Chapter 2  Theoretical methods of structural reliability

2.1 Introduction
This chapter will introduce the basic theory and methods of reliability analysis in detail, and will give suggestions on selecting methods for calculating structural reliability.
Structural reliability methods are classified according to Level, Moment and Order. ‘Level’ classifies reliability methods in the following groups:

Level I methods: deterministic reliability methods that use only one ‘characteristic’ value to describe each uncertain variable. Level I methods correspond to standard deterministic design methods.

Level II methods: reliability methods that use two values to describe each uncertain variable: such as its mean and variance, supplemented with a measure of correlation between the variables, usually the covariance. Reliability index methods, such as the first-order second-moment method, are examples of Level II methods.

Level III methods: reliability methods that use the joint probability distribution of all the uncertain variables to describe each uncertain variable. Level III methods include numerical integration, approximate analytical methods, such as first- and second-order reliability methods, and simulation methods.

Level IV methods: reliability methods that compare a structural prospect with a reference prospect according to the principles of engineering economic analysis under uncertainty.

‘Moment and Order’ classifies reliability methods into some approximate methods:

First-order and Second-moment method,
Second-order and Second-moment method, etc.

‘Exactness of calculation result’. Classifies the methods of reliability calculation into the following three types:

Approximate methods (such as the Mean-Value First-Order Second-Moment method etc).
Simulation methods (such as the Monte Carlo method).
The directly integral method.

2.2 Overview of reliability calculation
Structural reliability analysis, in general, can be simplified into calculating the following integral (calculating the failure probability of the structure).

\[ P_f = \int_{G(X) = 0} f_X(X) dX \]
where \( X \) represents the random variables associated with the assessed structure.

\( G(X) \) is the limit state function such that:
- \( G(X) > 0 \) for \( X \) in safe set of structure.
- \( G(X) = 0 \) for \( X \) on limit state surface.
- \( G(X) < 0 \) for \( X \) in failure set.

\( f(X) \) is the probability function of \( X \).

\( I[G(X)] \) is the indicator function.

The complement, \( 1 - P_f \), is the reliability of the structure, i.e.,

\[ P_r = 1 - P_f \]

The corresponding reliability index \( \beta \) is defined as

\[ \beta = -\Phi^{-1}(P_f) \text{ i.e., } P_f = \Phi(-\beta) \]  

(2.2.2)

where \( \Phi \) is the standard normal cumulative distribution function.

When the structure or component in question contains more than one limit state, the corresponding system reliability will have to be considered. For a marine structure, a system reliability analysis will usually be required, but under some conditions it can be dealt with as a simple reliability analysis.

### 2.3 Analytical methods (Approximate Methods)

The advantages of the analytical methods are mainly that they are simple, have apparent physical interpretations and do not usually require excessively large computation time. Analytical methods usually contain the First-order and Second-order reliability methods, i.e., FORM and SORM, respectively. The drawback of these methods is that they do not give exact results for the probabilities in question for most practical structures.

#### 2.3.1 Mean-Value First-Order Second-Moment (MVFOSM) method

In general, the limit state function can be presented as follows

\[ M = g(x_1, x_2, \ldots, x_n) \]  

(2.3.1)

where \( x_i \) are the load and strength parameters considered as variables, and the limit state function \( g(.) \) is a function that relates these variables for the limit state of interest (serviceability or ultimate state). The limit state is reached when:

\[ M = g(x_1, x_2, \ldots, x_n) \leq 0 \]  

(2.3.2)

The probability of failure can be calculated by integrating Equation (2.2.1). This reliability analysis can be very difficult to apply in practice. The two main reasons for this are:
the lack of information to determine the joint probability density function of the design variables and the difficulty associated with evaluation of the resulting multiple integrals because of the complexity of the limit state function. This is why analytical methods (i.e., level II reliability methods, approximate methods) were developed. In level II, the safety index concept, which was first introduced by Cornell in 1969, was further developed by several researchers. In the next few sections, the Level II reliability will be presented starting with the simple safety index concept, followed by several improvements of the concept.

The limit state function can be expanded using Taylor’s series, and if only the first order terms are retained, we get

\[ g(x_1, x_2, \cdots, x_n) \approx g(x_1^*, x_2^*, \cdots, x_n^*) + \sum_{i=1}^{n} (x_i - x_i^*) \left( \frac{\partial g}{\partial x_i} \right)_{x_i^*}, \]  

\[ (2.3.5) \]

where \( x_i^* \) is the linearization point, and the partial derivatives are evaluated at that point. In the MVFOSM method the linearization point is set at the mean value \((E(x_1), E(x_2), \cdots, E(x_n))\).

The mean and variance of \( M \) are then approximated by

\[ \mu_m = g(\mu_1, \mu_2, \cdots, \mu_n) \]  

\[ \sigma^2_m = \sum_{i} \sum_{j} (x_i - x_i^*) \left( \frac{\partial g}{\partial x_i} \right)_{x_i^*} \left( \frac{\partial g}{\partial x_j} \right)_{x_j^*} \rho_{x_i x_j} \sigma_{x_i} \sigma_{x_j}, \]  

\[ (2.3.6) \]

\[ (2.3.7) \]

where \( \rho_{x_i x_j} \) is the correlation coefficient and the \( \left( \frac{\partial g}{\partial x_i} \right)_{x_i^*}, \left( \frac{\partial g}{\partial x_j} \right)_{x_j^*} \) denote evaluation of the partial derivatives at the mean point. \( \mu_i = E(x_i) = \bar{x}_i \)

The accuracy of equations (2.3.4) and (2.3.5) depends on the effect of neglecting the higher-order terms in equation (2.3.3).

If the variables \( x_i \) are statistically independent, then (2.3.4) remains unchanged but (2.3.5) becomes

\[ \sigma^2_m = \sum_{i} \left( \frac{\partial g}{\partial x_i} \right)^2 \sigma_{x_i} \]  

\[ (2.3.6) \]

The standardized margin \( G_M \), which has a zero mean and a unit standard deviation, can be written as

\[ G_M = \frac{M - \mu_m}{\sigma_m} \]

\[ (2.3.7) \]

Failure occurs when \( M \leq 0 \) so that the probability of failure can be written as

\[ P_f = P[M \leq 0] \]

\[ (2.3.8) \]
where $\beta = \frac{\mu_m}{\sigma_m}$ is safety index, which is the inverse of the coefficient of variation of the safety margin.

If the distribution function $F_{G_u}(\cdot)$ is known, then the exact probability of failure associated with the safety index can be determined. But even for unknown or unspecified distribution function $F_{G_u}(\cdot)$, there will be a corresponding though unspecified probability of failure for each value of $\beta$. Thus $\beta$ can be taken as safety measure as is the case in the MVFOSM method.

As an example, if the margin $M$ is represented by the variables $S$ and $Z$ only, that is,

$$Z = S - Z,$$

then applying equation (2.3.4) and (2.3.5) for determining the mean and variance, the following results are obtained

$$\mu_m = \mu_S - \mu_Z,$$
$$\sigma_m^2 = \sigma_S^2 - \sigma_Z^2.$$

Equation (2.3.8) follows accordingly. The following is a geometric interpretation of the safety margin $M=S-Z$.

First we notice that $M>0$ represents a safe state or region, $M<0$ represents failure state and $M=0$ represents a limit state or failure surface (or line in the case of two variables). The standardized variables of $S$ and $Z$ can be written as

$$S' = \frac{S - \mu_S}{\sigma_S}; Z' = \frac{Z - \mu_Z}{\sigma_Z}.$$

Therefore, the limit state function $M=0$ can be written in the space of standard variables as:

$$M = \sigma_s S' - \sigma_z Z' + \mu_s - \mu_z = 0$$

Which is a straight line shown in (Figure 2.3.1).

![Figure 2.3.1 Limit state function in the standard space](image-url)

The region on one side of the straight line which contains the origin ‘0’ represents the safe state ($M>0$) and the other region represents the failure state ($M<0$). Thus the distance from the origin to the line $M=0$ can be used as a measure of reliability. In fact, from geometry, the minimum distance ‘D’ shown on figure 2.3.1 is given by
Notice that $D$ is equal to the safety index $\beta$ for the case of the normal variables and linear limit state function discussed earlier, i.e., for this case

$$\beta = D = \frac{\mu_m - \mu_Z}{\sigma_m}$$

and the probability of failure is thus

$$P_f = \Phi(-D)$$

### 2.3.2 Advanced First-Order Second-Moment (MVFORM) method

The MVFOSM method described previously has the following basic shortcomings:

First, if $g(\cdot)$ is non-linear and the linearization takes place at the mean values of $x_i$, errors may be introduced at increasing distance from the linearization points by neglecting higher-order terms.

Second, the method fails to be invariant to different equivalent formations of the same problem. In effect this means that the safety index $\beta$ depends on how the limit state equation is formulated. For example if the $M$ is set a nonlinear function of $S$ and $Z$ such as

$$M = S^2 - Z^2$$

then

$$P_f = P[M \leq 0] = F_M(0) = F_{\bar{g}_m}\left(\frac{-\mu_m}{\sigma_m}\right) = F_{\bar{g}_m}(-\beta)$$

however, when $\mu_m, \sigma_m$ are computed from (2.3.4) and (2.3.6) and substituted in

$$\beta = \frac{\mu_m}{\sigma_m}$$

the following $\beta$ is obtained

$$\beta = \frac{\mu_S^2 - \mu_Z^2}{2\sqrt{\mu_S^2 \sigma_S^2 + \mu_Z^2 \sigma_Z^2}}$$

which is different from the $\beta$ obtained when $M$ is taken as $M=S-Z$, even though the criterion of failure is still given by equation (2.3.8).

Third, in the MVFORM method the safety index $\beta$ can be related to a probability of failure in cases when the variables $x_i$ are normally distributed (and when the function $g(\cdot)$ is linear in $x_i$). It is known that wave bending moments in ships follow a Weibull or exponential distribution. Thus, one of the improvements in an advanced method over the MVFOSM method would be to include such distribution information.

#### 2.3.2.1 The Hasofer-Lind Index

The first two shortcomings discussed previously are avoided by using a procedure usually attributed to Hasofer and Lind (1974). Instead of expanding Taylor’s series about the mean value point, which cause the invariance problem, the linearization point
is taken at some point on the failure surface. On the failure surface, the limit state function \( g(.) \) and its derivatives are independent of how the problem is performed.

In the Hasofer-Lind procedure, the load and resistance variables, \( x_i \), are transformed to standardized variables with zero mean and unit variance by

\[
y_i = \frac{x_i - \mu_i}{\sigma_{x_i}}
\]  

(2.3.12)

The Hasofer-Lind reliability index is defined as the shortest distance from the origin to the failure surface in the reduced (standardized) space. This point is found by solving the following set of equations

\[
G(y_1^*, y_2^*, \ldots, y_n^*) = 0
\]

(2.3.13)

\[
y_i^* = -\alpha_i^* \beta
\]

(2.3.14)

\[
\alpha_i^* = \frac{\left( \frac{\partial G}{\partial y_i} \right)_{y_i^*}}{\sqrt{\sum_i \left( \frac{\partial G}{\partial y_i} \right)_{y_i^*}^2}}
\]

(2.3.15)

\( G(.) \) is the failure surface in the reduced space, and \( \mu_i^* \) are coordinates of point closest to the origin in the reduced space (the checking point, i.e., design point). All partial derivatives are evaluated at the checking point. In effect, this procedure is equivalent to linearizing the limit state function in the reduced variables space at the checking point and computing \( \beta \) associated with that point.

In the original space, the checking point is obtained from

\[
x_i^* = \bar{x}_i + \sigma_{x_i} \mu_i, x_i^* = \bar{x}_i - \sigma_{x_i} \alpha_i \beta
\]

(2.3.16)

In general, for a linear limit state function, the Hasofer-Lind method will yield the same result for \( \beta \) as the MVFOSM method. For nonlinear limit state functions, this method yields a safety index \( \beta \) which is invariant to the formulation of the performance function.

### 2.3.2.2 Inclusion of distribution information

The third and final refinement of the advanced method over the MVFOSM method is that the information about the variable distributions (if known) can be incorporated in the computation of the safety index and the corresponding probability of failure. In the MVFOSM method the index \( \beta \) can be related to the probability of failure in cases when the variables \( x_i \) are normally distributed and when \( g(.) \) is linear in \( x_i \). This relation is given by

\[
P_f = \Phi\left( -\beta \right)
\]

(2.3.17)

where \( \Phi(.) \) is the standard normal distribution function. In order to be able to use equation (2.3.16) approximately, in the case of non-normal variables, a transformation
of these variables into equivalent normal variables is necessary, prior to each iteration, in the solution of equations (2.3.13) to (2.3.15).

The tail of the distribution is usually the location where most of the contribution to the probability of failure comes from. It is, therefore, better to fit the normal distribution to the tail of the non-normal distribution at the linearization \( x^*_i \), which is where failure is most likely to occur (that is, minimum \( \beta \)). This is the basic idea of the Rackwitz/Fiessler method as given in 1978.

By requiring that the cumulative distributions and the probability density functions of both the actual distribution and the normal distribution be equal at the linearization point, one can determine the mean \( \mu^*_i \) and standard deviation \( \sigma^*_i \) of the equivalent normal variable, that is

\[
F^*_i(x^*) = \Phi\left(\frac{x^*_i - \mu^*_i}{\sigma^*_i}\right)
\]

and

\[
f^*_i(x^*) = \frac{1}{\sqrt{2\pi} \sigma^*_i} \exp\left[-0.5 \left(\frac{x^*_i - \mu^*_i}{\sigma^*_i}\right)^2\right] = \frac{1}{\sigma^*_i} \varphi\left(\Phi^{-1}[F^*_i(x^*)]\right)
\]

where \( \varphi(\cdot) \) is the standard normal probability density.

Since we are concerned with finding \( \mu^*_i \) and \( \sigma^*_i \), the parameters of the equivalent normal distribution once fitted at the linearization point, we can solve for them as follows:

\[
\sigma^*_i = \frac{\varphi\left(\Phi^{-1}[F^*_i(x^*)]\right)}{f^*_i(x^*)}
\]

\[
\mu^*_i = x^*_i - \Phi^{-1}\left[F^*_i(x^*)\right]\sigma^*_i
\]

Since the linearization point \( x^*_i \) changes with each iteration, \( \mu^*_i \) and \( \sigma^*_i \) must also be calculated for each iteration. These values are then used in equations (2.3.13) through (2.3.15) as before. Note that if the iteration is performed in a reduced space, then distribution transformation into reduced space has to be performed in each step.

So far, the random variables \( x_i \) have been assumed to be independent. In the structural reliability analysis of marine (ship) structures, basic variables are often considered as independent random variables. In cases of correlated variables, see SSC-351.

The following are the main steps to calculate structure reliabilities by this analytical method:

a) Transform the physical basic variables into the space of standard normal variables.

b) Calculate design points.

c) Calculate failure reliability

The following is an iterative solution scheme for calculating the reliability index and design point (Robert E. Melchers, 2001):
a) Standardize basic random variables X to the independent standardized normal variables Y, using (2.3.12) and the other methods if necessary.
b) Transform $G(X) = 0$ to $g(Y) = 0$
c) Select initial design (checking) point $(x^{(i)}, y^{(i)})$
d) Compute $\beta^{(i)} = \left[ y^{(i)^T} \cdot y^{(i)} \right]^{0.5}$, let $m=1$
e) Compute direction cosines $\alpha^{(m)}$ using following formulae
$$\alpha_i = \frac{c_i}{l}, \quad l = \left( \sum_i c_i^2 \right)^{0.5}, \quad c_j = \lambda \frac{\partial g}{\partial y_i}$$ (2.3.22)
f) Compute $g(Y^{(m)}) = 0$
g) Compute $y^{(m+1)}$ using
$$y^{(m+1)} = -\alpha^{(m)} \left[ \beta^{(m)} + \frac{g(y^{(m)})}{l} \right]$$ (2.3.23)
h) Compute $\beta^{(m+1)} = \left[ y^{(m+1)^T} \cdot y^{(m+1)} \right]^{0.5}$
i) Check whether $y^{(m+1)}$ and / or $\beta^{(m+1)}$ and have stabilized; if not go to e) and increase m by unity.

2.3.3 Second-Order methods

From the above discussions it should be clear that approximating the limit state surface by a linear surface (through a Taylor series expansion) may not be satisfactory if the limit state surface has significant curvature. Even for linear limit state functions in the original space, a non-linear limit state may result when the reliability problem is transferred from the original space to the standard normal space.

In the context of second moment methods and their developments, methods to deal with the non-linearity of the limit state function have been termed 'second order' methods. The most common approach has been to attempt to fit a parabolic, quadratic or higher order surface to the actual surface, centred on the design point. This requires some decision about the extent to which the approximation is valid away from the design point $y^*$. When the limit state surface in the standard normal space is approximated by a second order quadratic surface in the design point, the reliability calculation method is called a “second order reliability method”.

There are no easy ways to estimate the probability content enclosed by a quadratic surface, but the following are two ways to estimate it:
1. This method relies on sampling the space between the linear approximation and the quadratic approximation to estimate the probability content between these two approximations to the limit state surface. The majority of the probability content is estimated using first order theory (Hohenbichler and Rackwitz, 1988).
2. In the alternative method, the FORM result for a linear limit state can be taken as the starting point for a simulation about the design point to estimate the error in
failure probability between that given by the linear limit state approximation and the actual limit state (e.g. Mitteau, 1996).

Another way to evaluate the probability content for the second order approach uses asymptotic concepts. In the space of independent standard normal random variables (the y space), and for a limit state function which is not too highly non-linear, the failure probability can be estimated from a determination of the limit state surface curvatures \( K_i \) at the design point \( y^* \) and then applying the asymptotic expression (Breitung, 1984):

\[
P_f = \Phi(-\beta) \left[ \sum_{j=1}^{k} \prod_{i=1}^{n-1} (1 - \beta \cdot K_i) \right]^{-0.5}
\]

where \( K_i = -\left( \frac{\partial^2 y_n}{\partial y_j^2} \right) \) is the \( i \)th principal curvature of the limit state surface \( g(y^*) = 0 \) at \( y^* \). The error in the asymptotic approximation (2.3.24) is not defined (Robert E. Melchers, 2001).

The second order reliability methods (SORM) can be demanding for large problems with a large number of basic variables or where the limit state function is complex. In practical engineering applications, first order reliability methods are used.

### 2.4 Direct integration method

The direct integration method is based on the direct integration of the joint probability density function of the basic random variables involved in a design. The probability of failure or the probability of reaching a specified limit state is determined from (2.2.1):

\[
P_f = \int_{G(X) \leq 0} f(X) dX
\]

The domain of integration is over the unsafe region of a limit state associated with a structure. The above general equation can be simplified for some specified cases. The equation (2.2.1) is only a simple multivariable integration and the theoretical method is simple. In practical engineering, there are no easy ways to estimate the probability content when the limit state function \( G(X) \) is complex and the number of basic random variables is large. The multi-normal integral method by importance sampling is one of them (Mori, Y., 2003). The following is a traditional method.

#### 2.4.1 Standardized integral region

The equation (2.2.1) can be transformed into the following equation (2.4.1) by using the indicator function (2.4.2).

\[
P_f = \int \cdots \int f_X(X) dX = \int \cdots \int I[G(X)] f_X(X) dX
\]

\[
= \int \cdots \int F(X) dX \approx \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} F(X) dX
\]

where \( F(X) = I[G(X)] \cdot f_X(X) \),
I[G(X)] is the indicator function.

\[
I[G(X)] = \begin{cases} 
1 & \text{if } G(X) \leq 0 \\
0 & \text{if } G(X) > 0 
\end{cases}
\]  
(2.4.2)

For each variable, its \(a_i\) & \(b_i\) can be determined by its distribution type and the integral precision \(\alpha\) by equation (2.4.3).

\[
\int_{a_i}^{b_i} f(x)dx \geq 1 - \alpha 
\]  
(2.4.3)

2.4.2 Joint probability density function

In practical engineering projects, its joint probability function is unknown, but its margin probability density function and correlation between basic variables can often be obtained. For non-independent variables, some authorities supply some suggestions (Liu, 1986). The calculation modes are, for example, Morgenstern Mode and NATAF mode.

For marine structural reliability, basic variables are often considered as independent variables. For independent variables, their joint probability density function is the product of marginal density functions of all variables:

\[
f_X(x) = f_{x_1}(x_1)f_{x_2}(x_2)\cdots f_{x_n}(x_n) 
\]  
(2.4.4)

2.4.3 Multivariate integration

The method for calculation of structural reliability is to calculate the following multivariate integration:

\[
I = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} F(x_1,\ldots,x_n)dx_1\cdots dx_n 
\]  
(2.4.5)

From the definition of integral, the above integral can be transformed into the following summing formula:

\[
I = \sum_{i=1}^{k_1} \cdots \sum_{i=1}^{k_n} F(a_{i_1},\ldots,a_{i_n})\Delta a_{i_1}\cdots \Delta a_{i_n} 
\]  
(2.4.6)

where \(k_1,k_2,\ldots,k_n\) are the numbers of the small regions generated by dividing the regions \([a_1,b_1]\ldots[a_n,b_n]\);

\(a_{i_j}\) is a value in \([a_j,b_j]\);

\(\Delta a_{i_j}\) is \(b_j - a_j\), i.e. the length of \([a_j,b_j]\).

For this method, the key drawback is the excessive amount of computation time when the number of basic variables is large or the limit state function is complex (Zhang Wei, 1997).

For marine structural reliability, this method is a reasonable method, because for each of failure mode (refer to Chapter 4) there are not more than 7 basic variables.

2.4.4 Advanced method of integration
There are many integral definition methods developed in mathematics, such as the BF rule method (Genz and Malik 1980). Importance sampling method is also a good method for multi-normal integration (Mori, Y. 2003).

Although the direct integration method is usually limited in application to actual structures because of the complexity of the analysis, this method can be still applied to assess ship primary strength when the ship is considered as a beam.

2.5 Simulation methods
The analytical methods give approximate results, and should be applied for failure probability less than 0.05. For larger failure probabilities, direct integral method maybe better. However, because the direct integral method is not suitable where there is a large number of variables, practitioners have tried to develop other effective methods. Simulation methods are one of these effective methods.

This part will introduce simulation methods including basic Monte Carlo Simulation method, Importance Sampling simulation method, Adaptive importance simulation method, and Conditional expectation simulation method and so on.

In general, simulation is a technique for conducting experiments in a laboratory or on a digital computer in order to model the behaviour of a system. Usually simulation models result in ‘simulated’ data that must be treated statistically in order to predict the future behaviour of the system. In this broad sense, simulation has been used as a predictive tool for economic systems, business environment, war games and management games.

Monte Carlo simulation is usually used for problems involving random variables of known or assumed probability distributions. Using statistical sampling techniques, a set of values of the random variables is generated in accordance with the corresponding probability distributions. These values are treated as being similar to a sample of experimental observations and are used to obtain a ‘sample’ solution. By repeating the process and generating several sets of sample data, many sample solutions can be determined. Statistical analysis of the sample solutions is then performed.

The Monte Carlo method thus consists of the following basic steps:

a) Simulation of the random variables and generation of several sample data sets using statistical sampling techniques

b) Solutions using the sampled data

c) Statistical analysis of the results

Since the results from the Monte Carlo technique depend on the number of samples used, they are not exact and are subject to sampling errors. Generally the accuracy increase as the sample size increases.

Sampling from a particular probability distribution involves the use of random numbers. Random numbers are essentially random variables uniformly distributed over the unit interval [0,1]. Many codes are available for computers for generating sequences of
‘pseudo’ random digits where each digit occurs with approximately equal probability. The generation of such random numbers plays a central role in the generation of a set of values (or realizations) of a random variable that has a probability distribution other than uniform probability law.

The Monte Carlo method is considered now as one of the most powerful techniques for analysing complex problems. Since its chief constraint is computer capability, it is expected to become even more commonly used in the future, as computer capacities increase and become less expensive to use.

The following steps are necessary to apply Monte Carlo techniques to structural reliability problems (Robert E. Melchers, 2001):

a) to develop systematic methods for numerical ‘sampling’ of the basic variables $X$;
b) to select an appropriate economical and reliable simulation technique or ‘sampling strategy’;
c) to consider the effect of the complexity of calculating $g(X)$ and the number of basic variables on the simulation technique used;
d) for a given simulation technique to be able to determine the amount of ‘sampling’ required to obtain a reasonable estimate of $P_f$;
e) to deal with dependence between all or some of the basic variables if necessary.

2.5.1 Basic principle and method

As was discussed above, the reliability of a structure can be characterized by a limit state function $g(X) = g(x_1, x_2, \ldots, x_d)$, where $X$ is random variables representing the basic design variables. The inequality $g(X) \leq 0$ corresponds to failure, while $g(X) > 0$ represents the safe region. In the Monte Carlo approach a random sample, of values $x_i$ for the basic design variables, is generated numerically according to their probability distributions using a random number generator. The generated sample values are then substituted in the limit state function whose value is then computed to see if it is negative or positive, i.e., failure or no failure. By repeating this process many times, it is possible to simulate the probability distribution of $g(X)$. This will require a very large number of samples. The probability of failure can then be estimated from either of the following methods.

The probability of failure is given by

$$P_f = P[g(X) \leq 0] = \lim_{N \to \infty} \frac{n}{N}$$

where $N$ is the total number of trials or simulations and $n$ is the number of trials in which $g(X) \leq 0$.

The ratio $n/N$ is usually very small and the estimated probability of failure is subjected to considerable uncertainty. In particular the variance of $n/N$ depends on the total number of trials $N$, decreasing as $N$ increases. That is, the uncertainty in estimating $P_f$
decrease as N increases. Statistical rules can be used to establish the necessary number of trials which depends on the magnitude of $P_f$. Many variance reduction techniques have been developed to decrease the variance of $\frac{n}{N}$ with smaller number of trials than would have been necessary otherwise.

In the second method, the probability of failure is estimated by first fitting an approximate probability distribution for $g(X)$ using the trial values as described by Thoft-Christensen, P. and Baker, M., (1982). The moment or any other estimated statistical method may be used in the fitting process. Elderton and Johnson (1969) suggested some distributions that are suitable for fitting the $g(X)$ data. The probability of failure is then determined from

$$P_f = \int_{-\infty}^{M} f_M(m)dm$$

(2.5.2)

where $M = g(X)$ is a random variable representing the margin and $f_M(m)$ is its probability density function as estimated from the fitting process.

2.5.2 Generation of random numbers for a random variable with a prescribed continuous probability distribution

As mentioned earlier, the process of generating random numbers with a specified probability distribution may be accomplished by first generating uniformly distributed random numbers between 0 and 1. Through appropriate transformation, one may then obtain a corresponding random number with a specified probability distribution. This section will first discuss how to generate uniformly distributed random numbers, then how to obtain the corresponding random numbers with a specified probability distribution.

2.5.2.1 Random numbers with uniform distributions

In physical methods, special devices may be used to generate uniformly distributed random numbers within a given range. For example, by equally subdividing the circumference of a wheel into a number of intervals equal to the given range and spinning the wheel, the desired uniformly distributed random number can be generated. Uniformly distributed random numbers are also tabulated and are available in the literature for pencil-and-paper Monte Carlo simulations.

The most common practical approach is to employ a ‘pseudo’ random number generator (PRNG) such as is available on virtually all computer systems. They are termed ‘pseudo’ since they use a formula to generate a sequence of numbers. Although this sequence is reproducible and repeats after (normally) a long cycle interval, for many principal purposes it is indistinguishable from a sequence of strictly true random numbers (Rubinstein, 1981). The production of a reproducible sequence has an advantage in certain problems and in research work. If required, reproducibility can be destroyed simply by changing (randomly) the ‘seed number’ required as input for most PRNGs. A simple device is to use the local time as a seed value.
There are some mathematical reservations about the terminology ‘random sampling’. As soon as a table of ‘random numbers’ or a PRNG is used, the sequence of numbers is determined, and so no longer is truly random. It follows that ‘tests of randomness’ are therefore strictly misnomers. These tests usually are applied only for a one-dimensional sequence, and may not guarantee complete ‘randomness’ in more dimensions. To avoid the philosophical difficulties, it has been suggested that Monte Carlo methods using PRNGs be dubbed ‘quasi Monte Carlo’ methods and that the only justification for the use of sequences of pseudo-random numbers is the equi-distribution (or uniform distribution) property implied by their use. In essence, the numbers so elected give a reasonable guarantee of accuracy in the computations (Zaremba, 1968).

In computer simulation, methods for generating uniformly distributed random numbers are generally based on recursive calculations, which, because of cyclic effects, do not generate truly random numbers. The generated set eventually repeats itself after a very long cycle and, therefore, referred to as pseudo-random or quasi-random. An example of a recursive calculation of the residues of modulus ‘m’ that produce such a set of pseudo-random numbers is

\[ x_i = ax_{i-1} + b \pmod{m} \]  

(2.5.3)

where \( a, b \) and \( m \) are non-negative integers and the quotients \( x_i/m \) constitute the sequence of pseudo-random numbers (Elishakoff, I., 1983, Ang, A. H.-S. and Tang, W. H., 1984). Such a sequence repeats itself after almost \( m \) steps, i.e., cyclic. For this reason \( m \) must be set very large e.g. \( 10^8 \) or larger.

2.5.2.2 Random numbers with prescribed distributions

Based on a generated set of uniformly distributed random numbers between 0 and 1, one may obtain the corresponding random numbers with a specified probability distribution. This can be done using a method known as the ‘inverse-function’ method. The method is suitable if the inverse of the prescribed cumulative distribution function (CDF) of the random variable can be expressed analytically. The method is illustrated as follows.

Suppose that a set of random numbers is to be generated for a random variable \( Y \) which follows a prescribed distribution with CDF \( F_Y(y) \). The value of \( Y \) at \( F_Y(y) = x \) is

\[ y = F_Y^{-1}(x) \]  

(2.5.4)

where \( F_Y^{-1}(y) \) is the inverse of the CDF at ‘x’. If \( X \) is a uniformly distributed random variable between 0 and 1, then

\[ F_X(x) = x \quad 0 \leq x < 1 \]

Thus if \( x \) is an outcome of the random variable \( X \), the corresponding value of \( Y \) obtained from (2.5.4) will satisfy the following equations.

\[ P[Y \leq y] = P[F_Y^{-1}(X) \leq y] = P[X \leq F_Y(y)] = F_X[F_Y(y)] = F_Y(y) \]  

(2.5.5)

This means that if \( (x_1, x_2, \ldots, x_n) \) is a set of numbers of the random variable \( X \), the corresponding number obtained from equation (2.5.4), i.e.,
\[ y_i = F_y^{-1}(x_i) \quad i = 1, 2, \ldots, n \] (2.5.6)

will have the CDF \( F_y(y) \) as required. As an example consider the random variable \( Y \) to have a Weibull distribution with CDF given by
\[ F_y(y) = 1 - e^{-(y/k)^l} \quad y \geq 0 \] (2.5.7)

The inverse function is
\[ y = F_y^{-1}(x) = k[-\ln(1-x)]^{1/l} = k[-\ln x]^{1/l} \] (2.5.8)

since \( x \) and \((1-x)\) have identical distributions, i.e., uniform distribution. Thus one can generate the random numbers \( y_i, i = 1, 2, \ldots, n \) corresponding to uniformly distributed random numbers \( x_i \) according to (2.5.8) form:
\[ y_i = k[-\ln x_i]^{1/l} \] (2.5.9)

The Weibull distribution can be reduced to the Rayleigh and the exponential distributions. If the Weibull parameters \( k \) and \( l \) are equal to \( E^2 \) and 2, respectively, the Weibull distribution reduces to the Rayleigh distribution. If \( k \) and \( l \) are equal to \( \lambda \) and 1, it reduces to the exponential distribution. Thus, substitution for \( k \) and \( l \) in equation (2.5.9) will lead to a set of random numbers for these two special distributions as well.

### 2.5.3 Sample size and variance reduction techniques

#### 2.5.3.1 Number of samples required

As mentioned earlier, the simulated data according to Monte Carlo method should be treated as a sample of experimental observation and is, therefore, subjected to sampling error (SSC-351). If the probability of structural failure is to be computed from the simulated data, the underlying error becomes an important consideration since the sought probability of failure is usually small. Shooman (1968) developed the following expression for estimating the error in the estimated probability of failure:
\[ \text{error} = 2 \sqrt{\frac{1-P_f}{NP_f}} \] (2.5.10)

where \( N \) is the total number of simulations (sample size) and \( P_f \) is the probability of failure. There is a 95% chance that the actual error in the estimated probability is less than that given by equation (2.5.10). It is seen that the error is dependent on the number of simulations \( N \) and the probability of failure \( P_f \), it decreases by increasing \( N \) or \( P_f \). Therefore, if the estimated probability \( P_f \) is small which is usually the case, \( N \) should be large enough to decrease the error.

The following is an estimation method for the number of samples required (Robert E. Melchers, 2001), i.e., it is an estimate of the number of simulations required for a given confidence level. Since \( G(X) \) is a random variable in \( X \), the indicator function \( I[G(X) \leq 0] \) is also a random variable, albeit with only two possible outcomes. It follows from the central limit theorem that the distribution of \( J_t \) given by the sum of independent sample functions (2.5.11) approaches a normal distribution as \( N \to \infty \).
\[ P_f \approx J_1 = \frac{1}{N} \sum_{j=1}^{N} I[G(X_j) \leq 0] \]  

(2.5.11)

The mean of this distribution is

\[ E(J_1) = \frac{1}{N} \sum_{i=1}^{N} E[I(G \leq 0)] = E[I(G \leq 0)] \]  

(2.5.12)

which is equal to \( J \), while the variance is given by

\[ \sigma^2_{J_1} = \frac{1}{N^2} \text{var}[I(G \leq 0)] = \frac{\sigma^2_{I(G \equiv 0)}}{N} \]  

(2.5.13)

This shows that the standard deviation of \( J_1 \) and hence of the Monte Carlo estimate (2.5.11) varies directly with the standard deviation of \( I(.) \) and inversely with \( N^{0.5} \). These observations are important in determining the number of simulations required for a particular level of confidence. To actually calculate confidence levels, an estimate of \( \sigma_{I(.)} \) is required. The variance is given by

\[ \text{var}[I(.)] = \int \cdots \int [I(G \leq 0)]^2 dX - J^2 \]  

(2.5.14)

so that the sample variance is given by:

\[ S_{I(G \equiv 0)}^2 = \frac{1}{N-1} \left( \sum_{j=1}^{N} I[G(\tilde{X}_j) \leq 0] - N \left\{ \frac{1}{N} \sum_{j=1}^{N} I[G(\tilde{X}_j) \leq 0] \right\}^2 \right) \]  

(2.5.15)

where the last \( \{.\} \) term is simply the mean (2.5.12) or, by (3.5.11) the estimate \( J_1 \) for \( P_f \).

On the basis that the central limit theorem applies, the following confidence statement can be given for the number \( (J_1) \) of trials in which 'failure' occurs:

\[ P(-k\sigma < J_1 - \mu < k\sigma) = C \]  

(2.5.16)

where \( \mu \) is the expected value of \( J_1 \) given by (2.5.12) and \( \sigma \) is given by (2.5.13). For confidence interval \( C=95\% \), \( k=1.96 \), as can be verified from standard normal tables (Robert E. Melchers, 2001). As \( \sigma \) is not known, it may be estimated using (2.5.15).

However, this is not very helpful at the beginning of a Monte Carlo study.

Shooman (1968) has suggested that \( \sigma \) and \( \mu \) in (2.5.16) can be approximated by binomial parameters \( \sigma = (Nqp)^{0.5} \) and \( \mu = Np \), with \( q = 1 - p \), provided that \( Np \geq 5 \) when \( p \leq 0.5 \). If these are substituted in (2.5.16) the following is obtained:

\[ P\left[-k(Npq)^{0.5} < J_1 - np < k(Npq)^{0.5}\right] = C \]  

(2.5.17)

If the error between the actual value of \( J_1 \) and the observed value is denoted by

\[ \varepsilon = \frac{J_1 - np}{Np} \]  

and this is substituted into (2.5.17), it follows easily that

\[ \varepsilon = k \sqrt{\frac{1-p}{Np}} \].
Thus, for N=100000 samples, and \( p = P_f = 10^{-3} \) (expected), the error in \((J_i)\) and hence \(P_f\) will be less than 20% with 95% confidence (as then \(k=1.96\)).

Bording et al. (1964) suggested that a first estimate of the number \(N\) of simulations for a given confidence level \(C\) in the failure probability \(P_f\) can be obtained from

\[
N > \frac{-\ln(1-C)}{P_f} \tag{2.5.18}
\]

Thus, for 95% confidence level and \(P_f=0.001\), the required number of simulations is more than 3000. The actual number of variates to be calculated is, of course, \(N\) times the number of independent basic variables. Others have suggested that the number of simulations may need to be of the order of 10000-20000 for approximately 95% confidence limit, depending on the function being evaluated (Mann et al., 1974).

The above ‘rules’ while often useful, do not tell the analyst much about the accuracy being achieved in any particular Monte Carlo analysis. A useful tool for this purpose is to plot progressive results of the estimate of \(P_f\) and its estimated variance as obtained from (2.5.11) and (2.5.15) respectively.

Typically such plots will show that these measures are reduced as the number of samples is increased and that a degree of stability is reached at a sufficiently high number of samples. However, the rate of convergence, and its stability, depend, to some extent, on the quality of the random number generator being used. Hence it is possible for apparent stability to be attained under some (unfavourable) circumstances.

The following is a suggestion from DNV (DNV, Classification Notes-No.30.6, 1992). Reliability estimates by simulation methods are considered as verified if a sufficient number of simulations are carried out. Simulations by indicator-based Monte Carlo methods should be carried out with a number of simulation samples not less than \(100/P_f\). Simulations by other methods should be carried out such that the estimate of \(P_f\) has a coefficient of variation less than 10%.

### 2.5.3.2 Variance reduction techniques: ‘antithetic variates’

There are techniques, however, which may reduce the error (or variance) without increasing the sample size. These techniques are known as variance reduction techniques, and the one that is used often in the structural failure problems is called ‘antithetic variates’.

The following is an introduction of antithetic variates method.

Let \(Y_1\) and \(Y_2\) be two unbiased estimates of \(Y\) as determined from two separate sets of samples or simulation cycles.

The average of these two unbiased estimations \(Y_a = \frac{Y_1 + Y_2}{2}\) is also an unbiased estimator since its expected value \(E[Y_a]\) is equal to \(Y\). The variance \(\sigma^2_{Y_a}\) of the new estimator \(Y_a\) is determined from the individual variances \(\sigma^2_{Y_1}\) and \(\sigma^2_{Y_2}\) as:
If \( Y_1 \) and \( Y_2 \) are negatively correlated, i.e., the \( \text{cov}(Y_1, Y_2) < 0 \), it is seen from equation (2.5.19) that the third term becomes negative and

\[
\sigma_{y_v}^2 = \sigma_{y_1}^2 + \sigma_{y_2}^2 + 2\text{cov}(Y_1, Y_2)
\]

(2.5.19)

That is, the accuracy of the estimator \( Y_v \) can be improved (or its variance can be reduced) if \( Y_1 \) and \( Y_2 \) are negatively correlated estimators. The Antithetic Variates method is thus a procedure that ensures a negative correlation between \( Y_1 \) and \( Y_2 \). This can be accomplished in structural reliability problems as follows.

If \( X \) is a random variable uniformly distributed between 0 and 1, then \( 1-X \) is also a uniformly distributed random variable between 0 and 1 and the two random variables \( X \) and \( 1-X \) are negatively correlated. Each of these random variables can be then used to generate the basic random variables \( Y_i \), which have prescribed probability distributions as described earlier. This results in a pair of negatively correlated basic random variables. The procedure is repeated for all the random variables \( Y_i \) in the limit state equation. The limit state equation is then solved for each negatively correlated set of random variables separately and the results are averaged to estimate the population mean. Note that the error (or variance) of the result is reduced because of the negative correlation between the generated variables according to equation (2.5.20).

### 2.5.4 Importance sampling simulation method

For a given level of confidence, the basic (crude) Monte Carlo method requires a large number of samples in general. In order to look for a relative efficiency method, sample reduction techniques are developed. The following section will present an advanced sample reduction method, i.e., the importance sampling simulation method. This method requires far fewer sample points than using the ‘crude’ Monte Carlo method.

#### 2.5.4.1 Theory of importance sampling

The multiple integral (2.2.1) can be written, equivalently, as the following equation:

\[
J = \int \cdots \int \left[ I[G(X \leq 0)] \frac{f_X(X)}{h_v(X)} h_v(X) dX \right]
\]

(2.5.21)

where \( h_v(.) \) is termed the ‘importance-sampling’ probability density function whose definition will be considered in detail below. Again by comparison with

\[
E(X) \equiv \mu_x = \int -\infty^{\infty} x f_X(x) dx
\]

expression (2.5.21) can be written as an expected value:

\[
J = E \left[ I[G(X \leq 0)] \frac{f_X(X)}{h_v(X)} \right] = E \left( \frac{f}{h} \right)
\]

(2.5.22)

where \( V \) is any random vector with probability density function \( h_v(V) \). Obviously it is required that \( h_v(V) \) exists for all valid \( V \) and that \( f_v(V) \neq 0 \).
An unbiased estimate of $J$ is given by (cf. 2.5.11)

$$P_j \approx J_2 = \frac{1}{N} \sum_{j=1}^{N} I[G(\widetilde{V}_j) \leq 0] \frac{f_X(\widetilde{V}_j)}{h_v(V_j)}$$

where $\widetilde{V}_j$ is a vector of sample values taken from the importance sampling function $h_v(.)$.

It is evident that $h_v(.)$ governs the distribution of the samples. How this distribution should be chosen is quite important. By direct analogy with (2.5.13) and (2.5.14), the variance of $J_2$ is given by

$$\text{var}(J_2) = \frac{1}{N} \text{var}\left(\frac{f_f}{h}\right)$$

where

$$\text{var}\left(\frac{f_f}{h}\right) = \int \cdots \int \left\{ I[\cdot] \frac{f_x(X)}{h_v(X)} \right\}^2 h_v(X) dX - J^2$$

Clearly (2.5.25) should be minimized to minimize $\text{var}(J_2)$. If the function $h_v(.)$ is selected as

$$h_v(V) = \frac{\left[ f_x(V) \right]}{\int \cdots \int \left[ f_x(V) \right] dV}$$

then, upon substitution into (2.5.25) it is easily found that

$$\text{var}\left(\frac{f_f}{h}\right) = \left( \int \cdots \int \left[ f_x(V) \right] dV \right)^2 - J^2$$

If the integral $\int \left[ f_x(V) \right] dV$ does not change sign, the multiple integral is identical with $J$ and $\text{var}(f_f/h) = 0$. In this case the optimal function $h_v(.)$ is obtained from (2.5.26) as:

$$h_v(V) = \frac{\left[ I[G(V) \leq 0] f_x(X) \right]}{J}$$

At first sight this expression is not very helpful since the very integral to be determined, $J$, is needed. However, progress can be made even if only an approximate value can be estimated for $J$. Then the variance can be reduced using (2.5.25) or (2.5.27). Evidently, it is desirable that $h_v(.)$ should have the form of the integrand of (2.2.1), divided by the estimate for $J$. Similarly, the variance is reduced if $\int \left[ f_x(V) \right] h_v(.) = cons \tan t < 1$ [Shreider, 1966].

If the integrand does change sign, then the addition of a sufficiently large constant may adapt it to the above form. A weaker result when the integrand does change sign can also be given (Kahn, 1956; Rubinstein, 1981).

From (2.5.27) it follows that a good choice of $h_v(.)$ can actually produce zero variance in the estimate for $J$ for the special case of non-negative integrand. This may seem to be a surprising result. However, it demonstrates that the more effort is put into obtaining a close initial estimate of $J$ in equation (2.5.28), the better the Monte Carlo result will be.
Conversely, and importantly, the variance is likely to be increased by using a very poor choice of \( h_v(.) \) (Kahn, 1956).

2.5.4.2 Choice of importance sampling functions
In general, the derivation of optimal \( h_v(.) \) functions is difficult. In the n-dimensional reliability problem, the region of most interest is the “hyperzone” \( G(X) \leq 0 \) and, more particularly, the region of greatest probability density within that zone. In general the region cannot be easily identified, as there are no simple criteria with which to work. However, a surrogate for identifying the region of interest is to identify the design point \( x^* \). In this case the point \( x^* \) is the point having the largest ordinate \( f_x(X) \) on the limit state function. This point is known also as the point of ‘maximum likelihood’. For most probability density functions \( f_x(X) \), the point \( x^* \) may be found by the direct application of numerical maximization techniques. In some cases, such as when \( f_x(X) \) has a rectangular distribution, no unique maximum point \( x^* \) may exist. This is not as critical as it may seem at first, since \( P_f \) is usually quite small and therefore the region of \( f_x(X) \) satisfying \( G(X) \leq 0 \) is also small. In consequence, the choice of an approximate point \( x^* \) is restricted and may be made rather arbitrarily within \( G(X) \leq 0 \) although a point close to the maximum mass location of \( f_x(X) \) would be preferable.

Once \( x^* \) is identified, one approach for choosing \( h_v(.) \) is simply to use the distribution \( f_x(X) \) shifted so that its mean is at \( x^* \) (Harbitz, 1983). However, a more appropriate distribution for \( h_v(.) \) is

\[
V = \Phi(V, C_v)
\]

where \( C_v \) is a strictly diagonal matrix of \( \sigma_i^2 \) and with the mean of \( V \) placed at \( x^* \).

Such a distribution will produce sample points unbiased with respect to each variable. It will also give greater sampling density in the neighbourhood of the region where it is of most significance. The choice of \( x^* \) as the locating point for \( h_v(.) \) has the obvious advantage that it gives a systematic procedure for its selection. The shape or format of the limit state function \( G(X) = 0 \) is of little importance, since with \( h_v(.) \) centred at \( x^* \), a wide region is included in the sample space. This means that problems with nonlinear regions are handled as easily as those with linear limit state functions. Similarly, the format of \( f_x(X) \) is of little significance; dependence between the \( X_i \) does not affect the procedure (Robert E. Melchers, 2001).

2.5.4.3 Observations about importance sampling functions
The sampling function suggested in the previous section has been shown to be very effective and robust for a range of possible shapes of limit state function (Engelund and Rackwise, 1993; Melchers and Li, 1994). Nevertheless, there are some matters which must considered as possible limitations and which will be addressed bellows (Melchers, 1991):
a) $h_v(.)$ may not be well-chosen, such as being too ‘fat’ or being skewed.
b) Extremely concave limit state functions may lead to very low sampling efficiency.
c) A unique point of maximum likelihood $x^*$ may not be identifiable:
   (1) When $f_X(.)$ has a ‘flat’ contour region
   (2) When the limit state function $G(.) = 0$ coincides with a contour of $f_X(.)$
       in a region of interest, or
   (3) When the limit state function $G(.) = 0$ is non-smooth (e.g. ripple-like)
       and has a number of candidate locations for $x^*$
d) There may be more than one point of local maximum likelihood such as when
   $f_X(.)$ is not uni-model.

Examples of these matters can be found in some literatures such as Robert E. Melchers, (2001). It is important to note that not all the above issues are confined to importance sampling only. Thus points b), c) and d) also produce difficulties for the FOSM methods.

Generally, the $h_v(.)$ should attempt to mirror as closely as possible the shape of the failure domain. Thus when the limit state function forms a surface which is close to spherical, or partly so, an approximate $h_v(.)$ could be obtained by using a multi-normal probability density function centred at or near the origin of the ‘spherical’ limit state system and censoring all samples which would lie in the safe domain. In two dimensions the resulting $h_v(.)$ is then like a doughnut with a hole in the middle. However, this rather special problem probably is better handled by directional simulation (refer to Robert E. Melchers, 2001).

The above examples reinforce the comment (Kahn, 1956) that the choice of $h_v(.)$ involves a certain amount of intuition or ‘art’. This is not particularly satisfactory for problems in high-dimensional spaces and which therefore cannot be visualized. Various suggestions have been made to develop procedures to obtain an ‘optimal’ $h_v(.)$ for a given problem without the need for visualization. The more important and realistic of these are considered in the following section (Section 2.5.5).

### 2.5.5 Adaptive sampling simulation

As noted in section 2.5.4 the importance sampling progress becomes more efficient as the sampling function $h_v(.)$ becomes more closely proportional to $f_X(.)$ in the failure region. Various suggestions have been made to attempt to optimise $h_v(.)$ in some way.

One simple observation is that it may be possible to ‘censor’ part of the region included by $h_v(.)$, that is, sampling is prevented in a region for which it clear there will be no contribution to the estimate of the failure probability. The region $f_X(.) > f_X(X^*)$ can be censored for samples taken from $h_v(.)$ centred at $X^*$ as points in this region cannot make a contribution (Melchers, 1989). An exactly parallel case (Harbitz, 1986) exists in
the standardized normal space \( Y \) with the censored region described by the circle having radius \( \beta \), the safety index.

Another possibility for improving the importance sampling function is to select a trial \( h_v(.) \) and to perform some initial sampling. From this information, \( h_v(.) \) could be modified in some way in an attempt to improve the rate of convergence to the probability estimate. One proposal is to let \( h_v(.) \) be a composite of \( k \) elementary pre-selected sampling densities \( h_{\nu j}(.) \):

\[
h_{\nu j}(.) = \sum_{j=1}^{k} w_j h_{\nu j}(X)
\]  

(2.5.30)

where \( w_j, j=1,2,\cdots,k \) are weights whose values are selected to let \( h_{\nu j}(.) \) approach as closely as possible the shape of \( f_{X}(.) \) within the failure region. If the \( h_{\nu j}(.) \) are selected as components of \( f_{X}(.) \) but with smaller standard deviation, and each centred at some initial sample point \( \tilde{X}^* \), the weights \( w_j \) can be chosen as (Karamchandani et al., 1989):

\[
w_j = \frac{f_{X}(\tilde{X}_j)}{\sum_{i=1}^{k} f_{X}(\tilde{X}_i)}
\]  

(2.5.31)

where \( \tilde{X}_j \) is a sample vector. This means that the weights \( w_j \) are proportional to the contribution the sample vector \( \tilde{X}_j \) makes to the (current) estimate of the probability.

Once the weights are established from preliminary sampling, further sampling can be done and the process repeated to establish new weightings and a better estimate of the failure probability. To prevent clustering of newly obtained sampling densities \( h_{\nu j}(.) \) around previously obtained densities, sampling points located closer than some distance \( d_0 \) from the centre of the current sampling densities \( h_{\nu j}(.) \) might be ignored.

A slightly different approach is to modify \( h_v(.) \) by changing the weights \( w_j \) in (2.5.30) so as to minimize the variance in the estimate of the failure probability \( P_f \) (Ang et al., 1989; 1991). This can be done with a limited amount of Monte Carlo sampling to estimate \( P_f \) for setting the weights (or ‘kernels’). With an initial choice for the form of each \( h_{\nu j}(.) \), sampling from the constructed sampling density function \( h_v(.) \) obtained from (2.5.30) will allow refinement of the estimate for \( P_f \).

In both techniques the location of the initial \( h_{\nu j}(.) \) is based on the samples \( \tilde{X}_j \) obtained from an initial sampling scheme (either Monte Carlo or some importance sampling). This means that the effectiveness of further sampling in developing an improved sampling density function depends very much on the initial samples and therefore on the sampling distribution from which they are drawn. Moreover, it has been assumed that convergence is assured. In addition, assumptions need to be made at the beginning about appropriate initial or component sampling densities. More ambitious schemes, in which the sampling densities are modified as more information becomes available with additional sampling have been proposed to address these matters.
It has been proposed also that the form of \( h_v(.) \) can be based on asymptotic notions (Robert E. Melchers, 2001) about the shape and gradient of \( f_X(.) \) in the region of interest (Maes, et al., 1993). This approach is useful for very low probability situations and with explicit limit state functions.

It has suggested above that for any one limit state function the ‘best’ location for the importance sampling probability density is the point of maximum likelihood (i.e., design point). Usually, this point normally is not known and will require a prior analysis to locate it, particularly for problems of high dimensionality. An alternative approach is to seek out the point using a search algorithm with the constraint that the point must lie along the limit state. This would then fix the desirable location of \( h_v(.) \). It is possible also to modify \( h_v(.) \), depending on the information being obtained during the research process.

### 2.6 Response Surface Methods

In addition to the above structural reliability calculation methods, there are many other methods, such as Response Surface Methods, Directional Simulation Methods (Robert E. Melchers, 2001), etc. In this section the response surface method will be presented in detail.

In many practical problems the limit state function may not be known explicitly, such as through one or more equations to represent the limit state surface. Rather, it may be known only implicitly, through a procedure such as a finite element analysis. This means that the safe domain can be defined only through point-by-point ‘discovery’ such as through repeated numerical analysis with different input values. These values could be random, as in Monte Carlo analysis, or specifically ordered. In either case it is clear that the FOSM and related methods are not immediately applicable, as they require a closed, and preferably differentiable, form for the limit state function \( G(X) = 0 \).

A closed form and differential limit state surface can be constructed artificially using a polynomial or other suitable function fitted to the results obtained from a limited number of discrete numerical analyses. Clearly, these would need to represent the structural response most accurately in the area around the design point(s), with lower accuracy acceptable elsewhere. Provided the approximating surface fits the point responses reasonably well, a fairly good estimate of structural failure probability would be expected. This notion will be explored in a little more detail in this section.

#### 2.6.1 Basic theory

As is usual, let the structural response be represented by \( G(X) \). However, it is now an implicit function of the random variables \( X \). It can be evaluated only for discrete values of \( X \). Let \( \overline{X} \) represent a set of points in \( X \) space for which \( G(X) \) is evaluated. The ‘response surface’ approach is then to seek a function \( \overline{G}(X) \) which best fits the discrete set of values of \( G(\overline{X}) \). The usual approach is to let \( \overline{G}(X) \) be an \( n^{th} \) order polynomial.
The undetermined coefficients in this polynomial are then determined so as to minimize the error of approximation, particularly in the region around the design point. The order $n$ of the polynomial selected for fitting to the discrete point outcomes will affect both the number of such evaluations required and the number of derivatives which need to be estimated. Also, for a well-conditioned system it is desirable for $\bar{G}(X)$ to be of equal or lower degree than $G(X)$. Higher degree $\bar{G}(X)$ will lead to ill-conditioned systems of equations to be solved for the undetermined coefficients. Also, it is likely to have erratic behaviour.

Since the actual limit state function is known only through discrete outcomes, its form and degree are not known, nor can be design point be estimated, yet. This means there is little guidance for the selection of the approximating function $\bar{G}(X)$. However, a second order polynomial is most often employed for the response surface (Faravelli, 1989; Bucher and Bourgund, 1990; El-Tawil et al., 1992; Rajashekhar and Ellingwood, 1993; Maymon, 1993):

$$\bar{G}(X) = A + X^T B + X^T CX$$

where the undetermined (regression) coefficients are defined by $A$, $B^T = (B_1, B_2, \cdots, B_n)$ and

$$C = \begin{bmatrix}
    C_{11} & \cdots & C_{1n} \\
    \vdots & \ddots & \vdots \\
    C_{n1} & \cdots & C_{nn}
\end{bmatrix}_{sym}$$

The (regression) coefficients may be obtained by conducting of a series of numerical ‘experiments’, i.e. a series of structural analyses with input variables selected according to some ‘experimental design’. An approximate experimental design takes into account that the objective lies in estimating, reasonably accurately, the probability of failure. This implies that the main interest is in the region of the maximum likelihood within the failure domain, that is the design point. However, as noted, this will not be known initially. One simple approach is to select input variables for the experimental design around the mean value of the variables. Thus points located along the axes are appropriate. A simple experimental design is shown in Figure 2.6.1. More complex designs have been discussed by Rajashekar and Ellingwood (1993) and are also

Figure 2.6.1 Simple experimental design for a two-variable problem with $X_m$ as mean point
available in texts dealing with response surface methodology in general (e.g. Myers, 1971).

Once the experimental design has been selected and the function $G(X)$ evaluated, it will be found that there is a difference between the actual response $G(\bar{X})$ for the values $X = \bar{X}$ and the evaluation of (2.6.1) for the same values. This difference is due to intrinsic randomness and due to ‘lack of fit’ contributed by using the simpler function (2.6.1) to represent the actual limit state surface. It is not possible to separate them without more refined analysis. The important step, however, is to try select the (regression) coefficients $A, B, C$ such that the total error is minimized.

Let the vector $D$ represent the collection of regression coefficients $A, B, C$. Then (2.6.1) can be represented as $\bar{G}(D, X)$. Also, for each point $\bar{X}$ in the experimental design let $\varepsilon_i$ represent the error between the evaluation of the actual (but implicit) limit state function $G(\bar{X}_i)$ and the approximating surface (2.6.1) given now as $\bar{G}(D, \bar{X}_i)$. This would be done for each point in the experimental design. The total error can now be minimized. One simple approach is to use a least-squares-fit method, with $D$ selected to minimize the sum of squares of the errors:

$$S = \min_D \left[ \sum_i \varepsilon_i^2 \right]^{0.5} = \min_D \left[ \sum_i \left( \bar{G}(D, \bar{X}_i) - G(\bar{X}_i) \right)^2 \right]^{0.5} \quad (2.6.2)$$

### 2.6.2 Simplifications for large systems

In some problems of practical interest the number of variables is so large that the direct application of (2.6.1) is impractical. This might occur, for example, when the structural response is described using a finite element analysis. The number of random variables involved may be reduced in various ways.

One approach is simply to replace random variables of low uncertainty with deterministic equivalents.

Another approach is to reduce the set of random variables $X$ to the smaller set $X_A$ describing their spatial averages. This can be done, for example, by using the same (i.e. spatial average) yield strength at every point on a plate under stress, rather than specifically allowing for the variation of yield strength from point to point (or from finite element to finite element).

A third approach is to simplify the error effect of random variables (or spatial averages) as having only an additive effect on the response, rather than some more complex relationship (Faravelli, 1989). The enhanced expression (2.6.1) with additive error effects is then given by:

$$\bar{G}(X) = A + X^T B + X^T CX + \sum_j \left( e_j + \sum_k e_{jk} + \cdots \right) + \varepsilon \quad (2.6.3)$$

where the $e_j, e_{jk}, \cdots$ are the error terms due to spatial averaging (assumed here independent of the spatially averaged random variables) and $\varepsilon$ represents the remaining
errors (such as due to randomness) (Veneziano et al. 1983). This format is useful if the $e_j, e_{jk}, \cdots$ can be obtained separately, such as through an analysis of variance of the effect of spatial averaging. The error to be considered for fitting $\vec{G}(\vec{X})$ is then restricted only to the term $\varepsilon$.

### 2.6.3 Iterative solution scheme

The best points for fitting the approximating response surface to the actual (but implicit) limit state function are not usually known at first. It has been suggested that an iterative, search technique might be used to locate such points (Bucher and Bourgund, 1990). Let (2.6.1) have just sufficient undetermined coefficients to fit the evaluations $G(\vec{X})$ exactly – a so-called ‘fully saturated’ experimental design – so that the surface fits exactly at the evaluation points $X_i$.

Typically these might be at a mean point $X_m$ (see Figure 2.6.1) and at points given by $x_i = x_{mi} \pm h_i \sigma_i$ where $h_i$ is an arbitrary factor and $\sigma_i$ is the standard deviation of $X_i$. Using these points, the approximating surface $\vec{G}(\vec{X})$ for the assumed mean point $X_m$ can be determined exactly. If the approximating surface is located in the optimal position, the mean point $X_m$ would coincide with the point of maximum likelihood (the design point) and the distance from this point to the origin would be a minimum in standardized normal space (see Section 2.3).

If $X_m$ is not the design point, some other point, say $X_D$, can be found on the approximating surface $\vec{G}(\vec{X})$ which is closer to the origin and which is therefore the best estimate of the design point. With $X_D$ known, a new mean point $X'_m$ can be obtained by linear interpolation between $X_m$ and $X_D$, as:

$$X'_m = X_m + \frac{G(X_m)}{[G(X'_m) - G(X_D)]} (X_D - X_m)$$  \hfill (2.6.4)

The speed with which this technique approaches a sufficiently accurate fit depends on the selection of $h_i$ and the shape of the actual (but implicit) limit state function in the area being explored. It should be evident that the results are improved if the location of points is selected to correspond with points in the lower tail area for random variables representing resistance and in the upper tail area for random variables representing loads (Rajashekhar and Ellingwood, 1993). Nevertheless, convergence to all possibly relevant design points is not guaranteed, in general. It is possible to modify search algorithms to direct the research away from design points already obtained (e.g. Der Kiureghian and Dakessian, 1998).

### 2.6.4 Response surface and finite element analysis

Response surfaces have been proposed as a vehicle for incorporating finite element analysis into structural reliability computations when using FOSM/FORM techniques. For simulation techniques, of course, response surfaces are not specifically required for probability estimation.
In the present context it should be noted that the use of a quadratic (parabolic) response surface in standard normal space is essentially similar to the approach used for 'second order' second moment theory, with a quadratic (or parabolic) surface fitted either to a known non-linear limit state surface or to known discrete points, as in response surfaces. It follows that there is a close parallel between the discussion for these two approaches. In using the response surface method for finite element analysis, repeated solutions are made of the finite element code for a given set of selected points, i.e. those used for fitting the response surface (Liu and Der Kiurghian, 1991). If FORM is used, the fitted surface is linear and only a small number of points are required for determining the response surface, since, as for response surfaces generally, the work in fitting to the surface is independent of the probability to be estimated. More fitting points are required for SOSM/FORM approaches.

One requirement is that the gradients at the design point can be determined or estimated. Some finite element analysis codes can be modified to achieve this. Alternatively, an approximate finite difference-type approach can be used, at greater computational cost. Nevertheless, some commercial finite element codes have been integrated with response surface methods and FORM (e.g. Lemaire et al., 1997).

An important part of finite element modelling in a reliability context is the representation of random fields, such as might be required to represent the statistical variation of properties such as Young’s modulus across a plate. This is a rather specialized topic and various ways have been suggested for doing this (e.g. Vanmarcke et al., 1986; Li and Der Kiureghian, 1993; Zhang and Elingwood, 1995; Matthies et al., 1997).

### 2.6.5 Applications and basic idea of directional simulation method

A number of example applications of response surfaces have been given in the literature (e.g., Das and Zheng, 2000; Zheng and Das, 2000; Yu, Das, Zheng, 2002; Yu, Das and Zheng, 2001; Bucher and Bourgund, 1990; Faravelli, 1989; El-Tawil, Lemaire, and Muzeau, 1992; Rajashekhar and Ellingwood, 1993; etc.). These developments suggest that the concept of a response surface works well, provided the point of maximum likelihood can be identified and that reasonable decisions can be made about the points to be used for fitting the response surface. Various suggestions for this have been made (e.g., Kim and Na, 1997). They include the use of iterative methods (e.g. Liu and Moses, 1994), but the difficulty, particularly for large systems, is that such points cannot always be identified without subjective interference, that is, they require the input from the analyst. It is likely, however, that nearly flat limit state surfaces will be insensitive and that problems with highly curved limit state surfaces, or with intersecting limit state surfaces will be more sensitive. The response surface method is a useful supplementary for structural reliability analysis.
At the beginning of this section, another reliability analysis method was mentioned, i.e., directional simulation method. The following is the basic concept of directional simulation method.

So far in this chapter, the problem of estimating the probability of failure has been formulated exclusively in the Cartesian coordinate system. The idea of using a polar coordinate system arose from the specialized work of Deak (1980) in evaluating the multi-variate standard normal distribution function. This led to simulation in the polar coordinate standard normal space and was given the name ‘directional simulation’ (Ditlevsen, et al., 1987).

### 2.7 Choice of reliability methods

The above calculation methods each have their own advantages and drawbacks. A reliability method which gives acceptable estimates of the reliability for the structure or structural components should be used. The choice of the reliability methods must be justified. The justification may be based on verification by other relevant methods. When the limit state is a linear function, the FORM and SORM can be used to verify the results of the simulation and direct integral methods. Analytical FORM and SORM reliability estimate can generally be verified by simulation. When the number of basic random variables is under 5, the direct integral methods can be used to verify the results of analytical methods and simulation methods. In general, the simulation methods can be used to verify the results of the other methods. A local reliability estimate by FORM at a single design point can be verified by a SORM estimate. Simulation can then be used to verify if this local estimate is sufficient as an estimate of the global reliability when there is more than one design point.

For FORM and SORM solutions the design point shall be documented to have a sound physical interpretation.

To select the best method to calculate marine structural reliability, the following methods are suggested:

- For linear failure limit function or the failure probability less than 0.05, the analytical FORM and SORM reliability estimates are suggested.
- Where there are less than 5 variables, the directly integrated method is suggested.
- All others are calculated best by means of simulation methods.

In general, the simulation methods should be chosen first.

For implicit limit state function, response surface method should be selected.

For the same kinds of structures, such as marine structures, reliability analysis should be done by the same reliability analysis method, i.e., selecting one kind of reliability method only, such as simulation method or FORM, etc. This means a same standard should be used to same kind of structure, because:

- The analysis models of structural reliability are usually imperfect.
- The information about loads and resistance is usually incomplete.
c) Reliability methods are based on analysis models for the structure in conjunction with available information about loads and resistances and their associated uncertainties.
d) Reliability methods deal with the uncertainty nature of loads, resistance etc. and lead to assessment of the reliability.
e) The reliability, as assessed by reliability methods, is therefore generally not a purely physical property of the structure in its environment of actions, but rather a nominal measure of the safety of the structure, given a certain analysis model and a certain amount and quality of information.
f) Correspondingly, the estimated failure probability is also dependent on the analysis model and the level of information, and it can therefore usually not be interpreted as the frequency of occurrence of failure for the particular type of structure.

The reliability method used shall be capable of producing a full sensitivity analysis such as parametric sensitivities for changes in fixed variables and importance factors for uncertain variables. The sensitivity should be executed and documented as a part of the reliability analysis. For more details about sensitivity analysis see Chapter 3.

2.8 Conclusions
In this chapter, the basic theories and methods of reliability have been presented in detail.

First, analytical methods (approximate methods) are introduced. The advantages of the analytical methods are mainly that they are simple and have apparent physical interpretations and do not usually require excessively large computation time. Analytical methods usually contain the First-order and Second-order reliability methods, i.e., FORM and SORM, respectively. The drawback of these methods is that they do not give the exact results for the probabilities in question for most practical structures.

The Hasofer-Lind methods and the other advanced methods help to overcome the shortcomings of MVFOSM method.

Then, the direct integration method is presented, including the techniques of standardized integral region, joint probability density function, multivariate integration and the advanced method of integration. The analytical methods give approximate results, and should be applied for failure probability less than 0.05. For larger failure probability, direct integral method can be used, but because the direct integral method is not suitable for the failure modes with a large numbers of variables, other effective methods are also developed.

Third, the simulation methods are introduced. The Monte Carlo method consists of the following basic steps:

a) Simulation of the random variables and generation of several sample data using statistical sampling techniques
b) Solutions using the sampled data
c) Statistical analysis of the results
In this section also contains the following aspects:
a) Basic principle and method
b) Generation of random numbers for a random variable with a prescribed continuous
probability distribution
c) Sample size and variance reduction techniques
d) Importance sampling simulation method
e) Adaptive sampling simulation
In many practical problems the limit state function may not be known explicitly, such as
through one or more equations to represent the limit state surface. Rather, it may be
known only implicitly through a procedure such as a finite element analysis. This means
that the safe domain can be defined only through point-by-point ‘discovery’ such as
through repeated numerical analysis with different input values. These values could be
random, as in Monte Carlo analysis, or specifically ordered.

Fourth, in addition to the above calculation methods of structural reliability, response
surface methods are presented, including basic theory, simplifications for large systems,
iterative solution scheme and response surface and finite analysis etc.
Finally, suggestions are given for choice and justification of reliability methods. The
justification may be based on verification by other relevant methods. The following
suggestions are given for the method of choosing a structural reliability analysis
technique:
a) For linear failure limit function or the failure probability less than 0.05, the analytical
FORM and SORM reliability estimates are suggested.
b) For fewer than 5 variables, the directly integrated method is suggested.
c) The others except the above are calculated best by simulation methods.
d) In general, the simulation methods shall be chosen first.
e) For implicit limit state function, response surface method should be selected.
The reliability method used shall be capable of producing a full sensitivity analysis such
as parametric sensitivities for changes in fixed variables and importance factors for
uncertain variables. The sensitivity should be executed and documented as a part of the
reliability analysis.