# Optimal Adaptive Importance Sampling for Reliability Analysis

André J. Torii [1], Lucas A. Rodrigues da Silva[2], André T. Beck[2]

[1]*Latin American Institute of Technology, Infrastructure and Territory, Federal University of Latin American Integration*
*Av. Tarquínio Joslin dos Santos, 1000, 85870-650, Foz do Iguaçu, PR, Brazil*
*andre.torii@unila.edu.br*
[2]*Department of Structural Engineering, University of São Paulo*
*Avenida Trabalhador São-carlense, 400, 13566-590 - São Carlos, SP, Brazil*
*araujolucasrs@usp.br, 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). Here we consider Adaptive Importance Sampling (AIS) schemes, that are based on Importance Sampling (IS). The idea of AIS is to iteratively improve the sampling distribution employed in IS. Since AIS methods are iterative, the scheme leads to a sequence of IS estimates. In standard AIS schemes the final estimate is taken as the average among all iterations. However, in this work we demonstrate that this approach is not the optimal choice. This occurs because some iterations will be more accurate than others, and thus a weighted mean is able to give better results. In this context, we demonstrate that optimal weights should be inversely proportional to the variance of the IS estimates of each iteration.

**Keywords:** Reliability Analysis, Importance Sampling, Adaptive Importance Sampling, Optimal Weights

## 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 a previous work we investigated the use of Adaptive Importance Sampling (AIS) in reliability analysis [2]. AIS is a family of sampling-based algorithms based on the original Population Monte Carlo (PMC) method by [3]. A detailed review on AIS is presented by [4]. See also [2] for a brief overview on how to employ AIS in the context of reliability analysis.

AIS schemes are iterative methods that provide a sequence of IS estimates. In standard AIS, the final estimate is taken as the average among the estimates of each iteration. However, in this work we demonstrate that this is not the optimal choice, since the estimates of each iteration do not have the same accuracy. We thus demonstrate that an optimal result can be obtained using a weighted mean among the iterations. The optimal weights depend basically on the variance and on the sample size of the estimate of each iteration.

Before presenting the main results of this work in Section 6 we present a brief review on the subject. A numerical example is studied in Section 7. The main conclusions of this work are summarized in Section 8.

## 2 Reliability Analysis

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]

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

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.

## 3 Monte Carlo Simulation

Suppose we wish to evaluate

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

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 [5]

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

where $\{x_1, x_2, ..., x_N\}$ is a sample of size $N$ for the distribution $f_X$. It can be demonstrate (see [2] or some other reference on MCS) that the mean squared error of MCS results

$$e = \frac{1}{N} \mathbb{V}\left[h(X)\right], \tag{5}$$

i.e., the error of the MCS estimate can be reduced by increasing the sample size, but also depends on the variance of $h(X)$.

In the context of Reliability Analysis it can be demonstrated that (see [2] or some other reference on reliability analysis) the relative error of the estimate results

$$c = \frac{1}{\sqrt{N}} \frac{\sqrt{P_f - P_f^2}}{P_f} \tag{6}$$

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.

## 4 Importance Sampling

In Importance Sampling (IS) we rewrite the problem from Eq. (3) as (see [2] or some other reference on importance sampling)

$$J = \mathbb{E}_q\left[h(X) w(X)\right] = \int_{\Omega} h(x) w(x) q(x) dx, \tag{7}$$

where $q(x) > 0, x \in \Omega$ is a sampling distribution, $\mathbb{E}_q$ represents the expected value with respect to the sampling distribution $q$ and

$$w(x) = \frac{f_X(x)}{q(x)}. \tag{8}$$

From Eq. (7) we observe that the expected value is now evaluated with respect to the sampling distribution $q$ instead of the distribution $f_X$. This can be used to get more accurate estimates.

Consider then the IS estimate

$$\hat{J} = \frac{1}{N} \sum_{i=1}^{N} h(x_i) w(x_i), \tag{9}$$

where $\{x_1, x_2, ..., x_N\}$ is a sample of size $N$ for the sampling distribution $q(x)$. It can be demonstrated that (see [5]) the optimal sampling distribution $q(x)$ satisfies

$$q^*(x) = \frac{|h(x)|f_X(x)}{\int_\Omega |h(x)|f_X(x)dx}, \tag{10}$$

that is known as optimal importance sampling density. In the case of Reliability Analysis we have $h(x) = I(g(x))$ and thus

$$q^*(x) = \frac{I(g(x))f_X(x)}{P_f}. \tag{11}$$

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. Since the failure region is not explicitly known beforehand, several approaches have been proposed to find some efficient sampling distribution. This is the idea of AIS.

## 5 Adaptive Importance Sampling (AIS)

Consider we have a target sampling function $q^*(x)$ (i.e. the optimal density) and a sequence of sampling functions

$$\mathcal{Q} = \{q_1(x), q_2(x), ..., q_n(x)\}. \tag{12}$$

See [2] for more details on how to use AIS to build such a sequence of sampling functions in the context of reliability analaysis.

We then define the Adaptive Importance Sampling (AIS) estimate

$$\boxed{\hat{J} = \sum_{k=1}^{n} \alpha_k \hat{J}_k}, \tag{13}$$

with

$$\sum_{k=1}^{n} \alpha_k = 1, \tag{14}$$

where $\hat{J}_k$ is an IS estimate with sampling function $q_k(x)$, i.e.

$$\hat{J}_k = \frac{1}{N_k}\sum_{i=1}^{N_k} h(x_i)w_k(x_i), \tag{15}$$

$$w_k(x) = \frac{f_X(x)}{q_k(x)}. \tag{16}$$

Note that the AIS estimate $\hat{J}$ is a weighted mean of the sequential IS estimates $\hat{J}_k$ with weights $\alpha_k$. However, standard AIS schemes consider that all estimates $\hat{J}_k$ have the same weight, and thus the final estimated is a simple average of the estimates $\hat{J}_k$ (see [4] for more details).

Also note that the variance of each iteration of AIS (i.e. Eq. (15)) results

$$\mathbb{V}\left[\hat{J}_k\right] = \mathbb{V}\left[\frac{1}{N_k}\sum_{i=1}^{N_k} h(x_i)w_k(x_i)\right] = \frac{1}{N_k^2}\sum_{i=1}^{N_k}\mathbb{V}\left[h(x_i)w_k(x_i)\right] = \frac{\mathbb{V}\left[h(X)w_k(X)\right]}{N_k}. \tag{17}$$

## 6 Optimal AIS

Here we wish to evaluate the optimal weights $\alpha_1$, $\alpha_2$, ..., $\alpha_n$, .i.e. the weights that make $\hat{J}$ as accurate as possible. The variance of the AIS estimate from Eq. (13) results

$$\mathbb{V}[\hat{J}] = \mathbb{V}\left[\sum_{k=1}^{n} \alpha_k \hat{J}_k\right] = \sum_{k=1}^{n} \alpha_k^2 \mathbb{V}[\hat{J}_k]. \tag{18}$$

Since

$$\mathbb{V}[\hat{J}_k] = \frac{1}{N_k}\mathbb{V}\left[h(X)w_k(X)\right]. \tag{19}$$

we have

$$\mathbb{V}[\hat{J}] = \sum_{k=1}^{n}\frac{\alpha_k^2}{N_k}\mathbb{V}[h(X)w_k(X)]. \tag{20}$$

Since IS sampling schemes are unbiased, the optimal weights are the ones that minimize the variance of the estimate $\hat{J}$. Thus, the optimal weights are given by

$$\alpha = \text{argmin} \sum_{k=1}^{n}\frac{\alpha_k^2}{N_k}\mathbb{V}[h(X)w_k(X)] \tag{21}$$

subject to $\sum_{k=1}^{n}\alpha_k^2 = 1$.

The Lagrangian function of this optimization problem results

$$L = \sum_{k=1}^{n}\frac{\alpha_k^2}{N_k}\mathbb{V}[h(X)w_k(X)] + \lambda\left(\sum_{k=1}^{n}\alpha_k - 1\right), \tag{22}$$

where $\lambda$ is the Lagrange multiplier of the equality constraint. The first order stationary condition $\partial L/\partial\alpha_i = 0$ then gives (after rearrangement)

$$\alpha_i = -\frac{\lambda}{2}\frac{N_i}{\mathbb{V}[h(X)w_i(X)]}. \tag{23}$$

Since the Lagrange multiplier is unknown we can take $k = -\lambda/2$ and thus

$$\alpha_i = k\frac{N_i}{\mathbb{V}[h(X)w_i(X)]}. \tag{24}$$

This puts in evidence that the optimal weights must be proportional to $N_i/\mathbb{V}[h(X)w_i(X)]$. Thus, the optimal weights can be found by

$$\boxed{\alpha_k = \frac{\overline{\alpha}_k}{\sum_{i=1}^{n}\overline{\alpha}_k}}, \tag{25}$$

with

$$\overline{\alpha}_k = \frac{N_k}{\mathbb{V}[h(X)w_k(X)]}. \tag{26}$$

This result puts in evidence that, in general, taking all weights the same (i.e. taking $\hat{J}$ as the average value between the sequential estimates $\hat{J}_k$) is not an optimal choice. Finally, from Eq. (17) we conclude that

$$\boxed{\overline{\alpha}_k = \frac{1}{\mathbb{V}\left[\hat{J}_k\right]} = \frac{N_k}{\mathbb{V}[h(X)w_k(X)]}}. \tag{27}$$

The above results puts in evidence that the optimal weights should be inversely proportional to the variance of the IS estimate of each iteration. This means that optimal weights depend on the sample size $N_k$ and the variance of argument being estimated of each iteration. This is the main conclusion of this work.

Since the variance $\mathbb{V}[h(X)w_i(X)]$ is unknown in practice, we can estimate it using a sample variance $s_k^2$. This gives

$$\overline{\alpha}_k = \frac{N_k}{s_k^2}. \tag{28}$$

Since a sample estimate to $\mathbb{V}[h(X)w_i(X)]$ is being proposed here, some kind of error analysis must be considered. In this work we perform a very simple error analysis, that should be improved in future works.

### 6.1 Error analysis and sample size

It is known that the sample variance $s^2$ has standard error

$$\hat{\sigma}_{s^2} = s^2\sqrt{2/(n-1)}, \tag{29}$$

where $n$ is the sample size employed to estimate it. Thus, the relative error of the estimate results

$$e = \frac{\sigma_{s^2}}{\sigma^2} = \frac{\hat{\sigma}_{s^2}}{s^2} = \frac{\sqrt{2}}{\sqrt{n-1}}. \tag{30}$$

Consequently, we can rearrange the above equation to get

$$n = 1 + \frac{2}{e^2}. \tag{31}$$

This expression allow us to estimate the necessary sample size $n$ in order to get relative error $e$. Assuming, for example, that we require a relative error $e = 0.05$ we get a sample size $n = 801$. For this reason, in this work we recommend taking the sample size of each iteration of AIS as to satisfy

$$N_k \gtrapprox 800, \tag{32}$$

to ensure robustness of the proposed approach. This is obviously a preliminary error analysis, that should be further investigated in future studies.

## 7 Numerical Example

Here we consider an example taken from [6], that has two random variables $X_1, X_2$ with Standard Normal distribution. The limit state function is given by

$$g = 5 - x_2 - 0.5(x_1 - 0.1)^2. \tag{33}$$

The reference value for the probability of failure is $P_f = 3.01 \times 10^{-3}$ [6]. In this work we employed only the DM-PMC (Deterministic Mixture PMC) algorithm described in [2]. The parameters of the algorithm were taken as: $n = 5$, $N_k = 800$, $k^{(1)} = 2$, $k^{(t)} = 1/2, t \geq 1$. These are basically the same parameters employed in [2], but the sample size of each iteration has been increased from $N_k = 400$ to $N_k = 800$ (the total sample size has bee increased from 2,000 to 4,000), in order to roughly comply with the condition from Eq. (32). The accuracy of the results was measured running each algorithm 10 times and estimating the average probability of failure $\overline{P}_f$, the bias $b$, the coefficient of variation $c$ and the root mean square error $e_{rms}$. See [2] for a detailed description on how to evaluate these quantities.

The results obtained with standard DM-PMC and with optimal weights are presented in Table 1. From these results we observe that standard DM-PMC simply does not work for this example, since the estimates are completely wrong. The DM-PMC with optimal weights, on the other hand, is able to give accurate results. Thus, these results demonstrate that the proposed weights are able to improve the accuracy.

Table 1. Results for Example 2 (Total sample size equal to 4,000)

| Method | $\overline{P}_f$ | $b$ | $c$ | $e_{rms}$ | $e_{rms}/P_f$ |
|---|---|---|---|---|---|
| Standard DM-PMC | 1.0150e+06 | 1.0150e+06 | 3.1623 | 3.2097e+06 | 1.0663e+09 |
| DM-PMC with optimal weights | 0.0030 | -1.5669e-05 | 0.1437 | 4.0855e-04 | 0.1357 |

From the results from Table 1 we also observe that the standard DM-PMC provided very bad results, indicating that the algorithm became unstable in some way. In order to investigate what really happened, we present the results of each iteration of each run in Tables 2 and 3.

From Table 2 we observe that some IS estimates gave very bad results, namely at runs 7 and 10 (the values are indicated in red and magenta). The estimate of the third iteration of run 10 is fact very bad, indicating that the DM-PMC algorithm failed for some reason at that point. Thus, if standard DM-PMC is employed and we take the average value of the iterations as the final estimate, we observe that unsuccessful iterations may influence the result too much. This is the reason why standard DM-PMC gave such bad results in this case.

Table 2. IS estimates of each iteration and each run of the algorithm

| Run | $\hat{J}_1$ | $\hat{J}_2$ | $\hat{J}_3$ | $\hat{J}_4$ | $\hat{J}_5$ |
|---|---|---|---|---|---|
| 1 | 2.7212e-03 | 3.5827e-03 | 3.3617e-03 | 2.8745e-03 | 2.9510e-03 |
| 2 | 3.2096e-03 | 2.7931e-03 | 3.2212e-03 | 3.3917e-03 | 3.0087e-03 |
| 3 | 3.1344e-03 | 3.4998e-03 | 3.4195e-03 | 3.4020e-03 | 3.2703e-03 |
| 4 | 3.7343e-03 | 3.7986e-03 | 3.0019e-03 | 3.2253e-03 | 3.0866e-03 |
| 5 | 3.4235e-03 | 3.1159e-03 | 3.1599e-03 | 3.2099e-03 | 3.2614e-03 |
| 6 | 2.6711e-03 | 2.8397e-03 | 3.4325e-03 | 3.1004e-03 | 3.3965e-03 |
| 7 | 2.5228e-03 | 4.5020e-02 | 4.6297e-03 | 3.1796e-03 | 3.0379e-03 |
| 8 | 2.7653e-03 | 2.9092e-03 | 3.2981e-03 | 3.0898e-03 | 3.1159e-03 |
| 9 | 3.9200e-03 | 3.1840e-03 | 3.2795e-03 | 3.0589e-03 | 2.8810e-03 |
| 10 | 3.1782e-03 | 4.4611e-01 | 5.0749e+07 | 1.5332e-03 | 1.7153e-03 |

Table 3. Optimal weights for each run of the algorithm

| Run | $\alpha_1$ | $\alpha_2$ | $\alpha_3$ | $\alpha_4$ | $\alpha_5$ |
|---|---|---|---|---|---|
| 1 | 0.0329 | 0.1137 | 0.1866 | 0.3390 | 0.3278 |
| 2 | 0.0286 | 0.3108 | 0.1948 | 0.1755 | 0.2903 |
| 3 | 0.0295 | 0.2585 | 0.2819 | 0.1566 | 0.2734 |
| 4 | 0.0211 | 0.0605 | 0.2641 | 0.3052 | 0.3491 |
| 5 | 0.0246 | 0.3230 | 0.1509 | 0.2326 | 0.2690 |
| 6 | 0.0313 | 0.3617 | 0.2645 | 0.2818 | 0.0607 |
| 7 | 0.0929 | 0.0000 | 0.0035 | 0.3182 | 0.5855 |
| 8 | 0.0265 | 0.3056 | 0.1726 | 0.2852 | 0.2100 |
| 9 | 0.0195 | 0.3163 | 0.0672 | 0.2335 | 0.3634 |
| 10 | 0.0677 | 0.0000 | 0.0000 | 0.0450 | 0.8873 |

The optimal weights estimated with the proposed approach are presented in Table 3. We now observe that the weights given to those very bad estimates indicated in Table 2 are close to zero. This is the reason why the proposed optimal weights gave accurate results in Table 1, the approach gave very small weights to failed IS iterations. The resulting algorithm became more stable and accurate.

## 8 Conclusions

In this work we proposed a modification for AIS schemes in the context of reliability analysis. In this modification we try to pursue optimal weights to give to the IS estimate of each iteration, instead of given the same weight to all iterations as occurs in standard AIS. We concluded that these optimal weights should be inversely proportional to the variance the IS estimate of each iteration. The numerical example studied shows that this modification leads to a more stable algorithm that provides more robust results. We highlight that this is a preliminary work on the subject and that some aspects should be further investigated in the future.

**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] A. J. Torii, R. H. Lopez, L. F. F. Miguel, H. M. Kroetz, de W. J. Santana Gomes, and A. T. Beck. Adaptive importance sampling for reliability analysis. In *Proceedings of the XLIII Ibero-Latin-American Congress on Computational Methods in Engineering*, Foz do Iguaçu. ABMEC, ABMEC, 2022.

[3] 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.

[4] 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.

[5] R. Y. Rubinstein and D. P. Kroese. *Simulation and the Monte Carlo method*. John Wiley & Sons, Hoboken, 2nd edition, 2008.

[6] I. Papaioannou, C. Papadimitriou, and D. Straub. Sequential importance sampling for structural reliability analysis. *Structural Safety*, vol. 62, pp. 66–75, 2016.