An Experimental Approach to Statistical Mutation-Based Testing[1]

SERC TR–63–P

William Hsu
Software Engineering Research Center
Department of Computer Sciences
Purdue University
West Lafayette, Indiana 47907
(317) 494-7812
hsu@cs.purdue.edu

Mehmet Sahinoglu[2]
Department of Statistics
Purdue University
West Lafayette, Indiana 47907
(317) 494-7913
nse@l.cc.purdue.edu

Eugene H. Spafford
Software Engineering Research Center
Department of Computer Sciences
Purdue University
West Lafayette, Indiana 47907
(317) 494-7825
spaf@cs.purdue.edu

August 25, 1992

[1]An abbreviated form of this paper was presented as [11].

[2]Currently visiting Purdue University from Middle East Technical University (METU), Ankara, Turkey on a Fulbright Scholarship, support which is gratefully acknowledged.

Abstract

Mutation analysis is a well-known method of measuring test case adequacy. When using mutation testing to test software, it is the tester's goal to construct a test data set that is sufficient to distinguish all mutants from the original program. If the original program executes correctly against all elements of the resultant test data set, the tester has great confidence that the program is correct.

One of the drawbacks to full mutation testing involves the number of mutations that must be executed. Each mutation is equal to the size and complexity of the original program, and the number of mutants is proportional to the square of the number of lines of source code present in the original software. Each mutant may need to be run against many test cases before being distinguished from the original code. For routines in the hundreds of lines of code, this may result in millions of separate executions.

It would be beneficial to test only a small random sampling of mutants against the given test cases to determine, at some arbitrary level of confidence, that the program is correct. This would be of great benefit in situations where full mutation testing would be too expensive, too time consuming, or simply beyond the capabilities of the available testing software.

We propose a sequential decision procedure with a stopping and decision rule to be used with a prototype mutation analysis system, MOTHRA. This statistical sequential testing is sound and scientific with a simple (good) vs. a simple (bad) hypothesis specified by Type I and II error probabilities. One is also capable of conducting sensitivity analyses by varying associated values one at a time and observing the resultant Operating Characteristics curve. Overall, this novel approach in mutation-based testing of software is superior to the simple deterministic approach currently in use.

1 Introduction

Software testing is a critical phase of the software development life cycle. However, software testing can be an expensive proposition. One desires to obtain an adequate measure of the reliability of a software system while minimizing the expenditure of resources. Testing is always a trade-off between increased confidence in the correctness of the software under examination, and constraints on the amount of time and effort that can be spent testing that software.

For over a decade, researchers have been working with *program mutation* as a method of developing test cases to test software.[1] A basic goal of program mutation is to provide the user with a measure of test set *adequacy*, by executing that test set against a collection of program *mutations*. Mutations are simple changes introduced one at a time into the code being tested. These changes are derived empirically from studies of errors commonly made by programmers when translating requirements into code, although theoretical justification also can be found for their selection[8].

A mutant is *killed* if the execution of the mutated code against the test set distinguishes the behavior or output of the mutation from the unmutated code. The more mutants killed by a test set, the better the measured adequacy of the test set. By proper choice of mutant operators, comprehensive testing can be performed, [2] including path coverage [10] and domain analysis. [17] By examination of unkilled mutants, testers can add new test cases to better the adequacy score of the entire test set.

Mutation analysis is designed to substantiate the correctness of a program Φ. Mutation analysis strives to develop test data that, when applied to the program Φ, illustrates the equality of the *intended* and *observed* behaviors of Φ. This methodology is normally embedded in a test environment that enables a tester to test his program interactively.

Somewhat more formally, the mutation approach is to induce syntactically correct changes into a program Φ, thereby creating a set of *mutant programs*. Each mutant represents a possible error in Φ, and the goal of the tester is to construct a set of test data τ that distinguishes the output or behavior of $\Phi(\tau)$ from that of all mutant programs. Test data sensitive enough to distinguish all mutant programs is deemed adequate to infer the probable correctness of Φ.

The mutation analysis methodology is as follows: submit a program Φ and a set of test data τ whose adequacy is to be determined. The mutation system first executes Φ with respect to τ. If the results are incorrect, then certainly Φ is in error. However, if the results appear correct, it may still be that Φ is in error, for the test data set τ may not be adequate to distinguish this inherent incorrectness. In this case, a set of *mutant programs* are created, call them $\Phi_1, \ldots, \Phi_m$. Each mutant program differs from the original program Φ by one single-point, syntactically correct change. Such *mutant transformations* are statically defined for a language Λ and designed to expose errors frequently committed by programmers using Λ.

For each execution of a mutant against the test set, $\Phi_i(\tau)$, exactly one of two things happens:

[1] The mutant $\Phi_i(\tau)$ yields different results than $\Phi(\tau)$, or

[2] The mutant $\Phi_i(\tau)$ yield the same results as $\Phi(\tau)$.

[1] Program mutation has been well documented in the literature and will only be summarized here. The reader unfamiliar with mutation testing is directed to recent references on mutation for detailed descriptions and further references, e.g. [4, 15, 9, 3, 18, 12].

In the first case, the mutant is said to be *dead* since the mutant difference between Φ_i and Φ has been distinguished by τ. In the second case, either:

[a] τ is not adequate to distinguish the mutant that gave rise to Φ_i, or

[b] Φ and Φ_i are *equivalent programs* and no test data τ can distinguish their behavior (the "error" that gave rise to Φ_i was not an error at all, but an alternate encoding of Φ). Such a Φ_i is called an *equivalent mutant*.

Test data that leaves no live mutants or only equivalent mutants is said to be of adequate sensitivity to infer the probable correctness of Φ. Moreover, the ratio of dead mutants to the total number of nonequivalent mutants yields a relative measure of the adequacy of the test data τ in testing Φ. The test set τ may be augmented with additional test cases and the process repeated until an appropriate level of test case adequacy or threshold level in the cost of the test (in terms of dollars, resources, etc.) has been reached. The test set τ can then be carried into the maintenance stage of the software life cycle.

2 On MOTHRA and the Stopping Rule

The MOTHRA software testing environment [6, 5, 4] is an integrated set of tools and interfaces that support the planning, definition, preparation, execution, analysis, and evaluation of mutation-based tests of software systems. MOTHRA is designed to be used starting at the earliest stages of software development and continuing through the progressively later stages of system integration, acceptance testing, operation, and maintenance. It has been in use at Purdue and other locations for the last few years as a testbed for experimental work in software engineering.

One of the drawbacks to full mutation testing with a system such as MOTHRA involves the number of mutations that must be executed. Each mutation is equal to the size and complexity of the original program, and the number of mutants is proportional to the square of the number of lines of source code present in the original software. Each mutant may need to be run against many test cases before being distinguished from the original code. For routines in the hundreds of lines of code, this may result in millions of separate executions.

It would be beneficial to test only a small statistically random sampling of mutants against the given test cases to determine, at some arbitrary level of confidence, that the program is correct. This would be of great utility in situations where full mutation testing would be too expensive, too time consuming, or simply beyond the capabilities of the available testing software.

Conventionally, a typical mutation system user must specify test requirements of the following form: [9] "I must be X% sure that Y% of the mutants of type Z are killed in unit A." Here, X% is the degree of assurance MOTHRA will use to perform the random selection of mutants of type Z in unit A. The user must still monitor the status information to determine whether or not the stopping criteria Y% for the test of mutants of type Z in unit A has been achieved. Hence, one defines a stopping criterion based on a given "Y% of the mutants to be killed" with a certain confidence level X% over a multiplicity of test-cases.

This approach may be defined as a deterministic stopping rule-of-thumb, since the entire experiment is of a random nature and is not structured by statistical hypothesis testing.

The customary ad hoc approach, although straightforward and easy, may result in a waste of testing time and resources due to lack of a statistical risk analysis. The only definitive statement that can be made from such testing involves achieving a 100% level of mutant distinction.

3 Sequential Statistical Procedures

The usual statistical hypothesis-testing procedures ordinarily have fixed sample size. In particular, the statistical testing procedures have only answered the question, "Do we have sufficient evidence to declare a particular hypothesis false?" Other models address the question, "When have we corrected an 'optimum' number of errors? That is, should we now release the software?"[1] However, if we allow the sample size to increase without limit, we can obtain a test for any prespecified probability of occurrence of Type I and Type II errors.[2] Thus, if testing for $H_0 : \Theta = \Theta_1$ vs $H_1 : \Theta = \Theta_2$, we can assure ourselves of our choice given the probability of a wrong decision. We can thus test software to any prespecified level of confidence and criticality, rather than to some single, theoretical "optimum" point.

Our approach will be to sample $X_1 = x_1, \ldots, X_k = x_k$ and after each observation $X_k = x_k$, make a decision based on $x_1, \ldots, x_k$ whether or not to continue sampling. If we stop, we want to choose between H_0 and H_1.

Let $X_1, \ldots, X_n$ be a sequence of i.i.d. (identical independently distributed) random variables (r.v.) with p.d.f. (probability distribution function) $F(x;\Theta), \Theta \in \Omega$(parameter space), where the x_i's are observed sequentially. Let $\mathcal{A}$ be the space of actions available to the statistician.

Definition: A sequential decision procedure has two components:

1. A stopping rule which specifies for every set of values $(x_1, x_2, \ldots, x_n), n \geq 1$ whether to stop sampling and choose a decision in $\mathcal{A}$, or to continue sampling and take another observation x.
2. A decision rule $d_n(x_1, \ldots, x_n)$ which chooses the action in $\mathcal{A}$ to be taken for the set of values $(x_1, \ldots, x_n)$ when the sampling is stopped.

4 The Sequential Probability Ratio Test (SPRT)

Let $x_1, x_2, \ldots$ be i.i.d r.v.'s with p.d.f. f(x;Θ) $\Theta \in \Omega = \{\Theta_1, \Theta_2\}$. Let $f_1(x) = f(x;\Theta_1)$; $f_2(x) = f(x;\Theta_2)$. We want to test $H_0 : \Theta = \Theta_1$ vs $H_1 : \Theta = \Theta_2$ without fixing the sample size in advance.

For a fixed sample size k, the NP (Neyman Pearson) Lemma tells us to reject for large values of the ratio

$$\lambda_k = \frac{f_2(x_1)f_2(x_2)\ldots f_2(x_k)}{f_1(x_1)f_1(x_2)\ldots f_1(x_k)} \tag{1}$$

[2]In this context, a Type II error is mistakenly accepting a faulty program. A Type I error is mistakenly rejecting a good program as faulty.

Wald [14] tells us to consider the sequence of ratios $\lambda_1(x_1), \lambda_2(x_1, x_2), \ldots, \lambda(x_1, \ldots, x_k)$, i.e. consider the sequence:

$$\frac{f_2(x_1)}{f_1(x_1)}, \frac{f_2(x_1)f_2(x_2)}{f_1(x_1)f_1(x_2)}, \frac{f_2(x_1)\ldots f_2(x_k)}{f_1(x_1)\ldots f_1(x_k)} \tag{2}$$

Definition: The SPRT [16, 13, 7] (Test) for testing $H_0 : \Theta = \Theta_1$ versus $H_1 : \Theta = \Theta_2$ is a rule that states:

1. if $\lambda_k(x) >= \mathcal{A}$, stop sampling and reject H_0 (accept H_1).
2. if $\lambda_k(x) <= \mathcal{B}$, stop sampling and reject H_1 (accept H_0).
3. if $\mathcal{B} < \lambda_k(x) < \mathcal{A}$, continue sampling and take another observation x_{k+1}.

Here $\mathcal{A}$ and $\mathcal{B}$ are constants determined so that α = P{Type I error} = P{Reject $H_0 \,|\, \Theta_1$}, i.e. probability of rejecting H_0 when $H_0 : \Theta = \Theta_1$ is true. β = P{Type II error} = P{Fail to Reject $H_0 \,|\, \Theta_2$}, i.e. probability of accepting H_0 when H_1 is true.

Namely, if N is the stopping time for this procedure,

$$\alpha = P_{\Theta_1}(\lambda_N(x) >= \mathcal{A}) \tag{3}$$
$$\beta = P_{\Theta_2}(\lambda_N(x) <= \mathcal{B}) \tag{4}$$

5 Binomial SPRT

Let $x_1, x_2, \ldots$ be i.i.d. Bernoulli r.v.'s with p.d.f. f(x;Θ) = $\Theta^x(1-\Theta)^{1-x}$ x=0,1.

We now wish to test for $H_0 : \Theta = \Theta_1$ vs $H_1 : \Theta = \Theta_2$.

$$z_i = \log(\frac{f(x_i;\Theta_2)}{f(x_i;\Theta_1)}) = \begin{cases} \log(\frac{\Theta_2}{\Theta_1}) & \text{if } x_i = 1 \\ \log(\frac{1-\Theta_2}{1-\Theta_1}) & \text{if } x_i = 0 \end{cases}$$

Then

$$S_k = \sum_{i=1}^{k} Z_i = r_k \log(\frac{\Theta_2}{\Theta_1}) + (k - r_k)\log(\frac{1-\Theta_2}{1-\Theta_1}) \tag{5}$$

where r_k = number of 1's in $x_1, \ldots, x_k = \sum_{i=1}^{k} x_i$. We observe $S_k = s_k$ and continue sampling if $b < s_k < a$ i.e.

$$b < r_k \log(\frac{\Theta_2}{\Theta_1}) + (k - r_k)\log(\frac{1-\Theta_2}{1-\Theta_1}) < a, \tag{6}$$
$$b < r_k(\log(\frac{\Theta_2}{\Theta_1}) - \log(\frac{1-\Theta_2}{1-\Theta_1})) + k\log(\frac{1-\Theta_2}{1-\Theta_1}) < a \tag{7}$$

Solving the inequality for r_k, we obtain, as illustrated in figure 1:

$$\frac{b - k\log(\frac{1-\Theta_2}{1-\Theta_1})}{\log(\frac{\Theta_2}{\Theta_1}) - \log(\frac{1-\Theta_2}{1-\Theta_1})} < r_k < \frac{a - k\log(\frac{1-\Theta_2}{1-\Theta_1})}{\log(\frac{\Theta_2}{\Theta_1}) - \log(\frac{1-\Theta_2}{1-\Theta_1})} \quad \text{if } \Theta_2 < \Theta_1 \tag{8}$$

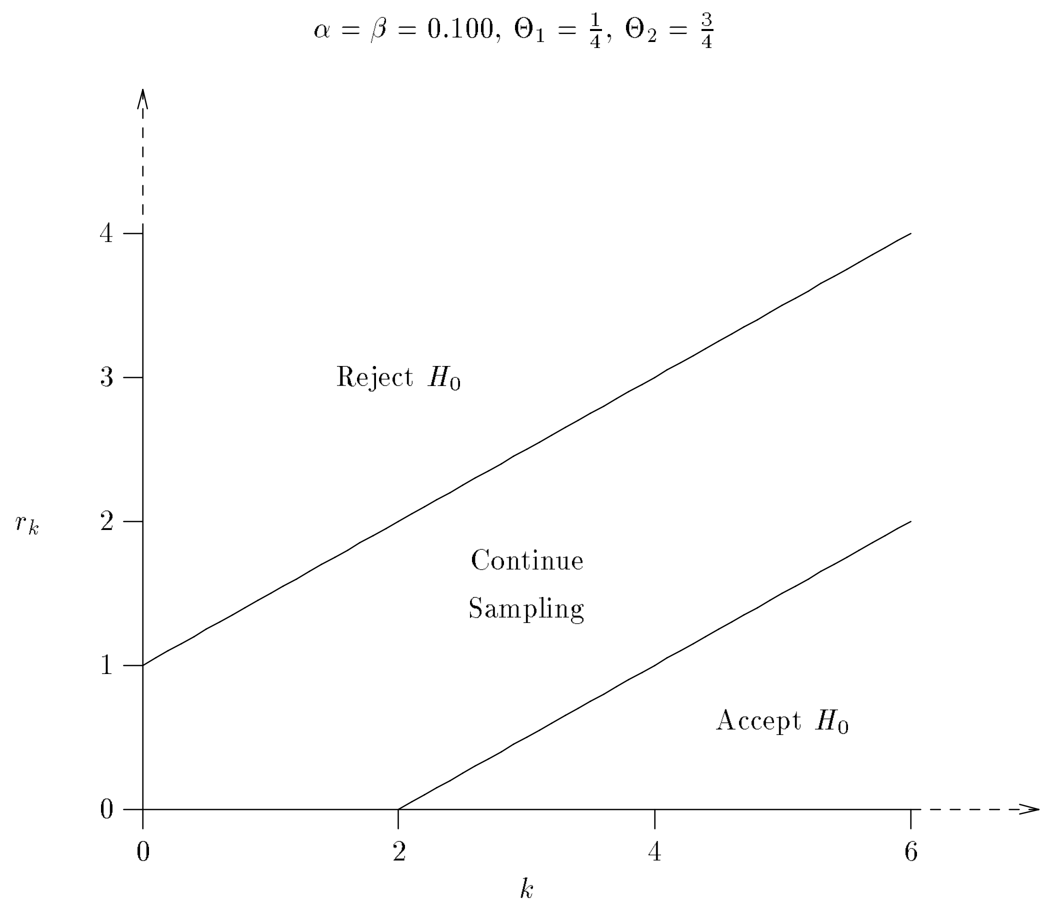

Figure 1: $r_k = \sum_{i=1}^{k} x_i$ vs. k

5.1 Operating Characteristics Curve

The outcome of sequential inspection is a decision to accept or reject a test set on the basis of evidence of a random sample. Hence, the primary consideration in the choice of the constants that determine the plan becomes the relative frequency with which the plan will lead to the acceptance of the test set that may be submitted for inspection. Given only the inspection plan, the probability L_p of accepting a test set that might not kill a given fraction p of the total mutations is mathematically determinable. When L_p is plotted on a vertical axis with p on the horizontal axis, the resulting curve is known as the operating characteristic (OC) curve, as shown in figure 2.

In practice, the choice of an OC curve always involves a gradation of preferences, a compromise between the number of incorrect decisions that can be tolerated, and the amount of inspection time that can be undertaken.

For a sequential inspection procedure, this compromise is affected in the following manner. In the scale of test set quality (as measured by a fraction live mutants p), two values p_1 and p_2 are chosen. These values define good and bad test sets in the sense that the user has a strong preference for accepting test sets of quality p_1 or better, and for rejecting test sets of

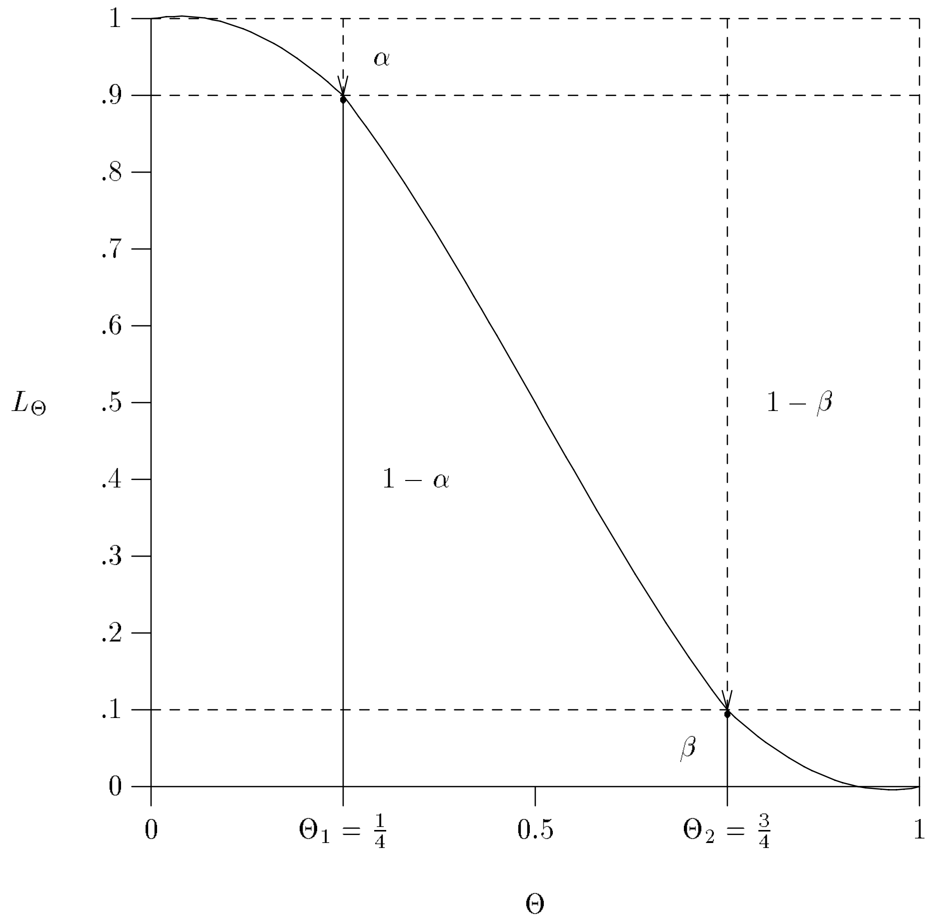

Figure 2: OC Curve for $p_1 = \frac{1}{4}$, $p_2 = \frac{3}{4}, \alpha = 0.10$, and $\beta = 0.10$

quality p_2 or worse. Quality here is the determination of the number of mutations a test set will distinguish.

For test sets with quality between zero and p_1, as well as between p_2 and unity, the user wants only small risks of making incorrect decisions. For test sets with quality between p_1 and p_2, he is less concerned over whether they are accepted or rejected. Thus the user may wish to tolerate a probability not greater than α of rejecting test sets of quality p_1 or better, and he may wish to tolerate a probability not greater than β of accepting test set of quality p_2 or worse. When the quantities α, β, p_1, and p_2 are selected, the three constants (a common slope and two intercepts for determining two parallel lines) defining a sequential inspection plan can be computed. The OC curve as well as the criterion of inspection are completely determined by α, β, p_1, and p_2.

If the user, in order to get the desired protection, is willing to pay for the amount of inspection required by the plan, it is put into operation. Otherwise, the constants α, β, p_1, and p_2 must be adjusted to give less protection at less cost. Also, the average number of samples will show the relationship between the quality of the test set p and the amount of inspection required on the average to reach a decision. It is obvious that the least inspection is required for very good or very bad test sets, and the most for "in-between" test sets.

5.2 Construction of SPRT: a software quality testing problem

Let us have a software test set whose quality testing will be carried out by a sequential procedure and mutation analysis. A sequential plan is completely determined by four numerical quantities that the user must select. These quantities specify:

1. What he considers to be a good (adequate) software test set
2. What he considers to be a bad (inadequate) software test set
3. What risk he is willing to run of rejecting a good software test set
4. What risk he is willing to run of accepting a bad software test set

Selecting these four preliminary quantities amounts to selecting the points on the OC curve, as (1) and (3) define one point, while (2) and (4) define another.

Even with the most economical inspection plan, a prohibitive amount of inspection may be required to obtain the desired quality and protection. In order to reduce the amount of inspection to an acceptable level, it may be necessary to make adjustments either in the quality tolerance limits or in the size of the risks. Hence, the quantities which completely determine the sequential inspection plan are

1. p_1: the "acceptable" quality limit for the software test set expressed as a fraction of live mutants.[3]
2. p_2: the "unacceptable" quality limit for the software test set expressed as a fraction of live mutants.
3. α: the maximum risk or probability of mistakenly rejecting test sets of quality p_1.
4. β: the maximum risk or probability of mistakenly accepting test sets of quality p_2.

A sequential sampling plan determined by these quantities accomplishes the following (see OC Curve):

1. On the average, software test sets resulting in a fraction live mutants p_1 are rejected[4] with a relative frequency equal to α. If test sets are submitted that leave a fraction live mutants less than p_1, they are rejected with a relative frequency less than α.
2. On the average, software test sets resulting in a fraction live mutants p_2 are accepted with a relative frequency equal to β. If test sets are submitted that leave a fraction live mutants greater than p_2, they are accepted with a relative frequency less than β.
3. Test sets leaving a fraction live mutants between p_1 and p_2 are accepted with a relative frequency between 1-α to β as the fraction of live mutants increases from p_1 to p_2.

[3] Note that this corresponds to the notion of "defective products" in traditional sequential decision procedures.

[4] To be discarded or augmented and tried again.

The special merit of a sequential plan is this: it accomplishes these purposes with substantially less inspection than other plans available.

The criterion for accepting or rejecting a software test data set can be given in sequential inspection by two parallel straight lines which are uniquely determined by α, β, p_1, and p_2. The equations of the lines may be written as

$$d_1 = -h_1 + \mathcal{S} * n \tag{9}$$

$$d_2 = h_2 + \mathcal{S} * n \tag{10}$$

where d stands for the number of defects; n, number of items; $\mathcal{S}$, the slope; and $-h_1$ and h_2, the intercepts. The slopes and intercepts are given by the following:

$$h_1 = \frac{\ln(\frac{1-\alpha}{\beta})}{\ln(\frac{p2}{p_1})(\frac{1-p_1}{1-p_2})} \tag{11}$$

$$h_2 = \frac{\ln(\frac{1-\beta}{\alpha})}{\ln(\frac{p2}{p_1})(\frac{1-p_1}{1-p_2})} \tag{12}$$

$$\mathcal{S} = \frac{\ln(\frac{1-p_1}{1-p_2})}{\ln(\frac{p2}{p_1})(\frac{1-p_1}{1-p_2})} \tag{13}$$

In summary, these may be computed as follows. Calculate

$$g_1 = \ln(\frac{p_2}{p_1}) \tag{14}$$

$$g_2 = \ln(\frac{1-p_1}{1-p_2}) \tag{15}$$

$$a = \ln(\frac{1-\beta}{\alpha}) \tag{16}$$

$$b = \ln(\frac{1-\alpha}{\beta}) \tag{17}$$

$$h_1 = \frac{b}{g_1+g_2} \tag{18}$$

$$h_2 = \frac{a}{g_1+g_2} \tag{19}$$

$$\mathcal{S} = \frac{g_2}{g_1+g_2} \tag{20}$$

In these computations, p_2 must be greater than p_1 and $\alpha + \beta$ must be less than one. Hence all the above calculated quantities are necessarily positive. The slope, $\mathcal{S}$, is always between p_1 and p_2. The quantities h_1 and h_2 are equal if and only if $\alpha = \beta$.

5.3 An Illustration (Applied to Mutation Testing)

After extensive *a priori* experimentation, percentages were decided as a suitable measure of good quality. That is, a certain test set is assumed to be satisfactory if the number of live mutants it leaves is not more than 100% $\times p_1$. The test set is not acceptable if the percentage is 100% $\times p_2$ or more. The program user is willing to take "b" chances in 100 of accepting an inadequate test set ($\beta = \frac{b}{100}$) and "a" chances in 100 of rejecting an adequate test set ($\alpha = \frac{a}{100}$).

As an example, consider if $p_1 = 0.005$, $p_2 = 0.3$, $\alpha = 0.01$, and $\beta = 0.02$, then $d_2 = 1.03 + 0.08n$ and $d_1 = -0.88 + 0.08n$. This would give the characteristics shown in figure 3.

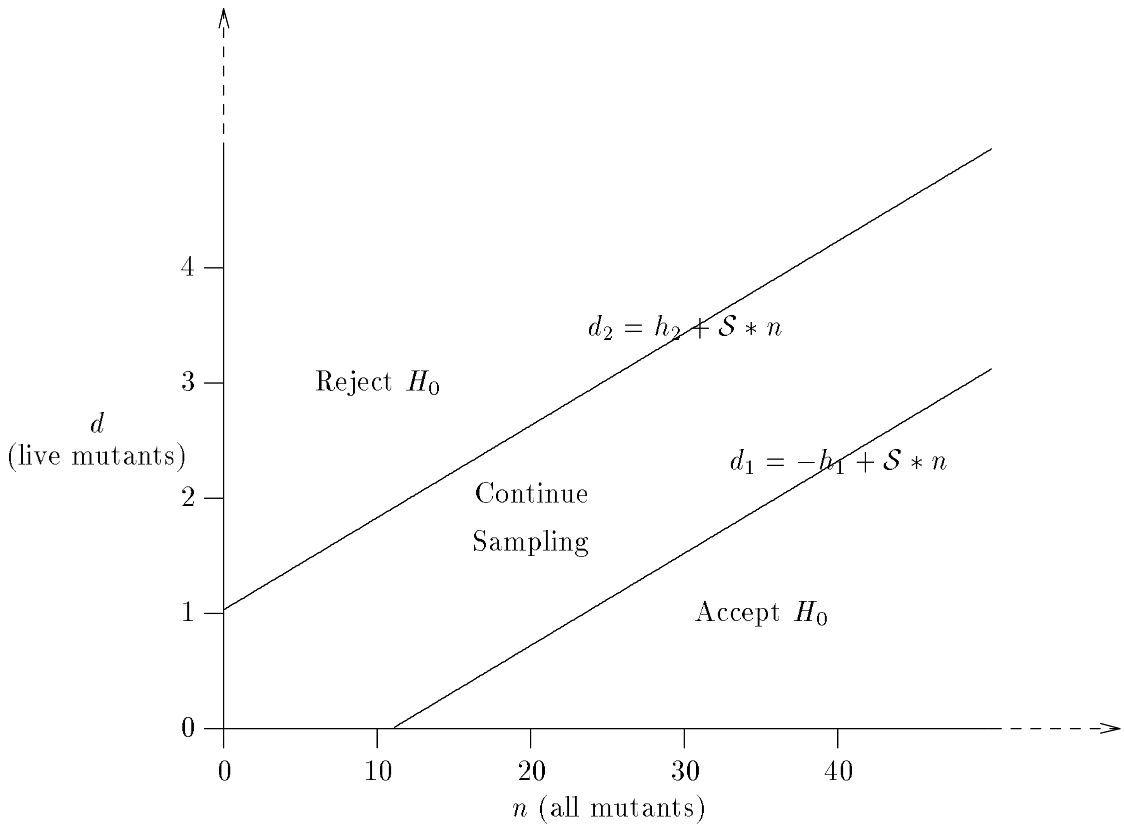

Figure 3: SPRT Graph for the Illustration of Section 5.3

Note that:

1. if the software is of such a nature that an inadequate test set may cause disaster[5], it is crucial not to accept an inadequate test set and the risk of such an error must be set very low.

2. if it is important not to reject an adequate test set, the risk of such an error must be set very low.

3. taking observations in groups is more desirable sometimes than one observation at a time.

For example, it may be convenient to examine 25 mutants at a time. In that case, the number of unkilled mutants (defective) found in the first group of 25 mutants is plotted against 25 on the same figure.

[5] By failing to expose errors.

To avoid any possible oversight of a decision rule because of grouping rather than examination on an item by item basis, one can decrease group size as one approaches prospective boundary situations, i.e., as the likelihood of a decision increases.

Remark: The same illustration can be replicated by taking a different α (or 1-α = confidence of the test) and β (or 1-β = power of the test). One can then generate a different set of decision rules for the same given p_1, and p_2.

6 Experiments with MOTHRA

We used the MOTHRA test environment to perform tests using the approach described in the previous section. The following experimental results were obtained for several sensitivity analyses of variations on the input-data α, β, Θ_1 and Θ_2 where changes were performed on one variable at a time. The discussion of the results is with respect to these changes, and in particular with the changes to Θ_1 (corresponding to mutation scores). This helps to illustrate the sensitivity of the method in conjunction with the range of possible goals set by the tester.

The first experiment, Expt 1 (Tables 1–6), concerns a mutation-based test of a Fortran program of 60 lines performed at 10% sampling intervals. The second experiment, Expt 2 (Tables 7–12), similarly is a mutation-based test performed on a Fortran program of 34 lines for 10, 25, 50, 75 and 100 sampling percentages.

6.1 Expt 1

In Table 1, α risk is twice as much as β risk with α=0.01 and β=0.005. Θ_2 is set at 0.2. Θ_1 can vary within the range $0 < \Theta_1 < \Theta_2$. However, all results for $\Theta_1 < 0.007$ are the same as if $\Theta_1 = 0.007$, hence lower tableaus for $\Theta_1 < 0.007$ are omitted. Thus $\Theta_1 >= 0.03$ will realize the unanimous acceptance of the product for all sampling percentages. The greatest shifts occur between $0.01 < \Theta_1 < 0.03$ which may be further subdivided for microscopic viewing. However, whereas the trend of decision along the sampling percentages column should pursue an R-C-A (Reject-Continue-Accept) trend, an earlier acceptance leading to a continue-sampling afterwards should still not be alarming. This is due to the fact that the analyst has not been able to choose a test case to kill mutants at a rate compatible with the advent of new mutants at the next sampling percentage. Conversely, a rejection message at a later sampling percentage after having already accepted is simply a sign of poor performance of the test case, whose ability to kill mutants may decrease as new mutants are enabled. This effect occurs when individual mutants are sampled as they are not uniform in behavior in the small. Thus, the authors believe, based on experience and observation, that larger samples of mutants behave closer to uniformity and do not exhibit this behavior.

In brief, the anomaly of test case choice should be studied if other than an R-C-A decision trend is observed with respect to the sampling percentage increase. The analyst may and should be able to halt and accept the software after 10% sampling at $\Theta_1 = 0.02$ and given α, β, and Θ_2, despite a unanimous acceptance at $\Theta_1 = 0.03$.

Table 2 is similar to Table 1 except that $\Theta_2 = 0.3$ instead of $\Theta_2 = 0.2$. Uniform acceptance is achieved at a more strict $\Theta_1 = 0.006$ than a less demanding $\Theta_1 = 0.03$. The analyst can stop after 10% sampling given a set of data at the $\Theta_1 = 0.003$ level.

In Table 3, α and β risks are identical, each with 1% and $\Theta_2 = 0.2$. The range $0.0025 < \Theta_1 < 0.03$ gives the interval of tableaus, where $\Theta_1 = 0.03$ is the overall acceptance level. The analyst may stop the experiment at $\Theta_1 = 0.02$. The doubling of β risk from 0.005 to 0.01 has not affected the results as observed.

In Table 4, where α and β risks are still identical with $\Theta_2 = 0.3$, a unanimous acceptance is obtained at $\Theta_1 = 0.006$, as in Table 2. The analyst may halt at $\Theta_1 = 0.003$, accepting the software.

In Table 5, the β risk is twice that of α risk, with $\Theta_2 = 0.2$, but the results are the same as those in Table 3. The change in the risk ratio has not altered anything. At $\Theta_1 = 0.03$, we achieve unanimous acceptance.

Table 6 results are identical to those of Table 4.

6.2 Expt 2

In Table 7, the values $\Theta_2 = 0.2$, $\alpha = 2\beta$, $\Theta_1 < 0.004$ will give a unanimous rejection for all sampling percentages. Hence we choose to investigate $0.0003 < \Theta_1 < 0.002$ where $\Theta_1 >= 0.02$ yields unanimous acceptance. The product is gradually granted total acceptance as sampling percentages increase to where $\Theta_1 = 0.02$. According to our model, one may halt the experiment at $\Theta_1 = 0.003$ and decide to accept after 50% sampling. The R-C-A trend is more consistent in this test (as will also be observed in the following tables) because of better choices of test cases.

Note that $\Theta_1 = 0.02$ possesses all the circumstances for an ideal rejection scenario. Again, the fact that the analyst rejects at 100% sampling for $\Theta_1 = 0.0045$ after sampling at 75% sampling is possible due to an inherent property of the data set (α, β, Θ_1 and Θ_2) in that the program cannot kill mutants as fast as it should with the number of mutants increasing due to sampling percentage increase.

In Table 8, for the same $\alpha = 2\beta$, $\Theta_2 = 0.3$ will tighten the restraint on Θ_1 for acceptance of the product. This time, the tester can accept at $\Theta_1 = 0.003$, instead of the former $\Theta_1 = 0.02$ after 50% sampling.

In Table 9, for $\alpha = \beta$, $\Theta_2 = 0.2$ as in Table 7, $\Theta_1 = 0.02$ makes the acceptance of the software unanimous, and one may stop and accept at $\Theta_1 = 0.003$.

In Tables 10 and 11, results are identical to those in Table 9. Table 12 results are the same as in Table 10.

7 Conclusion

The authors believe this statistical sequential testing is sound and scientific with a simple *good* vs. a simple *bad* hypothesis specified by Type I and II error probabilities. It is also possible to use this method to conduct sensitivity analyses by varying the values of $\Theta_1, \Theta_2, \alpha$, and β and observing the resultant Operating Characteristics curve. Overall, this novel approach in mutation-based testing of software appears to be superior to the deterministic traditional approach. It provides more information about the expected reliability of the software under test, and it requires less resource usage to perform non-critical testing.

When compared with our previous, conventional technique, the decision procedure in the proposed method should prove to be time-saving and cost effective. In the most extreme cases, the given procedure becomes the conventional procedure, requiring 100% mutation scores for acceptance.

As was shown in experiment 2, with better management of test cases, if a target mutation score of 0.90 is taken, then one accepts the program after performing 50% sampling. If a mutation score of 0.95 is assumed, then one cannot accept the software under test in this experiment (with the test data used) because that score is not achieved; considerable additional effort would be required to achieve a higher absolute mutation score. However, in the proposed statistical sequential procedure, one can observe in tables 7–12 for $\Theta_1 = 0.1$ (corresponding to a mutation score of 90%) that for an arbitrary α, β, and Θ_2, the product is accepted at only 10% sampling.

There are two major advantages to using this procedure as opposed to the conventional rule-of-thumb 90% target mutation score. Both are illustrated by the results of experiment 2:

1. Quality Aspect. This software is accepted at $\Theta_1 = 0.02$ ($\alpha = 0.01, \beta = 0.003, \Theta_2 = 0.2$), corresponding to an absolute mutation score of 98%. With the *ad hoc* approach, it is not possible to achieve even a 95% absolute mutation score without expanding the test data set or expending considerable effort identifying equivalent mutations. The statistical hypothesis testing approach takes into account the one-sided accumulation of the error of estimation.

2. Savings Aspect. There is no need to perform tests on the remaining sample percentages after 10% for the corresponding tableau with $\Theta_1 = 0.1$ (arbitrary α, β, Θ_2). This advantage is more pronounced if the target acceptance mutation score is high, e.g. 98%. The savings increases as the size of program under test increases, because the number of mutants increases at a rate proportional to the square of the size of the program.

Therefore, in many situations that do not require testing to 100% confidence, such as intermediate testing during development, this method may provide sufficient confidence in the software at greatly reduced cost to the tester.

8 Acknowledgements

We would like to thank Wynne Hsu for helping to obtain some of the experimental data presented here.

Table 1: Experiment 1

$\alpha = 0.0100$ $\beta = 0.0050$ $\Theta_2 = 0.2000$

Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0070$					
10	298	16.57	19.34	20	Reject
20	593	34.43	37.20	44	Reject
30	892	52.54	55.31	72	Reject
40	1185	70.28	73.06	94	Reject
50	1487	88.57	91.35	117	Reject
60	1781	106.38	109.15	132	Reject
70	2076	124.25	127.02	155	Reject
80	2373	142.23	145.00	176	Reject
90	2670	160.22	162.99	197	Reject
100	2966	178.15	180.92	213	Reject
$\Theta_1 = 0.0075$					
10	298	16.85	19.68	20	Reject
20	593	35.03	37.86	44	Reject
30	892	53.45	56.28	72	Reject
40	1185	71.51	74.34	94	Reject
50	1487	90.12	92.95	117	Reject
60	1781	108.24	111.06	132	Reject
70	2076	126.41	129.24	155	Reject
80	2373	144.72	147.54	176	Reject
90	2670	163.02	165.84	197	Reject
100	2966	181.26	184.08	213	Reject
$\Theta_1 = 0.0080$					
10	298	17.13	20.01	20	Continue Sampling
20	593	35.61	38.49	44	Reject
30	892	54.34	57.22	72	Reject
40	1185	72.69	75.57	94	Reject
50	1487	91.61	94.49	117	Reject
60	1781	110.03	112.90	132	Reject
70	2076	128.50	131.38	155	Reject
80	2373	147.11	149.99	176	Reject
90	2670	165.71	168.59	197	Reject
100	2966	184.26	187.14	213	Reject

$\alpha = 0.0100$		$\beta = 0.0050$			$\Theta_2 = 0.2000$
Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0085$					
10	298	17.39	20.32	20	Continue Sampling
20	593	36.16	39.10	44	Reject
30	892	55.19	58.12	72	Reject
40	1185	73.83	76.76	94	Reject
50	1487	93.05	95.98	117	Reject
60	1781	111.75	114.68	132	Reject
70	2076	130.52	133.45	155	Reject
80	2373	149.42	152.35	176	Reject
90	2670	168.32	171.25	197	Reject
100	2966	187.15	190.08	213	Reject
$\Theta_1 = 0.0090$					
10	298	17.65	20.63	20	Continue Sampling
20	593	36.70	39.68	44	Reject
30	892	56.01	58.99	72	Reject
40	1185	74.93	77.92	94	Reject
50	1487	94.44	97.42	117	Reject
60	1781	113.43	116.41	132	Reject
70	2076	132.48	135.46	155	Reject
80	2373	151.66	154.64	176	Reject
90	2670	170.84	173.82	197	Reject
100	2966	189.96	192.94	213	Reject
$\Theta_1 = 0.0095$					
10	298	17.90	20.93	20	Continue Sampling
20	593	37.22	40.26	44	Reject
30	892	56.81	59.84	72	Reject
40	1185	76.01	79.04	94	Reject
50	1487	95.79	98.82	117	Reject
60	1781	115.05	118.08	132	Reject
70	2076	134.37	137.41	155	Reject
80	2373	153.83	156.86	176	Reject
90	2670	173.29	176.32	197	Reject
100	2966	192.68	195.71	213	Reject

$\alpha = 0.0100$ $\beta = 0.0050$ $\Theta_2 = 0.2000$

Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0100$					
10	298	18.14	21.22	20	Continue Sampling
20	593	37.73	40.81	44	Reject
30	892	57.59	60.67	72	Reject
40	1185	77.05	80.13	94	Reject
50	1487	97.10	100.18	117	Reject
60	1781	116.63	119.71	132	Reject
70	2076	136.22	139.30	155	Reject
80	2373	155.94	159.02	176	Reject
90	2670	175.66	178.74	197	Reject
100	2966	195.32	198.40	213	Reject
$\Theta_1 = 0.0200$					
10	298	22.03	25.97	20	Accept
20	593	45.92	49.87	44	Accept
30	892	70.14	74.09	72	Continue Sampling
40	1185	93.87	97.82	94	Continue Sampling
50	1487	118.33	122.28	117	Accept
60	1781	142.15	146.09	132	Accept
70	2076	166.04	169.99	155	Accept
80	2373	190.10	194.04	176	Accept
90	2670	214.15	218.10	197	Accept
100	2966	238.13	242.07	213	Accept
$\Theta_1 = 0.0300$					
10	298	24.95	29.68	20	Accept
20	593	52.15	56.88	44	Accept
30	892	79.71	84.45	72	Accept
40	1185	106.73	111.46	94	Accept
50	1487	134.57	139.31	117	Accept
60	1781	161.68	166.41	132	Accept
70	2076	188.88	193.61	155	Accept
80	2373	216.27	221.00	176	Accept
90	2670	243.65	248.38	197	Accept
100	2966	270.94	275.67	213	Accept

Table 2: Experiment 1

$\alpha = 0.0100$		$\beta = 0.0050$			$\Theta_2 = 0.3000$
Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0015$					
10	298	17.79	19.54	20	Reject
20	593	36.32	38.07	44	Reject
30	892	55.10	56.85	72	Reject
40	1185	73.51	75.26	94	Reject
50	1487	92.48	94.23	117	Reject
60	1781	110.95	112.70	132	Reject
70	2076	129.49	131.24	155	Reject
80	2373	148.15	149.89	176	Reject
90	2670	166.80	168.55	197	Reject
100	2966	185.40	187.15	213	Reject
$\Theta_1 = 0.0020$					
10	298	18.71	20.56	20	Continue Sampling
20	593	38.21	40.06	44	Reject
30	892	57.98	59.82	72	Reject
40	1185	77.35	79.19	94	Reject
50	1487	97.31	99.16	117	Reject
60	1781	116.75	118.59	132	Reject
70	2076	136.25	138.09	155	Reject
80	2373	155.88	157.72	176	Reject
90	2670	175.51	177.36	197	Reject
100	2966	195.08	196.92	213	Reject
$\Theta_1 = 0.0025$					
10	298	19.50	21.42	20	Continue Sampling
20	593	39.82	41.74	44	Reject
30	892	60.41	62.34	72	Reject
40	1185	80.60	82.52	94	Reject
50	1487	101.40	103.32	117	Reject
60	1781	121.65	123.57	132	Reject
70	2076	141.97	143.90	155	Reject
80	2373	162.43	164.35	176	Reject
90	2670	182.89	184.81	197	Reject
100	2966	203.28	205.20	213	Reject

$\alpha = 0.0100$ $\beta = 0.0050$ $\Theta_2 = 0.3000$

Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0030$					
10	298	20.19	22.18	20	Accept
20	593	41.23	43.22	44	Reject
30	892	62.55	64.55	72	Reject
40	1185	83.45	85.44	94	Reject
50	1487	104.99	106.98	117	Reject
60	1781	125.96	127.95	132	Reject
70	2076	147.00	148.99	155	Reject
80	2373	168.18	170.17	176	Reject
90	2670	189.36	191.36	197	Reject
100	2966	210.47	212.47	213	Reject
$\Theta_1 = 0.0035$					
10	298	20.81	22.86	20	Accept
20	593	42.49	44.55	44	Continue Sampling
30	892	64.47	66.53	72	Reject
40	1185	86.01	88.07	94	Reject
50	1487	108.21	110.27	117	Reject
60	1781	129.83	131.88	132	Reject
70	2076	151.51	153.57	155	Reject
80	2373	173.34	175.40	176	Reject
90	2670	195.18	197.24	197	Continue Sampling
100	2966	216.94	219.00	213	Accept
$\Theta_1 = 0.0040$					
10	298	21.37	23.49	20	Accept
20	593	43.65	45.77	44	Continue Sampling
30	892	66.23	68.34	72	Reject
40	1185	88.35	90.47	94	Reject
50	1487	111.16	113.28	117	Reject
60	1781	133.36	135.48	132	Accept
70	2076	155.64	157.75	155	Accept
80	2373	178.06	180.18	176	Accept
90	2670	200.49	202.61	197	Accept
100	2966	222.85	224.96	213	Accept

$\alpha = 0.0100$ $\beta = 0.0050$ $\Theta_2 = 0.3000$

Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0045$					
10	298	21.89	24.07	20	Accept
20	593	44.72	46.89	44	Accept
30	892	67.85	70.02	72	Reject
40	1185	90.52	92.69	94	Reject
50	1487	113.88	116.06	117	Reject
60	1781	136.63	138.80	132	Accept
70	2076	159.45	161.62	155	Accept
80	2373	182.43	184.60	176	Accept
90	2670	205.41	207.58	197	Accept
100	2966	228.31	230.48	213	Accept
$\Theta_1 = 0.0050$					
10	298	22.38	24.61	20	Accept
20	593	45.71	47.94	44	Accept
30	892	69.36	71.59	72	Reject
40	1185	92.54	94.76	94	Continue Sampling
50	1487	116.43	118.65	117	Continue Sampling
60	1781	139.68	141.91	132	Accept
70	2076	163.01	165.24	155	Accept
80	2373	186.51	188.73	176	Accept
90	2670	210.00	212.22	197	Accept
100	2966	233.41	235.63	213	Accept
$\Theta_1 = 0.0055$					
10	298	22.84	25.11	20	Accept
20	593	46.65	48.93	44	Accept
30	892	70.79	73.06	72	Continue Sampling
40	1185	94.44	96.71	94	Accept
50	1487	118.82	121.09	117	Accept
60	1781	142.55	144.82	132	Accept
70	2076	166.36	168.64	155	Accept
80	2373	190.34	192.61	176	Accept
90	2670	214.31	216.59	197	Accept
100	2966	238.21	240.48	213	Accept

$\alpha = 0.0100$ $\beta = 0.0050$ $\Theta_2 = 0.3000$

Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0060$					
10	298	23.27	25.59	20	Accept
20	593	47.54	49.86	44	Accept
30	892	72.14	74.46	72	Accept
40	1185	96.24	98.56	94	Accept
50	1487	121.08	123.40	117	Accept
60	1781	145.27	147.59	132	Accept
70	2076	169.54	171.86	155	Accept
80	2373	193.97	196.29	176	Accept
90	2670	218.40	220.72	197	Accept
100	2966	242.75	245.07	213	Accept

Table 3: Experiment 1

$\alpha = 0.0100$ $\beta = 0.0100$ $\Theta_2 = 0.2000$

Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0075$					
10	298	17.05	19.68	20	Reject
20	593	35.23	37.85	44	Reject
30	892	53.65	56.28	72	Reject
40	1185	71.71	74.33	94	Reject
50	1487	90.32	92.94	117	Reject
60	1781	108.43	111.06	132	Reject
70	2076	126.61	129.24	155	Reject
80	2373	144.91	147.54	176	Reject
90	2670	163.22	165.84	197	Reject
100	2966	181.46	184.08	213	Reject
$\Theta_1 = 0.0080$					
10	298	17.33	20.01	20	Continue Sampling
20	593	35.81	38.48	44	Reject
30	892	54.54	57.21	72	Reject
40	1185	72.89	75.57	94	Reject
50	1487	91.81	94.49	117	Reject
60	1781	110.23	112.90	132	Reject
70	2076	128.71	131.38	155	Reject
80	2373	147.31	149.99	176	Reject
90	2670	165.92	168.59	197	Reject
100	2966	184.46	187.13	213	Reject

$\alpha = 0.0100$		$\beta = 0.0100$			$\Theta_2 = 0.2000$
Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0085$					
10	298	17.60	20.32	20	Continue Sampling
20	593	36.37	39.09	44	Reject
30	892	55.39	58.12	72	Reject
40	1185	74.04	76.76	94	Reject
50	1487	93.25	95.98	117	Reject
60	1781	111.96	114.68	132	Reject
70	2076	130.73	133.45	155	Reject
80	2373	149.63	152.35	176	Reject
90	2670	168.52	171.25	197	Reject
100	2966	187.36	190.08	213	Reject
$\Theta_1 = 0.0090$					
10	298	17.86	20.63	20	Continue Sampling
20	593	36.91	39.68	44	Reject
30	892	56.22	58.99	72	Reject
40	1185	75.14	77.92	94	Reject
50	1487	94.65	97.42	117	Reject
60	1781	113.63	116.41	132	Reject
70	2076	132.69	135.46	155	Reject
80	2373	151.87	154.64	176	Reject
90	2670	171.05	173.82	197	Reject
100	2966	190.16	192.94	213	Reject
$\Theta_1 = 0.0095$					
10	298	18.11	20.93	20	Continue Sampling
20	593	37.44	40.26	44	Reject
30	892	57.02	59.84	72	Reject
40	1185	76.22	79.04	94	Reject
50	1487	96.00	98.82	117	Reject
60	1781	115.26	118.08	132	Reject
70	2076	134.59	137.40	155	Reject
80	2373	154.04	156.86	176	Reject
90	2670	173.50	176.32	197	Reject
100	2966	192.89	195.71	213	Reject

$\alpha = 0.0100$ $\beta = 0.0100$ $\Theta_2 = 0.2000$

Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0100$					
10	298	18.36	21.22	20	Continue Sampling
20	593	37.95	40.81	44	Reject
30	892	57.80	60.67	72	Reject
40	1185	77.26	80.13	94	Reject
50	1487	97.32	100.18	117	Reject
60	1781	116.84	119.71	132	Reject
70	2076	136.43	139.30	155	Reject
80	2373	156.16	159.02	176	Reject
90	2670	175.88	178.74	197	Reject
100	2966	195.54	198.40	213	Reject
$\Theta_1 = 0.0200$					
10	298	22.30	25.97	20	Accept
20	593	46.20	49.87	44	Accept
30	892	70.42	74.08	72	Continue Sampling
40	1185	94.15	97.82	94	Accept
50	1487	118.61	122.28	117	Accept
60	1781	142.42	146.09	132	Accept
70	2076	166.32	169.98	155	Accept
80	2373	190.37	194.04	176	Accept
90	2670	214.43	218.10	197	Accept
100	2966	238.40	242.07	213	Accept
$\Theta_1 = 0.0300$					
10	298	25.28	29.68	20	Accept
20	593	52.48	56.87	44	Accept
30	892	80.05	84.44	72	Accept
40	1185	107.06	111.46	94	Accept
50	1487	134.91	139.30	117	Accept
60	1781	162.01	166.41	132	Accept
70	2076	189.21	193.61	155	Accept
80	2373	216.60	220.99	176	Accept
90	2670	243.98	248.38	197	Accept
100	2966	271.27	275.67	213	Accept

Table 4: Experiment 1

$\alpha = 0.0100$ $\quad \beta = 0.0100$ $\quad \Theta_2 = 0.3000$

Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0015$					
10	298	17.91	19.53	20	Reject
20	593	36.44	38.07	44	Reject
30	892	55.23	56.85	72	Reject
40	1185	73.63	75.26	94	Reject
50	1487	92.61	94.23	117	Reject
60	1781	111.08	112.70	132	Reject
70	2076	129.61	131.24	155	Reject
80	2373	148.27	149.89	176	Reject
90	2670	166.93	168.55	197	Reject
100	2966	185.52	187.15	213	Reject
$\Theta_1 = 0.0020$					
10	298	18.84	20.56	20	Continue Sampling
20	593	38.34	40.06	44	Reject
30	892	58.11	59.82	72	Reject
40	1185	77.48	79.19	94	Reject
50	1487	97.44	99.15	117	Reject
60	1781	116.88	118.59	132	Reject
70	2076	136.38	138.09	155	Reject
80	2373	156.01	157.72	176	Reject
90	2670	175.64	177.36	197	Reject
100	2966	195.21	196.92	213	Reject
$\Theta_1 = 0.0025$					
10	298	19.63	21.42	20	Continue Sampling
20	593	39.95	41.74	44	Reject
30	892	60.55	62.34	72	Reject
40	1185	80.73	82.52	94	Reject
50	1487	101.53	103.32	117	Reject
60	1781	121.79	123.57	132	Reject
70	2076	142.11	143.89	155	Reject
80	2373	162.57	164.35	176	Reject
90	2670	183.02	184.81	197	Reject
100	2966	203.41	205.20	213	Reject

$\alpha = 0.0100$		$\beta = 0.0100$		$\Theta_2 = 0.3000$	
Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0030$					
10	298	20.33	22.18	20	Accept
20	593	41.37	43.22	44	Reject
30	892	62.69	64.55	72	Reject
40	1185	83.59	85.44	94	Reject
50	1487	105.13	106.98	117	Reject
60	1781	126.10	127.95	132	Reject
70	2076	147.14	148.99	155	Reject
80	2373	168.32	170.17	176	Reject
90	2670	189.50	191.35	197	Reject
100	2966	210.61	212.47	213	Reject
$\Theta_1 = 0.0035$					
10	298	20.95	22.86	20	Accept
20	593	42.64	44.55	44	Continue Sampling
30	892	64.62	66.53	72	Reject
40	1185	86.16	88.07	94	Reject
50	1487	108.36	110.27	117	Reject
60	1781	129.97	131.88	132	Reject
70	2076	151.66	153.57	155	Reject
80	2373	173.49	175.40	176	Reject
90	2670	195.32	197.24	197	Continue Sampling
100	2966	217.08	219.00	213	Accept
$\Theta_1 = 0.0040$					
10	298	21.52	23.49	20	Accept
20	593	43.80	45.76	44	Continue Sampling
30	892	66.38	68.34	72	Reject
40	1185	88.50	90.47	94	Reject
50	1487	111.31	113.27	117	Reject
60	1781	133.51	135.48	132	Accept
70	2076	155.79	157.75	155	Accept
80	2373	178.21	180.18	176	Accept
90	2670	200.64	202.61	197	Accept
100	2966	222.99	224.96	213	Accept

$\alpha = 0.0100$ $\beta = 0.0100$ $\Theta_2 = 0.3000$

Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0045$					
10	298	22.05	24.06	20	Accept
20	593	44.87	46.89	44	Accept
30	892	68.00	70.02	72	Reject
40	1185	90.67	92.69	94	Reject
50	1487	114.04	116.05	117	Reject
60	1781	136.78	138.80	132	Accept
70	2076	159.60	161.62	155	Accept
80	2373	182.58	184.60	176	Accept
90	2670	205.56	207.58	197	Accept
100	2966	228.46	230.48	213	Accept
$\Theta_1 = 0.0050$					
10	298	22.54	24.60	20	Accept
20	593	45.87	47.94	44	Accept
30	892	69.52	71.59	72	Reject
40	1185	92.70	94.76	94	Continue Sampling
50	1487	116.58	118.65	117	Continue Sampling
60	1781	139.84	141.90	132	Accept
70	2076	163.17	165.24	155	Accept
80	2373	186.66	188.73	176	Accept
90	2670	210.15	212.22	197	Accept
100	2966	233.57	235.63	213	Accept

$\alpha = 0.0100$ $\beta = 0.0100$ $\Theta_2 = 0.3000$

Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0055$					
10	298	23.00	25.11	20	Accept
20	593	46.81	48.92	44	Accept
30	892	70.95	73.06	72	Continue Sampling
40	1185	94.60	96.71	94	Accept
50	1487	118.98	121.09	117	Accept
60	1781	142.71	144.82	132	Accept
70	2076	166.52	168.64	155	Accept
80	2373	190.50	192.61	176	Accept
90	2670	214.47	216.59	197	Accept
100	2966	238.37	240.48	213	Accept
$\Theta_1 = 0.0060$					
10	298	23.44	25.59	20	Accept
20	593	47.70	49.86	44	Accept
30	892	72.30	74.46	72	Accept
40	1185	96.40	98.56	94	Accept
50	1487	121.25	123.40	117	Accept
60	1781	145.43	147.59	132	Accept
70	2076	169.70	171.85	155	Accept
80	2373	194.13	196.29	176	Accept
90	2670	218.56	220.72	197	Accept
100	2966	242.91	245.07	213	Accept

Table 5: Experiment 1

$\alpha = 0.0100$ $\beta = 0.0200$ $\Theta_2 = 0.2000$

Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0070$					
10	298	16.95	19.33	20	Reject
20	593	34.82	37.20	44	Reject
30	892	52.93	55.31	72	Reject
40	1185	70.67	73.05	94	Reject
50	1487	88.96	91.34	117	Reject
60	1781	106.77	109.15	132	Reject
70	2076	124.63	127.01	155	Reject
80	2373	142.62	145.00	176	Reject
90	2670	160.61	162.99	197	Reject
100	2966	178.53	180.91	213	Reject

$\alpha = 0.0100$		$\beta = 0.0200$			$\Theta_2 = 0.2000$
Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0075$					
10	298	17.25	19.67	20	Reject
20	593	35.43	37.85	44	Reject
30	892	53.85	56.28	72	Reject
40	1185	71.91	74.33	94	Reject
50	1487	90.52	92.94	117	Reject
60	1781	108.63	111.06	132	Reject
70	2076	126.81	129.24	155	Reject
80	2373	145.11	147.54	176	Reject
90	2670	163.41	165.84	197	Reject
100	2966	181.65	184.08	213	Reject
$\Theta_1 = 0.0080$					
10	298	17.53	20.00	20	Continue Sampling
20	593	36.01	38.48	44	Reject
30	892	54.74	57.21	72	Reject
40	1185	73.09	75.57	94	Reject
50	1487	92.01	94.48	117	Reject
60	1781	110.43	112.90	132	Reject
70	2076	128.91	131.38	155	Reject
80	2373	147.51	149.98	176	Reject
90	2670	166.12	168.59	197	Reject
100	2966	184.66	187.13	213	Reject
$\Theta_1 = 0.0085$					
10	298	17.80	20.32	20	Continue Sampling
20	593	36.57	39.09	44	Reject
30	892	55.60	58.12	72	Reject
40	1185	74.24	76.76	94	Reject
50	1487	93.46	95.97	117	Reject
60	1781	112.16	114.68	132	Reject
70	2076	130.93	133.45	155	Reject
80	2373	149.83	152.35	176	Reject
90	2670	168.73	171.25	197	Reject
100	2966	187.56	190.08	213	Reject

$\alpha = 0.0100$ $\beta = 0.0200$ $\Theta_2 = 0.2000$

Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0090$					
10	298	18.07	20.63	20	Continue Sampling
20	593	37.12	39.68	44	Reject
30	892	56.43	58.99	72	Reject
40	1185	75.35	77.91	94	Reject
50	1487	94.86	97.42	117	Reject
60	1781	113.84	116.40	132	Reject
70	2076	132.90	135.46	155	Reject
80	2373	152.08	154.64	176	Reject
90	2670	171.26	173.82	197	Reject
100	2966	190.37	192.93	213	Reject
$\Theta_1 = 0.0095$					
10	298	18.32	20.93	20	Continue Sampling
20	593	37.65	40.25	44	Reject
30	892	57.24	59.84	72	Reject
40	1185	76.43	79.03	94	Reject
50	1487	96.21	98.82	117	Reject
60	1781	115.47	118.08	132	Reject
70	2076	134.80	137.40	155	Reject
80	2373	154.25	156.86	176	Reject
90	2670	173.71	176.31	197	Reject
100	2966	193.10	195.70	213	Reject
$\Theta_1 = 0.0100$					
10	298	18.57	21.22	20	Continue Sampling
20	593	38.16	40.81	44	Reject
30	892	58.02	60.67	72	Reject
40	1185	77.48	80.12	94	Reject
50	1487	97.53	100.18	117	Reject
60	1781	117.06	119.70	132	Reject
70	2076	136.65	139.29	155	Reject
80	2373	156.37	159.02	176	Reject
90	2670	176.09	178.74	197	Reject
100	2966	195.75	198.40	213	Reject

$\alpha = 0.0100$ $\beta = 0.0200$ $\Theta_2 = 0.2000$

Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0200$					
10	298	22.58	25.97	20	Accept
20	593	46.47	49.86	44	Accept
30	892	70.69	74.08	72	Continue Sampling
40	1185	94.42	97.81	94	Accept
50	1487	118.89	122.27	117	Accept
60	1781	142.70	146.09	132	Accept
70	2076	166.59	169.98	155	Accept
80	2373	190.65	194.04	176	Accept
90	2670	214.71	218.09	197	Accept
100	2966	238.68	242.07	213	Accept
$\Theta_1 = 0.0300$					
10	298	25.61	29.67	20	Accept
20	593	52.81	56.87	44	Accept
30	892	80.38	84.44	72	Accept
40	1185	107.39	111.45	94	Accept
50	1487	135.24	139.30	117	Accept
60	1781	162.34	166.41	132	Accept
70	2076	189.54	193.61	155	Accept
80	2373	216.93	220.99	176	Accept
90	2670	244.31	248.37	197	Accept
100	2966	271.60	275.67	213	Accept

Table 6: Experiment 1

$\alpha = 0.0100$		$\beta = 0.0200$			$\Theta_2 = 0.3000$
Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0015$					
10	298	18.03	19.53	20	Reject
20	593	36.56	38.07	44	Reject
30	892	55.35	56.85	72	Reject
40	1185	73.76	75.26	94	Reject
50	1487	92.73	94.23	117	Reject
60	1781	111.20	112.70	132	Reject
70	2076	129.73	131.23	155	Reject
80	2373	148.39	149.89	176	Reject
90	2670	167.05	168.55	197	Reject
100	2966	185.65	187.15	213	Reject
$\Theta_1 = 0.0020$					
10	298	18.97	20.55	20	Continue Sampling
20	593	38.47	40.05	44	Reject
30	892	58.24	59.82	72	Reject
40	1185	77.61	79.19	94	Reject
50	1487	97.57	99.15	117	Reject
60	1781	117.01	118.59	132	Reject
70	2076	136.51	138.09	155	Reject
80	2373	156.14	157.72	176	Reject
90	2670	175.77	177.35	197	Reject
100	2966	195.34	196.92	213	Reject
$\Theta_1 = 0.0025$					
10	298	19.77	21.42	20	Continue Sampling
20	593	40.09	41.74	44	Reject
30	892	60.68	62.34	72	Reject
40	1185	80.87	82.52	94	Reject
50	1487	101.67	103.32	117	Reject
60	1781	121.92	123.57	132	Reject
70	2076	142.24	143.89	155	Reject
80	2373	162.70	164.35	176	Reject
90	2670	183.16	184.81	197	Reject
100	2966	203.55	205.20	213	Reject

$\alpha = 0.0100$ $\beta = 0.0200$ $\Theta_2 = 0.3000$

Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0030$					
10	298	20.47	22.18	20	Accept
20	593	41.51	43.22	44	Reject
30	892	62.83	64.54	72	Reject
40	1185	83.73	85.44	94	Reject
50	1487	105.27	106.98	117	Reject
60	1781	126.24	127.95	132	Reject
70	2076	147.28	148.99	155	Reject
80	2373	168.46	170.17	176	Reject
90	2670	189.64	191.35	197	Reject
100	2966	210.75	212.46	213	Reject
$\Theta_1 = 0.0035$					
10	298	21.09	22.86	20	Accept
20	593	42.78	44.55	44	Continue Sampling
30	892	64.76	66.53	72	Reject
40	1185	86.30	88.07	94	Reject
50	1487	108.50	110.27	117	Reject
60	1781	130.11	131.88	132	Reject
70	2076	151.80	153.57	155	Reject
80	2373	173.63	175.40	176	Reject
90	2670	195.47	197.23	197	Continue Sampling
100	2966	217.23	218.99	213	Accept
$\Theta_1 = 0.0040$					
10	298	21.67	23.49	20	Accept
20	593	43.94	45.76	44	Continue Sampling
30	892	66.52	68.34	72	Reject
40	1185	88.65	90.47	94	Reject
50	1487	111.46	113.27	117	Reject
60	1781	133.66	135.47	132	Accept
70	2076	155.93	157.75	155	Accept
80	2373	178.36	180.18	176	Accept
90	2670	200.79	202.61	197	Accept
100	2966	223.14	224.96	213	Accept

$\alpha = 0.0100$ $\beta = 0.0200$ $\Theta_2 = 0.3000$

Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0045$					
10	298	22.20	24.06	20	Accept
20	593	45.02	46.89	44	Accept
30	892	68.15	70.02	72	Reject
40	1185	90.82	92.69	94	Reject
50	1487	114.19	116.05	117	Reject
60	1781	136.93	138.80	132	Accept
70	2076	159.76	161.62	155	Accept
80	2373	182.73	184.60	176	Accept
90	2670	205.71	207.58	197	Accept
100	2966	228.61	230.48	213	Accept
$\Theta_1 = 0.0050$					
10	298	22.69	24.60	20	Accept
20	593	46.03	47.94	44	Accept
30	892	69.68	71.59	72	Reject
40	1185	92.85	94.76	94	Continue Sampling
50	1487	116.74	118.65	117	Continue Sampling
60	1781	139.99	141.90	132	Accept
70	2076	163.33	165.24	155	Accept
80	2373	186.82	188.73	176	Accept
90	2670	210.31	212.22	197	Accept
100	2966	233.72	235.63	213	Accept
$\Theta_1 = 0.0055$					
10	298	23.16	25.11	20	Accept
20	593	46.97	48.92	44	Accept
30	892	71.11	73.06	72	Continue Sampling
40	1185	94.76	96.71	94	Accept
50	1487	119.14	121.09	117	Accept
60	1781	142.87	144.82	132	Accept
70	2076	166.68	168.63	155	Accept
80	2373	190.66	192.61	176	Accept
90	2670	214.63	216.58	197	Accept
100	2966	238.53	240.48	213	Accept
$\Theta_1 = 0.0060$					
10	298	23.60	25.59	20	Accept
20	593	47.87	49.86	44	Accept
30	892	72.46	74.45	72	Accept
40	1185	96.57	98.56	94	Accept
50	1487	121.41	123.40	117	Accept
60	1781	145.59	147.58	132	Accept
70	2076	169.86	171.85	155	Accept
80	2373	194.29	196.28	176	Accept
90	2670	218.72	220.72	197	Accept
100	2966	243.07	245.06	213	Accept

Table 7: Experiment 2

$\alpha = 0.0100$ $\beta = 0.0050$ $\Theta_2 = 0.2000$

Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0003$					
10	98	2.46	3.93	4	Reject
25	248	7.43	8.90	12	Reject
50	496	15.65	17.12	23	Reject
75	741	23.77	25.24	37	Reject
100	983	31.79	33.26	69	Reject
$\Theta_1 = 0.0004$					
10	98	2.57	4.11	4	Continue Sampling
25	248	7.76	9.30	12	Reject
50	496	16.34	17.88	23	Reject
75	741	24.82	26.35	37	Reject
100	983	33.19	34.73	69	Reject
$\Theta_1 = 0.0005$					
10	98	2.66	4.25	4	Continue Sampling
25	248	8.03	9.63	12	Reject
50	496	16.92	18.51	23	Reject
75	741	25.70	27.29	37	Reject
100	983	34.37	35.96	69	Reject
$\Theta_1 = 0.0006$					
10	98	2.74	4.38	4	Continue Sampling
25	248	8.27	9.91	12	Reject
50	496	17.42	19.06	23	Reject
75	741	26.46	28.10	37	Reject
100	983	35.39	37.03	69	Reject
$\Theta_1 = 0.0007$					
10	98	2.81	4.49	4	Continue Sampling
25	248	8.49	10.17	12	Reject
50	496	17.87	19.55	23	Reject
75	741	27.14	28.83	37	Reject
100	983	36.30	37.99	69	Reject
$\Theta_1 = 0.0008$					
10	98	2.87	4.59	4	Continue Sampling
25	248	8.68	10.40	12	Reject
50	496	18.28	20.00	23	Reject
75	741	27.76	29.49	37	Reject
100	983	37.13	38.85	69	Reject

$\alpha = 0.0100$		$\beta = 0.0050$			$\Theta_2 = 0.2000$
Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0009$					
10	98	2.93	4.69	4	Continue Sampling
25	248	8.86	10.61	12	Reject
50	496	18.65	20.41	23	Reject
75	741	28.33	30.09	37	Reject
100	983	37.89	39.65	69	Reject
$\Theta_1 = 0.0015$					
10	98	3.21	5.15	4	Continue Sampling
25	248	9.71	11.65	12	Reject
50	496	20.46	22.39	23	Reject
75	741	31.08	33.01	37	Reject
100	983	41.57	43.50	69	Reject
$\Theta_1 = 0.0020$					
10	98	3.39	5.44	4	Continue Sampling
25	248	10.27	12.32	12	Continue Sampling
50	496	21.63	23.68	23	Continue Sampling
75	741	32.86	34.91	37	Reject
100	983	43.95	45.99	69	Reject
$\Theta_1 = 0.0025$					
10	98	3.55	5.70	4	Continue Sampling
25	248	10.74	12.89	12	Continue Sampling
50	496	22.63	24.78	23	Continue Sampling
75	741	34.37	36.52	37	Reject
100	983	45.97	48.12	69	Reject
$\Theta_1 = 0.0030$					
10	98	3.68	5.92	4	Continue Sampling
25	248	11.16	13.39	12	Continue Sampling
50	496	23.51	25.75	23	Accept
75	741	35.71	37.95	37	Continue Sampling
100	983	47.76	50.00	69	Reject
$\Theta_1 = 0.0035$					
10	98	3.81	6.13	4	Continue Sampling
25	248	11.53	13.85	12	Continue Sampling
50	496	24.30	26.62	23	Accept
75	741	36.92	39.24	37	Continue Sampling
100	983	49.38	51.70	69	Reject
$\Theta_1 = 0.0040$					
10	98	3.92	6.31	4	Continue Sampling
25	248	11.87	14.27	12	Continue Sampling
50	496	25.03	27.42	23	Accept
75	741	38.03	40.42	37	Accept
100	983	50.86	53.26	69	Reject

$\alpha = 0.0100$ $\beta = 0.0050$ $\Theta_2 = 0.2000$

Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0045$					
10	98	4.02	6.49	4	Accept
25	248	12.19	14.66	12	Accept
50	496	25.71	28.17	23	Accept
75	741	39.05	41.52	37	Accept
100	983	52.24	54.70	69	Reject
$\Theta_1 = 0.0050$					
10	98	4.12	6.65	4	Accept
25	248	12.49	15.02	12	Accept
50	496	26.34	28.87	23	Accept
75	741	40.02	42.55	37	Accept
100	983	53.53	56.06	69	Reject
		⋮			
$\Theta_1 = 0.0200$					
10	98	5.83	9.77	4	Accept
25	248	17.98	21.92	12	Accept
50	496	38.06	42.01	23	Accept
75	741	57.91	61.86	37	Accept
100	983	77.51	81.46	69	Accept
$\Theta_1 = 0.1000$					
10	98	7.71	19.91	4	Accept
25	248	29.50	41.69	12	Accept
50	496	65.52	77.71	23	Accept
75	741	101.10	113.30	37	Accept
100	983	136.25	148.45	69	Accept

Table 8: Experiment 2

$\alpha = 0.0100$ $\beta = 0.0050$ $\Theta_2 = 0.3000$

Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0001$					
10	98	3.55	4.73	4	Continue Sampling
25	248	9.94	11.12	12	Reject
50	496	20.52	21.70	23	Reject
75	741	30.96	32.14	37	Reject
100	983	41.28	42.46	69	Reject
$\Theta_1 = 0.0002$					
10	98	3.87	5.15	4	Continue Sampling
25	248	10.84	12.13	12	Continue Sampling
50	496	22.36	23.65	23	Continue Sampling
75	741	33.75	35.04	37	Reject
100	983	45.00	46.29	69	Reject
$\Theta_1 = 0.0003$					
10	98	4.08	5.44	4	Accept
25	248	11.44	12.80	12	Continue Sampling
50	496	23.61	24.97	23	Accept
75	741	35.63	36.99	37	Reject
100	983	47.50	48.86	69	Reject
$\Theta_1 = 0.0004$					
10	98	4.25	5.66	4	Accept
25	248	11.91	13.32	12	Continue Sampling
50	496	24.57	25.99	23	Accept
75	741	37.08	38.50	37	Accept
100	983	49.44	50.86	69	Reject
$\Theta_1 = 0.0005$					
10	98	4.39	5.85	4	Accept
25	248	12.30	13.76	12	Accept
50	496	25.38	26.84	23	Accept
75	741	38.30	39.76	37	Accept
100	983	51.06	52.53	69	Reject
$\Theta_1 = 0.0006$					
10	98	4.51	6.01	4	Accept
25	248	12.63	14.14	12	Accept
50	496	26.07	27.58	23	Accept
75	741	39.35	40.86	37	Accept
100	983	52.47	53.97	69	Reject

$\alpha = 0.0100$ $\beta = 0.0050$ $\Theta_2 = 0.3000$

Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0007$					
10	98	4.61	6.15	4	Accept
25	248	12.93	14.48	12	Accept
50	496	26.69	28.23	23	Accept
75	741	40.29	41.83	37	Accept
100	983	53.71	55.25	69	Reject
$\Theta_1 = 0.0008$					
10	98	4.71	6.28	4	Accept
25	248	13.21	14.78	12	Accept
50	496	27.25	28.83	23	Accept
75	741	41.13	42.70	37	Accept
100	983	54.84	56.41	69	Reject
$\Theta_1 = 0.0009$					
10	98	4.80	6.40	4	Accept
25	248	13.45	15.06	12	Accept
50	496	27.77	29.37	23	Accept
75	741	41.90	43.51	37	Accept
100	983	55.87	57.47	69	Reject
$\Theta_1 = 0.0015$					
10	98	5.22	6.97	4	Accept
25	248	14.64	16.39	12	Accept
50	496	30.23	31.97	23	Accept
75	741	45.62	47.37	37	Accept
100	983	60.82	62.57	69	Reject
$\Theta_1 = 0.0020$					
10	98	5.49	7.34	4	Accept
25	248	15.41	17.25	12	Accept
50	496	31.80	33.65	23	Accept
75	741	48.00	49.84	37	Accept
100	983	64.00	65.84	69	Reject
$\Theta_1 = 0.0025$					
10	98	5.72	7.65	4	Accept
25	248	16.05	17.98	12	Accept
50	496	33.14	35.06	23	Accept
75	741	50.01	51.94	37	Accept
100	983	66.68	68.61	69	Reject

$\alpha = 0.0100$ $\beta = 0.0050$ $\Theta_2 = 0.3000$

Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0030$					
10	98	5.92	7.92	4	Accept
25	248	16.62	18.62	12	Accept
50	496	34.31	36.30	23	Accept
75	741	51.78	53.78	37	Accept
100	983	69.04	71.04	69	Accept
		⋮			
$\Theta_1 = 0.1000$					
10	98	14.33	21.65	4	Accept
25	248	42.25	49.58	12	Accept
50	496	88.42	95.75	23	Accept
75	741	134.03	141.36	37	Accept
100	983	179.09	186.41	69	Accept

Table 9: Experiment 2

$\alpha = 0.0100$ $\beta = 0.0100$ $\Theta_2 = 0.2000$

Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0003$					
10	98	2.56	3.93	4	Reject
25	248	7.53	8.90	12	Reject
50	496	15.75	17.12	23	Reject
75	741	23.87	25.24	37	Reject
100	983	31.89	33.26	69	Reject
$\Theta_1 = 0.0004$					
10	98	2.68	4.10	4	Continue Sampling
25	248	7.87	9.30	12	Reject
50	496	16.45	17.88	23	Reject
75	741	24.93	26.35	37	Reject
100	983	33.30	34.73	69	Reject
$\Theta_1 = 0.0005$					
10	98	2.77	4.25	4	Continue Sampling
25	248	8.15	9.62	12	Reject
50	496	17.03	18.51	23	Reject
75	741	25.81	27.29	37	Reject
100	983	34.48	35.96	69	Reject
$\Theta_1 = 0.0006$					
10	98	2.85	4.38	4	Continue Sampling
25	248	8.39	9.91	12	Reject
50	496	17.54	19.06	23	Reject
75	741	26.58	28.10	37	Reject
100	983	35.51	37.03	69	Reject
$\Theta_1 = 0.0007$					
10	98	2.93	4.49	4	Continue Sampling
25	248	8.60	10.17	12	Reject
50	496	17.99	19.55	23	Reject
75	741	27.26	28.83	37	Reject
100	983	36.42	37.99	69	Reject
$\Theta_1 = 0.0008$					
10	98	2.99	4.59	4	Continue Sampling
25	248	8.80	10.40	12	Reject
50	496	18.40	20.00	23	Reject
75	741	27.88	29.48	37	Reject
100	983	37.25	38.85	69	Reject

$\alpha = 0.0100$		$\beta = 0.0100$			$\Theta_2 = 0.2000$
Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0009$					
10	98	3.05	4.69	4	Continue Sampling
25	248	8.98	10.61	12	Reject
50	496	18.78	20.41	23	Reject
75	741	28.46	30.09	37	Reject
100	983	38.02	39.65	69	Reject
$\Theta_1 = 0.0015$					
10	98	3.35	5.15	4	Continue Sampling
25	248	9.85	11.65	12	Reject
50	496	20.60	22.39	23	Reject
75	741	31.21	33.01	37	Reject
100	983	41.70	43.50	69	Reject
$\Theta_1 = 0.0020$					
10	98	3.54	5.44	4	Continue Sampling
25	248	10.41	12.32	12	Continue Sampling
50	496	21.77	23.68	23	Continue Sampling
75	741	33.00	34.90	37	Reject
100	983	44.09	45.99	69	Reject
		⋮			
$\Theta_1 = 0.1000$					
10	98	8.57	19.90	4	Accept
25	248	30.35	41.69	12	Accept
50	496	66.37	77.71	23	Accept
75	741	101.96	113.29	37	Accept
100	983	137.11	148.44	69	Accept

Table 10: Experiment 1

$\alpha = 0.0100$ $\beta = 0.0100$ $\Theta_2 = 0.3000$

Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0001$					
10	98	3.63	4.73	4	Continue Sampling
25	248	10.02	11.12	12	Reject
50	496	20.60	21.70	23	Reject
75	741	31.04	32.14	37	Reject
100	983	41.36	42.46	69	Reject
$\Theta_1 = 0.0002$					
10	98	3.96	5.15	4	Continue Sampling
25	248	10.93	12.13	12	Continue Sampling
50	496	22.45	23.65	23	Continue Sampling
75	741	33.84	35.04	37	Reject
100	983	45.09	46.29	69	Reject
$\Theta_1 = 0.0003$					
10	98	4.18	5.44	4	Accept
25	248	11.53	12.80	12	Continue Sampling
50	496	23.70	24.97	23	Accept
75	741	35.72	36.99	37	Reject
100	983	47.59	48.86	69	Reject
$\Theta_1 = 0.0004$					
10	98	4.35	5.66	4	Accept
25	248	12.01	13.32	12	Accept
50	496	24.67	25.99	23	Accept
75	741	37.18	38.50	37	Accept
100	983	49.54	50.86	69	Reject
$\Theta_1 = 0.0005$					
10	98	4.49	5.85	4	Accept
25	248	12.40	13.76	12	Accept
50	496	25.48	26.84	23	Accept
75	741	38.40	39.76	37	Accept
100	983	51.17	52.53	69	Reject
$\Theta_1 = 0.0006$					
10	98	4.61	6.01	4	Accept
25	248	12.74	14.14	12	Accept
50	496	26.18	27.58	23	Accept
75	741	39.46	40.86	37	Accept
100	983	52.57	53.97	69	Reject

$\alpha = 0.0100$		$\beta = 0.0100$		$\Theta_2 = 0.3000$	
Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0007$					
10	98	4.72	6.15	4	Accept
25	248	13.04	14.47	12	Accept
50	496	26.80	28.23	23	Accept
75	741	40.39	41.83	37	Accept
100	983	53.82	55.25	69	Reject
$\Theta_1 = 0.0008$					
10	98	4.82	6.28	4	Accept
25	248	13.32	14.78	12	Accept
50	496	27.36	28.83	23	Accept
75	741	41.24	42.70	37	Accept
100	983	54.95	56.41	69	Reject
$\Theta_1 = 0.0009$					
10	98	4.91	6.40	4	Accept
25	248	13.57	15.06	12	Accept
50	496	27.88	29.37	23	Accept
75	741	42.02	43.51	37	Accept
100	983	55.98	57.47	69	Reject
$\Theta_1 = 0.0015$					
10	98	5.34	6.97	4	Accept
25	248	14.77	16.39	12	Accept
50	496	30.35	31.97	23	Accept
75	741	45.74	47.37	37	Accept
100	983	60.94	62.57	69	Reject
$\Theta_1 = 0.0020$					
10	98	5.62	7.33	4	Accept
25	248	15.54	17.25	12	Accept
50	496	31.93	33.64	23	Accept
75	741	48.13	49.84	37	Accept
100	983	64.12	65.84	69	Reject
$\Theta_1 = 0.0025$					
10	98	5.86	7.64	4	Accept
25	248	16.19	17.98	12	Accept
50	496	33.27	35.06	23	Accept
75	741	50.15	51.94	37	Accept
100	983	66.82	68.61	69	Reject

$\alpha = 0.0100$ $\beta = 0.0100$ $\Theta_2 = 0.3000$

Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0030$					
10	98	6.06	7.92	4	Accept
25	248	16.76	18.61	12	Accept
50	496	34.45	36.30	23	Accept
75	741	51.92	53.78	37	Accept
100	983	69.18	71.04	69	Accept
		⋮			
$\Theta_1 = 0.1000$					
10	98	14.84	21.65	4	Accept
25	248	42.77	49.57	12	Accept
50	496	88.94	95.74	23	Accept
75	741	134.55	141.36	37	Accept
100	983	179.60	186.41	69	Accept

Table 11: Experiment 1

$\alpha = 0.0100$ $\beta = 0.0200$ $\Theta_2 = 0.3000$

Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0001$					
10	98	3.71	4.73	4	Continue Sampling
25	248	10.11	11.12	12	Reject
50	496	20.68	21.70	23	Reject
75	741	31.13	32.14	37	Reject
100	983	41.45	42.46	69	Reject
$\Theta_1 = 0.0002$					
10	98	4.05	5.15	4	Accept
25	248	11.02	12.12	12	Continue Sampling
50	496	22.54	23.65	23	Continue Sampling
75	741	33.93	35.04	37	Reject
100	983	45.18	46.29	69	Reject
$\Theta_1 = 0.0003$					
10	98	4.27	5.44	4	Accept
25	248	11.63	12.80	12	Continue Sampling
50	496	23.80	24.96	23	Accept
75	741	35.82	36.98	37	Reject
100	983	47.69	48.86	69	Reject
$\Theta_1 = 0.0004$					
10	98	4.45	5.66	4	Accept
25	248	12.11	13.32	12	Accept
50	496	24.77	25.99	23	Accept
75	741	37.28	38.50	37	Accept
100	983	49.64	50.86	69	Reject
$\Theta_1 = 0.0005$					
10	98	4.59	5.85	4	Accept
25	248	12.50	13.76	12	Accept
50	496	25.58	26.84	23	Accept
75	741	38.50	39.76	37	Accept
100	983	51.27	52.52	69	Reject
$\Theta_1 = 0.0006$					
10	98	4.72	6.01	4	Accept
25	248	12.85	14.14	12	Accept
50	496	26.29	27.58	23	Accept
75	741	39.56	40.85	37	Accept
100	983	52.68	53.97	69	Reject

$\alpha = 0.0100$		$\beta = 0.0200$		$\Theta_2 = 0.3000$	
Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0007$					
10	98	4.83	6.15	4	Accept
25	248	13.15	14.47	12	Accept
50	496	26.91	28.23	23	Accept
75	741	40.50	41.82	37	Accept
100	983	53.93	55.25	69	Reject
$\Theta_1 = 0.0008$					
10	98	4.93	6.28	4	Accept
25	248	13.43	14.78	12	Accept
50	496	27.47	28.82	23	Accept
75	741	41.35	42.70	37	Accept
100	983	55.06	56.41	69	Reject
$\Theta_1 = 0.0009$					
10	98	5.02	6.40	4	Accept
25	248	13.68	15.06	12	Accept
50	496	27.99	29.37	23	Accept
75	741	42.13	43.51	37	Accept
100	983	56.10	57.47	69	Reject
$\Theta_1 = 0.0015$					
10	98	5.47	6.97	4	Accept
25	248	14.89	16.39	12	Accept
50	496	30.47	31.97	23	Accept
75	741	45.86	47.36	37	Accept
100	983	61.07	62.57	69	Reject
$\Theta_1 = 0.0020$					
10	98	5.75	7.33	4	Accept
25	248	15.67	17.25	12	Accept
50	496	32.06	33.64	23	Accept
75	741	48.26	49.84	37	Accept
100	983	64.25	65.84	69	Reject
$\Theta_1 = 0.0025$					
10	98	5.99	7.64	4	Accept
25	248	16.32	17.97	12	Accept
50	496	33.41	35.06	23	Accept
75	741	50.28	51.93	37	Accept
100	983	66.95	68.60	69	Reject

$\alpha = 0.0100$ $\beta = 0.0200$ $\Theta_2 = 0.3000$

Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0030$					
10	98	6.20	7.91	4	Accept
25	248	16.90	18.61	12	Accept
50	496	34.59	36.30	23	Accept
75	741	52.06	53.77	37	Accept
100	983	69.32	71.03	69	Accept
		⋮			
$\Theta_1 = 0.1000$					
10	98	15.35	21.64	4	Accept
25	248	43.28	49.57	12	Accept
50	496	89.45	95.74	23	Accept
75	741	135.06	141.35	37	Accept
100	983	180.11	186.40	69	Accept

Table 12: Experiment 1

$\alpha = 0.0100$ $\beta = 0.0200$ $\Theta_2 = 0.2000$

Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0001$					
10	98	2.30	3.38	4	Reject
25	248	6.57	7.66	12	Reject
50	496	13.64	14.73	23	Reject
75	741	20.63	21.71	37	Reject
100	983	27.52	28.61	69	Reject
$\Theta_1 = 0.0002$					
10	98	2.52	3.71	4	Reject
25	248	7.21	8.40	12	Reject
50	496	14.96	16.15	23	Reject
75	741	22.62	23.81	37	Reject
100	983	30.19	31.38	69	Reject
$\Theta_1 = 0.0003$					
10	98	2.67	3.93	4	Reject
25	248	7.64	8.90	12	Reject
50	496	15.86	17.12	23	Reject
75	741	23.97	25.24	37	Reject
100	983	31.99	33.25	69	Reject
$\Theta_1 = 0.0004$					
10	98	2.78	4.10	4	Continue Sampling
25	248	7.98	9.29	12	Reject
50	496	16.56	17.87	23	Reject
75	741	25.03	26.35	37	Reject
100	983	33.41	34.73	69	Reject
$\Theta_1 = 0.0005$					
10	98	2.88	4.25	4	Continue Sampling
25	248	8.26	9.62	12	Reject
50	496	17.14	18.51	23	Reject
75	741	25.92	27.29	37	Reject
100	983	34.59	35.96	69	Reject
$\Theta_1 = 0.0006$					
10	98	2.97	4.38	4	Continue Sampling
25	248	8.50	9.91	12	Reject
50	496	17.65	19.06	23	Reject
75	741	26.69	28.10	37	Reject
100	983	35.62	37.03	69	Reject

$\alpha = 0.0100$ $\beta = 0.0200$ $\Theta_2 = 0.2000$

Percentage	Generated	d_1	d_2	Live	Decision
$\Theta_1 = 0.0007$					
10	98	3.05	4.49	4	Continue Sampling
25	248	8.72	10.17	12	Reject
50	496	18.11	19.55	23	Reject
75	741	27.38	28.82	37	Reject
100	983	36.54	37.98	69	Reject
$\Theta_1 = 0.0008$					
10	98	3.11	4.59	4	Continue Sampling
25	248	8.92	10.40	12	Reject
50	496	18.52	20.00	23	Reject
75	741	28.00	29.48	37	Reject
100	983	37.37	38.85	69	Reject
$\Theta_1 = 0.0009$					
10	98	3.18	4.69	4	Continue Sampling
25	248	9.10	10.61	12	Reject
50	496	18.90	20.41	23	Reject
75	741	28.58	30.09	37	Reject
100	983	38.14	39.65	69	Reject
$\Theta_1 = 0.0015$					
10	98	3.48	5.14	4	Continue Sampling
25	248	9.98	11.64	12	Reject
50	496	20.73	22.39	23	Reject
75	741	31.35	33.01	37	Reject
100	983	41.84	43.50	69	Reject
$\Theta_1 = 0.0020$					
10	98	3.68	5.44	4	Continue Sampling
25	248	10.55	12.31	12	Continue Sampling
50	496	21.92	23.68	23	Continue Sampling
75	741	33.14	34.90	37	Reject
100	983	44.23	45.99	69	Reject
		⋮			
$\Theta_1 = 0.0500$					
10	98	8.30	13.75	4	Accept
25	248	24.85	30.29	12	Accept
50	496	52.20	57.65	23	Accept
75	741	79.22	84.67	37	Accept
100	983	105.91	111.36	69	Accept
$\Theta_1 = 0.1000$					
10	98	9.42	19.89	4	Accept
25	248	31.21	41.67	12	Accept
50	496	67.23	77.70	23	Accept
75	741	102.81	113.28	37	Accept
100	983	137.96	148.43	69	Accept

References

[1] S. Bai and Y. Yun. Optimum number of errors corrected before releasing a software system. *IEEE Transactions on Reliability*, 37(1):41–44, April 1988.

[2] T. A. Budd. *Mutation Analysis of Program Test Data.* PhD thesis, Yale University, 1980.

[3] C. Bullard and E. H. Spafford. Testing experience with Mothra. In *Proceedings of 25th Southeast ACM Conference*, Birmingham AL, April 1987.

[4] B. J. Choi, R. A. DeMillo, E. W. Krauser, R. J. Martin, A. P. Mathur, A. J. Offutt, H. Pan, and E. H. Spafford. The Mothra tools set. In *Proceedings of the 22nd Hawaii International Conference on Systems and Software*, pages 275–284, Kona, Hawaii, January 1989.

[5] R. A. DeMillo and A. J. Offutt. Experimental results of automatically generated adequate test sets. In *Proceedings of the Sixth Annual Pacific Northwest Software Quality Conference*, Portland, Oregon, September 1988.

[6] Richard A. DeMillo and Eugene H. Spafford. The Mothra software testing environment. In *11th Nasa Software Engineering Laboratory Workshop.* Goddard Space Center, December 1986.

[7] R. Govindarajalu. *Sequential Statistical Procedures.* Academic Press, 1975.

[8] W. E. Howden. *Functional Program Testing and Analysis.* McGraw-Hill, 1987.

[9] R. J. Martin. Using the Mothra software testing environment to ensure software quality. In *Proceedings of the IEEE National Aerospace and Electronics Conference*, Dayton, OH, May 1989.

[10] G. Myers. *The Art of Software Testing.* John Wiley and Sons, New York, NY, 1979.

[11] Mehmet Şahinoğlu and Eugene H. Spafford. A sequential statistical procedure in mutation-based testing. In *Procedings of the 28th Reliability Spring Seminar*, pages 127–148, Boston, MA, April 1990. Central New England Reliability Chapter of the IEEE.

[12] Eugene H. Spafford. Extending mutation testing to find environmental bugs. *Software Practice and Experience*, 20(2):181–189, February 1990.

[13] Statistics Research Group. *Sequential Analysis of Statistical Data: Applications.* Columbia University Press, 1945.

[14] Abraham Wald. *Selected Papers in Statistics and Probability.* McGraw Hill, 1945.

[15] P. J. Walsh. *A Measure of Test Case Completeness.* PhD thesis, Watson School of Engineering, State University of New York at Binghamton, Binghamton, NY, 1985.

[16] G. B. Wetherill and K. D. Glazebrook. *Sequential Methods in Statistics.* Chapman and Hall, 1986. Third edition.

[17] Lee J. White and E. K. Cohen. A domain strategy for computer program testing. *IEEE Transactions on Software Engineering*, 6(3):247–257, May 1980.

[18] M. R. Woodward and K. Halewood. From weak to strong: Dead or alive? an analysis of some mutation testing issues. In *Proceedings of the Second Workshop on Software Testing, Verification and Analysis*, pages 152–158, Banff, Alberta, Canada, July 1988.