Structural Eng. / Earthquake Eng. Vol. 5. No. 1, 31s-37s. April 1988
Japan Society of Civil Engineers (Proc. of JSCE No. 392/I-9)

UPDATING FIRST-AND SECOND-ORDER RELIABILITY ESTIMATES BY IMPORTANCE SAMPLING

By Munehisa FUJITA and Rüdiger RACKWITZ***

First- and second-order reliability methods have turned out to be efficient practical tools in structural reliability for direct probabilistic design or for the development of probability-based design codes. These methods are approximate but certain Monte Carlo techniques with importance sampling can make reliability estimates arbitrarily accurate. Three different methods are presented and tested at a suitable example with respect to their numerical efficiency. It is found that a method which also uses curvature information in the so-called most likely failure point usually is preferable to the alternatives if an update of first- or second-order estimates is necessary. However, that method becomes inadequate for very high problem dimensions and/or large failure probabilities.

Keywords : structural reliability, Monte Carlo method, importance sampling

1. INTRODUCTION

One of the crucial problems in structural reliability is the calculation of failure probability integrals of the type

$$P_f=P(F_x)=\int_{F_x} dF_{\underline{X}}(\underline{x}) \quad \cdots\cdots (1)$$

where $F_{\underline{X}}(\underline{x})$ is the joint distribution function of the random vector $\underline{X}=(X_1, \cdots, X_n)$ and $F_x=\{\underline{x} : h(\underline{x})\leq 0\}$ the failure domain. In recent years powerful methods have been developed to approximate that failure probability. In particular, let a piecewise continuous probability preserving transformation $\underline{X}=\underline{T}(\underline{U})$, for example, the Rosenblatt-transformation described in 1), exist, where $\underline{U}=(U_1, \cdots, U_n)$ is an independent, standard normal vector. Then, the probability in eq. (1) is

$$P_f=P(F_u)=\int_{F_u} \varphi_n(\underline{u})d\underline{u}\sim \Phi(-\beta)C \quad \cdots\cdots (2)$$

where $F_u=\{\underline{u} : h(\underline{T}(\underline{u}))=g(\underline{u})\leq 0\}$ with $g(\underline{u})$ at least twice differentiable in the solution point $\underline{u}^*$ (so-called β-point) defined by

$$\beta=\|\underline{u}^*\|=\min\{\|\underline{u}\|\} \text{ for } \{\underline{u} : g(\underline{u})\leq 0\} \quad \cdots\cdots (3)$$

$\Phi(\cdot)$ is the standard normal integral and, for convenience, $g(\underline{0})>0$ is assumed. The factor C is a second-order correction term for the first-order result $\Phi(-\beta)$. It can be given asymptotically correct as $(\beta\rightarrow\infty)$[2]

$$C=\prod_{i=1}^{n-1}(1-\beta\kappa_i)^{-1/2} \qquad (\beta\rightarrow\infty) \quad \cdots\cdots (4)$$

* Member of JSCE, Technical University of Munich, F. R. G., on leave from Shimizu Corporation, Tokyo, Japan (Laboratorium für den Konstruktiven Ingenieurbau, Technische Universität München, Arcisstr. 21, 8000 München 2, F. R. G.)
** Dr.-Ing. habil., Technical University of Munich, F. R. G.

where the κ_i are the main curvatures of $g(\underline{u})=0$ in $\underline{u}^{*2)}$. This correction factor takes account of the difference between the probability content of an approximating hyperplane and an approximating paraboloid in the β-point.

Although FORM (first-order reliability method) or SORM (second-order reliability method) estimates are considered as sufficient in many applications there have been repeated doubts about the validity and attempts to improve these results or at least quantify the error. The most promising approach appears to be a combination of first-or second-order concepts with certain Monte Carlo techniques with importance sampling. In the context of structural reliability these techniques have first been proposed by Shinozuka[3] and Harbitz[9] and have further been developed by Hohenbichler[5),6)], Melchers[7)], Ditlevsen et al.[8)], Harbitz[9)] and, recently, by Bjerager[10)]. The various methods differ especially with respect to the amount of information which is used from first-or second-order reliability (FORM or SORM) concepts. The methods in 8) and 10) use spherical sampling techniques which are not discussed here. In this study three alternative methods are briefly described and tested at a suitably chosen example in order to judge their efficiency.

2. IMPORTANCE SAMPLING

(1) Direct method

Based on Shinozuka[3)] and Rubinstein[11)], Melchers[7)] proposed the following failure probability estimate

$$P_f=\int l_{g(\underline{u})}\frac{\varphi_n(\underline{u})}{\Psi_n(\underline{u})}\Psi_n(\underline{u})d\underline{u}\sim\frac{1}{N}\sum_{i=1}^{N} l_{g(\underline{u}_i)}\frac{\varphi_n(\underline{u}_i)}{\Psi_n(\underline{u}_i)} \quad (5)$$

where $l_{g(\underline{u})}$ is the indicator function, such as

$l_{g(\underline{u}_i)}=0$: $g(\underline{u}_i)\geq 0$

$l_{g(\underline{u}_i)}=1$: $g(\underline{u}_i)<0$

and

$$\Psi_n(\underline{u})=\varphi_n(\underline{u};\underline{u}^*,\underline{\underline{D}}) \quad (6)$$

the sampling distribution with $\underline{u}^*$ the mean of $\underline{U}$ and $\underline{\underline{D}}$ a diagonal matrix which is usually set equal to $\underline{\underline{I}}$. It is clear that the estimation in eq. (5) has much smaller variance than the usual crude Monte Carlo estimator,

$$P_f=\int l_{g(\underline{u})}\varphi_n(\underline{u})d\underline{u}\sim\frac{1}{N}\sum_{i=1}^{N} l_{g(\underline{u}_i)} \quad (7)$$

with the components of $\underline{u}_i$ sampled independently from the standard normal distribution. The specific choice of $\Psi_n(\underline{u})$ in eq. (6) ensures that there are sufficiently many points where the indicator function is non-zero, i. e. in roughly 50 % of the simulation points, whereas much more sample points usually have to be generated in eq. (7) in order to find points in the failure region. The only information on the important sampling region used in eq. (5) is the location of the β-point. Simple g-function calls are necessary for each simulated point.

(2) Methods for estimating a correction factor

The next two methods rest on the idea that it is more efficient to estimate a correction factor for an approximate probability result, e. g. obtained by FORM or SORM, rather than to replace it by an importance sampling alternative as before.

We write

$$P_f=P(A)\frac{P(F)}{P(A)}=P(A)C \quad (8)$$

with A an approximation (linear or quadratic) to F. Then,

$$C\sim\frac{1}{N}\sum_{i=1}^{N} C_i \quad (9)$$

in which C_i is the probability ratio in eq. (8) and which is estimated by simulation. It is clear that the

simplest choice of an approximation to F is the plane going through $\underline{u}^*$ with gradient $\underline{\dot{g}}=-\underline{\alpha}$. $\underline{\alpha}$ is the unit normal vector of $\underline{u}^*$. If $g(\underline{u})$ is differentiable in $\underline{u}^*$, this plane corresponds to the tangent plane used in FORM. In this case

$$C\sim\frac{1}{N}\sum_{i=1}^{N}\frac{\Phi(-\|\underline{u}^*+b_i\underline{\alpha}\|)}{\Phi(-\beta)} \quad \cdots\cdots (10)$$

b_i is the root of $g(\underline{u}_i+b_i\underline{\alpha})=0$. It must be found by a suitable algorithm. The distance of a plane with gradient $-\underline{\alpha}$ through the solution point is $\|\underline{u}^*+b_i\underline{\alpha}\|$. The sampling points $\underline{u}_i$ can be generated in two different ways which yield very similar results. The first alternative is by eq. (6). The second alternative generates $n-1$ standard normal variables on the plane $\underline{\alpha}^T(\underline{u}-\underline{u}^*)=0$. The last variable is found by fulfilling that equation.

The estimate eq. (10) should have smaller variance than the ones discussed before because part of the integral is evaluated analytically (in the direction of $\underline{\alpha}$) and only the remaining dimensions are handled by Monte Carlo simulation. It should become exact with $N\to\infty$ because the variance of C then vanishes and the probability for the linearized surface is exact, i.e. :

$$P(A)=\Phi(-\beta) \quad \cdots\cdots (11)$$

If, on the other hand, a quadratic expansion

$$q(\underline{u})=(\underline{u}-\underline{u}^*)^T\underline{\underline{G}}(\underline{u}-\underline{u}^*)+2\underline{\dot{g}}^T(\underline{u}-\underline{u}^*)=0$$

with $\underline{\underline{G}}$ the Hessian matrix and $\underline{\dot{g}}$ the gradient of $g(\underline{u})$ in $\underline{u}^*$, respectively, is used it is easy to see that

$$C\sim\frac{1}{N}\sum_{i=1}^{N}\frac{\Phi(-\|\underline{u}^*+b_i\underline{\alpha}\|)}{\Phi(-\|\underline{u}^*+c_i\underline{\alpha}\|)} \quad \cdots\cdots (12)$$

where now the root of $q(\underline{u}_i+c_i\underline{\alpha})=0$ with respect to c_i must also be found. Unfortunately, exact results for the probability content of general quadratic forms in standard normal variates are difficult to obtain[12]. The most promising computation scheme has recently been proposed by Rice[13] and Helstrom[14] based on the inversion of the characteristic function of the distribution of a quadratic form. For convenience, those results are summarized in the appendix. For sufficiently small (or large) domain probabilities one may use the asymptotic approximation eq. (A·3) given in the appendix. Note that $q(\underline{u})=0$ needs not to be a quadratic Taylor expansion but can also be produced as a quadratic interpolation or even regression surface[15].

A variant of the quadratic expansion discussed before has been outlined in 6). It is based on the principal form of the fitting paraboloid of $g(\underline{u})$ in $\underline{u}^*$ which is also used for the derivation of the asymptotic correction factor eq. (4) for eq. (2). In the same reference it is shown that the corresponding correction factor for the probability of the fitting paraboloid

$$P(A)\sim\Phi(-\beta)\prod_{j=1}^{n-1}(1-\theta(-\beta)\kappa_j)^{-1/2}$$

is

$$C=\int\frac{\Phi(f(\underline{v}))}{\Phi(-\beta)}\frac{\varphi_{n-1}(\underline{v})}{\Psi_{n-1}(\underline{v})}\Psi_{n-1}(\underline{v})d\underline{v}\sim\frac{1}{N}\sum_{i=1}^{N}\frac{\Phi(f(\underline{v}_i))}{\Phi(-\beta)}\exp\left[-\frac{1}{2}\theta(-\beta)\sum_{j=1}^{n-1}\kappa_j v_{ij}^2\right] \quad \cdots\cdots (13)$$

where $\theta(-\beta)=\varphi(-\beta)/\Phi(-\beta)\sim\beta$, κ_j the main curvatures of $g(\underline{u})$ in $\underline{u}^*$ and the v_{ij} the components of the i-th sampling point. These components are independently, normally distributed with mean zero and standard deviation $(1-\theta(-\beta)\kappa_j)^{-1/2}$. $v_n=f(\underline{v})$ is the solution of $g(\underline{v},\ v_n)=0$ with $\underline{v}=(v_1,\ \cdots,\ v_{n-1})$ in a new coordinate system $(\underline{v},\ v_n)$ obtained by an orthogonal transformation such that in the new coordinate systems the β-point lies on the negative v_n-axis and all mixed derivatives of $g(\underline{v},\ v_n)=0$ in $\underline{v}^*=(0,\ \cdots,\ 0,\ -\beta)$ vanish. Note that in contrast to the method setting out from a linearization of $g(\underline{u})=0$ in $\underline{u}^*$ and where the probability content $P(A)$ is exact, no easily evaluated exact probability can be used for the quadratic approximation. (see, however, eq. (A·2)) Because additional curvature information is now used, that method is expected to be more efficient than the method using only gradient information.

3. EXAMPLE

It is not easy to construct test examples which have not only exact solutions but deviate significantly from plane or quadratic surfaces so that the merits of alternatives and/or the necessity of the proposed updates of FORM or SORM can be judged objectively. As an extreme example already used in 6) we take $F_x=\{g(\underline{X})=\pm\sum_{i=1}^{n} X_i \mp C\leq 0\}$ where the X_i are independently and identically exponentially distributed with parameter λ. Applying the necessary probability transformation produces a highly non-linear domain boundary, i. e.

$$F_u=\left\{ g(\underline{U})=\mp\frac{1}{\lambda}\sum_{i=1}^{n} \ln \Phi(-U_i)\mp C\leq 0\right\} \quad \cdots\cdots (14)$$

The exact result of eq. (14) is known to be $P_F=F_{Ga}(C; n, \lambda)$ with F_{Ga} the Gamma (Erlang) distribution function. The negative sign in eq. (14) corresponds to the lower tail of F_{Ga} and the positive sign to the upper tail of F_{Ga}, respectively.

The three methods, namely method A (eq. (5), (6)), method B (eq. (8), (10)) and method C (eq. (8), (13)), are applied to eq. (14) in order to compare their results in terms of the equivalent safety index $\beta_E=-\Phi^{-1}[P(F)]$ and the coefficient of variation (COV) of the probability estimates. The comparison is performed for three different probability levels ($P_t=10^{-2}$, 10^{-4}, 10^{-6}) corresponding to the

Table 1 β and COV for upper tail of F_{Ga} (Sample size : $N=50$).

β_t \ n			2	5	10	20
2.327	β_I (FORM)		2.541	2.889	3.239	3.709
	β_{II} (SORM)		2.242	2.101	1.899	1.485
	A	β	2.638	2.335	2.492	2.791
		COV(%)	27.23	30.16	27.25	31.41
	B	β	2.388	2.397	2.390	2.434
		COV(%)	12.25	16.82	14.25	13.24
	C	β	2.339	2.295	2.323	2.303
		COV(%)	6.08	7.77	10.86	14.12
3.722	β_I (FORM)		3.918	4.255	4.601	5.069
	β_{II} (SORM)		3.654	3.546	3.400	3.126
	A	β	4.000	3.667	3.948	4.146
		COV(%)	35.47	45.69	29.93	39.03
	B	β	3.789	3.816	3.828	3.884
		COV(%)	16.99	23.68	19.24	17.86
	C	β	3.727	3.692	3.715	3.701
		COV(%)	6.61	8.74	12.58	16.56
4.756	β_I (FORM)		4.942	5.270	5.612	6.078
	β_{II} (SORM)		4.698	4.605	4.480	4.254
	A	β	5.013	4.894	4.964	5.233
		COV(%)	41.36	33.16	33.48	43.60
	B	β	4.829	4.872	4.897	4.963
		COV(%)	19.11	27.12	22.09	20.51
	C	β	4.760	4.729	4.749	4.737
		COV(%)	6.88	9.32	13.67	18.29

target safety indices β_t=2.327, 3.722, 4.756 and four different dimensions of $\underline{X}$ (n=2, 5, 10, 20). For comparison, the sample size in all three methods is 50 and λ=1.0.

The results are shown in Table 1 for the case of the upper tail of F_{Ga} and in Table 2 for the case of the lower tail of F_{Ga}, respectively. In the tables the safety indices calculated according to FORM and SORM are also shown for comparison.

As expected the results according to FORM and SORM become worse with decreasing target safety index and increasing dimension n. As it should be, SORM produces almost exact results for large target safety indices. Method A produces the largest coefficient of variation. Because method A only needs the location of the β-point and only simple function calls during simulation, it must, nevertheless, be considered a rather robust method. Roughly half of that coefficient of variation is observed for method B which can be explained by the fact that it uses the precise locations of the failure surface for the simulation points. Because the corresponding probability $P(A)$ is exact and the variability of the correction factor vanishes as $N \to \infty$, it also reproduces the exact probability for any failure domain. Method C has the smallest coefficient of variation because it further uses curvature information of the failure surface in the β-point. Because $P(A)$ in eq. (13) is only accurate for $\beta \to \pm\infty$, it is especially suited for large or small probability domains.

A comparison of these methods should further be carried out with regard to the numerical effort. A

Table 2 β and COV for lower tail of F_{Ga} (Sample size : N=50).

β_t		n	2	5	10	20
2.327	β_I (FORM)		2.070	1.684	1.319	0.841
	β_{II} (SORM)		2.330	2.299	2.223	2.047
	A	β	2.280	2.381	2.414	2.227
		COV(%)	23.16	32.76	43.55	49.40
	B	β	2.306	2.301	2.319	2.354
		COV(%)	8.17	13.45	14.70	15.25
	C	β	2.326	2.329	2.347	2.328
		COV(%)	0.43	3.11	4.19	5.72
3.722	β_I (FORM)		3.469	3.068	2.692	2.209
	β_{II} (SORM)		3.731	3.723	3.680	3.567
	A	β	3.724	3.752	3.891	3.965
		COV(%)	27.37	44.31	52.94	69.34
	B	β	3.714	3.703	3.741	3.788
		COV(%)	12.05	22.61	21.44	26.47
	C	β	3.719	3.722	3.741	3.727
		COV(%)	1.54	5.08	4.28	5.85
4.756	β_I (FORM)		4.514	4.103	3.718	3.228
	β_{II} (SORM)		4.768	4.764	4.744	4.662
	A	β	4.776	4.879	4.905	5.003
		COV(%)	30.23	65.37	56.13	69.40
	B	β	4.754	4.739	4.816	4.861
		COV(%)	14.75	31.20	26.37	38.35
	C	β	4.753	4.757	4.777	4.765
		COV(%)	2.09	7.04	4.58	6.07

Table 3 Number of g-function calls (n=5, β_t=2.327).

	method	N	g-function call No.*	β	COV(%)
upper tail	FORM	0	60	2.889	-
	SORM	0	70	2.101	-
	A	2100	2160	2.348	7.68
	B	200	1260	2.389	7.65
	C	50	370	2.295	7.77
lower tail	FORM	0	60	1.684	-
	SORM	0	70	2.299	-
	A	7000	7060	2.331	3.06
	B	1000	6060	2.305	3.02
	C	50	370	2.329	3.11

* K_1 = 10, K_2 = 3 are assumed.

convenient measure of the effort is the number of g-function calls because these usually are most time-consuming in practical applications. Let K_1 be the number of iterations in FORM using a gradient algorithm and K_2 the number of iterations for the line search algorithm in method B and C. The numerical effort, then, is approximately :

$$\left.\begin{array}{ll} \text{FORM} & K_1(n+1) \\ \text{SORM} & K_1(n+1)+n(n-1)/2 \\ \text{Method A}: & K_1(n+1) \quad +N \\ \text{Method B}: & K_1(n+1) \quad +2NK_2 \\ \text{Method C}: & K_1(n+1)+n(n-1)/2+2NK_2 \end{array}\right\} \quad \cdots\cdots (15)$$

K_1, typically, is around 10 whereas K_2 is 2 or 3 depending on the initial solution of the algorithm. The analysis of the example now shows that the effort of method A and of method B is approximately the same for the same accuracy of the probability estimate (see Table 3). It can be seen that the number of function calls for method B is roughly one half of method A for the upper tail and the same for the lower tail. The numerical effort for method C is in the order of 10 % of the other methods for small dimension n. For very large problem dimension method C must, however, become inefficient due to its quadratic increase of function calls which in the other methods is only linear. Table 3 further demonstrates that FORM or SORM are much less expensive than any importance sampling simulation method.

4. SUMMARY AND CONCLUSIONS

Three importance sampling methods replacing or updating FORM or SORM probability estimates are compared at an extreme example. Method A may be called direct estimation and uses only the location of the β-point. Method B uses gradient information and involves some direct probability integration. Method C additionally uses curvature information. It was found that this method is the most efficient for not too high-dimensional uncertainty vectors and extreme probabilities. Method A or B with slight preference to B due to its higher efficiency should be applied in very high dimensional problems irrespective of the magnitude of the probabilities. In any case the updating by simulation requires much more effort than simple FORM or SORM estimates, so that such updates should only be made if there are serious doubts about the validity of FORM/SORM results.

APPENDIX : PROBABILITY CONTENT OF QUADRATIC FORMS

Let a quadratic form be given in its principal form which always can be achieved by suitable orthogonal transformations[12] :

$$Q=q(\underline{U})=\sum_{i=1}^{n}\lambda_i(u_i-\delta_i)^2 \quad \cdots\cdots (A\cdot1)$$

Helstrom[14] gave the following formula for the probability $P(Q>k)$ based on Rice[13] :

$$P(Q>k)=\int_{c-i\infty}^{c+i\infty}\exp[h(u)/(2\pi i)]du \quad \cdots\cdots (A\cdot2)$$

$$\sim\left\{\begin{matrix}0 \text{ for } u^*>0\\ 1 \text{ for } u^*<0\end{matrix}\right\}+(2\pi h''(u^*))^{-1/2}\exp[h(u^*)] \quad \cdots\cdots (A\cdot3)$$

where u^* is the (numerical) solution of $h'(u)=0$ and

$$h(u)=f(u)-uk-\ln u \quad \cdots\cdots (A\cdot4)$$

with

$$f(u)=\frac{1}{2}\sum_{i=1}^{n}[\delta_i^2(1-2u\lambda_i)^{-1}-\delta_i^2-\ln(1-2u\lambda_i)] \quad \cdots\cdots (A\cdot5)$$

the cumulant generating function with $\arg(\ln u)=0$ at $u=c$. For solving $h'(u)=0$, the starting value $u^{(0)}$ should be chosen such that $u^{(0)}<E[Q]$ for $E[Q]\leq k$ and $u^{(0)}>E[Q]$ otherwise with $E[Q]=\sum_{i=1}^{n}\lambda_i(1+\delta_i^2)$. Formula (A·3) is an asymptotic approximation.

REFERENCES

1) Hohenbichler, M. and Rackwitz, R. : Non-Normal Dependent Vectors in Structural Safety, Journ. of the Eng. Mech., ASCE, Vol. 107, No. 6, pp. 1227~1249, 1981.
2) Breitung, K. : Asymptotic Approximations for Multinormal Domain and Surface Integrals, Proc. ICASP 4, Florence, Pitagora Ed., Bologna, pp. 755~768, 1983.
3) Shinozuka, M. : Basic Analysis of Structural Safety, Journ. of Struct. Eng., ASCE, Vol. 109, 3, pp. 721~740, 1983.
4) Harbitz, A. : Efficient and Accurate Probability of Failure Calculation by Use of the Importance Sampling Technique, Proc. ICASP-4, Florence, Pitagora Ed., Bologna, pp. 825~836, 1983.
5) Hohenbichler, M. : An Asymptotic Formula for the Probability of Intersections, in : Berichte zur Zuverlaessigkeitstheorie der Bauwerke, Technische Universitaet Muenchen, SFB 96, Heft 69, pp. 21~48, 1984.
6) Hohenbichler, M. and Rackwitz, R. : Improvement of Second-order Reliability Estimates by Importance Sampling, Journ. of the Eng. Mech., ASCE, submitted. 1986.
7) Melchers, R. : Efficient Monte Carlo Probability Integration, Res. Rep. No. 7, Monash University, 1984.
8) Ditlevsen, O., et al. : Clipped Normal FBC Load Combination by Deak Simulation, DCAMM Reports, Technical University of Denmark, 1985.
9) Harbitz, A. : An Efficient Sampling Method for Probability of Failure Calculation, Structural Safety, Jan., 1986.
10) Bjerager, P. : Probability Integration by Directional Simulation, Journ. of the Eng. Mech., ASCE, submitted 1986.
11) Rubinstein, R. Y. : Simulation and the Monte Carlo Method, J. Wiley & Son, New York, 1981.
12) Fiessler, B., Neumann, H.-J. and Rackwitz, R. : Quadratic Limit States in Structural Reliability, Journ. of the Eng. Mech., ASCE, Vol. 105, No. 4, pp. 661~676, 1979.
13) Rice, S. O. : Distribution of Quadratic Forms in Normal Random Variables-Evaluation by Numerical Integration, Soc. for Industrial and Applied Mathematics, J. Sci. Stat. Comp., 1, 4, pp. 438~448, 1981.
14) Helstrom, C. W. : Comment to Rice, S. O., Distribution of Quadratic Forms in Normal Random Variables-Evaluation by Numerical Integration, Society for Industrial and Applied Mathematics, J. Sci. Stat. Comp., 4, 2, pp. 353~356, 1983.
15) Rackwitz, R. : Response Surfaces in Structural Reliability, in : Berichte zur Zuverlaessigkeitstheorie der Bauwerke, Technische Universitaet Muenchen, SFB 96, Heft 67, 1982.

(Received January 6 1987)