demand class ($k = 3$) are deemed relevant, but only a subset of the test demands in the other classes are included.

7. Compensating for Profile Uncertainty

If there is an uncertainty Δp probability of a demand class p_k, then for the worst case demand class, the worst possible scale factor would be:

$$S \leq= 1 + \frac{\Delta p}{\hat{p}_k}$$

This is a particular concern if $\hat{p}_k$ is very small relative to the uncertainty. For example, if there is a rare demand scenario which is estimated to be $\hat{p}_k = 10^{-6}$ but the uncertainty $\Delta p = 10^{-4}$ then it is possible that the bound would increase by two orders of magnitude in actual operation if the demand class occurs at the maximum possible rate.

Sensitivity to uncertainty in the demand probability can be reduced if extra tests Δn_k are performed such that:

$$\Delta n_k = N \Delta p \tag{34}$$

This test strategy is illustrated in illustrated in Figure 4.

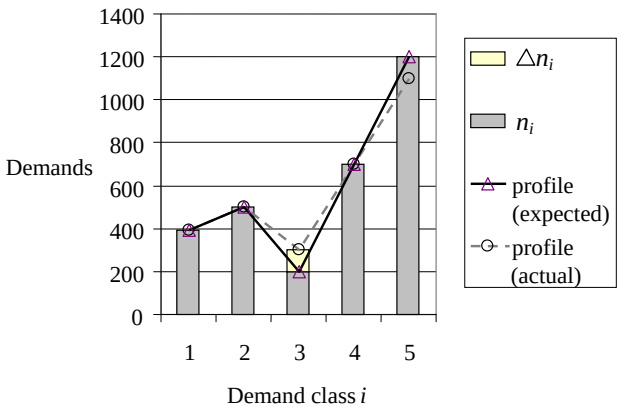

Figure 4: "Relevant" demands for different profiles

The extra Δn_k tests would not be relevant for the expected profile, so the bound based on N remains unchanged. But if the actual demand probability for the class lies within $\hat{p}_k$ and $\hat{p}_k + \Delta p$, then some or all of the extra Δn_k tests are included, and a corresponding number of tests on other demand classes are excluded, so the total number of relevant tests (and hence $\tilde{q}_s$) remains unchanged.

An alternative strategy is to set a lower bound for the number of tests n_{min}, so that extra tests are performed if the number of tests for a demand class (under the expected profile) falls below the lower limit, i.e. for all classes $i = \{1 \dots m\}$:

$$\Delta n_i = n_{min} - n_i, \qquad n_i < n_{min}$$

Provided the actual demand probability for all classes i

$$p_i \leq \hat{p}_i + \Delta n_i / N$$

the confidence bound will not increase relative to the bound derived with the original profile $\hat{\mathbf{p}}$ and test vector $\mathbf{n}$. In practice, the extra testing required to accommodate demand profile uncertainty is likely to be fairly modest. For example, if there are 50 very infrequent demand classes and $n_{min} = 5$, no more than 250 extra tests would be needed compared to the 4600 tests needed for a 99% confidence in a *pfd* of 10^{-3} based on the assumption that the test demands and the actual profile match perfectly.

If we know very little about the profile and can only specify an upper bound p_{max} for every demand class, then a much greater number of tests will be needed to assure that some target q_t will be met for all possible profiles $\mathbf{p}$ subject to this constraint. From equation (26)

$$q_t \leq \max_{i=1\dots m} \left(\frac{p_{max}}{n_i} \right) \ln \frac{1}{\alpha} \tag{35}$$

Hence the number of tests n_i required for every demand class i to have confidence in a target bound q_t is

$$n_i = \left(\frac{p_{max}}{q_t} \right) \ln \frac{1}{\alpha} \tag{36}$$

8. Numerical example

The effect of adding extra tests to compensate for profile uncertainty can be illustrated using the simple test vector shown in Table 1.

Table 1: Example test vector

Class i	Demands n_i	$\hat{p}_i$
1	234	0.078
2	540	0.180
3	6	0.002
4	720	0.240
5	1500	0.500
Total (N)	3000	

The approximate and exact 95% confidence ($\alpha = 0.05$) bounds are shown in Table 2 for the case where the test and operational profiles match, i.e. $\mathbf{p} = \hat{\mathbf{p}}$.

Table 2: Solution comparison for matching test and operational profiles

Model	Eqn.	Bound for $\hat{\mathbf{p}}$
Single bin approx $\tilde{q}_s$	(26)	0.9986×10^{-3}
Numerical solution q_s	(10)	0.9985×10^{-3}
Single bin Lagrange q_s^*	(29)	0.9958×10^{-3}

All the bound values are close to 10^{-3} at the 95% confidence level and the relative over-estimation error for $\tilde{q}_s$ is around 0.05%. The "single bin" Lagrange lower bound estimate q_s^* is 0.4% less than the true bound.

If there is uncertainty in the operational profile of $\Delta p = 0.1$, then from (34) the scaling of the confidence bound, S, is bounded by $\max(\hat{p}_i + \Delta p)/\hat{p}_i$. The maximum scale-up of the confidence bound occurs when Δp is applied to the lowest probability bin ($i = 3$) where the scale factor is $S = (0.002+0.1)/0.002 = 51$. The bound estimates for the revised profile (where $p_3 = \hat{p}_3 + \Delta p$ and $p_5 = \hat{p}_5 - \Delta p$) are shown in Table 3.

Table 3: Solution comparison for non-matching operational profiles

Model	Eqn.	Bound for $\hat{p}_3 + \Delta p$
Single bin approx $\tilde{q}_s$	(26)	50.9×10^{-3}
Numerical solution q_s	(10)	40.1×10^{-3}
Single bin Lagrange q_s^*	(29)	40.1×10^{-3}

It can be seen that, in this case, the true bound q_s is the same as the Lagrange lower bound estimate q_s^*. This is likely to occur whenever $p_i \gg \hat{p}_i$ as the Lagrange equivalent to the single bin solution represents the worst case bound. The example also shows that the single bin approximation $\tilde{q}_s$ is a conservative upper estimate. The upper estimate is consistent with the maximum error of 27% predicted in (30). The confidence bound and its relative error can be decreased dramatically if the test vector $\underline{\mathrm{n}}$ is padded so that demand class $i = 3$ is no longer the worst case bin. If every class is padded by $\Delta n_i = N\Delta p = 300$, the original confidence bound is guaranteed if the departures from $\mathbf{p}$ do not exceed Δp. Table 4 shows the bound equation results for these two scenarios.

Table 4: Worst case confidence bound under uncertainty (padded test profile)

Model	$n_3 + N\Delta p$	$n_{i=1\ldots m} + N\Delta p$
Single bin approx $\tilde{q}_s$	2.2279×10^{-3}	0.9986×10^{-3}
Numerical solution q_s	2.2264×10^{-3}	0.9980×10^{-3}
Single bin Lagrange q_s^*	2.2264×10^{-3}	0.9980×10^{-3}

We can see that if demand class $i = 3$ is padded with 300 extra tests (a 10% increase on the original total) the worst case bound only increases by a factor of 2.3 (rather than 40). With 300 extra tests for all classes (a 50% increase on the original total) the original confidence bound will always be met.

9. Generalization of the Conservative Bound Method

The approach outlined above is expressed in terms of a profile of "demands" that represent the occurrence of some event external to the system. However the theory can be applied more broadly if different interpretations of a "profile" are used. Some alternative profile definitions which extend the applicability of the conservative reliability bound estimation method for an arbitrary profile are discussed below.

As a result, the strategies used in Section 7 for reducing sensitivity to operational profile changes are equally applicable in these new contexts. In particular, it is desirable to perform extra testing on the worst element of the system k which dominates the bound estimate.

9.1. Equivalence Class Coverage

In this definition, each "demand class" i represents a specific equivalence class in the input space of the program. As equivalence classes are disjoint (like demands) the same parameters $\mathbf{p}$, $\mathbf{n}$ can be used to characterise the operational profile and the test vector, but the bound $\tilde{q}_s$ relates the probability of failure per program execution, where the bound $\tilde{q}_s$ at confidence level $(1 - \alpha)$ is

$$\tilde{q} = \max_{i=1\ldots m} \left(\frac{p_i}{n_i} \right) \ln \frac{1}{\alpha} \tag{37}$$

9.2. Structural Coverage

In this definition, each element i represents a different element within the software structure (such as a code segment). In that case the test vector $\mathbf{n}$ represents the number of executions of the segments during reliability testing. However, we cannot use an operational profile of probabilities $\mathbf{p}$ that sum to unity because the segment executions are not, in general, disjoint. For example, a sequence of code segments connected in series would all be executed at the same time. Furthermore, code segments can be executed inside a program loop and hence be executed multiple times for each invocation of the overall program. As a result, we need to define the profile as a vector of module executions $\mathbf{x}$ where x_i represents the number of executions of segment i for each execution of the overall program. With multiple executions of the same segment and a sequence of segments being executed, it is possible that several segments will fail during the same program execution cycle and hence be merged into a single failure at the program execution level. We make the conservative assumption that segment failures will never merge. As a result, equation (9) can be redefined as

$$Q_S(\mathbf{q}, \mathbf{x}) = \sum_{i=1}^{m} q_i x_i \tag{38}$$

As this equation is formally identical to (9), the demand-based analysis and approximations still apply, so the conservative bound on the probability of failure per program execution, $\tilde{q}_s$, at confidence level $(1 - \alpha)$ is

$$\tilde{q} = \max_{i=1\ldots m} \left(\frac{x_i}{n_i} \right) \ln \frac{1}{\alpha} \tag{39}$$

9.3. Execution Time

For continuous time where a test vector $\mathbf{t}$ represents the test execution times without failure for a set of components, the single-bin exponential model provides an *exact bound* rather than an approximation (provided the failures for the elements i are disjoint). As a result, the analysis is formally identical the previous analyses so the confidence bound for the system failure rate per unit time, λ_s, given an operational profile $\mathbf{p}$ of disjoint execution probabilities per unit time would be:

$$\lambda_s = \max_{i=1\ldots m} \left(\frac{p_i}{t_i} \right) \ln \frac{1}{\alpha} \tag{40}$$

This model can be extended to a concurrently executing set of components where the profile is expressed as a usage factor $\mathbf{u}$. Note that the individual terms u_i represent the proportion of time that the component is running. The usage factor can be greater than unity if multiple instances of the same component run concurrently. We can construct a normalised operational profile where there is a disjoint execution probability per unit time for component i such that $p_i = u_i / \sum_{j=1\ldots m} u_j$. Substituting this into (40) and rescaling by $\sum_{j=1\ldots m} u_j$, we obtain:

$$\lambda_s = \max_{i=1\ldots m} \left(\frac{u_i}{t_i} \right) \ln \frac{1}{\alpha} \tag{41}$$

The scaling of the operational profile means that different components will be executing at the same time. Since simultaneous (i.e. non-disjoint) component failures would be merged into a single failure at the system level, this effectively reduces the observed failure rate. So this equation remains a conservative upper bound even if the assumption of disjoint failures does not apply.

10. Relationship to Earlier Work

There has been extensive research on the use of statistical methods for estimating software reliability using realistic operational scenarios [12, 13]. Adaptive testing strategies have been used to estimate confidence intervals (such as [14, 15]) but these strategies are designed to adapt the test profiles once failures are observed, so they are not applicable to testing high integrity systems where no failures are expected.

Musa [16] and Crespo et al. [17] have modelled the impact of different operational profiles based on reliability growth during testing, but this is not directly applicable to high integrity systems where we do not expect to observe failures in the final test phase.

Bishop [18] used an operational profile characterised by the execution of code segments within the program to rescale a prior reliability bound, but in [18] the derivation of the reliability bound required an estimate of residual faults and was not explicitly related a confidence level.

Miller et al. [19] considered the impact of operational profile on reliability estimation and suggested discarding test results that did not conform to the new profile. However, there was no formal justification for discarding tests and this was proposed in the context of preparing input data for a Bayesian reliability analysis. Our analysis formally justifies the use of a "relevant" test subset in the context of a frequentist confidence bound model.

Ehrenberger [20] proposed a frequentist confidence bound model for a new operational profile which asserts that:

$$q_s = \ln 1/\alpha \sum_{i=1,m} \frac{p_i^2}{n_i}$$

However, it can be shown that the Ehrenberger model is only valid for cases where $p_i \propto n_i$ for a subset of the demand classes i and the remaining elements in the profile $\mathbf{p}$ are zero. For other non-matching profiles, the Ehrenberger model can produce non-conservative results. For example, if the model is applied to the example in Section 8 we obtain $q_s = 2.07 \times 10^{-3}$ which is significantly less than both the true confidence bound (40.1×10^{-3}) and our conservative approximation (50.9×10^{-3}).

11. Summary and Conclusions

This paper has presented a conservative analytic method for estimating the reliability bound given a specified confidence level, a set of test demands and an

arbitrary operational profile. Based on this model we show that the "scale-up" in failure rate can be highly sensitive to uncertainties in demand probability of infrequent demand classes. We also show that adding some "padding" tests for infrequent demand classes can ensure that the original confidence bound will remain valid for a range of demand probabilities for a given class.

We have also shown that the same conservative bound estimation method can be applied in other contexts, e.g. where testing is defined in terms of time rather than demands and equivalence domains rather than demand classes.

References

[1] W. Ehrenberger, Statistical testing of real time software, in: Verification and Validation of Real-Time Software, Berlin: Springer, 1985, pp. 147–178.

[2] D. L. Parnas, G. Asmis, J. Madey, Assessment of safety-critical software in nuclear power plants., Nuclear safety 32 (2) (1991) 189–198.

[3] IEC, Functional safety of electrical/electronical/programmable electronic safety-related systems: Part 7 Overview of techniques and measures, IEC 61508-7 (2010).

[4] J. May, G. Hughes, A. Lunn, Reliability estimation from appropriate testing of plant protection software, Software Engineering Journal 10 (6) (1995) 206–218.

[5] ONR, EDF and AREVA UK EPR Generic Design Assessment: GDA Issue - Protection System Independent Confidence Building Measures, GI-UKEPR-CI-02 Revision 2 (2011).

[6] J. Neyman, E. Pearson, On the use and interpretation of certain test criteria, Biometrika (20A) (1928) 175–240, 263–294.

[7] J. Neyman, On the two different aspects of the representative methods: the method of stratified sampling and the method of purposive selection, Journal of Royal Statistical Society (5) (1934) 58–606.

[8] C. Clopper, E. Pearson, The use of confidence or fiducial limits illustrated in the case of the binomial, Biometrika (26) (1934) 404–413.

[9] Y. Wang, Fiducial intervals: What are they?, The American Statistician 54 (2) (May, 2000) 105–111.

[10] H. Turnbull, An elementary derivation of the exponential limit and of euler's constant, Mathematical Notes 29 (1935) xxi–xxiv. doi:10.1017/S1757748900002401.

[11] P. Bishop, R. Bloomfield, B. Littlewood, P. Popov, A. Povyakalo, L. Strigini, A conservative bound for the probability of failure of a 1-out-of-2 protection system with one hardware-only and one software-based protection train, Reliability Engineering & System Safety 130 (2014) 61–68.

[12] J. D. Musa, Operational profiles in software-reliability engineering, Software, IEEE 10 (2) (1993) 14–32.

[13] J. Ai, F. Zheng, J. Shang, A scenario modeling method for software reliability testing, in: Systems and Informatics (ICSAI), 2012 International Conference on, New Jersey: IEEE, 2012, pp. 2429–2433.

[14] J. Lv, B.-B. Yin, K.-Y. Cai, Estimating confidence interval of software reliability with adaptive testing strategy, Journal of Systems and Software 97 (2014) 192–206.

[15] K.-Y. Cai, C.-H. Jiang, H. Hu, C.-G. Bai, An experimental study of adaptive testing for software reliability assessment, Journal of Systems and Software 81 (8) (2008) 1406–1429.

[16] J. D. Musa, Adjusting measured field failure intensity for operational profile variation, in: Software Reliability Engineering, 1994. Proceedings., 5th International Symposium on, New Jersey: IEEE, 1994, pp. 330–333.

[17] O. J. Da Silva, A. N. Crespo, M. L. Chaim, M. Jino, Sensitivity of two coverage-based software reliability models to variations in the operational profile, in: Secure Software Integration and Reliability Improvement (SSIRI), 2010 Fourth International Conference on, New Jersey: IEEE, 2010, pp. 113–120.

[18] P. G. Bishop, Rescaling reliability bounds for a new operational profile, in: ACM SIGSOFT Software Engineering Notes, Vol. 27, New York: ACM, 2002, pp. 180–190.

[19] K. W. Miller, L. J. Morell, R. E. Noonan, S. K. Park, D. M. Nicol, B. W. Murrill, J. M. Voas, Estimating the probability of failure when testing reveals no failures, Software Engineering, IEEE Transactions on 18 (1) (1992) 33–43.

[20] W. Ehrenberger, Anforderungsprofil von software wechseln (changing the demand profile of software), atp edition-Automatisierungstechnische Praxis 55 (01-02) (2013) 56–63.

[21] I. Vapnyarskii, Lagrange multipliers, in: M. Hazewinkel (Ed.), Encyclopedia of Mathematics, Berlin: Springer, 2001.

[22] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge: Cambridge University Press, 2004, p. 244.

Appendix A. Lagrange Multipliers Analysis

Appendix A.1. Problem Restatement

The system *pfd* can be redefined as

$$Q_S(\mathbf{p}, \mathbf{r}) = 1 - R(\mathbf{p}, \mathbf{r}),$$

where

$$R(\mathbf{p}, \mathbf{r}) = 1 - \sum_{i=1}^{m} r_i p_i,$$

and: $\mathbf{p} = \{p_1, p_2, ..., p_m\}$ is a vector of *input profile* probabilities; $\mathbf{r} = \{r_1, r_2, ..., r_m\}'$ is a vector of (unknown) conditional reliabilities for the demand class.

For a test vector $\mathbf{n}$, the probability of observing zero failures given $\mathbf{r}$ is

$$\tilde{P}(\mathbf{n}) = \prod_{i=1}^{m} (r_i)^{n_i} \tag{A.1}$$

Hence, to a confidence level $1 - \alpha$, the observation of $\mathbf{n}$ tests without failure defines a confidence area for $\mathbf{r}$ where

$$D(\mathbf{n}, \alpha) = \left\{ \mathbf{r}' \mid \tilde{P}(\mathbf{n}) \geq \alpha \right\} \tag{A.2}$$

and for an operational profile $\mathbf{p}$, the confidence area (A.2) induces a $(1 - \alpha) \times 100\%$ confidence interval for the system pfd with an upper bound:

$$\tilde{R}(\mathbf{n}, \mathbf{p}, \alpha) = \min_{\mathbf{r}' \in D(\mathbf{n}, \alpha)} R(\mathbf{p}, \mathbf{r}') \tag{A.3}$$

Appendix A.2. Lagrange Multipliers Analysis

A solution to (A.3) can be found with Lagrange multipliers [21] $\mu, \lambda_i, i = 1..m$. The Lagrange function is

$$L = \sum_{i=1}^{m} p_i r_i' - \mu \times \left(\prod_{i=1}^{m} (r_i')^{n_i} - \alpha \right) - \sum_{i=1}^{m} \lambda_i (r_i - 1),$$

and its stationary point satisfies the following system of equations (Kuhn-Tucker conditions [22]):

$$\begin{aligned}
&\frac{\partial L}{\partial \mu} = \prod_{i=1}^{m} (r_i')^{n_i} - \alpha = 0; \\
&\frac{\partial L}{\partial r_i'} = p_i - \mu \frac{\alpha n_i}{r_i'} - \lambda_i = 0; i = 1..m; \\
&\lambda_i (r_i' - 1) = 0, i = 1..m.
\end{aligned}$$

Thus, for every stationary point $\mathbf{r}'$, either $\lambda_i = 0$ or $r'_i = 1, i = 1 : m$ Therefore, we can identify every stationary point with a binary vector

$$\mathbf{b} = \{b_1, b_2, ..., b_m\}',$$

where $b_i = 0$ iff $r_i = 1$ and denote

$$n(\mathbf{b}) = \sum_{i=1}^{m} n_i b_i,$$

obtaining

$$r'_i = \left(\mu \frac{\alpha n_i}{p_i}\right)^{b_i}; i = 1..m \tag{A.4}$$

$$\prod_{i=1}^{m} (r'_i)^{n_i} = \alpha;$$

$$R(\mathbf{b}) = \mu \times \alpha \times n(\mathbf{b}) + \sum_{i=1}^{m} p_i(1 - b_i). \tag{A.5}$$

We find μ by solving the equation

$$\prod_{i=1}^{m} \left(\mu \frac{\alpha n_i}{p_i}\right)^{n_i b_i} = \alpha$$

or

$$\mu = \left(\alpha \times \prod_{i=1}^{m} \left(\frac{p_i}{\alpha n_i}\right)^{n_i b_i}\right)^{1/n(\mathbf{b})} = \frac{1}{\alpha}\left(\alpha \times \prod_{i=1}^{m} \left(\frac{p_i}{n_i}\right)^{n_i b_i}\right)^{1/n(\mathbf{b})}$$

$$R(\mathbf{b}) = n(\mathbf{b}) \times \left(\alpha \times \prod_{i=1}^{m} \left(\frac{p_i}{n_i}\right)^{n_i b_i}\right)^{1/n(\mathbf{b})} + \sum_{i=1}^{m} p_i(1 - b_i)$$

If we denote $\hat{p}_i = \frac{n_i}{n(\mathbf{b})}, i = 1..m$, then

$$r'_j = \left(\mu \cdot \alpha \cdot n(\mathbf{b}) \frac{\hat{p}_j}{p_j}\right)^{b_j}; j = 1..m \tag{A.6}$$

$$\mu = \frac{1}{\alpha n(\mathbf{b})} \alpha^{1/n(\mathbf{b})} \prod_{i=1}^{m} \left(\frac{p_i}{\hat{p}_i}\right)^{\hat{p}_i b_i} \tag{A.7}$$

$$r'_j = \left(\alpha^{1/n(\mathbf{b})} \frac{\hat{p}_j}{p_j} \prod_{i=1}^{m} \left(\frac{p_i}{\hat{p}_i}\right)^{\hat{p}_i b_i}\right)^{b_j}; j = 1..m \tag{A.8}$$

$$R(\mathbf{b}) = \alpha^{1/n(\mathbf{b})} \prod_{i=1}^{m} \left(\frac{p_i}{\hat{p}_i}\right)^{\hat{p}_i b_i} + \sum_{i=1}^{m} p_i(1 - b_i); \tag{A.9}$$

$$Q_S(\mathbf{b}) = \sum_{i=1}^{m} p_i b_i - \alpha^{1/n(\mathbf{b})} \prod_{i=1}^{m} \left(\frac{p_i}{\hat{p}_i}\right)^{\hat{p}_i b_i} \tag{A.10}$$

Equation (A.8) and the feasibility constraints $r_i \leq 1, i = 1..m$ define the feasibility area $\mathbf{B}$ for the binary vectors $\mathbf{b}$ identifying the stationary points:

$$\mathbf{B} = \left\{ \mathbf{b} : \alpha^{1/n(\mathbf{b})} \prod_{i=1}^{m} \left(\frac{p_i}{\hat{p}_i} \right)^{\hat{p}_i b_i} \leq \min_{i:b_i=1} \left(\frac{\hat{p}_i}{p_i} \right) \right\} \tag{A.11}$$

Thus, the exact Lagrangian upper $(1-\alpha) \times 100\%$ confidence bound for the system pfd is

$$\tilde{Q_S} = \min_{\mathbf{b} \in \mathbf{B}} \left(Q_S(\mathbf{b}) \right) \tag{A.12}$$

Appendix A.3. Relationship to the Approximate Solution

When, $p_i = \hat{p}_i, i = 1..m$, then

$$R(\mathbf{b}) = \alpha^{1/n(\mathbf{b})} + \sum_{i=1}^{m} p_i (1 - b_i),$$

or

$$Q_S(\mathbf{b}) = \sum_{i=1}^{m} p_i b_i - \alpha^{1/n(\mathbf{b})}.$$

If $\mathbf{b} = \{1, 1, \ldots, 1\}$, then

$$R(\mathbf{b}) = \alpha^{1/N},$$

or

$$Q_S(\mathbf{b}) = 1 - \alpha^{1/N},$$

where

$$N = \sum_{i=1}^{m} n_i.$$

Let's now assume that:

$$\begin{aligned} &\sum_{i=1}^{m} b_i = 1; \\ &b_k = 1; \\ &n(\mathbf{b}) = n_k; \\ &\hat{p}_k = 1, \end{aligned}$$

Then,

$$R(\mathbf{b}) = 1 - p_k (1 - \alpha^{1/n_k})$$

or

$$Q_S(\mathbf{b}) = p_k (1 - \alpha^{1/n_k}) \approx -\frac{p_k}{n_k} \log \alpha$$

Thus, the approximation to the conservative confidence bound found using (26) is very close to one of the stationary points for the *"exact"* optimisation problem.

Appendix A.4. Upper bound on the approximation error

We can use the suboptimal Lagrangian stationary points to estimate the error of the "bin-filling" solution.

If k is the last "filled" in the "bin-filling" solution, then

$$\Delta q_k = \frac{p_k}{n_k}\left(\ln\frac{1}{\alpha} - \sum_{i=1}^{k-1} p_i\right),$$

and

$$Q_S = \sum_{i=1}^{k-1} p_i + \frac{p_k}{n_k}\left(\ln\frac{1}{\alpha} - \sum_{i=1}^{k-1} p_i\right) = \sum_{i=1}^{k-1} p_i + \frac{p_k}{\hat{p}_k}\left(\frac{1}{N}\ln\frac{1}{\alpha} - \frac{1}{N}\sum_{i=1}^{k-1} p_i\right)$$

The corresponding binary vector $\mathbf{b}$ for this has the components

$$b_i = \begin{cases} 1, i = 1..k-1 \\ 0, i = k..m. \end{cases}$$

with the stationary (sub-optimal) "Lagrangian" value for Q_S:

$$Q_S(\mathbf{b}) = \sum_{i=1}^{k-1} p_i - \alpha^{1/n(\mathbf{b})} \prod_{i=1}^{k-1}\left(\frac{p_i}{\hat{p}_i}\right)^{\hat{p}_i}.$$

Thus, the absolute error from the approximation has the upper bound:

$$Q_S - Q_S(\mathbf{b}) = \frac{p_k}{\hat{p}_k}\left(\frac{1}{N}\ln\frac{1}{\alpha} - \frac{1}{N}\sum_{i=1}^{k-1} p_i\right) + \alpha^{1/n(\mathbf{b})} \prod_{i=1}^{k-1}\left(\frac{p_i}{\hat{p}_i}\right)^{\hat{p}_i}$$