Adaptive Optimal Sampling Methodology for Zone-Based Probabilistic Life Prediction

Michael P. Enright*
Southwest Research Institute, San Antonio, TX, 78238, USA

Harry R. Millwater†
University of Texas at San Antonio, San Antonio, TX, 78249, USA

Luc Huyse‡
Southwest Research Institute, San Antonio, TX, 78238, USA

Simulation-based system reliability prediction may require significant computations, particularly when the expected value of the system failure probability is relatively low. In this paper, a methodology is presented for variance reduction of sampling-based series system reliability predictions based on optimal allocation of Monte Carlo samples to the individual failure modes. An algorithm is presented for adaptively allocating samples to member failure modes based on initial estimates of the member failure probabilities $p_i$. The methodology is demonstrated for a simple series system and a gas turbine engine disk modeled using a zone-based series system approach. For the example considered, it is shown that the computational accuracy of the method does not appear to depend on the initial $p_i$ estimate. However, the computational efficiency is highly dependent on the initial $p_i$ estimate. The results can be applied to improve the efficiency of sampling-based series system reliability predictions.

Nomenclature

- $COV_i = \text{coefficient of variation of the failure probability for member } i$
- $E(\bullet) = \text{mean value}$
- $g = \text{limit state function}$
- $i, j = \text{member number}$
- $k_{\alpha/2} = \text{standard normal variate}$
- $m = \text{number of members in system}$
- $N = \text{number of samples associated with risk prediction of the system}$
- $n = \text{vector containing } n_i \text{ for all } m \text{ members in the system}$
- $n_i = \text{number of samples associated with risk prediction of member } i$
- $p_f, P_f = \text{system failure probability, estimator of } p_f$
- $p_i, P_i = \text{failure probability associated with member } i, \text{ estimator of } p_i$
- $R_i = \text{resistance of member } i$
- $S_i = \text{applied stress for member } i$
- $Var(\bullet) = \text{variance}$
- $Z(n) = \text{objective function}$
- $\varepsilon = \text{convergence tolerance}$
- $(1-\alpha) = \text{confidence level}$
- $\gamma = \text{relative sampling error}$
- $\lambda = \text{Lagrange multiplier}$

* Principal Engineer, Reliability and Materials Integrity, 6220 Culebra Road, Member AIAA.
† Assistant Professor, Mechanical Engineering, 6900 N. Loop 1604 W., Member AIAA.
‡ Senior Research Engineer, Reliability and Materials Integrity, 6220 Culebra Road, Member AIAA.
\[ \phi(n, \lambda) = \text{Lagrange function} \]

\[ \sigma_i = \text{standard deviation of the failure probability for member } i. \]

I. Introduction

Aircraft gas turbine engine rotors and disks may occasionally contain material anomalies that can form during the manufacturing process which can lead to uncontained failure of the engine\(^1\). The occurrence of metallurgical anomalies is relatively rare (i.e., expected value on the order of one anomaly per million pounds of disk material\(^2,3\)), so a probabilistic approach has been developed for the fracture mechanics-based life prediction of components which may contain the anomalies\(^4,8\). A zone-based method is used to address the location uncertainty associated with these anomalies in which a component is subdivided into a number of zones of approximately equal risk. Component failure is modeled as a series system of zones, in which failure of any zone is interpreted as failure of the component.

Reliability prediction of general systems may require significant computations, particularly if correlation among members and post-failure material behavior are considered\(^6,10\). Over the past several decades, a number of approaches have been used to improve the efficiency of computations associated with system reliability predictions, such as bounding methods (e.g., Refs. 11-12) and various variance reduction techniques (e.g., Refs. 13-15, among many others). Over the past several years, a number of methods have been developed to improve the efficiency of zone-based system reliability predictions, such as tailored response surface\(^16\), zone refinement\(^17\), importance sampling\(^16,18\), and parallel processing\(^19\), among other improvements.

In this paper, a methodology is presented for variance reduction of sampling-based series system reliability predictions based on optimal allocation of Monte Carlo samples to the individual failure modes. It is an extension of a technique developed previously for mutually exclusive failure modes\(^20,21\). An algorithm is presented for adaptively allocating samples to member failure modes based on initial estimates of the individual member failure probabilities. The methodology is demonstrated for a simple series system and a gas turbine engine disk modeled using the zone-based series system approach.

II. Optimal Sampling Method for Series Systems

Consider the series system shown in Fig. 1 which consists of five members with an applied load. Illustrative values of the bivariant probability densities (pdfs) for the individual member failure modes are shown in Fig. 2. Also shown is the member limit state

\[ g(R_i, S_i) = R_i - S_i \]  \hspace{1cm} (1)

where \( R_i \) and \( S_i \) are resistance and applied stress for member \( i \), respectively. From Fig. 2, it can be observed that members 1, 4, and 5 have a significant probability of occurrence in the failure region (\( R_i \leq S_i \)) and should have a larger contribution to system failure compared to members 2 and 3. An importance sampling strategy would place the samples near the limit state (i.e., near members 1, 4, and 5). However, some samples should be provided for members 2 and 3 to quantify their contribution to system failure. The optimal allocation of samples is identified by minimizing the variance of the system failure probability, described in the equations below.

The failure probability \( P_f \) of a series system of \( m \) independent members can be expressed as:

\[ P_f = 1 - \prod_{i=1}^{m} [1 - P_i] \]  \hspace{1cm} (2)

where \( 1 - P_i \) is the survival probability associated with member \( i \). The variance of \( P_f \) is
Figure 2. Illustrative bivariant probability densities reveal the relative values of member failure probabilities in a system.

\[
Var(P_i) = \prod_{i=1}^{m} E\left((1-P_i)^2\right) - \left\{ \prod_{i=1}^{m} E\left(1-P_i\right) \right\}^2
\] (3)

For a large number of samples \(n_i\), the member survival probability can be approximated as a normal distribution with the following mean and variance:

\[
E(1-P_i) = 1 - p_i
\] (4)

\[
Var(1-P_i) = p_i (1-p_i)/n_i
\] (5)

Noting that

\[
E\left((1-P_i)^2\right) = \left(E[1-P_i]\right)^2 + Var(1-P_i)
\] (6)
Eqn. (3) becomes

\[
Var(P_i) = \prod_{i=1}^{m} \left[ (1-p_i)^2 + \frac{p_i(1-p_i)}{n_i} \right] - \prod_{i=1}^{m} (1-p_i)^2
\]  

(7)

To identify values of \( n_i \) that minimize \( Var(P_i) \), the following optimization formulation is used:

\[
\min Z(n) = \prod_{i=1}^{m} \left[ (1-p_i)^2 + \frac{p_i(1-p_i)}{n_i} \right] - \prod_{i=1}^{m} (1-p_i)^2
\]  

(8)

subject to

\[
N = \sum_{i=1}^{m} n_i
\]  

(9)

where \( n_i \) is the number of Monte Carlo samples used to estimate the survival probability of member \( i \) and \( N \) is the total number of samples for the system.

Eqns. 8 and 9 are solved using the Lagrange-multiplier method.

\[
\phi \{ n_i, \lambda \} = \prod_{i=1}^{m} \left[ (1-p_i)^2 + \frac{p_i(1-p_i)}{n_i} \right] - \prod_{i=1}^{m} (1-p_i)^2 + \lambda \left( N - \sum_{i=1}^{m} n_i \right)
\]  

(10)

subject to the following conditions at the optimum:

\[
\frac{\partial \phi}{\partial n_i} = 0 \quad i = 1, 2, \ldots m
\]  

(11)

\[
\frac{\partial \phi}{\partial \lambda} = 0
\]  

(12)

Applying these conditions to Eqn. 10:

\[
\frac{\partial \phi}{\partial n_i} = -p_i(1-p_i) \prod_{i=1}^{m} \left[ (1-p_j)^2 + \frac{p_j(1-p_j)}{n_j} \right] - \lambda = 0
\]  

(13)

\[
\frac{\partial \phi}{\partial \lambda} = N - \sum_{i=1}^{m} n_i = 0
\]  

(14)

Combining Eqns. 13 and 14:

\[
n_i = \frac{p_i(1-p_i) \prod_{j=1}^{m} \left[ (1-p_j)^2 + \frac{p_j(1-p_j)}{n_j} \right]}{-\lambda}
\]  

(15)
Substituting Eqn. 15 into 14:

\[
N = \sum_{j=1}^{m} \sqrt{p_j (1-p_j) \prod_{j=1}^{m} \left(1-p_j \right)^2 + \frac{p_j (1-p_j)}{n_j}} - \lambda
\]  

(16)

Combining Eqn. 15 and 16, the optimal number of samples for each member failure mode is given by:

\[
n_j = N \sqrt{p_j (1-p_j) \prod_{j=1}^{m} \left(1-p_j \right)^2 + \frac{p_j (1-p_j)}{n_j}} \frac{1}{\sum_{j=1}^{m} \sqrt{p_j (1-p_j) \prod_{j=1}^{m} \left(1-p_j \right)^2 + \frac{p_j (1-p_j)}{n_j}}}
\]

(17)

III. Optimal Sampling Application: 5-Member Series System

To illustrate the methodology, Eqn. (17) is used to identify the optimal number of samples for the five member series system shown in Fig. 1. Member resistances \(R_i\) are modeled as independent lognormal random variables each subjected to a deterministic stress \(S_i\). The main descriptors for these variables are indicated in Table 1, and associated probability densities are shown in Fig. 3. The limit state is given by Eqn. (1) and can be solved analytically for each member failure mode. Analytical member failure probability results \(P_i\) are indicated in Table 2.

Eqn. (1) can also be solved using numerical simulation with the variance of the member and system failure probabilities given by Eqns. 5 and 7, respectively. Suppose that one million Monte Carlo samples are used to predict the failure of the system. For “uniform sampling”, the same number of samples is allocated to the failure prediction of each member (i.e., \(n_i = 200,000\) for each member, see Table 2). For optimal sampling, Eqn. (17) is used to specify the number of samples for each member based on the analytical solution for \(P_i\). The optimal number of samples for each member is specified by:

\[
\sum_{j=1}^{m} \sqrt{p_j (1-p_j) \prod_{j=1}^{m} \left(1-p_j \right)^2 + \frac{p_j (1-p_j)}{n_j}} \frac{1}{\sum_{j=1}^{m} \sqrt{p_j (1-p_j) \prod_{j=1}^{m} \left(1-p_j \right)^2 + \frac{p_j (1-p_j)}{n_j}}}
\]

Table 1. Applied and ultimate stress values associated with the 5-member series system example.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Units</th>
<th>Name</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>(S)</td>
<td>MPa</td>
<td>Applied Stress</td>
<td>100.0</td>
<td>0.0</td>
<td>Deterministic</td>
</tr>
<tr>
<td>(R_1)</td>
<td>MPa</td>
<td>Ultimate Stress, Member 1</td>
<td>190.0</td>
<td>38.0</td>
<td>Lognormal</td>
</tr>
<tr>
<td>(R_2)</td>
<td>MPa</td>
<td>Ultimate Stress, Member 2</td>
<td>200.0</td>
<td>40.0</td>
<td>Lognormal</td>
</tr>
<tr>
<td>(R_3)</td>
<td>MPa</td>
<td>Ultimate Stress, Member 3</td>
<td>215.0</td>
<td>43.0</td>
<td>Lognormal</td>
</tr>
<tr>
<td>(R_4)</td>
<td>MPa</td>
<td>Ultimate Stress, Member 4</td>
<td>225.0</td>
<td>45.0</td>
<td>Lognormal</td>
</tr>
<tr>
<td>(R_5)</td>
<td>MPa</td>
<td>Ultimate Stress, Member 5</td>
<td>240.0</td>
<td>48.0</td>
<td>Lognormal</td>
</tr>
</tbody>
</table>
Values for these variables are indicated in Table 2. The variance of the system failure probability is also indicated in Table 2. Comparing the values of the system failure probability $P_f$ to the individual member failure probabilities $P_i$, it can be observed that members 1 and 2 contribute significantly to system failure (over 90% of system failure can be attributed to these two members). When optimal sampling is used, the failure probability COV values of members 1 and 2 are smaller compared to the values for uniform sampling, leading to an overall reduction in the variance of the system failure probability. On the other hand, the failure probability COV is larger for the remaining members in the system when optimal sampling is used (the member 5 failure probability COV value associated with optimal sampling is more than double the value associated with uniform sampling). Probability densities for the $P_i$ estimates associated with the uniform and optimal sampling methods are shown in Fig. 4. The influence of optimal sampling on the pdf of $P_i$ is significant for all of the members considered, particularly for the strongest and weakest members (5 and 1, respectively).

![Figure 3. Probability densities associated with member resistances for the 5-member structural system shown in Fig. 1.](image)

<table>
<thead>
<tr>
<th>Member</th>
<th>$P_i$</th>
<th>Uniform Sampling</th>
<th>Optimal Sampling</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$n_i$</td>
<td>$\sigma_i$</td>
</tr>
<tr>
<td>1</td>
<td>9.38E-04</td>
<td>200000</td>
<td>6.84E-05</td>
</tr>
<tr>
<td>2</td>
<td>3.82E-04</td>
<td>200000</td>
<td>4.37E-05</td>
</tr>
<tr>
<td>3</td>
<td>9.68E-05</td>
<td>200000</td>
<td>2.20E-05</td>
</tr>
<tr>
<td>4</td>
<td>3.83E-05</td>
<td>200000</td>
<td>1.38E-05</td>
</tr>
<tr>
<td>5</td>
<td>9.46E-06</td>
<td>200000</td>
<td>6.88E-06</td>
</tr>
<tr>
<td>System</td>
<td>1.46E-03</td>
<td>1000000</td>
<td>8.55E-05</td>
</tr>
</tbody>
</table>
Figure 4. Probability densities associated with $p_i$ estimates for the uniform and optimal sampling strategies.

IV. Adaptive Optimal Sampling

Eqn. (17) indicates that the optimum number of samples for an individual member failure mode is dependent on knowledge of the failure probability $P_i$ for all of the member failure modes. However, since $p_i$ values are unknown prior to sampling, initial estimates of these values are required to make an estimate of the optimal number of samples in each zone. The efficiency and accuracy of the optimal sampling method is highly dependent on the number of samples $n_i$ used for the initial estimate of $p_i$. If $n_i$ is too large, the method is inefficient (i.e., provides no benefit over uniform sampling). On the other hand, if $n_i$ is too small, the method may be inaccurate. For example, if the number of samples for a member is so small that the limit state is not violated for any of the samples, then $p_i$ is estimated as equal to zero and $n_i$ is also zero. Accuracy of the initial estimate can be improved by setting a minimum value for $p_i$ when no samples violate the limit state (i.e., $p_i = 1/ n_i$). Depending on the limit state for a member, $p_i$ can sometimes be estimated using a non-sampling-based method (e.g., numerical integration of dominant random variable(s), among others).

The optimal number of samples for each failure mode can be computed adaptively using the following procedure:

1. Estimate $p_i$ using a small number of Monte Carlo samples or other efficient strategy.
2. Compute optimal number of samples $n_i$ for each member failure mode using Eqn. 17.
3. Compute $p_i$ and $P_f$ using Monte Carlo sampling with estimated optimal $n_i$.
4. Repeat steps 2-3 until $\Delta \sum n_i \leq \varepsilon$.

If the individual $p_i$ values are relatively small, the optimal number of samples for the system can be estimated as$^{20,21}$:
where $k_{a/2}$ = standard normal variate for a $(1 - \alpha)$ confidence level, and $\gamma = (p_j - p_f)/P_f =$ sampling error.$^{16}$

V. Application of Adaptive Optimal Sampling to Zone-Based System

The adaptive optimal sampling method is illustrated for the aircraft rotor disk shown in Fig. 5. The disk is discretized into 64 zones (i.e., it is modeled as a series system consisting of 64 members). A fracture mechanics-based limit state is applied to each zone, and three primary random variables are used to model the uncertainties associated with material properties, applied stress, and initial anomaly size. Individual $p_i$ values are relatively small, so Eqn. (20) can be used to estimate the optimal number of samples for the system with $k_{a/2} = 95\%$ and $\gamma = 10\%$.

Complete details regarding the disk and values of the associated deterministic and random variables are provided in Ref. 18.

The following methods are used to estimate initial $p_i$ values prior to sampling: (1) uniform sampling – same number of samples in each zone (i.e., 100, 1000, and 10000 samples per zone), (2) first failure – sampling is performed in a zone until either the limit state is violated or a specified number of samples (100 or 1000 samples per zone) have been applied ($p_i$ is estimated as one divided by the number of samples to first failure), and (3) critical defect – $p_i$ is estimated directly assuming that the dominant variable (initial anomaly size) is random and the remaining variables are deterministic (see Ref. 4 for details). For methods (1) and (2), $p_i$ is set equal to $1/n_i$ if no limit state violations occur within $n_i$ for a given zone.

The influence of the initial $p_i$ estimate on the predicted probability of failure is shown in Figs. 6, 7, and 8 for a relatively weak zone, a strong zone, and the disk, respectively. In Fig. 6, the error in the initial $p_i$ estimate for a weak zone is between 1% and 29% depending on the method used, but is within 8% for the converged solution (i.e., after the adaptive optimal sampling algorithm is applied). In Fig. 7 it is shown that the sampling error for a strong zone is 1-67% for the initial estimate and 5-65% for the converged solution. Since the strong zone has a small influence on $P_f$, the relatively large post convergence sampling error associated with some of the methods has a relatively small impact on the error associated with $P_f$. Sampling error for the system is shown in Fig. 8, with values of 0.5-660% and 2-8% for the initial and converged estimates, respectively. As shown in Fig. 8, all converged $P_f$ values fall within the 10% error bounds at 95% confidence, regardless of the method used for the initial prediction of $p_i$ values. This result suggests that sampling accuracy is somewhat independent of the method used to estimate initial $p_i$ values, provided that converged values are used to compute $P_f$.

The influence of the method used to estimate $p_i$ on computational efficiency is shown in Figs. 9, 10, and 11 for a weak zone, a strong zone, and the system, respectively. For the weak zone (Fig. 9), a wide range of values for the number of samples (i.e., limit state evaluations) is required to estimate $p_i$ for the various methods considered. However, converged solutions require a similar number of samples ranging from 3774 (uniform sampling with 10000 samples) to 10,600 (uniform sampling with 100 samples per zone). For the strong zone (Fig. 10), the range of samples for converged solutions ranges from 591 (critical defect) to 10,000 (uniform sampling with 10,000 samples per zone). The initial estimate appears to have a significant influence on the sampling efficiency of strong zones, with the most efficient methods requiring the fewest number of samples for the initial estimate. The number of samples associated with the system is shown in Fig. 11. For the converged solution, the number of samples ranges from

\[
N = \frac{k_{a/2}^2}{\gamma^2 P_f^2} \left[ \sum_{i=1}^{n} \sqrt{p_i} \right]^2
\]

Figure 5. The adaptive optimal sampling methodology is illustrated for an aircraft rotor disk (see Ref. 18 for details).
about 1.0E5 (critical defect) to over 2.0E6 (uniform sampling with 100 samples per zone). These results indicate that sampling efficiency is highly dependent on the method used to estimate initial $p_i$ values.

Figure 6. Influence of initial $p_i$ estimates on failure probability predictions for a relatively weak zone.

Figure 7. Influence of initial $p_i$ estimates on failure probability predictions for a relatively strong zone.
Figure 8. Influence of initial $p_i$ estimates on failure probability prediction of disk $P_f$ (system).

Figure 9. Influence of initial $p_i$ estimates on number of limit state evaluations required to accurately predict failure probability for a relatively weak zone.
Figure 10. Influence of initial $p_i$ estimates on number of limit state evaluations required to accurately predict failure probability for a relatively strong zone.

Figure 11. Influence of initial $p_i$ estimates on number of limit state evaluations required to accurately predict disk (system) failure probability.
VI. Summary

A methodology was presented for variance reduction of sampling-based series system reliability predictions based on optimal allocation of Monte Carlo samples to individual failure modes. An algorithm was presented for adaptively allocating samples to member failure modes based on initial estimates of the individual member failure probabilities. The optimal sampling methodology was demonstrated for a simple series system in which it was shown that the variance of the system failure probability is reduced compared to a uniform sampling approach, due to the reduced variances in the failure probabilities associated with the weakest members. The adaptive optimal sampling methodology was illustrated for a gas turbine engine disk modeled using several methods to estimate the failure probability $p_i$ in each zone prior to optimal sampling. For the example problem considered, it was shown that the computational accuracy of the method does not appear to depend on the initial $p_i$ estimate, whereas the computational efficiency is highly dependent on the initial $p_i$ estimate. The results can be applied to improve the efficiency of sampling-based series system reliability predictions.

Acknowledgments

This work was supported by the Federal Aviation Administration under Cooperative Agreement 95-G-041 and Grant 99-G-016. The authors wish to thank Joe Wilson (FAA Technical Center project manager) and Tim Mouzakis (FAA Engine and Propeller Directorate) for their continued support. The ongoing contributions of the Industry Steering Committee (Darryl Lehmann, Pratt & Whitney; Jon Tschopp, General Electric; Ahsan Jameel, Honeywell; Jon Dubke, Rolls-Royce) are also gratefully acknowledged.

References


