An Efficient method for estimating conditional failure probabilities

Domenico Altieri[1] and Edoardo Patelli.[1*]

[1] *Institute for Risk and Uncertainty, University of Liverpool, United Kingdom*
[*] Corresponding author: edoardo.patelli@liverpool.ac.uk

Abstract

Conditional reliability measures provide a more detailed description of the performance of a system, being representative of different initial configurations.
Commonly, since the failure region is characterized by a small probability of occurrence, advanced sampling techniques are required to reduce the computational effort of a simulation based approach. These techniques if on one hand decrease the number of samples needed to identify the failure domain, on the other hand do not generally allow a direct estimation of the conditional failure probability for different given inputs.
This study aims at providing an efficient and simple methodology to evaluate the conditional failure probability in the case of a static reliability analysis. In particular, under the assumption of probability density functions (PDFs) with a finite support, the failure region mapping process is carried out using surrogate PDFs associated with Sobol' sequences in order to reduce as much as possible the model evaluations. Finally, the integration of the failure region in the standard normal space employs probabilistic weights instead of a classic indicator function to account for the uncertainty associated with the failure region definition.
The approach is verified by comparing the results against those obtained from a Latin Hypercube Sampling. The performance of the proposed method is evaluated in terms of computational costs and accuracy.

Keywords: Conditional failure probability, Reliability analysis, Failure region

1 Introduction

Despite the constant evolution of modern computers, numerical models require significant computational efforts to achieve the desired level of accuracy, in particular, for detailed and complex systems efficient simulation techniques result to be often of paramount importance to reach the required solution in a feasible time.
In the reliability analysis field , many advanced approaches [1–4] have been proposed in the last decades, having as main purpose a reduction in the number of simulations required for the failure probability estimation. Unfortunately, the techniques adopted by the aforementioned methodologies [5, 6] do not allow a direct estimation of reliability measures conditioned to specific sub-domains of the probabilistic space. Conditional failure probability represents indeed an important sensitivity indicator, since partial derivatives $\partial P_f/\partial v_i$, that describe the effects of uncertain input variables v_i on the system reliability level, can be subsequently computed [7–9]. Moreover, an efficient approach is required especially when a reliability-based optimization is performed [7, 10, 11] and a full reliability analysis must be carried out for each sub-domain analyzed by the optimization algorithm.
The proposed approach aims at computing conditional failure probabilities starting from a failure region mapping based on a multinomial logistic regression [12]. The training of the classification algorithm employs quasi-random samples [13] that allow a further reduction in the model realizations required.
A simple case study that estimates the maximum displacement of a cantilever beam, with uncertain material and geometrical properties, subject to a random concentrated load is analyzed. Finally, a Latin Hypercube Sampling provides the reference results based on which a first comparison with the proposed method is carried out. The opensource Matlab toolbox OpenCossan [14, 15] is used for the probabilistic analysis.

2 Methodology

Given the failure event F and a general sub-domain B of the probabilistic space Ω, the conditional failure probability can be expressed following the Kolmogorov definition as:

$$P(F|B) = \frac{P(F \cap B)}{P(B)} \tag{1}$$

where $P(F \cap B)$ represents the joint probability of the events F and B , while the probability $P(B)$ of the event B can be computed through the integral of the joint probability function $\Phi(\mathbf{x})$ over the sub-domain B, with $\mathbf{x}$ the vector of coordinates in Ω. In general, given a failure event F, the associated probability of occurrence can be expressed as the integral of $\Phi(\mathbf{x})$ over F:

$$P(F) = P_f = \int_F \Phi(\mathbf{x})dx = \int_\Omega \Phi(\mathbf{x}) \cdot \text{Ĺ}I dx \tag{2}$$

where I represents an indicator function equal to 1 when the system fails and 0 otherwise.
Unfortunately, the multi-dimensional integration over the failure domain F may be often impossible to perform in a closed form [16], for this reason a simulation-based approach, that approximates the exact solution, is required.
The joint probability $P(F \cap B)$, according to Equations 1 and 2, can be viewed as the integral of $\Phi(\mathbf{x})$ in the standard normal space [17] over the sub-domain B and not Ω:

$$P(F \cap B) = \int_{F \cap B} \Phi(\mathbf{x})dx = \int_B \Phi(\mathbf{x}) \cdot \text{Ĺ}I dx \tag{3}$$

The mapping from the physical space to the standard normal space can be easily obtained by assuming that the cumulative distribution functions of the random variables remain the same after the transformation in standard normal distributions [18]. Equation 1 can be now rewritten as:

$$P(F|B) = \frac{P(F \cap B)}{P(B)} = \frac{\int_B \Phi(\mathbf{x}) \cdot \text{Ĺ}I dx}{\int_B \Phi(\mathbf{x})dx} \tag{4}$$

An approximated solution of $\int_B \Phi(\mathbf{x})dx$ can be easily reached without additional computational efforts by means of deterministic samples $\hat{\mathbf{x}} \in B$ in the standard normal space [17]. For instance, to compute the failure probability $P(F|V)$ conditioned to the variable V, a sequence of sub-domains $B_i = \{\mathbf{x}_i | V = \mathit{constant}\}$ is explored by the samples $\hat{\mathbf{x}}$. More complex shapes $s(\mathbf{x})$ of B may require a preliminary step to be defined, nonetheless $s(\mathbf{x})$ just represents a combination of $\mathbf{x}$ depending on the specific problem. An efficient solution of Equation 4 still needs an additional step to reduce the number of model realizations related to the indicator function I.

2.1 Failure region mapping

A multinomial logistic regression (MLR) [12] is adopted to predict the probability that an initial system configuration reaches the failure region or not. The chosen classification algorithm allows the definition of intermediate classes in case conditional failure probabilities with respect other final performance levels are required (multiple failure regions). The indicator function in Equation 2 is therefore replaced by the probabilistic score w_i associated with each prevision Y_i. In general, the probability that an observation i has K as predicted value is equal to:

$$w_i = P(Y_i = K) = \frac{e^{\beta_K \cdot \mathbf{x}_i}}{1 + \sum_{k=1}^{K} e^{\beta_k \cdot \mathbf{x}_i}} \tag{5}$$

where $\mathbf{x}_i$ is the vector of attributes describing each observation i, while β_K represents the vector of regression coefficients associated with the class K.
The initial dataset for the MLR training is obtained by means of Sobol' sequences from surrogate probability density functions in order to easily reach the failure region and getting at the same time a higher grade of coverage of the probabilistic space. Samples are generated from uniform distributions $z(\mathbf{x})$ with upper and lower boundaries (U, L) chosen according to the real distributions. In particular , given the original PDF $f(\mathbf{x})$ and the associated cumulative function $F(\mathbf{x})$, than $F(L)$ and $1 - F(U)$ result to be lower than 10^{-7}. An approximate solution of Equation 3 can now be obtained by sampling from the surrogate PDFs only in the sub-domain B and by evaluating the probabilistic score w_i associated with each realization:

$$P(F \cap B) = \int_B \Phi(\mathbf{x}) \cdot \text{Ĺ}w_i dx \tag{6}$$

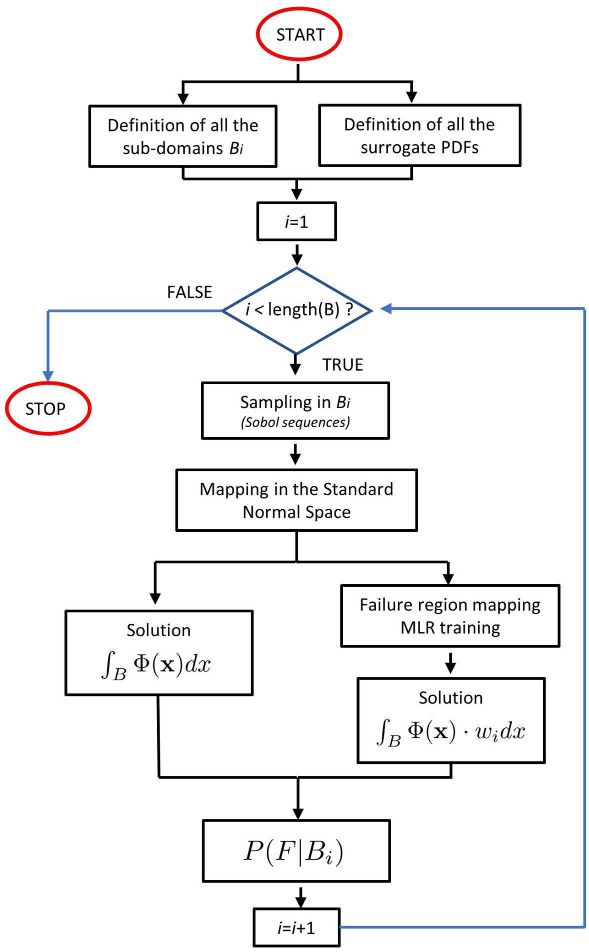

Figure 1: Flowchart representing the developed algorithm

Despite the MLR algorithm is a linear classifier based on a separating hyperplane, more complex and general failure region shapes can be still mapped by adopting alternative kernels [19]. In Figure 1 a flowchart describes step-by-step the algorithm employed to estimate the failure probability corresponding to different B_i.

3 Case study

A simple cantilever beam (Figure 2) subject to a random concentrated load, with uncertain material and geometric characteristics, has been analyzed as case study. The random variables are described in Table 1, with the associated uniform distributions employed for the failure region mapping.
The maximum displacement R can be easily computed as:

$$R = \frac{Q \cdot L^4}{8 \cdot E \cdot I} + \frac{P \cdot L^3}{3 \cdot E \cdot I} \tag{7}$$

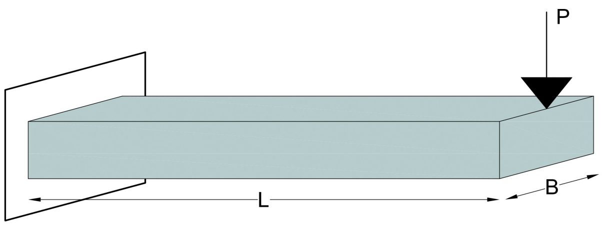

Figure 2: Cantilever beam

Table 1: Random variables (R.V.), parameters and surrogate PDFs employed in the failure region mapping

Quantity	Type	μ	σ	$z(\mathbf{x})$	L	U	Value
Load - P $[N]$	R.V.-Lognormal	5000	400	R.V.-Uniform	3900	6000	-
Elastic M.-E $[MPa]$	R.V.-Lognormal	10000	1600	R.V.-Uniform	1600	19000	-
Density - ρ $[Kg/m^3]$	R.V.-Lognormal	600	140	R.V.-Uniform	50	1350	-
Height - H $[m]$	R.V.-Normal	0.24	0.01	R.V.-Uniform	0.18	0.30	-
Length - L $[m]$	Parameter	-	-	-	-	-	1.8
Width - B $[m]$	Parameter	-	-	-	-	-	0.12

where $Q = f(\rho)$ represents the uniformly distributed beam's weight and I the moment of inertia.

4 Results

The system reliability has been analyzed assuming a maximum allowed displacement t equal to 8 mm and the conditional failure probability $P(R|E > t)$ has been evaluated for different given E values.
To asses the goodness of the result $g(\mathbf{x})$ provided by the proposed method, a reference fragility curve $fr(\mathbf{x})$, obtained through a Latin Hypercube Sampling, has been adopted. In particular, the comparison is carried out considering a different number of initial samples needed for the classification algorithm training.

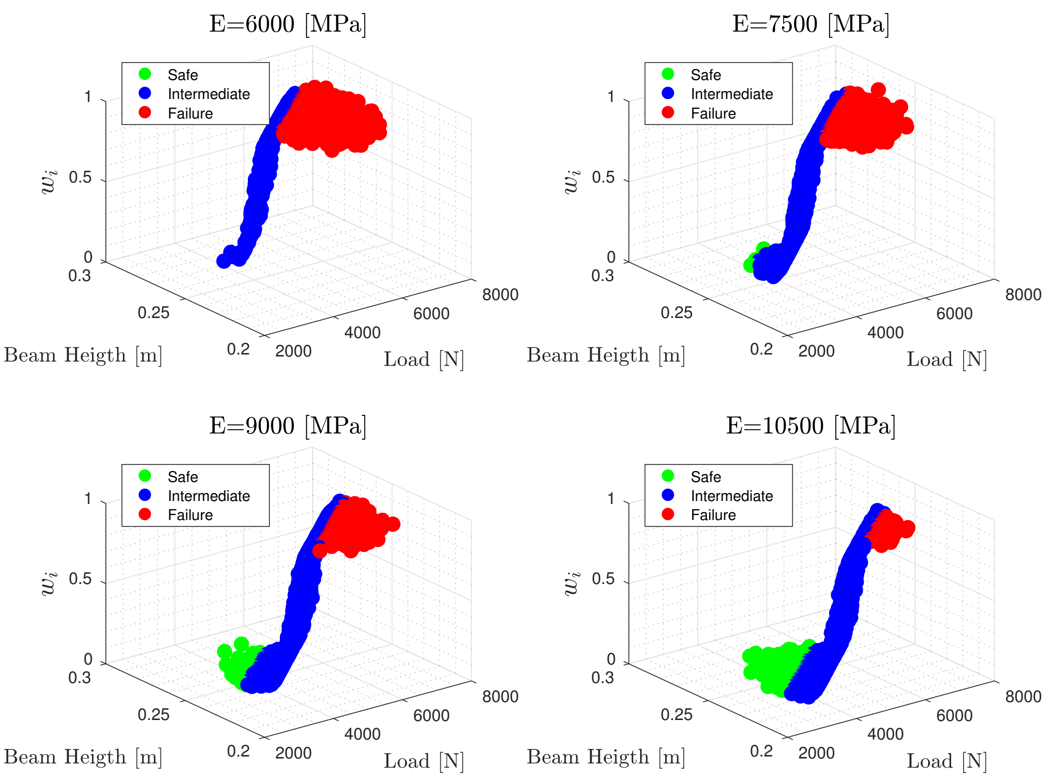

Figure 3: MLR probabilistic score associated with each prediction for different values of the elastic modulus

Figure 3 shows the probabilistic score of each sample involved in the solution of Equation 4 for different fixed values of the elastic modulus E. For illustrative purpose two only random variables have been considered.
Figure 4 reports the comparison with the reference curve obtained with a total of 7000 model realizations. The accuracy of the final match results to be dependent on the number of samples employed in the failure region mapping. In particular, by increasing the initial dataset the match improves reaching a good accuracy after only 250-350 samples.

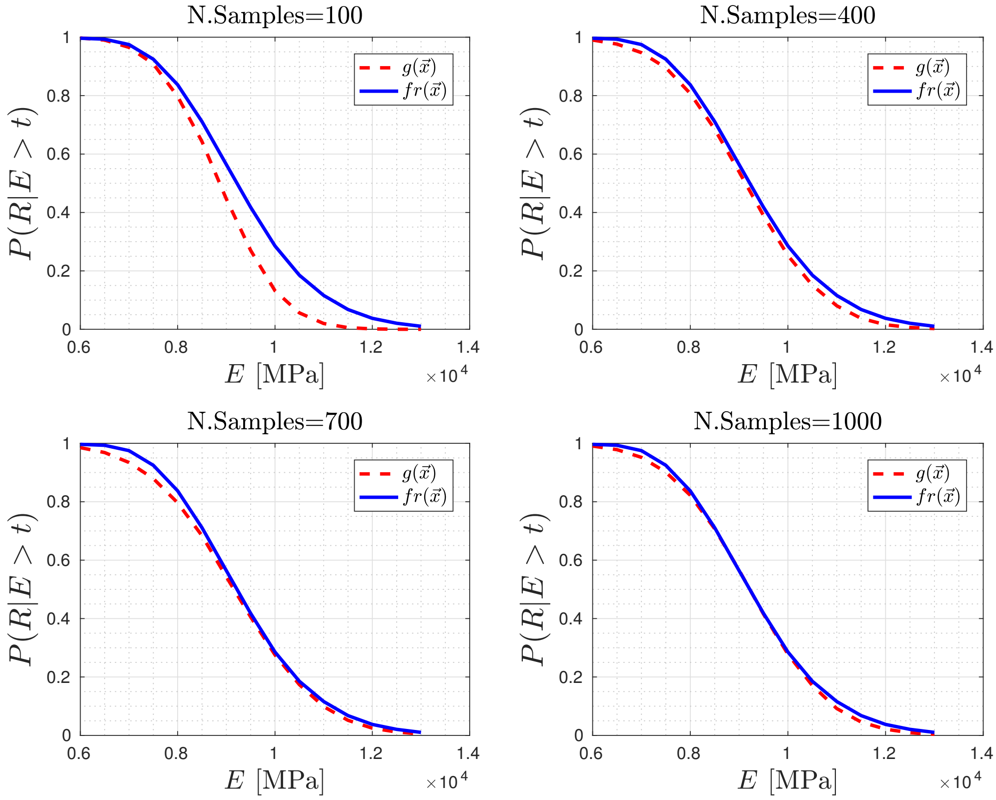

Figure 4: Comparison for different datasets between the result $g(\mathbf{x})$ provided by the method and the target solution $fr(\mathbf{x})$

Figure 5 describes the evolution of the mean absolute error D for different sizes of the dataset employed in the MLR training stage, where $D_N = \sum |g(\mathbf{x}_N) - fr(\mathbf{x}_N)| \ / \ L$, with N and L equal to the number of initial samples and elements in $g(\mathbf{x}_N)$ respectively. The variation of D for different initial datasets shows a quick decrease of the error after only 300 samples, even if the results seem to not fully converge to the target solution, even increasing the training dataset, due to the limits in the MLR mapping capability.

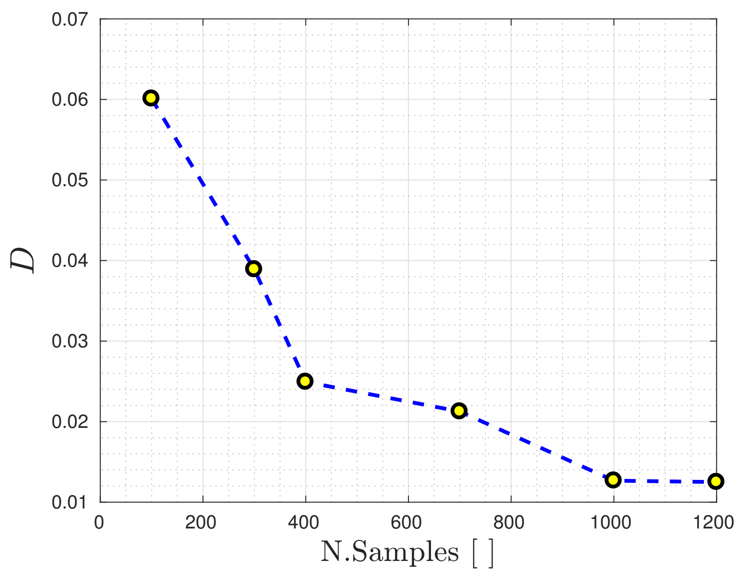

Figure 5: Evolution of the mean absolute error D for different training datasets

5 Conclusions

A simplified approach to estimate conditional failure probabilities has been proposed. The reduction in the computational expense is reached by adopting a multinomial logistic regression trained over a predefined dataset generated with surrogate PDFs. The use of a quasi-random sampling technique allows obtaining a further improvement in the failure region mapping efficiency, even if the goodness of the final comparison with the target solution depends on the dimension of the dataset used for the training stage. Thanks to the MLR algorithm, the proposed approach provides even the possibility to define multiple thresholds in order to evaluate after one single training stage the system conditional reliability with respect different performance levels. According to the analyzed case study, the results show a good match starting from only 250 model realizations and a further reduction in the confidence bounds is obtained by increasing the initial samples. Additional analyses are required to check how the dimensionality affects the efficiency of the method and additional applications needed to be tested to generalize the effectiveness of the failure region mapping.

References

[1] S.-K. Au and E. Patelli, Rare event simulation in finite-infinite dimensional space, *Reliability Engineering & System Safety* **148**, 67 (2016).

[2] M. de Angelis, E. Patelli, and M. Beer, Advanced line sampling for efficient robust reliability analysis, *Structural safety* **52**, 170 (2015).

[3] Z. Lu, S. Song, Z. Yue, and J. Wang, Reliability sensitivity method by line sampling, *Structural Safety* **30**, 517 (2008).

[4] C. Bucher, Asymptotic sampling for high-dimensional reliability analysis, *Probabilistic Engineering Mechanics* **24**, 504 (2009).

[5] W. K. Hastings, Monte carlo sampling methods using markov chains and their applications, *Biometrika* **57**, 97 (1970).

[6] H. R. Maier, B. J. Lence, B. A. Tolson, and R. O. Foschi, First-order reliability method for estimating reliability, vulnerability, and resilience, *Water Resources Research* **37**, 779 (2001).

[7] S. Au, Reliability-based design sensitivity by efficient simulation, *Computers & structures* **83**, 1048 (2005).

[8] Y.-T. Wu, Computational methods for efficient structural reliability and reliability sensitivity analysis, *AIAA journal* **32**, 1717 (1994).

[9] R. Melchers and M. Ahammed, A fast approximate method for parameter sensitivity estimation in monte carlo structural reliability, *Computers & Structures* **82**, 55 (2004).

[10] I. Enevoldsen and J. D. Sørensen, Reliability-based optimization in structural engineering, *Structural safety* **15**, 169 (1994).

[11] D. Altieri, E. Tubaldi, M. De Angelis, E. Patelli, and A. Dall'Asta, Reliability-based optimal design of nonlinear viscous dampers for the seismic protection of structural systems, *Bulletin of Earthquake Engineering* 1–20 (2018).

[12] D. Böhning, Multinomial logistic regression algorithm, *Annals of the Institute of Statistical Mathematics* **44**, 197 (1992).

[13] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola, *Global sensitivity analysis: the primer* (John Wiley & Sons, 2008).

[14] E. Patelli, R. Ghanem, D. Higdon, and H. Owhadi, Cossan: A multidisciplinary software suite for uncertainty quantification and risk management, *Handbook of Uncertainty Quantificationt* 1–69 (2016).

[15] E. Patelli, S. Tolo, H. George-Williams, J. Sadeghi, R. Rocchetta, M. De Angelis, and M. Broggi, Opencossan 2.0: an efficient computational toolbox for risk, reliability and resilience analysis, *Proceedings of the joint ICVRAM ISUMA UNCERTAINTIES Conference* (2018).

[16] M. Shinozuka, Basic analysis of structural safety, *Journal of Structural Engineering* **109**, 721 (1983).

[17] B. Huang and X. Du, Uncertainty analysis by dimension reduction integration and saddlepoint approximations, *Journal of Mechanical Design* **128**, 26 (2006).

[18] M. Rosenblatt, Remarks on a multivariate transformation, *The annals of mathematical statistics* **23**, 470 (1952).

[19] N. M. Nasrabadi, Pattern recognition and machine learning, *Journal of electronic imaging* **16**, 049901 (2007).