Assessment of MCMC algorithms for subset simulation

I. Papaioannou, K. Zwirglmaier & D. Straub

Engineering Risk Analysis Group, Technische Universität München, Germany
iason.papaioannou@tum.de

ABSTRACT: The subset simulation is an adaptive simulation method that efficiently solves structural reliability problems with a large number of random variables. The method includes sampling from conditional distributions, which is achieved through Markov Chain Monte Carlo (MCMC) algorithms. This paper investigates the performance of different MCMC algorithms for subset simulation. It is found that most of the MCMC algorithms proposed in the literature, based on the Metropolis-Hastings (M-H) sampler, do not present significant improvements over the component-wise M-H algorithm originally proposed for subset simulation in [Au & Beck, Prob Eng Mech, 16(4): 263-277, 2001]. Based on these findings, a novel approach for MCMC sampling in the standard normal space is introduced, which has the benefit of simplicity. Moreover, it is shown that an optimal scaling of either this new approach or the component-wise M-H algorithm can improve the accuracy of the original algorithm, without the need for additional model evaluations.

1 INTRODUCTION

Let $\mathbf{X}$ denote the basic random variable space of dimension n, which models the system variables that are expected to present an uncertain behavior. Assuming the usual case, where the probabilistic description of $\mathbf{X}$ comes in terms of marginal distributions and correlations, we can adopt the Nataf model (Gaussian copula) for the joint distribution of $\mathbf{X}$ and then define a mapping $\mathbf{U} = \mathbf{T}(\mathbf{X})$ to a transformed space $\mathbf{U}$ consisting of n independent standard normal random variables (Der Kiureghian & Liu 1986). In the case where the joint distribution of $\mathbf{X}$ is known, then the transformation $\mathbf{U} = \mathbf{T}(\mathbf{X})$ can be performed in a straightforward fashion (Hohenbichler & Rackwitz 1981). Let $F \subset \mathbb{R}^n$ be the failure domain in the $\mathbf{U}$-space, such that $\mathbf{u} \in F$ defines the event of unsatisfactory performance of the system. The probability of failure can then be expressed as follows:

$$P_f = P(F) = \int_{\mathbf{u}\in F} \varphi_n(\mathbf{u})\, d\mathbf{u} = \int_{\mathbf{u}\in\mathbb{R}^n} I_F(\mathbf{u})\varphi_n(\mathbf{u})\, d\mathbf{u} \tag{1}$$

where $\varphi_n(\mathbf{u}) = \prod_{i=1}^{n} \varphi(u_i)$ and $\varphi(.)$ is the standard normal PDF. The function $I_F(\mathbf{u})$ is the indicator function: $I_F(\mathbf{u}) = 1$ if $\mathbf{u} \in F$ and $I_F(\mathbf{u}) = 0$ otherwise.

The evaluation of the probability of failure is not a trivial task, especially when the performance of the system for a realization of the random variables is obtained through a computationally demanding model evaluation. The Monte Carlo method is a robust technique that is able to handle any model, independent of its complexity. In this method, the probability of failure is estimated by the sample mean of the indicator function:

$$P_f \approx \hat{P}_f = \frac{1}{N}\sum_{k=1}^{N} I_F(\mathbf{u}_k) \tag{2}$$

where $\{\mathbf{u}_k : k = 1, \dots, N\}$ are independent and identically distributed samples of the joint PDF $\varphi_n(\mathbf{u})$. Equation 2 gives an unbiased estimate of P_f. The coefficient of variation of the estimate reads:

$$\delta_{\hat{P}_f} = \sqrt{\frac{1 - P_f}{N P_f}} \tag{3}$$

As shown in Equation 3, the coefficient of variation, which serves as a measure of accuracy of the estimate, does not depend on the dimension n. Hence, the efficiency of the Monte Carlo method does not depend on the number of random variables. However, for small values of P_f, a large number of samples N is required to achieve a sufficiently small $\delta_{\hat{P}_f}$.

In order to overcome the inefficiency of the Monte Carlo method in estimating small failure probabilities, while maintaining its independency on the number of random variables n, a number of advanced simulation methods have been developed, including the subset simulation (Au & Beck 2001), the spherical subset simulation (Katafygiotis & Cheung 2007) and the asymptotic sampling method (Bucher 2009). Here, we focus on the subset simulation. This method expresses P_f as a product of conditional probabilities that are significantly larger than P_f. These conditional probabilities are then estimated by application of Markov Chain Monte Carlo (MCMC) sampling.

This paper discusses different MCMC algorithms that have been proposed for subset simulation. Moreover, a new approach is presented for MCMC in the **U**-space. In Section 2, the subset simulation is described. Section 3 reviews the considered MCMC methods and evaluates their performance.

2 SUBSET SIMULATION

The subset simulation is an adaptive Monte Carlo method proposed by Au & Beck (2001) for the estimation of small failure probabilities in high dimensional problems. The idea behind subset simulation is to express the failure domain F as the intersection of M larger intermediate failure domains:

$$F = \bigcap_{j=1}^{M} F_j \tag{4}$$

where $F_1 \supset F_2 \supset \cdots \supset F_M = F$, hence the name "subset simulation". The probability of failure is then estimated as a product of conditional probabilities:

$$P_f = P(F) = P\left(\bigcap_{j=1}^{M} F_j\right) = P(F_1) \prod_{j=2}^{M} P(F_j | F_{j-1}) \tag{5}$$

An appropriate selection of the intermediate failure domains can lead to large conditional probabilities. That is, the original problem of evaluating a small failure probability reduces to a sequence of M intermediate problems that correspond to the estimation of larger probabilities.

The probability $P(F_1)$ can be computed by application of crude Monte Carlo simulation through sampling of $\varphi_n(\mathbf{u})$. For estimating the probabilities $\{P(F_j | F_{j-1}) : j = 2, \dots, M\}$, we need to generate samples of the conditional PDFs $\{\varphi_n(\mathbf{u}|F_{j-1}) : j = 2, \dots, M\}$, where:

$$\varphi_n(\mathbf{u}|F_{j-1}) = \frac{\varphi_n(\mathbf{u}) I_{F_{j-1}}(\mathbf{u})}{P(F_{j-1})} \tag{6}$$

A direct sampling of $\varphi_n(\mathbf{u}|F_{j-1})$ by application of the acceptance-rejection method is inefficient, especially as the domain F_{j-1} approaches the actual failure domain. However, MCMC techniques can be applied for sampling $\varphi_n(\mathbf{u}|F_{j-1})$. MCMC methods produce samples of a target distribution, by constructing a Markov chain that has the target distribution as its stationary distribution (Rubinstein & Kroese 2007). The derived samples will be distributed according to $\varphi_n(\mathbf{u}|F_{j-1})$, however they will not be independent.

We now need to specify the intermediate failure domains $\{F_j: j = 1, \dots, M-1\}$. Let $G(\mathbf{u})$ denote a limit-state function with negative values defining the failure event, i.e. $F = \{\mathbf{u} \in \mathbb{R}^n: G(\mathbf{u}) \leq 0\}$. Without loss of generality, we assume that the function $G(\mathbf{u})$ can express any type of system failure in the **U**-space. The intermediate failure domain F_j can then be defined as $F_j = \{\mathbf{u} \in \mathbb{R}^n: G(\mathbf{u}) \leq c_j\}$, where $c_1 > c_2 \dots > c_M = 0$. The values of $\{c_j: j = 1, \dots, M-1\}$ can be chosen adaptively, so that the estimates of the conditional probabilities correspond to a chosen value p_0, e.g. $p_0 = 0.1$. This is achieved by successively sampling each conditional PDF $\varphi_n(\mathbf{u}|F_{j-1})$ and setting c_i equal to the $[(1-p_0)N+1]$-th largest value among the samples $\{G(\mathbf{u}_k): k = 1, \dots, N\}$. This procedure is repeated until the actual failure domain F_M is reached, for which the threshold $c_M = 0$ is given. We can then obtain an estimate of the failure probability as follows:

$$P_f \approx \hat{P}_f = p_0^{M-1}\hat{P}(F_M|F_{M-1}) \tag{7}$$

The estimate $\hat{P}(F_M|F_{M-1})$ of the conditional probability is as follows:

$$\hat{P}(F_M|F_{M-1}) = \frac{1}{N}\sum_{k=1}^{N} I_F(\mathbf{u}_k) \tag{8}$$

where $\{\mathbf{u}_k: k = 1, \dots, N\}$ are samples of $\varphi_n(\mathbf{u}|F_{M-1})$. It should be noted that the estimator $\hat{P}_f$ in Equation 7 is biased for a finite N, due to the correlation between the estimates of the conditional probabilities, but it is asymptotically unbiased (Au & Beck 2001).

3 MCMC ALGORITHMS FOR SUBSET SIMULATION

As discussed in Section 2, the subset simulation applies MCMC sampling to simulate each conditional PDF $\varphi_n(\mathbf{u}|F_{j-1})$. In this section, first the general principle of MCMC sampling for subset simulation is summarized, followed by a review of different proposed and new methods.

Let us define a stationary discrete-time vector random process $\{\mathbf{U}_t, t \in \mathbb{N}\}$ with marginal PDF $\varphi_n(\mathbf{u}|F_{j-1})$, which possesses the Markov property. That is, the random vector at position t is distributed according to a conditional PDF given the outcome of the random vector at position $t-1$. This conditional PDF is termed transition PDF and is denoted by $p(\mathbf{u}_1|\mathbf{u}_0)$, where $\mathbf{u}_0$, $\mathbf{u}_1$ are subsequent states of the chain. The transition PDF satisfies the following condition:

$$p(\mathbf{u}_1|\mathbf{u}_0)\varphi_n(\mathbf{u}_0|F_{j-1}) = p(\mathbf{u}_0|\mathbf{u}_1)\varphi_n(\mathbf{u}_1|F_{j-1}) \tag{9}$$

The above is termed reversibility condition and it is an essential property of the Markov process $\mathbf{U}_t$, since it ensures that the stationary PDF of the process is $\varphi_n(\mathbf{u}|F_{j-1})$.

MCMC methods produce samples of a distribution by simulating states of a stationary Markov process whose marginal distribution is the desired distribution. This can be achieved by simulating every new state of the process from a transition PDF $p(\mathbf{u}_1|\mathbf{u}_0)$ that satisfies the reversibility condition. Starting from a state that may or may not be distributed according to the target distribution, the Markov chain will asymptotically converge to the target (stationary) distribution (Rubinstein & Kroese 2007). The transient period that is required until the Markov chain reaches its stationary state is termed burn-in period. Moreover, the generated samples will be correlated according to the correlation of the Markov process which will depend on the particular choice of $p(\mathbf{u}_1|\mathbf{u}_0)$.

In the context of subset simulation, MCMC sampling is applied at subset j to sample $\varphi_n(\cdot\,|F_{j-1})$ through simulating states of a Markov chain using as starting point (or 'seed') each sample $\{\mathbf{u}_k: k = 1, \dots, N_f\}$ that fell in F_{j-1} at subset $j-1$. Since all the samples $\mathbf{u}_k$ are distributed according to $\varphi_n(\cdot\,|F_{j-1})$, all states of the Markov chains will be distributed according to the target distribution $\varphi_n(\cdot\,|F_{j-1})$. Hence, the Markov chains do not require a burn-in period to reach their stationary states. The coefficient of variation of the estimate of each conditional probability $\hat{P}_j = \hat{P}(F_j|F_{j-1})$ can be estimated in terms of the sample variance of the stationary process $I_{F_j}(\mathbf{U}_t)$ (Au & Beck 2001):

$$\delta_{\hat{P}_j} = \sqrt{\frac{1-P_j}{NP_j}\left(1+\gamma_j\right)} \tag{10}$$

where

$$\gamma_j = 2\sum_{k=1}^{N/N_f-1}\left(1-\frac{kN_f}{N}\right)\rho_j(k) \tag{11}$$

where N_f is the number of seeds, N/N_f is the length of each chain and $\rho_j(k)$ is the auto-correlation coefficient of the sequence $\{I_{F_j}(\mathbf{u}_t), t = 1, \ldots, N/N_f\}$. $\rho_j(k)$ can be estimated from the samples. The estimator of Equation 10 assumes that the different chains are uncorrelated through the indicator function, i.e. possible dependence between the different seeds is neglected. Equation 10 indicates that the efficiency of the subset simulation decreases if the chain correlation increases. A decreased chain correlation implies that the chain explores its state space faster. This motivates the introduction of a new measure of the efficiency of the estimator $\hat{P}_j$, based on the expected Euclidean distance between two successive samples, say $\mathbf{u}_0$ and $\mathbf{u}_1$:

$$\lambda_j = \frac{1}{\sqrt{n}}\mathrm{E}\left[\sqrt{\sum_{i=1}^{n}(u_{0i}-u_{1i})^2}\right] \tag{12}$$

where u_{0i} is the i-th coordinate of $\mathbf{u}_0$. λ_j can be viewed as the average velocity of the different chains. A maximum λ_j will give a minimum chain dependence and hence a minimum coefficient of variation $\delta_{\hat{P}_j}$.

In the following, we assess the efficiency of different MCMC algorithms for subset simulation.

3.1 *Metropolis-Hastings algorithm*

The Metropolis-Hastings (M-H) algorithm (Metropolis et al. 1953, Hastings 1970) is the most widely used MCMC method for sampling from distributions that are difficult to sample from directly. Consider a n-dimensional proposal PDF $q(\cdot\,|\mathbf{u}_0)$ that depends on the current state of the chain. The transition from the state $\mathbf{u}_0$ to the next state $\mathbf{u}_1$ is described by the following steps:

1. Generate a candidate $\mathbf{v}$.
   1.a. Generate a pre-candidate $\boldsymbol{\xi}$ by sampling from from the PDF $q(\cdot\,|\mathbf{u}_0)$
   1.b. Accept or reject $\boldsymbol{\xi}$

$$\mathbf{v} = \begin{cases} \boldsymbol{\xi}, & \text{with prob. } a(\mathbf{u}_0, \boldsymbol{\xi}) \\ \mathbf{u}_0, & \text{with prob. } 1-a(\mathbf{u}_0, \boldsymbol{\xi}) \end{cases} \tag{13}$$

   where

$$a(\mathbf{u}_0, \boldsymbol{\xi}) = \min\left\{1, \frac{\varphi_n(\boldsymbol{\xi})\, q(\mathbf{u}_0|\boldsymbol{\xi})}{\varphi_n(\mathbf{u}_0)\, q(\boldsymbol{\xi}|\mathbf{u}_0)}\right\} \tag{14}$$

2. If $\boldsymbol{\xi}$ was rejected set $\mathbf{u}_1 = \mathbf{u}_0$. Else, accept or reject $\mathbf{v}$

$$\mathbf{u}_1 = \begin{cases} \mathbf{v}, & \mathbf{v} \in F_{j-1} \\ \mathbf{u}_0, & \mathbf{v} \notin F_{j-1} \end{cases} \tag{15}$$

It can be shown that the transition PDF that results from the above procedure will satisfy Equation 9 independent of the choice of the proposal PDF $q(\cdot\,|\mathbf{u}_0)$ (Hastings 1970). If the proposal PDF has the symmetry property, i.e. $q(\mathbf{v}|\mathbf{u}) = q(\mathbf{u}|\mathbf{v})$, the algorithm reduces to the original Metropolis sampler (Metropolis et al. 1953).

As discussed in (Au & Beck 2001), the M-H algorithm becomes inefficient for high dimensional problems. This is due to the fact that the probability that the pre-candidate is rejected in step 1 increases rapidly with increasing number of dimensions n. This will lead to many repeated samples and hence to an increased correlation of the Markov chain. This is illustrated in Fig-

ure 1, where the acceptance rate of the pre-candidate state of the M-H algorithm is plotted against the random dimension n.

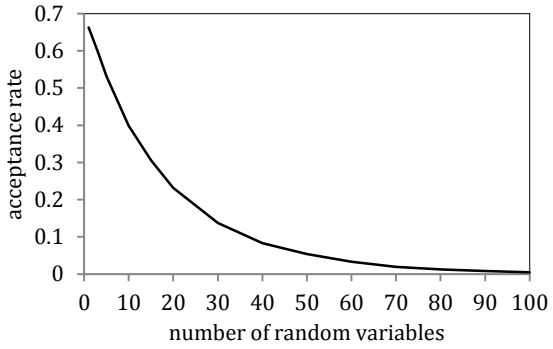

Figure 1: Acceptance rate of the original M-H algorithm applied to sampling from the independent standard normal distribution, as a function of the number of random variables n.

3.2 *Component-wise M-H*

The component-wise M-H algorithm was proposed by Au & Beck (2001) for sampling from high dimensional conditional distributions. The method requires that the random variable space be independent, however independence is achieved by the transformation of the original random variable space to the **U**-space. The method differs from the original M-H algorithm in the generation of the candidate state. That is, instead of using a n-dimensional proposal PDF, each coordinate v_i of the candidate state is generated from a one-dimensional proposal PDF $q(\cdot\,|u_{0i})$ that depends on the i-th coordinate u_{0i} of the current state. The algorithm is summarized as follows:

1. Generate a candidate $\mathbf{v} = (v_1, \cdots, v_n)$. For each $i = 1, \ldots, n$
   1.a. Generate a pre-candidate ξ_i from the PDF $q(\cdot\,|u_{0i})$
   1.b. Accept or reject ξ_i

$$v_i = \begin{cases} \xi_i, & \text{with prob. } a(u_{0i}, \xi_i) \\ u_{0i}, & \text{with prob. } 1 - a(u_{0i}, \xi_i) \end{cases} \tag{16}$$

   where

$$a(u_{0i}, \xi_i) = \min\left\{1, \frac{\varphi(\xi_i)\, q(u_{0i}|\xi_i)}{\varphi(u_{0i})\, q(\xi_i|u_{0i})}\right\} \tag{17}$$

2. Accept or reject **v**: apply Equation 15.

Due to the independence of the random vector **U**, the component-wise M-H algorithm satisfies the reversibility condition independent of the choice of the one-dimensional proposal PDF. Au & Beck (2001) suggest to choose $q(\cdot\,|u_{0i})$ as the uniform PDF centered at u_{0i} with width of 2, i.e. twice the standard deviation at the **U**-space – the optimal choice of the spread of the proposal PDF is discussed in Section 3.7. Moreover, due to the component-wise generation of the candidate state, the probability of repeated candidates decreases fast with increasing number of random variables n. Hence, the method is suitable for application to high-dimensional problems.

3.3 *M-H with repeated generation of pre-candidate states*

A different approach for reducing the correlation of the samples was proposed by Santoso et al. (2011). In this method, the candidate state is generated through a repeated generation of pre-candidate samples until acceptance of the pre-candidate is achieved. Hence, the algorithm avoids the generation of repeated candidates, thus reducing the chain correlation, as compared to the original M-H algorithm. The update of the Markov chain is as follows:

1. Generate a candidate $\mathbf{v}$.
   1.a. Generate a pre-candidate $\boldsymbol{\xi}$ from the PDF $q(\cdot\,|\mathbf{u}_0)$
   1.b. Accept or reject $\boldsymbol{\xi}$

$$\mathbf{v} = \begin{cases} \boldsymbol{\xi}, & \text{with prob. } a(\mathbf{u}_0, \boldsymbol{\xi}) \\ \mathbf{u}_0, & \text{with prob. } 1 - a(\mathbf{u}_0, \boldsymbol{\xi}) \end{cases} \tag{18}$$

   where

$$a(\mathbf{u}_0, \boldsymbol{\xi}) = \min\left\{1, \frac{\varphi_n(\boldsymbol{\xi})\, q(\mathbf{u}_0|\boldsymbol{\xi})}{\varphi_n(\mathbf{u}_0)\, q(\boldsymbol{\xi}|\mathbf{u}_0)}\right\} \tag{19}$$

   1.c. If $\boldsymbol{\xi}$ is rejected go to 1.a.
2. Accept or reject $\mathbf{v}$: apply Equation 15.

This approach is based on the M-H with delayed rejection (Tierney & Mira 1999), however in the latter method the acceptance probability of the pre-candidate sample is updated in each re-generation in order to ensure the satisfaction of the reversibility condition.

The method does not allow for an analytical expression of the transition PDF. Santoso et al. (2011) evaluated the transition PDF numerically for a one-dimensional truncated normal distribution using a uniform proposal PDF and showed that the reversibility condition is approximately satisfied. However, it turns out that the stationary distribution of the chain will differ from the target distribution. This is illustrated in Figure 2, where the CDF of the one-dimensional truncated normal distribution with different normalizing constants is compared to the empirical CDF from 10^4 samples.

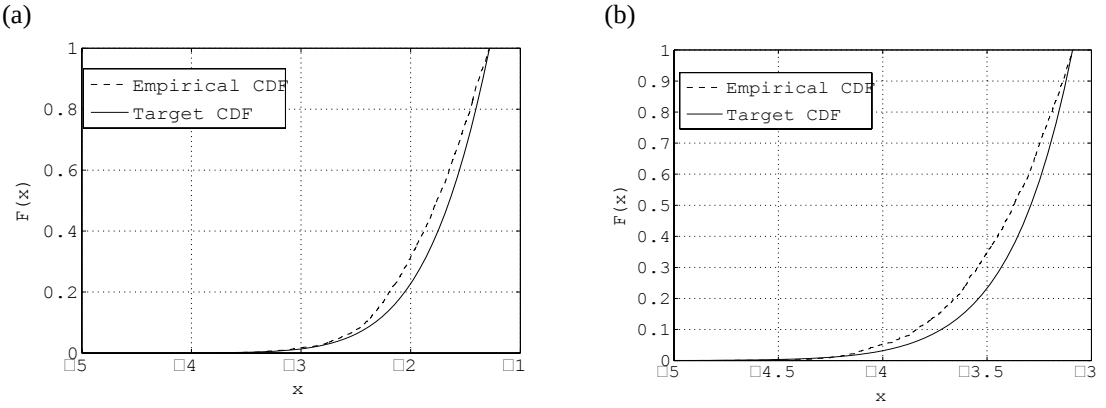

Figure 2: Empirical CDF of the M-H with repeated generation of pre-candidate states against target CDF of the Markov chain for the one-dimensional truncated normal distribution with probability normalizing constant (a) 10^{-1} and (b) 10^{-3}.

The fact that the target distribution is different from the stationary distribution of the chain may lead to biased probability estimates. Consider a reliability problem modeled by the following linear limit-state function at the **U**-space:

$$G(\mathbf{U}) = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i + \beta \tag{20}$$

The probability of failure for this function is $\Phi(-\beta)$, where $\Phi(\cdot)$ is the standard normal CDF. For a chosen conditional probability p_0, the failure probability at each subset level j is p_0^j. Hence, the threshold c_j at the corresponding subset level is $c_j = \beta + \Phi^{-1}(p_0^j)$. Figure 3 shows the relative bias for subset levels $j = 2$ and $j = 4$, i.e. the difference between the true probability P_f and the estimate $\hat{P}_f$ averaged over 500 independent subset simulation runs and divided by P_f, for $p_0 = 0.1$ and 500 samples per subset level. It is shown that the probability estimate is slightly biased and the bias tends to increase with increasing subset level, corresponding to a decreasing failure probability. On the other hand, the component-wise M-H gives a nearly unbiased estimate, since the produced samples follow the target distribution of the Markov chain.

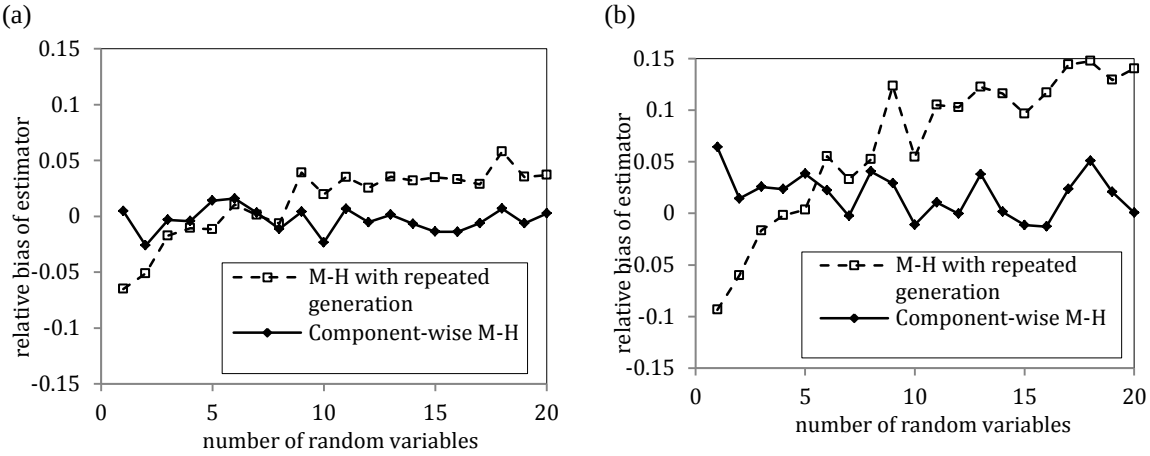

Figure 3: Relative bias of the estimator $\hat{P}_f$ from 500 independent simulation runs with the M-H with repeated generation of pre-candidate states and the component-wise M-H at subset level (a) $j = 2$ and (b) $j = 4$.

3.4 *Component-wise M-H with delayed rejection at the first acceptance level*

As discussed in Section 3.3, a repeated generation of pre-candidate states requires that the acceptance probability is adapted to account for the fact that the sample was rejected, in order to model the target distribution exactly. This procedure is called delayed rejection and was developed by Tierney & Mira (1999) for application to Bayesian statistics. Miao & Ghosn (2011) applied this approach in combination with the component-wise M-H algorithm leading to the following updating procedure:

1. Generate a candidate $\mathbf{v} = (v_1, \cdots, v_n)$. For each $i = 1, \ldots, n$
   1.a. Generate a pre-candidate ξ_{1i} from the PDF $q_1(\cdot\,|u_{0i})$
   1.b. Accept or reject ξ_{1i}

$$v_i = \begin{cases} \xi_{1i}, & \text{with prob. } a_1(u_{0i}, \xi_{1i}) \\ u_{0i}, & \text{with prob. } 1 - a_1(u_{0i}, \xi_{1i}) \end{cases} \tag{21}$$

   where

$$a_1(u_{0i}, \xi_{1i}) = \min\left\{1, \frac{\varphi(\xi_{1i})\, q_1(u_{0i}|\xi_{1i})}{\varphi(u_{0i})\, q_1(\xi_{1i}|u_{0i})}\right\} \tag{22}$$

   1.c. If ξ_{1i} was rejected, generate ξ_{2i} from $q_2(\cdot\,|u_{0i}, \xi_{1i})$
   1.d. Accept or reject ξ_{2i}

$$v_i = \begin{cases} \xi_{2i}, & \text{with prob. } a_2(u_{0i}, \xi_{1i}, \xi_{2i}) \\ u_{0i}, & \text{with prob. } 1 - a_2(u_{0i}, \xi_{1i}, \xi_{2i}) \end{cases} \tag{23}$$

   where

$$a_2(u_{0i}, \xi_{1i}, \xi_{2i}) = \min\left\{1, \frac{\varphi(\xi_{2i}) q_1(\xi_{1i}|\xi_{2i}) q_2(u_{0i}|\xi_{2i}, \xi_{1i})[1 - a_1(\xi_{2i}, \xi_{1i})]}{\varphi(u_{0i}) q_1(\xi_{1i}|u_{0i}) q_2(\xi_{2i}|u_{0i}, \xi_{1i})[1 - a_1(u_{0i}, \xi_{1i})]}\right\} \tag{24}$$

2. Accept or reject $\mathbf{v}$: apply Equation 15.

The algorithm allows for the second proposal PDF to depend not only on the current state of the chain but also on the rejected pre-candidate. It can be shown that the method satisfies the reversibility condition independent of the choices of the two proposal PDFs. The method will reduce the chain correlation, since fewer repeated pre-candidates will occur, however its benefit is limited to low-dimensional problems. For high-dimensional problems, the acceptance rate is high already for the first pre-candidate.

3.5 *Simulation of conditional samples in $\boldsymbol{U}$-Space*

In the first step of the M-H algorithm for subset simulation, one is sampling a candidate $\mathbf{v}$ from the joint Gaussian PDF $\varphi_n(\cdot)$, conditional on the previous sample $\mathbf{u}_0$. One is free to assume that $\mathbf{v}$ and $\mathbf{u}_0$ are jointly Gaussian with correlation coefficient ρ. Hence, the PDF of $\mathbf{v}$ will be the conditional normal distribution with mean value $\rho\mathbf{u}_0$ and covariance matrix $(1-\rho^2)\mathbf{I}$, where $\mathbf{I}$ is the unit diagonal matrix. It is possible to directly sample from this distribution, thus avoiding the generation of repeated candidates through rejection of pre-candidate states. This leads to the following updating scheme:

1. Generate a candidate $\mathbf{v} = (v_1, \cdots, v_n)$. For each $i = 1, \ldots, n$
   Generate v_i from the normal distribution with mean ρu_{0i} and standard deviation $\sqrt{1-\rho^2}$
2. Accept or reject $\mathbf{v}$: apply Equation 15.

It is trivial to see that the transition PDF between $\mathbf{u}_0$ and $\mathbf{u}_1$ satisfies the reversibility condition. Since we eventually sample from the conditional normal distribution $\varphi_n(\cdot\,|F_{j-1})$, a small correlation between the actual and the candidate state does not imply a small correlation of the final samples. This is due to the fact that a very small ρ will lead to many rejected samples. On the other hand, a large ρ will increase the acceptance rate but will lead to a larger correlation of the resulting samples. Section 3.7 comments on the optimal choice of ρ.

Besides its simplicity, the advantage of this approach lies with the fact that the candidate state is always accepted, without compromising the stationary distribution of the chain. Figure 4 compares the coefficient of variation of the probability estimate obtained by the algorithms described in Sections 3.2-3.5 for the limit-state function of Equation 20 in terms of the number of random variables n. For the component-wise M-H and the M-H with repeated generation of pre-candidates, the proposal PDF is chosen as the uniform PDF with width 2. The same PDF is chosen for both proposal PDFs for the component-wise M-H with delayed rejection. For the conditional sampling method, the correlation was chosen as 0.8. It is shown that the methods have similar performance for $n > 10$, while the conditional sampling algorithm performs better than the other methods in low dimensional problems. However, it should be noted that the algorithms with lower acceptance rate of the pre-candidate have the advantage that fewer limit-state function evaluations are required.

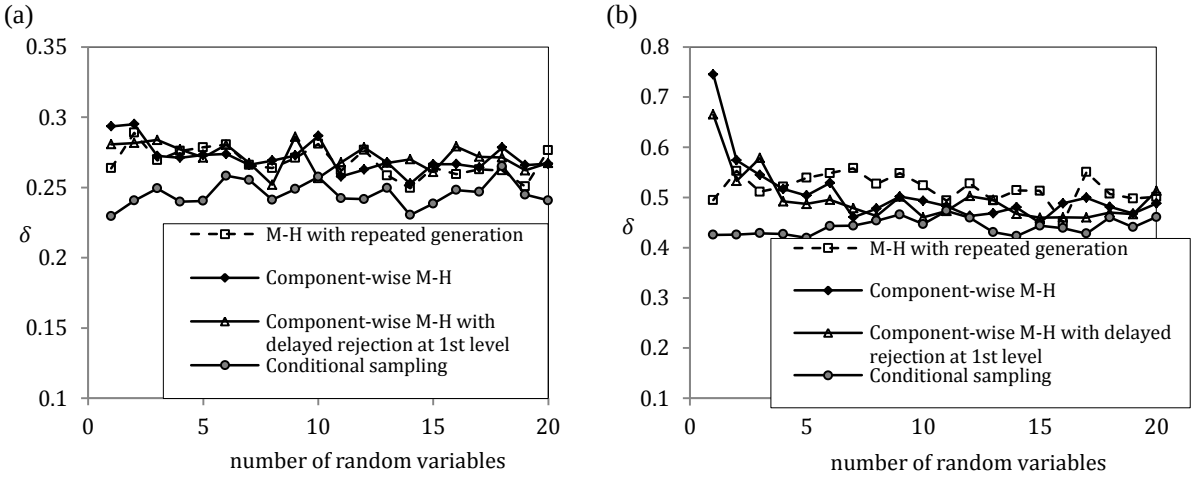

Figure 4: Coefficient of variation δ of the probability estimates evaluated from 500 independent subset simulation runs by the algorithms presented in Sections 3.2-3.5 at subset level (a) $j = 2$ and (b) $j = 4$.

3.6 *Component-wise M-H with delayed rejection at the second acceptance level*

The algorithms considered until now focus on the generation of the candidate state, i.e. they share the same second acceptance step. Therefore, they do not involve additional limit-state function evaluations. Zuev & Katafygiotis (2011) applied the delayed rejection concept at the second acceptance level, i.e. after the limit-state function has been evaluated to check whether the candidate state lies in F_{j-1}. If the candidate state is rejected, the accepted coordinates of the

pre-candidate state are re-sampled from a different one-dimensional proposal PDF and accepted or rejected with a suitably computed probability. The updating procedure reads:

1. Generate a candidate $\mathbf{v} = (v_1, \cdots, v_n)$. For each $i = 1, \ldots, n$
   1.a. Generate a pre-candidate ξ_{1i} from the PDF $q_1(\cdot\,|u_{0i})$
   1.b. Accept or reject ξ_{1i}

$$v_i = \begin{cases} \xi_{1i}, & \text{with prob. } a_1(u_{0i}, \xi_{1i}) \\ u_{0i}, & \text{with prob. } 1 - a_1(u_{0i}, \xi_{1i}) \end{cases} \tag{25}$$

   where

$$a_1(u_{0i}, \xi_{1i}) = \min\left\{1, \frac{\varphi(\xi_{1i})\, q_1(u_{0i}|\xi_{1i})}{\varphi(u_{0i})\, q_1(\xi_{1i}|u_{0i})}\right\} \tag{26}$$

2. Accept or reject $\mathbf{v}$: apply Equation 15.
3. If $\mathbf{v}$ was rejected. For each $i = 1, \ldots, n$, if ξ_{1i} was accepted
   3.a. Generate a new pre-candidate ξ_{2i} from the PDF $q_2(\cdot\,|u_{0i}, \xi_{1i})$
   3.b. Accept or reject ξ_{2i}

$$v_i = \begin{cases} \xi_{2i}, & \text{with prob. } a_2(u_{0i}, \xi_{1i}, \xi_{2i}) \\ u_{0i}, & \text{with prob. } 1 - a_2(u_{0i}, \xi_{1i}, \xi_{2i}) \end{cases} \tag{27}$$

   where

$$a_2(u_{0i}, \xi_{1i}, \xi_{2i}) = \min\left\{1, \frac{\varphi(\xi_{2i})\, q_1(\xi_{1i}|\xi_{2i}) q_2(u_{0i}|\xi_{2i}, \xi_{1i}) a_1(\xi_{2i}, \xi_{1i})}{\varphi(u_{0i})\, q_1(\xi_{1i}|u_{0i}) q_2(\xi_{2i}|u_{0i}, \xi_{1i}) a_1(u_{0i}, \xi_{1i})}\right\} \tag{28}$$

4. Accept or reject $\mathbf{v}$: apply Equation 15.

Zuev & Katafygiotis (2011) showed that the algorithm satisfies the reversibility condition. The method reduces the chain correlation, since the acceptance probability of the candidate state increases. Moreover, its efficiency is independent of the random dimension. However, the method requires additional limit-state function evaluations, as compared to all the previous approaches.

In Figure 5, the performance of the algorithm for the limit-state function of Equation 20 with $n = 100$ is compared to the one of the component-wise M-H. The proposal PDF is chosen as the uniform PDF with width of 2 for both levels. The conditional probability is chosen as $p_0 = 0.1$, while the number of samples N at each level for the component-wise M-H is chosen such that the two algorithms result in the same limit-state function evaluations, starting with $N = 500$ for the algorithm with delayed rejection. The coefficient of variation of the probability estimates is evaluated from 500 independent simulation runs. It is shown that the gain in efficiency is rather small, which agrees with the findings in (Zuev & Katafygiotis 2011). However, the authors show that a larger gain in efficiency might be achieved by choosing a different proposal PDF in the second level.

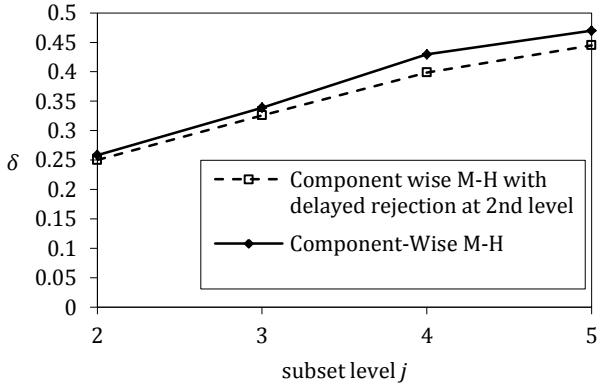

Figure 5: Coefficient of variation of the probability estimates by the component-wise M-H and the component wise M-H with delayed rejection at second acceptance level with the same LSF evaluations.

3.7 *Adaptive MCMC with optimal scaling*

It is shown above that the component-wise M-H and the conditional sampling method have similar performance in high dimensions. Also, it is discussed that the performance of the conditional sampling method depends on the choice of the correlation ρ between the actual and the candidate state. Similarly, the performance of the component-wise M-H depends on the variance of the proposal PDF. A large variance (resp. small ρ) will lead to many rejected candidates and a small variance (resp. large ρ) to a high correlation between states. Zuev et al. (2011) conjectured that an optimal variance of the proposal PDF at each subset level will give the minimum γ_j in Equation 11 and hence the minimum coefficient of variation of the conditional probability estimates. In their study, they varied the variance of the proposal PDF and evaluated γ_j and the second level acceptance rate of the algorithm. This led to the observation that γ_j is rather flat at the optimal acceptance rate, which lies in the interval $[0.3, 0.5]$.

Here, we perform a similar study by varying the variance of the proposal PDF of the component-wise M-H algorithm and the parameter ρ of the conditional sampling method. Moreover, we measure the performance of the methods in terms of the average velocity of the chains λ_j, defined in Equation 12. The curves obtained for the limit-state of Equation 20 at the subset levels $j = 2$ and $j = 4$ are shown in Figure 6. The results for the two algorithms agree with the ones in (Zuev et al. 2011). That is, the performance of the two algorithms can be optimized if the acceptance rate is kept between 0.3 and 0.5. This can be achieved by a scaling of the parameter of each algorithm after the simulation of each chain. If the acceptance rate of the chain is smaller than 0.3 then the variance of the proposal PDF is decreased (resp. the correlation parameter ρ is increased) and if it is larger than 0.5 the variance is increased (resp. ρ is decreased). Figure 7 shows the coefficient of variation of the probability estimates obtained from this adaptive procedure with the two considered algorithms for the limit-state function of Equation 20 with $n = 100$. The results are compared with the original version of the algorithms with a uniform proposal PDF with width 2 for the component-wise M-H algorithm and a correlation parameter $\rho = 0.9$ for the conditional sampling method. It is shown that the adaptive approaches give a smaller coefficient of variation at all subset levels.

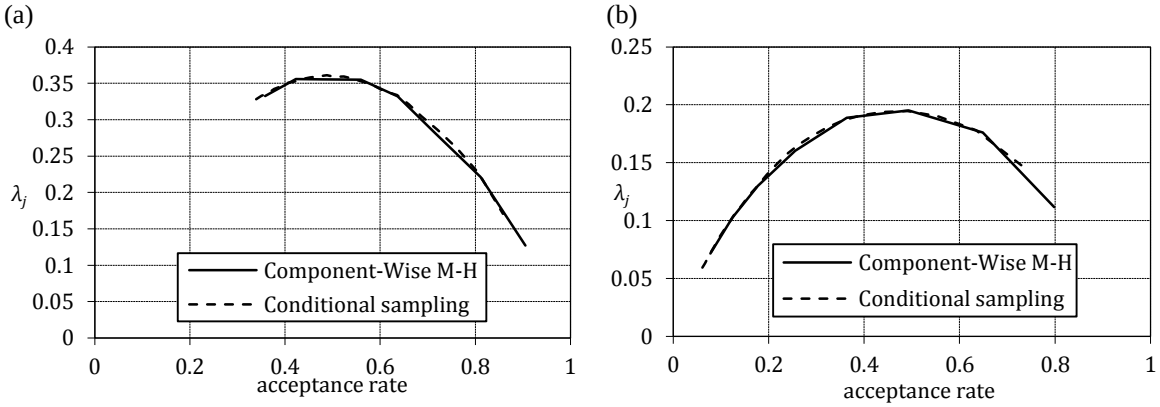

Figure 6: Chain velocity in terms of the second level acceptance rate for the component-wise M-H and the conditional sampling method at subset level (a) $j = 2$ and (b) $j = 4$.

4 CONCLUSION

This paper reviewed existing MCMC methods for subset simulation and proposed a new method that is based on sampling from a conditional normal distribution. The new approach is simpler and performs better than the other methods in low dimensional problems, since it accepts all candidate states of the Markov chain, without compromising the target distribution of the chain. In high-dimensional problems, the new method, together with all other algorithms that increase the first level acceptance rate, has a similar performance as the component-wise M-H algorithm, which was originally proposed for subset simulation. The component-wise M-H with delayed rejection at the second acceptance level provides better estimates at the expense of addi-

tional limit-state function evaluations. Finally, an adaptive procedure that adjusts the parameter of either the component-wise M-H or the conditional sampling method based on the chain acceptance rate provides better estimates without the need for further limit-state function evaluations.

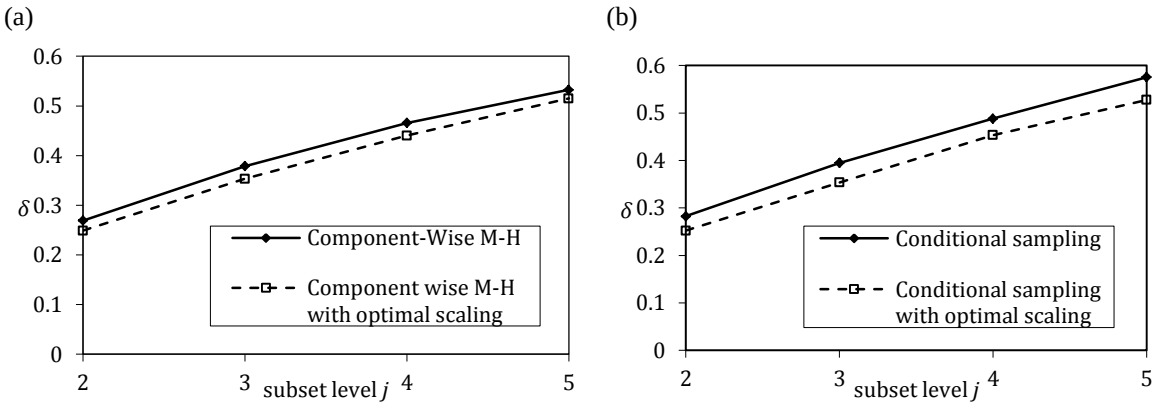

Figure 7: Coefficient of variation of the probability estimates by the (a) component-wise M-H and the (b) conditional sampling method and their optimal scaling variants.

REFERENCES

Au, S.K. & Beck, J.L. 2001. Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics*, 16(4): 263-277.

Bucher, C. 2009. Asymptotic sampling for high-dimensional reliability analysis. *Probabilistic Engineering Mechanics*, 24: 504-510.

Der Kiureghian, A. & Liu, P.-L. 1986. Structural reliability under incomplete probability information. *Journal of Engineering Mechanics*, 112(1): 85-104.

Hastings, W.K. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrica*, 57(1): 97-109.

Hohenbichler, M. & Rackwitz, R. 1981. Non-normal dependent vectors in structural safety. *Journal of Engineering Mechanics*, 107: 1227-1238.

Katafygiotis, L.S. & Cheung S.H. 2007. Application of the spherical subset simulation method and auxiliary domain method on a benchmark reliability study. *Structural Safety*, 29: 194-207.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. & Teller, E. 1953. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6): 1087-1092.

Miao, F. & Ghosn M. 2011. Modified subset simulation method for reliability analysis of structural systems. *Structural Safety*, 33: 251-260.

Rubinstein, R.Y. & Kroese, D.P. 2007. *Simulation and the Monte Carlo Method*, 2nd ed., Hoboken, John Wiley & Sons.

Santoso, A.M., Phoon, K.K. & Quek, S.T. 2011. Modified Metropolis-Hastings algorithm with reduced chain correlation for efficient subset simulation. *Probabilistic Engineering Mechanics*, 26: 331-341.

Tierney, L., & Mira, A. 1999. Some adaptive Monte Carlo methods for Bayesian inference. *Statistics in Medicine*, 18: 2507-2515.

Zuev, K.M., Beck, J.L. & Katafygiotis, L.S. 2011. On the optimal scaling of the modified Metropolis-Hastings algorithm. *Proceedings, 11th International Conference on Applications of Statistics and Probability in Civil Engineering, ICASP 11*, Faber, Köhler & Nishijima (eds), Zurich, Switzerland.

Zuev, K.M. & Katafygiotis, L.S. 2011. Modified Metropolis-Hastings algorithm with delayed rejection. *Probabilistic Engineering Mechanics*, 26: 405-412.