ELSEVIER

PII: S0266-8920(98)00003-4

Prob. Engng. Mech. Vol. 13, No. 4, pp. 299–308, 1998

0266-8920/98 $19.00+0.00

Multinormal probability by sequential conditioned importance sampling: theory and application

R. Ambartzumian[a], A. Der Kiureghian[b,*], V. Ohanian[a] & H. Sukiasian[a]

[a]*Institute of Mathematics of the Academy of Sciences of Armenia, Yerevan, Armenia*
[b]*Department of Civil & Environmental Engineering, University of California, Berkeley, CA 94720, USA*

An efficient Monte Carlo simulation algorithm is developed for estimating the probability content of rectangular domains in the multinormal probability space. The algorithm makes use of the properties of the multinormal distribution, as well as the concept of importance sampling. Accurate estimates of the probability are obtained with a relatively small number of simulations, regardless of its magnitude. The algorithm also allows easy computation of the sensitivities of the probability with respect to distribution parameters or the boundaries of the domain. Application of the algorithm to structural system reliability is demonstrated through a simple example.

1 INTRODUCTION

Let $\mathbf{X} = (X_1, \cdots, X_n)$ be a vector of jointly normal random variables with mean vector $\mathbf{m} = (m_1, \cdots, m_n)$ and positive-definite covariance matrix $\mathbf{C}$ having the elements c_{ij}, $i, j = 1, 2, \cdots, n$. Recall that the density function $\varphi(x_1, x_2, \cdots, x_n)$ of $\mathbf{X}$ has the form

$$\varphi(x_1, x_2, \cdots, x_n) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \mathbf{C}}} \exp\left[-\tfrac{1}{2}(\mathbf{x}-\mathbf{m})^{\mathrm{T}}\mathbf{C}^{-1}(\mathbf{x}-\mathbf{m})\right] = \frac{1}{(2)^{n/2}\sqrt{\det \mathbf{C}}} \exp\left[-\tfrac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}(x_i - m_i)(x_j - m_j)\right] \tag{1}$$

where d_{ij} are the elements of the inverse matrix $\mathbf{D} = \mathbf{C}^{-1}$. By $P(Q)$, we denote the probability that $\mathbf{X}$ lies in the n-dimensional rectangular domain $Q = [a_1,b_1] \times [a_2,b_2] \times \cdots \times [a_n,b_n]$, i.e.

$$P(Q) = \int_{a_1}^{b_1}\int_{a_2}^{b_2}\cdots\int_{a_n}^{b_n} \varphi(x_1, \cdots, x_n)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n \tag{2}$$

The above probability is of interest in many applications. We note that for $a_1 = \cdots = a_n = -\infty$, $P(Q) = P(\cap_{i=1}^{n} X_i \leq b_i) = \Phi(b_1, \cdots, b_n)$ is the n-variate cumulative normal probability function. Johnson and Kotz[1] (Chapter 35) devoted an entire section of their book to reviewing the analytical and numerical methods for computing this probability. Unfortunately, efficient methods for an arbitrary covariance matrix and a large number of random variables are not available.

The multinormal probability is of special interest in the theory of structural system reliability. As shown by Hohenbichler and Rackwitz,[2] in a first-order approximation, the reliability of series and parallel structural systems can be reduced to the standard multinormal cumulative probability (with zero means and unit variances). In such applications, the correlation matrix is defined in terms of the unit normal vectors at the points of linearization of the component limit-state surfaces of the system. The coordinates b_i correspond to the reliability indices associated with the individual components. For general structural systems, the reliability can be obtained in terms of the probabilities of parallel subsystems representing cut sets or link sets of the system.[2] Because of these relations, there has been continued interest in the structural reliability community in methods for computing the multinormal probability integral. Several current reliability codes (e.g., PROBAN by Det Norske Veritas) make use of an approximate algorithm, which is based on a recursive formula derived by Hohenbichler and Rackwitz[2] and an asymptotic approximation developed by Gollwitzer and Rackwitz.[3] Other approximations have been proposed

*To whom correspondence should be addressed.

by Genz[4] and Joe.[5] Later in this paper, we present an example that illustrates the use of the multinormal probability in structural system reliability analysis.

In this paper, we present a new, efficient and accurate Monte Carlo simulation approach for computing the multinormal probability of rectangular domains for arbitrary n, $\mathbf{m}$ and $\mathbf{C}$. Use is made of the properties of the multinormal probability distribution and the concept of importance sampling.[6] It is shown that sufficiently accurate estimates of $P(Q)$ can be obtained with a small number of simulations, regardless of the magnitude of the probability of interest. The algorithm is also effective in computing the sensitivities of the probability with respect to distribution parameters, or the boundaries of the domain.

In the following, by $\varphi_k(x_k|x_1, \cdots, x_{k-1})$ we denote the one-dimensional conditional density function of X_k for given $X_1 = x_1, \cdots, X_{k-1} = x_{k-1}$. We denote the corresponding conditional cumulative distribution function by $\Phi_k(x_k|x_1, \cdots, x_{k-1})$. It is well known[1] that this distribution is normal with mean

$$m_k(x_1, \cdots, x_{k-1}) = m_k - \sum_{i=1}^{k-1} d_{ki} \frac{x_i - m_i}{d_{kk}} \tag{3}$$

and variance $\nu_k = 1/d_{kk}$. Furthermore, by $\varphi_{1,\cdots,k-1,k+1,\cdots,n}(x_1, \cdots, x_{k-1}, x_{k+1}, \cdots, x_n|x_k)$ we denote the $(n-1)$-dimensional conditional density of $X_1, \cdots, X_{k-1}, X_{k+1}, \cdots, X_n$ given $X_k = x_k$. It is well known that this distribution is also normal with mean vector $\mathbf{m}^{(k)} = \mathbf{m}_1 + \mathbf{C}_{12} c_{kk}^{-2}(x_k - m_k)$ and covariance matrix $\mathbf{C}^{(k)} = \mathbf{C}_{11} - \mathbf{C}_{12} c_{kk}^{-2} \mathbf{C}_{21}$, where $\mathbf{m}_1$ is obtained from $\mathbf{m}$ by eliminating the kth row, $\mathbf{C}_{11}$ is obtained from $\mathbf{C}$ by eliminating the kth row and column, and $\mathbf{C}_{12} = \mathbf{C}_{21}^{\mathrm{T}}$ is the kth column of $\mathbf{C}$ excluding the kth row.

For explanatory and comparison purposes, in the following section we describe the Sequential Conditioned Sampling (SCS) algorithm. Our main algorithm, which we describe in the subsequent section, is a vastly more efficient version of the SCS. The main algorithm is denoted the Sequential Conditioned Importance Sampling (SCIS) algorithm.

2 THE SEQUENTIAL CONDITIONED SAMPLING ALGORITHM

The outcome of a trial by the SCS algorithm is random and can be a 'failure' or a 'success'. The trial consists of a random number of steps, which can be at most n. The trial proceeds as follows:

1. First we generate a random value $X_1 = x_1$ in accordance with the one-dimensional normal probability distribution with mean m_1 and variance c_{11}.
2. If $x_1 \notin [a_1, b_1]$, the trial is terminated and the outcome is 'failure'. If $x_1 \in [a_1, b_1]$, we proceed to the next step.
3. If in the first $k-1$ steps we have avoided a failure, or equivalently, the outcomes of the first $k-1$ steps satisfy the inequalities $a_i \le x_i \le b_i$, $i = 1, \cdots, k-1$, then we proceed to the kth step. For this we generate $X_k = x_k$ using the one-dimensional normal probability distribution with mean $m_k(x_1, \cdots, x_{k-1})$ and variance ν_k (see eqn (3)).
4. If $x_k \notin [a_k, b_k]$, the trial is terminated and the outcome is 'failure'.
5. If, finally, $(x_1, \cdots, x_n) \in Q$, then the trial is terminated and the outcome is 'success'.

We denote by P_1 the probability that the outcome of the SCS trial is a 'success'.

Proposition 1

$$P(Q) = P_1 \tag{4}$$

Proof. We have

$$\begin{aligned} P_1 &= \int_{a_1}^{b_1} \varphi_1(x_1)\, dx_1 \int_{a_2}^{b_2} \varphi_2(x_2|x_1)\, dx_2 \cdots \\ &\quad \int_{a_n}^{b_n} \varphi_n(x_n|x_1, \cdots, x_{n-1})\, dx_n \\ &= \int_{a_1}^{b_1} \int_{a_2}^{b_2} \cdots \int_{a_n}^{b_n} \varphi(x_1, \cdots, x_n)\, dx_1\, dx_2 \cdots dx_n = P(Q) \end{aligned} \tag{5}$$

The proof is complete.

Proposition 1 suggests the following algorithm for estimating $P(Q)$: perform N independent SCS trials and estimate $P(Q)$ using the formula

$$\hat{P}(Q) = \frac{\text{number of successes}}{N} \tag{6}$$

The estimated coefficient of variation (c.o.v.) of the above estimate is easily shown to be

$$\text{c.o.v.} = \sqrt{\frac{1 - \hat{P}(Q)}{N\hat{P}(Q)}} \tag{7}$$

3 THE SEQUENTIAL CONDITIONED IMPORTANCE SAMPLING ALGORITHM

The SCIS algorithm improves on SCS by using an importance sampling density at each step of the trial instead of the normal density. The outcome of a trial by the SCIS algorithm is a number. Each trial consists of n steps, as described below:

1. In the first step, a random value x_1 is sampled in accordance with the probability density function

$$\frac{\varphi_1(x_1)}{\Phi_1([a_1, b_1])} I_{[a_1, b_1]}(x_1) \tag{8}$$

where $\Phi_1([a_1,b_1]) = \Phi_1(b_1) - \Phi_1(a_1)$ and $I_{[a,b]}(x)$ is the indicator function

$$I_{[a,b]}(x) = \begin{cases} 1 & \text{if } x \in [a,b] \\ 0 & \text{if } x \notin [a,b] \end{cases} \tag{9}$$

The reader will recognize eqn (8) as the conditional density of X_1 given $X_1 \in [a_1,b_1]$.

2. In the kth step, after the values $x_1, \cdots, x_{k-1}$ have been sampled, we sample a random value x_k in accordance with the probability density function

$$\frac{\varphi_k(x_k|x_1, \cdots, x_{k-1})}{\Phi_k([a_k,b_k]|x_1, \cdots, x_{k-1})} I_{[a_k,b_k]}(x_k) \tag{10}$$

where $\Phi_k([a_k,b_k]|x_1, \cdots, x_{k-1}) = \Phi_k(b_k|x_1, \cdots, x_{k-1}) - \Phi_k(a_k|x_1 \cdots, x_{k-1})$. Clearly, eqn (10) is the conditional density of X_k given $X_1 = x_1 \cap \ldots \cap X_{k-1} = x_{k-1} \cap X_k \in [a_k,b_k]$.

3. The trial is terminated after we sample the nth value x_n.

Let Y be the random variable

$$Y = \prod_{k=1}^{n} \Phi_k([a_k,b_k]|x_1, \cdots, x_{k-1}) \tag{11}$$

Proposition 2

$$P(Q) = E[Y] \tag{12}$$

where $E[\cdot]$ denotes the expectation.

Proof. We can write the right-hand side of eqn (12) in the form

$$\begin{aligned} E[Y] &= \int_{-\infty}^{\infty} dx_1 \cdots \int_{-\infty}^{\infty} dx_n \prod_{k=1}^{n} \Phi_k([a_k,b_k]|x_1, \cdots, x_{k-1}) \\ &\quad \times \prod_{k=1}^{n} \frac{\varphi_k(x_k|x_1, \cdots, x_{k-1})}{\Phi_k([a_k,b_k]|x_1, \cdots, x_{k-1})} I_{[a_k,b_k]}(x_k) \\ &= \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} \varphi_1(x_1)\varphi_2(x_2|x_1)\cdots\varphi_n(x_n|x_1, \cdots, x_{n-1}) \\ &\quad \times dx_1 \cdots dx_n = P(Q) \end{aligned} \tag{13}$$

The proof is complete.

Proposition 2 suggests the following algorithm for calculating $P(Q)$: perform N independent SCIS trials and obtain the N outcomes $y_1, \cdots, y_N$ of the random variable defined in eqn (11). Estimate $P(Q)$ by the formula

$$\hat{P}(Q) = \frac{1}{N} \sum_{i=1}^{N} y_i \tag{14}$$

It is an easy matter to show that the coefficient of variation of the above estimate is

$$\text{c.o.v.} = \frac{1}{N\hat{P}(Q)} \sqrt{\sum_{i=1}^{N} (y_i - \hat{P}(Q))^2} \tag{15}$$

We note that the computation of y_i from eqn (11) involves repeated evaluations of the one-dimensional cumulative normal probability function. It is also worth noting that, when the random variables are statistically independent, eqn (11) gives the exact result with each trial.

Table 1. Estimates of $P(Q)$ by different methods

				SCS algorithm			SCIS algorithm		
n	r	Cube	Exact $P(Q)$	$\frac{\hat{P}(Q)}{P(Q)}$	N	C.o.v.	$\frac{\hat{P}(Q)}{P(Q)}$	N	C.o.v.
3	0·2	$[-5,-1]^3$	$5{\cdot}20 \times 10^{-3}$	1·12	68,517	0·05	1·01	10	0·02
	0·4	$[-5,-1]^3$	$9{\cdot}95 \times 10^{-3}$	1·13	35,243	0·05	1·03	21	0·05
	0·6	$[-5,-1]^3$	$2{\cdot}19 \times 10^{-2}$	1·02	17,506	0·05	0·96	60	0·05
	0·8	$[-5,-1]^3$	$4{\cdot}94 \times 10^{-2}$	0·94	8218	0·05	0·97	101	0·05
	0·2	$[-10,-2]^3$	$2{\cdot}22 \times 10^{-5}$	1·44	1,000,000	0·18	1·01	10	0·03
	0·4	$[-10,-2]^3$	$1{\cdot}00 \times 10^{-4}$	1·33	1,000,000	0·09	0·97	30	0·05
	0·6	$[-10,-2]^3$	$5{\cdot}63 \times 10^{-4}$	1·40	500,000	0·05	0·96	98	0·05
	0·8	$[-10,-2]^3$	$2{\cdot}90 \times 10^{-3}$	1·21	113,220	0·05	0·96	150	0·05
5	0·2	$[-5,-1]^5$	$2{\cdot}31 \times 10^{-4}$	1·09	1,000,000	0·06	0·99	10	0·04
	0·4	$[-5,-1]^5$	$1{\cdot}28 \times 10^{-3}$	1·16	267,703	0·05	0·92	57	0·05
	0·6	$[-5,-1]^5$	$6{\cdot}81 \times 10^{-3}$	1·11	52,284	0·05	0·99	228	0·05
	0·8	$[-5,-1]^5$	$2{\cdot}82 \times 10^{-2}$	1·01	13,596	0·05	1·00	271	0·05
	0·2	$[-10,-2]^5$	$4{\cdot}42 \times 10^{-8}$	—	1,000,000	—	0·99	10	0·04
	0·4	$[-10,-2]^5$	$2{\cdot}71 \times 10^{-6}$	0·92	1,000,000	0·71	0·99	166	0·05
	0·6	$[-10,-2]^5$	$7{\cdot}35 \times 10^{-5}$	1·96	1,000,000	0·08	0·99	353	0·05
	0·8	$[-10,-2]^5$	$1{\cdot}16 \times 10^{-3}$	1·32	260,701	0·05	1·01	483	0·05
7	0·2	$[-5,-1]^7$	$1{\cdot}32 \times 10^{-5}$	1·29	1,000,000	0·24	0·95	17	0·05
	0·4	$[-5,-1]^7$	$2{\cdot}53 \times 10^{-4}$	1·15	1,000,000	0·06	1·09	378	0·05
	0·6	$[-5,-1]^7$	$2{\cdot}98 \times 10^{-3}$	1·16	115,402	0·05	1·08	721	0·05
	0·8	$[-5,-1]^7$	$1{\cdot}95 \times 10^{-2}$	1·09	18,404	0·05	1·09	546	0·05
	0·2	$[-10,-2]^7$	$1{\cdot}59 \times 10^{-10}$	—	1000,000	-	0·94	21	0·05
	0·4	$[-10,-2]^7$	$1{\cdot}16 \times 10^{-7}$	—	1,000,000	-	1·08	793	0·05
	0·6	$[-10,-2]^7$	$1{\cdot}85 \times 10^{-5}$	2·70	1,000,000	0·14	0·99	2068	0·05
	0·8	$[-10,-2]^7$	$6{\cdot}48 \times 10^{-4}$	1·60	383,169	0·05	1·04	1310	0·05

4 NUMERICAL TESTS

Without loss of generality, in the following tests we consider the case of a standard multinormal probability distribution, i.e. the case with $\mathbf{m} = (0, \cdots, 0)$ and $\mathbf{C} = \mathbf{R}$, where $\mathbf{R}$ is a correlation matrix having the elements ρ_{ij}, $i, j = 1, \cdots, n$. A linear transformation of random variables can be used to transform a non-standard case into this standard form.

4.1 Comparison of algorithms

For the special class of correlation matrices with $\rho_{ij} = r_i r_j$ for $i \neq j$ and $\rho_{ii} = 1$, the probability $P(Q)$ can be represented by a one-dimensional integral.[1,7,8] For this class, denoted the 'D-S class', the exact value of $P(Q)$ can be computed by one-dimensional numerical integration. Table 1 compares the exact results computed by Simpson's rule with the estimates obtained by the SCS and SCIS algorithms for selected values of the correlation parameter $r_i = r_j = r$ (i.e. $\rho_{ij} = r^2$ for $i \neq j$), dimension n and cubes $[a,b]^n$. The following rule is used for terminating the simulation in each case: for the SCS algorithm, the simulation is terminated when c.o.v. $\leq$ 0.05 or $N = 10^6$; for the SCIS algorithm, the simulation is terminated when c.o.v. $\leq$ 0.05 and $N \geq 10$. For each algorithm, we have listed the ratio of the estimate $\hat{P}(Q)$ to the exact value $P(Q)$, as well as the corresponding c.o.v. and N. Note that, for three cases with small probabilities, the SCS algorithm did not produce a result within the maximum allowable N.

The results in Table 1 clearly demonstrate the accuracy of the SCIS algorithm and its vastly improved efficiency relative to the SCS algorithm. We note that the ratios $\hat{P}(Q)/P(Q)$ for the SCIS algorithm are all near unity and that the required number of trials to achieve c.o.v. $\leq$ 0.05 is manageably small in all cases. The results indicate that the required number of SCIS trials tends to increase with the correlation and dimension. However, no dependence on the magnitude of the estimated probability is observed.

Since the computational effort within each SCS or SCIS trial is nearly the same, the required computation time to achieve a desired level of accuracy is approximately proportional to the number of trials in each algorithm. Table 1 shows that the SCIS algorithm is several orders of magnitude faster than the SCS algorithm. Absolute measures of the computation time for the SCIS algorithm are given below.

4.2 Symmetry test of the SCIS algorithm

Under the choice of the correlation matrix from the D-S class with $r_i = r_j = r$, additional tests are possible based on the geometric symmetry properties of the corresponding multinormal probability space. Let Q_1 and Q_2 be two rectangular domains having the property that Q_1 transforms into Q_2 by means of changing the numbering of the coordinate axes. Then, necessarily $P(Q_1) = P(Q_2)$. We use this property to further examine the SCIS algorithm.

We represent the seven-dimensional cube $[0,2]^7$ as a union of small cubes of the form

$$Q_l = I_1 \times I_2 \times \cdots \times I_6 \tag{16}$$

where each interval I_k is either [0,1] or [1,2]. The label l is the corresponding binary sequence $l = (\varepsilon_1, \cdots, \varepsilon_7)$, wherein $\varepsilon_1 = 1$ if $I_1 = [0,1]$, $\varepsilon_1 = 2$ if $I_1 = [1,2]$, etc. There are $2^7 = 128$ small cubes Q_l. Using the SCIS algorithm with $N = 100$ and $r = 0.5$ (i.e. $\rho_{ij} = 0.25$ for $i \neq j$), we estimate the probabilities $P(Q_l)$ for each of the 128 cubes. Next, we define the groups of cubes

$$\mathcal{G}_k = \{Q_l : \text{ the number of 1's in the label sequence } l \text{ is } k \tag{17}$$

The number of members within each group $\mathcal{G}_k$ is given in Table 2. Due to the symmetry property described above $P(Q_l)$ is constant within each group $\mathcal{G}_k$. Table 2 lists the empirical mean and c.o.v. of the SCIS estimates $\hat{P}(Q_l)$ for the cubes within each group $\mathcal{G}_k$. The small c.o.v.s indicate the closeness of the estimates of $\hat{P}(Q_l)$ within each group, thus providing a further test of the accuracy of the SCIS algorithm

4.3 Additivity test of the SCIS algorithm

We can use the estimates $\hat{P}(Q_l)$ for the small cubes to perform an additivity test of the SCIS algorithm. We calculate independent SCIS estimates for 14 rectangular domains obtained as unions of subsets of the cubes:

$$B_i = \bigcup_{\varepsilon_i = 1} Q_l, \quad \bar{B}_i = \bigcup_{\varepsilon_i = 2} Q_l, \quad i = 1, 2, \cdots, 7 \tag{18}$$

($\bar{B}_i$ is the complement of B_i within $[0,2]^7$) and compare them with the values obtained by appropriate summation of the estimates $\hat{P}(Q_l)$. The results are shown in Table 3. Note that each rectangular domain B_i or $\bar{B}_i$ contains 64 cubes. Since the estimates of $\hat{P}(Q_l)$ were based on $N = 100$ trials each, the summation result for each B_i or $\bar{B}_i$ is effectively based on 6400 trials. The direct SCIS result in Table 3 for each B_i or $\bar{B}_i$ is based on $N = 1000$ trials. The results using the two methods are practically identical, once again demonstrating the remarkable accuracy of the SCIS algorithm.

4.4 Convergence test of the SCIS algorithm

The results in Table 1 indicated that the required number of SCIS trials for a given level of accuracy tends to increase

Table 2. Statistics of estimates $\hat{P}(Q_l)$ within each group $\mathcal{G}_k$

	Number of members within group	Mean of $\hat{P}(Q_l)$ estimates within group	C.o.v. of $\hat{P}(Q_l)$ estimate within group
$\mathcal{G}_0$	1	$8{\cdot}07 \times 10^{-5}$	—
$\mathcal{G}_1$	7	$8{\cdot}50 \times 10^{-5}$	0·035
$\mathcal{G}_2$	21	$9{\cdot}90 \times 10^{-5}$	0·034
$\mathcal{G}_3$	35	$1{\cdot}31 \times 10^{-4}$	0·020
$\mathcal{G}_4$	35	$1{\cdot}91 \times 10^{-4}$	0·022
$\mathcal{G}_5$	21	$3{\cdot}17 \times 10^{-4}$	0·017
$\mathcal{G}_6$	7	$5{\cdot}88 \times 10^{-4}$	0·011
$\mathcal{G}_7$	1	$1{\cdot}21 \times 10^{-3}$	—

Table 3. Additivity test of SCIS

	$P(B_i)$		$P(\bar{B}_i)$	
i	SCIS	Summation	SCIS	Summation
1	0·0160	0·0160	0·0101	0·0100
2	0·0161	0·0159	0·0101	0·0101
3	0·0160	0·0160	0·0101	0·0100
4	0·0160	0·0160	0·0101	0·0100
5	0·0160	0·0159	0·0101	0·0101
6	0·0160	0·0160	0·0101	0·0101
7	0·0160	0·0160	0·0101	0·0100

with increasing dimension and correlation between the variables. To further examine the efficiency of the SCIS algorithm, we compute the estimates of $P(Q)$ for rectangular domains defined by $a_i = -\infty$, $b_i = -2 + 0.05(i-1)$, $i = 1, 2, \cdots, n$ for $n = 5$, 10, 15 and 20. We use a correlation matrix from the D-S class with $r_i = r_j = r = \sqrt{0.5}$ (i.e. $\rho_{ij} = 0.5$ for $i \neq j$). Fig. 1 shows plots of estimates $\hat{P}(Q)$ and the corresponding c.o.v.s against the number of trials N, as well as the computation times required to achieve c.o.v. ≤ 0.05 on a desktop computer with a 100-MHz Pentium processor. It is noted that the simulation result quickly stabilizes in all cases and that the required computation time is trivial even for $n = 20$. In order to examine the effect of negative correlation, we repeated the same analysis for the transformed variables $z_i = (-1)^i x_i$. Note that each pair of the transformed variables with one odd and one even index has correlation coefficient $\rho_{ij} = -0.5$. The results are shown in Fig. 2. The probability contents within the rectangular domains are now extremely small. Nevertheless, the SCIS algorithm converges more rapidly than in the previous case. This has to do with the nearly uniform distribution of the probability density within the rectangular domains considered for the transformed variables.

5 PROBABILITY SENSITIVITIES

In certain applications, including structural system reliability, the sensitivities of the multinormal probability with respect to the parameters defining the rectangular domain Q or the means, variances and correlation coefficients of the variables are of interest (see Hohenbichler and Rackwitz[9] and Bjerager and Krenk[10] for applications in structural system reliability). The sensitivities with respect to the mean and variances are easily obtained in terms of the matrix defining the linear transformation to the standard normal space. The SCIS algorithm offers a convenient way for computing the sensitivities with respect to the boundaries of Q and the correlation coefficients ρ_{ij}, as described below.

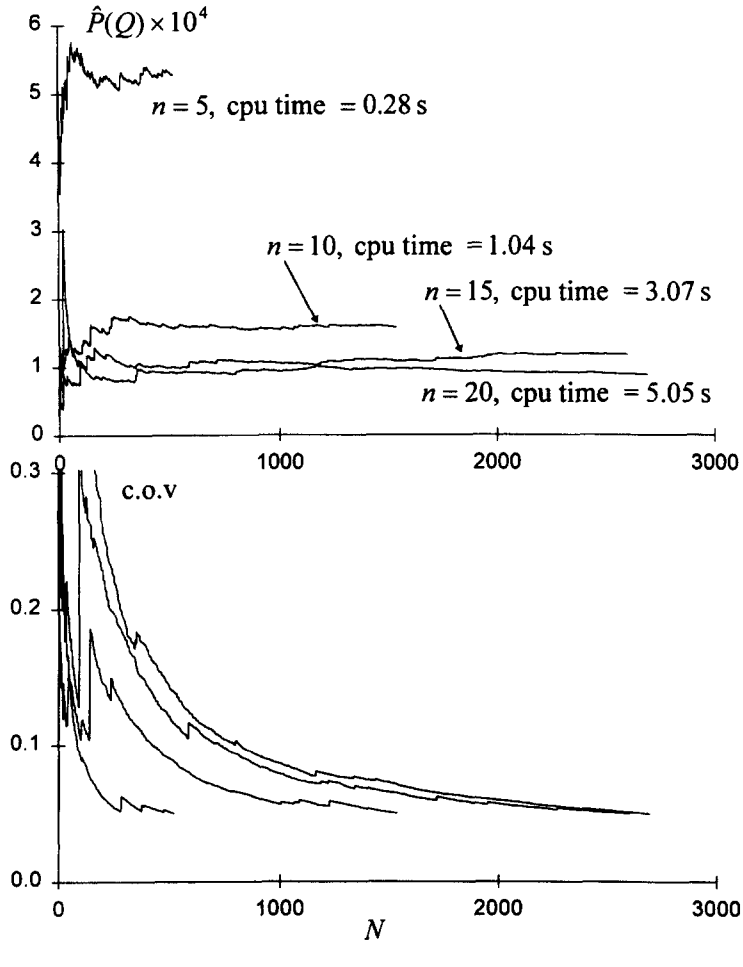

Fig. 1. Estimated $\hat{P}(Q)$ and c.o.v. as a function of number of trials for variables x_i.

Consider the sensitivities of $P(Q)$ with respect to the bounds a_k and b_k of X_k. Using eqn (2), these can be written as

$$\frac{\partial P(Q)}{\partial a_k} = -\int_{a_1}^{b_1}\cdots\int_{a_{k-1}}^{b_{k-1}}\int_{a_{k+1}}^{b_{k+1}}\cdots \int_{a_n}^{b_n}\varphi(x_1, \cdots, x_{k-1}, a_k, x_{k+1}, \cdots, x_n)\, dx_1\cdots dx_{k-1}\, dx_{k+1}\cdots dx_n = -\varphi_k(a_k)P_{a_k}(Q) \tag{19}$$

$$\frac{\partial P(Q)}{\partial b_k} = -\int_{a_1}^{b_1}\cdots\int_{a_{k-1}}^{b_{k-1}}\int_{a_{k+1}}^{b_{k+1}}\cdots \int_{a_n}^{b_n}\varphi(x_1, \cdots, x_{k-1}, b_k, x_{k+1}, \cdots, x_n)\, dx_1\cdots dx_{k-1}\, dx_{k+1}\cdots dx_n = -\varphi_k(b_k)P_{b_k}(Q) \tag{20}$$

where $\varphi_k(\cdot)$ is the marginal density of X_k, $Q = [a_1,b_1] \times \cdots \times [a_{k-1},b_{k-1}] \times [a_{k+1},b_{k+1}] \times \cdots \times [a_n,b_n]$ is a rectangular domain in the $(n-1)$-dimensional space, and

$$P_{x_k}(Q) = \int_{a_1}^{b_1}\cdots\int_{a_{k-1}}^{b_{k-1}}\int_{a_{k+1}}^{b_{k+1}}\cdots \int_{a_n}^{b_n}\varphi_{1,\cdots,k-1,k+1,\cdots,n}(x_1, \cdots, x_{k-1}, x_{k+1}, \cdots, x_n|x_k)\, dx_1\cdots dx_{k-1}\, dx_{k+1}\cdots dx_n \tag{21}$$

The required probabilities $P_{a_k}(Q)$ and $P_{b_k}(Q)$ can be estimated by SCIS using the conditional mean vector $\mathbf{m}^{(k)}$ and covariance matrix $\mathbf{C}^{(k)}$ for given $X_k = a_k$ and $X_k = b_k$, respectively. Hence, an SCIS simulation in the $(n-1)$-dimensional space yields the sensitivity with respect to each parameter a_k or b_k. Unfortunately, a similar result for the sensitivities with respect to the correlation coefficients is not possible. Instead, we explore the possibility of using finite differences.

Given that each SCIS trial is a realization of a continuous random variable (see eqn (11)), the sensitivities with respect to any set of parameters can be estimated by finite differences, provided the same sequence of random numbers are

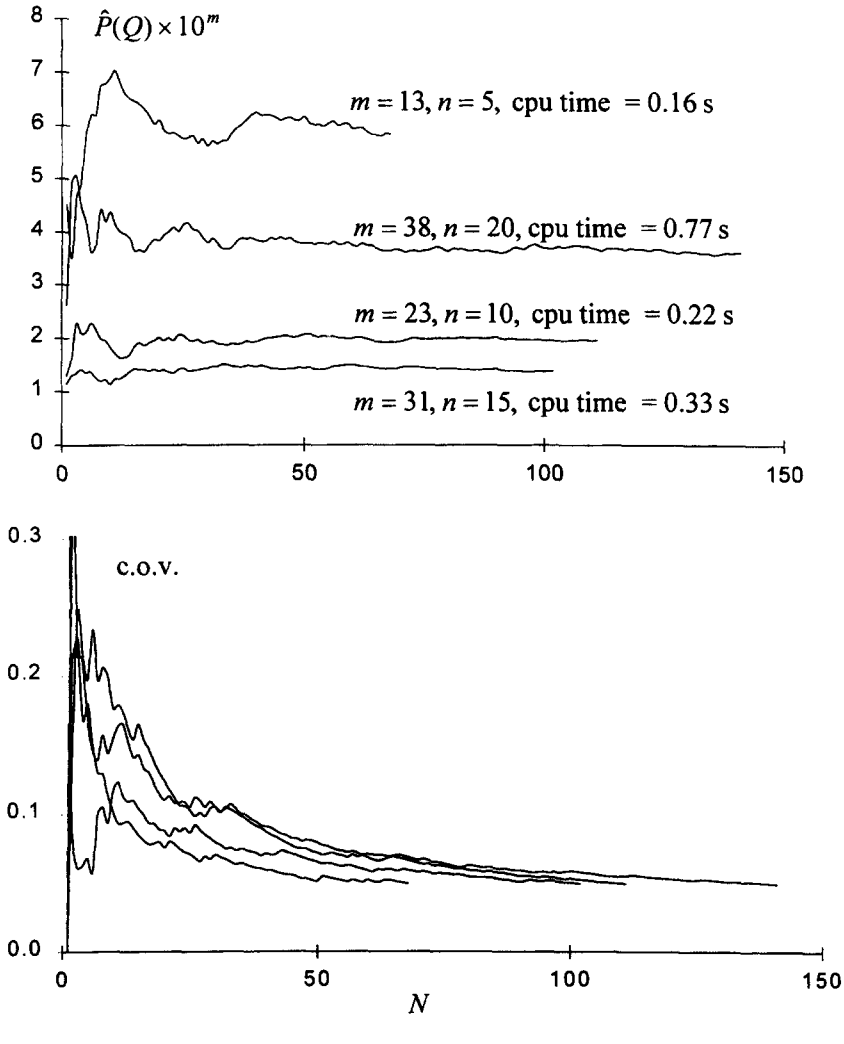

Fig. 2. Estimated $\hat{P}(Q)$ and c.o.v. as a function of number of trials for variables x_i.

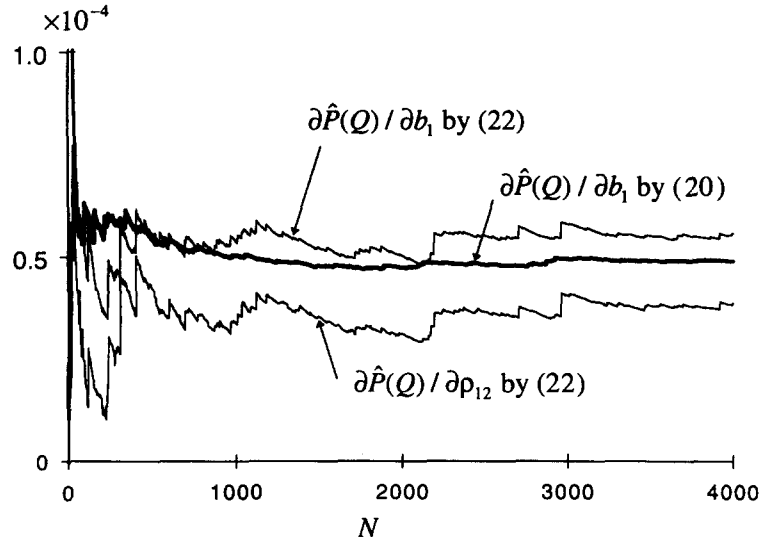

Fig. 3. Probability sensitivities as a function of number of trials.

used in the simulation, i.e.

$$\frac{\partial P(Q)}{\partial \theta} \cong \frac{\hat{P}(Q)_{\theta+\Delta\theta} - \hat{P}(Q)_{\theta}}{\Delta\theta} \tag{22}$$

where θ is the parameter of interest, $\Delta\theta$ is a small variation in it, and $\hat{P}(Q)_x$ is the SCIS estimate of $P(Q)$ for $\theta = x$. To generate the same sequence of random numbers, we initiate the simulation with a fixed seed.

To explore the effectiveness of the above approaches, the sensitivities of $P(Q)$ are computed with respect to b_1 and ρ_{12} for the case with $n = 20$ in Fig. 1, i.e. for $a_i = -\infty$, $b_i = -2 + 0.05(i-1)$, $i = 1, 2, \cdots, 20$, and a correlation matrix from the D-S class with $r_i = r_j = r = \sqrt{0.5}$. Fig. 3 shows plots of the estimates of $\partial P(Q)/\partial b_1$ based on eqns (20) and (22) and the estimate of $\partial P(Q)/\partial \rho_{12}$ based on eqn (22) as a function of the number of trials. It is seen that the estimates based on the finite difference approach are not as stable as those based on eqn (20), but they still provide reasonably accurate approximations of the probability sensitivities.

6 EXAMPLE APPLICATION

We present this example in order to demonstrate the use of multinormal probability in reliability analysis of structural systems.

Consider the one-bay frame in Fig. 4(a), which is subjected to random horizontal and vertical loads H and V. The frame has random plastic moment capacities M_i, $i = 1, \cdots, 5$, at the critical locations shown in the figure. Under the applied loads, this frame may fail in any of the three mechanisms shown in Fig. 4(b). Based on the principle of virtual work, the three mechanism are described by the limit-state functions

$$g_1(M_1, M_2, M_4, M_5, H) = M_1 + M_2 + M_4 + M_5 - 5H \tag{23}$$

$$g_2(M_2, M_3, M_4, V) = M_2 + 2M_3 + M_4 - 5V \tag{24}$$

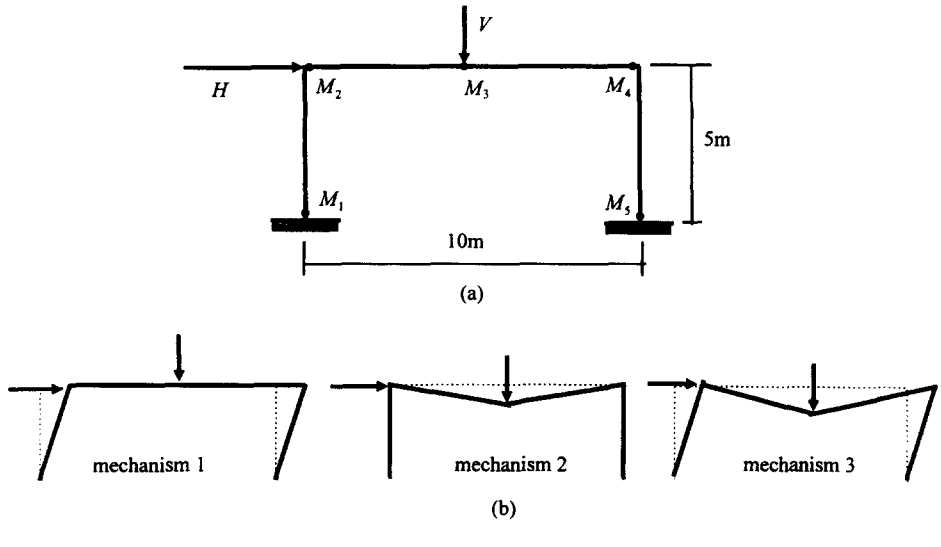

Fig. 4. Example frame structure and its failure mechanisms.

$$g_3(M_1, M_3, M_4, M_5, H, V) = M_1 + 2M_3 + 2M_4 + M_5 - 5H - 5V \quad (25)$$

such that the event $G_k = \{g_k \leq 0\}$ denotes the occurrence of the kth mechanism. Assume M_i have identical means equal to 200 kNm, identical standard deviations equal to 30 kNm and the correlation coefficients $\rho_{M_iM_j} = 0.3$, $i \neq j$. Furthermore, assume H has mean 60 kN and standard deviation 30 kN, and V has mean 60 kN and standard deviation 15 kN. H and V are assumed to be statistically independent of each other and of M_i. We first assume a joint normal distribution for these variables. The case of non-normal distributions is discussed subsequently.

The frame represents a series system reliability problem, as the failure event is the union of the 'component' failure events G_k, $k = 1, 2, 3$. We first introduce the standard normal variables $\mathbf{u} = (u_1, \cdots, u_7)$ by a linear transformation of the random variables $\mathbf{x} = (M_1, \cdots, M_5, H, V)$ such that $\mathbf{u}$ has a zero mean and a unit covariance matrix. In the space of $\mathbf{u}$, each of the hyper-planes $g_k(\mathbf{x}) = 0$, $k = 1, 2, 3$, can be written in the form $\beta_k - \alpha_k^T\mathbf{u} = 0$, where β_k denotes the distance from the origin to the kth hyper-plane (commonly known as the *reliability index*) and α_k denotes the corresponding unit normal vector. The solutions for β_k and α_k are listed in the first three rows of Table 4.

Let $\nu_k = \alpha_k^T\mathbf{u}$. It is easy to verify that ν_k are jointly normal random variables having zero means, unit variances and correlation coefficients $\rho_{kl} = \alpha_k^T\alpha_l$, $k, l = 1, 2, 3$. Using the α_k values in Table 4, we obtain $\rho_{\nu_1\nu_2} = 0.278$, $\rho_{\nu_2\nu_3} = 0.645$ and $\rho_{\nu_3\nu_1} = 0.875$. The probability of failure of the frame can now be formulated as follows:

$$P_f = P\left[\bigcup_{k=1}^{3} g_k(\mathbf{x}) \leq 0\right] = P\left[\bigcup_{k=1}^{3} (\beta_k - \alpha_k^T\mathbf{u}) \leq 0\right]$$
$$= P\left(\bigcup_{k=1}^{3} \beta_k \leq \nu_k\right) = 1 - P\left(\bigcap_{k=1}^{3} \nu_k \leq \beta_k\right) \quad (26)$$

We see that the failure probability of the frame is given as the complement of a trivariate cumulative normal probability. Using 1000 SCIS simulations in the space of variables ν_k, we obtain $\hat{P}_f = 0.329 \times 10^{-2}$ with a negligible c.o.v.

Now suppose the frame survives a proof test under a horizontal load $h = 80$ kN and a vertical load $v = 70$ kN. This observation gives us indirect information about the plastic moment capacities of the frame. We wish to update our estimate of the failure probability in light of this information. Let $g_4 = -g_1(M_1, M_2, M_4, M_5, h)$, $g_5 = -g_2(M_2, M_3, M_4, v)$ and $g_6 = -g_3(M_1, M_3, M_4, M_5, h, v)$. It should be clear that $\{g_4 \leq 0 \cap g_5 \leq 0 \cap g_6 \leq 0\}$ denotes the observed event. Using the notation $G_i \equiv \{g_i \leq 0\}$, the updated failure probability is

$$P_{f|\text{survival at proof test}} = P(G_i \cup G_2 \cup G_3 | G_4G_5G_6) = \frac{P(G_1G_4G_5G_6 \cup G_2G_4G_5G_6 \cup G_3G_4G_5G_6)}{P(G_4G_5G_6)} \quad (27)$$

The probability in the numerator can be expanded by use of the inclusion–exclusion rule to read

$$P(G_1G_4G_5G_6 \cup G_2G_4G_5G_6 \cup G_3G_4G_5G_6) = P(G_1G_4G_5G_6) + P(G_2G_4G_5G_6) + P(G_3G_4G_5G_6) - P(G_1G_2G_4G_5G_6) - P(G_1G_3G_4G_5G_6) - P(G_2G_3G_4G_5G_6) + P(G_1G_2G_3G_4G_5G_6) \quad (28)$$

Each of the above terms represents a parallel system reliability problem (intersection of events) and can be computed in terms of the multinormal cumulative probability For example,

$$P(G_1G_2G_3G_4G_5G_6) = P\left[\bigcap_{k=1}^{6} g_k(x) \leq 0\right] = P\left[\bigcap_{k=1}^{6} (\beta_k - \alpha_k^T\mathbf{u}) \leq 0\right] = P\left(\bigcap_{k=1}^{6} \beta_k \leq \nu_k\right) = P\left(\bigcap_{k=1}^{6} \nu_k \leq -\beta_k\right) \quad (29)$$

where, in the last equation, we have employed the symmetry property of the normal space. The result is a six-dimensional multinormal cumulative probability in terms of the variable ν_k. Each of the probability terms in eqn (28) and the denominator in eqn (27) are computed by 1000 SCIS simulations The required values for β_k and α_k are listed in Table 4. The results are used to estimate the updated probability as follows

$$\hat{P}_{f|\text{survival at proof test}} = \frac{0.174 \times 10^{-2} + 0.721 \times 10^{-5} + 0.210 \times 10^{-2} - 0.189 \times 10^{-6} - 0.790 \times 10^{-3} - 0.334 \times 10^{-5} + 0.187 \times 10^{-6}}{1 - 0.661^{-6}}$$
$$= 0.305 \times 10^{-2} \quad (30)$$

Table 4. β_k and α_k^T for the case of normal random variables

k	β_k	α_k^T						
1	2·92	−0·333	−0·244	−0·061	−0·185	−0·158	0·876	0·00
2	4·27	−0·307	−0·413	−0·520	−0·234	0·000	0·000	0·64
3	2·86	−0·357	−0·157	−0·340	−0·281	−0·129	0·714	0·35
4	−1·21	0·689	0·506	0·126	0·383	0·327	0·000	0·00
5	−3·33	0·400	0·538	0·677	0·304	0·000	0·000	0·00
6	−0·79	0·593	0·261	0·564	0·466	0·213	0·000	0·00

Table 5. β_k and α_k^T for the case of non-normal random variables

k	β_k	α_k^T						
1	2·32	−0·161	−0·123	−0·031	−0·093	−0·080	0·970	0·000
2	4·01	−0·203	−0·275	−0·331	−0·158	0·000	0·000	0·865
3	2·47	−0·203	−0·089	−0·190	−0·157	−0·073	0·909	0·243
4	−6·62	0·691	0·505	0·126	0·382	0·326	0·000	0·000
5	−7·33	0·408	0·568	0·632	0·335	0·000	0·000	0·000
6	−4·38	0·608	0·260	0·551	0·456	0·227	0·000	0·000

The updated probability of failure is found to be smaller than the previous failure probability ($= 0.329 \times 10^{-2}$). This is due to the 'good' news of survival during the proof test. The difference, however, is found to be rather small. This is because the failure probability of the frame is dominated by the uncertainty in the loads, for which the proof test does not provide any information.

The above solution is applicable when the random variables are normal. For the case of non-normal random variables, the first-order reliability method (FORM) can be used to obtain an approximation of the failure probability.[2] For this purpose, we first transform the random variables $\mathbf{x}$ into standard normal variables $\mathbf{u}$ through a non-linear transformation. Each of the hyper-planes $g_k(\mathbf{x}) = 0$ transforms into a curved surface in the space of $\mathbf{u}$. These surfaces are linearized (i.e. approximated by their tangent hyper-planes) at their respective points of minimum distance from the origin of the $\mathbf{u}$ space. β_k and α_k now represent the distance from the origin and the unit normal vector for the approximating hyper-planes. The remaining analysis is as before, except that the probability estimates must now be regarded as first-order approximations.

Suppose M_i, $i = 1, \cdots, 5$, are jointly lognormal, H has a type I extreme-value distribution for the largest values and V has a gamma distribution. The first and second moments are as before. For these distributions, the FORM approximation results in the β_k and α_k values listed in Table 5. Using these values, 1000 SCIS simulations produce $\hat{P}_f \cong 0.120 \times 10^{-1}$ and $\hat{P}_{f|\text{survival at proof test}} \cong 0.115 \times 10^{-1}$. It is interesting to note that these probabilities are much larger than the corresponding probabilities for the case of normal distributions. The reason is the heavier tails of the type I and gamma distributions relative to the tail of the normal distribution.

The above examples demonstrated the use of the multinormal cumulative probability function for calculation of probabilities for series and parallel systems (respectively representing unions and intersections of events). The sensitivities of the multinormal probability are also of interest in reliability analysis. For example, in the above problem one may be interested in the sensitivities of the updated probability with respect to the proof load values h and ν. By use of the chain rule, one first finds the sensitivities of β_k and α_k with respect to these parameters, and then computes the sensitivities of the multinormal probabilities with respect to β_k and α_k in the manner described in the previous section.

7 SUMMARY AND CONCLUSIONS

We have presented an efficient and accurate algorithm for computing the multinormal probability of rectangular domains by simulation. The required computational effort increases modestly with the dimension of the problem and the correlation between the random variables, but it is independent of the magnitude of the probability of interest and is easily manageable for large dimensions and strong correlations. The algorithm can also be used to compute the sensitivities of the probability with respect to correlation coefficients and parameters that define the boundaries of the domain. The multinormal probability and its sensitivities have important applications in the theory of structural system reliability. This is demonstrated through a simple example.

ACKNOWLEDGEMENTS

A part of this study was performed at the American University of Armenia in Yerevan, Armenia, an affiliate of the University of California, where the second author serves as the Dean of Engineering concurrently with his position at Berkeley. The study was supported by a grant from the engineering Research Center of AUA.

REFERENCES

1. Johnson, N. L. and Kotz, S., *Distributions in Statistics: Continuous Multivariate Distributions*. John Wiley, New York, 1972.
2. Hohenbichler, M. and Rackwitz, R. First-order concepts in system reliability. *Structural Safety*, 1983, **1**(3), 177–188.
3. Gollwitzer, S. and Rackwitz, R. An efficient numerical solution to the multinormal integral. *Probabilistic Engineering Mechanics*, 1988, **3**(2), 98–101.
4. Genz, A. Numerical computation of multivariate normal probabilities. *J. Computational and Graphical Statistics*, 1992, **1**, 141–149.
5. Joe, H. Approximations to multivariate normal rectangle probabilities based on conditional expectations. *J. American Statistical Association*, 1995, **90**(431), 957–964.
6. Rubenstein, R. Y., *Simulation and the Monte Carlo Method*. John Wiley, New York, 1981.

7. Dunnett, C. W. and Sobel, M. Approximations to the probability integral and certain percentage points of a multivariate analogue of Student's t-distribution. *Biometrika*, 1955, **42**, 258–260.
8. Melchers, R., *Structural Reliability, Analysis and Prediction*. John Wiley, New York, 1987.
9. Hohenbichler, M. and Rackwitz, R. Sensitivity and importance measures in structural reliability. *Civil Engineering Systems*, 1986, **3**, 203–210.
10. Bjerager, P. and Krenk, S. Parametric sensitivity in first-order reliability theory. *J. Engineering Mechanics, ASCE*, 1989, **115**(7), 1577–1582