Lecture 9: Monte Carlo Simulation

Lecturer: Prof. F. Jalayer
Notes prepared by: Prof. F. Jalayer and Dr. H. Ebrahimian

1. General

The *Limit States* describe specific stages of structural behavior and its consequences; for instance, states associated with consequences in terms of costs, loss of lives, impact on the environment, or serviceability. Reaching a limit state denotes that an unacceptable behavior for the structure is going to be met. This condition is generically called as "*Failure*". The limit state exceedance is verified by comparing a measurable quantity, associated with the structural response, denoted as the *Demand* "*D*"(also referred to as the engineering demand parameter EDP, e.g., maximum inter-story drift ratio, maximum chord rotation, maximum component shear, etc.), with the corresponding limit value, denoted as the *Capacity* "*C*"(e.g., yielding rotation, ultimate rotation, shear strength, etc.). As a result, the following logical statements can be defined:

$F \equiv$ having Failure $\equiv D > C$

Thus, the probability term $P(F|\mathbf{I})$ can be defined as:

$$P(F|\mathbf{I}) \triangleq P_F = P(D > C|\mathbf{I}) \tag{1}$$

2. Sources of Uncertainty

The vector of N uncertain parameters can be defined as:

$$\mathbf{\Theta} = [\Theta_1, \Theta_2, \cdots, \Theta_N] \tag{2}$$

Therefore, the reliability problem can be defined as:

$$P_F = P\left(F|\mathbf{I}\right) = P\left[D\left(\mathbf{\Theta}\right) > C\left(\mathbf{\Theta}\right)|\mathbf{I}\right] \tag{3}$$

3. General Formulation of Reliability Problem

For a realization of the vector of uncertain parameters $\mathbf{\Theta}$, denoted as $\boldsymbol{\theta}$, consider the following indicator:

$$\mathrm{I}_F\left(\boldsymbol{\theta}\right) = \begin{cases} 1 & D\left(\boldsymbol{\theta}\right) > C\left(\boldsymbol{\theta}\right) \\ 0 & D\left(\boldsymbol{\theta}\right) \leq C\left(\boldsymbol{\theta}\right) \end{cases} \tag{4}$$

Thus, Eq. (3) can be written as follows (considering the total probability theorem):

$$\begin{aligned} P_F = P\left(F|\mathbf{I}\right) = P\left[D\left(\mathbf{\Theta}\right) > C\left(\mathbf{\Theta}\right)|\mathbf{I}\right] &= \int_{\Omega_{\mathbf{\Theta}}} P\left[D\left(\boldsymbol{\theta}\right) > C\left(\boldsymbol{\theta}\right)|\boldsymbol{\theta},\mathbf{I}\right] \cdot f_{\mathbf{\Theta}}\left(\boldsymbol{\theta}|\mathbf{I}\right) \mathrm{d}\boldsymbol{\theta} \\ &= \int_{\Omega_{\mathbf{\Theta}}} \mathrm{I}_F\left(\boldsymbol{\theta}\right) \cdot f_{\mathbf{\Theta}}\left(\boldsymbol{\theta}|\mathbf{I}\right) \mathrm{d}\boldsymbol{\theta} \\ &\triangleq \mathbb{E}_{\mathbf{\Theta}}\left[\mathrm{I}_F\left(\boldsymbol{\theta}\right)|\mathbf{I}\right] \end{aligned} \tag{5}$$

4. The expected value and variance of the average of *i.i.d.* uncertain parameters

Suppose we are looking at n independent and identically distributed random variables, X_i, i=1:n. Since they are *i.i.d.*, each random variable X_i has the same mean and variance, as follows:

$$\begin{aligned} &\mathbb{E}\left(X_i|\mathbf{I}\right) = \mu \\ &\mathbb{VAR}\left(X_i|\mathbf{I}\right) = \sigma^2 \ , \ \ i = 1,\cdots,n \end{aligned} \tag{6}$$

Suppose that we want to look at the average value of our n random variables:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \tag{7}$$

Now, we want to find the expected value and variance of the average:

$$\begin{aligned}
&\mathbb{E}\left(\bar{X}\,|\mathbf{I}\right)=\mathbb{E}\left(\frac{\sum_{i=1}^{n}X_i}{n}\Bigg|\mathbf{I}\right)=\frac{\sum_{i=1}^{n}\mathbb{E}\left(X_i\,|\mathbf{I}\right)}{n}=\frac{n\mu}{n}=\mu\\
&\mathbb{VAR}\left(\bar{X}\,|\mathbf{I}\right)=\mathbb{VAR}\left(\frac{\sum_{i=1}^{n}X_i}{n}\Bigg|\mathbf{I}\right)=\frac{\sum_{i=1}^{n}\mathbb{VAR}\left(X_i\,|\mathbf{I}\right)}{n^2}=\frac{n\sigma^2}{n}=\frac{\sigma^2}{n}
\end{aligned} \tag{8}$$

5. The Monte Carlo Simulation

Recalling from Section 3,

$$P_F=P\left(F\,|\mathbf{I}\right)=\mathbb{E}_{\boldsymbol{\Theta}}\left[\mathrm{I}_F\left(\boldsymbol{\theta}\right)|\mathbf{I}\right] \tag{9}$$

Now, consider that we have N_{sim} realization (simulation) of $\boldsymbol{\Theta}$, which are denoted as $\boldsymbol{\theta}_i$, $i=1,\cdots,N_{\mathrm{sim}}$. Thus, as estimator for the probability of failure, $\tilde{P}\left(F\,|\mathbf{I}\right)$, can be defined as:

$$P\left(F\,|\mathbf{I}\right)=\mathbb{E}_{\boldsymbol{\Theta}}\left[\mathrm{I}_F\left(\boldsymbol{\theta}\right)|\mathbf{I}\right]\approx\tilde{P}\left(F\,|\mathbf{I}\right)=\frac{\sum_{i=1}^{N_{\mathrm{sim}}}\mathrm{I}_F\left(\boldsymbol{\theta}_i\right)}{N_{\mathrm{sim}}}\triangleq\frac{N_{\mathrm{failure}}}{N_{\mathrm{sim}}} \tag{10}$$

According to Eq. (4), $\mathrm{I}_F\left(\boldsymbol{\theta}_i\right)$ are *i.i.d.* Bernoulli random variables, as follows:

$$\mathrm{I}_F\left(\boldsymbol{\theta}\right)=\begin{cases}1 & P\left[D\left(\boldsymbol{\theta}\right)>C\left(\boldsymbol{\theta}\right)\right]=P\left(F\,|\mathbf{I}\right)=P_F\\ 0 & P\left[D\left(\boldsymbol{\theta}\right)\le C\left(\boldsymbol{\theta}\right)\right]=1-P\left(F\,|\mathbf{I}\right)=1-P_F\end{cases} \tag{11}$$

Therefore,

$$\begin{aligned}
&\mathbb{E}\left[\mathrm{I}_F\left(\boldsymbol{\theta}_i\right)|\mathbf{I}\right]=P_F\\
&\mathbb{VAR}\left[\mathrm{I}_F\left(\boldsymbol{\theta}_i\right)|\mathbf{I}\right]=P_F\left(1-P_F\right),\quad i=1,\cdots,N_{\mathrm{sim}}
\end{aligned} \tag{12}$$

Then, with respect to Eq. (8):

$$\mathbb{E}\left[\tilde{P}\left(F|\mathbf{I}\right)\right] = P_F$$
$$\mathbb{VAR}\left[\tilde{P}\left(F|\mathbf{I}\right)\right] = \frac{P_F\left(1-P_F\right)}{N_{\text{sim}}} \tag{13}$$

The COV of estimator can be calculated as:

$$\mathbb{COV}\left[\tilde{P}\left(F|\mathbf{I}\right)\right] = \frac{\sqrt{\dfrac{P_F\left(1-P_F\right)}{N_{\text{sim}}}}}{P_F} = \sqrt{\frac{\left(1-P_F\right)/P_F}{N_{\text{sim}}}} \tag{14}$$

In seismic reliability of structures, we usually have small probability of failure in order of $P_F \le 10^{-3}$. Therefore, Eq. (14) can be approximates as:

$$\mathbb{COV}\left[\tilde{P}\left(F|\mathbf{I}\right)\right] \cong \sqrt{\frac{1}{P_F N_{\text{sim}}}} \tag{15}$$

Example: Consider the COV to be equal to 0.30, and the probability of failure equal to 10^{-3}. How many simulations are required for having a proper estimate by Monte Carlo simulation?

$$0.30 \cong \sqrt{\frac{1}{10^{-3} \times N_{\text{sim}}}} \Rightarrow N_{\text{sim}} \cong 10000$$

6. Method of simulation

The standard Monte Carlo (MC) simulation uses the random sampling approach in order to generate a large amount of realizations of uncertain parameters. The probability of failure can then be estimated by directly implementing Eq. (10) regardless of the complexity of the problem. To simulate from the joint density function $f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}|\mathbf{I})$ (see Eq. 5) in standard MC simulation, we consider two different conditions:

6.1. $\boldsymbol{\theta}_i$'s are independent

Here, we have:

$$f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}|\mathbf{I}) = f_{\boldsymbol{\Theta}}(\theta_1, \theta_2, \cdots, \theta_N|\mathbf{I}) = f_{\Theta_1}(\theta_1|\mathbf{I}) f_{\Theta_2}(\theta_2|\mathbf{I}) \cdots f_{\Theta_N}(\theta_N|\mathbf{I}) = \prod_{i=1}^{N} f_{\Theta_i}(\theta_i|\mathbf{I}) \quad (16)$$

For each $f(\theta_i|\mathbf{I})$, we calculate the associated CDF, $F(\theta_i|\mathbf{I})$, as shown schematically in Fig. 1. Subsequently, a random number, denoted as Z_j between 0 and 1 is generated. Then, the random numbers Z_j is mapped to $\theta_{i,j}$ by:

$$\theta_{i,j} = F_{\Theta_i}^{-1}(Z_j) \quad (17)$$

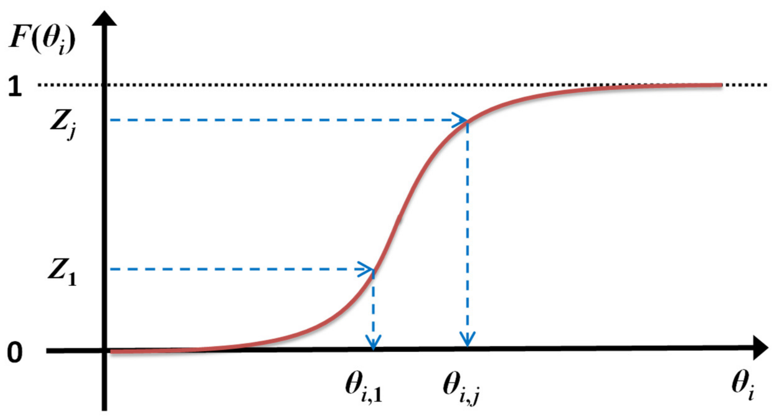

Figure 1: Sampling from a CDF

This procedure is repeated for $j = 1, \cdots, N_{\text{sim}}$.

6.2. $\boldsymbol{\theta}_i$'s are not independent

In this case, we have:

$$\begin{aligned} f_{\boldsymbol{\Theta}}\left(\boldsymbol{\theta} \middle| \mathbf{I}\right) &= f_{\boldsymbol{\Theta}}\left(\theta_1, \theta_2, \cdots, \theta_N \middle| \mathbf{I}\right) \\ &= f_{\Theta_1}\left(\theta_1 \middle| \mathbf{I}\right) f\left(\theta_2, \cdots, \theta_N \middle| \theta_1, \mathbf{I}\right) \\ &= f_{\Theta_1}\left(\theta_1 \middle| \mathbf{I}\right) f\left(\theta_2 \middle| \theta_1, \mathbf{I}\right) f\left(\theta_3, \cdots, \theta_N \middle| \theta_1, \theta_2, \mathbf{I}\right) \end{aligned} \tag{16}$$

Therefore, in this condition, we simulate first from $f_{\Theta_1}\left(\theta_1 \middle| \mathbf{I}\right)$; then, knowing θ_1, we simulate from $f\left(\theta_2 \middle| \theta_1, \mathbf{I}\right)$; and the procedure will continue until we simulate from the distribution $f\left(\theta_N \middle| \theta_1, \theta_2, \cdots, \theta_{N-1}, \mathbf{I}\right)$.