7-5-2005

Evaluation of Accuracy and Efficiency of some Simulation and Sampling Methods in Structural Reliability Analysis

Christopher D. Eamon
Mississippi State University, Starkville, MS, christopher.eamon@wayne.edu

Michel D. Thompson
Mississippi State University, Starkville, MS

Zhenyu Liu
Mississippi State University, Starkville, MS

Evaluation of Accuracy and Efficiency of some Simulation and Sampling Methods in Structural Reliability Analysis

Christopher D. Eamon[1], Michel Thompson[2], and Zhenyu Liu[2]

[1]Assistant Professor of Civil Engineering

[2]Graduate Student, Department of Civil Engineering

Mississippi State University, MS, 39762-9546

Abstract

Numerous simulation and sampling methods can be used to estimate reliability index or failure probability. Some point sampling methods require only a fraction of the computational effort of direct simulation methods. For many of these methods, however, it is not clear what trade-offs in terms of accuracy, precision, and computational effort can be expected, nor for which types of functions they are most suited. This study uses nine procedures to estimate failure probability and reliability index of approximately 200 limit state functions with characteristics common in structural reliability problems. The effects of function linearity, type of random variable distribution, variance, number of random variables, and target reliability index are investigated. It was found that some methods have the potential to save tremendous computational effort for certain types of limit state functions. Recommendations are made regarding the suitability of particular methods to evaluate particular types of problems.

Introduction

For many practical problems in structural reliability, an analytical determination of failure probability is unattainable. This is especially so when complex structures are modeled with numerical techniques such as the finite element method, where generation of an explicit expression of load or resistance, and thus limit state function, is typically not possible. Although a direct Monte Carlo simulation is a valid approach for such cases, it is often practically infeasible to conduct due to the large number of simulations required for sufficient accuracy. To address this problem, numerous simulation and sampling methods were developed by various researchers that can aid the determination of failure probability or reliability index. Currently, there exists a large number of competing schemes to estimate statistical parameters from a significantly reduced number of samples. A review of many of these methods can be found in [1]. Unfortunately, when presented in the literature, the accuracy and efficiency of these schemes are typically evaluated for only a small number of simple example cases and without significant comparison to other methods [1-12].

For many schemes, the relationship between computational effort and accuracy is not apparent. Nor is it clear which schemes are most suitable for a particular type of problem, in terms of number of random variables, degree of nonlinearity, random variable distributions and variances, and so on. The lack of a readily-available evaluation and comparison among competing methods presents an obstacle to proper method selection, particularly for those without extensive experience using these methods. This paper attempts to fill this gap by examining a number of simulation and sampling methods that may be used to estimate failure probability (P_f) or reliability index (β) in structural reliability analysis. In this study, a subset of methods with specific typical formulations is considered. Of particular interest in this study are point sampling methods, which promise to provide large reductions in computational effort. These methods require a fixed number of simulations based on the number of random variables in the problem. Three point sampling methods are considered: Rosenblueth's 2n+1 point estimation method [6]

(ROS) and Zhou's n+1 (N+1) and 2n (2N) point integration methods [7]. These methods are characterized by having a fixed number of pre-determined samples and are used to calculate β. Five simulation methods are also considered: traditional Monte Carlo Simulation (MCS), Latin Hypercube (LH) stratification [8], Importance Sampling [9] (IS), Plane-based Adaptive Importance Sampling [13] (AISP), and Curvature-based Adaptive Importance Sampling (AISC) [10]. These methods have an open-ended number of samples, and are used to calculate P_f directly. The Rackwitz-Feissler procedure [5] (RF), a well-known first order analytical method suitable for non-normal distributions, is considered as well. By evaluating 198 different functions, these seven methods are compared in terms of accuracy, precision, and computational effort for parameters that include linearity of the limit state function, number of random variables (RVs), type of RV distribution, variance, and target reliability index.

As the results presented here are empirical, they cannot be used to make universal conclusions for a method's effectiveness for all possible functions. However, the results may serve to aid the process of initially selecting a simulation or sampling method appropriate for a particular type of problem, given specific accuracy, precision, or computational effort constraints.

Methods Considered

Thorough discussions of MCS are presented elsewhere [4, 15-17]. With MCS, as the accuracy of the estimate of P_f depends on the number of simulations, a desired degree of precision must be chosen. In this study, the number of MCS samples needed is estimated by [18]:

$$COV = \sqrt{\frac{(1 - P_f)}{N\,P_f}} \qquad (1)$$

Where P_f is the failure probability of the limit state function to be estimated and N the number of samples. To keep the number of simulations reasonable, N was chosen such that the expected COV of the result is 5%.

There are several stratification techniques available [8, 16, 19]. In general, the Latin Hypercube (LH) technique involves partitioning the probability density function of each RV into vertical strata. For each simulation, a random strata is sampled for each RV, with the condition that every strata is sampled only once in the entire procedure. In this study, strata are formed from equal probability weights. To reduce the possibility of spurious correlations occurring, a correlation reduction technique as described by Iman and Conover [19] is employed. The number of strata (samples) was chosen such that the accuracy results were close to those obtained from MCS. This number varies depending on function failure probability. Specific quantitative results are discussed later.

The Rackwitz-Feissler Procedure [5] is a commonly used first-order iterative method that calculates β rather than P_f. It accounts for non-normal distributions by converting these to 'equivalent normal' distributions at the design point on the failure boundary. It provides excellent results for linear problems, but has two potential sources of inaccuracies for nonlinear limit states. First, as it computes β, the shortest distance from the failure region to the origin of reduced coordinates, the same reliability index may be achieved for failure regions of different (hyper)volumes and thus failure probabilities. The conversion of reliability index to failure probability ($P_f = \Phi(-\beta)$) is therefore not always accurate. A second source of error may occur if the failure (hyper)surface is not consistently convex or concave, but has multiple local minimums. Since the iteration is begun at a single trial design point (typically at the mean values), the search algorithm will stop when it finds the first local minimum. This may or may not be the global minimum.

Note that RF first requires one evaluation of the limit state function (g) to insure that the design point is on the failure boundary (this will require more than 1 evaluation if at least 1 RV is not linear with respect to g), then a small number of iterations as the search algorithm attempts to find the minimum distance from the failure surface to the origin of reduced coordinates. Each iteration requires the computation of one derivative per RV in the problem (the partial derivative of g with respect to the reduced RV), as well as another evaluation of g to place the updated design point on the failure boundary once again. If the limit state is not explicitly given, such as in the case where a finite element procedure is used, RF must be evaluated numerically. Derivatives can be evaluated with a finite difference approach, which would require at minimum one evaluation of g (if a forward or backwards difference scheme is used), or call to the finite element code, per RV. In this case, the total number of samples required for RF is a minimum of $1 + (n+1)i$, where n is the number of RVs and i is the number of iterations required for β convergence. This does not include the special case where if all RVs are nonlinear with respect to g. Here, the unit values in the above expression must be replaced with an unknown number of iterations, based on the specific problem and nonlinear solver used, to allow the design point to converge on the failure boundary. The number of iterations i needed for convergence varies, but is often small. For this study, three iterations were sufficient for most cases. Thus, $4+3n$ samples would be required to evaluate the limit states considered.

The fundamental approach of importance sampling (IS) is to shift the probability densities of the RVs close to the design point, such that the modified limit state will produce a high rate of failures. These failures can then be captured with a much-reduced number of samples, and a failure probability can be calculated. The final failure probability is then adjusted to account for the shifted distributions. Two key needs of IS are 1) to identify the design point (the point of maximum likelihood), or a point close to it, and 2) to choose an appropriate importance sampling function, which is used to adjust the final calculated failure probability. There are many ways proposed in the literature to satisfy 1) and 2) [3, 9, 20-26]. For this study, the design point is found with the RF method, while the importance sampling function is taken as the joint probability density function of the original limit state, with RV means shifted to the design point. The failure probability is evaluated with 50 MCS runs. This procedure was found to give close results to the crude MCS method for most functions studied. Quantitative results are discussed later. Allowing for three RF iterations, this procedure required [(4+3n) + 50] total samples. Since RF is used to identify the design point for use in IS, this procedure is subject to the same errors as RF.

Adaptive Importance Sampling (AIS) schemes adjust the sampling region as the analysis progresses, refining the sampling space to maximize sampling in the failure region. For highly non-linear or non-normal limit states, depending of the shape of the failure surface, sampling uniformly around the MPP, as with IS, may not produce a high failure rate and thus may give a poor estimate of failure probability. For AIS, after the MPP is identified, a small number of initial samples are taken. If few samples will fall within the failure region, the sampling boundary is adjusted, and a new set of samples is taken. The process repeats until a satisfactory number of failures is achieved. As with IS, the computed failure probability is then adjusted to account for the shifted sampling region. Many adaptive methods exist [2, 9, 10, 13, 14, 27-30]. In this study, two AIS methods are considered, as described by Wu [10, 13 ,14]. In one method, the limiting surface is parabolic, and is refined by adjusting curvatures at the MPP (AISC), while the other method uses a planar surface (AISP), and is refined by shifting the plane closer or further away from the MPP. The AIS process considered here is as follows. After the MPP is found (such as from RF), the limit state is linearized at that point then converted to a parabolic surface as described by Tvedt [31]. Failure probability is then estimated using a second-order method [10]. Ten initial MCS samples are then taken by evaluating the actual limit state function at the estimated MPP within the confines of the estimated parabolic failure surface. If samples fall outside of the true failure region, for AISC, the curvatures of the limiting surface are reduced to further enclose the failure region (or distance of the planar surface in the case of AISP), such that the change in estimated failure probability is approximately 10%. Here additional samples are taken only within the changed region. The failure probability estimate is then updated. This process of adjusting limiting surface curvatures and resampling in the incremented region is repeated until the failure probability estimates converge. Allowing three increments and ten samples per increment for AISC and 5 increments and 15 samples for AISP (AISP typically requires more samples for equivalent results), the scheme requires $44 + 3n + n(n-1)/2$ samples for AISC and $79+3n$ samples for AISP. As with IS, since the AIS methods depend on an initial location of the MPP, the method used to find the MPP may influence the quality of the final results.

Many point sampling methods exist [6-8, 32]. The focus of this study is on those methods that specify a small number of simulations, which realize the least computational costs. The N+1 and 2N point integration methods are based on Gauss quadrature. The derivation of the methods is given elsewhere [7]. Point values are given as a function of the number of random variables in the limit state, the position of an RV in the considered set, and the simulation run number. N+1 point values are given in Table 1. In the table, Z_{ij} refers to the simulation run number *i* and random variable j, while *n* refers to the number of random variables in the problem. Here *i* is from 1 to *n*+1 while *j* is from 1 to *n*. For 2N, each sample is conducted by setting all RVs equal to 0 (i.e. the mean value in standard normal space) except one, which is set to $n^{1/2}$. To conduct the simulation, the RV that takes the value of $n^{1/2}$ is alternated until all RVs are considered, to generate *n* samples. *n* additional samples are generated by repeating the above procedure but by using $-n^{1/2}$ as the value for the non-zero RV. Application of the point integration methods generates standard normal space values for each of the system random variables. These values are then transformed to basic variable space values for input to the limit state function using standard transformations, just as with MCS or LH. From the total sample of limit state evaluations, the mean value and standard deviation of the function are calculated. β is then estimated by dividing mean value by standard deviation. Zhou [7] provides no evaluation as to the expected effectiveness or limitations of the methods, nor could this be found in the available literature. However, based on the small number of samples, it is expected that accuracy will be degraded for nonlinear functions.

The derivation and use of Rosenblueth's 2n+1 point estimate is substantially different from N+1 and 2N [6]. Although it is referred to as "2n+1", if β is to be estimated, only 2n samples are needed (a single additional sample is required if the function's mean value is needed). It is used by first evaluating the limit state function with one of the RVs shifted upward in value by one standard deviation (y_i^+), while the remaining RVs in the function are kept at their mean values (y). This is repeated until the function is evaluated *n* times, where each time a different RV has its value shifted. This process is then repeated *n* times again, but now with the value of each RV shifted downward by one standard deviation (y_i^-).   The COV of the function can then be calculated as [6, 32]:

$$COV_Y = \sqrt{\left[\prod_{i=1}^{n}\left(1+V_{yi}^2\right)\right]-1} \qquad (2)$$

Where

$$V_{yi} = \frac{y_i^+ - y_i^-}{y_i^+ + y_i^-} \qquad (3)$$

Reliability index can then be estimated by taking the inverse of the function's COV. The method reportedly works best for functions that are linear with low skew and low variation. As neither nonlinearities nor distributions are taken into account, reliability index computed from this method may not provide an accurate indication of failure probability for non-normal and nonlinear functions.

Limit State Functions Considered

Each of the methods discussed above will be used to evaluate the failure probability and reliability index of a general limit state function under the influence of a variety of parameters. The parametric cases considered are given in Table 2. In the table, 'mixed' means that the RVs are given different distributions or coefficients of variation (COV). For the non-mixed cases, RVs are given identical distributions or COVs.   Each of the 22 cases in Table 2 is repeated three times, varying the number of elements in *g* that are RVs, to create a total of 66 subcases. In turn, each of these subcases is repeated three times, with a different target reliability index, for a total of 198 different limit state functions. A summary of the parameters considered is as follows:

- Linearity: ½ of the functions are linear, ½ are nonlinear.
- Distributions: Approximately ¼ have all normal RVs, ¼ are all lognormal, ¼ are all extreme I, and ¼ have mixed distribution types.
- Variance: Approximately $^1/_3$ have all RVs at 5% COV, $^1/_3$ have all RVs at 35% COV, and $^1/_3$ have different RV COVs (mixed among 5, 15, and 35%). One exception is the extreme I distributions, in which no mixed COV cases were considered.
- Random Variables: $^1/_3$ have 2 RVs, $^1/_3$ have 5 RVs, and $^1/_3$ have 15 RVs.
- Target Reliability Index: $^1/_3$ have a 'low' β, $^1/_3$ have 'moderate' β, and $^1/_3$ have 'high' β. The low group has an average β of 1.3 with range of 0.3 to 2. The moderate group has an average β of 3.5 with range of 2 to 5. The high group has an average β of 10 with range of 5 to 15.

Specific parameter values were chosen such that the mean reliability index is consistent among different parametric comparison groups. This is important because a function's failure probability influences the accuracy at which it can be estimated. For example, the normal and non-normal distribution groups have the same average β, as do functions with 2, 5, and 15 RVs. Two exceptions to this rule are the COV groups (the low COV functions have, on average, a higher β than the higher COV functions), and linearity (the linear functions have a higher average β than the nonlinear functions). When making data comparisons for β, COV, and linearity parameters, the bias that target reliability index brings to the results is accounted for and is discussed below in the results section.

To assess the effectiveness of the methods, a general function *g* was developed in which parameters such as number of random variables, target reliability index, and linearity could be adjusted in a consistent way. The limit state functions have the general form:

$$g = k\sum_{i=1}^{n} d_i - c\sum_{j=1}^{k} \frac{L_j^4 w_j}{E_j I_j} \qquad (4)$$

Although *g* represents specific numerical problems for this study, function *g* was chosen such that its form is mathematically similar to many typical limit states encountered in structural reliability; those containing sums of products and quotients of RVs. Depending on the choice of RVs, three forms of g (with respect to linearity) were considered: linear, nonlinear with positive exponents, and nonlinear with positive and negative exponents (i.e. inverse) problems. The RVs, distributions, and COVs for each RV subcase (2, 5, 15) are given in Table 3. Some reliability problems may involve hundreds or thousands of RVs, but many practical problems in structural reliability, particularly those used for code calibration, and even some system reliability problems, involve a dozen or less RVs. Solving reliability problems with a large number of RVs generally requires tremendous computational effort and are beyond the scope of this work. However, some important trends can be seen based on the number of RVs considered here. For the mixed cases in Table 3, distribution type and COV values chosen for the RVs were influenced based on the problem that *g* represents, as discussed in the next paragraph.

As specific formulations of *g* must be chosen for evaluation, constants as well as RV values were defined such that *g* represents one (for 2 RV problems), two (for 5 RV problems), or five (for 15 RV problems) simple, uniformly-loaded beams in parallel with multiple load effects (such as dead load and live load) applied, and failure defined as exceeding a non-deterministic midspan deflection limit. For all problems, c = 5/384, L = 6.1 (m), E = $2x10^8$ (kPa), and I = $6.452x10^{-4}$ (m^4), and w = 73600 (N/m), w_{DL} = 19300 (N/m) and w_{LL} = 54300 (N/m). These values may be either constants or the mean values of RVs, depending on how the element is used in the specific problem. Values for d and k are problem-specific and chosen such that three different target reliability indices (within the ranges of low, medium, and high, quantified above) could be examined for each subcase. These values are given in tables 4 and 5. Clearly, as COV or distribution changes, the range of possible β values also changes, and β is particularly limited for high COV or extreme I distributions. In each case, the intent was to examine the effectiveness of the seven considered procedures over a wide range of failure probabilities. For the 2 RV problems, n=k=1. For the 2 RV linear problems, d_1 and w_1 are the RVs. For the 2 RV nonlinear problems, d_1 and L_1 are the RVs. For the 5 RV linear problems, n=2 and k=3, while d_i and w_i are taken as RVs. For 5 RV nonlinear problems, n=k=1 and d_1, w_1, E_1, I_1, and L_1 are RVs. For the 15 RV linear problems, n=5, k=10, and RVs are d_i and w_i. For the 15 RV nonlinear problems, n=k=3 and RVs are d_i, w_i, E_i, I_i, and L_i. Substituting these values into equation (4) results in the following, where random variables are in bold and subscripted:

The 2 RV Linear Case is:

$$g = k\mathbf{d_1} - \frac{cL^4}{EI}\mathbf{w_1} \qquad (5)$$

The 5 RV Linear Case is:

$$g = k(\mathbf{d_1} + \mathbf{d_2}) - \frac{cL^4}{EI}(\mathbf{w_{DL1}} + \mathbf{w_{LL1}} + \mathbf{w_{LL2}}) \quad (6)$$

The 15 RV Linear Case is:

$$\begin{aligned} g &= k(\mathbf{d_1} + \mathbf{d_2} + \mathbf{d_3} + \mathbf{d_4} + \mathbf{d_5}) \\ &\quad - \frac{cL^4}{EI}(\mathbf{w_{DL1}} + \mathbf{w_{LL1}} + \mathbf{w_{DL2}} + \mathbf{w_{DL2}} + \mathbf{w_{DL3}} + \mathbf{w_{LL3}} + \mathbf{w_{DL4}} + \mathbf{w_{LL4}} + \mathbf{w_{DL5}} + \mathbf{w_{LL5}}) \end{aligned} \quad (7)$$

The 2 RV Nonlinear Case is:

$$g = k\mathbf{d_1} - c\frac{w\mathbf{L_1^4}}{EI} \quad (8)$$

The 5 RV Nonlinear Case is:

$$g = k\mathbf{d_1} - c\frac{\mathbf{w_1 L_1^4}}{\mathbf{E_1 I_1}} \quad (9)$$

The 15 RV Nonlinear Case is:

$$g = k(\mathbf{d_1} + \mathbf{d_2} + \mathbf{d_3}) - c\left(\frac{\mathbf{w_1 L_1^4}}{\mathbf{E_1 I_1}} + \frac{\mathbf{w_2 L_2^4}}{\mathbf{E_2 I_2}} + \frac{\mathbf{w_3 L_3^4}}{\mathbf{E_3 I_3}}\right) \quad (10)$$

Special Problems

Engelund and Rackwitz [9] identified six limit states that posed potential problems for various importance sampling schemes. These limit states have been included in this study for evaluation as well. One limit state was a linear function that considered various numbers of random variables and failure probabilities, the characteristics of which are fully covered in the linear limit states above. The remaining five limit states are listed below.

1. Exponential nonlinearity.

This highly nonlinear limit state is given by:

$$g = \pm \sum_{i=1}^{20} \ln\left[\Phi\left(-u_i\right)\right] \pm C \qquad (11)$$

Where u_i are normally distributed random variables. Although Engelund and Rackwitz considered various values for C, in this study the function of highest curvature is considered, C=41.05, which gave the poorest reported results [9]. RVs are given means of 1.0 and standard deviations of 0.15. Reliability index is approximately 4.

2. Multiple reliability indices.

This limit state is hyperbolic with two minimum and one maximum beta values. It is given by:

$$g = x_1 x_2 - 146.14 \qquad (12)$$

Where the means of x_1 and x_2 are 78064.4 and 0.0104, while their standard deviations are: 11709.7, 0.00156. Both are normally distributed. Reliability index is approximately 4.2.

3. Series System.

The series system is given by:

$$g = \min\left(g_1, g_2, g_3\right) \qquad (13)$$

Where:

$$g_1 = x_1 + 2x_2 + 2x_4 + x_5 - 5x_6$$
$$g_2 = x_1 + 2x_2 + x_4 + x_5 - 5x_6$$
$$g_3 = x_2 + 2x_3 + x_4 + x_5 - 5x_6$$

x_1-x_5 are lognormally distributed with means of 60.0 and standard deviations of 6.0, while x_5 and x_6 have extreme type I distributions with means of 20 and 25 and standard deviations of 6.0 and 7.5, respectively. Reliability index is approximately 2.1.

4. Parallel System.

This limit state is given by:

$$g = \max\left(g_1, g_2, g_3, g_4\right) \tag{14}$$

Where:

$$\begin{aligned} g_1 &= 2.677 - u_1 - u_2 \\ g_2 &= 2.500 - u_2 - u_3 \\ g_3 &= 2.323 - u_3 - u_4 \\ g_4 &= 2.250 - u_4 - u_5 \end{aligned}$$

All g_i are standard normal random variables. Reliability index is approximately 3.5.

5. Noisy Limit State.

This limit state has a fluxuating boundary and is:

$$g = x_1 + 2x_2 + 2x_3 + x_4 - 5x_5 - 5x_6 + 0.001\sum_{i=1}^{6} \sin(100x_i) \tag{15}$$

All x_i are lognormal. x_1-x_4 have means of 120 and standard deviations of 12; x_5 has mean of 50 and standard deviation of 15; and x_6 has mean of 40 and standard deviation of 12. Reliability index is approximately 2.3.

Results

Tables 6-8 and figures 1-10 present results in terms of β, while tables 9-11 and figures 11-20 give results in terms of P_f. In figures 1-10, results are presented for each method based on a subset of cases, to examine the effect of the various parameters considered (linearity, normality, COV, number of RVs, target β). Results are measured in two ways: accuracy and precision. Accuracy refers to closeness to the exact value, and is measured by mean value. This is calculated by first normalizing the β computed by a particular method to the exact value (calculated β / exact β) for each limit state in the subset, then the mean value of this fraction is taken for the subset. Precision refers to the degree of consistency of the results. This is measured by coefficient of variation of the mean (COV). Note that an exact solution would have a (normalized) mean of 1.0 and a COV of 0. For figures 11-20, the same procedure is used to report results in terms of P_f. "Overall" in the figures refers to the results of all of the limit states (198) evaluated by a particular method.

The tables present data specifically for each of the 66 subcases and results are calculated for the subset of 3 limit states composing each subcase (i.e. the mean and COV are computed from the results consisting of functions with the same number of random variables, distribution types, and COV, but different βs). For the simulation methods, some functions had P_f beyond the practical range of applicability. These cases are designated with a "n/a" in the tables. Because such a small set of data (3) is considered for mean and COV results presented in the tables, results should not be considered as reliable as those presented in the figures. However, results in the tables are useful to identify trends as well as specific problem areas, as discussed in detail below.

General Observations

For all of the methods considered, it is clear by observing the patterns in the tables that some types of functions are inherently easier or more difficult to estimate well. In terms of β, most of the point sampling methods gave high precision results for cases 1-4, 9,10, and 14 (table 6). These cases are primarily linear and contain distributions other than extreme I. Cases estimated with high accuracy by most sampling methods were similar: 1-3, 9, and 14. These are primarily linear and contain normal distributions. Low precision as well as low accuracy results for most sampling methods were predominately found for nonlinear cases with non-normal distributions (5, 8, 12, 13, 15, 16, 18-21 in table 7). The simulation methods give good results for most cases. There is some degradation in accuracy and precision for two cases, however (13 and 21). These are nonlinear functions with high variation. Overall, MCS, IS, and AISC gave best results for accuracy, while AISP, AISC, and RF gave best results for precision.

The analytical method (RF) also provides very good overall results, with nearly all cases estimated with high accuracy and precision (table 8). Quality does degrade as number of RVs increases. The 15 RV functions for cases 13, 16, 19, and 21 (nonlinear with high variation) are particularly troublesome. As the number of RVs increases, the failure surface becomes more complex, with potentially many local minimum. As such, the search algorithm may get 'stuck' in a local but not global minimum, thereby introducing error in β estimation. This is evidenced by the fact that RF always overestimates β if it errs, but never underestimates it.

If P_f is considered, the differences between the point sampling methods and simulation methods are magnified, the former providing poor performance. For RF, as with β results, the most inaccurate and imprecise cases are 13, 16, 18, 19, and 21. These are primarily nonlinear with high variation. The cases that have primarily low target β are 2, 10, 13, 16, and 21. The most accurate and precise cases as estimated by most of the point sampling methods considered tend to fall into this group. For example, cases 2, 10, 13, 18, and 21 are the most accurate of the sampling methods, while case 2 is the most precise (while the remaining cases show little difference; see table 9). Although this is not as clearly defined for the simulation methods (table 10), there is still significant overlap; cases 21 and 7 are the most precise, while cases 2, 21, 11, and 18 are the most accurate. Trends become clearer when inter-dependent factors are controlled for, as discussed below.

Although β precision results are quite reasonable in many cases, overall precision of P_f results was low. Although superior to all other methods except RF, the precision of MCS was also lower than expected, having an overall COV of P_f results of 0.18 (but 0.06 for β) rather than the 0.05 predicted using eq. (1). It should be noted that this approximate formula does not adjust for function characteristics (RV distribution, linearity, or RV COV), which, in addition to P_f, also affect precision. With regard to LH, for lower failure probabilities ($P_f \geq 0.00001$), the method could reduce the number of MCS simulations by a factor of 50 for similar accuracy, but for higher P_f, proportionally more samples (relative to MCS) were needed. This varied from no reduction (for $P_f \leq 0.10$), while for $0.10 \leq P_f \leq 0.00001$, a reduction factor of about 10 was achieved. It should be noted that, although the LH approach used here provided equivalent accuracy to MCS, a higher degree of variation was observed than with MCS (overall COV for β

= 0.08; overall COV for P_f = 0.34). Clearly, increasing the number of samples for both of these methods would reduce variation. Overall, MCS, LH, AISC, and RF gave best results for accuracy, while MCS, AISP, AISC, and RF gave best results for precision.

Note that although many of the procedures studied produced large errors in P_f in some cases, to keep the results in practical perspective, most civil engineering structures designed by codes that lack a reliability-based calibration (pre-Load and Resistance Factor Design versions) typically have variations in P_f of orders of magnitude. Even with calibrated codes, the expected variation of P_f is typically greater than a factor of 2 [33].

Linearity

Because the nonlinear functions, taken as a whole, have a lower average β than the linear functions (mean β of linear subset = 5.6; mean β of nonlinear subset = 4), results may be biased, as accuracy and precision are not independent of target reliability index (low β functions are in general estimated with higher accuracy). To eliminate this bias, a smaller set of functions was also considered (about half the of the original), such that the limit states which have a significant difference in β between their linear and nonlinear forms were not included in the set (the average β of the controlled linear and nonlinear sets are both equal to 3.1). The linear and nonlinear β-controlled results are indicated as “lin B ctrl” and “nonlin, B ctrl” in figures 1-2 and 11-12. It is these results that should be compared to understand the true effect that linearity has on accuracy and precision. As shown in figures 1-2 and 11-12, for the β -controlled results, all methods display better accuracy for linear functions and worse accuracy for nonlinear functions. However, in terms of β (figure 1), differences due to linearity are minor for the most part.

Exceptions are ROS and RF, which display a significant bias in accuracy toward linearity. When failure probability is considered, these differences are magnified, particularly with N+1, which demonstrates a significant advantage in accuracy with the linear functions. For precision, the same trend is realized, with a few exceptions. It appears that MCS (figures 2 and 12) results are worse for the β -controlled linear cases than the overall results. This is not a general trend, however, but due to a high-error outlying sample which throws off the results. If this point (within case 3, 5 RV) is eliminated from the set, the trend returns to that which is expected (linear cases are more precise).

Normality

In figures 2-4 and 13-14, functions are grouped into those that have all RVs with normal distributions ("normal") and those that have any or all RVs with a non-normal distribution ("non-normal"). The mean β for each of these groups is identical. For the point sampling methods, RV distribution clearly affect accuracy as well as precision, where all-normal functions have a distinct advantage. This is not so with the simulation or RF methods, however, where there is no difference in accuracy. In terms of precision, the simulation and RF methods appear to have a slight advantage for all-normal functions. Two exceptions are ROS (figure 3), and MCS (figures 4 and 14), which show a reverse trend. This again, is due to a single outlying datum (within case 13, 15 RV for ROS and case 13, 2 RV for MCS), and if this datum is removed, ROS and MCS also shows better results for all-normal functions.

Number of RVs

In figures 5-6 and 15-16, functions are grouped into those which have 2 and 15 RVs. The mean β for each of these groups is identical. For all of the point sampling methods, accuracy decreases as number of RVs increases. In each, there is a significant difference in the 2 RV and 15 RV cases. Interestingly, for precision, however, the reverse trend is true, such that the high RV cases are most precise.

MCS appears to decrease in accuracy for a high number of RVs if failure probability is measured but to increase in accuracy for a high number of RVs if β is measured. This discrepancy occurs because of the nonlinear relationship between β and failure probability (figure 21, discussed further below under the effect of target β). For the 2 RV cases, data in two outlying cases (10, 12) contain large errors. These have low average β (1.4). For the 15 RV cases, functions with the worst results (within cases 5, 6, 10, 12, 14, 15, 20), have a much higher average β (5.0). The errors in this latter group, once converted to P_f, are greatly increased relative to those of the low β set. Thus, a different trend with regard to β and P_f is observed. If these worst cases are not included in the data sets, there is no difference is results between the 2 and 15 RV cases. In terms of precision, it appears that MCS has an increased precision for a low number of RVs. This is also due to the results of two outlying data points (within cases 13, 21). Eliminating these two points results in the low and high RV case having the same COV result (0.02). Thus for MCS, the number of RVs does not appear to affect accuracy or precision. When LH is considered, both accuracy and precision degrade as the number of RVs increases. This is an overall trend that can be seen in tables 7 and 10 and figures 5-6 and 15-16 and is not due to one or two very poor results. This result is interesting, as the expectation is that LH would show the same trends as MCS. It appears that the particular process of PDF stratification used in this study [19] may cause results to degrade for a higher number of RVs. For RF and IS, as expected, both accuracy and precision decrease for the high RV case. This is also true for AISP and AISC, though to a lesser extent.

<u>COV</u>

Regardless of the limit state load and resistance values, each limit state has a upper bound of reliability index depending on the function variance. Although functions with low variance can have high βs, functions with high variance cannot. The resulting effect is that, the low variance functions have both high and low β results, but the high variance functions only have (relatively) low β results. As target reliability index affects accuracy and precision, this effect must be controlled for. In figures 7-8 and 17-18, the category "5, low B" represents the set of limit state functions in which all RVs have a COV of 5%, and for which the high β cases were removed, such that the mean β of this set is equal to the mean β of the set of limit states with high variation (COV for all RVs = 35%). To determine the unbiased effect of COV on accuracy and precision, these two cases ("5, low B" and "35") should be compared. In general, the trend is that as COV increases, accuracy and precision decrease.

The figures dealing with failure probability (17-18) clearly show that the low COV cases, controlled for target β ("5, low B") are better than the high COV cases, as well as the uncontrolled low COV cases (5% uncontrolled, which includes high failure probability cases). One exception is LH, which has an outlying datum (case 18, 15 RVs). If this point is eliminated, the trend matches those of the other methods. Considering reliability index (figures 7-8), notice that for most results, the results for all low COV cases ("5") are better than the low COV cases for which β is controlled ("5, low B"). Low β cases in general show a worsening of results due to the nonlinear transformation between failure probability and reliability index. This occurs because the low COV cases have a higher average reliability index than the β-controlled low-COV cases. Errors for the higher β cases, when converted to failure probability, are reduced much more than those of the lower β set. However, the low COV, controlled β cases ("5, low B") are of course still more accurately and precisely predicted than the high COV cases (35%). Here there are two exceptions in terms of accuracy, LH and IS, which again are caused by outliers (case 18, 15 RV for LH, and case 7, 15 RV for IS).

Target Reliability Index

In the figures (9-10, 19-20), functions are grouped into the low target β set (β of 2 and below, with mean of 1.3), a high β set (all β greater than 2, with mean of 6.75). Since all high COV functions must be within the low β group, results may be biased, as COV affects accuracy and precision as well. Therefore, an additional data set is considered which controls for COV (low, low V). This is the set of functions that have low COV (i.e. COVs other than 35%) and low β. These should be compared to the high β set (which also all have low COV) to ascertain the affect that target reliability index has on the results. As can be seen from the failure probability figures (19-20), target reliability index may have a significant effect on accuracy and precision. In general, as failure probability decreases (and reliability index increases), results worsen. For the simulation methods (MCS, LH), this is because failure probability is calculated as the number of failures sampled (i.e. where $g < 0$) divided by the total number of samples. Clearly, at low failure probabilities there are fewer failures to sample, so an error here can have significant effect on the overall computed result. For low target βs, since many more failures are expected as a proportion of the total samples, errors in assessing the number of failures is not as significant. The point sampling methods also display a decrease in precision and accuracy for low failure probability functions, but for a different reason. As the point sampling methods are used to compute β, the result must be transformed to failure probability. As reliability index increases, small errors in β cause increasingly larger errors in failure probability (fig 21). For accuracy and precision, the low β, COV-controlled case (low, low V), as expected, is better than the high β case. The difference is extreme for N+1 and 2N. An exception is LH, which is again due to a outlying datum.

Because of the reliability index-failure probability transformation, when β is considered (figs. 9 and 10), the expectation is that the high β cases would show an increase in accuracy and the low β cases a decrease in accuracy. In general, the expected trend occurs. This is particularly so with N+1 and 2N. For precision (figure 10), four of the seven methods (N+1, 2N, LH, RF) gives poor results for the low- β, COV-controlled cases. For each method, the same single function within case 20, for 15 RVs and a low target reliability index (0.80), is very poorly predicted. Eliminating this outlying datum returns the trends to match that of the three other methods.

<u>Special Problems</u>

Tables 12 and 13 give results for the five special problems considered. Here the calculated reliability index is divided by the exact value for comparison. For the nonlinear limit state (problem 1), all simulation methods gave good results, most within a few percent of the exact value. All three point estimation methods, however, failed to provide a meaningful result, as each predicted a coefficient of variation very close or equal to zero, leading to a unreasonably high reliability index (200+). This is because each of the 20 RVs in the problem have identical mean, COV, as well as low sensitivity of the limit state for both positive and negative perturbations. In the case of ROS, the sum of the positive and negative perturbations cancel out (see equations 2 and 3), resulting in zero COV. For N+1 and N+2, the low sensitivity of the limit state to any single RV results in a gross under-prediction of COV, as the specified perturbation value leaves the limit state practically unchanged.

The multiple beta value case (problem 2) gave trouble for each of the simulation methods that use an analytical method to locate the MPP. Just as with RF, the maximum rather than the minimum MPP was found, and thus results were sampled about a local rather than a global minimum, over-predicting reliability index by about 35% in this case. Here use of prior knowledge of the proper MPP location would have prevented this. Note that MCS and LH, which sample without reference to the MPP, provided accurate results. The point estimation methods gave results of varying quality, with ROS close to the exact value.

The series system (problem 3) was accurately evaluated by all simulation methods, as well as RF, although the point methods uniformly over-predicted reliability index by about 30%. The parallel system (problem 4) was likewise accurately identified by the simulation methods. Here RF significantly under-predicted reliability index, while the point estimation methods gave varying results.

The noisy limit state (problem 5) was accurately predicted by all methods except N+1. It appears that small fluxuations in the limit state boundary are not nearly so difficult for the numerical algorithms to deal with than large-scale shape nonlinearities.

Efficiency

If accuracy and precision were the only considerations for structural reliability calculations, it is clear that point estimation methods would have little usefulness, as in most cases they are out-performed by the simulation methods (as well as RF) by a wide margin. However, the performance of the sampling methods varies greatly with the type of limit state considered (in terms of linearity, number of RVs, distribution type, variance, and target reliability index), and in some instances, may provide suitable accuracy and precision to solve practical problems. Tables 14 and 15 present such data. First, the results for all functions considered in this study are given for a particular method ("OVERALL"). The next columns list a subset of functions for which a particular method was found to give best results in accuracy and precision. These results are highlighted in the table for each method. Note some methods (LH, IS, AISP, AISC, RF) estimate several different types of limit states equally well, and so have more than one result highlighted in the table. The data reveals that the point sampling methods can be very accurate and precise for certain types of limit states. In terms of reliability index (table 14), the N+1, 2N, and ROS methods are equivalent to MCS (overall) accuracy for limit states that are: normal with low COV (N+1); have low COV (2N); and have a low number of RVs (ROS). Moreover, these functions may even exceed MCS (overall) in terms of precision for limit states that are normal and linear, based on the number of MCS samples conducted in this study. Such results make it clear that, at least as important as the number of samples chosen is how the samples are chosen (i.e. randomly, as in MCS, versus selected, as with the point estimation methods).

When considering failure probability, the point methods do not show as great of an improvement, relative to MCS (overall), as with β. N+1 and 2N, for example, for low β, low COV functions, still do not match MCS (overall) accuracy or precision. However, considering the range of variation expected with failure probability, results from these methods are very reasonable. For normal, linear functions, ROS provides excellent results, and may exceed MCS (overall). Most precise results for N+1 and 2N appear for low COV, low RV functions, which appears different from the best results for β (normal, linear). Here the nonlinear transformation from failure probability to reliability index effects results. Although the very best results are highlighted in the tables, notice that the actual differences between some types of functions are small.

The total number of samples used for a particular type of limit state given a particular method is shown in Table 16. All methods considered except MCS and LH have a lower limit of the number of required simulations based on the number of random variables in the problem. As shown in the table, the number of simulations is held constant for all limit states of a given number of RVs for statistical comparisons to be consistent and meaningful. The basis for the number of samples taken is described previously in *Methods Considered*, but in general these were chosen such that most limit states evaluated by a particular method gave results close to that of MCS. **Note that for MCS and the AIS methods, the number of simulations indicated is approximate.** Of course, increasing the number of simulations for all methods for which this is possible (i.e. all but the point sampling methods), should produce more accurate and precise results.

The difference in computational time needed to execute a single loop in any of these methods for most complex problems is negligible. That is, for practical problems where minimizing computational effort would be a concern (such as those which call a FEA code to evaluate g) the computational effort needed to evaluate the limit state far exceeds that for the reliability algorithm to utilize that information (no more than a fraction of a second). Thus, computational effort is directly proportional to the number of simulations used as given in Table 16, independent of reliability method. By comparing the number of simulations used and the results of the various methods for certain types of limit states, it is clear that, in some cases, the point methods can reduce the required number of samples tremendously, often by several orders of magnitude, and yet may retain equivalent accuracy and precision as the simulation methods.

Conclusions and Recommendations

For certain types of functions, it was found that it is possible to tremendously reduce computational effort using point sampling methods and yet still maintain competitive accuracy and precision. It was also found that all problems could be accurately evaluated by at least several methods. To select an appropriate method for use, several important factors must be considered:

1) what computational resources are available?

2) what accuracy and precision are required?

3) what are the characteristics of the limit state function (number of RVs, distribution type, variance, linearity, and target reliability index)?

4) is reliability index or failure probability of primary interest?

Based on the data presented in tables 14 and 15, some results and recommendations can be summarized. The versions of MCS, LH, IS, and AIS investigated in this study can be expected to have similar and good overall results for a wide range of limit states, and are recommended for nearly all problem types. For all types of limit states, MCS predicts average reliability index within 1%, with mean COV of 6%, while LH over predicts reliability index by 4%, with COV of 8%. For LH, results appear to degrade for a higher number of RVs. For reliability index, maximum COV is 15% for 2 RV cases with a typical COV of 3% or less, with a typical beta value within of 1-2% of the true value (Tables 7, 14). The 15 RV case, however, has a maximum COV of 37% with an value of 15%. Results are similar for beta value. For high RV problems, it may be prudent to avoid the formulation of LH used in this study. IS predicts reliability index within 1% with COV of 6%, while AISP over predicts reliability index by 4%, with 3% COV. AISP almost always gives a slightly higher (usually within a few percent) estimate of beta value than AISC, which predicts average reliability index within 1% and COV of 3%.

RF can excel at nearly all types of limit states (typical beta and COV are nearly exact) except those that have a high number of RVs or that are highly nonlinear; here results may be very poor (15 RV forms of cases 13,16,19, and 21 have an estimated beta over twice the actual value and COVs of 26-45%; see table 8). Thus it can be recommended for all but these types of problems.

Considering all types of limit states, the point methods give mixed results. Overall, N+1 under predicts reliability index by 20%, with COV of 15%. 2N under predicts reliability index 12%, with COV of 15%, while ROS over predicts reliability index by 12%, with COV of 12%. Limiting results to specific types of problems, if reliability index is of primary interest, all three point methods appear to give excellent results for functions that have normally distributed RVs, and are linear and/or have low variance. Specifically, for normal, low COV problems, all methods give nearly-exact results, generally within a few percent for beta, and all simulation methods are within 2%. ROS is the worse performer here, on average over predicting beta by 13%. For COV, all simulation methods provide an average COV within 2% and a beta value no higher than 2%. All three point methods result in a COV no higher than 6%. If failure probability is needed and will not be converted to reliability index, in addition to the characteristics above, functions that have low target β (around 2 or less, or expected failure probabilities of about 0.02 or more) may be estimated well with the point methods. For normal, linear problems, all methods likewise give excellent results within 1% on average for beta. Here the worst method is N+1, under predicting average beta by 10%. Most also give good estimates for COV, as all methods are within 4%, where all point methods are within 1%. Therefore, for these specific types of problems, point methods, including n+1, can be useful and accurate.

For the special types of problems considered, results are mixed. Point methods will fail or give very poor results for cases like special problem 1 (many RVs with the same mean and COV and the limit state has a low sensitivity to each), and thus cannot be used for this problem type. For problems with multiple beta values, RF and importance sampling methods may focus on a local rather than a global minimum beta, over-estimating reliability index. This occurred in this study when these methods over-predicted beta by about 35% for special problem 2. Unless the actual minimum MPP is known, RF, IS, and AIS methods that use RF-like searches should be used cautiously if at all for these problems. The series system evaluated here gave no trouble for the simulation methods, although point methods over predicted beta by about 30%. Similarly, the simulation methods gave good results for the parallel system, while RF under predicted beta by 54% and the point methods gave unreliable results. Clearly, some caution should be used if RF or point methods are applied to system problems. The noisy limit state was well-predicted by all methods except N+1 (40% over prediction of beta).

Note that in structural reliability analysis, precision is often more important than accuracy. That is, it is often of interest to compare the differences in reliability among structures. This is typically true in code calibration efforts, for example, where a small error in accuracy may not be important, as long as the error is consistent from one structure to the next. Further, in many cases, good accuracy with poor precision has little meaning. That is, although the mean value for a set of different functions may be estimated well, the expected result for any specific function may be poor, as indicated by a high COV in the results. Good accuracy in these cases (i.e. high accuracy but low precision) comes from high and low values averaging out, not a consistently good estimation of the functional values, which would be indicated by a low COV. If the object is to estimate the value of a single function, as often occurs, rather than a group of different functions, then the accuracy results must be considered with reference to COV.

This study is intended to be used to aid the initial consideration of a method appropriate for a specific type of problem. For some problems, point sampling methods may save tremendous computational effort. It must be kept in mind that these results are based not on the infinite set of all possible limit states, but for a sample of functions with characteristics commonly occurring in structural reliability analysis. Some particular functions are not typical to the overall trends; some are better, while some are worse. This must be considered when a method is chosen. Some guidance can be provided by observing the 'outliers' in the data discussed earlier, as indicated in tables 6-8 and 9-11. In any case, it is strongly recommended that, once a method(s) is initially selected, some degree of verification for the specific problem at hand is conducted to ensure that the method provides suitable results.

References

1. Melchers, R.E. "Structural Reliability Analysis and Prediction", 2nd Ed. John Wiley & Sons, New York, 1999.

2. Bucher, Christian G. "Adaptive Sampling - An Interative Fast Monte Carlo Procedure." Structural Safety, v 5, n 2, Jun, 1988, pp 119-126.

3. Bayer, V. and Bucher, C.G. "Importance sampling for first passage problems of nonlinear structures." Probabilistic Engineering Mechanics, v 14, n 1-2, Jan-Apr, 1999, pp 27-32.

4. Melchers, Robert E. and Ahammed, Mukshed. "Gradient estimation for applied Monte Carlo analyses." Reliability Engineering and System Safety, v 78, n 3, December, 2002, pp 283-288.

5. Rackwitz, R. and Fiessler. B. "Structural Reliability Under Combined Random Load Sequence," Computers and Structures, No. 9, 1978.

6. Rosenblueth, E. "Two-Point Estimates in Probabilities," Journal of Applied Mathematical Modeling, May, pp 327-333, 1981.

7. Zhou, J.H. and Nowak, A.S. "Integration Formulas to Evaluate Functions of Random Variables." Structural Safety, No 5, 1988, pp. 267-284.

8. McKay, M.D., Beckman, R.J, and Conover, W.J. "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," Technometrics, Vol 21, No. 2, May 1979, pp. 239-245.

9. Engelund, S. and Rackwitz, R. "A benchmark study of Importance Sampling Techniques in Structural Reliability." Structural Safety, Vol 12, no 4, pp 255-276, 1993.

10. Wu, Y.-T. "An Adaptive Importance Sampling Method for Structural System Reliability Analysis," Reliability Technology, Proceedings of the Symposium on Reliability Technology. Ed. T.A. Cruse, AD-Vol. 28, American Society of Mechanical Engineers, New York, 1992, pp. 217-231.

11. Pandey, M.D., and Sarkar, A. "Comparison of a simple approximation for multinormal integration with an importance sampling-based simulation method." Probabilistic Engineering Mechanics, v 17, n 2, April, 2002, p 215-218

12. Robinson, D. and Atcitty, C. "Comparison of quasi-and pseudo-Monte Carlo sampling for reliability and uncertainty analysis." Collection of Technical Papers - AIAA/ ASME/ ASCE/ AHS/ ASC Structures, Structural Dynamics and Materials Conference, v 4, 1999, p 2942-2949

13. Millwater, H.R., Wu, Y.-T., Torng, Y,., Thacker, B., Riha, D. and Leung, C. "Recent Developments of the NESSUS Probabilistic Structural Analysis Computer Program," Proceedings of the AIAA/ASME/ASCE/AHS/ASC 33rd Structures, Structural Dynamics, and Materials Conference, AIAA, Washington, DC, 1992, pp 614-624.

14. Wu, Y.-T. "Computational Methods for Efficient Structural Reliability and Reliability Sensitivity Analysis. " AIAA Journal, Vol. 32, No. 8, August 1994, pp 1717-1723.

15. Hammersley, J.M. and Handscomb, D.C. "Monte Carlo Methods." London: Methuen & Co Ltd, 1964.

16. Ripley, B.D. "Stochastic Simulation." New York: Wiley & Sons, 1987.

17. Rubinstein, B.Y. "Simulation and the Monte Carlo Method." New York: Wiley & Sons, 1981.

18. Soong, T.T. and Grigoriu, M. "Random Vibration of Mechanical and Structural Systems." Englewood Cliffs, NJ: Prentice Hall., 1993.

19. Iman, R.L., and Conover, W.J. "A Distribution-Free Approach to Inducing Rank Correlation Among Input Variables." Communications in Statistics, 11(3), pp. 311-334, 1982.

20. Harbriz, A. "An Efficient Sampling Method for Probability of Failure Calculation," Structural Safety, Vol 3, no. 2, pp. 109-115. 1986.

21. Melchers, R.E. "Simulation in Time-Invariant and Time-Variant Reliability Problems," Proc., 4th IFIP Conference on Reliability and Optimization of Structural Systems, Rackwitz, R. and Thoft-Christensen, P. (Eds), Springer, Berlin, pp 39-82, 1991.

22. Ambartzumian, R., Der Kiureghian, A., Ohanian, V., and Sukiasian, H. "Multinormal probability by sequential conditioned importance sampling: theory and application." Probabilistic Engineering Mechanics, v 13, n 4, Oct, 1998, p 299-308.

23. Rackwitz, R. "Reliability analysis-a review and some perspective." Structural Safety, v 23, n 4, 2001, p 365-395.

24. Mori, Y. and Kato, T. "Multinormal integrals by importance sampling for series system reliability." Structural Safety, v 25, n 4, October, 2003, p 363-378

25. Au, S.K. and Beck, J.L "Important sampling in high dimensions." Structural Safety, v 25, n 2, April, 2003, p 139-163

26. Au., S.K. "Probabilistic failure analysis by importance sampling Markov chain simulation." Journal of Engineering Mechanics, v 130, n 3, March, 2004, p 303-311

27. Mori, Yasuhiro and Ellingwood, Bruce R. "Time-dependent system reliability analysis by adaptive importance sampling." Structural Safety, v 12, n 1, Apr, 1993, p 59-73

28. Wang, G.S., Ang, A.H.-S., and Lee, J.-C. "Adaptive kernel method for evaluating structural system reliability." Structural Engineering and Mechanics, v 5, n 2, Mar, 1997, p 115-126

29. Au, S.K. and Beck, J.L. “New adaptive importance sampling scheme for reliability calculations.” Structural Safety, v 21, n 2, 1999, p 135-158

30. Mahadevan, S. and Raghothamachar, P. “Adaptive simulation for system reliability analysis of large structures.” Computers and Structures, v 77, n 6, Oct, 2000, p 725-734

31. Tvedt, L. “Distribution of Quadratic Forms in Normal Space--Application to Structural Reliability.” Journal of Engineering Mechanics, Vol. 116, No. 6, 1990, pp. 1183-1197.

32. Nowak, A.S. and Collins, K.R. “Reliability of Structures.” McGraw Hill, 2000.

33. Nowak, A.S. “Calibration of LRFD Bridge Design Code,” NCHRP Report 368, 1999.

Figure Captions

Figure 1. Effect of Linearity on Accuracy (β)

Figure 2. Effect of Linearity on Precision (β)

Figure 3. Effect of Normality on Accuracy (β)

Figure 4. Effect of Normality on Precision (β)

Figure 5. Effect of Normality on Accuracy (β)

Figure 6. Effect of Normality on Precision (β)

Figure 7. Effect of Normality on Accuracy (β)

Figure 8. Effect of Normality on Precision (β)

Figure 9. Effect of Normality on Accuracy (β)

Figure 10. Effect of Normality on Precision (β)

Figure 11. Effect of Linearity on Accuracy (P_f)

Figure 12. Effect of Linearity on Precision (P_f)

Figure 13. Effect of Normality on Accuracy (P_f)

Figure 14. Effect of Normality on Precision (P_f)

Figure 15. Effect of Normality on Accuracy (P_f)

Figure 16. Effect of Normality on Precision (P_f)

Figure 17. Effect of Normality on Accuracy (P_f)

Figure 18. Effect of Normality on Precision (P_f)

Figure 19. Effect of Normality on Accuracy (P_f)

Figure 20. Effect of Normality on Precision (P_f)

Table 1. Point Values for N+1.

$$Z_{1j} = \left(\sqrt{n}, 0,0,...0\right)$$

$$Z_{2j} = \left(-\sqrt{\frac{1}{n}}, \sqrt{\frac{(n+1)(n-1)}{n}}, 0,0,...0\right)$$

$$Z_{3j} = \left(-\sqrt{\frac{1}{n}}, \sqrt{\frac{(n+1)}{n(n-1)}}, \sqrt{\frac{(n+1)(n-2)}{(n-1)}}, 0,0,...0\right)$$

$$Z_{4j} = \left(-\sqrt{\frac{1}{n}}, \sqrt{\frac{(n+1)}{n(n-1)}}, \sqrt{\frac{(n+1)}{(n-1)(n-2)}}, \sqrt{\frac{(n+1)(n-3)}{(n-2)}}, 0,0,...0\right)$$

....

$$Z_{nj} = \left(-\sqrt{\frac{1}{n}}, \sqrt{\frac{(n+1)}{n(n-1)}}, \sqrt{\frac{(n+1)}{(n-1)(n-2)}}, \sqrt{\frac{(n+1)}{(n-2)(n-3)}}, \ldots, \sqrt{\frac{(n+1)}{2}}\right)$$

$$Z_{(n+1)j} = \left(-\sqrt{\frac{1}{n}}, \sqrt{\frac{(n+1)}{n(n-1)}}, \sqrt{\frac{(n+1)}{(n-1)(n-2)}}, \sqrt{\frac{(n+1)}{(n-2)(n-3)}}, \ldots, -\sqrt{\frac{(n+1)}{2}}\right)$$

Table 2. Parameters Considered

Case Number	Function Type	RV Distribution Type	RV COV
1	linear	normal	5
2	linear	normal	35
3	linear	normal	mixed
4	linear	lognormal	5
5	linear	lognormal	35
6	linear	lognormal	mixed
7	linear	extreme I	5
8	linear	extreme I	35
9	linear	mixed	5
10	linear	mixed	35
11	linear	mixed	mixed
12	nonlinear	normal	5
13	nonlinear	normal	35
14	nonlinear	normal	mixed
15	nonlinear	lognormal	5
16	nonlinear	lognormal	35
17	nonlinear	lognormal	mixed
18	nonlinear	extreme I	5
19	nonlinear	extreme I	35
20	nonlinear	mixed	5
21	nonlinear	mixed	35
22	nonlinear	mixed	mixed

Table 3. Random Variables, Distributions and COVs

Type of Problem	RVs	Mixed cases*	
		distribution	COV (%)
Linear 2RV	d_1	norm	15
	w_1	ext I	5
Nonlinear 2RV	d_1	log	15
	L_1	norm	5
Linear 5RV	d_1, d_2	norm	15
	w_1	norm	5
	w_2	ext I	35
	w_3	log	15
Nonlinear 5RV	d_1	norm	15
	w_1	ext I	35
	E_1	log	5
	I_1, L_1	norm	5
Linear 15RV	d_1-d_5	norm	15
	w_1-w_5	norm	5
	w_6-w_8	ext I	35
	w_9, w_{10}	log	15
Nonlinear 15RV	d_1-d_3	norm	15
	w_1-w_3	ext I	35
	E_1-E_3	log	5
	I_1-I_3, L_1-L_3	norm	5

*For non-mixed cases, RV distributions for a particular limit state are either all normal, all lognormal, or all extreme I, while COVs are either all 5% or all 35%, as shown in Table 2.

Table 4. Values of d and k for all cases except 7, 8, 18, 19.

#of RVs	Target β	mean d (mm)	k (linear cases)*
2	high	112	0.925
	medium	26.5	0.829
	low	16.1	0.875
5	high	138	0.216
	medium	38.1	0.447
	low	20.5	0.602
15	high	37.7	0.541
	medium	22.5	0.676
	low	15.5	0.809

*k = 1.0 for all nonlinear problems.

Table 5. Values of d and k for cases 7, 8, 18, 19.

Case	#of RVs	Target β	mean d (mm)	k (linear cases)*
7, 18	2	high	24.6	0.532
	2	low	12.5	0.879
	5	high	25.4	0.404
	5	low	12.8	0.728
	15	high	15.8	0.703
	15	low	11.5	0.917
8, 19	2	high	605.0	0.072
	2	low	34.5	0.478
	5	high	902.9	0.026
	5	low	41.0	0.298
	15	high	127.1	0.142
	15	low	21.7	0.573

*k = 1.0 for all nonlinear problems.

Table 6. Accuracy and Precision of Point Sampling Methods, Reliability Index

case	2 RV						5 RV						15 RV					
	N+1		2N		ROS		N+1		2N		ROS		N+1		2N		ROS	
	mean	COV	mean	COV	mean	COV	mean	COV	mean	COV	mean	COV	mean	COV	mean	COV	mean	COV
1	0.82	0.00	0.87	0.00	1.00	0.00	0.91	0.01	0.95	0.00	1.00	0.00	0.98	0.02	0.99	0.01	1.01	0.01
2	0.82	0.00	0.87	0.00	0.98	0.02	0.91	0.00	0.95	0.00	0.97	0.02	0.97	0.00	0.98	0.00	0.96	0.03
3	0.82	0.00	0.87	0.00	1.00	0.00	0.91	0.00	0.95	0.00	0.98	0.02	0.97	0.00	0.99	0.00	0.98	0.03
4	0.74	0.06	0.80	0.06	0.92	0.06	0.82	0.05	0.89	0.05	0.94	0.05	0.87	0.05	0.95	0.05	0.97	0.03
5	0.54	0.33	0.62	0.33	0.69	0.32	0.59	0.12	0.71	0.12	0.77	0.10	0.51	0.10	0.64	0.09	0.86	0.06
6	0.55	0.32	0.61	0.33	0.70	0.32	0.86	0.06	1.01	0.07	1.07	0.08	0.71	0.05	0.85	0.06	1.03	0.05
7	0.80	0.12	0.92	0.13	0.99	0.18	0.73	0.02	0.85	0.01	0.90	0.11	0.57	0.05	0.69	0.02	0.92	0.09
8	1.07	0.54	1.24	0.54	0.83	0.01	0.63	0.13	0.75	0.10	0.76	0.01	0.54	0.10	0.67	0.07	0.82	0.08
9	0.99	0.12	0.96	0.05	1.22	0.17	0.96	0.05	1.04	0.06	1.11	0.06	0.93	0.03	1.00	0.03	1.06	0.02
10	1.03	0.23	1.12	0.25	0.96	0.05	0.84	0.03	0.91	0.03	0.93	0.04	0.80	0.00	0.91	0.02	0.95	0.01
11	0.78	0.02	0.88	0.06	0.99	0.01	0.91	0.10	1.02	0.09	1.08	0.10	0.79	0.08	0.91	0.07	1.06	0.06
12	0.95	0.08	1.02	0.08	1.18	0.08	1.14	0.12	1.21	0.13	1.29	0.13	1.11	0.14	1.22	0.13	1.31	0.11
13	0.65	0.39	0.74	0.42	0.95	0.39	0.87	0.35	0.78	0.59	1.29	0.16	0.42	0.26	0.66	0.10	2.54	0.16
14	0.81	0.01	0.88	0.01	1.01	0.01	0.99	0.06	1.01	0.05	1.06	0.05	1.06	0.03	1.05	0.04	1.08	0.05
15	0.97	0.13	1.07	0.13	1.24	0.13	1.12	0.19	1.24	0.19	1.34	0.20	1.06	0.17	1.24	0.17	1.39	0.15
16	0.61	0.51	0.67	0.65	0.92	0.56	0.45	0.78	0.51	0.82	1.18	0.23	0.57	0.75	0.41	0.80	2.04	0.10
17	0.64	0.32	0.71	0.33	0.82	0.33	0.86	0.11	1.02	0.11	1.10	0.11	0.77	0.10	0.97	0.09	1.19	0.08
18	1.15	0.46	1.26	0.49	1.46	0.50	0.92	0.50	1.13	0.48	1.44	0.49	0.37	0.69	0.52	0.49	1.22	0.38
19	0.63	0.17	0.69	0.28	0.88	0.08	0.44	0.67	0.52	0.78	0.99	0.08	-0.41	- 1.94	-0.27	- 2.35	1.96	0.11
20	0.95	0.08	1.02	0.08	1.18	0.08	1.13	0.13	1.22	0.14	1.30	0.14	1.09	0.13	1.23	0.12	1.33	0.10
21	0.60	0.38	0.68	0.41	0.90	0.38	0.82	0.37	0.81	0.49	1.26	0.15	0.07	0.17	0.45	0.84	2.40	0.09
22	0.78	0.08	0.88	0.01	1.01	0.01	0.90	0.15	1.05	0.09	1.11	0.10	0.84	0.14	1.02	0.11	1.20	0.08

Table 7. Accuracy and Precision of Simulation Methods, Reliability Index

case	2 RV										5 RV								
	MCS		LH		IS		AISP		AISC		MCS		LH		IS		AISP		AIS
	mean	COV	mean	COV	mean	COV	mean	COV	mean	COV	mean	COV	mean	COV	mean	COV	mean	COV	m
1	n/a		n/a		1.01	0.01	1.02	0.01	1.00	0.00	n/a		n/a		1.00	0.00	1.01	0.01	
2	0.99	0.04	1.03	0.02	1.01	0.03	1.01	0.01	1.00	0.01	1.01	0.03	0.98	0.01	1.00	0.07	1.02	0.01	
3	1.00	0.00	0.98	0.04	0.97	0.07	1.00	0.00	1.01	0.00	0.94	0.08	1.02	0.01	0.97	0.01	1.00	0.00	
4	n/a		n/a		1.00	0.02	1.01	0.00	1.00	0.00	n/a		n/a		1.00	0.00	1.01	0.00	
5	0.94	n/a	1.01	n/a	1.02	0.02	1.02	0.01	1.00	0.00	1.08	n/a	1.00	n/a	1.06	0.03	1.02	0.01	
6	1.00	n/a	1.01	n/a	0.97	0.06	1.02	0.01	1.00	0.00	1.00	0.02	1.05	0.10	1.26	0.40	1.03	0.01	
7	1.01	0.01	1.00	0.00	1.00	0.00	1.02	0.08	0.86	0.12	0.97	0.00	1.04	0.06	1.01	0.01	1.04	0.09	
8	0.99	0.00	0.96	0.05	1.00	0.10	1.05	0.06	0.99	0.06	0.95	0.05	1.03	0.06	1.03	0.10	1.08	0.03	
9	n/a		0.94	n/a	1.00	0.01	1.01	0.01	1.00	0.00	n/a		n/a		0.97	0.04	1.02	0.01	
10	0.98	0.05	1.08	0.15	0.97	0.07	1.05	0.04	1.03	0.04	0.97	0.00	0.97	0.03	0.94	0.01	1.03	0.00	
11	0.98	0.01	0.99	0.02	0.99	0.02	1.02	0.01	1.00	0.00	0.99	0.04	1.00	0.06	0.98	0.05	1.02	0.01	
12	0.99	0.02	1.03	0.02	0.99	0.00	1.02	0.01	1.00	0.01	1.00	0.02	1.01	0.02	1.00	0.00	1.02	0.00	
13	1.72	0.62	1.03	0.03	0.93	0.14	1.13	0.04	1.10	0.05	0.85	0.35	1.00	0.01	1.07	0.10	1.16	0.01	
14	0.98	0.01	0.98	0.00	1.01	0.05	1.00	0.00	1.03	0.01	0.99	0.02	1.02	0.01	1.02	0.05	1.00	0.00	
15	1.00	0.03	1.05	0.03	0.99	0.00	1.03	0.02	1.01	0.01	1.00	0.03	1.10	0.03	1.00	0.01	1.03	0.01	
16	0.98	0.10	1.00	0.02	0.93	0.11	0.91	0.05	0.85	0.05	1.01	0.04	1.00	0.02	0.98	0.04	1.00	0.00	
17	0.99	n/a	1.01	n/a	0.99	0.06	1.01	0.00	1.00	0.00	1.00	0.00	1.07	0.06	1.03	0.03	1.03	0.01	
18	1.01	0.02	1.05	0.03	1.05	0.04	1.12	0.03	1.05	0.02	1.03	0.04	1.16	0.18	1.02	0.00	1.00	0.05	
19	1.00	0.01	1.01	0.00	1.03	0.04	0.99	0.03	1.01	0.01	1.01	0.02	1.14	0.18	1.01	0.03	1.00	0.01	
20	0.99	0.03	1.03	0.05	0.99	0.00	1.02	0.01	0.99	0.01	0.98	0.03	0.98	0.05	1.00	0.00	1.00	0.02	
21	1.81	0.66	1.09	0.12	0.88	0.13	1.17	0.14	1.15	0.17	0.87	0.28	1.01	0.02	1.04	0.03	0.97	0.09	
22	0.98	0.01	1.00	0.02	1.00	0.02	1.02	0.03	1.03	0.02	0.99	0.00	1.00	0.04	1.02	0.03	1.00	0.00	

Table 8. Accuracy and Precision of Analytical Procedure (RF), Reliability Index

case	2 RV		5 RV		15 RV	
	mean	COV	mean	COV	mean	COV
1	1.00	0.00	1.00	0.00	1.00	0.00
2	1.00	0.00	1.00	0.00	1.00	0.00
3	1.00	0.00	1.00	0.00	1.00	0.00
4	1.00	0.00	1.00	0.00	1.00	0.00
5	1.00	0.00	1.00	0.00	1.03	0.03
6	1.00	0.00	1.00	0.00	1.04	0.06
7	1.00	0.00	1.03	0.04	1.10	0.02
8	1.00	0.00	1.00	0.00	1.07	0.00
9	1.00	0.00	1.00	0.00	1.00	0.00
10	1.00	0.00	1.00	0.00	1.05	0.07
11	1.00	0.00	1.00	0.00	1.07	0.07
12	1.00	0.00	1.00	0.00	1.00	0.00
13	1.10	0.04	1.13	0.00	2.68	0.45
14	1.00	0.00	1.00	0.00	1.00	0.00
15	1.00	0.00	1.00	0.00	1.00	0.00
16	1.00	0.00	1.00	0.00	2.23	0.26
17	1.00	0.00	1.00	0.00	1.08	0.07
18	1.00	0.00	1.00	0.00	1.23	0.06
19	1.00	0.00	1.00	0.00	2.14	0.41
20	1.00	0.00	1.00	0.00	1.02	0.03
21	1.04	0.05	1.11	0.00	2.52	0.39
22	1.00	0.00	1.00	0.00	1.09	0.08

Table 9. Accuracy and Precision of Point Sampling Methods, Failure Probability

case	2 RV						5 RV						15 RV					
	N+1		2N		ROS		N+1		2N		ROS		N+1		2N		ROS	
	mean	COV	mean	COV	mean	COV	mean	COV	mean	COV	mean	COV	mean	COV	mean	COV	mean	COV
1	26.21	n/a	11.06	n/a	0.88	n/a	34.80	n/a	8.18	n/a	0.76	n/a	3.18	n/a	1.62	n/a	0.56	n/a
2	2.52	0.55	1.95	0.42	1.04	0.06	1.71	0.26	1.36	0.16	1.18	0.02	1.24	0.11	1.12	0.07	1.24	0.01
3	165.2	1.67	35.90	1.57	1.00	0.01	9.07	1.32	3.20	0.89	1.17	0.02	1.93	0.57	1.36	0.31	1.19	0.03
4	54.14	n/a	17.53	n/a	1.40	n/a	895.4	n/a	55.48	n/a	4.85	n/a	352.0	n/a	10.54	n/a	2.42	n/a
5	80.00	1.41	15.00	1.41	3.00	1.41	122.9	1.36	40.17	1.32	20.51	1.26	112.7	1.35	42.15	1.30	5.46	1.01
6	1510	1.41	563.0	1.41	115.1	1.40	6.55	1.00	0.74	0.76	0.47	1.18	372.3	1.66	8.26	0.86	0.63	0.81
7	2.52	0.59	1.16	0.15	0.75	0.89	6.53	1.11	2.61	0.74	1.31	0.05	12.60	1.23	5.97	1.08	1.21	0.02
8	8.97	1.35	3.98	1.32	2.87	0.78	17.87	1.29	7.97	1.18	4.88	1.00	16.54	1.27	8.31	1.16	2.17	0.53
9	0.79	1.40	0.34	n/a	0.03	n/a	0.67	n/a	0.03	n/a	0.00	n/a	61.60	n/a	2.82	n/a	0.06	n/a
10	1.63	0.91	1.08	0.88	1.14	0.09	2.62	0.32	1.71	0.14	1.43	0.03	4.29	0.72	2.08	0.48	1.54	0.32
11	20.78	1.73	2.88	1.73	1.09	0.03	1.88	0.08	0.68	0.93	0.48	1.26	23.37	1.36	2.03	0.19	0.49	1.05
12	1.06	0.78	0.57	1.31	0.21	1.41	0.29	n/a	0.11	n/a	0.03	n/a	0.76	1.40	0.26	1.41	0.15	n/a
13	1.75	0.23	1.31	0.12	0.81	0.79	1.09	0.46	1.20	0.60	0.50	0.96	1.55	0.39	1.25	0.17	0.36	0.88
14	191.8	1.68	35.51	1.59	0.88	0.17	6.94	1.53	2.15	1.18	0.58	0.55	0.44	1.00	0.47	1.04	0.39	1.24
15	0.72	1.28	0.45	1.40	0.17	1.41	0.16	1.41	0.09	n/a	0.02	n/a	1.69	1.29	0.41	1.41	0.11	n/a
16	1.85	0.05	1.45	0.37	0.92	0.98	2.68	0.12	2.25	0.11	0.60	0.96	7.13	1.05	6.58	1.05	0.33	1.40
17	18.77	1.25	6.71	1.08	1.51	0.48	922.6	1.73	9.78	1.63	1.08	1.16	61.33	1.56	0.96	0.69	0.24	1.63
18	0.69	1.41	0.64	1.41	0.54	1.41	0.91	1.31	0.73	1.41	0.55	1.41	8.83	1.08	4.62	0.85	0.57	1.40
19	7.53	1.07	3.49	0.67	1.58	0.28	11.45	1.10	4.24	0.55	1.35	0.45	8.71	1.13	8.28	1.13	0.36	1.41
20	1.06	0.78	0.57	1.31	0.21	1.41	0.41	n/a	0.11	n/a	0.04	n/a	0.73	1.28	0.17	1.41	0.09	n/a
21	2.15	0.42	1.55	0.07	0.90	0.65	1.21	0.36	1.16	0.56	0.52	0.92	1.71	0.22	1.37	0.16	0.37	0.89
22	191.8	1.68	35.51	1.59	0.88	0.17	8.45	1.49	1.40	1.09	0.41	0.82	4.21	0.52	0.66	1.05	0.26	1.58

Table 10. Accuracy and Precision of Simulation Methods, Failure Probability

case	2 RV										5 RV								
	MCS		LH		IS		AISP		AISC		MCS		LH		IS		AISP		AIS
	mean	COV	mean	COV	mean	COV	mean	COV	mean	COV	mean	COV	mean	COV	mean	COV	mean	COV	m
1	n/a	n/a	n/a	n/a	0.73	n/a	0.78	n/a	0.96	n/a	n/a	n/a	n/a	n/a	1.07	n/a	0.56	n/a	
2	0.98	0.17	0.92	0.04	0.92	0.18	0.94	0.02	0.99	0.01	0.97	0.13	1.09	0.00	1.16	0.42	0.91	0.02	
3	1.01	0.02	1.36	0.40	1.04	0.41	1.01	0.01	0.82	0.14	2.52	0.84	0.85	0.18	1.38	0.17	1.02	0.04	
4	n/a	n/a	n/a	n/a	0.69	n/a	0.81	n/a	1.00	n/a	n/a	n/a	n/a	n/a	0.81	n/a	0.54	n/a	
5	0.94	n/a	1.01	n/a	0.86	0.03	0.84	0.16	0.98	0.04	1.08	n/a	1.00	n/a	0.57	0.64	0.88	0.04	
6	1.00	n/a	1.01	n/a	1.20	0.38	0.81	0.11	0.98	0.03	0.97	0.10	0.66	0.92	0.71	0.93	0.68	0.37	
7	0.97	0.00	1.00	0.00	1.00	0.00	1.13	0.31	1.63	0.23	1.23	0.22	0.96	0.11	0.91	0.14	1.12	0.33	
8	1.02	0.01	1.09	0.03	0.79	0.57	0.92	0.08	1.34	0.40	1.15	0.03	1.06	0.21	0.63	0.45	0.63	0.67	
9	1.12	n/a	2.24	n/a	0.87	0.28	0.62	n/a	0.97	n/a	n/a	n/a	n/a	n/a	1.06	n/a	0.68	n/a	
10	1.01	0.20	0.65	0.97	1.03	0.25	0.85	0.02	0.94	0.09	1.21	0.08	1.12	0.12	1.56	0.28	0.84	0.11	
11	1.16	0.06	1.04	0.11	1.05	0.22	0.74	0.31	1.02	0.04	0.97	0.27	0.89	0.43	0.99	0.34	0.77	0.36	
12	1.05	n/a	0.83	n/a	1.13	0.13	0.73	0.28	0.96	0.06	1.04	n/a	0.90	n/a	0.96	0.06	0.63	0.39	
13	0.51	1.01	0.97	0.01	1.02	0.15	0.75	0.23	0.81	0.14	1.12	0.42	1.01	0.02	0.94	0.13	0.69	0.31	
14	1.10	0.02	1.14	0.09	1.23	0.33	0.99	0.01	0.74	0.31	1.11	0.17	0.88	0.03	1.02	0.23	1.00	0.00	
15	1.00	n/a	0.77	n/a	1.12	0.09	0.79	0.02	0.96	0.05	1.03	n/a	0.36	n/a	0.75	n/a	0.50	0.45	
16	1.18	0.28	1.04	0.07	1.07	0.12	1.35	0.29	1.56	0.38	0.95	0.11	0.98	0.06	1.10	0.14	1.00	0.00	
17	1.04	n/a	0.95	n/a	0.92	0.13	0.90	0.06	0.99	0.02	1.04	0.05	0.59	0.77	0.78	0.15	0.67	0.48	
18	1.00	0.04	0.86	0.06	0.82	0.10	0.57	0.53	0.72	0.44	0.92	0.02	0.74	0.27	0.88	0.14	0.96	0.08	
19	1.04	0.08	0.98	0.01	0.97	0.09	0.96	0.13	0.98	0.02	1.02	0.07	0.83	0.34	0.87	0.24	1.05	0.07	
20	1.05	n/a	0.83	n/a	1.13	0.13	0.82	0.02	1.16	0.26	1.19	n/a	1.29	n/a	1.06	0.13	0.98	0.30	
21	0.67	1.08	0.94	0.13	1.21	0.04	0.84	0.13	0.93	0.03	1.11	0.35	0.97	0.07	0.88	0.15	0.97	0.11	
22	1.10	0.02	1.10	0.20	1.07	0.21	0.88	0.11	0.67	0.33	1.11	0.05	0.93	0.26	0.97	0.30	1.04	0.07	

Table 11. Accuracy and Precision of Analytical Procedure (RF), Failure Probability

case	2 RV		5 RV		15 RV	
	mean	COV	mean	COV	mean	COV
1	1.00	0.00	1.00	0.00	1.00	0.00
2	1.00	0.00	1.00	0.00	1.00	0.00
3	1.00	0.00	1.00	0.00	1.00	0.00
4	1.00	0.00	1.00	0.00	1.00	0.00
5	1.00	0.00	1.00	0.00	0.90	0.16
6	1.00	0.00	1.00	0.00	0.89	0.21
7	1.00	0.00	0.96	0.06	0.66	0.43
8	1.00	0.00	1.00	0.00	0.73	0.36
9	1.00	0.00	1.00	0.00	1.00	0.00
10	1.00	0.00	1.00	0.00	0.83	0.29
11	1.00	0.00	1.00	0.00	0.68	0.41
12	1.00	0.00	1.00	0.00	1.00	0.00
13	0.83	0.13	0.75	0.23	0.37	0.29
14	1.00	0.00	1.00	0.00	1.00	0.00
15	1.00	0.00	1.00	0.00	1.00	0.00
16	1.00	0.00	1.00	0.00	0.24	1.23
17	1.00	0.00	1.00	0.00	0.61	0.59
18	1.00	0.00	1.00	0.00	0.37	1.37
19	1.00	0.00	1.00	0.00	0.25	1.10
20	1.00	0.00	1.00	0.00	0.84	0.33
21	0.97	0.05	0.77	0.19	0.38	0.32
22	1.00	0.00	1.00	0.00	0.62	0.56

Table 12. Special Problems, Reliability Index Results

Problem	N+1	2N	ROS	MCS	LH	IS	IASP	IASC	RF
1 nonlinear	*	*	*	0.97	1.01	0.91	0.99	0.96	1.03
2 multiple B	0.75	0.86	0.98	1.00	0.98	1.31	1.36	1.36	1.37
3 series	1.31	1.28	1.36	0.96	0.95	1.00	1.00	0.99	1.00
4 parallel	1.26	0.67	0.79	1.01	0.99	1.03	1.03	1.01	0.54
5 noisy	1.40	1.11	1.14	1.02	1.05	1.00	1.03	1.00	1.05

*all three point methods predicted approximately zero coefficient of variation.

Table 13. Special Problems, Failure Probability Results

Problem	N+1	2N	ROS	MCS	LH	IS	IASP	IASC	RF
1 nonlinear	*	*	*	1.47	0.92	13.63	1.12	1.77	0.67
2 multiple B	26.7	7.34	1.31	1.00	1.41	0.005	0.002	0.002	0.001
3 series	0.10	0.14	0.07	1.28	1.34	1.00	1.03	1.05	1.03
4 parallel	0.03	29.6	10.1	0.84	1.07	0.65	0.70	0.90	76.63
5 noisy	0.09	0.55	0.47	0.89	0.78	1.00	0.85	1.00	0.78

*all three point methods predicted approximately zero coefficient of variation.

Table 14. Accuracy and Precision by Type of Limit State and Evaluation Method, Reliability Index

Method	Type of Problem													
	OVERALL		norm, low COV		low RV		low RV, linear		normal, linear		low COV		low COV & RV	
	mean	COV	mean	COV	mean	COV	mean	COV	mean	COV	mean	COV	mean	COV
n+1	0.79	0.14	0.99	0.06	0.80	0.20	0.81	0.16	0.90	0.00	0.92	0.14	0.92	0.13
2n	0.88	0.15	1.04	0.06	0.88	0.21	0.89	0.16	0.94	0.00	1.01	0.13	0.99	0.13
ROS	1.12	0.12	1.13	0.06	0.99	0.17	0.93	0.10	0.99	0.01	1.16	0.14	1.15	0.15
MCS	1.01	0.06	0.98	0.02	1.07	0.10	0.99	0.02	0.99	0.04	0.99	0.01	1.00	0.02
LH	1.04	0.08	1.01	0.02	1.02	0.04	1.00	0.05	1.00	0.02	1.04	0.04	1.02	0.03
IS	1.01	0.06	1.00	0.01	0.99	0.05	0.99	0.04	1.00	0.04	1.02	0.02	1.01	0.01
AISP	1.04	0.03	1.02	0.01	1.03	0.03	1.02	0.03	1.01	0.02	1.03	0.02	1.03	0.04
AISC	1.01	0.03	1.00	0.00	1.01	0.03	1.00	0.03	1.01	0.03	1.00	0.02	0.99	0.03
RF	1.10	0.03	1.00	0.00	1.01	0.00	1.00	0.00	1.00	0.00	1.02	0.01	1.00	0.00

(grey) for best accuracy (purple) for best precision (grey + purple) best for both accuracy and precision

Table 15. Accuracy and Precision by Type of Limit State and Evaluation Method, Failure Probability

Method	Type of Problem											
	OVERALL		low RV		normal		low B,V		normal, linear		low COV & RV	
	mean	COV	mean	COV	mean	COV	mean	COV	mean	COV	mean	COV
n+1	82.00	1.02	104.0	1.08	25.00	0.85	1.78	0.39	27.00	0.75	10.90	1.04
2n	15.00	0.91	33.70	1.04	6.00	0.77	1.31	0.34	7.30	0.57	4.04	1.12
ROS	3.03	0.84	6.23	0.73	0.72	0.47	0.96	0.29	1.00	0.02	0.52	1.31
MCS	1.11	0.18	1.00	0.23	1.16	0.28	1.02	0.06	1.26	0.25	1.03	0.02
LH	0.95	0.34	1.04	0.16	1.01	0.24	0.85	0.38	1.11	0.17	1.09	0.03
IS	1.20	0.36	0.99	0.19	1.13	0.31	0.98	0.26	0.91	0.30	0.94	0.12
AISP	0.78	0.22	0.87	0.15	0.82	0.09	0.77	0.22	0.88	0.19	0.78	0.22
AISC	0.95	0.20	1.01	0.16	0.94	0.14	0.95	0.18	0.94	0.20	0.98	0.21
RF	0.90	0.13	0.99	0.01	0.94	0.04	1.04	0.07	1.00	0.00	1.00	0.00

for best accuracy (grey) — for best precision (purple) — best for both accuracy and precision (grey + purple)

Table 16. Number of Samples

Method	Type of Problem 2 RV	5 RV	15 RV
n+1	3	6	16
2n	4	10	30
ROS	4	10	30
MCS		3700 / 1.7x10^6*	
LH		370 / 34000*	
IS	60	69	99
AISP	85	94	124
AISC	51	69	194
RF	10	19	49