Optimal Importance-Sampling Density Estimator

By George L. Ang,[1] Alfredo H-S. Ang,[2] and Wilson H. Tang[3]

Abstract: Importance-sampling technique has been used in recent years in conjunction with Monte Carlo simulation method to evaluate the reliability of structural systems. Since the efficiency of the importance-sampling method depends primarily on the choice of the importance-sampling density, the use of the kernel method to estimate the optimal importance-sampling density is proposed. This method deviates from the current practice of prescribing the importance-sampling density from a given parametric family of density functions. Instead, the data obtained from an initial Monte Carlo run are utilized to determine the required importance-sampling density. The kernel method yields unbiased estimates of the probability of failure. Two measures are developed to quantify the efficiency of the kernel method relative to the basic Monte Carlo method. The first measure, called the *marginal efficiency*, is used as an indicator of the effectiveness of the kernel method, whereas the second measure, the *overall efficiency*, defines the advantage of the kernel method over the basic Monte Carlo method. Finally, a variety of example problems are used to examine the characteristics of the proposed kernel method and its efficiency over the basic Monte Carlo method.

Introduction

In structural reliability theory, the calculation of the probability of failure, p_F, of an engineering system requires the evaluation of the following multidimensional integral:

$$p_F = P[g(\mathbf{x}) \leq 0] = \int_{g(\mathbf{x})\leq 0} f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

where $\mathbf{X}$ = a vector of random variables (bold is used to denote a vector, and a random variable is indicated by an upper-case character); $g(\mathbf{x})$ = the performance function of the system. For any realization $\mathbf{x}$ of the random variables $\mathbf{X}$, $g(\mathbf{x}) \leq 0$ implies the system has failed; and $f_{\mathbf{X}}(\mathbf{x})$ = the joint probability density function of the random variables $\mathbf{X}$.

In general $f_{\mathbf{X}}(\mathbf{x})$ is a multidimensional density function with complex mathematical form, and the failure domain $g(\mathbf{x}) \leq 0$ is irregular in shape with piecewise, highly nonlinear boundaries. Hence, an analytical evaluation of the failure probability defined by (1) is usually impossible. This leads to the need for approximate approaches.

Among the approximate methods, a class called the Monte Carlo sampling method is widely used (Rubinstein 1981). The basic Monte Carlo method is a numerical sampling technique that simulates a process involving realizations of random variables, and determining if a particular event (e.g., failure) occurs for each simulation. The ratio of the number of failures to the total number of simulations is an estimate of the failure probability.

[1]Engr., Hibbitt, Karlsson & Sorensen, Inc., Pawtucket, RI 02860.
[2]Prof., Dept. of Civ. Engrg., Univ. of California, Irvine, CA 92717.
[3]Prof., Dept. of Civ. Engrg., Univ. of Illinois, Urbana, IL 61801.

Note. Discussion open until November 1, 1992. To extend the closing date one month, a written request must be filed with the ASCE Manager of Journals. The manuscript for this paper was submitted for review and possible publication on February 12, 1991. This paper is part of the ***Journal of Engineering Mechanics***, Vol. 118, No. 6, June, 1992. ©ASCE, ISSN 0733-9399/92/0006-1146/$1.00 + $.15 per page. Paper No. 1380.

Because low failure probabilities are typically expected or required of structural systems, the total number of simulations necessary to obtain a sufficient number of "failure samples" can be extremely large, and for most realistic problems these methods are not affordable. To reduce the number of simulations required, a variance reduction technique called *importance sampling* has been introduced (Bourgund and Bucher 1986; Harbitz 1986; Schueller and Stix 1987). In this paper, however, the importance-sampling density is not prescribed to be from a given family of distributions (Schueller and Stix 1987). Instead, data obtained from an initial basic Monte Carlo run are entirely utilized to determine the required importance-sampling density. Following sections present the concepts underlying the importance-sampling method in reliability computation, followed by the development of the bases for the kernel method of constructing the importance-sampling density. The kernel method was described by the writers in an earlier paper (Ang et al. 1989). Karamchandani (1990) also presented a somewhat similar procedure, i.e., similar in the sense that the importance-sampling density is constructed as a weighted sum of probability densities. In this paper, two measures are developed to quantify the efficiency of the kernel method relative to the basic Monte Carlo method. A few examples are presented to illustrate the applicability of the kernel method.

Importance-Sampling Technique

The problem in the basic Monte Carlo method is that when the probability of failure is small, most of the simulations will not result in system failure. Therefore, the number of simulations required to obtain a sufficient number of failure samples can be extremely large, and for most realistic problems the method can become unaffordable. One solution to this problem is to bias the generated realizations of the random variables using an importance-sampling density function such that the number of occurrences of system failure will increase. The results are then scaled to account for the bias in the sampling distribution. The mathematical basis for this latter technique may be described as follows.

The integral in (1) can be rewritten as

$$p_F = \int_{\text{all } \mathbf{x}} I[g(\mathbf{x}) \leq 0] f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x} \quad \ldots\ldots (2)$$

where

$$I[g(\mathbf{x}) \leq 0] = 0 \text{ if } g(\mathbf{x}) > 0 \quad \ldots\ldots (3a)$$

$$I[g(\mathbf{x}) \leq 0] = 1 \text{ if } g(\mathbf{x}) \leq 0 \quad \ldots\ldots (3b)$$

An importance-sampling density, $h_{\mathbf{X}}(\mathbf{x})$, can be introduced into (2) to obtain

$$p_F = \int_{\text{all } \mathbf{x}} I[g(\mathbf{x}) \leq 0] \frac{f_{\mathbf{X}}(\mathbf{x})}{h_{\mathbf{X}}(\mathbf{x})} h_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x} \quad \ldots\ldots (4)$$

A Monte Carlo algorithm to evaluate the integral in (4) would be to sample a series of $\mathbf{x}_i$ from $h_{\mathbf{X}}(\mathbf{x})$ and esimate p_F through

$$\hat{p}_F = \frac{1}{N} \sum_{i=1}^{N} I[g(\mathbf{x}_i) \leq 0] \frac{f_{\mathbf{X}}(\mathbf{x}_i)}{h_{\mathbf{X}}(\mathbf{x}_i)} \quad \ldots\ldots (5)$$

A caret above p_F is used to denote that it is an estimate of p_F. The variance of $\hat{p}_F$ is given by

$$\text{var}_{h_{\mathbf{X}}(\mathbf{x})}(\hat{p}_F) = \frac{1}{N}\left[\int_{\text{all } \mathbf{x}} \left\{I[g(\mathbf{x}) \le 0]\frac{f_{\mathbf{X}}(\mathbf{x})}{h_{\mathbf{X}}(\mathbf{x})}\right\}^2 h_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x} - p_F^2\right] \quad \ldots\ldots (6)$$

Observe that if the importance-sampling density is chosen as (Rubinstein 1981)

$$h_{\mathbf{X}}(\mathbf{x}) = \frac{I[g(\mathbf{x}) \le 0]\, f_{\mathbf{X}}(\mathbf{x})}{p_F} \quad \ldots\ldots (7)$$

then the variance of $\hat{p}_F$ becomes zero, regardless of the number of simulations. However, since p_F is not known a priori, it is impossible to obtain the optimal importance-sampling density, $h_{\mathbf{X}}(\mathbf{x})$. Therefore, the knowledge of the form of the optimal importance-sampling density defined by (7) is at best only a guide toward effective sampling.

The importance-sampling density should be selected properly to yield reliable estimates of p_F with ony a small number of samples. If a "bad" importance-sampling density is chosen, the procedure can be very inefficient, i.e., it can require a large number of samples for a reliable estimate of p_F, and the resulting $\hat{p}_F$ may be biased. For example, if the importance-sampling density $h_{\mathbf{X}}(\mathbf{x})$ is chosen such that it emphasizes only a small part of the failure domain, as illustrated in Fig. 1, the estimate of p_F using this procedure will usually be less than the true value of p_F. Caution should be exercised when using the estimated standard deviation, $\hat{\sigma}_{\hat{p}_F}$, as a measure of the accuracy of the estimator $\hat{p}_F$, because a low estimated $\hat{\sigma}_{\hat{p}_F}$ may be an indication of potential bias in $\hat{p}_F$ (Ibrahim and Cornell 1988). A properly chosen sampling density is essential to obtain a good (unbiased) estimate with a low $\hat{\sigma}_{\hat{p}_F}$. In the following sections, a method of determining the importance-sampling density $h_{\mathbf{X}}(\mathbf{x})$ is developed.

KERNEL SAMPLING DENSITY

The kernel sampling density is defined by

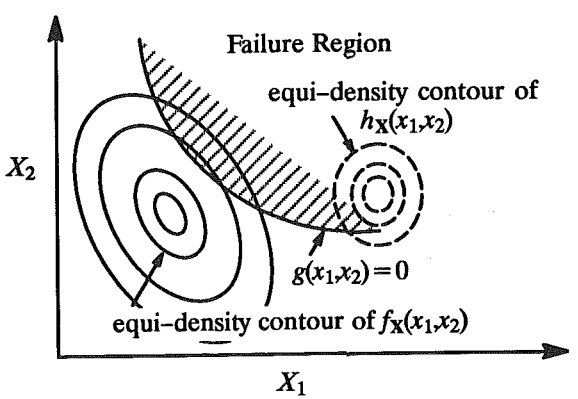

FIG. 1. Importance-Sampling Density Function $h_{\mathbf{X}}(x_1,x_2)$ Covering Small Part of Failure Region

$$\hat{h}_{\mathbf{x}}(\mathbf{x}) = \frac{1}{M}\sum_{i=1}^{M}\frac{1}{w^d}K\left(\frac{\mathbf{x}-\mathbf{y}_i}{w}\right) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (8)$$

where M = the number of samples; $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_M\}$ = the samples; w = the window width; d = the number of variables, and $K(.)$ = any function that satisfies the conditions

$$K(\mathbf{y}) \geq 0 \text{ and } \int_{-\infty}^{\infty} K(\mathbf{y})\, d\mathbf{y} = 1 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (9)$$

With the conditions in (9), the kernel sampling density is itself a probability density, and it will inherit all the continuity and differentiability properties of the kernel function $K(.)$. If $d = 1$ (one random variable case) and $K(.)$ is the standard normal density function, then the kernel sampling density is given by

$$\hat{h}_X(x) = \frac{1}{M}\sum_{i=1}^{M}\frac{1}{w}\left[\frac{1}{\sqrt{2\pi}}e^{-1/2(x-y_i/w)^2}\right] \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (10)$$

This is illustrated in Fig. 2, where the sum of all the individual "bumps," $M^{-1}w^{-1}K[(x - y_i)/w]$, yields the estimator, $\hat{h}_X(x)$. Also notice that the individual bumps are of equal shape. Consequently, if (8) is used to estimate a long-tailed distribution, $\hat{h}_X(x)$ will have a rough tail, as shown in Fig. 3. This difficulty may be resolved by introducing λ_is to vary the shape of each bump. The main idea is to use flatter bumps in regions of low density (Abramson 1982). The general strategy is as follows:

1. Use the sampling density defined in (8) as an initial estimate of the optimal sampling density, and denote this by $\tilde{h}_{\mathbf{x}}(\mathbf{x})$.
2. Define the local window width factors, λ_i, by

$$\lambda_i = [\tilde{h}_{\mathbf{x}}(\mathbf{y}_i)]^{-\alpha} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (11)$$

where α = the sensitivity parameter, a number between 0 and 1. A value of $\alpha = 0.5$ is suggested (see the following).
3. Define the adaptive kernel sampling density as

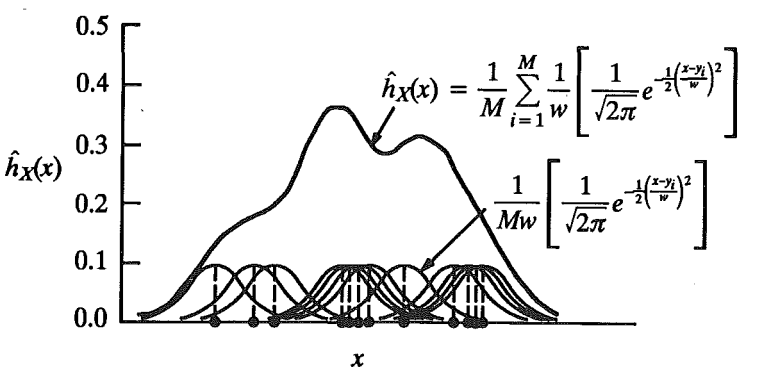

FIG. 2. Kernel Sampling Density and Individual Kernels

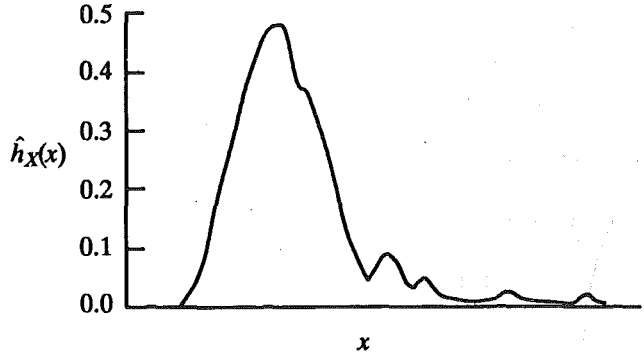

FIG. 3. Kernel Sampling Density with Rough Tail

$$\hat{h}_{\mathbf{X}}(\mathbf{x}) = \frac{1}{M}\sum_{i=1}^{M}\frac{1}{(w\lambda_i)^d}K\left(\frac{\mathbf{x}-\mathbf{y}_i}{w\lambda_i}\right) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (12)$$

The sensitivity factor α is introduced to give flexibility in the design of the method. The value of $\alpha = 0$ will reduce (12) back to (8); whereas $\alpha = 1$ will ensure that the number of observations within each window width will be approximately the same in all parts of the density. To see this, substitute $\alpha = 1$ into (11) and multiply the result by w to obtain the window width $w/\hat{h}_{\mathbf{X}}(\mathbf{y}_i)$ centered at $\mathbf{y}_i$. The mean number of observations within this window width centered at $\mathbf{y}_i$ is proportional to $[w/\hat{h}_{\mathbf{X}}(\mathbf{y}_i)] \times \hat{h}_{\mathbf{X}}(\mathbf{y}_i) = w$, which is a constant. Practical experience, reported by Abramson (1982), suggests that the choice of $\alpha = 0.5$ gives good results because the bias in $\hat{h}_{\mathbf{X}}(\mathbf{x})$ is minimized. Finally, it has been confirmed, also by Abramson (1982), that the method is insensitive to the choice of the initial estimate of $\hat{h}_{\mathbf{X}}(\mathbf{x})$ in step 1.

The effect of varying the window width w is illustrated in Fig. 4 for the one random variable case and all λ_is $= 1.0$. The limit as w tends to zero is (in a sense) a sum of Dirac delta function spikes at each sample point, as illustrated in Fig. 4(a), whereas as w becomes large, all details are obscured, as shown in Fig. 4(b), i.e., $\hat{h}_X(x)$ is smoother. The appropriate choice of the value of w will always be influenced by the purpose for which $\hat{h}_{\mathbf{X}}(\mathbf{x})$ is to be used. Two criteria for choosing the value of w are discussed in the following.

Minimizing Integrated Square Error (Abbreviated ISE)

Given an estimator $\hat{h}_{\mathbf{X}}(\mathbf{x})$ of $h_{\mathbf{X}}(\mathbf{x})$, the integrated square error is defined by

$$\mathrm{ISE}(\hat{h}_{\mathbf{X}}) = \int [\hat{h}_{\mathbf{X}}(\mathbf{x}) - h_{\mathbf{X}}(\mathbf{x})]^2\, d\mathbf{x} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (13a)$$

$$\mathrm{ISE}(\hat{h}_{\mathbf{X}}) = \int [\hat{h}^2_{\mathbf{X}}(\mathbf{x}) - 2\hat{h}_{\mathbf{X}}(\mathbf{x})h_{\mathbf{X}}(\mathbf{x}) + h^2_{\mathbf{X}}(\mathbf{x})]\, d\mathbf{x} \quad \ldots\ldots\ldots\ldots\ldots\ldots (13b)$$

Therefore, a choice of window width w can be obtained by minimizing the first two terms in (13), since last term does not depend on w. Denoting the first two terms by $R(.)$ and substituting (12) into the integer, we can obtain [see Ang (1991)]

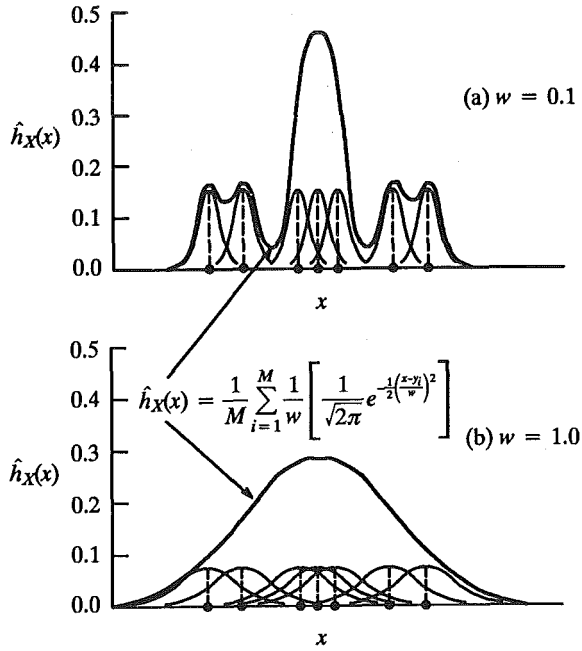

FIG. 4. Effect of Varying Window Width on Kernel Sampling Density

$$R(\hat{h}_{\mathbf{x}}) = \frac{1}{M^2}\sum_{i,j} K^{(2)}(\mathbf{y}_i,\mathbf{y}_j;\, \lambda_i w,\lambda_j w) - \frac{2}{M}\sum_{i=1}^{M} \hat{h}_{-i}(\mathbf{y}_i) \quad \ldots\ldots\ldots\ldots (14)$$

where

$$\hat{h}_{-i}(\mathbf{y}_i) = \frac{1}{M-1}\sum_{\substack{j=1\\ j\neq i}}^{M} \frac{1}{(w\lambda_j)^d} K\left(\frac{\mathbf{y}_i - \mathbf{y}_j}{w\lambda_j}\right) \quad \ldots\ldots\ldots\ldots\ldots\ldots (15)$$

If $K(.)$ is the multidimensional normal density

$$K(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d|\mathbf{S}|}}\, e^{-1/2\mathbf{x}^T\mathbf{S}^{-1}\mathbf{x}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (16)$$

then

$$K^{(2)}(\mathbf{y}_i,\mathbf{y}_j;\, \lambda_i w,\lambda_j w) = \frac{1}{\sqrt{(2\pi)^d|\mathbf{S}|[w^2(\lambda_i^2+\lambda_j^2)]^d}}\, e^{-(\mathbf{y}_i-\mathbf{y}_j)^T\mathbf{S}^{-1}(\mathbf{y}_i-\mathbf{y}_j)/2w^2(\lambda_i^2+\lambda_j^2)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (17)$$

where $\mathbf{S}$ = the sample covariance matrix of the sample set $\{\mathbf{y}_1,\mathbf{y}_2, \ldots ,\mathbf{y}_M\}$; and $|\mathbf{S}|$ is the determinant of $\mathbf{S}$. The sample covariance matrix is introduced in (16) to take into account the spread of data in each direction and the coordinate scale used for each random variable. Therefore, to obtain w using this criterion, minimize (14) with respect to the window width.

Minimizing Variance of $\hat{p}_F$

Since $\hat{h}_{\mathbf{X}}(\mathbf{x})$ is used to estimate p_F, it is logical to choose w that will minimize the variance of $\hat{p}_F$ defined by replacing $h_{\mathbf{X}}(\mathbf{x})$ in (6) with $\hat{h}_{\mathbf{X}}(\mathbf{x})$, i.e.

$$\mathrm{var}_{\hat{h}_{\mathbf{X}}(\mathbf{x})}(\hat{p}_F) = \frac{1}{N}\left[\int_{-\infty}^{\infty}\left\{I[g(\mathbf{x}) \le 0]\frac{f_{\mathbf{X}}(\mathbf{x})}{\hat{h}_{\mathbf{X}}(\mathbf{x})}\right\}^2 \hat{h}_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x} - p_F^2\right] \quad \ldots\ldots (18)$$

Substitute (7) into (18) to obtain [see Ang (1991)]

$$\hat{\mathrm{var}}_{\hat{h}_{\mathbf{X}}(\mathbf{x})}(\hat{p}_F) = \frac{1}{N}\hat{p}_F(\hat{V}_{\mathrm{kernel}} - \hat{p}_F) \quad \ldots\ldots (19)$$

where

$$\hat{V}_{\mathrm{kernel}} = \frac{1}{M}\sum_{i=1}^{M}\frac{f_{\mathbf{X}}(\mathbf{y}_i)}{\hat{h}_{-i}(\mathbf{y}_i)} \quad \ldots\ldots (20)$$

and $\hat{h}_{-i}(\mathbf{y}_i)$ is defined in (15). The window width, w, may then be selected to minimize (19). It is shown in Ang (1991) that minimizing the variance of $\hat{p}_F$ is similar to minimizing the integrated square error given by (13), except for a weighting factor $1/\hat{h}_{\mathbf{X}}(\mathbf{x})$. In regions where the value of $\hat{h}_{\mathbf{X}}(\mathbf{x})$ is smaller than $h_{\mathbf{X}}(\mathbf{x})$, the contribution of $[h_{\mathbf{X}}(\mathbf{x}) - \hat{h}_{\mathbf{X}}(\mathbf{x})]$ to the expression for the variance is greater because the weight, $1/\hat{h}_{\mathbf{X}}(\mathbf{x})$, is large. Consequently, minimizing the variance with respect to the window width would result in $\hat{h}_{\mathbf{X}}(\mathbf{x})$ that approximates $h_{\mathbf{X}}(\mathbf{x})$ better in regions where the value of $\hat{h}_{\mathbf{X}}(\mathbf{x})$ is small.

In this paper, the kernel sampling density defined by (12) is proposed to estimate the optimal importance sampling, $h_{\mathbf{X}}(\mathbf{x})$, defined in (7).

Estimating Failure Probability

The two-step approach to estimate p_F using the adaptive kernel sampling density is as follows:

1. Construct the adaptive kernel sampling density as follows.
   A. A sample of size M is generated using the acceptance-rejection sampling algorithm presented. The essence of the algorithm can be described as follows: A sample $\mathbf{x}$ is generated from the original probability density function $f_{\mathbf{X}}(\mathbf{x})$; if the inequality $g(\mathbf{x}) \le 0$ holds, then accept $\mathbf{x}$ as a sample. Accepted samples will be denoted by $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_M\}$. The total number of samples generated to obtain the M samples is denoted by N_{basic}.

      An initial estimate for the failure probability denoted by $(\hat{p}_F)_{\mathrm{basic}}$ and an estimate of its sample variance, denoted by $\hat{\sigma}^2_{\mathrm{basic}}$ are also computed using

$$(\hat{p}_F)_{\mathrm{basic}} = \frac{M}{N_{\mathrm{basic}}} \quad \ldots\ldots (21)$$

      and

$$\hat{\sigma}^2_{\text{basic}} = N_{\text{basic}} \text{ var } [(\hat{P}_F)_{\text{basic}}] = \frac{1}{N_{\text{basic}} - 1}$$

$$\cdot [M - N_{\text{basic}}(\hat{p}_F)^2_{\text{basic}}] \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (22)$$

B. These M samples are then used to compute the window width factors, $\{\lambda_1, \lambda_2, \ldots, \lambda_M\}$, as follows.
   1. Use the density defined in (8) as an initial estimate of the optimal sampling density, and denote this by $\hat{h}_{\mathbf{X}}(\mathbf{x})$.
   2. Compute the local window width factors

$$\lambda_i = [\hat{h}_{\mathbf{X}}(\mathbf{y}_i)]^{-0.5} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (23)$$

   for $i = 1,2, \ldots, M$.

C. The optimal window width is computed by minimizing the ISE's R function or the variance of $\hat{p}_F$ with respect to w. The ISE's R function is given by (14) and the variance of $\hat{p}_F$ is given by (19).

D. The adaptive kernel sampling density is then given by (12), where d = the number of random variables.

2. Generate samples using the constructed adaptive kernel sampling density, and estimate p_F as follows:

A. A sample of size N_{kernel} is generated from the kernel sampling density $\hat{h}_{\mathbf{X}}(\mathbf{x})$ defined previously. These samples, denoted by $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{N_{\text{kernel}}}\}$, are generated from the kernel sampling density by first generating a number at random from $\{1, 2, \ldots, M\}$, say j, then sampling from the density function

$$\frac{1}{(w\lambda_j)^d} \mathbf{K}\left(\frac{\mathbf{x} - \mathbf{y}_j}{w\lambda_j}\right) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (24)$$

B. Compute

$$(\hat{p}_F)_{\text{kernel}} = \frac{1}{N_{\text{kernel}}} \sum_{i=1}^{N_{\text{kernel}}} I[g(\mathbf{x}_i) \leq 0] \frac{f_{\mathbf{X}}(\mathbf{x}_i)}{\hat{h}_{\mathbf{X}}(\mathbf{x}_i)} \quad \ldots\ldots\ldots\ldots (25)$$

and its variance

$$\text{var}[(\hat{p}_F)_{\text{kernel}}] = \frac{1}{N_{\text{kernel}}(N_{\text{kernel}} - 1)} \sum_{i=1}^{N_{\text{kernel}}}$$

$$\cdot \left\{ I[g(\mathbf{x}_i) \leq 0] \frac{f_{\mathbf{X}}(\mathbf{x}_i)}{\hat{h}_{\mathbf{X}}(\mathbf{x}_i)} - \hat{p}_F \right\}^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots (26)$$

C. The failure probability is then estimated as

$$\hat{p}_F = \frac{N_{\text{basic}}(\hat{p}_F)_{\text{basic}} + N_{\text{equiv}}(\hat{p}_F)_{\text{kernel}}}{N_{\text{basic}} + N_{\text{equiv}}} \quad \ldots\ldots\ldots\ldots\ldots (27)$$

where

$$N_{\text{equiv}} = N_{\text{kernel}} \frac{\hat{\sigma}^2_{\text{basic}}}{\hat{\sigma}^2_{\text{kernel}}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (28a)$$

$$\hat{\sigma}^2_{\text{kernel}} = N_{\text{kernel}} \text{ var}[(\hat{p}_F)_{\text{kernel}}] \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (28b)$$

in which $\hat{\sigma}^2_{\text{basic}}$ and $\hat{\sigma}^2_{\text{kernel}}$ are the estimated sample variances of the failure probability obtained using the basic Monte Carlo and kernel methods, respectively.

The estimator $\hat{p}_F$ defined by (27) utilizes the information from the initial simulation (to obtain the samples used to construct the kernel sampling density) and that from the simulation runs using the kernel sampling density. This estimator is found to be unbiased, with a smaller variance than that of $(\hat{p}_F)_{\text{kernel}}$ for small sample sizes N_{kernel} (Ang 1991).

Efficiency of Kernel Method

To assess the efficiency of the kernel method relative to the basic Monte Carlo, two measures are developed. The first measure is the marginal efficiency, E_{marginal}. This measure can be used to indicate the effectiveness of the kernel method after a sample size M has been generated. It can be obtained by equating the variance estimate of $\hat{p}_F$ by the basic Monte Carlo and (19), and solving for the sample ratio, i.e., the marginal efficiency is defined by

$$E_{\text{marginal}} = \frac{N_{\text{basic}}}{N_{\text{kernel}}} = \frac{\hat{V}_{\text{basic}} - \hat{p}_F}{\hat{V}_{\text{kernel}} - \hat{p}_F} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (29)$$

where $\hat{V}_{\text{kernel}}$ is given by (20) and

$$\hat{V}_{\text{basic}} = \frac{1}{M} \sum_{i=1}^{M} I[g(\mathbf{y}_i) \leq 0] = 1 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (30)$$

To obtain the same variance of $\hat{p}_F$ between the kernel method and the basic Monte Carlo method, (29) can be interpreted as the equivalent number of basic Monte Carlo samples for each sample generated using the kernel method.

The second measure is the overall efficiency

$$E_{\text{overall}} = \frac{\text{No. of } g(\mathbf{x}) \text{ calls using basic Monte Carlo}}{\text{No. of } g(\mathbf{x}) \text{ calls using kernel method}}$$

$$= \frac{N_{\text{total}}}{N_{\text{basic}} + N_{\text{kernel}}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (31)$$

where N_{total} = used to denote the sum $N_{\text{basic}} + (N_{\text{basic}})_2$. The variable $(N_{\text{basic}})_2$ denotes the number of additional basic Monte Carlo simulation runs required to obtain the same accuracy of p_F as the kernel method. This measure is dependent on the required accuracy $\delta_{\hat{p}_F}$ and sample size M. The accuracy, defined as the coefficient of variation of $\hat{p}_F$, affects the total number of simulations required, whereas M affects the value of N_{kernel}. The significance of this measure is that a value of $E_{\text{overall}} = 5$ means that the kernel method needs only one-fifth of the number of simulations required by the basic Monte Carlo method to obtain the same accuracy. The E_{overall} is computed as follows:

1. Generate a sample of size N_{basic} from the original probability density function $f_{\mathbf{X}}(\mathbf{x})$ for a given M.

2. With a prescribed value for the accuracy $\delta_{\hat{p}_F}$, solve for N_{total} from the definition of $\delta_{\hat{p}_F}$, i.e.

$$N_{\text{total}} = \frac{1 - (\hat{p}_F)_{\text{basic}}}{\delta^2_{\hat{p}_F}(\hat{p}_F)_{\text{basic}}} \quad \ldots\ldots (32)$$

3. Compute $(N_{\text{basic}})_2$ from

$$(N_{\text{basic}})_2 = N_{\text{total}} - N_{\text{basic}} \quad \ldots\ldots (33)$$

4. Compute N_{kernel} from

$$N_{\text{kernel}} = \frac{(N_{\text{basic}})_2}{E_{\text{marginal}}} \quad \ldots\ldots (34)$$

where E_{marginal} is defined by (29).

5. Compute E_{overall} from (31).

The idea behind the development of the measure defined in (31) is motivated by the fact that the evaluation of the performance function, $g(\mathbf{x})$, is generally the part requiring a major portion of the computer time in a simulation. For a typical structure, the computer time required to set up the kernel sampling density is assumed to be negligible, compared with that required for a single evaluation of its performance function.

Basic Example

Consider the following single-random-variable problem with a simple performance function given by

$$g(x) = \alpha - x \quad \ldots\ldots (35)$$

where α = a parameter. The failure probability is then given by

$$p_F = \int_{-\infty}^{\infty} I[(\alpha - x) \leq 0] f_X(x)\, dx \quad \ldots\ldots (36)$$

in which $f_X(x)$ = the standard normal density function. This problem is convenient for examining the following: (1) The best criterion for choosing the window width; and (2) the effect of the value of p_F on the efficiency of the kernel method.

Best Criterion for Choosing Window Width

The objective is to obtain the best estimate of the optimal importance-sampling density, which is

$$h_X(x) = \frac{I[(\alpha - x) \leq 0] f_X(x)}{p_F} \quad \ldots\ldots (37)$$

where $f_X(x)$ = the standard normal density. For this simple example, p_F can be easily determined to be 0.02275; hence, the optimal importance sampling density, $h_X(x)$, can be determined from (37). In Fig. 5, the optimal importance-sampling density, and its kernel sampling densities obtained with window widths chosen by minimizing (14) and (19) with respect to w, are plotted together. The kernel sampling densities are constructed using M = 10 for the case α = 2.0. The kernel sampling density with w chosen by

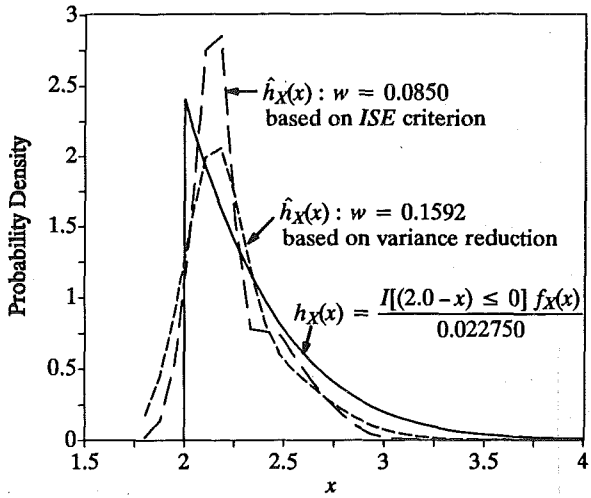

FIG. 5. Importance-Sampling Density Function

using the variance reduction criterion approximates the optimal importance-sampling density better than that based on the ISE criterion. The two kernel sampling densities also can be compared on the basis of the estimated failure probability. The p_F is estimated using (25) for different samples each of size N generated from the constructed $\hat{h}_X(x)$ based on ISE or the variance reduction criterion. The corresponding variance of $\hat{p}_F$ is computed using (26). The $\hat{p}_F$ and its $\hat{\sigma}_{\hat{p}_F}$ are plotted in Fig. 6 with respect to N. This figure shows that values of $\hat{\sigma}_{\hat{p}_F}$ based on variance reduction are consistently lower than those using the ISE criterion for choosing the window width w. Therefore, the variance reduction criterion is preferred since it yielded a smaller variance in the estimate of the failure probability.

Effect of p_F on Efficiency of Kernel Method

The marginal efficiency, E_{marginal}, defined in (29) is computed using only the sample set $\{y_1, y_2, \ldots, y_M\}$ and an initial estimate of p_F; no additional simulation run is performed. Hence, the marginal efficiency for future simulation can be predicted after any given value of M. However, the actual marginal efficiency could differ from this predicted value, as Fig. 7 shows. In this figure, the marginal efficiency is computed for different values of M and α. The actual E_{marginal} is computed by first generating samples from the kernel sampling density, then computing the sample variance $\hat{\sigma}^2_{\text{kernel}}$; the ratio $\hat{\sigma}^2_{\text{basic}}/\hat{\sigma}^2_{\text{kernel}}$ is the actual marginal efficiency. The following observations can be made from the figure:

1. The difference between the marginal efficiency of (29) and the actual efficiency is not significant. In general, the comparison may not be good for small M, because only a small sample set $\{y_1, y_2, \ldots, y_M\}$ is used to evaluate the actual marginal efficiency, as shown for the case with $p_F = 0.001350$. However, good agreement is observed for the case with $p_F = 0.022750$.
2. The marginal efficiency increases as M increases and as p_F decreases.

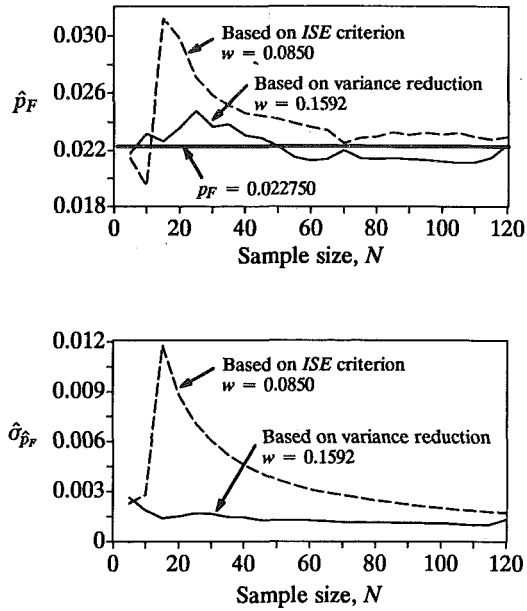

FIG. 6. Estimates of p_F and Its Standard Deviations: (a) Comparison of p_F Estimates; (b) Standard Deviations of p_F Estimates

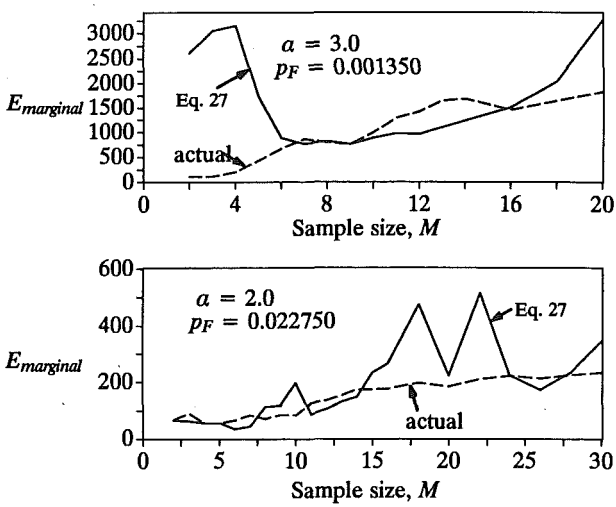

FIG. 7. Effect of p_F on Marginal Efficiency

The overall efficiency is computed for different value of M, $\delta_{\hat{p}_F}$, and α (or p_F). The computed values of E_{overall} are plotted in Figs. 8 and 9. The following can be observed from the figures:

1. For a given level of p_F and M, the overall efficiency tends to increase as

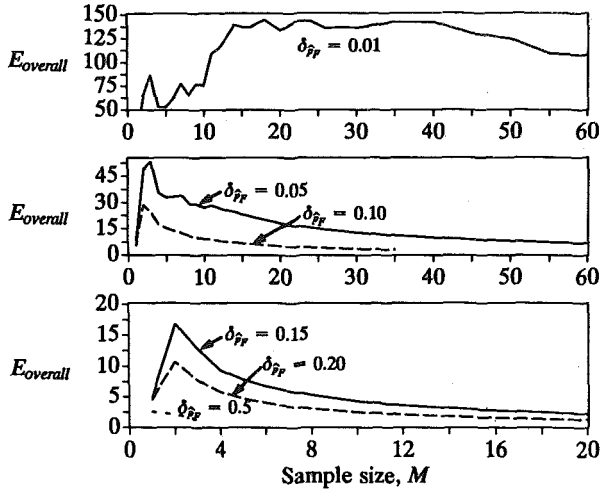

FIG. 8. Effect of Accuracy on Overall Efficiency ($p_F = 0.2275$; $\alpha = 2.0$)

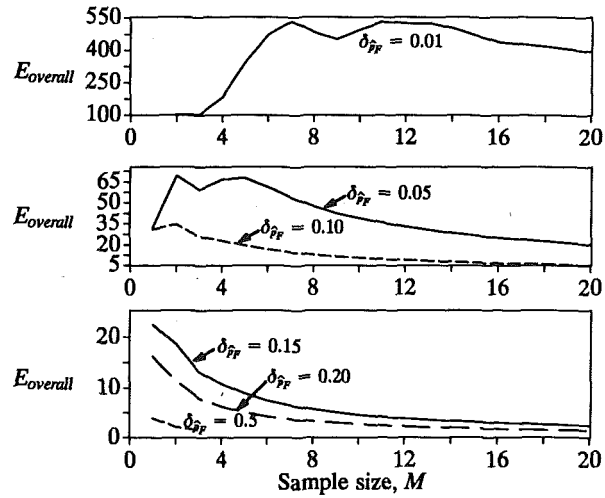

FIG. 9. Effect of Accuracy on Overall Efficiency ($p_F = 0.00135$; $\alpha = 3.0$)

$\delta_{\hat{p}_F}$ decreases, i.e., the overall efficiency increases with the degree of accuracy required on the estimate $\hat{p}_F$. This conclusion also applies to the maximum E_{overall} for a given $\delta_{\hat{p}_F}$ and p_F (see Table 1). The table also indicates that the smaller the value of p_F, the higher the maximum E_{overall}, for the same accuracy.

2. An optimal sample size M, denoted by M_{optimal}, exists, depending on the accuracy of the p_F estimate. It is defined as the number of samples used to construct the kernel sampling density that yields a maximum overall efficiency. As the required accuracy of the p_F estimate is increased (i.e., $\delta_{\hat{p}_F}$ is decreased), M_{optimal} increases, as shown in Table 2. The table also indicates the optimal

TABLE 1. Effect of Accuracy on Maximum $E_{overall}$

	$\delta_{\hat{p}_F}$					
p_F (1)	0.50 (2)	0.20 (3)	0.15 (4)	0.10 (5)	0.05 (6)	0.01 (7)
0.022750	2.55	10.55	16.61	28.20	52.92	143.31
0.001350	3.74	16.08	22.18	33.98	68.62	532.64

TABLE 2. Effect of Accuracy on $M_{optimal}$

	$\delta_{\hat{p}_F}$					
p_F (1)	0.50 (20	0.20 (3)	0.15 (4)	0.10 (5)	0.05 (6)	0.01 (7)
0.022750	1	2	2	2	3	18
0.001350	1	1	1	2	2	7

sample size required to construct the kernel sampling density is smaller for lower p_F.

The main drawback of the kernel method is that the M samples to construct the kernel sampling density must be obtained through the basic Monte Carlo method, e.g., in order to obtain $M = 1$ sample, one needs 1,000,000 samples on the average of $p_F = 1 \times 10^{-6}$.

Multivariable Example

The purpose of this example is to illustrate the application of the kernel method to a multivariable problem. The effect of the covariance matrix of the random variables on the efficiency of the kernel method also is examined.

The safety of a structural element may be evaluated on the basis of a linear performance function; for example

$$g(R,D,L) = R - D - L$$

where R = the resistance of the element; D = the dead-load effect; and L = the maximum live-load effect. The statistics of the random variables, including the correlations, are shown in Table 3.

The overall efficiency is computed for the case $\rho = 0.0$ for different values of M and $\delta_{\hat{p}_F}$, and plotted in Fig. 10 with respect to the sample size M. The underlying probability of failure is 0.00196. Similar observations made for the single random variable problem also can be made for this multivariable problem, i.e., the overall efficiency tends to increase as $\delta_{\hat{p}_F}$ decreases, and the optimal sample size M increases as $\delta_{\hat{p}_F}$ is increased.

To isolate the effect of the correlation ρ on the efficiency of the kernel method, the mean of L is adjusted for different values of ρ to maintain the same value of p_F (= 0.001960), as summarized in Table 4. The overall efficiency of the kernel method with $\delta_{\hat{p}_F} = 0.10$ is computed for different values of ρ and M, and plotted in Fig. 11. The results indicate that the efficiency of the kernel method is not affected by ρ provided that p_F is kept constant (the apparent difference in the value of $E_{overall}$ for $M < 3$ is due to statistical sampling error).

TABLE 3. Statistics of Random Variables

Variable (1)	Distribution type (2)	Mean (3)	Coefficient of variation (4)
R	Normal	2.831	0.11
D	Normal	1.000	0.10
L	Normal	0.745	0.25

Note: $\rho_{R,D} = \rho$ and $\rho_{R,L} = \rho_{D,L} = 0.0$.

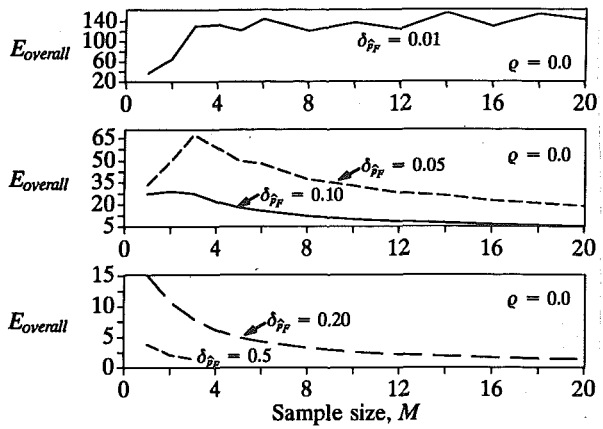

FIG. 10. Effect of Accuracy on Overall Efficiency

TABLE 4. $\bar{L}$ for Different Correlation Coefficient

ρ (1)	$\bar{L}$ (2)
0.0	0.7450
0.2	0.7804
0.6	0.8524
0.8	0.8891

d-Dimensional Spherical Example

To show the effect of the number of random variables on the overall efficiency of the kernel method, consider the following problem with the performance function given by

$$g(x_1,x_2, \ldots ,x_d; \alpha) = \alpha - \sum_{i=1}^{d} x_i^2$$

where α = a parameter. The random variables $x_1,x_2, \ldots ,x_d$ are standard normal and statistically independent.

The overall efficiency of the kernel method with $M = 4$ and $\delta_{\hat{P}_F} = 0.10$ is computed for different numbers of random variables, d. To isolate the effect of d, the parameter α needs to be adjusted for different d to maintain

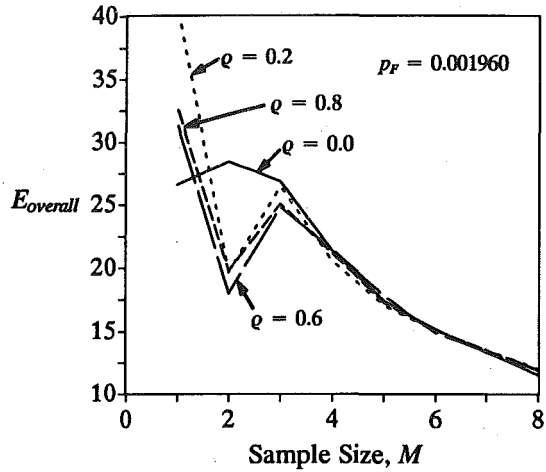

FIG. 11. Effect of Correlation Coefficient of E_{overall}

TABLE 5. Values of α for Different d

d (1)	α (2)
1	10.07
2	13.00
3	15.40
4	17.56
6	21.48
8	25.08
10	28.49
15	36.49
20	44.00

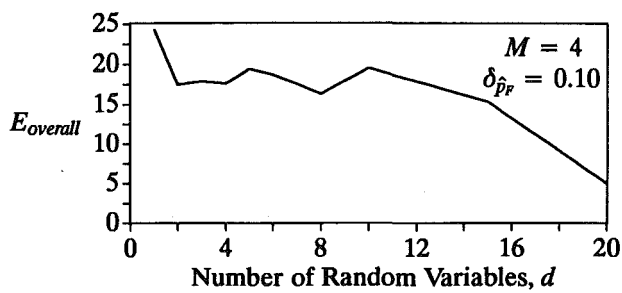

FIG. 12. Effect of d on E_{overall}

a constant p_F ($=0.00150$). The values of α corresponding to different values of d are listed in Table 5. The results, plotted in Fig. 12, reveal a decrease in the overall efficiency as d is increased. The figure shows that the overall efficiency decreases sharply for $d > 15$. The drop in overall efficiency with

respect to d is expected because of the nature of estimating the optimal importance-sampling density. Fewer points are needed to fit a curve in a univariate problem than are required for a surface in a bivariate problem for a given accuracy; and fitting a hypersurface in a multivariable problem requires many more points or samples to obtain the same accuracy.

Conclusions

The proposed kernel method is much more efficient than the basic Monte Carlo method. In the few examples studied, the overall efficiency ranges from 2.5–500.0, depending on the problem and the accuracy required for the estimated p_F. The main reason for the high efficiency is that the proposed procedure to construct the kernel sampling density ensures that this sampling density resembles the optimal sampling density. Moreover, a most important advantage of the kernel method is that it is completely automatic, i.e., the choice of the number of samples, M, required to construct the kernel sampling density, and the number of simulations, N, to be generated from the kernel sampling density, can be determined automatically by the program based on the level of accuracy desired.

Based on the results from the numerical examples, the following conclusions can be observed:

1. The p_F estimated with the kernel method is sensitive to the method of choosing the window width. The variance reduction criterion is preferred, since it yielded a smaller variance in the estimated failure probability.
2. The difference between the marginal efficiency expressed by (29) and the actual efficiency is not significant. Therefore, (29) can be used to estimate the actual marginal efficiency and compute the overall efficiency. The examples show that both the marginal and overall efficiencies of the kernel method increase as p_F decreases. Furthermore, it is shown that the maximum overall efficiency occurs at small values of M (< 10). This means that to obtain the maximum benefit of the kernel method, a small sample of size M is required to construct the kernel sampling density. However, the required size of M depends on the accuracy desired. The greater the accuracy desired, the larger the sample size M required to obtain the maximum overall efficiency of the kernel method.
3. The degreee of correlation between random variables does not affect the overall efficiency of the kernel method, provided that p_F remains the same. If the change in the covariance matrix results in a decrease in p_F, the efficiency of the kernel method will increase, and vice versa.
4. The overall efficiency of the kernel method decreases with the number of random variables. The reason for the decrease in efficiency with the number of random variables is that fewer samples are needed to construct the kernel sampling density for a univariate density function than for a multivariate density function, for a given accuracy.

Acknowledgments

This study was supported in part by the National Science Foundation, under grant MSM 85-18005, and a fellowship from the Kajima Corp. Support for this work is gratefully acknowledged.

APPENDIX I. REFERENCES

Ang, G. L., Ang, A. H-S., and Tang, W. H. (1989). "Kernel method in importance sampling density function." *Structural safety and reliability*, A. H-S. Ang, M. Shinozuka, and G. I. Schueller, eds., Proc. of ICOSSAR '89, ASCE, New York, N.Y. Vol. II, 1193–1200.

Ang, G. L. (1991). "Kernel method in Monte Carlo importance sampling," PhD thesis, University of Illinois, at Urbana, Illinois.

Abramson, I. S. (1982). "On bandwidth variation in kernel estimates—a square root law." *Ann. Statist.*, 10(4), 1217–1223.

Bourgund, U., and Bucher, C. G. (1986). "Importance sampling procedure using design points (ISPUD)—a user's manual." *Report No. 8*, Institute of Engineering Mechanics, University of Innsbruck, Innsbruck, Austria.

Harbitz, A. (1982). "A level 3 accurate probability of failure calculation method." *Proc. of Euromech 155 on Reliability Theory of Structural Engineering System*, Technical University of Denmark, Lyngby, Denmark, 337–346.

Harbitz, A. (1986). "An efficient sampling method for probability of failure calculation." *Struct. Safety*, 3(2), 109–115.

Ibrahim, Y., and Cornell, C. A. (1988). "Experiences with applications of importance sampling in structural reliability computation." *Probabilistic methods in civil engineering*, P. D. Spanos, ed., ASCE, New York, N.Y., 344–347.

Karamchandari, A. (1990). "New methods in system reliability." *Report No. RMS-7*, Reliability of Marine Structures Program, Department of Civil Engineering, Stanford University, Stanford, Calif.

Rubinstein, R. Y. (1981). *Simulation and the Monte-Carlo method*. John Wiley & Sons, New York, N.Y.

Schueller, G. I., and Stix, R. (1987). "A critical appraisal of methods to determine failure probabilities" *Struct. Safety*, 4(4), 293–309.