Proceedings of the ASME 2014 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference
IDETC/CIE 2014
August 17-20, 2014, Buffalo, USA

DETC2014-34051

A GENERALIZED "MAX-MIN" SAMPLE FOR RELIABILITY ASSESSMENT WITH DEPENDENT VARIABLES

Sylvain Lacaze
Computational Optimal Design of Engineering Systems
Aerospace and Mechanical Engineering Department
University of Arizona
Tucson, Arizona, 85721
Email: lacaze@email.arizona.edu

Samy Missoum
Computational Optimal Design of Engineering Systems
Aerospace and Mechanical Engineering Department
University of Arizona
Tucson, Arizona, 85721
Email: smissoum@email.arizona.edu

ABSTRACT

This paper introduces a novel approach for reliability assessment with dependent variables. In this work, the boundary of the failure domain, for a computational problem with expensive function evaluations, is approximated using a Support Vector Machine and an adaptive sampling scheme. The approximation is sequentially refined using a new adaptive sampling scheme referred to as generalized "max-min". This technique efficiently targets high probability density regions of the random space. This is achieved by modifying an adaptive sampling scheme originally tailored for deterministic spaces (Explicit Space Design Decomposition). In particular, the approach can handle any joint probability density function, even if the variables are dependent. In the latter case, the joint distribution might be obtained from copula. In addition, uncertainty on the probability of failure estimate are estimated using bootstrapping. A bootstrapped coefficient of variation of the probability of failure is used as an estimate of the true error to determine convergence. The proposed method is then applied to analytical examples and a beam bending reliability assessment using copulas.

1 Introduction

The calculation of probability of failure is often done using sampling techniques such as Monte-Carlo simulations, or using approximations such as moment-based methods (FORM, SORM) [1, 2]. It is now well known that the computational cost associated with computer simulations as well as the potential complexity of the failure domain (e.g., nonlinearity of the limit states) are major hurdles to an efficient calculation of probabilities of failure. For this reason, surrogate-based methods, whereby a limit state is approximated by surrogates such as Kriging [3] or Support Vector Machines (SVM) [4], are often used since these approximations are computationally efficient and enable the use of many Monte-Carlo samples. In order to tackle the difficulties stemming from nonlinear limit-states, adaptive sampling techniques have been developed for Kriging (e.g., Efficient Global Reliability Assessment (EGRA) [5], AK-MCS [6]) as for SVM (e.g., Explicit Design Space Decomposition(EDSD) [7], ^{2}SMART [8]). Loosely presented, these approaches aim at refining the approximation of the limit states by locating and evaluating samples in regions where information is needed.

If major strides have been achieved in the areas of surrogates and adaptive sampling for reliability assessment, existing approaches are still mostly limited in the important case of dependent variables. Indeed, it was shown in [9] that it is essential to tailor adaptive sampling schemes for reliability assessment when the variables are dependent. This stems from the fact that the approximation of the failure domain must be accurate in places where the probability densities are the highest. This can only be obtained if the adaptive sampling scheme includes information on the joint distribution of the variables.

We propose to extend the original EDSD sampling scheme to the case of dependent variables. The original scheme used a “max-min” sample whose purpose was to populate sparse regions while updating an SVM approximation of the failure domain [7, 10]. This was done by locating the sample on the approximated domain boundary. However, this sampling scheme did not include any information on the distributions of random variables. This work generalizes the “max-min” sample by including the joint probability density functions. This generalized “max-min” has proven to be beneficial in the case of independent variables of any distribution [11]. In this paper, we investigate its use in the case of correlated variables. It is shown that only the joint probability density function needs to be known. An examples is shown for an analytical joint probabilistic density function and an example shows the derivations for use of copulas [12].

This paper is structured as follows. Section 2 presents the generalization of the “max-min” sample for independent variables along with its key features. Section 3 further extends the formulation to dependent variables, either with known joint PDF or using copulas. Section 4 briefly describes an additional feature using bootstrapping to assess the error on the estimated probability of failure due to SVM approximation. Section 5 presents results for an analytical example and a simple mechanical example.

2 Generalized “max-min”

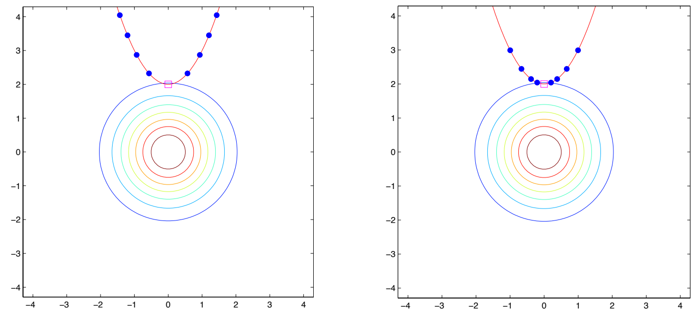
This section introduces the key elements of the proposed approach: a new distance-based adaptive sampling scheme able to account for the joint distribution of the random variables. For the sake of simplicity, it is first introduced for independent variables. It is subsequently extended to the case of dependent variables.

2.1 Classic “max-min”

The generalized “max-min” is an extension of a previous adaptive sampling scheme referred to as “Explicit Design Space Decomposition” (EDSD) [7,10]. In its most basic form, the sample is based on the identification of a point in the space that maximizes the distance to its closest neighbor while lying on the current approximation of the limit state:

$$\begin{aligned} \mathbf{x}_{mm} = \arg\max_{\mathbf{x}} \quad & \min_{j=1,\ldots,N_s} \left|\left|\mathbf{x}-\mathbf{x}^{(j)}\right|\right| \\ \text{s.t.} \quad & s^{(k)}(\mathbf{x}) = 0 \\ & l_i \le x_i \le u_i \\ & i = 1,\ldots,n \end{aligned} \tag{1}$$

where N_s is the number of existing samples, l_i and u_i are the lower and upper bounds for the i^{th} random variable. $s^{(k)}$ is the SVM approximation at iteration k (used in EDSD as approximation of the limit state). Figure 1(a) shows the distribution of such adaptive samples along a given limit state.



(a) Classic “max-min” (b) Generalized “max-min”

FIGURE 1. Distribution of classic and generalized “max-min” samples along a given limit state with independent normal random variables.

Because EDSD was developed for deterministic spaces, it does not take into account the probabilistic definition of the problem. More specifically, the distribution of samples shown on Figure 1(a) would remain the same regardless of the stochastic model. This leads to scalability problems and waste of function calls during the adaptive sampling process. This is due to the fact that a sample that maximizes the distance to its neighbors in low probabilistic content regions does not bring much information for estimating the probability of failure. In this work, we propose a generalization of this sample that takes advantage of the probabilistic definition of the problem.

2.2 Generalization to account for arbitrary joint PDF

In order to tackle the aforementioned hurdles, the “max-min” problem is reformulated as:

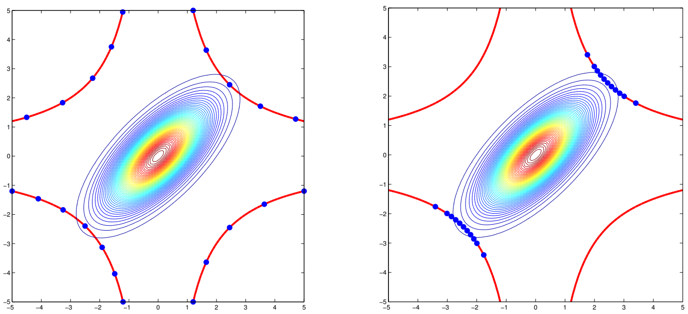
$$\begin{aligned} \mathbf{x}_{gmm} = \arg\max_{\mathbf{x}} \quad & \sqrt[n]{\mathbf{f}_{\mathbf{X}}(\mathbf{x})} \min_{j=1,\ldots,N_s} \left|\left|\mathbf{x}-\mathbf{x}^{(j)}\right|\right| \\ \text{s.t.} \quad & s(\mathbf{x}) = 0 \end{aligned} \tag{2}$$

This formulation weights the Euclidean distance with probabilistic information. Hence, the density of adaptive samples is higher in regions that are statistically important. Note that there are no more side constraints in this definition. This comes from the fact that the information about the domain is embedded within the joint PDF $\mathbf{f}_{\mathbf{X}}$. Figure 1(b) shows the distribution of such adaptive samples along a given limit state.

The key argument behind this formulation is that adaptive samples drawn from:

$$\max_{\mathbf{x}} \quad \sqrt[n]{\mathbf{f}_{\mathbf{X}}(\mathbf{x})} \min_{j=1,\ldots,N_s} \left|\left|\mathbf{x}-\mathbf{x}^{(j)}\right|\right| \tag{3}$$

will actually follow the distribution $\mathbf{f}_{\mathbf{X}}$. A proof of concept of the ability of the proposed formulation to actually follow the joint



(a) Classic "max-min" (b) Generalized "max-min"

FIGURE 3. Distribution of classic and generalized "max-min" samples along a given limit state with correlated normal random variables.

distribution $\mathbf{f_X}$ is proposed using numerical experiments of up to 30 dimensions in the case of a joint normal PDF. This was done both graphically (Figure 2) and numerically by means of a Kolmogorov-Smirnov test at the 5% significance level. This approach has shown promising results for reliability assessment using independent random variables [11].

3 The case of dependent variables

As introduced in the previous section, the generalized "max-min" works with any arbitrary joint PDF, without any assumptions of independence. Hence, as long as dependent joint PDFs can be provided, the proposed approach can be used.

3.1 Known joint PDF

If an analytical joint PDF is known for the given problem, the formulation of the generalized max-min can be used in a straightforward manner. An example of correlated normally distributed variables is provided in Figure 3.

Unfortunately, only few analytical dependent joint PDFs exist. In most real world scenarios, it is unlikely that one would be able to fit a joint distribution accurately. However, rather recently, copulas have become an attractive tool to model dependency and obtain an approximation of a joint distribution.

3.2 Use of copulas

Copulas [12] have received a large amount of attention in the past years in fields such as economics [13], biostatistics [14], hydrology [15], as well as engineering design [16]. The theoretical foundation comes from the Sklar's theorem [17]. It essentially states that for any joint CDF $\mathbf{F_X}$, there exists a unique copula C such that:

$$\mathbf{F_X}(x_1,\ldots,x_n) = C(\mathbf{F}_{X_1}(x_1),\ldots,\mathbf{F}_{X_n}(x_n)) \tag{4}$$

where $\mathbf{F}_{X_i}$ is the i^{th} marginal CDF. The joint PDF function is defined as:

$$\mathbf{f_X}(x_1,\ldots,x_n) = \frac{\mathrm{d}^n\mathbf{F_X}(x_1,\ldots,x_n)}{\mathrm{d}x_1\ldots\mathrm{d}x_n} \tag{5}$$

$$= \prod_{i=1}^{n}\mathbf{f}_{X_i}(x_i)\,\frac{\mathrm{d}^nC(v_1,\ldots,v_n)}{\mathrm{d}v_1\ldots\mathrm{d}v_n}\bigg|_{v_j=\mathbf{F}_{X_j}(x_j)} \tag{6}$$

Maybe the most widely used families of copulas are referred to as elliptical or Archimedean. As long as one is able to compute the copula, the PDF can be obtained numerically through successive finite differences. In the case of elliptical copulas, the derivation of the joint PDF can be carried out analytically. For example, consider a Gaussian copula defined as:

$$C(v_1,\ldots,v_n) = \Phi_R(\Phi^{-1}(v_1),\ldots,\Phi^{-1}(v_n)) \tag{7}$$

where R is a correlation matrix. By carrying out the derivation, we obtain:

$$\mathbf{f_X}(x_1,\ldots,x_n) = \prod_{i=1}^{n}\left[\frac{\mathbf{f}_{X_i}(x_i)}{\phi\left(\Phi^{-1}(v_i)\right)}\right]\phi_R\left(\Phi^{-1}(v_i)\right) \tag{8}$$

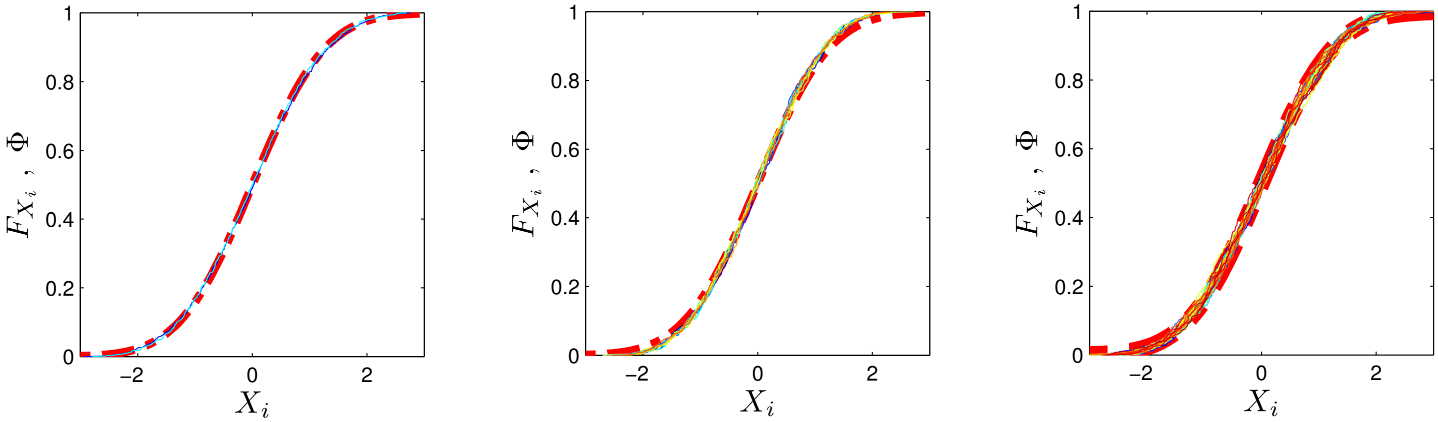
where $v_i = \mathbf{F}_{X_i}(x_i)$. This process is illustrated on Figure 4.

4 Estimation of the probability of failure

Recall that the proposed approach is an iterative process, for which at each iteration k, a new SVM is constructed and then refined by the addition of a generalized "max-min" sample. At iteration k, the probability of failure is estimated using Monte Carlo simulations such that:

$$P_f^{(k)} = \mathbb{E}\left[\mathbf{I}\left[s^{(k)}(\mathbf{x}) \le 0\right]\right] \tag{9}$$

10^6 Monte Carlo samples are used so that the error is minimal. However, the error on the estimated probability of failure due to the SVM approximation of the failure domain should not be ignored. In order to account for it, bootstrapping [18] is used. This is achieved by bootstrapping the training samples and obtaining a new SVM for each bootstrap. More specifically, at iteration k, let $\underline{\mathbf{x}}^{(k)} = [\mathbf{x}^{(1)},\ldots,\mathbf{x}^{(N_s)}]$ be the training samples set. A bootstrap training set $\underline{\mathbf{x}}^{(k,j)}$ is obtained, a new SVM $s^{(k,j)}(\mathbf{x})$ is constructed on this new training set and a new probability of failure $P_f^{(k,j)} = \mathbb{E}\left[\mathbf{I}\left[s^{(k,j)}(\mathbf{x}) \le 0\right]\right]$ is estimated. This process is repeated n_{bs} times (i.e., $j = [1,\ldots,n_{bs}]$). Based on the bootstrap sample of the probability of failure $P_f^{(k,\star)} = [P_f^{(k,1)},\ldots,P_f^{(k,n_{bs})}]$,



(a) 2D (b) 10D (c) 30D

FIGURE 2. Graphical proof of concept of the ability of the generalized formulation to follow a given joint distribution. Empirical marginal CDFs $\mathbf{F}_{X_i}$ compared to standard normal CDF Φ (red thick dotted line) for 2, 10, and 30 dimensions based on 400 generalized max-min samples.

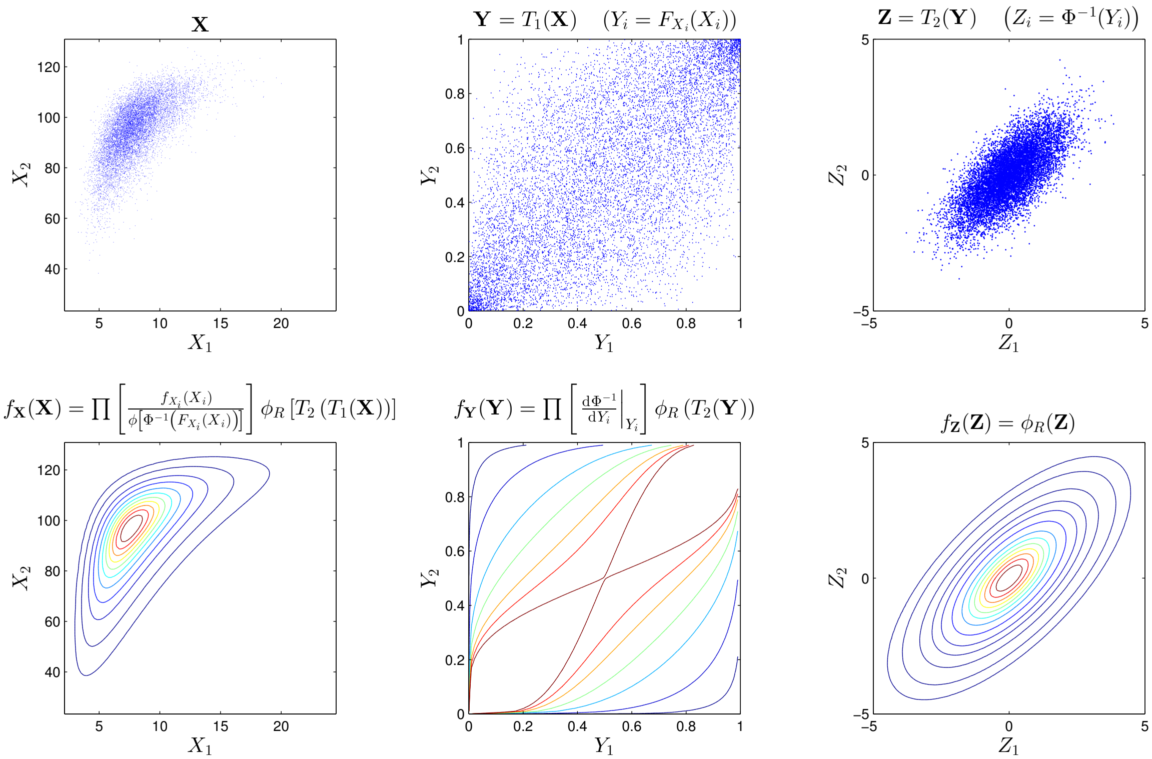


FIGURE 4. Graphical illustration of the derivation of the joint PDF for a Gaussian copula.

an empirical confidence interval can be derived. In addition, the median of the repetition $P_f^{(k,\star)}$ is used as the estimated probability for iteration k. Finally, the coefficient of variation $cv_{bs}^{(k)}$ of $P_f^{(k,\star)}$ is used as an estimate of the error due to SVM. In this work $n_{bs} = 200$ repetitions and a 95% confidence interval are used.

5 Results

The proposed approach is applied to an analytical example and a cantilever beam. For the sake of comparison, the results using the generalized "max-min" are compared to results using the classic "max-min" sample. In addition, to motivate the idea of

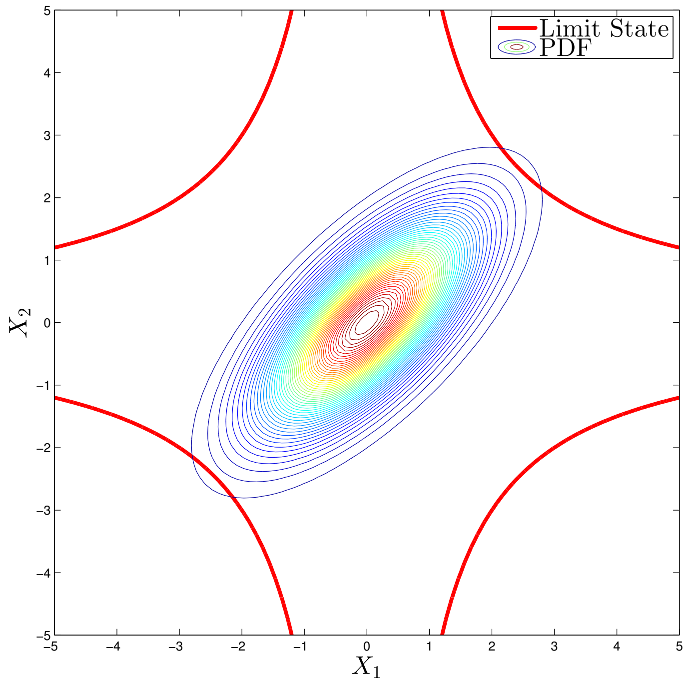


FIGURE 5. Probability density function and limit state for the demonstrative example.

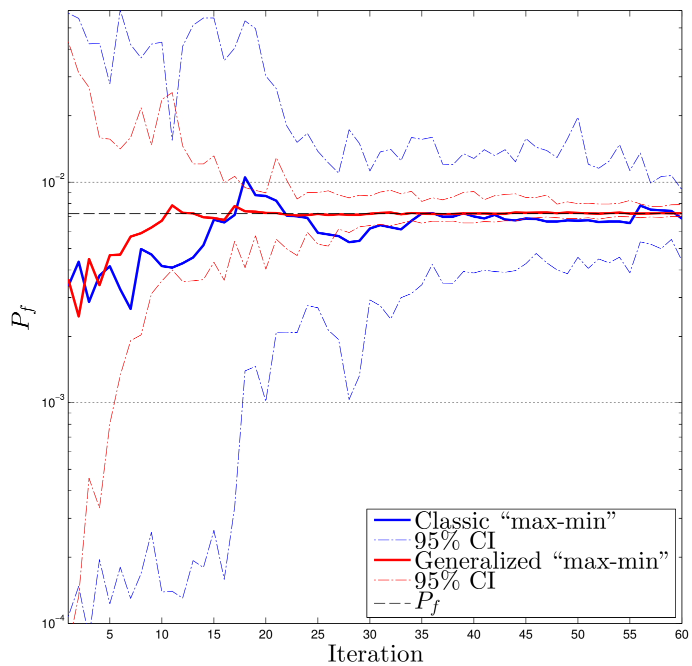


FIGURE 6. Convergence of the estimated probability of failure using two adaptive sampling schemes, classic or generalized "max-min" sample for the demonstrative example.

using the coefficient of variation $cv_{bs}^{(k)}$ of the bootstrapped $P_f^{(k,\star)}$ as an estimate of the error, it is compared to the true absolute relative error defined as:

$$\varepsilon^{(k)} = 100 \times \frac{\left|P_f - P_f^{(k)}\right|}{P_f} \tag{10}$$

5.1 Demonstrative example: Two dimensional correlated Gaussian variables

The first example used in this paper features a complex 2D limit state defined as:

$$\frac{6 \times \mathrm{sgn}(x_1)}{x_1} - x_2 \times \mathrm{sgn}(x_2) \leq 0 \tag{11}$$

where sgn is the sign or signum function. For this example, the random variables X_1 and X_2 follow a Gaussian correlated joint PDF with $\rho = 0.7$. Figure 5 provides an overview of the stochastic space along with the true limit state.

Figure 6 shows the convergence of the estimated probability of failure for the two different schemes. Figure 7 shows the evolution of the true error ε along with the bootstrapped coefficient of variation cv_{bs} when generalized "max-min" samples are used for the refinement scheme. In addition, Table 1 reports values of the estimated probability of failure at iterations 30, 60, and 100. As a reference, the confidence interval of the Monte Carlo simulation itself on the true limit state is provided. It can clearly be seen that using the generalized "max-min" drastically improves the convergence of the algorithm in addition to the coefficient of variation. When the generalized "max-min" sample is used, a coefficient of variation below 10% is first achieved at iteration 28 and below 5% at iteration 45. The actual errors made on the estimated probability of failure at these iterations are respectively 1.1% and 0.9%.

5.2 Cantilever Beam: Three dimensional example using copula

$$1.5 - \frac{PL^3}{3EI} \leq 0 \tag{12}$$

The marginal distributions of b, h and P are shown in Table 2. The copula for this problem is a Gaussian with $\rho_{12} = \rho_{13} = \rho_{23} = 0.7$. A scatter plot of the projected dependent joint PDF is provided on Figure 9.

Figure 10 shows the convergence of the estimated probability of failure for the two different schemes. Figure 11 shows the

TABLE 1. Estimated probability of failure (i.e., 0.5 quantile) and associated 95% CI at iteration 30, 60 and 100 for both scheme using either classic or generalized "max-min" for the demonstrative example.

Quantile	Classic			Generalized			Monte Carlo
	30	60	100	30	60	100	10^6 samples
0.025	0.0029	0.0044	0.0055	0.0064	0.0070	0.0070	$0.0070^{\star}$
0.5	**0.0062**	**0.0068**	**0.0071**	**0.0072**	**0.0072**	**0.0072**	$\mathbf{0.0072}^{\star}$
0.975	0.0112	0.0092	0.0088	0.0087	0.0079	0.0076	$0.0074^{\star}$
cv_{bs} (%)	31.90	16.83	11.52	8.49	3.38	1.83	$1.2^{\star\star}$
ε (%)	14.37	4.95	1.92	0.08	0.40	0.65	

* Obtained using 200 independent Monte Carlo simulation

** Obtained using (**??**)

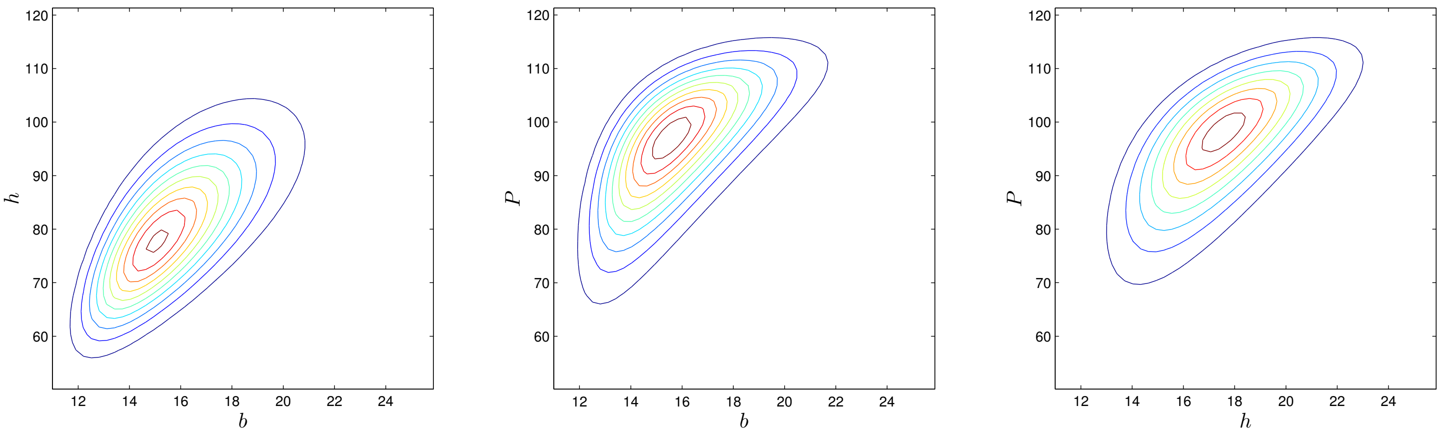


FIGURE 9. Scatter plot of the dependent joint distributions for the cantilever beam.

TABLE 2. Marginal distribution of the parameters for the cantilever beam.

Parameter	Distribution
b (mm)	$\ln\mathcal{N}(m = 2.0491, v = 0.2462)$
h (mm)	$\Gamma(\alpha = 53, \beta = 0.33)$
P (N)	$Weibull(a = 100, b = 10)$

evolution of the true error ε along with the bootstrapped coefficient of variation cv_{bs} when generalized "max-min" samples are used for the refinement scheme. In addition, Table 3 reports values of the estimated probability of failure at iterations 30, 60 and 100. For reference, the confidence interval of the Monte Carlo simulation itself on the true limit state is provided. Once again, a faster convergence is observed when the generalized "max-min" sample is used. When the generalized "max-min" sample is used, a coefficient of variation below 10% is first achieved at iteration 34 and below 5% at iteration 39. The actual errors made on the estimated probability of failure at these iterations are respectively 4.1% and 5.3%.

6 Conclusion

In this paper, a novel adaptive sampling scheme for reliability assessment with any PDF, including dependent ones, was introduced. Promising results have been shown on two test cases. In addition, a bootstrapping-based confidence interval was derived for the estimated probability of failure, in order to account for the error due to the SVM approximation.

The main focus of future work will be to derive a formal convergence criterion and investigate some possible tighter bounds for the confidence interval.

TABLE 3. Estimated probability of failure (i.e., 0.5 quantile) and associated 95% CI at iteration 30, 60 and 100 for both schemes using either classic or generalized “max-min” for the cantilever beam.

Quantile	Classic			Generalized			Monte Carlo
	30	60	100	30	60	100	10^6 samples
0.025	0.0001	0.0018	0.0027	0.0025	0.0029	0.0030	0.0030*
0.5	**0.0018**	**0.0036**	**0.0032**	**0.0029**	**0.0030**	**0.0031**	**0.0031**$\star$
0.975	0.0266	0.0086	0.0037	0.0045	0.0031	0.0031	0.0032*
cv_{bs} (%)	428.86	44.89	11.59	21.67	1.89	1.09	1.8**
ε (%)	40.52	18.75	5.84	2.31	0.63	2.06	

* Obtained using 200 independent Monte Carlo simulation

** Obtained using (**??**)

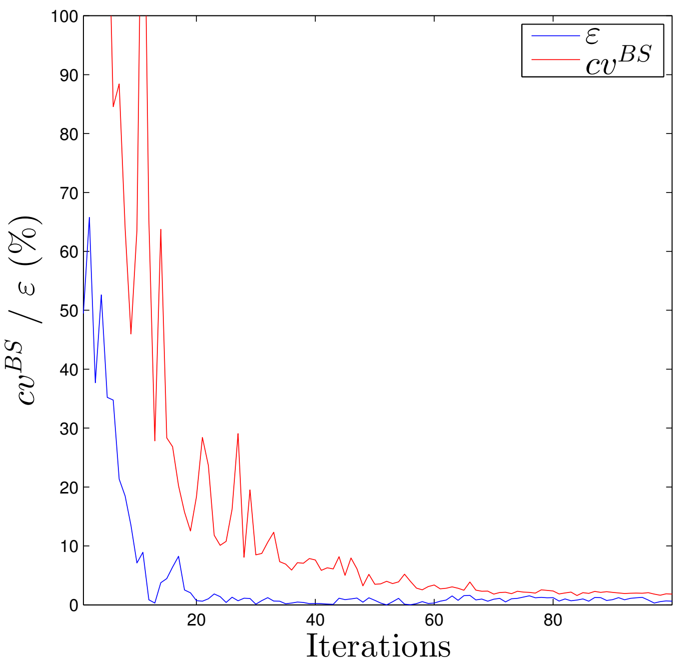


FIGURE 7. Evolution of the true error ε along with the bootstrapped coefficient of variation cv_{bs} for the generalized “max-min” scheme for the demonstrative example.

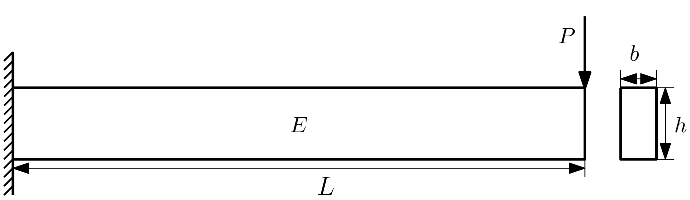


FIGURE 8. Description of the cantilever beam.

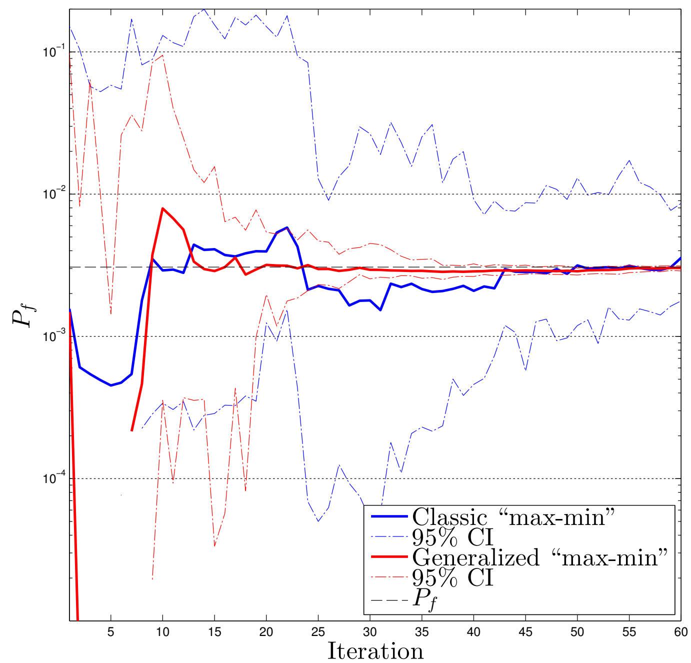


FIGURE 10. Convergence of the estimated probability of failure using two adaptive sampling schemes, classic or generalized “max-min” sample for the cantilever beam.

ACKNOWLEDGMENT

Support from the National Science Foundation (award CMMI-1029257) is gratefully acknowledged.

REFERENCES

[1] Melchers, R. E., and Melchers, R. E., 1999. “Structural reliability analysis and prediction”.

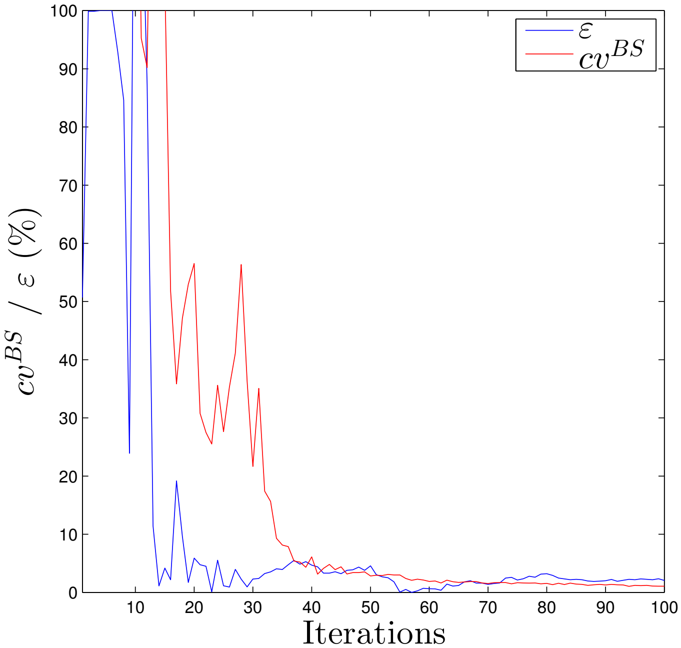


FIGURE 11. Evolution of the true error ε along with the bootstrapped coefficient of variation cv_{bs} for the generalized "max-min" scheme for the cantilever beam.

[2] Lemaire, M., 2010. *Structural reliability*, Vol. 84. John Wiley & Sons.

[3] Cressie, N., 1990. "The origins of kriging". *Mathematical Geology,* ***22***(3), pp. 239–252.

[4] Vapnik, V., 2000. *The nature of statistical learning theory*. Springer Verlag.

[5] Bichon, B., Eldred, M., Swiler, L., Mahadevan, S., and McFarland, J., 2008. "Efficient global reliability analysis for nonlinear implicit performance functions". *AIAA journal,* ***46***(10), pp. 2459–2468.

[6] Echard, B., Gayton, N., and Lemaire, M., 2011. "Ak-mcs: An active learning reliability method combining kriging and monte carlo simulation". *Structural Safety,* ***33***(2), pp. 145–154.

[7] Basudhar, A., and Missoum, S., 2010. "An improved adaptive sampling scheme for the construction of explicit boundaries". *Structural and Multidisciplinary Optimization,* ***42***(4), pp. 517–529.

[8] Bourinet, J.-M., Deheeger, F., and Lemaire, M., 2011. "Assessing small failure probabilities by combined subset simulation and support vector machines". *Structural Safety,* ***33***(6), pp. 343–353.

[9] Jiang, P., Basudhar, A., and Missoum, S., 2011. "Reliability assessment with correlated variables using support vector machines". In Proceedings of the 52nd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference.

[10] Basudhar, A., Missoum, S., and Harrison Sanchez, A., 2008. "Limit state function identification using support vector machines for discontinuous responses and disjoint failure domains". *Probabilistic Engineering Mechanics,* ***23***(1), pp. 1–11.

[11] Lacaze, S., and Missoum, S., 2013. "A generalized max-min sample for surrogate update". *Structural and Multidisciplinary Optimization*, pp. 1–5.

[12] Nelsen, R. B., 2006. *An introduction to copulas*. Springer.

[13] Frees, E. W., and Valdez, E. A., 1998. "Understanding relationships using copulas". *North American actuarial journal,* ***2***(1), pp. 1–25.

[14] Li, D., 2000. "On default correlation: a copula function approach". *Journal of Fixed income,* ***9***(4), pp. 43–54.

[15] Dupuis, D., 2007. "Using copulas in hydrology: Benefits, cautions, and issues". *Journal of Hydrologic Engineering,* ***12***(4), pp. 381–393.

[16] Noh, Y., Choi, K., and Du, L., 2009. "Reliability-based design optimization of problems with correlated input variables using a gaussian copula". *Structural and multidisciplinary optimization,* ***38***(1), pp. 1–16.

[17] Sklar, A., 1959. "Fonctions de répartition à n dimensions et leurs marges". *Publications of the Institute of Statistics, University of Paris,* ***8***, pp. 229–231.

[18] Efron, B., and Tibshirani, R., 1993. *An introduction to the bootstrap*, Vol. 57. CRC press.