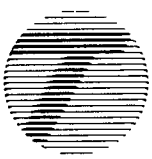
REPORT FHWA/NY/SR-93/112

Improved Direct Importance Sampling Method For System Reliability

GONGKANG FU



SPECIAL REPORT 112

ENGINEERING RESEARCH AND DEVELOPMENT BUREAU
NEW YORK STATE DEPARTMENT OF TRANSPORTATION
Mario M. Cuomo, Governor/John C. Egan, Commissioner

STATE OF NEW YORK

Mario M. Cuomo, Governor

DEPARTMENT OF TRANSPORTATION

John C. Egan, Commissioner
Michael J. Cuddy, Assistant Commissioner for Engineering and Chief Engineer
Paul J. Mack, Deputy Chief Engineer, Technical Services Division
Robert J. Perry, Director of Engineering Research and Development

The Engineering Research and Development Bureau conducts and manages the engineering research program of the New York State Department of Transportation. The Federal Highway Administration provides financial and technical assistance for these research activities, including review and approval of publications.

Contents of research publications are reviewed by the Bureau's Director, and the appropriate section head. However, these publications primarily reflect the views of their authors, who are responsible for correct use of brand names and for the accuracy, analysis, and inferences drawn from the data.

It is the intent of the New York State Department of Transportation and the Federal Highway Administration that research publications not be used for promotional purposes. This publication does not endorse or approve any commercial product even though trade names may be cited, does not necessarily reflect official view or policies of either agency and does not constitute a standard, specification, or regulation.

ENGINEERING RESEARCH PUBLICATIONS

A. D. Emerich, Editor
Donna L. Noonan, Graphics and Production
Nancy A. Troxell and Jeanette M. LaClair, Copy Preparation

IMPROVED DIRECT IMPORTANCE SAMPLING METHOD FOR SYSTEM RELIABILITY ANALYSIS

Gongkang Fu, Engineering Research Specialist II

A Report Published under Research Project 1200
Conducted in Cooperation With
The U.S. Department of Transportation
Federal Highway Administration

Special Report 112
September 1993

ENGINEERING RESEARCH AND DEVELOPMENT BUREAU
New York State Department of Transportation
State Campus, Albany, New York 12232

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.		3. Recipient's Catalog No.
FHWA/NY/SR-93/112	PB99-100224	

4. Title and Subtitle	5. Report Date
IMPROVED DIRECT IMPORTANCE SAMPLING METHOD FOR SYSTEM RELIABILITY	September 1993
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
Gongkang Fu	Special Report 112
9. Performing Organization Name and Address Engineering Research and Development Bureau New York State Department of Transportation Albany, New York 12232	10. Work Unit No.
	11. Contract or Grant No.
	13. Type of Report and Period Covered
12. Sponsoring Agency Name and Address Offices of Research, Development & Technology HRD-10 Federal Highway Administration U.S. Department of Transportation Washington, DC 20590	14. Sponsoring Agency Code

15. Supplementary Notes
Prepared in cooperation with the U.S. Department of Transportation, Federal Highway Administration.

16. Abstract

This report presents an importance-sampling method for the first-order problem of reliability analysis of structural systems, having a failure domain defined by linear or linearized functions. Truncated multimodal simulation is suggestsed as a new technique, offering an advantage of locating all samples in the failure domain and thus increasing computation efficiency. Variance of estimator is evaluated by an analytically derived upper bound. It is compared with that of the conventional Monte Carlo method by a variance change factor for conservative estimation of increase in accuracy and efficiency. The upper bound of variance can be used for a priori determination of required sample size, given an acceptable maximum error associated with a confidence level. Various application examples of both series and parallel systems are included for illustration.

17. Key Words	18. Distribution Statement
Structures, system reliability, general importance sampling, direct importance sampling, tuncated multimodal simulation, Monte Carlo method, variance reduction	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
		v + 23	

Form DOT F 1700.7 (8-69)

Preceding page blank iii

CONTENTS

1. INTRODUCTION

Application of the importance sampling method to structural system reliability problems has recently attracted attention. Various approaches have been presented in the literature [e.g., Melchers 1984, 1987, 1990a, 1990b; Fu and Moses 1986, 1987a, 1987b, 1988; Schuëller and Stix 1987; Bjerager 1988], and a comparison of some methods has been initiated by Engelund and Rackwitz [1992]. Accuracy and efficiency of these methods can only be evaluated on a case-by-case basis and after the simulation. This paper presents an importance sampling method for the first-order problems to be defined here. Major advantages of this method are its abilities of locating all samples in the failure region and evaluating its accuracy and efficiency before the simulation, as an advance to the method suggested by Fu and Moses [1987a, 1988].

2. FORMULATION AND DEFINITIONS

A structural system reliability problem can be formulated as assessment for system failure probability P_f:

$$P_f = \int_x G(\mathbf{x}) f(\mathbf{x})\, d\mathbf{x} \tag{1}$$

where $f(\mathbf{x})$ is assumed to be the normal probability density function of random variable vector $\mathbf{X}$ with mean vector $\underline{\mathbf{X}}$ and symmetric covariance matrix $\mathbf{C}$; $\mathbf{x} \varepsilon \mathbf{R}^n$ is realization of $\mathbf{X}$, with n being the dimension; and $G(\mathbf{x})$ is an indicator for the structure's failure state:

$$G(\mathbf{x}) = \begin{cases} 0 & \mathbf{x} \varepsilon \underline{S} \quad \underline{S} = \{\mathbf{x}: \text{ structural survival}\} \\ 1 & \mathbf{x} \varepsilon \underline{F} \quad \underline{F} = \{\mathbf{x}: \text{ structural failure}\} \end{cases} \tag{2}$$

It is understood that $\mathbf{C}$ is positive definite. $\mathbf{X}$ can be orthogonally transformed to independent normal variables $\mathbf{X}'$ with zero mean vector $\mathbf{0}$ and diagonal covariance matrix $\mathbf{C}'$:

$$\mathbf{X} = \underline{\mathbf{X}} + \mathbf{B}^t \mathbf{X}' \tag{3}$$

where $\mathbf{B}$ is the orthogonal transformation matrix having the eigenvectors of $\mathbf{C}$ as its columns; $\mathbf{C}'$ has the eigenvalues of $\mathbf{C}$ as its diagonal terms; and superscript t denotes transpose. The assumption of normal variables $\mathbf{X}$ imposes no restriction on generality of the problem, since random variables can always be transformed to normal variables [Hohenbichler and Rackwitz 1981].

Let $g_m(\mathbf{x})$ be the failure state indicator of mode m:

$$g_m(\mathbf{x}) = \begin{cases} 0 & \mathbf{x}\in\underline{S'}_m;\ \underline{S'}_m = \{\mathbf{x}:\ z_m(\mathbf{x})>0\} = \{\mathbf{x}:\text{ Failure mode m does not occur}\} \\ 1 & \mathbf{x}\in\underline{F'}_m;\ \underline{F'}_m = \{\mathbf{x}:\ z_m(\mathbf{x})\le 0\} = \{\mathbf{x}:\text{ Failure mode m occurs}\} \end{cases} \tag{4}$$

where $z_m(\mathbf{x})$ is the corresponding limit state function. A series structural system can be expressed as

$$G(\mathbf{x}) = \text{Min}\ \{g_m(\mathbf{x})\}\ \ (m=1,2,\ldots,M)\ \text{ or }\ \underline{F} = \underline{F'}_1\cup\underline{F'}_2\cup\ldots\cup\underline{F'}_{M-1}\cup\underline{F'}_M \tag{5}$$

and a parallel system can be modeled by

$$G(\mathbf{x}) = \text{Max}\ \{g_j(\mathbf{x})\}\ \ (j=1,2,\ldots,J)\ \text{ or }\ \underline{F} = \underline{F'}_1\cap\underline{F'}_2\cap\ldots\cap\underline{F'}_{J-1}\cap\underline{F'}_J \tag{6}$$

Note that a parallel system is treated here as a special case of series system with M=1 (with one design point). A more general problem is considered here as a series of parallel systems defined in Eq. 5, with $\underline{F'}_m$ being intersection of V_m subfailure sets:

$$\underline{F'}_m = \underline{F'}_{m,1}\cap\underline{F'}_{m,2}\cap\ldots\cap\underline{F'}_{m,Vm-1}\cap\underline{F'}_{m,Vm}\quad (m=1,2,\ldots,M) \tag{7}$$

P_f in Eq. 1 can be analytically calculated for only limited cases, mainly because of the irregular failure domain $\underline{F}$ in a space of multiple dimensions. The Monte Carlo simulation method is thus often used, although its variance of estimator can only be reduced by increasing the sample size. To reduce the variance, the importance sampling method [Kahn 1956, Kleijnen 1974] employs a new sampling distribution function $p(\mathbf{x})$:

$$P_f = \int_{\mathbf{x}} \frac{G(\mathbf{x})f(\mathbf{x})}{p(\mathbf{x})}\,p(\mathbf{x})d\mathbf{x} \tag{8}$$

The mean and variance of its estimator P_f' are, respectively [e.g., Fu 1988]

$$E[P_f'] = P_f \tag{9a}$$

$$\text{Var}[P_f'] = \frac{1}{N}\left\{\int_{\mathbf{x}} \frac{G^2(\mathbf{x})f^2(\mathbf{x})}{p^2(\mathbf{x})}\,p(\mathbf{x})d\mathbf{x} - P_f^2\right\} \tag{9b}$$

showing that if $p(\mathbf{x})$ is properly selected, the variance may decrease or vanish without increasing sample size N.

This report presents an importance sampling method for the first-order problems defined as those with linear or linearized failure surfaces:

$$g_m(\mathbf{x}) = \begin{cases} 0 & \mathbf{x}\in\underline{S}_m;\ \underline{S}_m = \{\mathbf{x}:\ \mathbf{a}_m^t(\mathbf{x}-\mathbf{x}_m^*)>0\} \\ 1 & \mathbf{x}\in\underline{F}_m;\ \underline{F}_m = \{\mathbf{x}:\ \mathbf{a}_m^t(\mathbf{x}-\mathbf{x}_m^*)\le 0\} \end{cases} \quad (m=1,2,\ldots,M) \tag{10a}$$

where $\mathbf{x}_m^*$ is the maximum likelihood point (or design point) [Hasofer and Lind 1974, Shinozuka 1983] for mode m, and $\mathbf{a}_m$ is the gradient vector of $z_m(\mathbf{x})$ at $\mathbf{x} = \mathbf{x}_m^*$:

$$\mathbf{a}_m = \nabla z_m(\mathbf{x})\big|_{\mathbf{x} = \mathbf{x}_m^*} \tag{10b}$$

Problems with linear failure surfaces are often of practical interest. Furthermore, a problem with nonlinear failure surface(s) can be always approximated by a series of piecewise linear ones, becoming a first-order problem defined here.

3. TRUNCATED MULTIMODAL IMPORTANCE SAMPLING

A multimodal simulation scheme is suggested for the first-order problems:

$$p(\mathbf{x}) = \sum_{m=1,M} w_m \, p_m(\mathbf{x}) \tag{11}$$

where $p_m(\mathbf{x})$ is the normal probability density function with mean vector $\mathbf{x}_m^*$ and covariance matrix $\mathbf{C}$, and truncated by the hyperplane $\mathbf{a}_m^t(\mathbf{x} - \mathbf{x}_m^*)=0$:

$$p_m(\mathbf{x}) = \begin{cases} \dfrac{2}{(2\pi)^{n/2}|\mathbf{C}|^{1/2}} \exp\{-0.5(\mathbf{x} - \mathbf{x}_m^*)^t \mathbf{C}^{-1}(\mathbf{x} - \mathbf{x}_m^*)\} & \mathbf{x}\varepsilon \underline{F}_m \\ 0 & \mathbf{x}\varepsilon \underline{S}_m \end{cases} \tag{12}$$

where $|\mathbf{C}|$ is the determinant of $\mathbf{C}$. The weight coefficients w_m are to be determined by solving:

$$\frac{f(\mathbf{x}_1^*)}{f(\mathbf{x}_m^*)} = \frac{p(\mathbf{x}_1^*)}{p(\mathbf{x}_m^*)} \qquad (m=2,3,\ldots,M)$$

$$\sum_{m=1,M} w_m = 1; \qquad w_m \geq 0 \quad (m=1,2,\ldots,M) \tag{13}$$

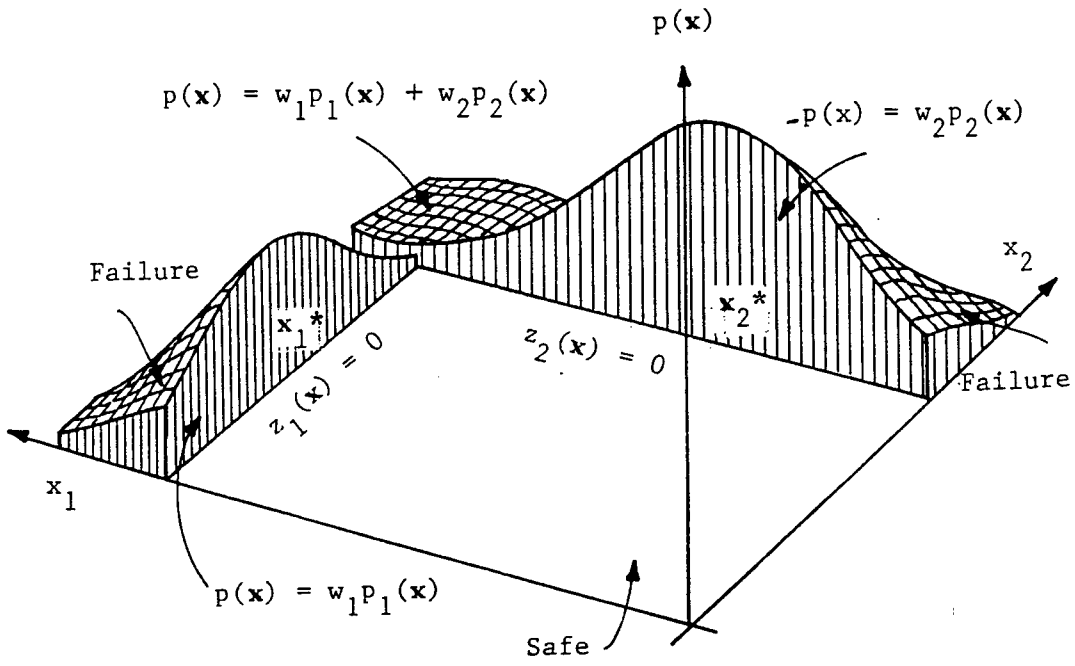
Computation of Eq. 1 using this sampling scheme is done in M groups of simulation. Each group generates correlated normal samples $\mathbf{x}_{k\_m}$ ($k\_m=1,2,\ldots,N_m$; $m=1,2,\ldots,M$) from $p_m(\mathbf{x})$, i.e.

$$P_f = \sum_{m=1,M} w_m \left\{ \int_{\mathbf{x}} \frac{G(\mathbf{x})f(\mathbf{x})}{p(\mathbf{x})} p_m(\mathbf{x})d\mathbf{x} \right\} \tag{14}$$

$$\approx \sum_{m=1,M} \frac{w_m}{N_m} \left\{ \sum_{k\_m=1,Nm} \frac{G(\mathbf{x}_{k\_m})f(\mathbf{x}_{k\_m})}{p(\mathbf{x}_{k\_m})} \right\} = P_f'$$

where P_f' denotes the estimator for P_f. The computation in parentheses is one of the M groups of simulation, corresponding to mode m with samples $\mathbf{x}_{k\_m}$ ($k\_m=1,2,\ldots,N_m$) from subdensity $p_m(\mathbf{x})$. To reflect relative importance of respective failure modes, the sample size for mode m, N_m, is determined by

Figure 1. Sampling distribution for the first-order system reliability problems.



$$N_m \approx w_m N \quad (m=1,2,\ldots,M) \quad \text{and} \quad \sum_{m=1,M} N_m \approx N \tag{15}$$

with N_m (m=1,2,...,M) being integers. This proposed distribution $p(\mathbf{x})$ for a series system of two modes (M=2) with two variables (n=2) is shown in Fig. 1.

Note that the effort of solving the M simultaneous linear equations in Eq. 13 is insignificant, as M is usually small when only significant failure modes are included. Further note that solution to Eq. 13 may not exist, when some of the maximum likelihood points are too close to one another in the space of $\mathbf{x}$ (for example, w_m may be found to be negative). These close points (and corresponding failure modes) should then be covered by a single subdensity. In that case, the problem is treated as if it had correspondingly fewer failure modes in order to determine the weights for the respective modes.

Random variable samples $\mathbf{x}_{k\_m}$ from the halfspaces defined by Eq. 12 can be generated by rejection-and-transformation. This method requires evaluation of $g_m(\mathbf{x})$. It generates a typical sample in the complete space, accepts it as a valid sample from $p_m(\mathbf{x})$ if $\mathbf{x}_{k\_m} \in \underline{F}_m$ ($g_m(\mathbf{x})<0$) or transforms it into $\underline{F}_m$ if $\mathbf{x}_{k\_m} \notin \underline{F}_m$ ($g_m(\mathbf{x})<0$). When the evaluations of $g_m(\mathbf{x})$ are excessively costly, they may be avoided using a direct method given in Appendix A.

4. VARIANCE REDUCTION

4.1 Upper Bound for Variance

The variance of estimator by the importance sampling method in Eq. 9b is bounded by [Fu 1988]

$$\mathrm{Var}[P_f'] \le \mathrm{Var}[P_f']_{max} \underset{N}{=} -\{P_f[\frac{f(\mathbf{x})}{p(\mathbf{x})}]_{max} - P_f^2\} \tag{16}$$

This upper bound for the present method is found to be

$$\mathrm{Var}[P_f']_{max} \underset{N}{=} -\{P_f \frac{f(\mathbf{x}^*)}{p(\mathbf{x}^*)} - P_f^2\} \tag{17}$$

using

$$[\frac{f(\mathbf{x})}{p(\mathbf{x})}]_{max} = \frac{f(\mathbf{x}^*)}{p(\mathbf{x}^*)} \tag{18}$$

where $\mathbf{x}^*$ stands for any one of the design points $\mathbf{x}_m^*$ $(m=1,2,\ldots,M)$. Eq. 18 is proved by inspecting the gradient of likelihood ratio $\nabla f(\mathbf{x})/p(\mathbf{x})$ as follows. Consider each domain $\underline{F}_a$ where $f(\mathbf{x})/p(\mathbf{x})$ is continuous and M' failure modes are active ($\mathbf{x} \in \underline{F}_a = \underline{F}_1 \cap \underline{F}_2 \cap \ldots \cap \underline{F}_{M'-1} \cap \underline{F}_{M'}$; $M' \le M$):

$$\nabla f(\mathbf{x})/p(\mathbf{x})\big|_{\mathbf{x}\in\underline{F}_a} = \frac{f(\mathbf{x})}{p^2(\mathbf{x})} \sum_{m=1,M'} w_m p_m(\mathbf{x}) \mathbf{C}^{-1}(\underline{\mathbf{X}}-\mathbf{x}_m^*)\big|_{\mathbf{x}\in\underline{F}_a} = \sum_{m=1,M'} d_m(\mathbf{x}) \mathbf{a}_m\big|_{\mathbf{x}\in\underline{F}_a} \tag{19a}$$

Appendix B proves that

$$d_m(\mathbf{x})\big|_{\mathbf{x}\in\underline{F}_a} > 0; \quad (m=1,2,\ldots,M') \tag{19b}$$

and Appendix C further proves that $\nabla f(\mathbf{x})/p(\mathbf{x})$ never becomes zero within the individual domains. It follows that the extreme values of $f(\mathbf{x})/p(\mathbf{x})$ do not occur within these domains but on their boundaries.

Therefore, the global maximum of $f(\mathbf{x})/p(\mathbf{x})$ can be found by comparing its values on the boundaries. For any two adjacent domains $\underline{F}_a = \underline{F}_1 \cap \underline{F}_2 \cap \ldots \cap \underline{F}_{M'-1} \cap \underline{F}_{M'}$ and $\underline{F}_b = \underline{F}_1 \cap \underline{F}_2 \cap \ldots \cap \underline{F}_{M'-1}$ where $f(\mathbf{x})/p(\mathbf{x})$ is respectively continuous, their common boundary is $\mathbf{a}_{M'}{}^t(\mathbf{x} - \mathbf{x}_{M'}{}^*)=0$. For $\mathbf{x}$ in a vicinity arbitrarily close to this boundary respectively from both sides ($\mathbf{a}_{M'}{}^t(\mathbf{x}-\mathbf{x}_{M'}{}^*)<0$ and $\mathbf{a}_{M'}{}^t(\mathbf{x} - \mathbf{x}_{M'}{}^*)>0$), one has

$$\lim_{(\mathbf{a}_{M'}{}^t(\mathbf{x}-\mathbf{x}_{M'}{}^*)\to 0} \frac{f(\mathbf{x})}{p(\mathbf{x})}\Big|_{\mathbf{x}\in\underline{F}_a} < \lim_{\mathbf{a}_{M'}{}^t(\mathbf{x}-\mathbf{x}_{M'}{}^*)\to 0} \frac{f(\mathbf{x})}{p(\mathbf{x})}\Big|_{\mathbf{x}\in\underline{F}_b} \tag{20a}$$

simply because the denominator on the left side is larger:

$$\lim_{\mathbf{a}_{M'}{}^t(\mathbf{x}-\mathbf{x}_{M'}{}^*)\to 0} p(\mathbf{x})\,|\,\mathbf{x}\varepsilon\underline{F}_a = \frac{\sum_{m=1,M'} w_m\, p_m(\mathbf{x})}{\mathbf{a}_{M'}{}^t(\mathbf{x}-\mathbf{x}_{M'}{}^*)\to 0}$$

$$> \lim_{\mathbf{a}_{M'}{}^t(\mathbf{x}-\mathbf{x}_{M'}{}^*)\to 0} p(\mathbf{x})\,|\,\mathbf{x}\varepsilon\underline{F}_b = \frac{\sum_{m=1,M'-1} w_m\, p_m(\mathbf{x})}{\mathbf{a}_{M'}{}^t(\mathbf{x}-\mathbf{x}_{M'}{}^*)\to 0} \tag{20b}$$

Generality of Eq. 20 is not restricted, since ordering of the failure mode subscripts in the definitions of $\underline{F}_a$ and $\underline{F}_b$ is arbitrary. By deduction, the global maximum value of $f(\mathbf{x})/p(\mathbf{x})$ must occur on the boundaries of domains where only one failure mode, say m, is active and $p(\mathbf{x}) = w_m p_m(\mathbf{x})$. In this case, $\nabla f(\mathbf{x})/p(\mathbf{x})$ has the same direction as $\mathbf{a}_m$ (Eq. 19), and it follows that the maximum point is on $\mathbf{a}_m{}^t(\mathbf{x}-\mathbf{x}_m{}^*) = 0$, namely $\mathbf{x}=\mathbf{x}_m{}^*$. This proves Eq. 18.

4.2 Variance Reduction Compared to the Monte Carlo Method

To evaluate the variance reduction by the importance sampling method, the ratio of estimator variances of the two methods is used as a variance change factor (VCF):

$$\mathrm{VCF} = \frac{\mathrm{Var}[P_f']_{IS}}{\mathrm{Var}[P_f']_{MC}} = \frac{\frac{1}{P_f}\int_{\mathbf{x}} \frac{G^2(\mathbf{x})f(\mathbf{x})}{p(\mathbf{x})} f(\mathbf{x})d\mathbf{x} - P_f}{1-P_f} \tag{21}$$

where subscripts IS and MC indicate the respective methods. The estimator variance by the conventional Monte Carlo method is readily obtained by setting $p(\mathbf{x}) = f(\mathbf{x})$ in Eq. 9b. VCF provides a means of evaluating the importance sampling method with respect to variance reduction. This single factor also represents changes of several governing factors of simulation discussed later.

Using the central limit theorem, three important factors in simulation (confidence index k, error ε', and sample size N) are related as follows [Shooman 1968, Melchers 1984, Fu 1988]:

$$\mathrm{Prob}\left[\frac{\varepsilon'}{\mathrm{COV}} \le k\right] = P \tag{22a}$$

where

$$\varepsilon' = \left|\frac{P_f'}{P_f} - 1\right| \tag{22b}$$

$$\mathrm{COV} = \frac{\mathrm{Var}^{1/2}[P_f']}{E[P_f']} \tag{22c}$$

and P is the confidence level in probability and k is the associated confidence index according to the standard normal cumulative function $\Phi()$:

$$k = -\Phi^{-1}[0.5(1-P)] \tag{22d}$$

An acceptable error ε (associated with the confidence level given by k) is defined by

$$\varepsilon = k\ COV \tag{23}$$

The changes in efficiency, confidence, and accuracy associated with N, k, and ε can be given in terms of VCF in simple manners, respectively:

$$\varepsilon_{IS}/\varepsilon_{MC}=VCF^{1/2}; \quad k_{IS}/k_{MC}=VCF^{-1/2}; \quad N_{IS}/N_{MC}=VCF \tag{24}$$

where subscripts IS and MC indicate the importance sampling and the conventional Monte Carlo methods, respectively. Each of these relationships can be readily derived by separately keeping the other two factors unchanged in the two simulation methods.

4.3 Variance Reduction of the Proposed Method

For the present method, substitute Eq. 18 into Eq. 21, and one has

$$VCF_{max} = \frac{\dfrac{f(\mathbf{x}^*)}{p(\mathbf{x}^*)} - P_f}{1-P_f} \tag{25}$$

Similarly,

$$COV_{max} = \left\{-\left[\frac{1}{N}\frac{f(\mathbf{x}^*)}{p(\mathbf{x}^*)P_f} - 1\right]\right\}^{1/2} \quad \text{and} \quad \varepsilon_{max} = k\ COV_{max} \tag{26}$$

The ratios of error level, confidence index, and sample size by the importance sampling and the conventional Monte Carlo methods are accordingly bounded by:

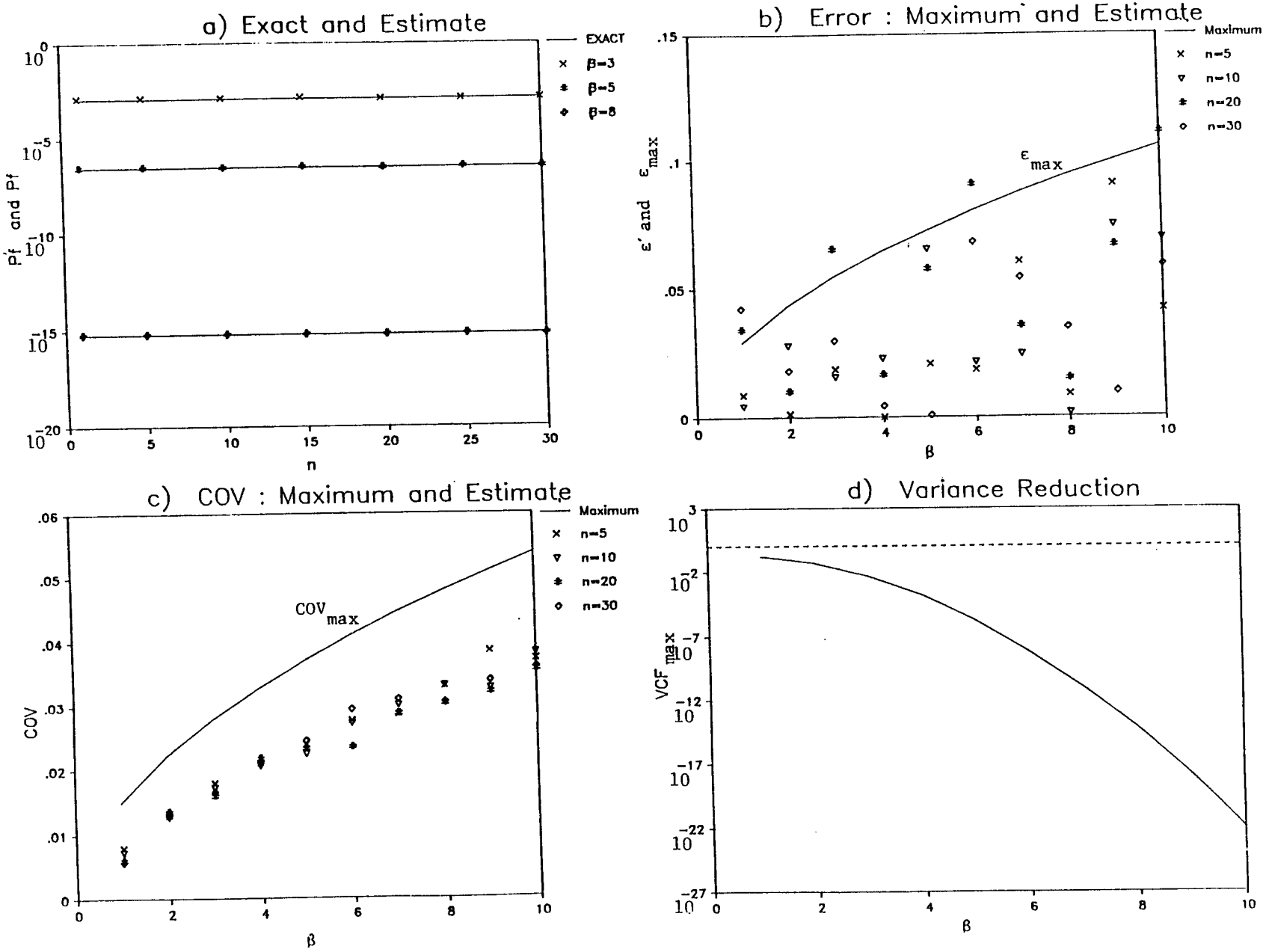
$$\varepsilon_{IS}/\varepsilon_{MC}\le VCF_{max}^{1/2}; \quad k_{IS}/k_{MC}\ge VCF_{max}^{-1/2}; \quad N_{IS}/N_{MC}\le VCF_{max} \tag{27}$$

These relations can be used to determine the sample size for given requirements of confidence level and associated error. This can be done even before the simulation if the range of P_f is estimated in Eqs. 25 to 27.

5. APPLICATION EXAMPLES

Simulation approaches generally use pseudo-random number generators to produce random samples, whose quality is critically important. It has been found that

Figure 2. Results of Example 1.



some congruential generators produce random numbers on certain hyperplanes [Marsaglia 1968]. Thus they are not suitable for multidimensional integrations. The Generalized Feedback Shift Register generator was developed to avoid this problem [Ripley 1987]. A generator of this kind [Fushimi and Tezuka 1983] is suggested for high dimensional problems, and it has been implemented on an IBM PS/2 Model 80 for computation of such cases as those now presented.

5.1 Example 1

This example is given by Engelund and Rackwitz [1992] in their comparison of importance sampling methods. The limit state function z_1 is defined as

$$z_1(\mathbf{x}) = \beta n^{1/2} - \sum_{i=1,n} x_i \tag{28}$$

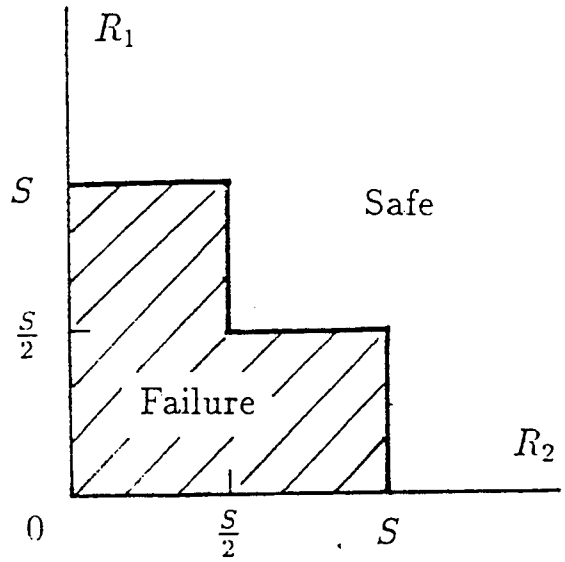
where X_i are independent standard normal variables ($\underline{X}$=0, C=I), and n and ß are constants. This problem can be used to test the random number generator's quality, especially in higher dimensional cases. It is also a relatively simple problem with a single failure mode (M=1), as a first step of illustration.

Fig. 2 shows the results by the present method for ß=1,2,...,10 and n=1,5,10,..., 30, respectively, using N=4,000 samples. Fig. 2a provides an overview comparing the estimates and exact values. Fig. 2b exhibits the involved errors in more detail, contrasted with ε_{max} by Eq. 26 and associated with a confidence level of 95% (k=1.96). The actual errors do not exceed this maximum level in all but a few cases. The actual error increases insignificantly with ß (or equivalently, with decrease of P_f). Fig. 2c shows similar comparison for COV vs. COV_{max}. The solid line is obtained by Eq. 26. Estimated COVs by Eq. 22c are marked points and they are bounded by COV_{max}. This also indicates the satisfactory quality of the pseudo-random number generator employed. Fig. 2d shows decrease of VCF_{max} with ß over a range of practical applications, which indicates that simulation efficiency and accuracy do not decrease with P_f to be estimated. As a matter of fact, Fig. 2c shows a small variation range of COV (0 to 6%) over a wide range of ß (1 to 10), using the same practically affordable sample size. These cases also show that the upper bounds for error and COV are fairly close to those estimated by simulation. Thus these bounds can be used for sample size determination without significant overestimation. This will be further demonstrated by the rest of the examples.

5.2 Example 2

Consider a series system problem with $z_1(\mathbf{x}) = x_1 - x_2$ and $z_2(\mathbf{x}) = 61 - 1.44x_1 - x_2$ (M=2) and independent variables X_1 = N(25,2.5) and X_2 = N(10,3.0). The modal failure probabilities and weights are respectively found to be $P_{f,1} = 0.00006126$, $P_{f,2} = 0.0006850$, $w_1 = 0.0953$, and $w_2 = 0.9047$, using $\mathbf{x}_1^{*t} = (18.85, 18.85)$ and $\mathbf{x}_2^{*t} = (31.15, 16.15)$. The exact P_f is computed by $P_f = P_{f,1} + P_{f,2} - P_{f,1} * P_{f,2} = 0.0007462$, since the two modes are independent of one another.

Figure 3. Example 3: failure domain of a series system of parallel systems.



The simulation is performed using N = 4,000, and a confidence level of 95% (k = 1.96) is used to obtain ε_{max} = 5.7% by Eq. 26, with COV_{max} = 2.9%. P_f' is equal to 0.0007522 with an error ε' = 0.8% and COV = 1.9%. Compared with the conventional Monte Carlo method, sample size is at least reduced by a factor VCF_{max} = 0.002536, without sacrifice in accuracy (ε') or confidence (k). In other words, almost 400 times more samples (about 1.7 million) would have to be used if the conventional Monte Carlo method were employed.

5.3 Example 3

This is a problem with combination of series and parallel systems. It is treated as a series of parallel systems as defined earlier. The structural system consists of a deterministic load S and two parallel bars made of a brittle material, which results in the failure region shown in Fig. 3 [Moses 1982]. The two nondifferentiable failure surfaces represent two symmetrical sequences of component failures that lead to system failure. They are equally important and thus equally weighted: $w_1 = w_2 = 0.5$. Axial strengths of bars R_1 and R_2 are assumed normally distributed with a correlation coefficient ρ and equal coefficients of variation of 20%. Their mean values are given by a safety factor SF: $\underline{R}_1 = \underline{R}_2$ = S*SF/2.

For SF = 1.7, 2.2, and 2.7, the results by the present method using N = 4,000 are plotted in Fig. 4. Exact P_f for comparison is obtained using an integration table [National Bureau of Standards 1959]. A good agreement between the two methods is observed in Fig. 4a. Error ε' of these cases is shown in Fig. 4b, most being within the range of 2.5% with a maximum of 5.4%. ε_{max} is based on Eq. 26 and a confidence level of 95% (k = 1.96). Fig. 4c shows COV within 3.7%. Note that they are not significantly dependent on SF (or equivalently P_f) as the

Figure 4. Results of Example 3.

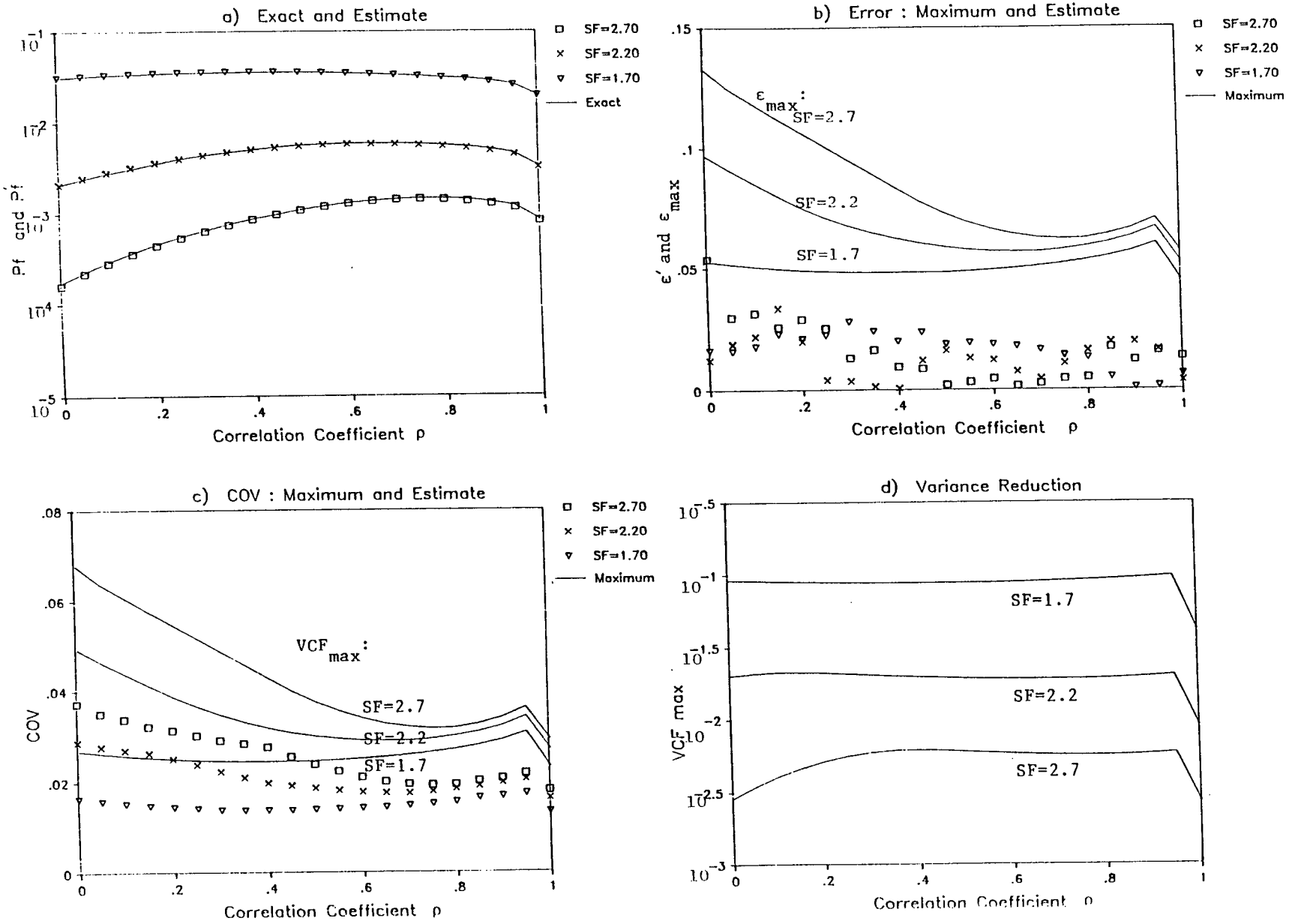
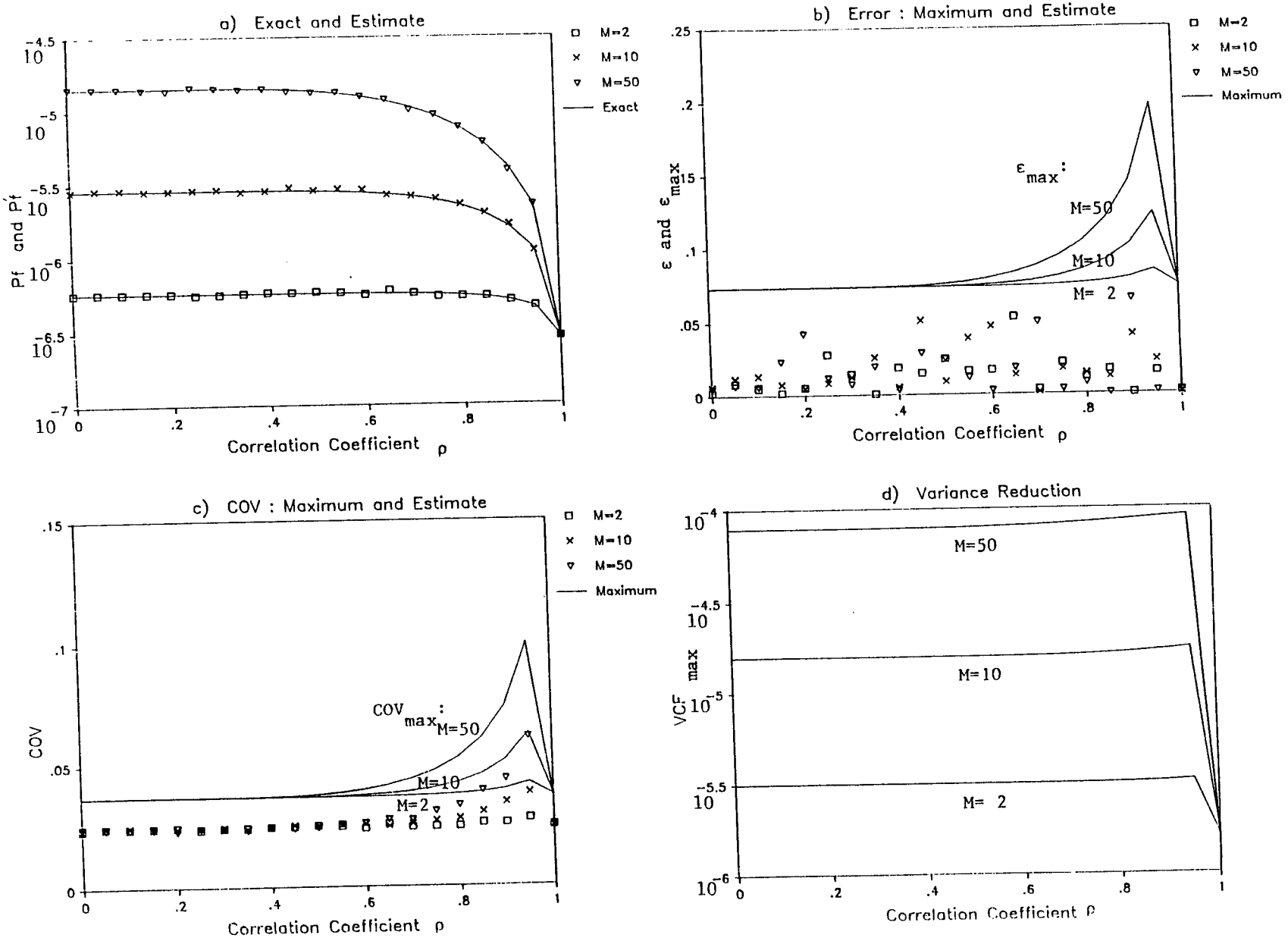


Figure 5. Results of Example 5.



conventional Monte Carlo method would be, using the same sample size for all the cases. This indicates that the standard deviation of the estimates is reduced almost proportionally to P_f. VCF_{max} in Fig. 4d demonstrates again its decrease with P_f.

At ρ=1, the subdensities of simulation become identical due to full modal correlation. This fact causes drop of $f(\mathbf{x}^*)/p(\mathbf{x}^*)$ and, in turn, drop of ε_{max}, COV_{max}, and VCF_{max} in Figs. 4b, 4c, and 4d.

5.4 Example 4

This example shows an application to a nonlinear problem with linearized failure surfaces. The original failure function is $z(\mathbf{x}) = 6 + x_1 - 0.622\ x_2^2 = 0$ with $\mathbf{X}^t = (X_1, X_2)$ and independent $X_1 = X_2 = N(0,1)$. Linearized failure functions are formed according to Eq. 10, using $\mathbf{x}_1^{*t} = (-0.804, 2.890)$ and $\mathbf{x}_2^{*t} = (-0.804, -2.890)$ and $\mathbf{a}_1^t = (1, -3.596)$ and $\mathbf{a}_2^t = (1, 3.596)$. The symmetric failure modes give $w_1 = w_2 = 0.5$. Exact $P_f = 0.002816$ is obtained by a one-dimensional integration using the Gamma distribution [Schuëller and Stix 1987].

Simulation results using N = 4,000 are: $VCF_{max} = 0.008316$, $P_f' = 0.002761$, $\varepsilon' = 2.0\%$ vs. $\varepsilon_{max} = 5.3\%$, and COV = 1.7% vs. $COV_{max} = 2.7\%$. Apparently the linearization represents a good approximation. P_f' by the present method based on the linearization is reasonably close to the exact P_f. The estimated COV and ε' are bounded by the analytically derived upper bounds, respectively. Maximum error ε_{max} is based on a confidence level of 95% (k = 1.96).

5.5 Example 5

This example exhibits applications to series systems of higher dimensions. A system of M components in series is considered, with the components designed to be equally reliable with a component reliability index $\beta = 5$ (equivalently a component failure probability of 0.287×10^{-6}) and equal weights $w_m = 1/M$ (m=1,2,...,M). For a deterministic load S and normally distributed component strengths R_m (m=1,2,...,M) with an equal correlation coefficient ρ among them, P_f' is plotted in Fig. 5, using N = 4,000 for M = 2, 10, and 50. Exact P_f is calculated for comparison by a one-dimensional integration [Grigoriu and Turkstra 1979] and shown in Fig. 5a. Error ε' is displayed in Fig. 5b, most being within the range of 5% with a maximum of 6.4%. COV of the estimates is shown in Fig. 5c, being around 2.5% and again almost independent of P_f. VCF_{max} in Fig. 5d has similar behavior to that in Fig. 5d for Example 3, within the range of 6.1%. The drop of ε_{max}, COV_{max}, and VCF_{max} at ρ=1 in Figs. 5b, 5c, and 5d is due to the same cause discussed in Example 3.

6. CONCLUSIONS

A general importance sampling method has been introduced. Its sampling distribution is proportional to the ideally optimal one at the maximum likelihood

points and can locate all samples in the failure domain. This method can be employed in integrals for extremely small probability when assessing system reliability, expected damage, etc. Upper bounds for variance, error, and COV of estimator are derived analytically, which have been used to evaluate the present method and can be used in general applications to determine required sample size for an acceptable maximum error associated with a confidence level. Example applications show the assured improvement on efficiency and accuracy for general cases, compared to the conventional Monte Carlo method. These include problems of both series and parallel systems and of higher dimensional space. They also show that the required sample size is not highly dependent on the failure probability P_f to be estimated.

7. ACKNOWLEDGMENTS

The Federal Highway Administration, U.S. Department of Transportation, provided partial support. Dr. J. Tang, E.W. Dillon, and I.A. Aziz of the New York State Department of Transportation assisted in partial preparation of the examples and figures.

8. REFERENCES

[1] P.Bjerager, Probability Integration by Directional Simulation, ASCE J.Eng. Mech., Vol.114, No.8 (Aug.1988), pp.1285-1302.

[2] S.Engelund and R.Rackwitz, Comparison of Some Importance Sampling Techniques in Structural Reliability, in Y.K.Lin (Ed.), Probabilistic Mechanics and Structural and Geotechnical Reliability (Proc. ASCE Specialty Conference, Denver, CO, USA, July 8-10, 1992) pp.108-111.

[3] G.Fu, Error Analysis for Importance Sampling Method, in D.M.Frangopol (Ed.), New Directions in Structural System Reliability (Proc. NSF Workshop on Research Needs for Applications of System Reliability Concepts and Techniques in Structural Analysis, Design and Optimization, Boulder, CO, Sept.12-14, 1988), pp.132-146.

[4] G.Fu and F.Moses, Application of Lifetime System Reliability, Preprint 52-1, ASCE Structures Congress '86, New Orleans, LA, Sept.15-18, 1986.

[5] G.Fu and F.Moses, A Sampling Distribution for System Reliability Assessment, Proc. IFIP 1st Working Conference on Reliability and Optimization of Structural Systems, Aalborg, Denmark, May 6-8, 1987 (1987a), pp.141-155.

[6] G.Fu and F.Moses, Lifetime System Reliability Models with Application to Highway Bridges, in N.C.Lind (Ed.), Reliability and Risk Analysis in Civil Engineering, Proc.ICASP5, Vancouver, Canada, May 25-29, 1987 (1987b), pp. 71-78.

References

[7] G.Fu and F.Moses, Importance Sampling in Structural System Reliability, in P.D.Spanos (Ed.), Probabilistic Methods in Civil Engineering, 5th ASCE Specialty Conference, Blacksburg, VA, May 1988, p.340-343.

[8] M.Fushimi and S.Tezuka, The k-Distribution of Generalized Feedback Shift Register Pseudorandom Numbers, Communications of the ACM, Vol.26, No.7 (July 1983), pp.516-523.

[9] M.Grigoriu and C.Turkstra, Safety of Structural Systems with Correlated Resistances, Appl. Math. Modelg., Vol.3 (1979), pp.130-136.

[10] A.M.Hasofer and N.C.Lind, An Exact and Invariant First Order Reliability Format, ASCE J.Eng.Mech., 1974, pp.111-121.

[11] M.Hohenbichler and R.Rackwitz, Non-Normal Dependent Vectors in Structural Safety, ASCE J.Eng.Mech., Vol.107, No.6, (Dec.1981), pp.1227-1238.

[12] H.Kahn, Use of Different Monte Carlo Sampling Techniques, in H.A.Meyer (Ed.), Symposium on Monte Carlo Methods, John Wiley & Sons, Inc., 1956) pp.146-190.

[13] J.P.C.Kleijnen, Statistical Techniques in Simulation: Parts I and II, Marcell Dekker, Inc., 1974.

[14] G.Marsaglia, Random Numbers Fall Mainly in the Planes, Proc. Nat. Acad. Sci. USA 61, 1968, pp.25-28.

[15] R.E.Melchers, Efficient Monte-Carlo Probability Integration, Civil Engineering Research Report No.7/1984, Monash University, 1984.

[16] R.E.Melchers, Structural Reliability: Analysis and Prediction, Ellis Horwod/Wiley, Chichester, 1987.

[17] R.E.Melchers, Radial Importance Sampling for Structural Reliability, ASCE J.Eng.Mech., Vol.116, No.1 (Jan. 1990a), pp.189-203.

[18] R.E.Melchers, Search-Based Importance Sampling, Structural Safety, Vol.9, (1990b), pp.117-128.

[19] F.Moses, System Reliability Developments in Structural Engineering," Structural Safety, Vol.1 (1982), pp.3-13.

[20] National Bureau of Standards, Tables of the Bivariate Normal Distribution Function and Related Functions, Appl.Math.Series 50, June 1959.

[21] B.D.Ripley, Stochastic Simulation, John Wiley & Sons, New York, 1987.

[22] G.I.Schuëller and R.Stix, A Critical Appraisal of Methods to Determine Failure Probabilities, Structural Safety, Vol.4 (1987), pp.293-309.

[23] M.Shinozuka, Basic Analysis of Structural Safety, ASCE J.Struc.Eng., Vol.109, No.3 (March 1983), pp.721-740.

[24] M.L.Shooman, Probabilistic Reliability: An Engineering Approach, McGraw-Hill Book Company, 1968.

Appendix A. A Direct Method for Generating Random Samples in a Halfspace

1. Find a lower triangular matrix $\mathbf{\Delta}$ such that $\mathbf{\Delta}\mathbf{\Delta}^t = \mathbf{C}$, say, by Cholesky decomposition. ($\mathbf{\Delta}$ always exists since $\mathbf{C}$ is symmetric and positive definite.)

2. Assemble an orthogonal transformation matrix $\mathbf{A}_m$ so that $\mathbf{\Delta}^t\mathbf{a}_m$ is in the plane spanned by the first two axes of the new coordinate system:

   $\mathbf{A}_m\mathbf{\Delta}^t\mathbf{a}_m = \mathbf{a'}_m = (a'_{m,1}, a'_{m,2}, 0, 0, \ldots, 0)^t$

3. Generate a sample $\mathbf{y}_{k\_m}$ of standard independent normal vector in a halfspace with its first element in only half of its complete space, i.e., $\mathbf{y}_{k\_m} = (y_{k\_m,1}, y_{k\_m,2}, \ldots, y_{k\_m,n})^t$ and $y_{k\_m,1} \varepsilon (-\infty, 0]$

4. Transform $\mathbf{y}_{k\_m}$ into the new coordinate system defined by $\mathbf{A}_m$: $\mathbf{y}_{k\_m}' = \mathbf{A}_m\mathbf{y}_{k\_m}$

5. Form $\mathbf{y}_{k\_m}''$ by rotating the projection of $\mathbf{y}_{k\_m}'$ on the plane defined by the first two axes, by the angle α between $\mathbf{a'}_m$ and the first axis: $\mathbf{y}_{k\_m}'' = \mathbf{D}_m\mathbf{y}_{k\_m}'$

6. Transform $\mathbf{y}_{k\_m}''$ back to the original space: $\mathbf{y}_{k\_m}''' = \mathbf{A}_m^{-1}\mathbf{y}_{k\_m}''$. Note that $\mathbf{y}_{k\_m}'''$ is a sample within the halfspace now ($\mathbf{a}_m^t\mathbf{\Delta}\mathbf{y}_{k\_m}''' \leq 0$)

7. Transform $\mathbf{y}_{k\_m}'''$ to correlated $\mathbf{x}_{k\_m}$ using mean $\mathbf{x}_m^*$ and covariance matrix $\mathbf{C}$: $\mathbf{x}_{k\_m} = \mathbf{\Delta}\mathbf{y}_{k\_m}''' + \mathbf{x}_m^*$

8. Go to Step 3 until $k\_m = N_m$

9. Go to Step 2 until $m = M$

Steps 3 to 7 are shown in Fig. 6 in the plane defined by the first two axes of the variable vector (Step 2). Note that Step 5 involves only the first two elements of $\mathbf{y}_{k\_m}'$, since others are orthogonal to this plane.

A numerical example is given below for illustration. For simplicity, $M = 1$, $n = 3$, $\mathbf{C} = \mathbf{I}$, $\mathbf{x}_1^* = \mathbf{0}$, and $\mathbf{a}_1 = (0.267, 0.535, 0.802)^t$. Steps 8 and 9 are omitted below because they simply indicate recurrence steps for more samples.

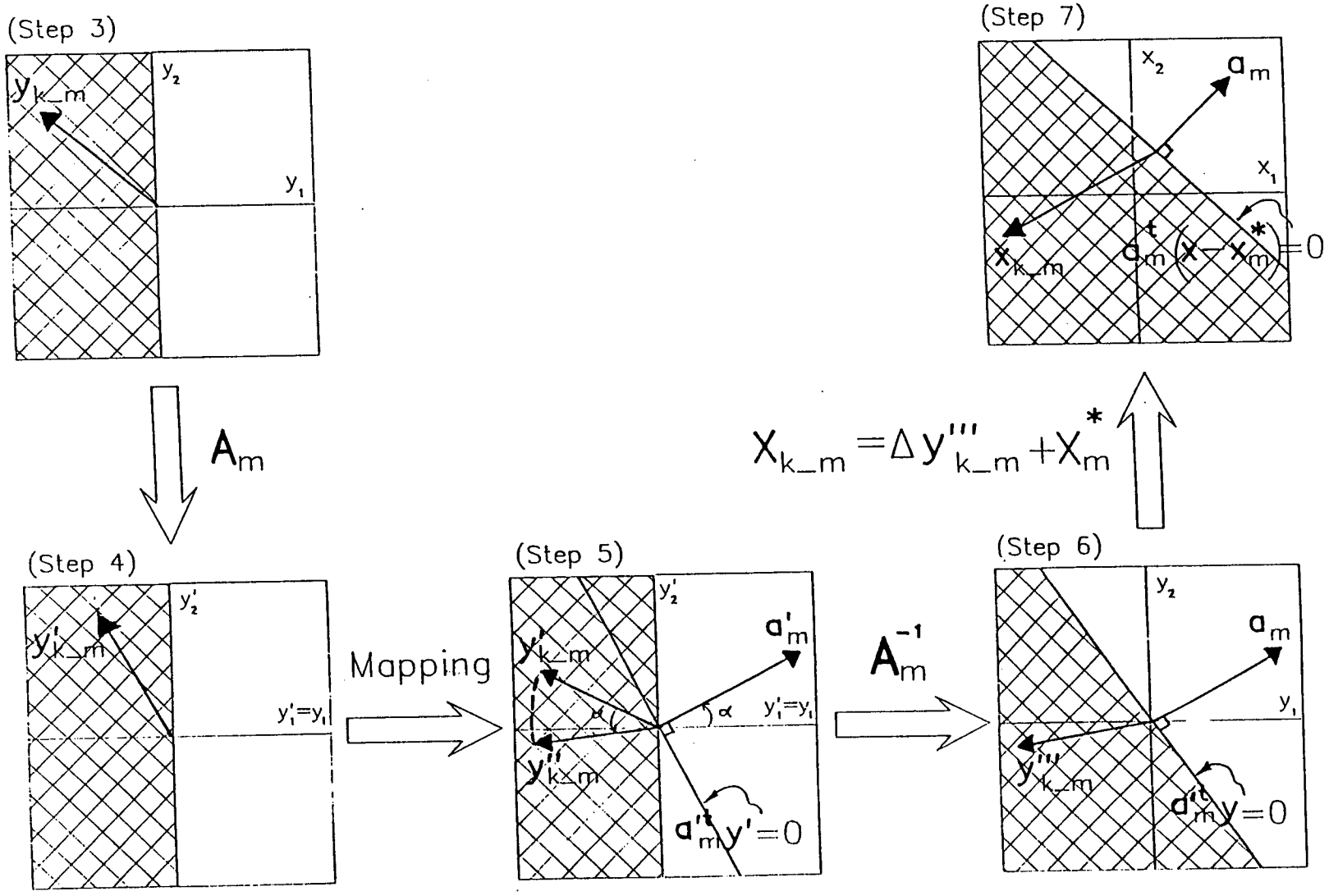
1. Triangular matrix $\mathbf{\Delta} = \begin{bmatrix} 1 & & \\ 0 & 1 & \\ 0 & 0 & 1 \end{bmatrix}$

2. Orthogonal transformation matrix $\mathbf{A}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\gamma & \sin\gamma \\ 0 & -\sin\gamma & \cos\gamma \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.555 & 0.832 \\ 0 & -0.832 & 0.555 \end{bmatrix}$

   where $\tan\gamma = 0.535/0.267 = 1.5$ and $\gamma = 0.982$. Note: $\mathbf{A}_1\mathbf{\Delta}^t\mathbf{a}_1 = \mathbf{a'}_1 = (0.267, 0.964, 0)^t$

3. $\mathbf{y}_0 = (-1.135, 2.014, 0.033)^t$

Figure 6. Generation of x_{k_m} in Failure Domain F_m.



4. $\mathbf{y}_0' = \mathbf{A}_1\mathbf{y}_0 = (-1.135, 1.145, -1.657)^t$

5. In-plane rotating matrix $\mathbf{D}_1 = \begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0.267 & -0.964 & 0 \\ 0.964 & 0.267 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

   where $\cos\alpha = 0.267$ and $\alpha = 1.300$. $\mathbf{y}_0'' = \mathbf{D}_1\mathbf{y}_0' = (-1.406, -0.788, -1.657)^t$

6. $\mathbf{y}_0''' = \mathbf{A}_1^{-1}\mathbf{y}_0'' = (-1.406, 0.942, -1.575)$. Note: $\mathbf{a}_1^t\Delta\mathbf{y}_0''' = -1.135 \leq 0$

7. $\mathbf{x}_0 = \Delta\mathbf{y}_0''' + \mathbf{x}_1^* = \mathbf{y}_0''' = (-1.406, 0.942, -1.575)^t$

Appendix B. Proof of $d_m(\mathbf{x})>0$ $(j=1,2,\ldots,m)$
where $\sum_{m=1,M'} d_m(\mathbf{x})\mathbf{a}_m = f(\mathbf{x})/p^2(\mathbf{x})\sum_{m=1,M'} w_m p_m(\mathbf{x})\mathbf{C}^{-1}(\underline{\mathbf{X}}-\mathbf{x}_m^*)$

For the optimization problem defined as

$$\text{Max } f(\mathbf{x}); \text{ Sub } \mathbf{a}_m^t(\mathbf{x}-\mathbf{x}_m^*) = 0 \tag{B1}$$

and for $\mathbf{x}_m^*$ to be the maximum likelihood point (or design point), the necessary condition is

$$\nabla f(\mathbf{x}) + \lambda_m \nabla[\mathbf{a}_m^t(\mathbf{x}-\mathbf{x}_m^*)] = \mathbf{0} \text{ at } \mathbf{x}=\mathbf{x}_m^* \tag{B2}$$

Eq. B2 leads to

$$f(\mathbf{x}_m^*)\mathbf{C}^{-1}(\underline{\mathbf{X}} - \mathbf{x}_m^*) + \lambda_m \mathbf{a}_m = \mathbf{0} \tag{B3}$$

where λ_m $(m=1,2,\ldots,M')$ are the Lagrange multipliers:

$$\lambda_m = -\frac{\mathbf{a}_m^t(\underline{\mathbf{X}}-\mathbf{x}_m^*)f(\mathbf{x}_m^*)}{\mathbf{a}_m^t\mathbf{C}\mathbf{a}_m} \tag{B4}$$

Substitute Eq. B3 into the definition

$$\frac{f(\mathbf{x})}{p^2(\mathbf{x})}\sum_{m=1,M'} w_m p_m(\mathbf{x})\mathbf{C}^{-1}(\underline{\mathbf{X}}-\mathbf{x}_m^*) = \sum_{m=1,M'} d_m(\mathbf{x})\mathbf{a}_m \tag{B5}$$

and use Eq. B4 subsequently, one has

$$\begin{aligned} d_m(\mathbf{x}) &= -\frac{f(\mathbf{x})}{p^2(\mathbf{x})}\frac{w_m p_m(\mathbf{x})}{f(\mathbf{x}_m^*)}\lambda_m \\ &= \frac{f(\mathbf{x})}{p^2(\mathbf{x})}\frac{w_m p_m(\mathbf{x})}{\mathbf{a}_m^t\mathbf{C}\mathbf{a}_m}\mathbf{a}_m^t(\underline{\mathbf{X}}-\mathbf{x}_m^*) \qquad (m=1,2,\ldots,M') \end{aligned} \tag{B6}$$

Since $\mathbf{a}_m^t(\underline{\mathbf{X}}-\mathbf{x}_m^*)>0$ (by assuming $\underline{\mathbf{X}}\varepsilon\underline{S}$ because $P_f<1$), $\mathbf{a}_m^t\mathbf{C}\mathbf{a}_m>0$ (because $\mathbf{C}$ is symmetric and positive definite), and the rest of terms in Eq. B6 are all positive by definition, $d_m(\mathbf{x})$ are positive scalar functions of $\mathbf{x}$. This completes the proof.

Appendix C. Proof of $\nabla f(\mathbf{x})/p(\mathbf{x}) = \sum_{m=1,M'} d_m(\mathbf{x})\mathbf{a}_m \neq \mathbf{0}$, with $d_m(\mathbf{x})>0$ $(m=1,2,\ldots,M')$ proved in Appendix B.

For $M'=1$, since $d_1(\mathbf{x})>0$ and $\mathbf{a}_1\neq\mathbf{0}$, then $d_1(\mathbf{x})\mathbf{a}_1\neq\mathbf{0}$.
For $M'>1$, use contradiction. Assume $\sum_{m=1,M'} d_m(\mathbf{x})\mathbf{a}_m=\mathbf{0}$. For a point $\mathbf{x}_0\varepsilon\underline{F}_1\cap\underline{F}_2\cap\ldots\cap\underline{F}_{M'-1}\cap\underline{F}_{M'}$ and another arbitrary point $\mathbf{x}_1\varepsilon R^n$, we have

$$\sum_{m=1,M'} d_m(\mathbf{x})\mathbf{a}_m{}^t\mathbf{x}_0 = \sum_{m=1,M'} d_m(\mathbf{x})\mathbf{a}_m{}^t\mathbf{x}_1 = 0 \qquad (C1)$$

Add a common term $-\sum_{m=1,M'} d_m(\mathbf{x})\mathbf{a}_m{}^t\mathbf{x}_m{}^*$ to both sides of Eq. C1

$$\sum_{m=1,M'} d_m(\mathbf{x})\mathbf{a}_m{}^t(\mathbf{x}_0-\mathbf{x}_m{}^*) = \sum_{m=1,M'} d_m(\mathbf{x})\mathbf{a}_m{}^t(\mathbf{x}_1-\mathbf{x}_m{}^*) < 0 \qquad (C2)$$

since $d_m(\mathbf{x})\mathbf{a}_m{}^t(\mathbf{x}_0-\mathbf{x}_m{}^*)<0$ according to the definition of $\mathbf{x}_0$. This also indicates that $\mathbf{x}_1\varepsilon\underline{F}$, since it belongs to at least one of the subfailure region $\underline{F}_m$ $(m=1,2,\ldots,M')$. It follows that the entire space belongs to the failure domain, i.e. $R^n=\underline{F}$ and $P_f=1$, which contradicts with the basic assumption that $P_f<1$. The proof is completed.