ELSEVIER

Structural Safety 25 (2003) 139–163

STRUCTURAL SAFETY

www.elsevier.com/locate/strusafe

Important sampling in high dimensions

S.K. Au[a,*], J.L. Beck[b]

[a]School of Civil and Environmental Engineering, Nanyang Technological University, Singapore 639798, Singapore
[b]Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125, USA

Abstract

This paper draws attention to a fundamental problem that occurs in applying importance sampling to 'high-dimensional' reliability problems, i.e., those with a large number of uncertain parameters. This question of applicability carries an important bearing on the potential use of importance sampling for solving dynamic first-excursion problems and static reliability problems for structures with a large number of uncertain structural model parameters. The conditions under which importance sampling is applicable in high dimensions are investigated, where the focus is put on the common case of standard Gaussian uncertain parameters. It is found that importance sampling densities using design points are applicable if the covariance matrix associated with each design point does not deviate significantly from the identity matrix. The study also suggests that importance sampling densities using random pre-samples are generally not applicable in high dimensions.

Keywords: Importance sampling; Monte Carlo simulation; Reliability; Relative entropy

1. Introduction

The proper assessment of the performance reliability of engineering structures is an important component in a modern performance-based engineering framework [1–3]. In the presence of uncertainties, the reliability, or equivalently, the failure probability of a system is of paramount importance [4,5]. Let $\underline{\Theta} \epsilon \mathbb{R}^n$ be the vector of uncertain parameters of a system with joint parameter probability density function (PDF) $q:\mathbb{R}^n \mapsto [0,\infty)$ that reflects the relative plausibility of its possible values [6] and let $F \subset \mathbb{R}^n$ denote the failure region such that $\underline{\Theta} \epsilon F$ corresponds to a failure event. Then the failure probability $P(F)$ can be formulated as

* Corresponding author. Tel.: +65-6790-6276; fax: +65-6791-0676..
E-mail address: cskau@ntu.edu.sg (S.K. Au).

0167-4730/02/$ - see front matter © 2002 Elsevier Science Ltd. All rights reserved.
PII: S0167-4730(02)00047-4

$$P(F) = \int_F q(\underline{\theta})d\underline{\theta} = \int \mathbb{I}_F(\underline{\theta})q(\underline{\theta})\mathrm{d}\underline{\theta} = \mathrm{E}_q[\mathbb{I}_F(\Theta)] \tag{1}$$

where the second integral is over the whole parameter space and $\mathbb{I}_F : \mathbb{R}^n \mapsto \{0, 1\}$ is the indicator function: $\mathbb{I}_F(\underline{\theta}) = 1$ when $\underline{\theta} \in F$ and $\mathbb{I}_F(\underline{\theta}) = 0$ otherwise. We denote this reliability problem by the ordered pair $\mathcal{R}$ (q, F). Since the dimension of the integral is equal to the number of uncertain parameters, n is also referred here to as the *dimension* of the reliability problem.

Efficient methods for computing the failure probability are required to facilitate the implementation of a reliability-based engineering framework. Standard Monte Carlo simulation [7–9] provides the most robust way for computing the failure probability, but it is well known to be inefficient when the failure probability is small. Essentially, estimating small probabilities requires information from 'rare' samples which lead to failure, and on average it requires many samples before one such failure sample occurs. In view of this, the importance sampling method [7,8,10–12] has been introduced, which basically chooses an *importance sampling density* (ISD) f: $\mathbb{R}^n \mapsto [0, \infty)$ to generate samples that lead to failure more frequently so as to gain more information about failure for more efficient estimation of the failure probability. It is assumed that efficient methods for evaluating f and simulating random samples according to f are available. In this method, the failure probability is written as

$$P(F) = \int \frac{\mathbb{I}_F(\underline{\theta})q(\underline{\theta})}{f(\underline{\theta})} f(\underline{\theta})\mathrm{d}\underline{\theta} = \mathrm{E}_f[\mathbb{I}_F(\underline{\Theta})R(\Theta)] \tag{2}$$

where the subscript f in $\mathrm{E}_f[\cdot]$ denotes that the expectation is taken with $\underline{\Theta}$ distributed as f, and

$$R(\underline{\Theta}) = \frac{q(\underline{\Theta})}{f(\underline{\Theta})} \tag{3}$$

is called the *importance sampling quotient*. It is assumed that F is a subset of the support $G = \{\underline{\theta} \in \mathbb{R}^n : f(\underline{\theta}) > 0\}$ of the ISD, so that $f(\underline{\theta}) > 0$ for all $\underline{\theta} \in F$ and the integral in (2) is valid over $\mathbb{R}^n$. The theoretical mean in (2) is estimated by a sample mean as

$$P(F) \approx \tilde{P} = \frac{1}{N}\sum_{k=1}^{N} \mathbb{I}_F(\underline{\Theta}_k)R(\underline{\Theta}_k) \tag{4}$$

where $\{\underline{\Theta}_k : k = 1, \ldots, N\}$ are independent and identically distributed (i.i.d.) samples simulated according to f instead of from q.

The variability of the importance sampling estimate $\tilde{P}$ is commonly measured by its coefficient of variation (c.o.v.), δ_{IS}, defined as the ratio of its standard deviation to its mean. Since the samples are i.i.d., $\delta_{IS} = \Delta_{IS}/\sqrt{N}$, where Δ_{IS} is called the *unit c.o.v.* of the importance sampling estimator, defined as the c.o.v. of the importance sampling estimator with $N = 1$ on the R.H.S. of (4), so

$$\Delta_{IS}^2 = \frac{\mathrm{Var}_f[\mathbb{I}_F(\underline{\Theta})R(\underline{\Theta})]}{P(F)^2} = \frac{\mathrm{E}_f\left[\mathbb{I}_F(\underline{\Theta})^2\right]}{P(F)^2} - 1 \tag{5}$$

The unit c.o.v. of the importance sampling quotient R can be viewed as corresponding to the special case when $F=G$:

$$\Delta_R^2 = \frac{\mathrm{Var}_f[R(\underline{\Theta})]}{E_f[R(\underline{\Theta})]^2} = \frac{\mathrm{Var}_f[\mathbb{I}_G(\underline{\Theta})R(\underline{\Theta})]}{P(G)^2} = \frac{\mathrm{E}_f\left[R(\underline{\Theta})^2\right]}{P(G)^2} - 1 \tag{6}$$

since $\mathrm{E}_f[R(\underline{\Theta})] = \int_G f(\underline{\theta}) \times q(\underline{\theta})/f(\underline{\theta})\mathrm{d}\underline{\theta} = \int_G q(\underline{\theta})\mathrm{d}\underline{\theta} = P(G)$.

The main problem in importance sampling is how to choose an ISD f that results in a small unit c.o.v. Δ_{IS} and hence leads to an efficient simulation procedure. It is noted that the optimal ISD, f_{opt}, is simply the conditional PDF given that failure occurs:

$$f_{\mathrm{opt}}(\underline{\theta}) = q(\underline{\theta} \mid F) = \frac{\mathbb{I}_F(\underline{\theta})q(\underline{\theta})}{P(F)} \tag{7}$$

The optimality of f_{opt} can be easily verified by noting that it leads to zero unit c.o.v. when substituted into (5). The use of the optimal ISD is not feasible, however, due to two basic reasons. The first is that its evaluation involves knowledge of the failure probability $P(F)$, which is the quantity to be computed in the reliability problem. The second reason is that an efficient method for simulating samples according to f_{opt} is often not available.

Although the optimal choice of the ISD in (7) is generally not feasible, it points out the important philosophy that an efficient simulation procedure is intimately related to efficient generation of condition samples according to $q(\cdot \mid F)$. Many schemes for constructing the ISD have appeared in the engineering reliability literature. Most schemes involve first finding the important parts of the failure region which give significant contribution to the failure probability, and then constructing an ISD based on information about such important failure regions. For example, a popular strategy is to construct the ISD as a mixture distribution among one or more design point(s) that have the highest probability density, at least locally, among all other points in their neighborhood within the failure region [10,13–18]. Another popular strategy, called adaptive importance sampling, is to construct the ISD as a mixture distribution among some random *presamples* which are generated in the failure region by some pre-designed stochastic algorithm [19–23]. Generally speaking, importance sampling has been successfully applied to static reliability problems with a small to medium number of uncertain structural parameters. Applications to 'high-dimensional' reliability problems, such as problems with a large number of uncertain structural model parameters or dynamic problems where the stochastic excitation is explicitly represented, are still at their early stage of exploration [24–26].

This paper is concerned with whether it is possible to apply importance sampling in high-dimensional problems, even if adequate information about the failure region, such as design points, is available. The study is motivated by a basic problem that we have discovered, namely, when the choice of the form of the importance sampling density is not appropriate, the variability of the importance sampling estimator may grow systematically without bound as the dimension increases.

The structure of this paper is as follows. We first give a formal definition of *applicability in high dimensions*. The concept of 'relative entropy' is then introduced, which gives a useful measure of the difference between two probability density functions. This facilitates a simple illustration of the basic problem of concern when importance sampling is applied to high-dimensional problems. The rest of the paper focuses on the common case where the parameter PDF q is a standard Gaussian PDF and the ISD f is a Gaussian PDF centered at one or more points. These points can be either design points or random pre-samples. The basic approach is to investigate the asymptotic probabilistic behavior of the importance sampling quotient as $n\to\infty$. Examples are given to illustrate the applicability issues of importance sampling in high-dimensional problems.

2. Definition of applicability in high dimensions

To address formally the issue of concern here, we need to define what we mean by importance sampling being applicable in high dimensions. For the question of applicability to be meaningful, assume that we have a generic reliability problem with n uncertain parameters, $n \in \mathbb{Z}^+$, from which a sequence of similar problems of increasing number of uncertain parameters can be induced by increasing n by admissible increments. For example, consider computing the failure probability of a deterministic single-degree-of-freedom oscillator subjected to Gaussian white noise discretized in the time domain by n i.i.d. standard Gaussian random variables. Then a legitimate sequence of problems with increasing dimension n can be generated by refining the discretization in the time-domain. In particular, if each refinement corresponds to subdividing each existing time interval by half, then an admissible increment of dimension n may be taken as n (assuming the first point of the time horizon is not represented). Starting with n_1 discrete time instants (and hence uncertain parameters), the sequence of dimensions associated with this sequence of reliability problems will be $\mathcal{N} = \{n_k : k = 1, 2, \ldots\} = \{n_1, 2n_1, 4n_1, .\}$.

Let a reliability problem with n uncertain parameters be defined by (1) and the ordered pair $\mathcal{R}(q_n, F_n)$, where q_n is the joint PDF for the uncertain parameters and $F_n \subset \mathbb{R}^n$ is the failure region. For a given sequence of admissible dimensions $\mathcal{N} = \{n_k : k = 1,2,\ldots\}$, consider a sequence of reliability problems $\{\mathcal{R}(q_{n_k}, F_{n_k}) : k = 1, 2, .\}$. For the k-th problem in the sequence, let $\mathcal{P}_{\mathcal{IS}}(n_k)$ denote a class of n_k-dimensional joint PDFs from which the ISD f_{n_k} is chosen. The failure probability is estimated by an importance sampling estimator using the ISD f_{n_k}:

$$P(F_{n_k}) \approx \tilde{P}_{n_k} = \frac{1}{N}\sum_{r=1}^{N} \mathbb{I}F_{n_k}(\underline{\Theta}_r) R_{n_k}(\underline{\Theta}_r) \tag{8}$$

where $\{\underline{\Theta}_r : r = 1, \ldots, N\}$ are i.i.d. samples simulated according to the ISD f_{n_k} and $R_{n_k}(\cdot) = q_{n_k}(\cdot)/f_{n_k}(\cdot)$ is the importance sampling quotient in the k-th problem of the sequence. Let $\Delta_{IS}(n_k)$ be the unit c.o.v. of the importance sampling estimator $\tilde{P}_{n_k}$, that is, according to (5),

$$\Delta_{IS}(n_k)^2 = \frac{\mathrm{E}_f\left[\mathbb{I}_{F_{n_k}}(\Theta) R_{n_k}(\Theta)^2\right]}{P\left(F_{n_k}\right)^2} - 1 \tag{9}$$

Definition 1. Importance sampling is applicable in high dimensions for the reliability problem $\mathcal{R}(q_n, F_n)$ with ISD chosen from the class of PDFs $\mathcal{P}_{\mathcal{IS}}(n)$, if $\Delta_{IS}(n_k) < \infty$ as $k \to \infty$ for some increasing sequence $\mathcal{N} = \{n_k \in \mathbb{Z}^+ : k = 1, 2, \ldots\}$ of admissible dimensions with $n_k \to \infty$ and some sequence of ISDs $\{f_{n_k} \in \mathcal{P}_{IS}(n_k) : k = 1, 2, \ldots\}$

In our context, 'applicability' does not imply 'efficiency,' that is, if according to Definition 1, importance sampling is found to be applicable in high dimensions, it is not necessary that the importance sampling procedure will be efficient. This is because the unit c.o.v. of the failure probability estimate may be large even if it remains bounded as the dimension increases. Also, the study of applicability does not offer an explicit answer as to whether the importance sampling estimate is biased or not. Rather, it is assumed that the estimate is unbiased in the analysis. The issue of bias is related to whether the ISD has accounted for all the parts in the failure region which give the major contribution to the failure probability. Efficiency and bias depend on which particular member from the class of ISDs $\mathcal{P}_{\mathcal{IS}}(n)$ is chosen, rather than on what general properties $\mathcal{P}_{\mathcal{IS}}(n)$ should possess. In short, the concern with 'applicability' is whether it is possible at all to apply importance sampling, leaving aside the issues of how to gain information about the failure region to avoid bias or whether the resulting ISD will lead to an efficient estimate. Applicability is the first concern when one applies importance sampling to high dimensional problems, however, since if the chosen class of ISDs already implies that the variability of the failure probability estimate will generally increase without bound as n increases, the effort spent on searching for a suitable ISD from $\mathcal{P}_{\mathcal{IS}}(n)$ will be in vain.

3. Concept of relative entropy

The relative entropy of a PDF p_2 relative to a PDF p_1, is defined as:

$$H(p_1, p_2) = \int p_1(\underline{\theta}) \log \frac{p_1(\underline{\theta})}{p_2(\underline{\theta})} \mathrm{d}\underline{\theta} \tag{10}$$

The relative entropy $H(p_1,p_2)$ is a useful measure for the difference between two PDFs [27–29]. It is always non-negative, and is equal to zero if and only if $p_1 \equiv p_2$. Applying the concept of relative entropy to the choice of ISD, one can expect that if the relative entropy of the ISD f to $q(\cdot \mid F)$ (the optimal ISD) is small, then the unit c.o.v. Δ_{IS} will also be small, and vice versa. In particular, the relative entropy is zero if and only if Δ_{IS} is zero, since both statements require $f(\cdot) = q(\cdot \mid F)$. These statements are indeed true and quite intuitive. In general, it can be shown [30] that

$$\Delta_{IS}^2 \geqslant \exp[H(q(\cdot \mid F), f)] - 1 \tag{11}$$

where $H(q(\cdot\mid F),f)$ is the relative entropy of the ISD $f(\cdot)$ to the conditional density $q(\cdot\mid F)$. This relates the unit c.o.v. of the importance sampling estimator to the relative entropy $H(q(\cdot\mid F),f)$. Thus, a necessary condition for importance sampling to be applicable in high dimensional problems for $\mathcal{R}(q_n, F_n)$ is that there exists $f_n \in \mathcal{P}_{\mathcal{IS}}(n)$ such that $H(q(\cdot\mid F),f_n) < \infty$ as $n \to \infty$. This provides the basic platform for our discussion here.

4. Basic problem of concern

We are now ready to illustrate the basic problem of concern that may occur in high dimensional problems. Consider the i.i.d. case where $q(\underline{\theta}) = \Pi_{i=1}^{n} q_1(\theta_i)$ and $f(\underline{\theta}) = \Pi_{i=1}^{n} f_1(\theta_i)$, where q_1 and f_1 are the one-dimensional PDF for each component Θ_i of $\underline{\Theta} = [\Theta_1, \ldots, \Theta_n]$ distributed according to the parameter PDF q and the ISD f, respectively. Then

$$H(q,f) = \mathrm{E}_q\left[\log\frac{q(\underline{\Theta})}{f(\underline{\Theta})}\right] = \sum_{i=1}^{n}\mathrm{E}_q\left[\log\frac{q_1(\Theta_i)}{f_1(\Theta_i)}\right] = n\mathrm{E}_{q1}\left[\log\frac{q_1(\Theta)}{f_1(\Theta)}\right] = nH(q_1,f_1) \tag{12}$$

where Θ is a generic random variable distributed as q_1 and $H(q_1,f_1)$ is the relative entropy of f_1 to q_1. This means that, unless $H(q_1,f_1)$ is at most of the order of $1/n$, $H(q,f)$ will grow with n. In fact, if $f_1 \neq q_1$ and the choice of f_1 does not depend on n, then when $\underline{\Theta}$ is distributed as $f(\underline{\theta}) = \Pi_{i=1}^{n} f_1(\theta_i)$, by the Strong Law of Large Numbers,

$$\frac{1}{n}\log\frac{q(\underline{\Theta})}{f(\underline{\Theta})} = \frac{1}{n}\sum_{i=1}^{n}\log\frac{q_1(\Theta_i)}{f_1(\Theta_i)} \;\to\; \mathrm{E}_{f_1}\left[\log\frac{q_1(\Theta)}{f_1(\Theta)}\right] = -H(f_1, q_1) \text{ as } n \to \infty \tag{13}$$

with probability 1. Consequently, with probability 1, when $\underline{\Theta}$ is distributed as f,

$$R(\underline{\Theta}) = \frac{q(\underline{\Theta})}{f(\underline{\Theta})} \;\to\; \exp[-nH(f_1, q_1)] \text{ as } n \to \infty \tag{14}$$

and hence R is exponentially small as $n\to\infty$. By noting that $E_f[R(\underline{\Theta})] = 1$ theoretically, one can infer that when n is large, R is exponentially small for most of the time, but on some rare occasions, it assumes extremely large values, so that its theoretical mean is maintained. This phenomenon stems from the difference between the one-dimensional PDFs q_1 and f_1, which is amplified linearly in the relative entropy and hence exponentially in the unit c.o.v. Δ_R of R as n increases. Although nothing has been said about Δ_{IS}, this observation suggests that Δ_{IS} will also grow without bound as $n \to \infty$ and hence importance sampling may not be applicable in high dimensions. Intuition suggests that the importance sampling estimate will be practically biased as well as having large variability and this turns out to be correct.

The question now is whether it is feasible in practice to choose an ISD that remains close to the parameter PDF, in the sense that $H(q,f)$ remains bounded as n$\to\infty$. Suppose $q(\underline{\theta}) = \Pi_{i=1}^{n}\phi(\theta_i)$ is the standard Gaussian PDF with independent components, where $\phi(\theta) = \exp\left(-\theta^2/2\right)/\sqrt{2\pi}$ is the one-dimensional standard Gaussian PDF. Let the ISD be a Gaussian PDF centered at a single design point $\underline{\theta}^* = \left[\theta_1^*, \ldots, \theta_n^*\right]$ with independent components that have unit variance, that is, $f(\underline{\theta}) = \Pi_{i=1}^n \phi\left(\theta_i - \theta_i^*\right)$, then

$$H(q,f) = \mathrm{E}_q\left[\log\frac{q(\Theta)}{f(\Theta)}\right] = \sum_{i=1}^n \mathrm{E}_\phi\left[\log\frac{\phi(\Theta_i)}{\phi\left(\Theta_i - \theta_i^*\right)}\right] = \frac{1}{2}\sum_{i=1}^n \theta_i^{*2} = \frac{1}{2}\left\|\underline{\theta}^*\right\|^2 \tag{15}$$

To determine if $H(f,q)$ is bounded, we note that the Euclidean norm $\left\|\underline{\theta}^*\right\|$ is intimately related to the failure probability. For example, if F is a half-space defined by a hyperplane with the design point $\underline{\theta}^*$, then the failure probability is $P(F) = \Phi\left(-\left\|\underline{\theta}^*\right\|\right)$, where $\Phi(x)$ is the standard Gaussian cumulative distribution function. Since we are investigating similar problems, it is reasonable to postulate that $P(F)$ remains nonzero as n increases. This means $\left\|\underline{\theta}^*\right\|$ remains bounded as $n \to \infty$ and so does $H(f, q)$.

This example suggests that importance sampling using design points may still be applicable in high dimensions, as the design point automatically adjusts itself so that $H(q, f)$ remains bounded. In general, however, this comes with some conditions and may not be taken for granted. One counter example for this is the case when f_1 has fixed standard deviation $s \neq 1$, that is, $f_1(\theta_i) = \exp\left(-\left(\theta_i - \theta_i^*\right)^2/2s^2\right)/\sqrt{2\pi}\,s$. In this case,

$$H(q,f) = \frac{n}{2}\left(\frac{1}{s^2} + \log s^2 - 1\right) + \frac{1}{2}\left\|\underline{\theta}^*\right\|^2 \tag{16}$$

The first term in (16) comes from the fact that a standard deviation $s \neq 1$ is used in the ISD, while the second term is due to the shift of ISD from the origin to the design point $\underline{\theta}^*$. Note that the first term is equal to the relative entropy of f to q if $\underline{\theta}^* = \underline{0}$, which can be easily verified by setting $\underline{\theta}^* = \underline{0}$ in (16). It is non-negative and is equal to zero if and only if $s = 1$. For fixed $s \neq 1$, the first term is $O(n)$. The second term is bounded, as in the last example. This means $H(q,f)$ grows in a linear fashion with n as long as s is fixed and not equal to unity. Thus, although the shift of ISD to the design point does not render importance sampling inapplicable in high dimensions, the use of standard deviation s different from that of the original PDF (equal to unity) does. Intuitively, one may expect that importance sampling is not applicable when $s < 1$ even when n is not large, since then the ISD decays faster than the parameter PDF at its 'tail' where $R(\underline{\theta}) = q(\underline{\theta})/f(\underline{\theta})$ grows without bound. The surprising observation from this example is that, although importance sampling with $s > 1$ is applicable when n is not large, the same is not true in high dimensions (mathematically as $n \to \infty$). This is a warning that intuition and guidelines for constructing the ISD in low dimensions may not give a complete picture of what phenomenon dominates in high dimensions, which leads to interest in conditions under which importance sampling is applicable in high dimensions.

5. Analysis of applicability in high dimensions

In what follows, we investigate the conditions under which importance sampling is applicable in high dimensions. To streamline our discussion, the detailed proofs are omitted here and for these the reader is referred to [30]. For convenience in analysis, we assume that the reliability problem $\mathcal{R}(q_n, F_n)$ is defined for every $n \in \mathbb{Z}^+$, so that we take the sequence of admissible dimensions as $\mathcal{N} = \{n_k : k = 1, 2, \ldots\} = \{1, 2, \ldots\}$. Since the subscript k now becomes redundant, it is dropped in our notation. We also no longer denote the dependence of quantities on the dimension n, so there is an implicit understanding that all quantities under consideration are specific to a simulation problem with n uncertain parameters. For example, $f_n(\underline{\theta})$ is abbreviated to $f(\underline{\theta})$, and F_n to F. We assume the following postulate for the failure regions, which essentially reflects that we are studying problems of non-vanishing failure probabilities.

Postulate 1. For every $n \in \mathbb{Z}^+$, $P(F_n) > \varepsilon$ for some fixed $\varepsilon > 0$ independent of n, that is, $P(F_n)$ is non-zero for every n and does not vanish as $n \to \infty$

The focus will be placed on the case where the uncertain parameters are i.i.d. standard Gaussian, that is, for a given $n \in \mathbb{Z}^+$,

$$q(\underline{\theta}) = \phi_n(\underline{\theta}) = (2\pi)^{-n/2}\exp\left(-\frac{1}{2}\underline{\theta}^T\underline{\theta}\right) \tag{17}$$

which is a common PDF used in applications. The independence assumption does not introduce any loss of generality, since the dependent variables can always be transformed to independent ones.

Ideally, the question of applicability can be answered if we know either analytically or numerically how Δ_{IS} behaves with increasing n. The analysis of Δ_{IS}, given by (5) or otherwise, is difficult in general, due to the fact that the expression of the importance sampling quotient could be complicated depending on the form of ISD used. Also, the failure region can have a complicated structure. The evaluation of Δ_{IS} by simulation is not computationally favorable, since it involves evaluating the indicator function $\mathbb{I}_F(\cdot)$ during the averaging process which requires system analyses. Realizing that the applicability problem arises basically due to the variability of the importance sampling quotient $R(\cdot) = q(\cdot)/f(\cdot)$, one is interested to see whether the behavior of Δ_{IS} can be inferred from that of Δ_R. If the answer is positive, then the applicability problem may be solved, at least numerically in an efficient manner, since one can estimate Δ_R by simulation which only involves the evaluation of $R(\cdot)$ and so does not require any system analysis. In fact, since $\mathbb{I}_F(\cdot) \leqslant 1$, we get from (5) and (6), $\Delta_{IS}^2 \leqslant \mathrm{E}_f\left[R(\underline{\Theta})^2\right]/P(F)^2 - 1 = \left(\Delta_R^2 + 1\right)P(G)^2/P(F)^2 - 1$ and therefore

Proposition 1. *As* $n \to \infty$, if $\Delta_R < \infty$, then $\Delta_{IS} < \infty$ *also and the reliability problem is applicable in high dimensions.*

Thus, if under certain conditions on the ISD we know that Δ_R is asymptotically bounded, then we can conclude that Δ_{IS} is also asymptotically bounded, and hence importance sampling is applicable in high dimensions in the given situation. It is also important, however, to examine whether $\Delta_R < \infty$ is a necessary condition for $\Delta_{IS} < \infty$, since the sufficient condition could be so restrictive that it excludes a large class of ISDs which are applicable. This is what motivates the logic throughout the analysis to follow, where an attempt will be made to investigate the relationship between Δ_{IS} and Δ_R.

5.1. ISD centered at a single point

The applicability aspects of ISDs chosen from the class of Gaussian PDFs centered at a single point and with a positive definite covariance matrix is investigated first. An ISD f chosen from this class will then have the form:

$$f(\underline{\theta}) = (2\pi)^{-n/2}\sqrt{|\underline{C}^{-1}|}\exp\left[-\frac{1}{2}\left(\underline{\theta}-\tilde{\underline{\theta}}\right)^T\underline{C}^{-1}\left(\underline{\theta}-\tilde{\underline{\theta}}\right)\right] \tag{18}$$

where $\tilde{\underline{\theta}}$ and $\underline{C}$ are the mean vector and covariance matrix of the Gaussian PDF, respectively. In applications, the point $\tilde{\underline{\theta}}$ at which the ISD is centered may correspond to a design point or a random pre-sample, for example. We will leave it general at this stage, and consider some special cases after the main results have been established.

We first present an expression for Δ_{IS} [30] that is central to the determination of the conditions for high dimensional applicability. The expression depends on the eigenvalues (or principal variances) $\{s_i^2 : i = 1, \dots, n\}$ of the covariance matrix $\underline{C}$. Specifically, if $s_i > 1/\sqrt{2}$ for all $i = 1, \dots, n$,

$$\Delta_{IS}^2 = \frac{Q\left(\tilde{\underline{\theta}}, \underline{C}, F\right)}{P(F)^2}\left(\prod_{i=1}^{n}\frac{s_i^2}{\sqrt{2s_i^2-1}}\right)\exp\left[\sum_{i=1}^{n}\frac{\tilde{z}_i^2}{2s_i^2-1}\right]-1 \tag{19}$$

where $[\tilde{z}_1, \dots, \tilde{z}_n]^T = \underline{\Psi}^T\tilde{\underline{\theta}}$; $\underline{\Psi}$ is the eigenmatrix of $\underline{C}$ with normalization $\underline{\Psi}^T\underline{\Psi} = \underline{I}$;

$$Q\left(\bar{\underline{\theta}}, \underline{C}, F\right) = \int \phi_n(\underline{z})\mathbb{I}_F\left(\underline{\Psi}\left(\Lambda_{\hat{s}}\underline{z} + \hat{\underline{z}}\right)\right)\mathrm{d}\underline{z} \tag{20}$$

is the probability that the vector $\underline{\Psi}(\Lambda_{\hat{s}}\underline{Z} + \hat{\underline{z}})$ lies in the failure region F, where $\underline{Z}$ is an n-dimensional standard Gaussian vector; $\hat{\underline{z}} = [\hat{z}_1, \dots, \hat{z}_n]^T$ with $\hat{z}_i = -\tilde{z}_i/\left(2s_i^2-1\right)$; and $\Lambda_{\hat{s}} \in \mathbb{R}^{n\times n}$ is a diagonal matrix with the i-th diagonal element equal to $\hat{s}_i = s_i/\sqrt{2s_i^2-1}$.

The situation is less determinate if $s_j \leqslant 1/\sqrt{2}$ for some $j \in \{1, \dots, n\}$. In this case, Δ_{IS} may not be bounded, depending on the structure of the failure region F. In the special case when $F = \mathbb{R}^n$, $\Delta_{IS} = \Delta_R$ is always unbounded when there exists $s_j \leqslant 1/\sqrt{2}$ for some $j \in \{1, \dots, n\}$. In general, the situation depends on the structure of F in the j-th eigendirection of $\underline{C}$ for which $s_j \leqslant 1/\sqrt{2}$, although it can be argued that Δ_{IS} is generally unbounded except for some special F. This information about F is usually not available when importance sampling is applied, and therefore choosing some $s_j \leqslant 1/\sqrt{2}$ may render Δ_{IS} unbounded, even for finite n. This case violates the usual rule of choosing an ISD with a 'heavier tail' than the parameter PDF and therefore is of little practical interest. We thus focus on the case when $s_i > 1/\sqrt{2}$ for all $i = 1, \ldots, n$.

By taking $F = G = \mathbb{R}^n$ in (19) and noting that $Q\left(\tilde{\underline{\theta}}, \underline{C}, \mathbb{R}^n\right) = 1$ and $P(G) = P(\mathbb{R}^n) = 1$, one readily obtains that

$$\Delta_R^2 = \left(\prod_{i=1}^{n} \frac{s_i^2}{\sqrt{2s_i^2 - 1}}\right)\exp\left(\sum_{i=1}^{n} \frac{\tilde{z}_i^2}{2s_i^2 - 1}\right) - 1 \tag{21}$$

and hence

$$\Delta_{IS}^2 = \frac{Q\left(\bar{\underline{\theta}}, \underline{C}, F\right)}{P(F)^2}\left(\Delta_R^2 + 1\right) - 1 \tag{22}$$

which relates Δ_{IS} and Δ_R through the term $Q(\tilde{\theta}, \underline{C}, F)/P(F)^2$.

Using the expressions for Δ_{IS} and Δ_R in (19) and (21), respectively, it can be shown [30] that:

Proposition 2. *Let the parameter PDF q be a standard Gaussian PDF given by (17) and the ISD f be a Gaussian PDF centered at* $\tilde{\underline{\theta}}$ *with covariance matrix* $\underline{C}$, *as given by* (18). *Let* $\left\{s_i^2 : i = 1, \ldots, n\right\}$ *be the eigenvalues of* $\underline{C}$, *where* $s_i > 1/\sqrt{2}$ *for all* $i=1,\ldots,n$, *and* s_i *does not depend on n. The following statements are equivalent:*

1. $\Delta_{IS} < \infty$ as $n \to \infty$
2. $\Delta_R < \infty$ as $n \to \infty$
3. As $n \to \infty, s_i \neq 1$ for at most a finite number of $i \in \mathbb{Z}^+$ and $\left\|\tilde{\underline{\theta}}\right\| < \infty$

Thus, in the case of an ISD centered as a single point, the asymptotic boundedness of Δ_R is a necessary and sufficent condition for that of Δ_{IS}, which supports our intuitiuon that the issue of applicability is solely related to R regardless of the failure region F (under our postulate about F). In terms of the conditions on the ISD, for an ISD centered at a design point $\underline{\theta}^*$ so that $\tilde{\underline{\theta}} = \underline{\theta}^*$ where $\left\|\underline{\theta}^*\right\|\infty$, one can only choose a small number of the principal standard deviations s_i differing from unity. A more general condition on the ISD can be found in [30].

5.2. ISD centered at multiple points

The case of ISDs constructed with multiple points is next considered. We focus on the case when the ISD is constructed as a mixture distribution with Gaussian kernels using the points $\underline{\theta}_1, \ldots, \underline{\theta}_{m_n}$ and identity covariance matrix:

$$f\left(\underline{\theta}\right) = \sum_{i=1}^{m_n} w_i \phi\left(\underline{\theta}; \tilde{\underline{\theta}}_i, \underline{I}\right) = \sum_{i=1}^{m_n} w_i (2\pi)^{-n/2} \exp\left[-\frac{1}{2}\left(\underline{\theta} - \tilde{\underline{\theta}}_i\right)^T \left(\underline{\theta} - \tilde{\underline{\theta}}_i\right)\right] \tag{23}$$

The number of points from which the ISD is constructed, m_n, can possibly depend on the dimension n, which is the case frequently encountered in high dimensional simulation problems.

It can be shown [30] that a sufficient condition for importance sampling to be applicable in high dimensions is

$$\sum_{i=1}^{m_n} w_i \left\| \tilde{\underline{\theta}}_i \right\|^2 < \infty \quad \text{as } n \to \infty \tag{24}$$

For example, it is sufficient to have $\left\|\tilde{\underline{\theta}}_i\right\| < \infty$ for all $i = 1, \ldots, m_n$, since then $\sum_{i=1}^{m_n} w_i \left\|\tilde{\underline{\theta}}_i\right\|^2 \leqslant \left(\sum_{i=1}^{m_n} w_i\right) \max_i \left\|\tilde{\underline{\theta}}_i\right\|^2 = \max_i \left\|\tilde{\underline{\theta}}_i\right\|^2 < \infty$. In the case where $\left\|\tilde{\underline{\theta}}_n\right\| \to \infty$ as $n \to \infty$, (24) requires that the weights should decrease for large n so that the L.H.S. of (24) is bounded.

5.3. ISD with random pre-samples

When the ISD is constructed using design points, it is often true that $\left\|\tilde{\underline{\theta}}_i\right\| < \infty$ for every $i = 1, \ldots, m_n$ and n, although it is possible that $\left\|\tilde{\underline{\theta}}_{m_n}\right\|$ becomes unbounded as $n \to \infty$. In this case, the sufficient condition in (24) can often be achieved by choosing the weights to properly decay with n, as just described. However, the same may not be true for ISDs constructed using random pre-samples simulated by some prescribed stochastic procedure intended to populate the important parts of the failure region. In this case, the ISD f is given by (23) with $\tilde{\underline{\theta}}_i = \hat{\underline{\Theta}}_i$, where $\left\{\hat{\underline{\Theta}} : i = 1, \ldots, m_n\right\}$ are the random pre-samples whose distribution is preferred to be as close as possible to the conditional PDF $q(\cdot \mid F)$ [22,23]. The Euclidean norm of the random pre-samples then often grows with n and becomes unbounded as $n \to \infty$, even if the design point of the failure region remains bounded. Specifically, it is shown in the Appendix that if $\underline{\Theta}$ is distributed as the original parameter PDF q, then $\lim_{n \to \infty} P\left(\left\|\underline{\Theta}\right\| > x \mid F\right) = 1$ for every $x > 0$ and $\mathrm{E}\left[\left\|\underline{\Theta}\right\|^2 \mid F\right] = O(n)$. For example, consider the case where the failure region is a half-space defined by a hyperplane with the design point $\underline{\theta}^*$. It can be readily shown that a random vector $\hat{\underline{\Theta}}$ distributed according to the conditional distribution $q(\cdot \mid F)$ can be represented as [25]

$$\hat{\underline{\Theta}} = \underline{Z} + \left(A - \left\langle \underline{Z}, \underline{u}^* \right\rangle\right) \underline{u}^* \tag{25}$$

where $\underline{u}^* = \underline{\theta}^* / \left\|\underline{\theta}^*\right\|$ is a unit vector in the direction of the design point $\underline{\theta}^*$, $\underline{Z}$ is a standard Gaussian vector and A is a standard Gaussian variable conditional on $A > \left\|\underline{\theta}^*\right\|$. It can be shown [30] that

$$\mathrm{E}\left\|\hat{\underline{\Theta}}\right\|^2 = n - 1 + \mathrm{E}A^2 \tag{26}$$

which is unbounded as $n \to \infty$. The condition in (24) is then violated when the ISD with only one pre-sample ($m_n = 1$) is used.

The remaining question is whether importance sampling is applicable in high dimensions when the ISD is constructed using more than one random pre-sample, which is the usual case of interest. Note that violation of (24) does not immediately imply inapplicability, since it is only a sufficient condition. The unknown factor here is whether the 'interaction' arising from the pre-samples can help prevent Δ_{IS} from growing without bound as n increases. To answer this question, one needs to study Δ_{IS} and hence the variability of $R(\underline{\Theta}) \mathbb{I}_F\left(\underline{\Theta}\right)$ when $\underline{\Theta}$ is distributed as $f = \sum_{i=1}^{m_n} w_i f_i$.

Due to the structure of the ISD, a definite answer to this question has not been sought. It is possible to obtain the c.o.v. of the reciprocal of $R(\underline{\Theta})$, however, which gives insights into the variability of R, assuming that $R(\underline{\Theta})$ has finite variance as $n \rightarrow \infty$ if and only if its reciprocal does. It has been shown [30] that

$$\mathrm{E}_f\left[R\left(\underline{\Theta}\right)^{-1}\right] = \sum_{i,j=1}^{n} w_i w_j \exp\left(\left\langle \tilde{\underline{\theta}}_i, \tilde{\underline{\theta}}_j \right\rangle\right) \tag{27}$$

where $<\cdot,\cdot>$ denotes the Euclidean scalar product, and

$$\mathrm{E}_f\left[R(\underline{\Theta})^{-2}\right] = \sum_{i,j,k=1}^{n} w_i w_j w_k \exp\left(\left\langle \tilde{\underline{\theta}}_i, \tilde{\underline{\theta}}_j \right\rangle + \left\langle \tilde{\underline{\theta}}_j, \tilde{\underline{\theta}}_k \right\rangle + \left\langle \tilde{\underline{\theta}}_i, \tilde{\underline{\theta}}_k \right\rangle\right) \tag{28}$$

By viewing the sets of numbers $\left\{w_i w_j : i, j = 1, \ldots, n\right\}$ and $\left\{w_i w_j w_k : i, j, k = 1, \ldots, n\right\}$ as discrete sets of probabilities and noting that $\left|\left\langle \tilde{\underline{\theta}}_i, \tilde{\underline{\theta}}_j \right\rangle\right| = O\left(\rho^2\right)$ where ρ is some representative scale among $\left\|\tilde{\underline{\theta}}_i\right\|$, $i = 1, \ldots, m_n$, it can be argued [30] that $\mathrm{E}_f[1/R] = O\left(\exp\left(\rho^2\right)\right)$ and $\mathrm{E}_f\left[1/R^2\right] = O\left(\exp\left(3\rho^2\right)\right)$. Consequently, $\Delta^2_{R^{-1}} = \mathrm{E}_f\left[R^{-2}\right]/\mathrm{E}_f\left[R^{-1}\right]^2 - 1 = O\left(\exp\left(\rho^2\right)\right)$. This shows that the variability of $R^{-1}\delta$ increases exponentially with the order of the Euclidean norm of the pre-samples, which suggests that importance sampling using ISD constructed from random pre-samples is not applicable in high dimensions.

6. *Diagnosis for applicability in high dimensions*

The foregoing analysis focuses on the case of a Gaussian parameter PDF q and an ISD f constructed from Gaussian PDFs or their mixtures. In general, it is desirable in a particular application to check whether importance sampling using an ISD from a chosen class of ISDs is applicable or not, before the actual simulation is started. Theoretically, one can estimate Δ_{IS} for the sequence of reliability problems in increasing dimensions and check if it grows without bound with the dimension n. However, this is not computationally favorable, since the estimation of Δ_{IS} involves the evaluation of the indicator function $\mathbb{I}_F$, which requires system analyses. A better strategy is to estimate Δ_R, which only involves estimating the variability of the importance sampling quotient and not the indicator function. Then, if Δ_R remains bounded as n increases, it can be guaranteed by Proposition 1 that Δ_{IS} that is bounded too, and hence importance sampling is applicable in high dimensions for the particular problem. On the other hand, if Δ_R is unbounded as n increases, then it is likely that Δ_{IS} is unbounded too, although the answer is not definite. In this case, one may try to implement importance sampling in high dimensions, and stop the process if Δ_{IS} estimated during the simulation process is large. In the latter case, however, one is cautioned that the importance sampling quotient is likely to be exponentially small. The resulting failure probability estimate may be practically biased, whose large variability may not be detected when the sample size is not sufficiently large. The advice here is that one should be extra cautious when it is found that Δ_R is unbounded as n increases.

Regarding the estimation of Δ_R, it is noted from (6) that,

$$\Delta_R^2 = \frac{1}{P(G)^2}\int_G \frac{q(\underline{\theta})^2}{f(\underline{\theta})^2} f(\underline{\theta})\mathrm{d}\underline{\theta} - 1 = \frac{1}{P(G)}\int_G \frac{q(\underline{\theta})}{f(\theta)}\frac{q(\underline{\theta})\delta\mathbb{I}_G(\underline{\theta})}{P(G)}\mathrm{d}\underline{\theta} - 1 = \frac{\mathrm{E}_{q|G}[R(\underline{\Theta})]}{P(G)} - 1 \tag{29}$$

and so Δ_R can be obtained by estimating $\mathrm{E}_{q|G}[R]$ with samples $\left\{\underline{\Theta}_1, \ldots, \underline{\Theta}_N\right\}$ simulated according to q that lie in G, that is,

$$\mathrm{E}_{q|G}[R] \approx \frac{1}{N}\sum_{k=1}^{N} R(\underline{\Theta}_k) \tag{30}$$

Assuming that checking whether a sample lies in G is not computationally expensive, this approach is preferable to obtaining Δ_R from (6) by estimating $\mathrm{E}_f\left[R^2\right]$ with samples simulated according to f, although the latter is commonly adopted. This is because when R has large variability, the variance of the estimate for $\mathrm{E}_{q|G}[R]$ in (30) is

$$\frac{\mathrm{E}_{q|G}\left[R^2\right] - \mathrm{E}_{q|G}[R]^2}{N} = O\left(\mathrm{E}_{q|G}\left[R^2\right]\right) \tag{31}$$

while in the latter case, the variance is

$$\frac{\mathrm{E}_f\left[R^4\right] - \mathrm{E}_f\left[R^2\right]^2}{N} = \frac{P(G)\mathrm{E}_{q|G}\left[R^3\right] - P(G)^2\mathrm{E}_{q|G}[R]^2}{N} = O\left(\mathrm{E}_{q|G}\left[R^3\right]\right) \tag{32}$$

This means that the variance in the latter could be an order of magnitude greater than the former. The intuitive reason for this reduction of estimation error when Δ_R is obtained by estimating $\mathrm{E}_{q|G}[R]$ is that in this case the samples are simulated from q, and populate in the region within G where $R = q/f$ assumes large values that give the major contribution to $\mathrm{E}_{q|}G[R]$. In contrast, when $\mathrm{E}_f\left[R^2\right]$ is estimated, the samples are simulated from f which are concentrated in the region where R is small. In fact, when the variability of R is very large, this could give a practically biased estimate for its variance.

7. Illustrative examples

7.1. Example 1

This example illustrates our results for the high-dimensional applicability of importance sampling for the case of an ISD constructed with a single point. The parameter PDF q is standard Gaussian, and the failure region F is a half space defined by a design point $\underline{\theta}^*$:

$$F = \left\{\underline{\theta} \in \mathbb{R}^n : \left\langle \underline{\theta}, \underline{\theta}^* \right\rangle \geqslant \beta^2 \right\} \tag{33}$$

where $\underline{\theta}^* = [1, \ldots, 1] \times \beta/\sqrt{n}$ is the design point. The exact failure probability is $P(F) = \Phi(-\beta)$, regardless of the number of uncertain parameters n. The ISD is a Gaussian PDF centered at the design point $\underline{\theta}^*$ with covariance matrix $\underline{C}$:

$$f(\underline{\theta}) = (2\pi)^{-n/2}\sqrt{|\underline{C}^{-1}|}\exp\left[-\frac{1}{2}\left(\underline{\theta}-\underline{\theta}^*\right)^T\underline{C}^{-1}(\underline{\theta}-\underline{\theta}^*)\right] \tag{34}$$

The covariance matrix is assumed to be diagonal with all diagonal elements equal to s^2, so all the n eigenvalues of $\underline{C}$ are equal to s^2. Three cases, cases 1, 2 and 3, corresponding to the values of $s=1$, 0.9 and 1.1, respectively, are considered. Thus, case 1 adopts a covariance matrix equal to that of the original parameter PDF, while cases 2 and 3 adopt a covariance matrix with a slightly smaller and slightly larger spread, respectively.

7.1.1. Typical simulation runs. Figs. 1–3 show the typical importance sampling simulation history for $n=10$, 100, 1000, respectively, where $\beta=3$. When $n=10$ (Fig. 1), the variability of the importance sampling estimate is similar in all the three cases. The estimates for cases 2 and 3 deteriorate as n is increased to 100 (Fig. 2), and become practically biased when $n=1000$ (Fig. 3). The sample c.o.v.s estimated in the same simulation runs are also practically biased and do not reflect truly the large variability of the corresponding failure probability estimates in high dimensions. Note that the estimate in case 1 converges to the exact value regardless of the value of n, indicating that importance sampling is applicable in high dimensions for case 1.

7.1.2. Analysis of c.o.v.. For the current example, it is possible to derive analytical expressions for Δ_{IS} and Δ_R using (19) and (21), respectively [30]. Using these expressions, the variation of Δ_{IS} and Δ_R with the dimension n is compared for different cases in Figs. 4–6. Except for case 1, both Δ_{IS} and Δ_R increase exponentially with n, indicating that importance sampling is not applicable in high dimensions in these cases. This conclusion is precisely what is predicted by Proposition 2 and is consistent with our observation from the typical simulation runs (Figs. 1–3). On the other hand, for given n, although Δ_R increases exponentially with β, the same is not true for Δ_{IS}. The plots of Δ_{IS} for different values of β are very close to each other. In fact, it can be shown [30] that Δ_{IS} increases only linearly with β. The values of Δ_R are seen to be orders of magnitude greater than those of Δ_{IS}, which indicates that the magnitude of Δ_{IS} cannot be inferred from that of Δ_R.

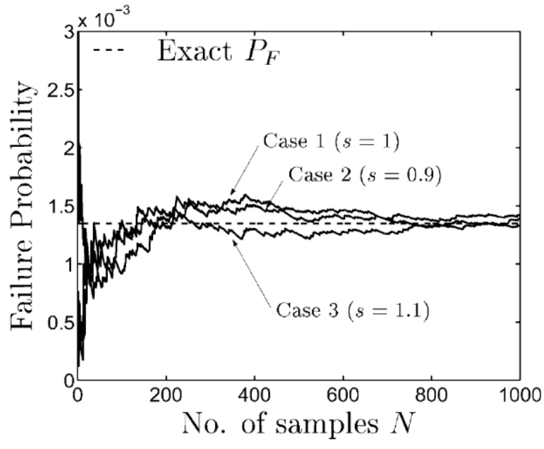

Fig. 1. Simulation histories for $\beta=3$ and $n=10$ (example 1).

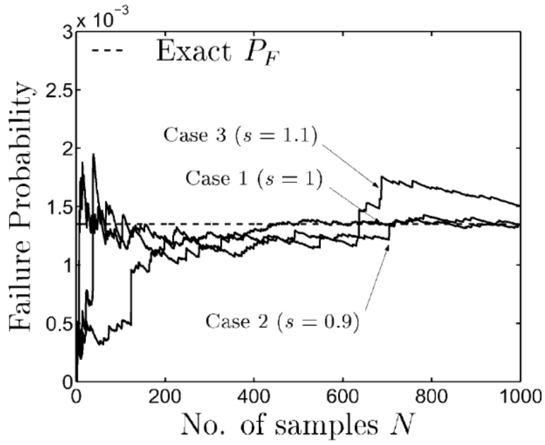

Fig. 2. Simulation histories for $\beta = 3$ and $n = 100$ (example 1).

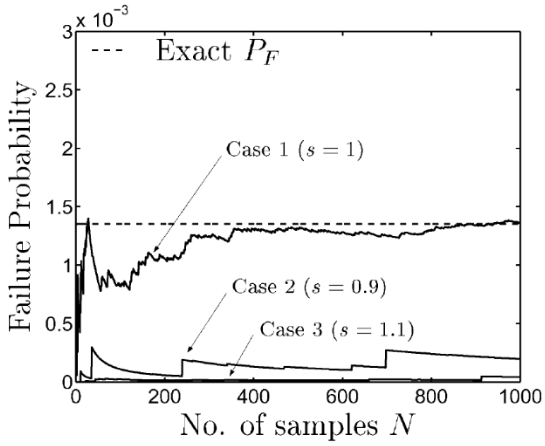

Fig. 3. Simulation histories for $\beta = 3$ and $n = 1{,}000$ (example 1).

Nevertheless, the trend of Δ_{IS} and Δ_R with n are similar; Δ_{IS} remains bounded as n increases whenever Δ_R does (case 1), and Δ_{IS} grows exponentially with n whenever Δ_{IS} does (cases 2–4). This shows that the behavior of Δ_R with n can be used for concluding the behavior of Δ_{IS}, and hence for diagnosing applicability in high dimensions.

7.2. Example 2

This example illustrates our results for the high-dimensional applicability of importance sampling for the case of an ISD constructed from multiple design points. We consider a single-degree-of-freedom oscillator with natural frequency $\omega_1 = 2\pi$ rad/s (1 Hz) and damping ratio $\zeta_1 = 5\%$ that is initially at rest and then subjected to white noise excitation $W(t)$:

$$\ddot{Y}(t) + 2\zeta_1\omega_1\dot{Y}(t) + \omega_1^2 Y(t) = W(t) \qquad (35)$$

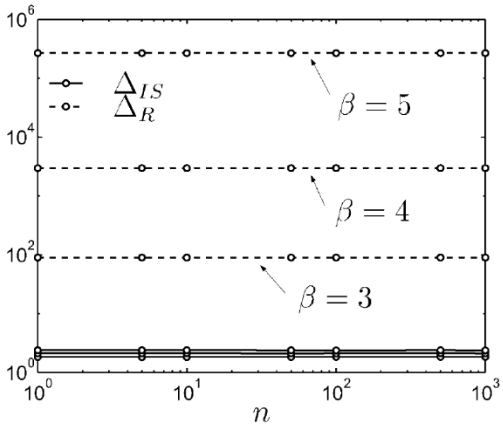

Fig. 4. Variation of Δ_R and Δ_{IS} with n for case 1 ($s=1$) (example 1).

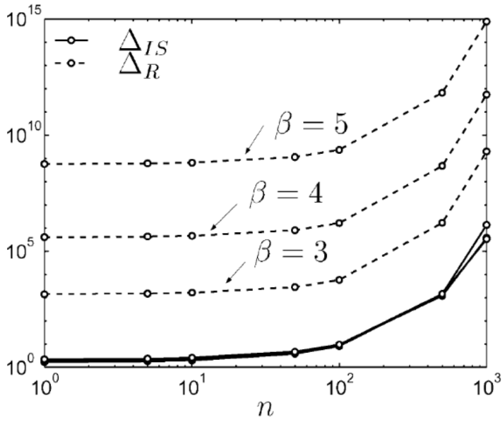

Fig. 5. Variation of Δ_R and Δ_{IS} with n for case 2 ($s=0.9$) (example 1).

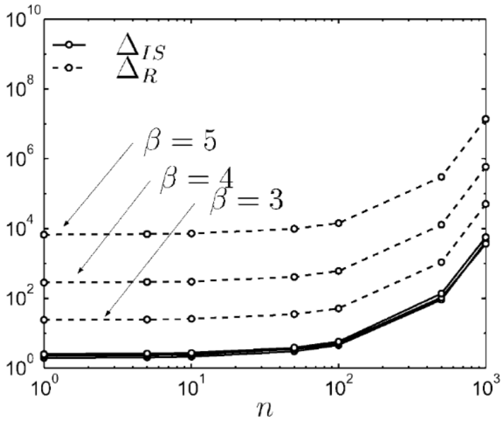

Fig. 6. Variation of Δ_R and Δ_{IS} with n for case 3 ($s=1.1$) (example 1).

The response is computed numerically at the discrete time instants $\{t_k = k\Delta t : k = 1, \ldots, n\}$ where $\Delta t = T_d/n$, $T_d = 10$ sec. and n is the number of time instants. The white noise excitation is assumed to have unit spectral intensity and has the following discrete-time representation:

$$W(t_k) = \sqrt{\frac{2\pi}{\Delta t}}\Theta_k \tag{36}$$

where $\{\Theta_k : k = 1, \ldots, n\}$ are i.i.d. standard Gaussian random variables. The number of uncertain parameters in the problem is thus equal to the number of time instants n considered. The vector $\underline{\Theta} = [\Theta_1, \ldots, \Theta_n]$ collects the n uncertain parameters of the problem, with parameter PDF $q(\underline{\theta}) = \phi_n(\underline{\theta})$ (n-dimensional standard Gaussian joint PDF).

Failure is defined as the exceedence of the displacement response over the threshold level b within the duration of study, that is, the failure region is:

$$F = \bigcup_{k=1}^{n}\left\{\underline{\theta} \in \mathbb{R}^n : Y(t_k; \underline{\theta}) > b\right\} = \bigcup_{k=1}^{n} F_k \tag{37}$$

where $F_\mathrm{k} = \left\{\underline{\theta} \in \mathbb{R}^n : Y(t_k; \underline{\theta}) > b\right\}$ is the 'elementary failure region' corresponding to the failure at time t_k [25]. The failure region is thus a union of n elementary failure regions. Each elementary failure region F_k (k = 1,. . .,n) is a half-space defined by a hyperplane, with the design point $\underline{\theta}_k^*$ given by

$$\underline{\theta}_k^*(j) = \sqrt{2\pi\Delta t}\, b\, U(k-j)\frac{h\left(t_k - t_j\right)}{\sigma_k^2} \tag{38}$$

where $U()$ is the unit step function, $h()$ is the unit impulse response of the oscillator:

$$h(t) = \frac{\exp(-\omega_1\zeta_1 t)}{\omega_1\sqrt{1-\zeta_1^2}}\sin\left(\omega_1\sqrt{1-\zeta_1^2}\, t\right) \tag{39}$$

and σ_k is the standard deviation of the response at time t_k, given by

$$\sigma_k^2 = 2\pi\Delta t\sum_{j=1}^{k} h\left(t_j\right)^2 \tag{40}$$

The importance sampling density is constructed as a weighted sum of Gaussian distributions centered among the design points $\left\{\underline{\theta}_k^* : k = 1, \ldots, n\right\}$ with covariance matrix $\underline{C}$:

$$f(\underline{\theta}) = \sum_{k=1}^{n} w_k (2\pi)^{-n/2}\sqrt{\left|\underline{C}^{-1}\right|}\exp\left[-\frac{1}{2}\left(\underline{\theta} - \underline{\theta}_k^*\right)^T \underline{C}^{-1}\left(\underline{\theta} - \underline{\theta}_k^*\right)\right] \tag{41}$$

where the weights are chosen to be proportional to the probability content of the elementary failure regions, that is, $w_k = \Phi(-\|\underline{\theta}_k^*\|)/\sum_{j=1}^{n}\Phi(-\|\underline{\theta}_j^*\|)$. Similar to Example 1, the covariance matrix $\underline{C}$ is assumed to be a diagonal matrix with all diagonal entries equal to s. Three cases, cases 1–3 are considered, corresponding to $s=1$, 0.9 and 1.1, respectively. The threshold level b is taken to be a multiple of the response standard deviation at time T_d, i.e., $b=d\times\sigma_{n_t}$, where the values $d=3,4,5$ will be studied. Note that with $q(\underline{\theta}) = \phi_n(\underline{\theta})$ and $f(\underline{\theta})$ given by (41) with $\underline{C} = s\underline{I}$, the importance sampling quotient $R(\underline{\theta}) = q(\underline{\theta})/f(\underline{\theta})$ becomes

$$R(\underline{\theta}) = s^n\left[\sum_{k=1}^{n} w_k \exp\left(-\frac{1}{2s^2}\|\underline{\theta}_k^*\|^2 - \frac{1}{2}\left(\frac{1}{s^2}-1\right)\|\underline{\theta}\|^2 + \frac{1}{s^2}\langle\underline{\theta}_k^*, \underline{\theta}\rangle\right)\right]^{-1} \tag{42}$$

7.2.1. Typical simulation runs. Fig. 7 shows the variation of the failure probability estimates with n computed using $s=1$ and $N=10{,}000$ samples. As expected, for each threshold level, the failure probability estimates always converge as n increases. Figs. 8–10 show the typical importance sampling simulation history for $n=100$, 500, 1000, respectively, where $d=4$. As in the last example, these figures clearly show that importance sampling is applicable in high dimensions in case 1 but not in cases 2 or 3. In particular, it can be seen that when n is large ($n=500$ in Fig. 9 and $n=1000$ in Fig. 10), the estimates for case 3 develop an intermittent trend of $1/N$ after a big jump in the beginning. This is because, after a significantly large value of the importance sampling quotient R is sampled which gives rise to the big jump in the beginning, most of the subsequent sampled values of R are exponentially small, and thus the estimates essentially behave as a $1/N$ decay rather than as a converging sum.

7.2.2. Analysis of c.o.v.. In this example, as is common in applications, analytical expressions for Δ_R and Δ_{IS} are not available, and so they are evaluated by simulation. The value of Δ_R is

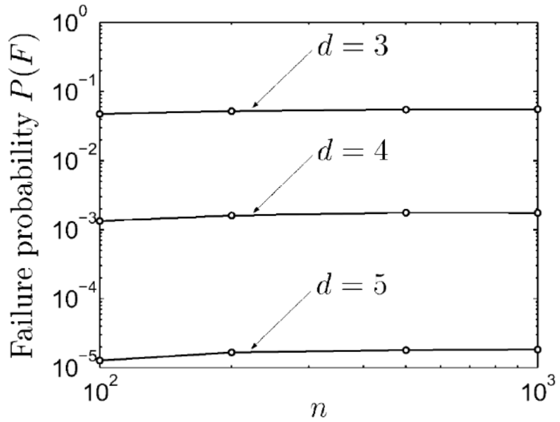

Fig. 7. Variation of failure probability estimates with n for different values of threshold level d in example 2 (computed using $s=1$ and $N=10{,}000$ samples).

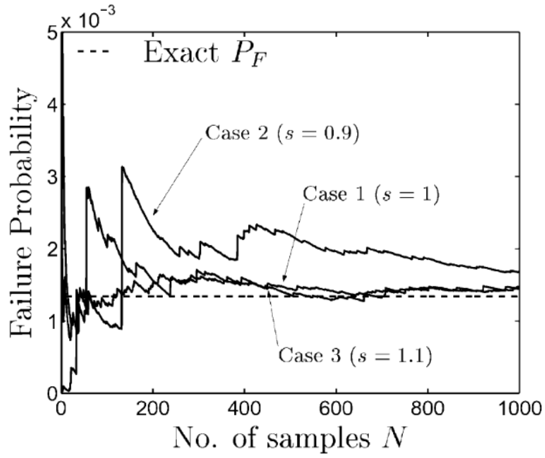

Fig. 8. Simulation histories for $d = 4$ and $n = 100$ (example 2).

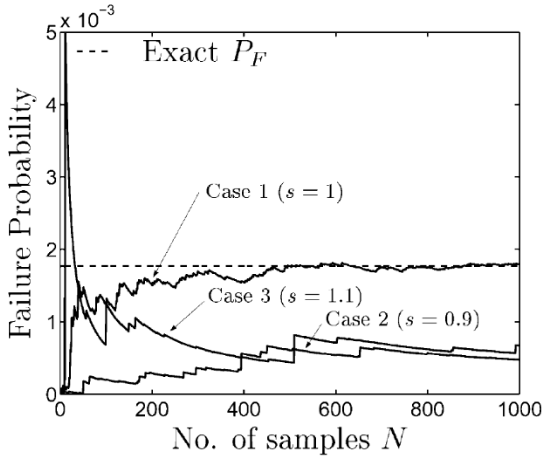

Fig. 9. Simulation histories for $d = 4$ and $n = 500$ (example 2).

computed based on (29), where it is noted that $P(G) = P(\mathbb{R}^n) = 1$ and $\mathrm{E}_{q|G}[R]$ is estimated based on (30) and (42) with 10,000 samples simulated from $q = \phi_n$. The unit c.o.v. Δ_{IS} is evaluated based on (5) using 10,000 samples simulated according to the ISD f in (41). Note that the demanding computation of Δ_{IS} is performed here for illustration purposes. In practice, only the computation of Δ_R is recommended.

Figs. 11 to 13 show the variation of the estimates of Δ_R and Δ_{IS} with n. In case 1 (Fig. 11), the unit c.o.v. Δ_{IS} is approximately 2 for different values of d and is practically independent of n. In case 2 (Fig. 12) and case 3 (Fig. 13), both Δ_R and Δ_{IS} grow exponentially with n, indicating that importance sampling is not applicable in these cases. In particular, for $n = 1000$, the values of Δ_{IS} in cases 2 and 3 are, respectively, 5 and 3 orders of magnitude greater than that in case 1. The variation of Δ_{IS} with n is similar to the variation of Δ_R; Δ_{IS} remains bounded as n increases whenever Δ_R does (case 1), and Δ_{IS} grows exponentially with n whenever Δ_R does (cases 2 and 3). This suggests that, in this example, the asymptotic boundedness of Δ_R could be a necessary

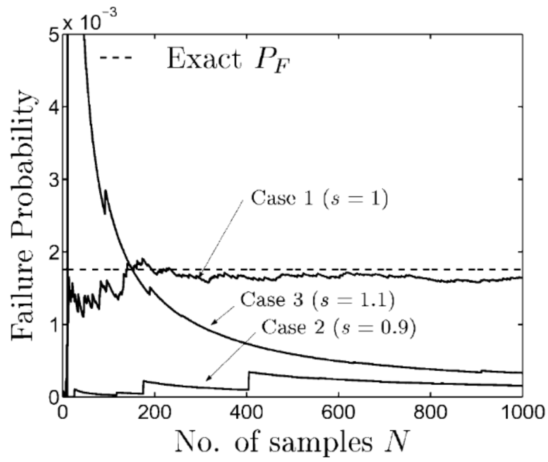

Fig. 10. Simulation histories for $d=4$ and $n=1000$ (example 2).

and sufficient condition for that of Δ_{IS}, although only the sufficiency part has been proven for the multiple design point case in this work.

8. Conclusions

A pioneering study has been devoted to the applicability of importance sampling in high-dimensional problems. A formal definition for this applicability has been given, and the conditions for it have been investigated for the common case where the PDF (probability density function) for the uncertain parameters is a standard multi-dimensional Gaussian PDF and the ISD (importance sampling density) is a mixture distribution of Gaussian PDFs centered among one or more points.

In summary, for ISDs centered at a single design point, it is required to have the covariance matrix almost equal to the identity matrix (in the standard Gaussian case). The same is true for ISDs centered on more than one design point. The use of an ISD centered among one or more

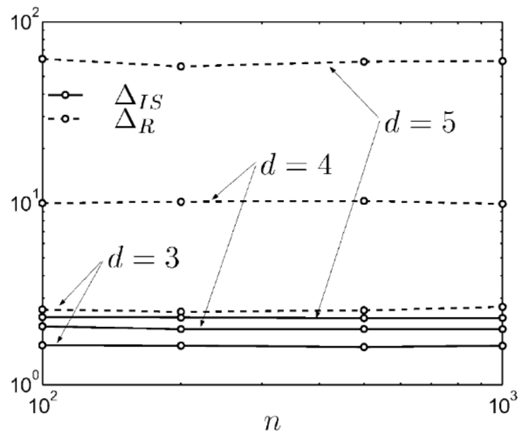

Fig. 11. Variation of Δ_R and Δ_{IS} with n for case 1 ($s=1$) (example 2).

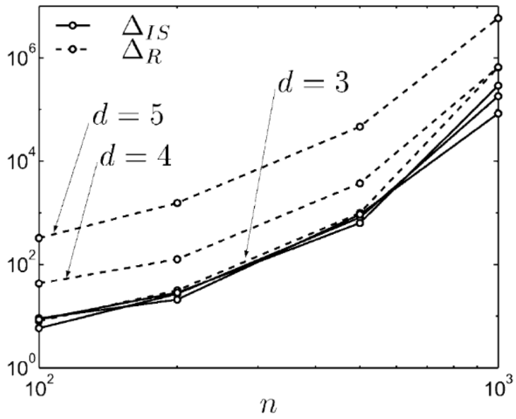

Fig. 12. Variation of Δ_R and Δ_{IS} with n for case 2 ($s=0.9$) (example 2).

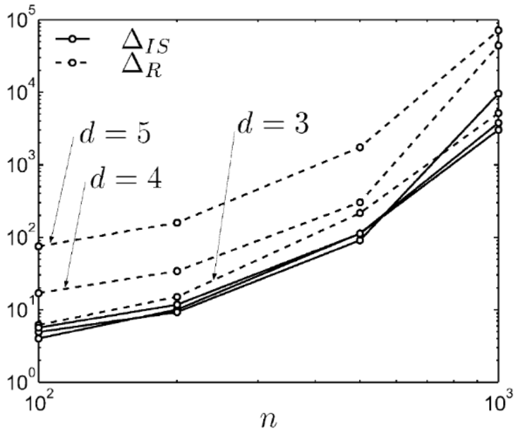

Fig. 13. Variation of Δ_R and Δ_{IS} with n for case 3 ($s=1.1$) (example 2).

random pre-samples is not applicable in high dimensions since the Euclidean norm of the pre-samples is generally unbounded as the dimension of the problem increases.

The numerical example on the computation of the first-passage probability of an oscillator suggests that there appears to be little flexibility in the choice of the covariance matrices associated with each design point. It discourages the use of covariance matrices significantly different from the identity matrix (in the standard Gaussian case) that is intended to yield more efficient estimates. It is doubtful whether the use of other functional forms for the ISD other than that for original parameter PDF (Gaussian) may improve applicability. It is apparent that the ISD and the original parameter PDF have to be of the same functional form, because the relative entropy of the PDFs will likely grow without bound in high dimensions, rending importance sampling inapplicable because of (11).

We have proven that the asymptotic boundedness of the c.o.v. of the importance sampling quotient is a sufficient condition for that of the c.o.v. of the importance sampling estimator. Although the necessity part of the statement has been proved only for the case of importance sampling density centered at a single point and has not been given for the general case, we speculate that it is true in general, possibly with some additional assumption on the failure region. It is important to confirm this statement so that the problem of diagnosing the applicability of importance sampling in high dimensions is reduced to investigating the behavior of the importance sampling quotient, which can be handled numerically, if not analytically, without system analysis.

Acknowledgements

This paper is based upon work partly supported by the Pacific Earthquake Engineering Research Center under National Science Foundation Cooperative Agreement No. CMS-9701568. This support is gratefully acknowledged. Part of this work was prepared when the first author was at graduate standing at the California Institute of Technology, whose generous financial support is gratefully acknowledged.

Appendix

This appendix proves two propositions which show that the Euclidean norm of a random vector with i.i.d. components is unbounded as $n \to \infty$, regardless of the conditioning by the failure region F. In these propositions, we deliberately stress the condition that $P(F)$ is assumed to be non-vanishing as $n \to \infty$, although this is already implicit throughout our study. The propositions can be extended to the general case when the components of $\underline{\Theta}$ are not identically distributed and/or when the one-dimensional distribution of each component depends on n, under the assumption that the Lindeberg Condition [31] is satisfied, in addition to the conditions given in the propositions. However, this generalization is not necessary in our discussion and will not be further pursued.

Proposition 3. *Let $\underline{\Theta} = [\Theta_1, \ldots, \Theta_n]$ be an n-dimensional random vector with i.i.d. components where the distribution of each component of $\underline{\Theta}$ does not depend on n. Let Θ be a generic random variable identically distributed as each component Θ_i. Let the failure region F be such that $P(F)$ is non-vanishing as $n \to \infty$, then for any non-negative sequence $\{x_n \geqslant 0 : n = 1, 2, \ldots\}$ with $x_n \leqslant c\sqrt{n}$ for sufficiently large n and $c \geqslant \sqrt{\mathrm{E}[\Theta^2]}$ is a fixed positive constant,*

$$\lim_{n \to \infty} P\left(\|\underline{\Theta}\| \leqslant x_n \mid F\right) = 0 \tag{43}$$

Proof. According to Bayes' Theorem,

$$P\left(\|\underline{\Theta}\| < x_n \mid F\right) = \frac{P\left(F \mid \|\underline{\Theta}\| \leqslant x_n\right) P\left(\|\underline{\Theta}\| \leqslant x_n\right)}{P(F)} \leqslant \frac{P\left(\|\underline{\Theta}\| \leqslant x_n\right)}{P(F)} \tag{44}$$

since $P\left(F \mid \|\underline{\Theta}\| \leqslant x_n\right) \leqslant 1$. By the Central Limit Theorem, $\|\underline{\Theta}\|^2/n = \sum_{i=1}^{n} \Theta_i^2/n$ is asymptotically. Normally distributed with mean $\mathrm{E}\left[\Theta^2\right]$ and variance $\mathrm{Var}\left[\Theta^2\right]/n$ as $n \to \infty$, so

$$P\left(\|\underline{\Theta}\| \leqslant x_n\right) = P\left(\|\underline{\Theta}\|^2/n \leqslant x_n^2/n\right) = \Phi\left(\frac{x_n^2/n - \mathrm{E}\left[\Theta^2\right]}{\mathrm{Var}\left[\Theta^2\right]/n}\right)$$
$$= \Phi\left(\frac{x_n^2 - n\mathrm{E}\left[\Theta^2\right]}{\mathrm{Var}[\Theta^2]}\right) \tag{45}$$

For sufficiently large n, $x_n \leqslant c\sqrt{n}$, so

$$\frac{x_n^2 - n\mathrm{E}\left[\Theta^2\right]}{\mathrm{Var}[\Theta^2]} \leqslant n\frac{c^2 - \mathrm{E}\left[\Theta^2\right]}{\mathrm{Var}[\Theta^2]} \tag{46}$$

which tends to $-\infty$ since $c^2 - \mathrm{E}\left[\Theta^2\right] < 0$. Consequently, according to (45),

$$\lim_{n\to\infty} P\left(\|\underline{\Theta}\| \leqslant x_n\right) = 0 \tag{47}$$

and the proof follows by applying (47) to (44) and using the fact that $P(F)$ does not vanish.

Corollary 1. *In the context of Proposition 3*,

$$\lim_{n\to\infty} P\left(\|\underline{\Theta}\| < x \mid F\right) = 0 \tag{48}$$

for every $x > 0$.

Proof. Take $x_n = x$ for $n = 1,2,\ldots$, in Proposition 3. □

Proposition 4. *In the context of Proposition 3*,

$$\mathrm{E}\left[\|\underline{\Theta}\|^2 \mid F\right] = O(n) \quad as\ n \to \infty \tag{49}$$

Proof. First of all, since $\mathrm{E}\left[\|\underline{\Theta}\|^2 \mid F\right]P(F) \leqslant \mathrm{E}\left[\|\underline{\Theta}\|^2\right] = n\mathrm{E}\left[\Theta^2\right]$, we have

$$\lim_{n\to\infty} \frac{\mathrm{E}\left[\|\underline{\Theta}\|^2 \mid F\right]}{n} \leqslant \frac{\mathrm{E}\left[\Theta^2\right]}{P(F)} \tag{50}$$

On the other hand, by Markov's inequality, for every $x_n > 0$,

$$\frac{\mathrm{E}\left[\|\underline{\Theta}\|^2 \mid F\right]}{x_n^2} \geqslant P\left(\|\underline{\Theta}\|^2 > x_n^2 \mid F\right) = P\left(\|\underline{\Theta}\| > x_n \mid F\right) \tag{51}$$

Take $x_n = c\sqrt{n}$ with $c < \sqrt{E[\Theta^2]}$ fixed, and let $n \to \infty$ in (51),

$$\lim_{n\to\infty} \frac{E\left[\|\underline{\Theta}\|^2 | F\right]}{c^2 n} \geqslant \lim_{n\to\infty} P\left(\|\underline{\Theta}\| > x_n \mid F\right) = 1 - \lim_{n\to\infty} P\left(\|\underline{\Theta}\| \leqslant x_n \mid F\right) = 1 \tag{52}$$

since $\lim_{n\to\infty} P\left(\|\underline{\Theta}\| \leqslant x_n \mid F\right) = 1$ by Proposition 3. Combining (50) and (52), we have

$$0 < c^2 \leqslant \lim_{n\to\infty} \frac{E\left[\|\underline{\Theta}\|^2 | F\right]}{n} \leqslant \frac{E\left[\Theta^2\right]}{P(F)} < \infty \tag{53}$$

and hence $E\left[\|\underline{\Theta}\|^2 | F\right] = O(n)$ as $n \to \infty$. □

References

[1] Vision, 2000. Performance based seismic engineering of buildings. Technical report, Sacramento, California, 2000.

[2] Wen YK. Reliability and performance based design. In Proceedings of 8th ASCE Specialty Conference on Probabilistic Mechanics and Structural Reliability. Indiana: Notre Dame; 2000.

[3] Cornell CA. Reliability-based earthquake-resistant design: the future. In: *Proceedings of 11th World Conference on Earthquake Engineering*, Acapulco, Mexico; 1996.

[4] Freudenthal AM. The safety of structures. Transactions of ASCE 1947;112:125–80.

[5] Freudenthal AM. Safety and the probability of structural failure. Transactions of ASCE 1956;12:1337–97.

[6] Cox RT. The algebra of probable inference. Baltimore: Johns Hopkins Press; 1961.

[7] Hammersley JM, Handscomb DC. Monte-Carlo methods. London: Methuen; 1964.

[8] Rubinstein RY. Simulation and the Monte-Carlo method. New York, NY: John Wiley & Sons Inc; 1981.

[9] Fishman GS. Monte Carlo: concepts, algorithms and applications. New York: Springer–Verlag; 1996.

[10] Schueller GI, Stix R. A critical appraisal of methods to determine failure probabilities. Structural Safety 1987;4: 293–309.

[11] Engelund S, Rackwitz R. A benchmark study on importance sampling techniques in structural reliability. Structural Safety 1993;12:255–76.

[12] Hohenbichler M, Rackwitz R. Improvement of second-order reliability estimates by importance sampling. Journal of Engineering Mechanics, ASCE 1988;114(12):2195–8.

[13] Harbitz A. An efficient sampling method for probability of failure calculation. Structural Safety 1986;3:109–15.

[14] Papadimitriou C, Beck JL, Katafygiotis LS. Asymptotic expansions for reliabilities and moments of uncertain dynamic systems. Journal of Engineering Mechanics, ASCE 1997;123(12):12–29.

[15] Au SK, Papadimitriou C, Beck JL. Reliability of uncertain dynamical systems with multiple design points. Structural Safety 1999;21:113–33.

[16] Liu PL, Der Kiureghian A. Optimization algorithms for structural reliability. Structural Safety 1991;9:161–78.

[17] Melchers RE. Importance sampling in structural systems. Structural Safety 1989;6:3–10.

[18] Der Kiureghian A, Dakessian T. Multiple design points in first and second-order reliability. Structural Safety 1998;20:37–49.

[19] Bucher CG. Adaptive sampling—an iterative fast Monte Carlo procedure. Structural Safety 1988;5:119–26.

[20] Karamchandani, A, Bjerager, P. Cornell CA. Adaptive importance sampling. In: Proc. 5th ICOSSAR, San Francisco; 1989: 855–862.

[21] Melchers RE. Search-based importance sampling. Structural Safety 1990;9:117–28.

[22] Ang GL, Ang AHS, Tang WH. Optimal importance sampling density estimator. Journal of Engineering Mechanics, ASCE 1992;118(6):1146–63.

[23] Au SK, Beck JL. A new adaptive importance sampling scheme. Structural Safety 1999;21:135–58.
[24] Schueller, GI, Pradlwarter, HJ, Pandey MD. Methods for reliability assessment of nonlinear systems under stochastic dynamic loading-a review. In: Proceedings of EURODYN'93, Balkema; 1993. p. 751–759.
[25] Au SK, Beck JL. First-excursion probabilities for linear systems by very efficient importance sampling. Probabilistic Engineering Mechanics 2001;16:193–207 [Software available at URL http://www.ntu.edu.sg/home/cskau].
[26] Au SK, Beck JL. Estimation of small failure probabilities in high dimensions by subset simulation. Probabilistic Engineering Mechanics 2001;16(4):263–77.
[27] Kullback S. Information theory and statistics. New York: Wiley; 1959.
[28] Renyi A. Probability theory. North-Holland series in applied mathematics and mechanics (Vol. 10). Amsterdam: North-Holland; 1970.
[29] Jumarie G. Relative information: theories and applications. *Springer Series in Synergetics* (vol. 47). Springer-Verlag; New York: 1990.
[30] Au SK. On the solution of first excursion problems by simulation with applications to probabilistic seismic performance assessment. PhD thesis in Civil Engineering, EERL Report No. 2001–02, California Institute of Technology, Pasadena, 2001. [Available at URL http://www.ntu.edu.sg/home/cskau].
[31] Gnedenko BV, Kolmogorov AN. Limit distributions for sums of independent random variables. Addison-Wesley; 1968.