XLIII CILAMCE
21–25, November, 2022
FOZ DO IGUAÇU – PR – BRAZIL

Adaptive Importance Sampling for Reliability Analysis

André Jacomel Torii[1], Rafael Holdorf Lopez[2], Leandro Fleck Fadel Miguel[2], Henrique Machado Kroetz[3], Wellison José de Santana Gomes[2], André Teófilo Beck[4]

[1]*Latin-American Institute of Technology, Infrastructure and Territory, Federal University for Latin-American Integration, Avenida Tancredo Neves, 6731, 85867-970, Foz do Iguaçu/PR, Brasil, andre.torii@unila.edu.br*
[2]*Department of Civil Engineering, Federal University of Santa Catarina,Rua João Pio Duarte Silva, s/n, 88040-900, Florianópolis/SC, Brasil, rafael.holdorf@ufsc.br, leandro.miguel@ufsc.br, wellison.gomes@ufsc.br*
[3]*Center for Marine Studies, Federal University of Paraná, Av Beira Mar, s/n, 83255-976, Pontal do Paraná/PR, Brasil, henrique.kroetz@ufpr.br*
[4]*Department of Structural Engineering, University of São Paulo, Av. Trabalhador Sãocarlense, 400, 13566-590, São Carlos/SP, Brasil, atbeck@sc.usp.br*

Abstract. Several numerical schemes have been proposed in the last decades to address the problem of reliability analysis (i.e. evaluation of the probability of failure). Among these, a very popular method is Importance Sampling (IS), where the probability distribution employed for sampling is different from that of the random variables. It is known that by appropriate choice of the sampling distribution it is possible to reduce the variance of the estimate and thus obtain more accurate results. However, it is often difficult to known beforehand what is an appropriate sampling distribution for IS in practice. In order to avoid this difficulty, in the last years researchers developed Adaptive Importance Sampling (AIS) techniques. The idea of AIS is to take an initial sampling distribution, draw a sample, evaluate the required statistical moments and then improve the sampling distribution using some update rule. In this work we compare some of the existing AIS techniques in the context of Reliability Analysis.

Keywords: Reliability Analysis, Importance Sampling, Adaptive Importance Sampling

1 Introduction

Several methods for evaluation of the probability of failure have been proposed over the years [1]. Among these, sampling-based schemes, such as Monte Carlo Simulation and Importance Sampling (IS), are some of the most popular ones.

In this work we review some fundamental concepts regarding Adaptive Importance Sampling (AIS), describe two algorithms and employ them solve a benchmark Reliability Analysis problem. AIS is a family of sampling-based algorithms based on the original Population Monte Carlo (PMC) method by [2]. A detailed review on AIS is presented by [3].

It should be emphasized that AIS algorithms were conceived mainly in the field of Signal Processing [3]. This seems to be the reason why such techniques have not yet been applied for Reliability Analysis yet. Thus, the main goal of this work is to investigate if AIS can be applied successfully for Reliability Analysis problems.

2 Reliability Analysis and Monte Carlo Simulation

Consider a limit state function $g(X) : \mathbb{R}^m \to \mathbb{R}$, where X is a vector of random variables with density f_X and support $\Omega \subseteq \mathbb{R}^m$, such that $g < 0$ indicates failure of the system under analysis. In Reliability Analysis we are generally interested in evaluation of the probability of failure [1]

$$P_f = \mathbb{P}\left[g(X) < 0\right] = \int_\Omega I(g(x)) f_X(x) dx = \mathbb{E}\left[I(g(X))\right], \tag{1}$$

where $\mathbb{P}\left[\cdot\right]$ indicates the probability of occurrence of a given event, $\mathbb{E}[\cdot]$ represents the expected value and

$$I(t) = \begin{cases} 1, & t < 0 \\ 0, & t \geq 0 \end{cases}, \tag{2}$$

is the Indicator Function.

The probability of failure from Eq. (1) can be evaluated with several methods, such as transformation methods (e.g. FORM/SORM), sampling methods (e.g. Monte Carlo simulation), expansion methods (e.g. Stochastic Expansion, Polynomial Caos), surrogate methods (e.g. Kriging), among others. In this work we are interested in AIS, that are sampling methods based on Importance Sampling (IS).

Suppose we wish to evaluate

$$J = \int_{\Omega} h(x) f_X(x) dx = \mathbb{E}\left[h(X)\right], \tag{3}$$

i.e., the expected value of some function $h(X)$. In the case of Monte Carlo Simulation (MCS), the above expected value is estimated as [4]

$$\hat{J} = \frac{1}{N} \sum_{i=1}^{N} h(x_i), \tag{4}$$

where $\{x_1, x_2, ..., x_N\}$ is a sample of size N for the distribution f_X. In other words, the expected value J is approximated by a sample average $\hat{J}$. From the law of large numbers we know that $\hat{I} \to I$ as $N \to \infty$ under very weak conditions [4]. Thus, MCS is a robust method, that converges under very weak conditions as long as sample that is large enough is employed.

The mean squared error of MCS is given by

$$e = \mathbb{E}\left[(\hat{J} - J)^2\right], \tag{5}$$

that results

$$e = \mathbb{E}\left[\hat{J}^2\right] + J^2 - 2J\mathbb{E}\left[\hat{J}\right] = \mathbb{E}\left[\hat{J}^2\right] - \mathbb{E}\left[\hat{J}\right]^2 + \mathbb{E}\left[\hat{J}\right]^2 + J^2 - 2J\mathbb{E}\left[\hat{J}\right] = \mathbb{V}[\hat{J}] + (\mathbb{E}[\hat{J}] - J)^2, \tag{6}$$

where $\mathbb{V}\left[\cdot\right]$ represents the variance [5]. Since the MCS estimate is unbiased [4] we have $\mathbb{E}\left[\hat{J}\right] = J$ and thus

$$e = \mathbb{V}\left[\hat{J}\right]. \tag{7}$$

We thus conclude that the mean square error of the MCS estimate is actually its variance. Since the estimate $\hat{J}$ is a linear combination of $h(x_i)$ we also have [4]

$$e = \mathbb{V}\left[\hat{J}\right] = \frac{1}{N^2} \sum_{i=1}^{N} \mathbb{V}\left[h(x_i)\right] = \frac{1}{N} \mathbb{V}\left[h(X)\right]. \tag{8}$$

From the above result we conclude that the error of the MCS estimate can be reduce by increasing the sample size. However, the error also depends on the variance of $h(X)$. Thus, large samples will be required if $\mathbb{V}\left[h(X)\right]$ is large.

A more convenient form of estimating the accuracy of the MCS estimate is by means of the coefficient of variation of $\hat{J}$, given by

$$c = \frac{\sqrt{e}}{J} = \frac{\sqrt{\mathbb{V}\left[\hat{J}\right]}}{\mathbb{E}\left[\hat{J}\right]} = \frac{1}{\sqrt{N}} \frac{\sqrt{\mathbb{V}\left[h(X)\right]}}{\mathbb{E}\left[h(X)\right]}, \tag{9}$$

i.e. the relative error of MCS is basically proportional to the coefficient of variation of $h(X)$ and inversely proportional to $\sqrt{N}$. Thus, we can reduce the error of estimate by increasing the sample size N, but the relative accuracy obtained also depends on the quantity $\sqrt{\mathbb{V}\left[h(X)\right]}/\mathbb{E}\left[h(X)\right]$.

In the context of Reliability Analysis we have $h(X) = I(g(X))$ and the coefficient of variation of the P_f estimated using MCS results

$$c = \frac{1}{\sqrt{N}} \frac{\sqrt{P_f - P_f^2}}{P_f} \tag{10}$$

We observe that this relative error increases when $P_f \to 0$. For this reason, MCS is generally inefficient (i.e. requires very large samples) for Reliability Analysis when the probability of failure is small.

3 Importance Sampling

In Importance Sampling (IS) we rewrite the problem from Eq. (3) as

$$J = \int_\Omega h(x) \frac{f_X(x)}{q(x)} q(x) dx, \tag{11}$$

where $q(x) > 0, x \in \Omega$ is a sampling distribution. The above equation can then be written as

$$J = \int_\Omega h(x) w(x) q(x) dx = \mathbb{E}_q \left[h(X) w(X) \right], \tag{12}$$

$$w(x) = \frac{f_X(x)}{q(x)}, \tag{13}$$

where $\mathbb{E}_q$ represents the expected value with respect to the sampling distribution q.

From Eq. (12) we observe that the expected value is now evaluated with respect to the sampling distribution q instead of the distribution f_X of the random variables. The relation between q and f_X is taken into account by means of the weight function w. It can be demonstrated (see [4]) that by careful choice of the sampling distribution h it is possible to reduce the variance of the estimate, thus increasing its accuracy for a given sample size.

Consider then the IS estimate

$$\hat{J} = \frac{1}{N} \sum_{i=1}^{N} h(x_i) w(x_i), \tag{14}$$

where $\{x_1, x_2, ..., x_N\}$ is a sample of size N for the sampling distribution $q(x)$. It can be demonstrated that this estimate is unbiased (see [4]), i.e., $\mathbb{E}\left[\hat{J}\right] = J$. For this reason, following the same procedure employed for the MCS we conclude that the mean square error of the IS estimate results

$$e = \frac{1}{N} \mathbb{V} \left[h(X) w(X) \right]. \tag{15}$$

The optimal choice of the sampling density $q(x)$ can then be found from the problem: Find $q(x)$ that

$$\min \mathbb{V} \left[h(X) w(X) \right], \tag{16}$$

with $w(x) = f_X(x)/q(x)$. It can be demonstrated that this problem has solution [6]

$$q^*(x) = \frac{|h(x)| f_X(x)}{\int_\Omega |h(x)| f_X(x) dx}, \tag{17}$$

that is known as optimal importance sampling density.

In the case of Reliability Analysis we have $h(x) = I(g(x))$. Since $I(g(X)) = |I(g(X))|$ we conclude that

$$q^*(x) = \frac{I(g(x)) f_X(x)}{P_f}. \tag{18}$$

This demonstrates that $q^*(x)$ should be proportional to $I(g(x)) f_X(x)$. In other words, in the context of Reliability Analysis the optimal sampling density should only sample the failure region. This result puts in evidence that truly optimal IS requires explicit knowledge of the failure region defined by $I(g(x))$. Since this information is not known beforehand, several approaches have been proposed to find some efficient sampling distribution. In the context of Reliability Analysis, for example, a very common approach for defining $q(x)$ is to displace the original density $f_X(x)$ to a failure point, generally obtained with FORM.

4 Adaptive Importance Sampling (AIS)

Adaptive Importance Sampling (AIS) is a family of iterative methods conceived to build a sample for a target distribution $\pi(X)$ that is not explicitly known. AIS methods are based on the works by [7] and [2]. A detailed review on the subject is presented by [3]. An important characteristic of AIS methods is that they avoid Markov Chain Monte Carlo (MCMC) [3]. Note that several adaptive sampling schemes based on MCMC have already been proposed over the years, such as those described by [8] and [9]. However, as pointed out by [3], MCMC algorithms present delicate issues regarding computational implementation and convergence. Since AIS avoids the need for MCMC, the resulting algorithms are easier to implement and analyze.

4.1 Population Monte Carlo (PMC)

In order to understand AIS methods it is interesting to start with the Population Monte Carlo (PMC), presented by [2], that is considered the basis of AIS methods. In this case we wish to evaluate the integral

$$J = \int_{\Omega} h(x) \frac{\pi(x)}{Z} dx, \tag{19}$$

where $\pi(x)/Z$ is a probability distribution that is not known explicitly. Besides, the scalar $Z > 0$ is given by

$$Z = \int_{\Omega} \pi(x) dx, \tag{20}$$

i.e., Z is a normalization constant that ensures that $\pi(x)/Z$ is a probability distribution. Since $\pi(x)$ is not known explicitly, the normalization constant Z is unknown.

The PMC considers a set of n of proposal distributions $q_i^{(t)}(X)$ that are iteratively updated to approximate the target distribution $\pi(X)$, where the iterations of PMC are represented by indexes $t = 1, 2, ..., T$ and each proposal distribution is represented by indexes $i = 1, 2, ..., n$.

The proposal distributions $q_i^{(t)}(X)$ can be defined in several ways. Here we consider the most fundamental form encountered in the literature. Consider the vector with the mean and the standard deviation of the random vector $X = \{X_1, X_2, ..., X_m\}$, given by

$$\mu = \{\mathbb{E}[X_1], \mathbb{E}[X_2], ..., \mathbb{E}[X_m]\}, \quad \sigma = \left\{\sqrt{\mathbb{V}[X_1]}, \sqrt{\mathbb{V}[X_2]}, ..., \sqrt{\mathbb{V}[X_m]}\right\}. \tag{21}$$

Here we consider that each proposal distribution $q_i^{(t)}(x)$ is associated with a point $x_i^{(t)} \in \mathbb{R}^m$ as

$$q_i^{(t)}(X) = \phi\left(x, x_i^{(t)}, k^{(t)}\sigma\right) \tag{22}$$

where $\phi(x, \mu, \sigma)$ is the Normal Multivariate density with mean μ and standard deviation σ [5]. In other words, we assume that each proposal $q_i^{(t)}(x)$ associated to a point $x_i^{(t)}$ at iteration t is a Normal Multivariate distribution centered on $x_i^{(t)}$ with standard deviation $k^{(t)}\sigma$. The parameter $k^{(t)}$ is a scale factor for the standard deviation employed, that depends on the iteration t.

The basic PMC algorithm is as follows:

Step a) Take $t = 0$ and set an initial sample $S^{(0)} = \left\{x_1^{(0)}, x_2^{(0)}, ..., x_n^{(0)}\right\}$.

Step b) Update the iteration $t \leftarrow t + 1$.

Step c) For $i = 1, 2, ..., n$. Define the proposal $q_i^{(t)}(X)$. Sample a new point $x_i^{(t)}$ with $q_i^{(t)}(X)$. Include $x_i^{(t)}$ in the temporary sample $\overline{S}$ and evaluate its weight

$$w_i^{(t)} = \frac{\pi\left(x_i^{(t)}\right)}{q_i^{(t)}\left(x_i^{(t)}\right)}, \tag{23}$$

that is basically a IS weight with sampling distribution $q_i^{(t)}$.

Step d) Resample n points from $\overline{S}$ with replacement, with probability proportional to $w_i^{(t)}$, to create a new sample $S^{(t)} = \left\{x_1^{(t)}, x_2^{(t)}, ..., x_n^{(t)}\right\}$.

Step e) Return to Step b) until the total number of iterations is performed.

The resampling algorithm builds a new sample $S^{(t)}$ using the temporary sample $\overline{S}$ such that points with larger weights will likely be included several times in the new sample. At each step, a point is drawn from $\overline{S}$ and included in the new sample $S^{(t)}$ with probability proportional to $w_i^{(t)}$. A basic resampling algorithm is as follows:

Step a) Take a temporary sample $\overline{S} = \left\{x_1^{(t)}, x_2^{(t)}, ..., x_n^{(t)}\right\}$ with weights $w_i^{(t)}$.

Step b) Evaluate the cumulative sums $s_i = \sum_{j=1}^{i} w_j^{(t)} / \sum_{j=1}^{n} w_j^{(t)}$, $(i = 1, 2, ..., n)$, $s_0 = 0$

Step c) For $i = 1, 2, ..., n$. Draw a random number u with uniform distribution on $[0, 1]$. Find k that satisfies $s_{k-1} \leq u \leq s_k$. Include $x_k^{(t)}$ in the sample $S^{(t)}$

A detailed discussion on the theoretical basis of PMC is presented by [2]. It can be demonstrated that the sequence of samples $S^{(1)}, S^{(2)}, ..., S^{(T)}$ approximate the target distribution $\pi(X)$. This occurs because at each iteration, a new sample is created based on the weights of the existing points.

At each iteration t, PMC estimates to J and Z can be evaluated as [2, 10]

$$\hat{J}^{(t)} = \frac{1}{nt} \frac{1}{\hat{Z}^{(t)}} \sum_{i=1}^{n} \sum_{j=1}^{t} w_i^{(j)} h\left(x_i^{(j)}\right), \tag{24}$$

$$\hat{Z}^{(t)} = \frac{1}{nt} \sum_{i=1}^{n} \sum_{j=1}^{t} w_i^{(j)}. \tag{25}$$

4.2 Deterministic Mixture PMC (DM-PMC)

As commented previously, several modifications and improvements to PMC have been proposed over the years, leadong to several AIS methods [3]. Here we will consider a modification proposed by [10], known as Deterministic Mixture PMC (DM-PMC). According to [3], the DM-PMC is more efficient than the original PMC and yet very simple from the computational point of view. The basic idea of DM-PMC is to evaluate the weights as

$$w_i^{(t)} = \frac{\pi\left(x_i^{(t)}\right)}{\sum_{j=1}^{n} q_j^{(t)}\left(x_i^{(t)}\right)}. \tag{26}$$

The estimate to J remains that of Eq. (24), but now considering the modified weights.

The modification from Eq. (23) to Eq. (26) may seem small, but is conceptually significant. The original PMC weights (Eq. (23)) are evaluated assuming we have n sampling distributions $q_i^{(t)}$. The DM-PMC weights, on the other hand (Eq. (26)), are evaluated assuming there is a single sampling distribution, given by $\overline{q}^{(t)}(x) = \sum_{j=1}^{n} q_j^{(t)}(x)$. This means that the DM-PMC considers there is a single sampling distribution $\overline{q}^{(t)}(x)$ at each iteration, given by the sum of all $q_j^{(t)}(x)$. This sampling distribution $\overline{q}^{(t)}(x)$ is the one that approximates the target $\pi(x)$. As discussed by [10], the DM-PMC is able to reduced the variance of the estimate and allows a better exploration of the random space, improving the computational efficiency of the algorithm.

4.3 AIS for Reliability Analysis

In the context of Reliability Analysis, the optimal sampling distribution is given by Eq. (18). Thus, we can take the target distribution

$$\pi(x) = I(g(x)) f_X(x) \tag{27}$$

with $h(x) = 1$. From Eq. (20) we observe that Z is given by

$$Z = \int_{\Omega} I(g(x)) f_X(x) = P_f. \tag{28}$$

Thus, an estimate to the probability of failure can be obtained from the estimate to Z. From Eq. (25) we get the PMC/DM-PMC estimate

$$\hat{P}_f^{(t)} = \frac{1}{nt} \sum_{i=1}^{n} \sum_{j=1}^{t} w_i^{(j)}, \tag{29}$$

that is the mean of all weights obtained until iteration (t).

5 Numerical Example

The example studied here was taken from [9] and considers two random variables X_1, X_2 with Standard Normal distribution. Here we consider the parameters: $T = 4$, $n = 400$, $k^{(0)} = 2$, $k^{(t)} = 1/2, t \geq 1$. This means that we carry an initial sampling step plus 4 iterations with 400 points. This leads to a total of 2,000 points evaluated. For the initial sampling step we take $k^{(0)} = 2$, i.e. the initial sample is built with a standard deviation twice that of the random variables. This allows for a better initial exploration of the random space. For iterations $t = 1, 2, ..., 4$ we then take $k^{(t)} = 1/2$, i.e. we sample with half the standard deviation of the random variables, in order to sample close to the interest regions.

The accuracy of the results was measured running each algorithm $N_R = 100$ times. For each algorithm, the following statistics were evaluated:

$$\overline{P}_f = \frac{1}{N_R} \sum_{i=1}^{N_R} \hat{P}_{fi}, \tag{30}$$

$$s^2 = \frac{1}{N_R - 1} \sum_{i=1}^{N_R} \left(\hat{P}_{fi} - \overline{P}_f \right)^2, \tag{31}$$

$$c = \frac{s}{\overline{P}_f}, \tag{32}$$

$$b = \overline{P}_f - P_f, \tag{33}$$

where $\hat{P}_{fi}$ are the estimates obtained each time the algorithm is run and P_f is the reference value. Note that $\overline{P}_f$ is the average, c is the coefficient of variation, s^2 is the variance and b is the bias of the estimates obtained. From Eq. (6) we also note that an estimate for the root mean square (RMS) error can be obtained as

$$e_{rms} = \sqrt{s^2 + b^2}. \tag{34}$$

The limit state function is taken as

$$g = 0.1(X_1 - X_2)^2 - \frac{1}{\sqrt{2}}(X_1 + X_2) + 2.5. \tag{35}$$

The reference value for the probability of failure is $P_f = 4.21 \times 10^{-3}$ [9]. The results obtained with PMC and DM-PMC are presented in Table 1. Results obtained with SubSet Simulation (SuS) and two versions of Sequential Importance Sampling (SIS-I and SIS-II[1]), taken from [9], are also presented.

Table 1. Results (Total sample size equal to 2,000)

Method	$\overline{P}_f$	b	c	e_{rms}
PMC	4.177E-03	-3.29E-05	0.082	3.46E-04
DM-PMC	4.182E-03	-2.80E-05	0.067	2.82E-04
SuS	4.26E-03*	+5.00E-05	0.130*	5.56E-04
SIS-I	4.13E-03*	-8.00E-05	0.070*	3.00E-04
SIS-II	4.20E-03*	-1.00E-05	0.060*	2.52E-04

* Results taken from [9].

A general accuracy measure is given by the RMS errors, that are also presented in Figure 1a). We observe that DM-PMC, SIS-I and SIS-II obtained very similar accuracy. Although the SIS-II is the most accurate one, the DM-PMC also obtained very competitive results in this example. We finally observe that the DM-PMC obtained better results than the PMC, as expected. Besides, the PMC performed better than SuS in this example.

[1]The approach represented here by SIS-II is the one obtained by taking $n_c = 0.1n_s$ in the work by [9], while SIS-I is the one obtained by taking $n_c = n_s$.

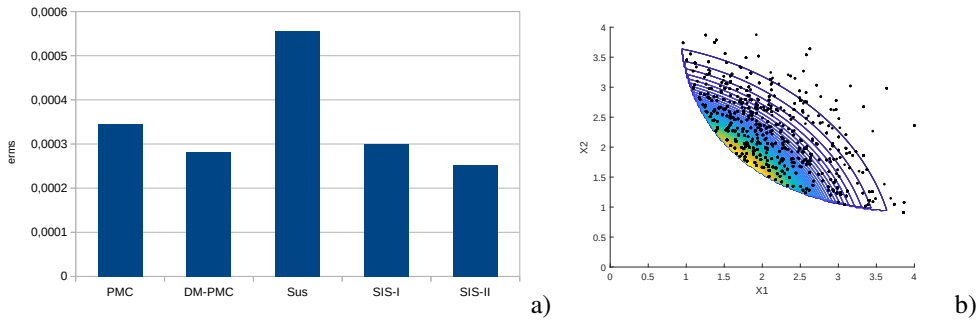

Figure 1. a) RMS error e_{rms} and b) Final sample obtained with a run of DM-PMC

The final sample obtained with a run of the DM-PMC is presented in Figure 1b). The contours represent the target sampling distribution from Eq. (27). We observe that the method indeed produces a sample that represents the target distribution very well. Note that the final sample has no points on the safety region (i.e. $g > 0$), since the target distribution from Eq. (27) only samples the failure region.

6 Conclusions

In this work we presented the application of Adaptive Importance Sampling (AIS) methods for Reliability Analysis. The DM-PMC was observed to be very competitive in comparison to SubSet Simulation and Sequential Importance Sampling. This is an important conclusion because DM-PMC is easier to implement, since it does not requires MCMC algorithms. These results indicate that further studies should be carried concerning AIS techniques for Reliability Analysis.

Acknowledgements. The authors would like to thank CNPq and FAPESP for financial support of this work.

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

References

[1] R. E. Melchers and A. T. Beck. *Structural Reliability Analysis and Prediction*, volume 1. John Wiley and Sons, 2018.
[2] O. Cappé, A. Guillin, J. M. Marin, and C. P. Robert. Population monte carlo. *Journal of Computational and Graphical Statistics*, vol. 13, n. 4, pp. 907–929, 2004.
[3] M. F. Bugallo, V. Elvira, L. Martino, D. Luengo, J. Miguez, and P. M. Djuric. Adaptive importance sampling: The past, the present, and the future. *IEEE Signal Processing Magazine*, vol. 34, n. 4, pp. 60–79, 2017.
[4] R. Y. Rubinstein and D. P. Kroese. *Simulation and the Monte Carlo method*. John Wiley & Sons, Hoboken, 2nd edition, 2008.
[5] M. Loève. *Probability Theory I*. Springer-Verlag, New York, 4th edition, 1977.
[6] R. Y. Rubinstein and B. Melamed. *Modern Simulation and Modeling*. John Wiley & Sons, New York, 1998.
[7] Y. Iba. Population monte carlo algorithms. *Transactions of the Japanese Society for Artificial Intelligence*, vol. 16, pp. 279–286, 2001.
[8] S.-K. Au and J. L. Beck. Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics*, vol. 16, n. 4, pp. 263–277, 2001.
[9] I. Papaioannou, C. Papadimitriou, and D. Straub. Sequential importance sampling for structural reliability analysis. *Structural Safety*, vol. 62, pp. 66–75, 2016.
[10] V. Elvira, L. Martino, D. Luengo, and M. F. Bugallo. Improving population monte carlo: Alternative weighting and resampling schemes. *Signal Processing*, vol. 131, pp. 77–91, 2017.