13th International Conference on Applications of Statistics and Probability in Civil Engineering, ICASP13
Seoul, South Korea, May 26-30, 2019

Distribution with Independent Components for Uncertainty Quantification and Structural Reliability Analysis

Umberto Alibrandi
Associate Professor, Dept. of Engineering, Aarhus University, Aarhus, Denmark

Khalid M. Mosalam
Taisei Professor of Civil Engineering & PEER Director, Dept. of Civil & Env. Eng., Berkeley, CA, USA

ABSTRACT: This paper presents a novel method based on the Information Theory, Machine Learning and Independent Component analysis for Uncertainty Quantification and Structural Reliability Analysis. At first, it is shown that the optimal probabilistic model may be determined through minimum relative entropy and the theory of statistical learning; it is also discussed that methods based on the maximum entropy may perform well for the evaluation of the marginal distributions, including the tails. To determine the joint distribution of the basic random variables it is introduced the multivariate probabilistic model of Distributions with Independent Components (DIC). It has same computational simplicity of Nataf, but it is more accurate, since it does not pursue any assumption about the tail dependency. The proposed framework is applied to determine the joint distribution of wave height and period of wave data. Its extension for high dimensional reliability analysis of complex structural systems is straightforward.

1. INTRODUCTION

The failure probability P_f with respect to an assigned limit state is defined as [Madsen et al. (1986); Ditlevsen and Madsen (1996)]

$$P_f = \int_{G(\mathbf{x})\leq 0} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \tag{1}$$

where $\mathbf{x}$ is an n-vector collecting the basic random variables, $G(\mathbf{x})$ is the limit state function, $G(\mathbf{x}) = 0$ is the limit state surface separating the failure set $G(\mathbf{x}) \leq 0$ from the safe set $G(\mathbf{x}) > 0$ while $f_{\mathbf{X}}(\mathbf{x})$ is the joint probability density function of the basic random variables $X_1, X_2, \ldots, X_n$.

The failure probability can be evaluted by using approximate methods such as the First Order Reliability Method (FORM) or the Monte Carlo Simulation (MCS). The MCS is the most robust procedure, but in particular in its crude form requires excessive computational effort for the evaluation of very small failure probabilities. To reduce the computational cost an alternative strategy is given by the Response Surface Methodology, which builds a surrogate model of the target limit state function, defined in a simple and explicit mathematical form [Bucher and Burgound (1990); Dubourg et al. (2011); Alibrandi et al. (2015)]. Alternatively, data-driven methods, including statistical and machine learning approaches, are recently arising, to determine probability distributions of the performance functions from small samples of data [Alibrandi and Mosalam (2017, 2018)].

A critical step in eq.(1), sometimes underevaluated, is represented by the evaluation of the joint Probability Density Function (PDF) $f_{\mathbf{X}}(\mathbf{x})$ of the basic random variables. This challenging task is of great interest, because it has well recognized the high sensitivity of the reliability of the engineering systems to the distribution functions of the input [Elishakoff (1991)].

Unfortunately, usually there is lack of data

to characterize the joint distribution [Beer et al. (2013)]. Under incomplete information, it is assumed that only marginal distributions $f_{X_1}(x_1), f_{X_2}(x_2), ..., f_{X_n}(x_n)$, together with their correlations, are available. This assumption gives rise to an approximate joint PDF $\hat{f}_{\mathbf{X}}(\mathbf{x}) \approx f_{\mathbf{X}}(\mathbf{x})$. It is claimed that through eq.(1) it is possible to determine the failure probability for any limit state function $G(\mathbf{x})$. However if $G(\mathbf{x}) = [(X_1 \leq x_1) \cap (X_2 \leq x_2)]$, then $P_f = Prob[G(\mathbf{x}) \leq 0] = F_{X_1X_2}(x_1, x_2)$, $F_{\mathbf{X}}(\mathbf{x})$ being the joint Cumulative Distribution Function (CDF) of $\mathbf{X}$. But of course, if the approximate $\hat{f}_{\mathbf{X}}(\mathbf{x})$ is adopted, then eq.(1) will provide $\hat{P}_f \neq P_f$.

The Nataf model [Der Kiureghian and Liu (1986)] is widely adopted in reliability analysis under incomplete information. It is applicable to an arbitrary number of random variables and is proven to be quite effective for structural reliability analysis. However, the obtained distributions may have undesirable behavior if the marginal distributions are highly non-normal. Moreover, some codes, e.g. [DNV (2010)] recognize that Nataf should be used with caution for Uncertainty Quantification (UQ) due to the simplified modeling of the dependency between variables. In fact, it assumes implicitly a Gaussian dependence structure, which implies the tail independency between the random variables. This may negatively affect the joint distribution in its extreme values [Lebrun and Dutfoy (2009)].

Dutfoy and Lebrun (2009) show that a natural extension of the Nataf model is represented by the copula, which from a theoretical point of view is the exact concept of stochastic dependence. However, they note that from a practical point of view the identification of the copula of a multivariate distribution may be as challenging as the direct identification of the full joint distribution.

The joint PDF may be also well modeled in terms of a marginal distribution and a series of conditional density functions. However, especially in the presence of a high number of random variables, the definition of the conditional distributions is not straightforward.

The main objective of this paper is to reinterpret uncertainty quantification and structural reliability analysis through *information theory* and *machine learning*. At first, it is shown through information theory that methods based on the maximum entropy may perform well for the evaluation of the marginal distributions, including the tails. Then, we introduce the Distributions with Independent Components (DIC) for the direct evaluation of the joint distribution $f_{\mathbf{X}}(\mathbf{x})$. Finally, by simulating samples from $f_{\mathbf{X}}(\mathbf{x})$, the distribution probability $f_G(g)$ of the performance function G is evaluated through the Kernel Density Maximum Entropy (KDMEM) [Alibrandi and Mosalam (2017)]. The accuracy and performances of the proposed formulation are shown through the application to the joint distribution of wave data.

2. INFORMATION THEORY

2.1. Entropy

In information theory, the entropy of a random variable is a "measure of uncertainty" and it can be interpreted as the degree of information that the observation of the variable gives. Let us consider a discrete-valued random variable X with probability distribution given by p_i evaluated at x_i, $i = 1, 2, \ldots, N$. The Shannon's entropy functional [Jaynes (1957, 1978)] of the discrete distribution $\mathbf{p}$ is defined as follows:

$$H(\mathbf{p}) = -\sum_{i=1}^{N} p_i \, log p_i \tag{2}$$

while the (differential) entropy of a continuous-valued random variable X with PDF $f_X(x)$ is

$$H(X) = -\int f(x) log\, f(x)\, dx \tag{3}$$

The Kullback-Leibler (KL) divergence $D(X_1, X_2)$ (also called relative entropy) measures the entropy difference between the PDFs $f_1(x)$ and $f_2(x)$ of two random variables X_1 and X_2

$$\begin{aligned} D(X_1, X_2) &= \int f_1(x) log \left[\frac{f_1(x)}{f_2(x)}\right] dx = \\ &= H(X_1, X_2) - H(X_1) \end{aligned} \tag{4}$$

where $H(X_1, X_2)$ is the joint entropy of X_1 and X_2. The *relative entropy* may be considered a measure

of the 'distance' between two distributions, since it is always non-negative for all the possible distributions $f_1(x)$ and $f_2(x)$, while $D(X_1,X_2)=0$ only when $X_1 \equiv X_2$.

Let us consider a continuous-valued random variable X with PDF $f_X(x)$. The best (minimally prejudiced) probability distribution, subjected to the satisfaction of the available information, maximizes the entropy. It minimizes the relative entropy, see eq.(4), and it can be interpreted as the density that is compatible with the measurements and imposes the minimum number of assumptions on the data. Therefore, it "represents the most honest description of our state of knowledge" [Jaynes (1978)].

2.2. *Mutual Information*

The mutual information $I(X_1,X_2)$ of two random variables X_1 and X_2 measures the information that they share. Thus, it is a measure of their dependence and it is defined as

$$I(X_1,X_2) = \iint f(x_1,x_2) log\left[\frac{f(x_1,x_2)}{f_1(x_1)f_2(x_2)}\right] dx_1\,dx_2 \tag{5}$$

where $f(x_1,x_2)$ is the joint PDF of X_1 and X_2, while $f_1(x_1)$ and $f_2(x_2)$ are the corresponding marginals. It is seen that the mutual information is equivalent to the KL divergence between the joint density $f(x_1,x_2)$ and the product of its marginal densities. It is always non-negative and it is equal to zero only if the variables are statistically independent. Alternatively it may be expressed alternatively in terms of entropies

$$\begin{aligned} I(X_1,X_2) &= H(X_1) - H(X_1|X_2) = \\ &= H(X_2) - H(X_2|X_1) = \\ &= H(X_1) + H(X_2) - H(X_1,X_2) \end{aligned} \tag{6}$$

where $H(X_i|X_j)$ is the conditional entropy of given X_j, which quantifies the amount of information needed to describe the outcome of X_i given the value of X_j. Eq.(6) shows that the mutual information measures how much the knowledge of one variables reduces the uncertainty about the other. It is of interest to note, see Figure 1, the analogy between the union and of the intersection of two sets with joint entropy and mutual information, respectively.

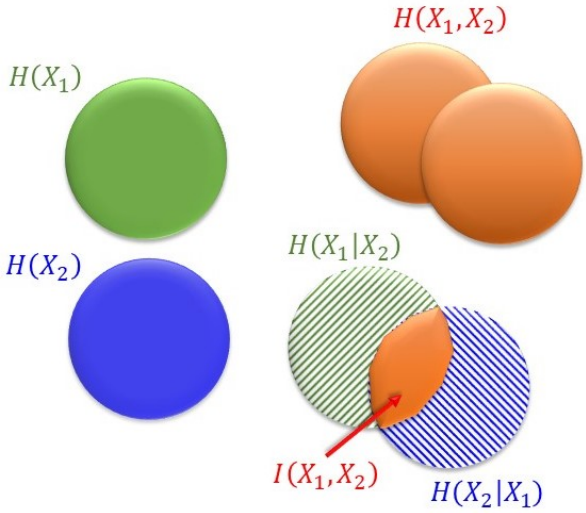

Figure 1: Mutual Information, Venn's diagram

3. UNCERTAINTY QUANTIFICATION

3.1. *Maximum Likelihood and KL Divergence*

Let consider now a set of n_s data $\mathbf{x} = \{x^{(1)},x^{(2)},\ldots,x^{(n)}\}$ drawn independently from the true but unknown target distribution $f(x)$. Let $f_m(x;\theta)$ the adopted probability model of parameters $\theta = \{\theta_1,\theta_2,\ldots,\theta_p\}$. The Maximum Likelihood Estimation (MLE) gives the model θ_{MLE} with the highest probability to the observed data

$$\begin{aligned} \theta_{MLE} &= \max_{\theta}\left(\frac{1}{n_s}\sum_{i=1}^{n_s}\log\left[f_m(x^{(i)};\theta)\right]\right) \\ &\approx \max_{\theta} E\left[\log\left(f_m(x;\theta)\right)\right] \end{aligned} \tag{7}$$

The KL divergence provides the lost information in representing $f(x)$ with the model $f_m(x)$, while the minimum divergence is

$$\begin{aligned} \theta^* &= \min_{\theta} D\left[f(x), f_m(x;\theta)\right] \\ &= \max_{\theta}\int f(x)\log\left[f_m(x;\theta)\right]dx \equiv \theta_{MLE} \end{aligned} \tag{8}$$

and the optimal parameters θ coincide. However, differenty from the maximum likelihood, the KL divergence has a known minimum value of zero. This can be helpful in the choice of the optimal statistical model. Different probability models are compared, including parametric distributions (e.g. Gaussian, Lognormal, Weibull, etc) and nonparametric ones. To choose the best model, we use the theory of the statistical learning [Vapnik (1995)]: the original dataset of n_s data is split into the *training set* $\mathbf{x}_{tr}$ of n_{tr} data and the *test set* $\mathbf{x}_{ts}$ of n_{ts} data, with $n_s = n_{tr} + n_{ts}$. Typically the training set collects 80% data of the entire dataset. The parameters θ_i^* of the probabilistic models $f_i(x, \theta_i^* | \mathbf{x}_{tr})$ are evaluated through eq.(7) or eq.(8) applied on the training set $\mathbf{x}_{tr}$. For each model $f_{m,i}(x)$ obtained, the minimum relative entropy with respect to the empirical distribution $\hat{f}(x)$based on the test set $\mathbf{x}_{ts}$ is evaluated

$$D_i = D\left[\hat{f}(x|\mathbf{x}_{ts}), f_{m,i}(x;\theta|\mathbf{x}_{tr})\right] \tag{9}$$

The optimal model for the given dataset $\mathbf{x}$ has the minimum divergence D_i. For structural reliability analysis, an attractive model is represented from the Kernel Density Maximum Entropy (KDME) [Alibrandi and Mosalam (2017)] since it provides the least biased distribution given the available information, including the tails, from a reduced sample of data.

3.2. Kernel Density Maximum Entropy Method

Let us consider a random variable X, whose PDF is $f(x)$ with support Ω. The target PDF, is expressed as a linear superposition of Kernel Density Functions (KDFs) as follows:

$$f(x) \approx f_{KD}(x;\mathbf{p}) = \sum_{i=1}^{N} p_i f_i^K(x;x_i,h) \tag{10}$$

where the coefficients p_i satisfy the constraints $0 \le p_i \le 1$, $\sum_i p_i = 1$, while $f_i^K(x;x_i,h)$ is the ith basis KDF, centered in x_i with bandwith h. If the Gaussian distribution is chosen as KDF, then x_i and h are the mean value and the standard deviation σ of the Gaussian KDF, respectively. The centers x_i, $i = 1,2,\ldots,N$, are uniformly spaced with a constant step $\Delta x = x_{i+1} - x_i$ in the range $[x_{min}, x_{max}]$. The bandwidth is $h = (2/3)\Delta x$, which is shown to be a suitable value under uniform spacing of the centers [Alibrandi and Ricciardi (2008); Alibrandi and Mosalam (2017)]. It is noted that when $N \to \infty$, then $h \to 0$, and Eq.(10) gives

$$f_{KD}(x;\mathbf{p}) = \sum_{i=1}^{N} p_i \delta(x - x_i) \tag{11}$$

where δ is the Dirac delta function. Therefore, the representation (10),(11) can reconstruct any kind of distribution. Let us consider a set of M independent functions $g_k(x;\alpha)$ of parameters α, representing the available information. Multiplying both sides of Eq.(11) by $g_k(x;\alpha)$, $k = 1,2,\ldots,M$, and integrating over the domain, the following relationship holds:

$$\begin{aligned} \mathbf{1}^T\mathbf{p} &= 1 \\ \mathbf{M}(\alpha)\mathbf{p} &= \mu(\alpha) \end{aligned} \tag{12}$$

where $\mathbf{1}$ is a vector of N unit entries, and $\mathbf{p}$ collects the N parameters $p_1, p_2, \ldots, p_N$, while $\mathbf{M}(\alpha)$ and $\mu(\alpha)$ are defined as

$$\begin{aligned} M_{ki}(\alpha) &= \int g_k(x;\alpha)\delta(x - x_i)dx = g_k(x_i;\alpha) \\ \mu_k(\alpha) &= \int g_k(x;\alpha) f(x)dx = E[g_k(x;\alpha)] \end{aligned} \tag{13}$$

According to Jaynes (1957), the Maximum Entropy (ME) probability distribution, $\mathbf{p}_{ME}$, is the least biased distribution, given the satisfaction of the available information. It is obtained through the maximization of the Shannon's entropy [Eq.(2)]:

$$\begin{aligned} \max_{\mathbf{p}} \quad & H(\mathbf{p}) \\ \mathbf{1}^T\mathbf{p} &= 1 \\ \mathbf{M}(\alpha)\mathbf{p} &= \mu(\alpha) \end{aligned} \tag{14}$$

The optimization problem (14) is convex, which implies the uniqueness of the ME distribution $\mathbf{p}_{ME}$

$$p_i^{ME}(\lambda;\alpha) = \exp\left[-\lambda_0(\lambda) - \sum_{k=1}^{M} \lambda_k g_k(x_i;\alpha)\right] \qquad i = 1,2,\ldots,N \tag{15}$$

where λ collects the M Lagrange multipliers $\lambda_1, \lambda_2, \ldots, \lambda_M$ of the dual optimization problem, while

$$\lambda_0(\lambda;\alpha) = \log\left[\sum_{i=1}^{N} \exp\left(-\sum_{k=1}^{M} \lambda_k g_k(x_i;\alpha)\right)\right] \tag{16}$$

The Lagrange parameters λ can be determined as a solution of a linear system of M equations [Alibrandi and Mosalam (2017)]

$$\Theta(\alpha)\lambda = \rho(\alpha) \tag{17}$$

From Eqs.(15)-(17), the parameters $\mathbf{p}_{ME}$ are determined, and after substitution into Eq.(10), the Kernel Density Maximum Entropy (KDME) distribution is obtained, i.e. $f_{KDME}(x) = f_{KD}(\mathbf{x};\mathbf{p}_{ME})$.

The KDMEM may be summarized as follows:

1. The original dataset of n_s data is split into the training set $\mathbf{x}_{tr}$ of n_{tr} data, the validating set $\mathbf{x}_{tv}$ of n_{tv} data and the test set $\mathbf{x}_{ts}$ of n_{ts} data, with $n_s = n_{tr} + n_{tv} + n_{ts}$.
2. Differents sets of hyperparameters $\alpha^{(1)}, \alpha^{(2)}, \ldots$ are chosen
3. For each set of hyperparameters $\alpha^{(k)}$
   - the optimal parameters $\mathbf{p}_{ME}(\alpha^{(k)})$ are evaluated through eq.(14) or eqs.(15)-(17);
   - the minimum relative entropy with respect to the empirical distribution based on the validation set $\mathbf{x}_{tv}$ is evaluated

   $$D_k = D\left[\hat{f}(x|\mathbf{x}_{tv}), f_{KDME}(x|\alpha^{(k)}, \mathbf{x}_{tr})\right] \tag{18}$$
4. The KDME model minimizes the divergence (18)
5. The performances of KDMEM are evaluated by determining the minimum relative entropy with respect to the empirical distribution based on the test set $\mathbf{x}_{ts}$, see eq.(9)

   $$D_{KDME} = D\left[\hat{f}(x|\mathbf{x}_{ts}), f_{KDME}(x)\right] \tag{19}$$

3.3. Distributions with Independent Components (DIC)

Let $\mathbf{X}$ a vector collecting the n dependent random variables $X_1, X_2, \ldots, X_n$. The aim is the determination of its joint Probability Density Function $f_{\mathbf{X}}(\mathbf{x}) = f_{X_1 X_2 \ldots X_n}(x_1, x_2, \ldots, x_n)$. Let $\mu_{\mathbf{X}}$ and $\Sigma_{\mathbf{X}}$ be the mean value and the covariance matrix of $\mathbf{X}$. First, a normalization of the random variables is developed as follows:

$$\mathbf{Z} = \Psi(\mathbf{X} - \mu_{\mathbf{X}}) \tag{20}$$

where $Z_1, Z_2, \ldots, Z_n$ are n uncorrelated (but not independent) random variables with zero mean and unit variance, i.e. $\mu_{\mathbf{Z}} = \mathbf{0}$ and $\Sigma_{\mathbf{Z}} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix of order n. Let us assume that the covariance matrix is written as $\Sigma_{\mathbf{X}} = \mathbf{D}_{\mathbf{X}}\mathbf{R}\mathbf{D}_{\mathbf{X}}$, where $\mathbf{D}_{\mathbf{X}} = diag[\sigma_{X_i}]$ is the diagonal matrix of standard deviations of $\mathbf{X}$ and $\mathbf{R} = [\rho_{ij}]$ is the correlation matrix. In such case $\Psi = \mathbf{L}^{-1}\mathbf{D}^{-1}$, where $\mathbf{L}$ is the lower triangular decomposition of the correlation matrix, i.e. $\mathbf{R} = \mathbf{L}\mathbf{L}^T$. Alternatively, Ψ can be obtained through eigenvalue and eigenvector analysis.

Second, we assume $Z_1, Z_2, ..., Z_n$ can be expressed as a linear combination of n unknown independent and usually non-Gaussian random variables $Y_1, Y_2, ..., Y_n$, called Independent Components (IC):

$$\mathbf{Z} = \mathbf{A}\mathbf{Y} = \sum_{j=1}^{n} \mathbf{a}_j Y_j \tag{21}$$

where $\mathbf{A} = [\mathbf{a}_1 \mathbf{a}_2 ... \mathbf{a}_n]$ is the mixing matrix of order $n \times n$, while the vectors $\mathbf{a}_j$ are its columns. It is assumed that both the n ICs $Y_1, Y_2, ..., Y_n$ and matrix $\mathbf{A}$ are unknown, while we only have samples of the vector $\mathbf{X}$, and consequently $\mathbf{Z}$ through Eq. (20). The fundamental assumption of ICA is that the components $Y_1, Y_2, ..., Y_n$ are statistically independent and that at least one of them follows a non-Gaussian distribution. It is also assumed that $\mathbf{A}$ is invertible, so that an alternative formulation of Eq.(21) is

$$\mathbf{Y} = \mathbf{W}\mathbf{Z} = \sum_{j=1}^{n} \mathbf{w}_j Z_j \tag{22}$$

where $\mathbf{W} = \mathbf{A}^{-1} = \mathbf{W}^T$. The main objective is the evaluation of the directions $Y_1, Y_2, ..., Y_n$ with maximum independence. To this aim, we minimize their mutual information $I(\mathbf{Y})$

$$I(Y_1, Y_2, ..., Y_n) = \sum_{i=1}^{n} H(Y_i) - H(\mathbf{Y}) \tag{23}$$

It is noted that $I(\mathbf{Y})$ takes into account the whole probabilistic dependence of the random variables, and not only the covariances. Combining Eqs.(20) to (22), one obtains:

$$\mathbf{Y} = \mathbf{P}(\mathbf{X} - \mu_{\mathbf{X}}) \tag{24}$$

$$\mathbf{X} = \mu_{\mathbf{X}} + \mathbf{QY} \tag{25}$$

where $\mathbf{P} = \mathbf{W}\Psi = \mathbf{W}\mathbf{L}^{-1}\mathbf{D}^{-1}$, while $\mathbf{Q} = \mathbf{P}^{-1} = \mathbf{DLA}$. From Eq.(25) it follows that the generation of samples of the ICs $\mathbf{Y}$ allows to determine samples of the joint random variables $X_1, X_2, ..., X_n$. The joint distribution of $\mathbf{X}$ is obtained by the elementary probability transformation rule $f_{\mathbf{X}}(\mathbf{x})d\mathbf{x} = f_{\mathbf{Y}}(\mathbf{y})d\mathbf{y}$ and considering that $Y_1, Y_2, ..., Y_n$ are independent, we obtain the following:

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= f_{\mathbf{Y}}(\mathbf{y})|\mathbf{J}_{\mathbf{y},\mathbf{x}}| \\ &= |\mathbf{P}| f_{Y_1}(y_1) f_{Y_2}(y_2) ... f_{Y_n}(y_n) \end{aligned} \tag{26}$$

where $\mathbf{J}_{\mathbf{y},\mathbf{x}}$ is the Jacobian of the transformation (24). Eq.(26) shows that the evaluation of the joint PDF $f_{\mathbf{X}}(\mathbf{x})$ is simply reduced to the evaluation of the univariate PDFs of $Y_1, Y_2, ..., Y_n$. DIC may model the exact full dependence of the random variables including the tails. This, differently from Nataf, where the dependence is described only in terms of correlations.

The following main particular cases of DIC may be detected: (i) *Parametric Distributions with Independent Components* (PDIC) where the ICs are modelled through the classical parametric distributions (Gaussian, Lognormal, Weibull, Gamma, etc.), (ii) *Kernel Density with Independent Components* (KDIC) where ICs are modelled through the Kernel Density Estimation, (iii) *Kernel Density Maximum Entropy Method with Independent Components* (KDMEMIC) where ICs are modelled through the Kernel Density Maximum Entropy. This classification of course is not exhaustive, since any univariate probabilistic model can be converted into the multivariate case, through the DIC framework.

4. STRUCTURAL RELIABILITY: ISOPROBABILISTIC TRANSFORMATION

In structural reliability analysis, it is common to develop an isoprobabilstic transformation toward the normal standard space, so that Eq.(1) becomes

$$P_f = \int_{g(\mathbf{u})\leq 0} \varphi_n(\mathbf{u}) d\mathbf{u} \tag{27}$$

Typically the Nataf transformation is adopted, but as discussed above, it assumes the tail independence of the distribution, which however it is not always realistic. In this regard DIC is an attractive tool because not only it may provide the exact failure probability, but also the transformation from the space of the IC $y_1, y_2, \ldots, y_n$ toward the normal standard space $u_1, u_2, \ldots, u_n$ is straightforward, pursued component by component

$$\Phi(u_i) = F_{Y_i}(y_i), \quad i = 1, 2, \ldots, n \tag{28}$$

5. NUMERICAL APPLICATION

The joint distribution $f_{T_z,H_s}(t,h)$ of the significant wave height and period is considered. The DNV code (2010) suggests to use the conditional model approach, i.e.

$$f_{T_z H_s}(t,h) = f_{T_z|H_s}(t|h) f_{H_s}(h_s) \tag{29}$$

where the significant wave height H_s is modeled by the following 3-parameter Weibull PDF:

$$f_{H_s}(h) = \frac{\beta_{H_s}}{\alpha_{H_s}} \left(\frac{h - \gamma_{H_s}}{\alpha_{H_s}} \right) \exp\left[- \left(\frac{h - \gamma_{H_s}}{\alpha_{H_s}} \right)^{\beta_{H_s}} \right] \tag{30}$$

The zero-crossing wave period T_z conditional on H_s is modeled by the following lognormal distribution

$$f_{T_z|H_s}(t|h) = \frac{1}{\sigma t \sqrt{2\pi}} \exp\left[- \frac{(\log t - \mu)^2}{2\sigma^2} \right] \tag{31}$$

where the distributions of parameters μ and σ depend on the significant wave height

$$\begin{aligned} \mu &= E[\log T_z] = a_0 + a_1 h^{a_2}, \\ \sigma &= std[\log T_z] = b_0 + b_1 h^{b_2} \end{aligned} \tag{32}$$

where a_i, b_i, $i = 0,1,2$ are estimated from the data. The DNV code suggests the parameters appearing in Eqs.(29)-(32) for various world-wide nautical zones. The nautical zone 63, close to Singapore, has been chosen, whose parameters are $\alpha_{H_s} = 1.88$, $\beta_{H_s} = 1.70$, $\gamma_{H_s} = 0.0$, $a_0 = 0.70$, $a_1 = 1.026$, $a_2 = 0.155$, $b_0 = 0.07$, $b_1 = 0.1477$, and $b_2 = -0.0224$. The joint PDF (29) is taken as the target and it is used to benchmark DIC. In Figure 2, we represent 10,000 samples generated from the target model, Eqs.(29) to (32).

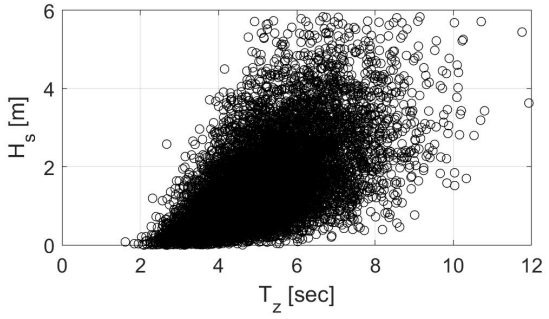

Figure 2: Wave data

Fig.3 shows the joint probability of exceeding

$$P = Prob\left[(T_z \geq 2\rho) \cap (H_s \geq \rho)\right] \qquad (33)$$

where ρ is a threshold ranging from 0.0 to 5.5 with steps of 0.25 for different models: target, Nataf and KDMEMIC trained with $m = 1,000$ samples. Each probability level is estimated through MCS by assuming a coefficient of variation $v_P = 2\%$.

In Fig.4 the relative error incurred by the two models is described. In the Nataf model, the marginal distributions $f_{T_z}(t)$ and $f_{H_s}(h)$ of T_z and H_s are described through a Lognormal and a Weibull distribution, respectively. It is seen that Nataf is in good agreement with the marginal empirical distributions. Consequently, as expected, the approximation given by Nataf for the lowest threshold is excellent, while the error incurred for the highest thresholds is to be attributed to the assumption of the tail independency. The results show the good

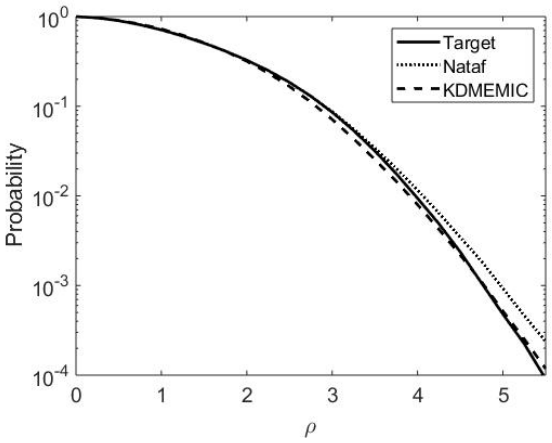

Figure 3: Joint probability of exceeding of $\{H_s, T_p\}$

prediction capabilities of KDMEMIC for the joint extreme values of the distributions, even if trained over samples of reduced size.

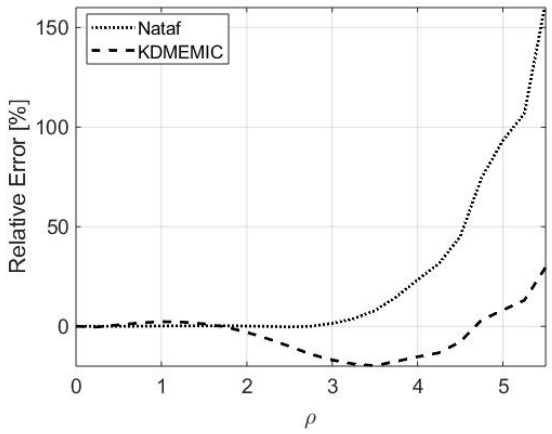

Figure 4: Joint probability of exceeding of $\{H_s, T_p\}$, relative error

6. CONCLUSIONS

In this paper a novel framework for Uncertainty Quantification and Structural Reliability analysis has been presented. It is based on Information Theory and Machine Learning. In the earlier stages of structural reliability, there was lack of available information, so the uncertainty quantification of the basic random variables under incomplete information was developed through statistical approaches, and it involved the knowledge of the marginal distributions and their correlation. Nowadays, the availability of data from experiments or networks

of sensors installed on the built environment, high performance computing and advanced probabilistic computational tools may allow enhanced modelling of the uncertain quantities and better estimates of the failure probability. This work goes into this direction. The effectiveness of the proposed framework has been shown through the prediction of joint extremes of wave data, and its extension to high dimensional reliability analysis of complex structures is straightforward. Thus, it has the potential of provide an enhanced tool of risk-based design for the fourth (digital) revolution industrial.

7. ACKNOWLEDGMENTS

This research was funded by the Republic of Singapore's National Research Foundation through a grant to the Berkeley Education Alliance for Research in Singapore (BEARS) for the Singapore Berkeley Building Efficiency and Sustainability in the Tropics (SinBerBEST) program. BEARS has been established by the University of California, Berkeley, as a center for intellectual excellence in research and education in Singapore. K.M. Mosalam is a core principal investigator of Tsinghua-Berkeley Shenzhen Institute (TBSI). The authors acknowledge the funding support from SinBerBEST and the partial support from TBSI.

8. REFERENCES

Alibrandi, U., Alani, A., and Ricciardi, G. (2015). "A new second-order svm-based response surface for structural reliability analysis." *Probabilistic Engineering Mechanics*, 41, 1:12.

Alibrandi, U. and Mosalam, K. (2017). "Kernel density maximum entropy with generalized moments for evaluating probability distributions, including tails, from a small sample of data." *International Journal for Numerical Methods in Engineering*, 113(13), 1904–28.

Alibrandi, U. and Mosalam, K. (2018). "Code-conforming peer performance based earthquake engineering using stochastic dynamic analysis and information theory." *KSCE Journal of Civil Engineering*, 22(3), 1002–15.

Alibrandi, U. and Ricciardi, G. (2008). "Efficient evaluation of the pdf of a random variable through the kernel-density maximum entropy approach." *International Journal for Numerical Methods in Engineering*, 75(13), 1511–48.

Beer, M., Zhang, Y., Quek, S., and Phoon, K. (2013). "Reliability analysis with scarce information: comparing alternative approaches in a geotechnical engineering context." *Structural Safety*, 41, 1:10.

Bucher, C. and Burgound, U. (1990). "A fast and efficient response surface approach for structural reliability problems." *Structural Safety*, 7, 57:66.

Der Kiureghian, A. and Liu, P. L. (1986). "Structural reliability under incomplete probability information." *Journal of Engineering Mechanics*, 111(1), 85–104.

Ditlevsen, O. and Madsen, H. O. (1996). *Structural Reliability Methods*. Wiley.

DNV (2010). *Recommended practice DNV-RP-C205, Environmental conditions and environmental loads.*

Dubourg, C., Sudret, B., and Bourinet, J. (2011). "Reliability-based design optimization using kriging surrogates and subset simulation." *Structural and Multidisciplinary Optimization*, 44(5), 673:690.

Dutfoy, A. and Lebrun, R. (2009). "Practical approach to dependence modelling using copulas." *Proc. IMech. E, Part O: J. Risk and Reliability*, 223, 347–361.

Elishakoff, I. (1991). "Essay on reliability index, probabilistic interpretation of safety factor and convex models of uncertainty." *Reliability Problems: General Principles and Applications in Mechanics of Solids and Structures*.

Jaynes, E. T. (1957). "Information theory and statistical mechanics." *Phys. Rev.*, 106, 620–30.

Jaynes, E. T. (1978). "Where do we stand on maximum entropy?." *Maximum Entropy formalism*, 15–118.

Lebrun, R. and Dutfoy, A. (2009). "An innovating analysis of the nataf transformation from the copula viewpoint." *Probabilstic Engineering Mechanics*, 24, 312–320.

Madsen, H. O., Krenk, S., and Lind, N. C. (1986). *Methods of Structural Safety*. Prentice-Hall.

Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. 10.1109/tnn.1997.641482 edition.