

Probabilistic Reanalysis Using Monte Carlo Simulation

Efstathios Nikolaidis, Sirine Saleem and Farizal

The University of Toledo

Geng Zhang and Zissimos P. Mourelatos

Oakland University

Copyright © 2008 SAE International

ABSTRACT

An approach for Probabilistic Reanalysis (PRA) of a system is presented. PRA calculates very efficiently the system reliability or the average value of an attribute of a design for many probability distributions of the input variables, by performing a single Monte Carlo simulation. In addition, PRA calculates the sensitivity derivatives of the reliability to the parameters of the probability distributions. The approach is useful for analysis problems where reliability bounds need to be calculated because the probability distribution of the input variables is uncertain or for design problems where the design variables are random. The accuracy and efficiency of PRA is demonstrated on vibration analysis of a car and on system reliability-based optimization (RBDO) of an internal combustion engine.

INTRODUCTION

PROBLEM DEFINITION AND SIGNIFICANCE

A designer often needs to estimate repeatedly the reliability or the average utility of a design for different probability distributions of the input variables. Two such cases are described below:

- 1) Because there is considerable variability in the design variables in an optimization problem, a designer wants to optimize the average values of these variables. For example, the designer wants to optimize the average dimensions of the structural members of a car body in order to maximize the probability of acceptable performance in a crash test.
- 2) A designer can only estimate lower and upper bounds for the parameters of the probability distributions of the random variables in a probabilistic analysis problem or he/she does not know the type of the probability distribution of the random variables. Consequently, the designer wants to estimate how low or high the system reliability can be.

It is often impractical to solve these problems because this requires repeated probabilistic analyses of the system for different probability distributions of the input variables and because these analyses are expensive.

Many methods have been presented for circumventing the cost of probabilistic analysis. Yet, the cost of probabilistic analysis and design is still a significant issue hindering the application of probabilistic methods to large scale, complex systems. This paper presents an approach for reducing the computational cost of solving probabilistic design problems with random design variables by several orders of magnitude.

BACKGROUND

In probabilistic design, a designer seeks the best configuration to maximize the average utility $U(\mathbf{d}, \mathbf{X}, \mathbf{P})$ of a design. Utility depends on input variables that can be categorized into deterministic design variables $\mathbf{d}$, random design variables $\mathbf{X}$, and random parameters $\mathbf{P}$. Designers can only change the deterministic design variables and the average values of the random design variables. In a design optimization problem, a designer seeks the optimum values of the deterministic design variables and the average values of the random design variables in order to maximize the expected utility,

$$\text{Find } \mathbf{d}, \mu_{\mathbf{X}} \quad (1a)$$

$$\text{To maximize } E[U(\mathbf{d}, \mathbf{X}, \mathbf{P})]$$

$$\text{s. t. } \mathbf{d}^L \leq \mathbf{d} \leq \mathbf{d}^U, \text{ and } \mu_{\mathbf{X}}^L \leq \mu_{\mathbf{X}} \leq \mu_{\mathbf{X}}^U$$

Symbol $E(\cdot)$ is the average value operator, while $\mathbf{d}^L$ and $\mathbf{d}^U$, and $\mu_{\mathbf{X}}^L$ and $\mu_{\mathbf{X}}^U$ represent lower and upper bounds of the deterministic design variables and the average values of the random design variables, respectively. A bold letter indicates a vector, an upper case letter indicates a random variable or parameter and a lower case letter indicates a realization of a random variable or parameter. In subsequent formulations of the design problem, the side constraints on the design variables in the third line of Eq. (1a) will be omitted.

Often, it is difficult to construct the utility function for a design. Therefore, alternative formulations of the above design problem are considered. A designer could minimize (or maximize) the average value of an attribute $L(\mathbf{d}, \mathbf{X}, \mathbf{P})$, such as the cost or weight, and impose

constraints on other attributes. Failure of some systems is catastrophic, and the boundary between success and failure is crisp (for example, failure of an aircraft wing due to aeroelastic divergence or collapse of a bridge). In this case, it is common practice to seek to minimize one attribute (for example the weight or cost) and satisfy minimum acceptable requirements for the remaining attributes. The system is idealized so that it fails under a finite number of failure modes and the probability of system failure is calculated from the probabilities of the modes. The probability of system failure is required to be acceptably low. Let $I(\mathbf{d}, \mathbf{X}, \mathbf{P})$ be the system failure indicator function, which is defined to be one if the system fails and zero if it survives. The design problem in Eq. (1a) reduces to a system Reliability-Based Design Optimization (RBDO) problem that seeks the most efficient design whose system probability of failure (violation of the minimum requirements for the remaining attributes) does not exceed an allowable value p_f^{all} . The formulation of the system RBDO problem is as follows,

Find $\mathbf{d}, \mu_{\mathbf{X}}$ (1b)

To minimize $E[L(\mathbf{d}, \mathbf{X}, \mathbf{P})]$

s. t. $p_{sys} = P[I(\mathbf{d}, \mathbf{X}, \mathbf{P}) = 1] \leq p_f^{all}$

where p_{sys} is the system probability of failure. Often, $E[L(\mathbf{d}, \mathbf{X}, \mathbf{P})]$ is approximated by the value of this attribute for the average values of the design variables and parameters $I[\mathbf{d}, E(\mathbf{X}), E(\mathbf{P})]$.

Probabilistic design problems are often very expensive to solve because the expected utilities, or failure probabilities, of tens to thousands of alternative designs should be calculated. Moreover, each calculation of the expected utility or system failure probability requires from few tens to several million deterministic analyses. For a complex structure, such as a car body in crash, each deterministic analysis could require several minutes or hours of CPU time.

Studies on reducing the computational cost of probabilistic design have employed the following approaches:

- a) Reduce the cost of repeated deterministic analyses by using inexpensive approximations of the attributes. These can be categorized into global, local or combined approximations (Kirsch, 2006).
- b) Reduce the number of repeated deterministic analyses. For example, if the system failure probability is estimated using Monte Carlo simulation, many techniques can be used to reduce the number of replications. These include Importance Sampling (Melchers, 1999), Adaptive Sampling (Au and Beck, 1999, Karamchandani et al.

1989) and the methods of Antithetic Variates, Control Variates, Indirect Estimation and Conditioning (Law, 2007).

- c) Estimate efficiently the moments of the attributes and use the moments to estimate the probability of failure. Rahman and Xu (2004), approximated the moments of the attributes by expressions requiring one-dimensional integration. Youn et al. (2006) used this method for efficient estimation of system reliability.
- d) Approximate the tail of the cumulative probability distribution of the performance function using a probability distribution that is suitable for representing rare events, such as the Extreme I to III types of distributions and the Generalized Pareto Distribution (Kim and Ramu, 2006).

Farizal and Nikolaidis (2006) presented a PRA approach for estimating lower and upper bounds of the probabilities of failure when the distribution parameters of the random variables are known up to their lower and upper bounds. They demonstrated the approach on a single degree of freedom system with a dynamic vibration absorber with uncertain natural frequencies. They showed that PRA can estimate bounds for the probability of failure at one fiftieth of the cost of Monte Carlo simulation.

A method for computing the sensitivity derivatives of the system failure probability to the average values of the random design variables can reduce the computational cost and improve the robustness of the algorithm for optimum search. Karamchandani et al. (1988) presented a finite difference method for estimating these sensitivities for simulation-based schemes by using importance sampling.

OBJECTIVES AND SCOPE

This paper proposes to circumvent the high computational cost of probabilistic design by employing PRA. This approach requires only one Monte Carlo simulation and estimates the probabilities of failure or expected utilities of tens or hundreds of designs very efficiently using importance sampling. The objectives of this paper are to; a) present the PRA approach, b) integrate this approach in a methodology for probabilistic design, and c) demonstrate the efficacy of the methodology on design problems. The study is confined to design problems where all design variables are random.

PRA calculates very efficiently the system reliability or the average value of an attribute for many probability distributions of the input variables, using the results of a single Monte Carlo simulation analysis for a sampling distribution of these input variables. The method is very efficient because it uses the sample values of the attribute that were already calculated in this Monte Carlo simulation in order to estimate the reliability or the average value of the attribute for other distributions.

addition, the method estimates efficiently the sensitivity derivatives of the reliability with respect to the distribution parameters, which helps identify the important design variables, reduces the cost of optimization and improves the convergence of optimization. The estimates of the reliability and their derivatives with respect to the distribution parameters vary smoothly with the latter parameters, which helps compare alternative designs.

OUTLINE

First, the PRA approach is described. The equations for PRA and for sensitivity analysis are presented. The selection of a suitable sampling probability distribution is also discussed. Then, the paper demonstrates the PRA approach on probabilistic analysis of a complex vehicle model and on RBDO of an internal combustion engine.

EFFICIENT PRA FOR DESIGN UNDER UNCERTAINTY

This section reviews the importance sampling method for estimation of the probability of failure of a system and the average value of a function of random variables. Then it presents the efficient PRA approach and explains its use in probabilistic design. Finally, the selection of a sampling probability distribution is discussed.

IMPORTANCE SAMPLING

The probability of failure of the system in Eq. (1b) is the integral of the joint probability density function (PDF) $f_{\mathbf{X},\mathbf{P}}(\mathbf{x},\mathbf{p})$ of the random design variables $\mathbf{X}$, and parameters $\mathbf{P}$, over the range of $\mathbf{X}$ and $\mathbf{P}$,

$$p_{\text{sys}} = \int_{\mathbf{X},\mathbf{P}} I(\mathbf{x},\mathbf{p}) f_{\mathbf{X},\mathbf{P}}(\mathbf{x},\mathbf{p}) d\mathbf{x}d\mathbf{p} \quad (2)$$

where $I(\mathbf{x},\mathbf{p})$ is the failure indicator function of the system.

In importance sampling, sample values of the random variables are generated using a PDF $f^S_{\mathbf{X},\mathbf{P}}(\mathbf{x},\mathbf{p})$ that yields many failures with high probability of occurrence, instead of the true PDF, $f_{\mathbf{X},\mathbf{P}}(\mathbf{x},\mathbf{p})$. The estimator of the probability of failure in this case is derived as follows. If sample values of the random variables are generated from $f^S_{\mathbf{X},\mathbf{P}}(\mathbf{x},\mathbf{p})$ then equation (2) becomes,

$$p_{\text{sys}} = \int_{\mathbf{X},\mathbf{P}} I(\mathbf{x},\mathbf{p}) \frac{f_{\mathbf{X},\mathbf{P}}(\mathbf{x},\mathbf{p})}{f^S_{\mathbf{X},\mathbf{P}}(\mathbf{x},\mathbf{p})} f^S_{\mathbf{X},\mathbf{P}}(\mathbf{x},\mathbf{p}) d\mathbf{x}d\mathbf{p} \quad (3)$$

Then an unbiased estimator of the probability of failure is,

$$\begin{aligned} \hat{p}_{\text{sys}} &= \frac{1}{n} \sum_{i=1}^n I(\mathbf{x}_i, \mathbf{p}_i) \frac{f_{\mathbf{X},\mathbf{P}}(\mathbf{x}_i, \mathbf{p}_i)}{f^S_{\mathbf{X},\mathbf{P}}(\mathbf{x}_i, \mathbf{p}_i)} \\ &= \frac{1}{n} \sum_{i=1}^n \frac{f_{\mathbf{X},\mathbf{P}}(\mathbf{x}_i, \mathbf{p}_i)}{f^S_{\mathbf{X},\mathbf{P}}(\mathbf{x}_i, \mathbf{p}_i)} \end{aligned} \quad (4a)$$

where $\mathbf{x}_i$, and $\mathbf{p}_i$, $i=1, \dots, n$ are the sample values of the random design variables and random parameters generated from sampling PDF $f^S_{\mathbf{X},\mathbf{P}}(\mathbf{x},\mathbf{p})$. Note that the sum of the right hand side is only for those replications in which the system fails.

An unbiased estimator of the standard deviation of the system failure probability is,

$$s_{\hat{p}_{\text{sys}}} = \frac{1}{\sqrt{n(n-1)}} \sqrt{\sum_{i=1}^n \left(\frac{f_{\mathbf{X},\mathbf{P}}(\mathbf{x}_i, \mathbf{p}_i)}{f^S_{\mathbf{X},\mathbf{P}}(\mathbf{x}_i, \mathbf{p}_i)} \right)^2 - n \cdot \hat{p}_{\text{sys}}^2} \quad (5a)$$

The $1-\alpha$ level confidence interval of the system failure probability is,

$$\hat{p}_{\text{sys}} \pm t_{n-1, 1-\frac{\alpha}{2}} \cdot s_{\hat{p}_{\text{sys}}} \quad (6a)$$

where $t_{n-1, 1-\frac{\alpha}{2}}$ is the point that has below it probability

$1-\alpha/2$ for Student's t -distribution with $n-1$ degrees of freedom. This interval covers the true system probability of failure with probability $1-\alpha$. Usually, the 95% confidence interval is calculated, that is $\alpha = 0.05$.

Similar equations are presented below for the average value of a design attribute Y , which is a function of random variables $\mathbf{X}$ and parameters $\mathbf{P}$:

$$\hat{E}(Y) = \frac{1}{n} \sum_{i=1}^n Y(\mathbf{x}_i, \mathbf{p}_i) \frac{f_{\mathbf{X},\mathbf{P}}(\mathbf{x}_i, \mathbf{p}_i)}{f^S_{\mathbf{X},\mathbf{P}}(\mathbf{x}_i, \mathbf{p}_i)} \quad (4b)$$

$$s_{\hat{E}(Y)} = \frac{1}{\sqrt{n(n-1)}} \sqrt{\sum_{i=1}^n \left(Y(\mathbf{x}_i, \mathbf{p}_i) \frac{f_{\mathbf{X},\mathbf{P}}(\mathbf{x}_i, \mathbf{p}_i)}{f^S_{\mathbf{X},\mathbf{P}}(\mathbf{x}_i, \mathbf{p}_i)} \right)^2 - n[\hat{E}(Y)]^2} \quad (5b)$$

$$\hat{E}(Y) \pm t_{n-1, 1-\frac{\alpha}{2}} \cdot s_{\hat{E}(Y)} \quad (6b)$$

EFFICIENT PRA

Estimation of Failure Probability and Average Utility

Assume that the random design variables and random parameters are statistically independent, $f_{\mathbf{X},\mathbf{P}}(\mathbf{x},\mathbf{p}) = f_{\mathbf{X}}(\mathbf{x})f_{\mathbf{P}}(\mathbf{p})$. The key idea of PRA is that it is

sufficient to calculate the failure indicator function of a system only for a single sample of values, $\{\mathbf{x}_i, i = 1, \dots, n\}$ of random design variables $\mathbf{X}$ in order to estimate the system failure probability for other PDFs. Indeed, if one generates a sample from $f^S_{\mathbf{X}, \mathbf{P}}(\mathbf{x}, \mathbf{p})$ and calculates and saves those sample values that caused system failure $\{\mathbf{x}_i, i = 1, \dots, n_f\}$, then one can reuse these sample values to estimate the probability of failure p_{sys} for another PDF $f_{\mathbf{X}, \mathbf{P}}(\mathbf{x}, \mathbf{p})$,

$$\hat{p}_{sys} = \frac{1}{n} \sum_{i=1}^{n_f} \frac{f_{\mathbf{X}}(\mathbf{x}_i) f_{\mathbf{P}}(\mathbf{p}_i)}{f^S_{\mathbf{X}, \mathbf{P}}(\mathbf{x}_i, \mathbf{p}_i)} \quad (7)$$

The above equation enables one to estimate the failure probabilities of many designs very efficiently because it only requires calculation of ratios of PDFs; it does not require calculation of the failure indicator function. This reduces the cost of system reliability analysis by several orders of magnitude because calculation of the failure indicator function dominates the cost of simulation. The probability of system failure as a function of the average values of the design variables $E(\mathbf{X})$ is,

$$\hat{p}_{sys}(\mu_{\mathbf{X}}) = \frac{1}{n} \sum_{i=1}^{n_f} \frac{f_{\mathbf{X}/E(\mathbf{X})}(\mathbf{x}_i / \mu_{\mathbf{X}}) f_{\mathbf{P}}(\mathbf{p}_i)}{f^S_{\mathbf{X}, \mathbf{P}}(\mathbf{x}_i, \mathbf{p}_i)} \quad (8)$$

where $f_{\mathbf{X}/E(\mathbf{X})}(\mathbf{x} / \mu_{\mathbf{X}})$ is the conditional PDF of $\mathbf{X}$ given that the average value of the random design variables is $E(\mathbf{X}) = \mu_{\mathbf{X}}$. Equations (5a) and (6a) can be easily modified to calculate the standard deviation and a confidence interval for the failure probability.

A similar approach can be employed to estimate the average value of a function Y of the random design variables, such as the utility of a design,

$$\hat{E}[Y(\mu_{\mathbf{X}})] = \frac{1}{n} \sum_{i=1}^n Y(\mathbf{x}_i, \mathbf{p}_i) \frac{f_{\mathbf{X}/E(\mathbf{X})}(\mathbf{x}_i / \mu_{\mathbf{X}}) f_{\mathbf{P}}(\mathbf{p}_i)}{f^S_{\mathbf{X}, \mathbf{P}}(\mathbf{x}_i, \mathbf{p}_i)} \quad (9)$$

The following is an algorithm for efficient estimation of the system failure probability for many different average values of the random design variables. A similar algorithm can be developed for estimation of the average value of a dependent variable.

- 1) Select a sampling PDF $f^S_{\mathbf{X}, \mathbf{P}}(\mathbf{x}, \mathbf{p})$ for the random design variables. This step is discussed in subsection 2.2.4.
- 2) Generate a sample of values of the design variables and the random parameters.
- 3) Calculate the system failure indicator function, $I(\mathbf{x}_i, \mathbf{p}_i)$, for the sample values in step 2. Select the subset of sample values that caused system failure.

- 4) Estimate the system failure probability from Eq. (8), the standard deviation of the estimate and a confidence interval from Eqs. (5a) and (6a).

Figure 1 shows the flow of information of the algorithm for PRA. Inputs are the sampling PDF, the sample of values that caused failure and the PDF for which the failure probability is to be estimated. Output is the probability of failure and confidence bounds of this probability. PRA is non intrusive – it does not require modifications of the computer codes used for deterministic analysis and reliability analysis.

The estimates of the system failure probability and the average value vary smoothly with the design parameters, provided that the sampling and true PDFs are smooth. The reason is that the same sample of values of the failure indicator function is used to calculate the failure probability (or the average value of a variable) for different values of the design variables (see Eqs. (8) and (9)). Figure 2 shows that the system failure probability of an internal combustion engine obtained from PRA varies smoothly with the cylinder bore. The system failure probability varies irregularly when calculated from Monte Carlo simulation and different samples of random values are used in each simulation (Fig. 2a). If the same sample of random values is used in all simulations then the estimates vary smoothly with bore (Fig. 2b). However, it is not always possible to synchronize a random number generator in Monte Carlo simulation because each simulation may require a different sample size. An example is the simulation of a system that can fail under different progressive collapse scenarios. In this case, PRA has an important advantage over Monte Carlo simulation.

In order to estimate accurately the system failure probability or the average value of a variable, the sample drawn from the sampling PDF should cover adequately the range corresponding to the true PDF. The reanalysis method will perform poorly if the mass of the PDF in the numerator of Eqs. (8) and (9) is concentrated in a region where the sampling PDF is almost zero.

Sensitivity Analysis

It is important to calculate the sensitivity derivatives of the probability of failure or the average utility with respect to the design variables in order to assess the importance of these variables and find the optimum design using gradient-based optimization. It is impractical to calculate these sensitivity derivatives by using a finite difference approach because this requires calculation of the system failure probability of designs obtained by perturbing the design variables from their nominal values by Monte Carlo simulation. Each Monte Carlo simulation can be very expensive and numerical noise in the estimates of the failure probability can

amplify the error in the estimates of the sensitivity derivatives.

In PRA, the sensitivity derivatives of the system failure probability with respect to the design variables can be calculated analytically by differentiating Eq. (8) for the system failure probability with respect to the average values of the design variables. The sensitivity derivative with respect to the j^{th} design variable μ_{X_j} is,

$$\frac{\partial p_{\text{sys}}}{\partial \mu_{X_j}} = \frac{1}{n} \sum_{i=1}^n \frac{\partial f_{\mathbf{X}/E(\mathbf{X})}(\mathbf{x}_i / \boldsymbol{\mu}_{\mathbf{X}})}{\partial \mu_{X_j}} \frac{f_{\mathbf{P}}(\mathbf{p}_i)}{f^S_{\mathbf{X},\mathbf{P}}(\mathbf{x}_i, \mathbf{p}_i)} \quad (10)$$

In most probabilistic analysis problems, analytical expressions of the PDFs of the random design variables are available, and these expressions can be differentiated analytically with respect to the average values. Therefore, the calculation in Eq. (10) can be performed very efficiently.

Similar expressions can be derived for the sensitivity derivatives of the average value of a function, such as the utility of a design, with respect to the design variables,

$$\frac{\partial E(Y)}{\partial \mu_{X_j}} = \frac{1}{n} \sum_{i=1}^n Y(\mathbf{x}_i, \mathbf{p}_i) \frac{\partial f_{\mathbf{X}/E(\mathbf{X}_j)}(\mathbf{x}_i / \boldsymbol{\mu}_{\mathbf{X}})}{\partial \mu_{X_j}} \frac{f_{\mathbf{P}}(\mathbf{p}_i)}{f^S_{\mathbf{X},\mathbf{P}}(\mathbf{x}_i, \mathbf{p}_i)} \quad (11)$$

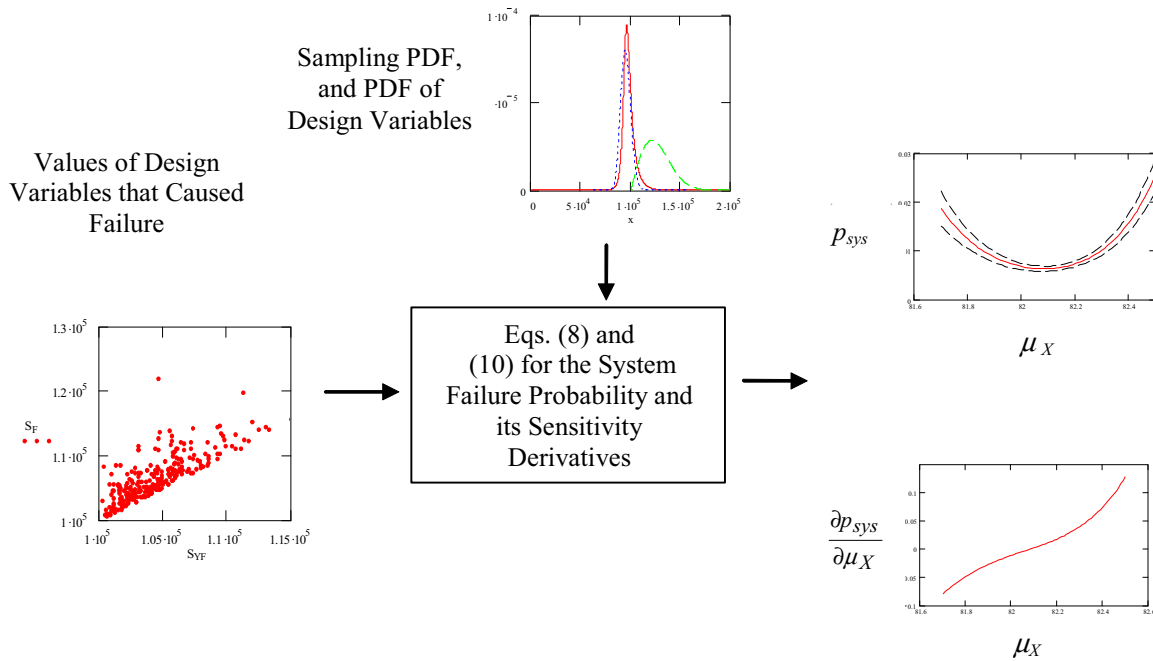


Figure 1. Flow of Information in Efficient PRA