Received: 14 November 2016 | Revised: 2 October 2017 | Accepted: 29 October 2017

DOI: 10.1002/nme.5725

RESEARCH ARTICLE

WILEY

Kernel density maximum entropy method with generalized moments for evaluating probability distributions, including tails, from a small sample of data

Umberto Alibrandi[1] | Khalid M. Mosalam[2]

[1]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 138062, Singapore

[2]Department of Civil and Environmental Engineering, University of California at Berkeley, Berkeley, California 94720-1710, USA

Correspondence
Umberto Alibrandi, Berkeley Education Alliance for Research in Singapore Limited (BEARS), CREATE Tower, 1 Create Way, #11-01, 138062 Singapore.
Email: umbertoalibrandi@gmail.com

Funding information
Republic of Singapore's National Research Foundation

Abstract

In this paper, a novel method to determine the distribution of a random variable from a sample of data is presented. The approach is called generalized kernel density maximum entropy method, because it adopts a kernel density representation of the target distribution, while its free parameters are determined through the principle of maximum entropy (ME). Here, the ME solution is determined by assuming that the available information is represented from *generalized* moments, which include as their subsets the power and the fractional ones. The proposed method has several important features: (1) applicable to distributions with any kind of support, (2) computational efficiency because the ME solution is simply obtained as a set of systems of linear equations, (3) good trade-off between bias and variance, and (4) good estimates of the tails of the distribution, in the presence of samples of small size. Moreover, the joint application of generalized kernel density maximum entropy with a bootstrap resampling allows to define credible bounds of the target distribution. The method is first benchmarked through an example of stochastic dynamic analysis. Subsequently, it is used to evaluate the seismic fragility functions of a reinforced concrete frame, from the knowledge of a small set of available ground motions.

KEYWORDS

fractional moments, fragility functions, kernel density estimation, maximum entropy, reinforced concrete frame, stochastic dynamic analysis

1 | INTRODUCTION

A random variable is fully characterized from a probabilistic point of view by the knowledge of its probability density function (PDF) or cumulative distribution function (CDF). Therefore, the evaluation of the probability distribution is of concern. Generally, a sample of data is available, collected from experimental data or a network of sensors. It is assumed that the distribution of the population is known (eg, Gaussian, lognormal, or Weibull) whose parameters are estimated from the sample data, through the "method of the moments" or the "maximum likelihood method." However, in some cases, the known parametric distributions cannot accurately fit the data and/or have good prediction capabilities over the unseen data.

In literature, several techniques have been proposed to determine a PDF from the knowledge of the first 4 moments, eg, the Pearson family of distributions[1] or the recently proposed shifted generalized lognormal distribution.[2] However, it is known that further information about the distribution can be obtained by also considering higher-order moments or

other quantities derived from the data. The first general representations for a PDF have adopted Hermite polynomials, eg, A-type and C-type Gram Charlier series and the Edgeworth series expansion.[3,4] An improvement is represented by the model proposed by Winterstein,[5] and it is largely applied in engineering.

The choice of the distribution type is an open issue, from a conceptual and practical point of view. In the presence of limited information (sample of small size and/or lower-order moments), in the most general case, there is no theoretical justification to prefer one distribution over another. In such case, an attractive technique is based on the principle of maximum entropy (ME).[6-8] The ME distribution represents the least biased distribution given the available information. However, the application of the ME principle, as typically adopted in literature, has some practical and theoretical limitations.

If the power moments represent the available information, then the ME PDF $f_{ME}(x)$ may require a large number M of moments ($M \geq 4$) to accurately describe the tails of the distribution. However, the moment problem is an ill-posed problem.[9] Thus, for large M, the entropy maximization algorithm experiences numerical instability. Moreover, at the tails, the ME distribution oscillates because of the nonmonotonic nature of the polynomial embedded in the $f_{ME}(x)$. Thus, only the lower-order moments are typically considered, but in such case, $f_{ME}(x)$ hardly models tails fatter than the Gaussian. Therefore, the tails of many distributions cannot be well fitted by the ME distribution with $M \leq 4$.[10] Furthermore, from a practical point of view, we do not have the *true moments*, but samples of data, from which the *sample moments* are derived. Depending on the size of the sample, the estimates of the higher-order moments provide large statistical errors, which imply high bias in the estimate of the tails of the ME PDF. Consequently, the modeling of the tails of the distributions from samples of data is challenging.

To this aim, we consider the application of the ME principle with the fractional moments, defined as powers of real numbers and not only integers.[11-13] It is known that (1) a reduced number of fractional moments may accurately model the tails of $f_{ME}(x)$ and (2) the fractional moments can be well estimated from a reduced sample of data. However, several issues remain open. First, while an infinite sequence of power moments uniquely characterizes most random variables of practical interest (with some notable exceptions, eg, lognormal distribution), an infinite sequence of fractional moments only characterizes positive random variables. Second, in the existing literature, the optimal powers of the fractional moments are determined within the entropy maximization procedure; however, the resulting optimization procedure is not convex, so that, given a sample of data, the uniqueness of the ME distribution may not be guaranteed. Third, the determination of $f_{ME}(x)$ requires the numerical evaluation of integrals, which is generally complicated.

The main objective of this paper is to develop a unified framework based on the ME, applicable to random variables with any kind of support (bounded, semibounded, or unbounded) and able to provide the least biased reconstruction of the distribution including its tails, from the knowledge of a small number of available data. This is achieved by discretizing the moment problem representing the target PDF $f_X(x)$ as a sum of kernel densities $f_{KDME}(x;\boldsymbol{p}) \cong f_X(x)$[14,15] whose free parameters collected in the vector $\boldsymbol{p}$ are obtained by applying the ME principle to the discretized moment problem. Differently from Alibrandi and Ricciardi,[15] here, the constraints of the ME method are the generalized moments, which include, as their subsets, the classical power and the fractional moments. The proposed approach is referred to as generalized kernel density maximum entropy method (GKDMEM) because it adopts a kernel density (KD) representation (KDR) of the target PDF, while the parameters are determined through the ME method, by using as constraints the generalized moments.

The paper is devoted in particular to engineering problems where only samples of data are available, and not the corresponding population generalized moments. In such cases, the main task is the evaluation of a good bias-variance tradeoff, such that the kernel density maximum entropy (KDME) distribution follows the data, but it keeps good prediction capabilities over the unseen data. This issue is explored through the Akaike information criterion,[16] which allows to define the optimal $f_{KDME}(x;\boldsymbol{p})$, for the given data sample.

After describing the main features of the proposed method, the GKDMEM is benchmarked to a Duffing oscillator, whose response distribution is known in a closed form. Subsequently, the method is applied to a reinforced concrete frame subjected to seismic excitation, to evaluate the corresponding fragility functions. This example shows that the GKDMEM gives an optimal distribution with respect to the available information.

2 | ME FORMALISM

2.1 | Entropy

In information theory, the entropy of a random variable 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, $i = 1, 2, ..., N$. The *Shannon's entropy functional*[6,7] of the discrete distribution $\boldsymbol{p}$ is defined as follows:

$$H(\boldsymbol{p}) = H(p_1, p_2, ..., p_N) = -\sum_{i=1}^{N} p_i \log p_i. \tag{1}$$

The entropy is larger when the random variable X is more "random," ie, more unpredictable and unstructured. This is seen in Equation 1 where it is noted that the entropy is small if the values of p_i are all close to zero or to one, and large otherwise.

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

$$H(f) = -\int f_X(x) \log f_X(x) dx. \tag{2}$$

Let us assume that X follows a uniform distribution $f_X(x) = 1/\xi$, defined in the interval $X \in [0, \xi]$. In view of Equation 2, its entropy is $H(f) = \log(\xi)$, and it is large for ξ large. This is expected since a small value of ξ implies that X is concentrated on a small interval, and its entropy (ie, randomness) is small. It is here noted that, differently from the Shannon's entropy $H(\boldsymbol{p})$ for discrete-valued random variables, the differential entropy $H(f)$ is not a correct information measure for continuous random variables, since it can assume negative values and lacks invariance under a change of variables. Jaynes has shown[17,18] that a good information measure for a continuous random variables is

$$H_c(f) = -\int f(x) \log\left[\frac{f(x)}{m(x)}\right] dx, \tag{3}$$

where $m(x)$ is an "invariant measure" function. Since $f(x)$ and $m(x)$ transform in the same manner under a change of variables $x \to y(x)$, it follows that $H_c(f)$ is invariant. It is also closely related to the Kullback-Leibler (KL) divergence $D(f_1, f_2)$ (also called relative entropy) measuring the entropy difference between 2 PDFs

$$D(f_1, f_2) = \int f_1(x) \log\left[\frac{f_1(x)}{f_2(x)}\right] dx = \int f_1(x) \log[f_1(x)] dx - \int f_1(x) \log[f_2(x)] dx = H(f_1, f_2) - H(f_1). \tag{4}$$

By comparing Equations 3 and 4, it is seen that $H_c(f) = -D(f_1, f_2)$ by setting $f(x) \equiv f_1(x)$ and $m(x) \equiv f_2(x)$. The KL divergence may be considered a measure of the "distance" between 2 distributions, since it is always nonnegative for all the possible distributions $f_1(x)$ and $f_2(x)$, while $D(f_1, f_2) = 0$ only when $f_1(x) \equiv f_2(x)$.

2.2 | ME probability distribution of a discrete random variable

Let us consider a discrete-valued random variable X with probability distribution given by p_i evaluated at x_i, $i = 1, 2, ..., N$, and $\sum_{i=1}^{N} p_i = 1$. The constraints imposed by the available information of X are given by $G_k \equiv E[g_k] = \sum_{i=1}^{N} g_k(x_i) p_i$, $k = 1, 2, ..., M$. According to Jaynes,[19] the best (minimally prejudiced) assignment probabilities, subjected to the satisfaction of the available information, maximize the entropy as follows:

$$\begin{cases} \max_{\boldsymbol{p}} H(\boldsymbol{p}) \\ \sum_{i=1}^{N} p_i = 1 \\ \sum_{i=1}^{N} g_k(x_i) p_i = G_k, \quad k = 1, 2, ..., M \end{cases}, \tag{5}$$

where $H(\boldsymbol{p})$ is the Shannon's entropy defined in Equation 1. The solution of Equation 5 is known as the ME *probability distribution*, 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."[18] The extended unconstrained functional of Equation 5 reads as

$$\mathcal{L}_p(\boldsymbol{p},\boldsymbol{\lambda}) = H(\boldsymbol{p}) - (\lambda_0 - 1)\left(\sum_{i=1}^{N} p_i - 1\right) - \sum_{k=1}^{M} \lambda_k \left(\sum_{i=1}^{N} g_k(x_i) p_i - G_k\right), \tag{6}$$

where $\boldsymbol{\lambda}$ collects the Lagrange multipliers λ_1, λ_2, ..., λ_M corresponding to the constraints containing, respectively, the averages G_1, G_2, ..., G_M. By imposing the stationarity conditions with respect to p_i of $\mathcal{L}_p(\boldsymbol{p},\boldsymbol{\lambda})$ (Equation 6), one obtains

$$p_i^{ME}(\lambda_1, \lambda_2, ..., \lambda_M) = N_0(\lambda_1, \lambda_2, ..., \lambda_M) \exp\left(-\sum_{k=1}^{M} \lambda_k g_k(x_i)\right), \quad i = 1, ..., N, \tag{7}$$

where $N_0 = \exp(-\lambda_0)$ is a normalization constant. The Lagrange multipliers can be obtained as a solution of a system of M nonlinear equations, obtained by substituting $p_i^{ME}(\boldsymbol{\lambda})$ of Equation 7 into the M constraints of Equation 5. Typically, a more efficient numerical procedure is used. By substituting $p_i^{ME}(\boldsymbol{\lambda})$ into the definition of the Shannon's entropy (Equation 1), it follows that the Lagrange multipliers are evaluated through the minimization of the following unconstrained convex functional[8,15]:

$$\Gamma^{ME}(\lambda_1, \lambda_2, ..., \lambda_M) = \lambda_0(\lambda_1, \lambda_2, ..., \lambda_M) + \sum_{k=1}^{M} \lambda_k G_k, \tag{8}$$

where

$$\lambda_0(\lambda_1, \lambda_2, ..., \lambda_M) = \log\left\{\sum_{i=1}^{N} exp\left(-\sum_{k=1}^{M} \lambda_k g_k(x_i)\right)\right\}. \tag{9}$$

The convexity of $\Gamma^{ME}(\boldsymbol{\lambda})$ implies the uniqueness of the ME solution in terms of λ_1, λ_2, ..., λ_M. The distribution of X in terms of p_i^{ME}, $i = 1, 2, ..., N$, is obtained by substituting λ_1, λ_2, ..., λ_M into Equation 7. It is also noted that it is always possible to determine the ME discrete distribution for the generalized moment problem, see Alibrandi and Ricciardi[15] and Tagliani.[20]

2.3 | ME probability distribution of a continuous random variable

Let us consider a continuous-valued random variable X with PDF $f_X(x)$ whose available information given by the constraints are $G_k \equiv E[g_k] = \int g_k(x) f_X(x) dx$, and the normalization condition $\int f_X(x) dx = 1$. In such case, the ME principle gives rise to

$$\begin{cases} \max\limits_{f} H(f) \\ \int f_X(x) dx = 1 \\ \int g_k(x) f_X(x) dx = G_k, \quad k = 1, 2, ..., M \end{cases}, \tag{10}$$

where $H(f)$ is the differential entropy defined in Equation 2, while the Lagrangian is

$$\mathcal{L}_f(f, \boldsymbol{\lambda}) = H(f) - (\lambda_0 - 1)\left(\int f_X(x) dx - 1\right) - \sum_{k=1}^{M} \lambda_k \left(\int g_k(x) f_X(x) dx - G_k\right). \tag{11}$$

The ME PDF is

$$f_{ME}(x) = \mathrm{N}_0(\lambda_1, \lambda_2, ..., \lambda_M) \exp\left(-\sum_{k=1}^{M} \lambda_k g_k(x)\right), \tag{12}$$

where λ_1, λ_2, ..., λ_M are the Lagrange multipliers, while $N_0 = \exp(-\lambda_0)$ is the normalization constant, with

$$\lambda_0(\lambda_1, \lambda_2, ..., \lambda_M) = log\left[\int \exp\left(-\sum_{k=1}^{M} \lambda_k g_k(x)\right) dx\right]. \tag{13}$$

It can be shown that a correspondence one to one exists between the averages G_1, G_2, ..., G_M and the Lagrange multipliers λ_1, λ_2, ..., λ_M. Therefore, Equation 12 can be considered as a family of distributions of the ME $f_{ME}(x;\boldsymbol{\lambda})$ of

parameters $\boldsymbol{\lambda}$. By substituting Equation 12 into $H(f)$, it is seen that $\lambda_1, \lambda_2, ..., \lambda_M$ can be determined as a solution of the following convex unconstrained optimization problem

$$\boldsymbol{\lambda}^* = \begin{cases} \max\limits_{\boldsymbol{\lambda}} H[f_{ME}(x;\boldsymbol{\lambda})] \\ \int g_k(x) f_{ME}(x;\boldsymbol{\lambda}) dx = G_k \end{cases} \equiv \min_{\boldsymbol{\lambda}} \Gamma^{ME}(\boldsymbol{\lambda}), \tag{14}$$

where the functional $\Gamma^{ME}(\lambda_1, \lambda_2, ..., \lambda_M)$ is defined by Equation 8, with λ_0 given by Equation 13 for the continuous case.

As discussed above, the differential entropy $H(f)$ is not a correct information measure for continuous random variables. However, it can be successfully adopted to determine the ME PDF, defined by Equation 12. To show this, the goal is to determine, inside the family of distributions $f_{ME}(x;\boldsymbol{\lambda})$ drawn by Equation 12, the closest distribution $f_{ME}(x;\boldsymbol{\lambda}^*)$ to $f_X(x)$ in terms of KL divergence. By assuming $f_1(x) \equiv f_X(x)$ and $f_2(x) \equiv f_{ME}(x;\boldsymbol{\lambda})$, Equation 4 gives

$$\min_{\boldsymbol{\lambda}} D[f_X(x), f_{ME}(x;\boldsymbol{\lambda})] = \min_{\boldsymbol{\lambda}} H[f_X(x), f_{ME}(x;\boldsymbol{\lambda})] = \min_{\boldsymbol{\lambda}} \Gamma^{ME}(\boldsymbol{\lambda}), \tag{15}$$

which provides the same optimal distribution of Equation 14.

2.4 | Discussions on ME distribution with power moments

Typically, the constraints representing the available information are expressed through the population power moments, ie, $g_k(x) = x^k$ and $G_k = E[X^k] \equiv \mu_k$. In such case, the ME distribution is

$$f_{ME}(x) = N_0 \exp\left(-\sum_{k=1}^{M} \lambda_k x^k\right) = N_0 \exp\left(-\lambda_1 x - \lambda_2 x^2 - ... - \lambda_M x^M\right). \tag{16}$$

The main problem with this choice is that an accurate description of the tails of the distribution requires a high number M of moments. But the moment problem is ill posed, so that, when M is high, numerical instability can arise inside the optimization algorithm of ME. Thus, only lower-order moments are typically considered, which may lead to inaccurate modeling of the tails of a broad family of distributions. As an example, consider the ME PDF for $M = 4$, which is $f_{ME}(x) = N_0 \exp(-\lambda_1 x - \lambda_2 x^2 - \lambda_3 x^3 - \lambda_4 x^4)$. For the sake of clarity, let us assume that the target distribution is symmetric with respect to the origin, so that $\lambda_1 = \lambda_3 = 0$ and $f_{ME}(x) = N_0 \exp(-\lambda_2 x^2 - \lambda_4 x^4)$. In such cases, when $|x| \to \infty$, then $f_{ME}(x)$ decays like $\exp(-\lambda_4 x^4)$. This implies that (1) $\lambda_4 \geq 0$ so that $f_{ME}(x) \to 0$ when $|x| \to \infty$ and (2) since $\lambda_4 \geq 0$, $f_{ME}(x)$ decays *at least as fast as the Gaussian density*. This makes the ME PDF of little use for softening behaviors, which are common cases of interest in structural engineering. For further details and numerical examples, see Winterstein and Mackenzie.[10]

In practical applications, we do not have the *true moments* μ_k, but samples of data $\{x^{(1)}, x^{(2)}, ..., x^{(n_s)}\}$, from which the *sample moments* m_k are derived. Typically, we assume that they coincide, that is,

$$\mu_k \cong m_k = \frac{1}{n_s} \sum_{j=1}^{n_s} \left[x^{(j)}\right]^k. \tag{17}$$

However, see Baker,[21] this assumption raises some questions: (1) the entropy is a measure of the information, and it should be dependent on the number n_s of sample data, and (2) an accurate description of the tails requires higher-order moments ($k > 2$). However, in practice, the sample size is usually small, and the estimates of the higher-order sample moments provide large statistical errors. Summarizing, from a given sample of data, if only lower-order sample moments are considered, there is inadequate information to model the tails of the distribution. Conversely, if higher-order sample moments are determined, these can represent unreliable estimates of the corresponding population moments. In both cases, the ME estimates of the tails of the target distribution may be significantly biased.

2.5 | ME distribution with fractional moments of positive continuous-valued random variables

The fractional moment of a *positive* continuous-valued random variable is defined as

$$\mu_\alpha = E[X^\alpha] = \int x^\alpha f_X(x) dx, \tag{18}$$

with α is a real number.[11-13] By developing a Taylor series expansion of x^{α} around a real constant c, the fractional moments (Equation 18) can be defined as

$$E[X^{\alpha}] = \sum_{i=0}^{\infty} \sum_{k=0}^{i} (-1)^{i-k} \binom{\alpha}{i} \binom{i}{k} c^{\alpha-k} E[X^k], \tag{19}$$

which shows that only few fractional moments can give the same information of several power moments. It is therefore expected that the adoption of M fractional moments as constraints may give more detailed information about the tails of the distribution compared to the use of M power moments.

Moreover, the fractional moments guarantee good performance from a numerical point of view. For a given sample of n_s data, the derived sample fractional moments are

$$m_{\alpha} \cong \mu_{\alpha} = \frac{1}{n_s} \sum_{j=1}^{n_s} \left[x^{(j)}\right]^{\alpha}. \tag{20}$$

This implies that the fractional moments can be defined for low values of α (eg, $\alpha \leq 2$), so that their estimates can be reliable even in the presence of samples of small size. Moreover, in view of Equation 19, they may give information over the tails of the distribution. The ME distribution is given by Equations 12 and 13, by setting $g_k(x) = x^{\alpha_k}$, that is,

$$f_{ME}(x;\boldsymbol{\lambda},\boldsymbol{\alpha}) = N_0 \exp\left(-\sum_{k=1}^{M} \lambda_k x^{\alpha_k}\right) = N_0 \exp(-\lambda_1 x^{\alpha_1} - \lambda_2 x^{\alpha_2} - \ldots - \lambda_M x^{\alpha_M}). \tag{21}$$

By comparing Equation 21 with Equation 16, it is seen that the ME PDF with fractional moments is representative of a broader family of distributions with respect to the case where only power moments are adopted, while they coincide only when $\alpha_k = k$, $k = 1, 2, \ldots, M$.

In the existing literature, the evaluation of the parameters $\boldsymbol{\lambda}$ and $\boldsymbol{\alpha}$ is obtained by minimizing the KL divergence between the target PDF $f_X(x)$ and $f_{ME}(x;\boldsymbol{\lambda},\boldsymbol{\alpha})$, as described above, Equation 15. It follows

$$\min_{\boldsymbol{\lambda},\boldsymbol{\alpha}} D[f_X(x), f_{ME}(x;\boldsymbol{\lambda},\boldsymbol{\alpha})] = \min_{\boldsymbol{\lambda},\boldsymbol{\alpha}} H[f_X(x), f_{ME}(x;\boldsymbol{\lambda},\boldsymbol{\alpha})] = \min_{\boldsymbol{\lambda},\boldsymbol{\alpha}} \Gamma_{\alpha}^{ME}(\lambda_1, \lambda_2, \ldots, \lambda_M, \alpha_1, \alpha_2, \ldots, \alpha_M). \tag{22}$$

In Equation 22, the functional $\Gamma_{\alpha}^{ME}(\boldsymbol{\lambda},\boldsymbol{\alpha})$ is defined by

$$\Gamma_{\alpha}^{ME}(\boldsymbol{\lambda},\boldsymbol{\alpha}) = \lambda_0(\boldsymbol{\lambda},\boldsymbol{\alpha}) + \sum_{k=1}^{M} \lambda_k \mu_{\alpha_k}, \tag{23}$$

where μ_{α_k} is the kth fractional moment while

$$\lambda_0(\boldsymbol{\lambda},\boldsymbol{\alpha}) = log\left[\int \exp\left(-\sum_{k=1}^{M} \lambda_k x^{\alpha_k}\right) dx\right]. \tag{24}$$

It is noted that while $\Gamma^{ME}(\boldsymbol{\lambda})$ is a convex functional in $\boldsymbol{\lambda}$, the convexity of $\Gamma_{\alpha}^{ME}(\boldsymbol{\lambda},\boldsymbol{\alpha})$ is not guaranteed. This implies that the initial choice of the parameters $\boldsymbol{\alpha}$ inside the optimization procedure of Equation 22 may significantly affect the ME distribution. Consequently, the increased flexibility of $f_{ME}(x;\boldsymbol{\lambda},\boldsymbol{\alpha})$ has the following shortcomings: (1) the fractional moments can be applied only to *positive* continuous-valued random variables and (2) the optimization procedure to determine the ME PDF is not well defined.

3 | MOMENT-BASED KD ESTIMATION

3.1 | The KDR

Let us consider a random variable X, whose PDF is $f_X(x)$ with support Ω. Assume that the event $[E] \equiv [X \in \Omega]$ is divided into n events E_1, E_2, ..., E_N mutually exclusive and collectively exhaustive, ie, $\bigcup_{i=1}^{N} E_i = E$, and $E_i \cap E_j = \varnothing$ for $i \neq j$. Let $[A] \equiv [X \leq x]$ and $[E_i] \equiv [x_{i-1} \leq X \leq x_i]$ with probability p_i, by applying the total probability theorem; the probability of the event A is as follows:

$$Prob[X \le x] = \sum_{i=1}^{N} Prob[X \le x | x_{i-1} \le X \le x_i] Prob[x_{i-1} \le X \le x_i]. \tag{25}$$

The CDF, $F_X(x)$, of X and corresponding PDF, $f_X(x)$, are thus obtained as follows:

$$F_X(x) = \sum_{i=1}^{N} p_i F_{X|E_i}(x|E_i), \tag{26}$$

$$f_X(x) = \sum_{i=1}^{N} p_i f_{X|E_i}(x|E_i) \cong \sum_{i=1}^{N} p_i f_X^K(x; x_i, h), \tag{27}$$

where $f_{X|E_i}(x|E_i)$ is the conditional PDF of X, given E_i. By approximating the target conditional PDFs with suitable KD functions (KDFs), $f_X^K(x; x_i, h) \cong f_{X|E_i}(x|E_i)$, the KDR of the target PDF is obtained, $f_{KD}(x; \boldsymbol{p}, \boldsymbol{\vartheta}) \cong f_X(x)$. In $f_{KD}(x; \boldsymbol{p}, \boldsymbol{\vartheta})$, $\boldsymbol{p}$ is an N-vector collecting the probabilities $p_1, p_2, ..., p_N$, while $\boldsymbol{\vartheta}$ is another N-vector collecting the parameters of the N KDFs. Each KDF is identified by its location parameter x_i and bandwidth h, which gives the spreading of the KDF around its location x_i. The main feature of the KDR is its extreme flexibility, see Athanassoulis and Gavriliadis,[14] since the type of KDF and its support can be chosen in accordance to the problem under consideration, ie, the centers x_i can be adaptively distributed and the bandwidth may also be variable, ie, $h = h_i$. This implies that the KDR (Equation 27) has $2N + 1$ free parameters: the N probabilities $p_1, p_2, ..., p_N$, the N centers $x_1, x_2, ..., x_N$, and the bandwidth h. The properties of the KDFs are described in detail in Athanassoulis and Gavriliadis[14] and Alibrandi and Ricciardi.[15] Here, it is underlined that when $h \to 0$, the KDF converges toward a Dirac delta function centered at x_i, ie,

$$\lim_{h \to 0} f_X^K(x; x_i, h) = \delta(x - x_i). \tag{28}$$

Thus, when $h \to 0$, the KDR of Equation 27 gives rise to the N-point distribution

$$f_X(x) \cong f_{KD}(x; \boldsymbol{p}, \boldsymbol{\vartheta}) = \sum_{i=1}^{N} p_i \delta(x - x_i). \tag{29}$$

It follows that the KDR may approximate the target PDF $f_X(x)$ as close as required. It is noted that the representation of Equation 29 has $2N$ free parameters represented by the N ascissas x_i and the N probabilities p_i.

Multiplying both sides of Equation 27 by a set of linearly independent functions $g_k(x)$, $k = 0, 1, ..., M$, with $g_0(x) = 1$, $G_0 = 1$ and integrating over the support, we obtain the following relationship between the parameters and the generalized moments of the target PDF:

$$\boldsymbol{M}\boldsymbol{p} = \boldsymbol{\mu}, \tag{30}$$

where $\boldsymbol{p}$ is a vector of order N collecting the probabilities p_i while $\boldsymbol{M}$ and $\boldsymbol{\mu}$ are defined component-wise as follows:

$$\begin{aligned} M_{ki} &= \mu_{k,i}^K = \int g_k(x) f_X^K(x; x_i, h) dx \\ \mu_k &= \int g_k(x) f_X(x) dx = E[g_k(X)] \end{aligned}, \tag{31}$$

where M_{ki} are generalized moments of order k of the ith KDF $f_X^K(x; x_i, h)$, ie, $\boldsymbol{M} = [M_{ki}]$ is a matrix of dimensions $(M + 1) \times N$ and $\boldsymbol{\mu}$ is a vector of order $M + 1$ collecting the assigned generalized moments.

3.2 | Choice of the KDFs

In the KDR (see Equation 27), a key issue is represented by the choice of the basis KDF $f_X^K(x; x_i, h)$. The support of the target PDF $f_X(x)$ can be used as a driving factor in the choice. In fact, there is no physical reason to use a Gaussian kernel (which is defined over the whole real axis) to reconstruct a PDF that is known to be defined over a finite interval $\Omega \equiv [a, b]$. It is noted that the moments of the Gaussian distribution appearing in the matrix $\boldsymbol{M}$ of Equations 30 and 31 are evaluated in a support $(-\infty, \infty)$ different from the original finite support for which the moments $\mu_0, \mu_1, ..., \mu_M$ are assigned. Likewise, there is no reason to adopt a uniformly distributed kernel when the support of $f_X(x)$ is unbounded

or semibounded. In this case, $f_{KD}(x)$ would provide a truncated distribution, which is significantly different from the target PDF, especially at the tails. Starting from this last consideration, a further driving factor in choosing the KDF is represented from the assumed tail behavior of $f_X(x)$. If the latter is sub-Gaussian, then the Gaussian kernel may be adopted. Conversely, if it is super-Gaussian, then it may be preferable to choose kernel with fat tail, eg, Laplace, lognormal, or Weibull.

For the sake of illustration, if the target distribution has unbounded support $\Omega \equiv (-\infty, \infty)$, a possible kernel is Gaussian $f_X^K(x; x_i, h) \equiv f_X^G(x; x_i, \sigma)$ whose parameters are the mean value coinciding with the center x_i and the standard deviation σ coinciding with the bandwidth h, ie,

$$f_{KD}(x) = \sum_{i=1}^{N} p_i f_X^G(x; x_i, \sigma) = \sum_{i=1}^{N} p_i \frac{1}{\sigma\sqrt{2\pi}} exp\left\{-\frac{1}{2}\left(\frac{x-x_i}{\sigma}\right)^2\right\}. \tag{32}$$

When the support is bounded, ie, $\Omega \equiv [a, b]$, a typical choice for the KDF is the beta distribution $f_X^K(x; x_i, h) \equiv f_X^B(x; \delta_i, \beta_i)$ of parameters δ_i and β_i, ie,

$$f_{KD}(x) = \sum_{i=1}^{N} p_i f_X^B(x; \delta_i, \beta_i) = \sum_{i=1}^{N} p_i \frac{\Gamma(\delta_i + \beta_i)}{\Gamma(\delta_i)\Gamma(\beta_i)} x^{\delta_i - 1}(1-x)^{\beta_i - 1}, \tag{33}$$

where

$$\delta_i(x_i, h) = \frac{x_i^2 - x_i^3 - h^2 x_i}{h^2}, \quad \beta_i(x_i, h) = \frac{(1-x_i)(x_i - x_i^2 - h^2)}{h^2}. \tag{34}$$

Finally, when the support is semibounded $\Omega \equiv [0, \infty)$, the lognormal distribution $f_X^K(x; x_i, h) \equiv f_X^{LN}(x; \lambda_i, \zeta_i)$ of parameters λ_i and ζ_i may be adopted,[15] ie,

$$f_{KD}(x) = \sum_{i=1}^{N} p_i f_X^{LN}(x; \lambda_i, \zeta_i) = \sum_{i=1}^{N} p_i \frac{1}{x\zeta_i\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\ln(x) - \lambda_i}{\zeta_i}\right)^2\right\}, \tag{35}$$

where

$$\lambda_i(x_i, h) = \ln\left(\frac{x_i^2}{\sqrt{x_i^2 + h^2}}\right), \quad \zeta_i^2(x_i, h) = \ln\left[1 + \left(\frac{h^2}{x_i^2}\right)\right] \tag{36}$$

In Yuan and Fox,[22] a gamma distribution is used as a kernel in the presence of semibounded support. It is here underlined that, provided that the number N of KDFs is high enough, the reconstruction of $f_X(x)$ in its central part is not significantly affected by the choice of the specific KDF (eg, lognormal, gamma, or Weibull). Conversely, differences may arise in the tails of the distributions if the number M of moments is low, as it is typically the case.

3.3 | Existing KD methods

Kernel density functions represent a key tool within the framework of nonparametric statistical estimation.[23,24] Kernel density estimation is a data smoothing problem and, starting from the knowledge of a finite data sample, provides inferences about the population. In the KD estimation, it is assumed that $p_i = 1/N$ in Equation 27, so only the free parameters h has to be determined. It is known that in a typical sample-based KD reconstruction, the choice of the KD does not affect the quality of the estimate provided that the sample set is large enough. In contrast, the choice of an inappropriate value of the bandwidth can lead to an oversmoothed or undersmoothed estimated PDF. Cross-validation may be used to estimate h. The main shortcoming of the classical KD estimation is its strong dependence on the sample data. Consequently, it does not have good prediction properties over the unseen data, and the reconstruction of the tails is typically poor unless a very large dataset is available.

Moment-based KD methods are used in stochastic dynamic analysis,[25-27] in turbulence and multiphase fluid mechanics. To the best of the authors' knowledge, the first moment-based KD method based on the (population) moments has been presented in Athanassoulis and Gavriliadis[14] and called KD element method (KDEM) where the Hausdorff moment problem is studied with the random variable X having a bounded support $\Omega \equiv [a, b]$. In the KDEM,

the free parameters are the N probabilities p_1, p_2, ..., p_N, which are determined through a minimization of the L_2-norm $\|\boldsymbol{Mp}-\boldsymbol{\mu}\|_2^2$ (see Equation 30), under the constraints $\boldsymbol{p} \geq 0$ and the exact reproduction of the lowest-order moments. In a least squares minimization approach, it is required that $M + 1 > N$ and the authors in previous study[14] find that $M = 10 - 20$ moments are sufficient for an accurate reconstruction. They also show that the maximum number of (population) moments is constrained by the computer arithmetic, namely, $M \leq 10$ and $M \leq 20$ in single and double precision, respectively. An important shortcoming of the KDEM is that only few lower-order moments are exactly reproduced by the KD PDF.

In some applications, PDFs need to be reconstructed and integrated repeatedly, so that there is interest in minimizing the number of points required for the reconstruction. To this aim, a widely used closure for the moment transport equation is the quadrature method of moments (QMOM).[28] In the QMOM, the representation in Equation 29 is adopted, and the $2N$ free parameters p_i and x_i are described by a set of $M = 2N - 1$ moments μ_1, μ_2, ..., μ_{2N-1} (in addition to the zeroth-order moment $\mu_0 = 1$).[29] The moment-inversion algorithm used in the QMOM is accurate for $N \leq 10$ (and $M \leq 19$). To capture the response at the boundary, a continuous distribution is needed. To this aim, in Chalons et al[30] and Yuan et al,[31] the extended QMOM (EQMOM) is proposed, where the representation in Equation 27 is adopted. In such case, the $2N + 1$ free parameters (p_i, x_i where $i = 1, 2, ..., N$ and h) are determined from the knowledge of $M = 2N$ moments. To minimize the numerical difficulties for moment inversion, the adaptive Wheeler algorithm is used.[32,33] A reasonable accuracy is obtained only for $N \leq 5 - 10$ (and $M \leq 10 - 20$) where the EQMOM outperforms both QMOM and KDEM, and numerical examples have demonstrated its good performance in several applications. However, the focus of the EQMOM is devoted mostly to the prediction of the moment up to order $2N + 1$, and not to the reconstruction of the target PDF.

In this paper, following Jaynes,[18] we focus on the evaluation of the less "subjective" reconstruction of the distribution of a random variable, including its tails, from the knowledge of a sample data. This cannot be obtained through the existing KD methods, because the sample-based KD does not have good prediction capabilities over the unseen data, while with the moment-based KD methods, the evaluation of the tails is affected by the choice of the KDF. Moreover, as discussed above, the moment-based KD methods assume that the first $M = 10 - 20$ population moments are known. Unfortunately, with most sample data, only the first $M = 2 - 4$ sample moments can be determined with reasonable accuracy.

4 | KDME METHOD WITH GENERALIZED MOMENTS

The support of the random variable X can be bounded, $\Omega \equiv [a, b]$, semibounded, $\Omega \equiv [0, \infty)$, or unbounded, $\Omega \equiv (-\infty, \infty)$. For the KDR, an important issue stems from the selection of the window of analysis $\widehat{\Omega}_x \subseteq \Omega$, to which the centers x_i of the KDFs belong. If X has $\Omega \equiv [a, b]$, then clearly, $\widehat{\Omega}_x \equiv \Omega$. Conversely, if X has semibounded or unbounded support, then $\widehat{\Omega}_x \equiv [x_{\min}, x_{\max}]$ is selected and the approximated PDF $f_{KD}(x;\boldsymbol{p})$ is evaluated. To satisfy the normalization condition, the target PDF should be zero toward infinity, ie, $\lim_{x\to\infty} f_X(x) = 0$ and $\lim_{x\to\pm\infty} f_X(x) = 0$ for variables with semibounded or unbounded support, respectively. Consequently, at the boundary of $\widehat{\Omega}_x$, $f_{KD}(x;\boldsymbol{p})$ has to be lower than a chosen tolerance level $\varepsilon > 0$. If this is not the case, the window of analysis $\widehat{\Omega}_x$ is enlarged as much as needed. In this way, the problem is reduced to the estimation of the distribution of a random variable over the bounded support $\widehat{\Omega}_x$.

A coordinate transformation, $z = (x - x_{\min})/(x_{\max} - x_{\min})$, is developed where Z is a continuous-valued random variable whose support is $\Omega_z \equiv [0, 1]$. The KDR described above is applied to Z to determine its PDF $f_Z(z)$. The location parameters z_i, $i = 1, 2, ..., N$, are N points belonging to the domain, and they can be chosen in several different manners. For simplicity, a constant step $\Delta z \equiv \Delta z_i = z_{i+1} - z_i$, $i = 1, 2, ..., N - 1$, is chosen. Similarly, we choose $h = \rho\Delta x$ with $0 < \rho < \Delta x$, such that when $N \to \infty$, $h \to 0$. It is noted that a reduced number of KDFs (eg, $N = 5 - 10$) implicitly assume the shape of the KD PDF, especially with reference to the tails. To obtain the most objective distribution given the available information, N is chosen as a very high number (eg, $N \geq 100$) so that $f_Z^K(z; z_i, h) \to \delta(z - z_i)$ and the representation in Equation 29 may be assumed valid. The KDR provides the generalized constraints $\boldsymbol{Mp} = \boldsymbol{\mu}$, see Equation 30, where $\boldsymbol{p}$ is a vector of order N collecting the probabilities $p_i = Prob[z_{i-1} \leq Z \leq z_i] \equiv Prob[x_{i-1} \leq X \leq x_i]$ while $\boldsymbol{M}$ and $\boldsymbol{\mu}$ are defined component-wise as follows:

$$\begin{aligned} M_{ki} &= \mu_{k,i}^K = \int_{\Omega_z} g_k(z)\delta(x - x_i)dz = g_k(z_i) \\ \mu_k &= \int_{\Omega_z} g_k(z) f_Z(z) dz = E[g_k(Z)]. \end{aligned} \tag{37}$$

Let us consider the discrete random variable $\widehat{Z}$ defined in $\Omega_z \equiv [0,1]$ with probability distribution given by p_i evaluated at the centers z_i, $i = 1, 2, ..., N$, and $\sum_{i=1}^{n} p_i = 1$. The generalized moments of $\widehat{Z}$ are given by $E(g_k) \equiv \mu_k = \sum_{i=1}^{n} g_k(z_i) p_i$. The ME probability distribution of $\widehat{Z}$ is

$$p_i^{ME}(\lambda_1, \lambda_2, ..., \lambda_M) = N_0(\lambda_1, \lambda_2, ..., \lambda_M) \exp\left(- \sum_{k=1}^{M} \lambda_k g_k(z_i) \right), \quad i = 1, ..., N. \tag{38}$$

This ME probability distribution is given by the solution of

$$\begin{cases} \max\limits_{\boldsymbol{p}} H(\boldsymbol{p}) \\ \sum\limits_{i=1}^{N} p_i = 1 \\ \sum\limits_{i=1}^{N} g_k(z_i) p_i = \mu_k, \quad k = 1, 2, ..., M \end{cases} \equiv \quad \left\{ \min_{\boldsymbol{\lambda}} \Gamma_{ME}(\boldsymbol{\lambda}) = \lambda_0(\boldsymbol{\lambda}) + \sum_{k=1}^{M} \lambda_k \mu_k. \right. \tag{39}$$

Once $p_i^{ME}(\boldsymbol{\lambda})$ are determined, the reconstruction of the target PDF $f_Z(z)$ is developed through the KDME PDF $f_{KDME}(z)$ (Equation 27). Since N ≥ 100 and only the region inside the bounded domain Ω_z is considered, the choice of the KDF (eg, Gaussian, lognormal, Weibull, or beta) does not significantly affect the approximation.

Algorithm 1. *Let* X *be a random variable with support Ω; we seek to recover its unknown PDF* $f_X(x)$.

1. *Choose a support* $\widehat{\Omega}_x \subseteq \Omega$, *such that outside* $\widehat{\Omega}_x \equiv [x_{\min}, x_{\max}]$, *the values of* $f_X(x)$ *may be negligible.*
2. *Develop the coordinate transformation* $z = (x - x_{min})/(x_{max} - x_{min})$, *by introducing the random variable* Z *defined in* $\Omega_z \equiv [0, 1]$ *where* $\mu_1, \mu_2, ..., \mu_M$ *are the first* M *generalized moments of* Z.
3. *Apply the KDR to the random variable* Z. *In all the numerical examples, the considered number of kernel densities is high,* N ≥ *100, such that Equation 29 for the KDR is considered valid.*
4. *Thus, a discrete random variable* $\widehat{Z}$ *is defined in* Ω_z, *whose probability distribution is characterized by the probabilities* p_i, i = *1, 2, ...,* N, *determined from the ME solution* $p_i^{ME} = p_i^{ME}(\lambda_1, \lambda_2, ..., \lambda_M)$ *(Equation 38) by maximizing the Shannon's entropy subject to the available information or, which is the same, by minimizing the free convex functional* $\Gamma^{ME}(\boldsymbol{\lambda})$. *It is noted that, since* $\widehat{Z}$ *is a discrete-valued random variable, the minimization of* $\Gamma^{ME}(\boldsymbol{\lambda})$ *does not require the numerical evaluation of complicated integrals.*
5. *Choose the KDF that best fits the target PDF* $f_z(z)$ *(Equations 32-36). From the knowledge of the probabilities* $p_i^{ME} \cong p_i$, *determine the KDME PDF* $f_{KDME}(z) \cong f_Z(z)$.
6. *Obtain KDME approximation of* $f_X(x)$ *by the inverse coordinate transformation from* z *to* x.
7. *Check the goodness of the window of analysis* $\widehat{\Omega}_x$, *and, if necessary, enlarge/reduce it and go back to step 2; otherwise, stop.*

5 | GKDMEM WITH FRACTIONAL MOMENTS

In this section, the GKDMEM is applied by choosing a suitable class of fractional moments as generalized ones. The fractional moment is defined as $\mu_k = E[Z^{\alpha_k}] \equiv E\left[\widehat{Z}^{\alpha_k}\right]$, where α_k is a real number, including integer and fractional numbers. The corresponding generalized functions becomes $g_k(z) = z^{\alpha_k}$. If the powers are integer numbers, then $g_1(z) = z$, $g_2(z) = z^2$, ..., $g_M(z) = z^M$, and the functions $g_k(z)$ are linearly independent, which is the classical approach with power moments. As a generalization to the fractional moments, we choose $\alpha_k = k\Delta\alpha$, such that the functions $g_1(z) = z^{\Delta\alpha}$, $g_2(z) = z^{2\Delta\alpha}$, ..., $g_M(z) = z^{M\Delta\alpha}$ are also linearly independent. This mathematical structure has some important properties.

First, it is shown that if an infinite sequence of classical power moments $\mu_k = E\left[\widehat{Z}^k\right]$ or of fractional moments $\mu_k = E\left[\widehat{Z}^{\alpha_k}\right] = E\left[\widehat{Z}^{k\Delta\alpha}\right]$ determines p_i uniquely, then the ME approximation p_i^{ME}, constrained by the same moment sequence $\{\mu_k\}_{k=0}^{M}$, converges in entropy to p_i, ie, for $M \to \infty$, then $H[p_i^{ME}] \to H[p_i]$.[20,34-36] Second, the convergence in entropy entails convergence in the L_1-norm and thus in distribution, ie, $p_i^{ME} \to p_i$. In view of Equation 29, for $N \to \infty$,

the convergence in p_i implies the convergence of Z and consequently of X, $f_{KDME}(x) \to f_X(x)$. Third, for M finite, the existence and uniqueness of the ME solution $p_i^{ME}(\lambda_1, \lambda_2, ..., \lambda_M)$ is guaranteed, and from Equation 38 by setting $g_k(z) = z^{\alpha_k}$, with $\alpha_0 = 0$, we obtain the following:

$$p_i^{ME}(\boldsymbol{\lambda}, \boldsymbol{\alpha}) = \exp\left(-\lambda_0 - \sum_{k=1}^{M} \lambda_k z_i^{\alpha_k}\right) = \exp\left(-\sum_{k=0}^{M} \lambda_k z_i^{\alpha_k}\right), \quad i = 1, ..., N. \tag{40}$$

From a practical point of view, the number of moments is limited, for both cases, fractional and integers,[37] because of numerical instabilities. Thus, the choices of how many and which powers to consider are important. Let us consider now the KL divergence between the true but unknown distribution p_i and the ME solution p_i^{ME}, defined as follows:

$$D(\boldsymbol{p}, \boldsymbol{p}^{ME}) = \sum_{i=1}^{N} p_i \, log \, \frac{p_i}{p_i^{ME}} = H[\boldsymbol{p}^{ME}] - H[\boldsymbol{p}] \geq 0. \tag{41}$$

In Equation 41, the definition of Shannon's entropy (Equation 1) is used. In view of Equation 40, $\boldsymbol{p}^{ME} = \boldsymbol{p}^{ME}(\boldsymbol{\lambda}, \boldsymbol{\alpha})$. Thus, the minimum divergence between the 2 distributions is given by the distribution of ME $\boldsymbol{p}^{ME}(\boldsymbol{\lambda}^*, \boldsymbol{\alpha}^*)$, determined from a nested optimization procedure:

$$\min_{\boldsymbol{p}, \boldsymbol{p}^{ME}} D(\boldsymbol{p}, \boldsymbol{p}^{ME}) = \min_{\boldsymbol{\lambda}, \boldsymbol{\alpha}} H[\boldsymbol{p}^{ME}(\boldsymbol{\lambda}, \boldsymbol{\alpha})] = \min_{\alpha_1, ..., \alpha_M} \left\{ \min_{\lambda_1, ..., \lambda_M \ \alpha} \Gamma_{\alpha}^{ME}(\boldsymbol{\lambda}, \boldsymbol{\alpha}) \right\}, \tag{42}$$

where

$$\Gamma_{\alpha}^{ME}(\lambda_1, ..., \lambda_M, \alpha_1, ..., \alpha_M) = \log\left\{ \sum_{i=1}^{N} \exp\left(-\sum_{k=1}^{M} \lambda_k z_i^{\alpha_k}\right) \right\} + \sum_{k=1}^{M} \lambda_k E[Z^{\alpha_k}]. \tag{43}$$

This formulation, originally proposed in Novi Inverardi and Tagliani,[11] is involved because of its nonuniqueness of the solution with respect to $\boldsymbol{\alpha}$, ie, different initial conditions of $\boldsymbol{\alpha}$ may give different ME solutions. In this paper, starting from considerations in previous studies,[12,38,39] we present an effective computational procedure for a unique ME solution given sample data.

Using the definition in Equation 40 and integrating by parts the fractional moments, the internal optimization procedure of Equation 42 may be substituted by the solution of a simple linear system of M equations in the M unknowns $\lambda_1, \lambda_2, ..., \lambda_M$:

$$\boldsymbol{\Theta}(\boldsymbol{\alpha})\boldsymbol{\lambda} = \boldsymbol{\rho}(\boldsymbol{\alpha}), \tag{44}$$

where

$$\Theta_{kj}(\boldsymbol{\alpha}) = \alpha_k E[Z^{\alpha_k + \alpha_j}], \quad \rho_j(\boldsymbol{\alpha}) = (\alpha_j + 1) E[Z^{\alpha_j}], \tag{45}$$

with $k = 1, 2, ..., M$ and $j = 0, 1, ..., M - 1$. Thus, for given values of $\alpha_1, \alpha_2, ..., \alpha_M$, the Lagrange multipliers $\lambda_1, \lambda_2, ..., \lambda_M$ are given by the solution of the linear system in Equations 44 and 45.

A crucial point in the definition of $g_k(z) = z^{\alpha_k} = z^{k\Delta}\alpha$ is the choice of $\alpha_{\max}$. It is underlined once again that when $N \to \infty$, in view of Equation 29, $f_{KDME}(z_i) = p_i^{ME}$. The approach presented here implies that the GKDMEM with fractional moments should give the exact solution for the distributions that may be expressed as an exponential of polynomials with real powers. If, for example, the target PDF $f_X(x)$ is Gaussian, then the KDME approach should give $f_{KDME}(z) = N_0 \exp(-\lambda_1 z - \lambda_2 z^2)$. If it is reasonable to assume that the tail of $f_X(x)$ is fatter than Gaussian, then one may choose $\alpha_{\max} = 2$; otherwise, $\alpha_{\max} = 4$ can be selected. Thus, a simple choice is to define a sequence of Q-powers $\alpha_q = q\Delta\alpha$, with $q = 1, 2, \ldots, Q$ and $Q \geq M$ such that the sequence also contains integer numbers. Let us consider the collection $\mathbb{Q}_M$ of M real powers taken from the extended set $\alpha_1, \alpha_2, ..., \alpha_Q$. The cardinality of this set is $N_{Q,M} = \binom{Q}{M} = \frac{Q!}{M!(Q-M)!}$. For illustration, assume $M = 2$ and $Q = 4$, with $\alpha_1 = 0.5$, $\alpha_2 = 1.0$, $\alpha_3 = 1.5$, $\alpha_4 = 2.0$. In this case, $\mathbb{Q}_2 \equiv \left\{ \binom{0.5}{1.0}, \binom{0.5}{1.5}, \binom{0.5}{2.0}, \binom{1.0}{1.5}, \binom{1.0}{2.0}, \binom{1.5}{2.0} \right\}$ with $N_{4,2} = \binom{4}{2} = 6$. Define $\boldsymbol{\alpha}^{(r)}$ as the rth element of the collection $\mathbb{Q}_M$, $r = 1, 2, ..., N_{Q,\ M}$. From Equations 44 and 45, $\boldsymbol{\lambda}^{(r)} = \boldsymbol{\lambda}(\boldsymbol{\alpha}^{(r)})$ is determined, and in turn from Equations 40 and 1, one obtains

$$H^{(r)} = H\left[\boldsymbol{p}^{ME}\left(\boldsymbol{\alpha}^{(r)}\right)\right] \equiv H\left[\boldsymbol{\alpha}^{(r)}\right] = \sum_{k=0}^{M} \lambda_k\left(\boldsymbol{\alpha}^{(r)}\right) E\left[Z^{\alpha_k^{(r)}}\right]. \tag{46}$$

Therefore, the nested optimization procedure (Equations 42 and 43) is reduced to

$$H_{Q,M} = \min_{\boldsymbol{\alpha}} H\left[\boldsymbol{p}^{ME}(\boldsymbol{\alpha})\right] = \min\left\{H\left[\boldsymbol{\alpha}^{(1)}\right], ..., H\left[\boldsymbol{\alpha}^{(r)}\right]\right\}, r = 1, ..., N_{Q,M}. \tag{47}$$

This reduces to determining the solution of $N_{Q,\,M}$ linear systems in M unknowns. It is noted that the presented procedure is effective and robust, but it gives the optimal ME solution within the class of fractional powers belonging to the collection $\mathbb{Q}_M$. However, this is not a limitation, because it is enough to reduce $\Delta\alpha$ to obtain an extended class $\mathbb{Q}'_M \supseteq \mathbb{Q}_M$ whose solution of Equation 47 is $H'_{Q,M} \leq H_{Q,M}$.

Algorithm 2. *The GKDMEM with fractional moments is summarized as the following algorithm:*

1. *For a given number* M *of fractional moments, choose* $\Delta\alpha$ *and* α_{max}*, to define a set of* Q *fractional powers* α_1*,* α_2*, ...,* α_Q*, with* $Q \geq M$*. A good choice is* $\alpha_{max} = 2$ *or 4, while* $\Delta\alpha$ *should be selected such that the fractional powers also include the integer ones.*
2. *Construct the collection* $\mathbb{Q}_M$ *of* M *real powers taken from the extended set* α_1*,* α_2*, ...,* α_Q*.*
3. *Select the* r*th element* $\boldsymbol{\alpha}^{(r)}$ *of the collection* $\mathbb{Q}_M$*,* $r = 1, 2, ..., N_{Q,\,M}$*.*
4. *The solution of the linear system in Equation 44 gives the Lagrange multipliers* $\boldsymbol{\lambda}^{(r)} = \boldsymbol{\lambda}(\boldsymbol{\alpha}^{(r)})$*.*
5. *Determine the* r*th entropy* $H^{(r)} \equiv H[\boldsymbol{\alpha}^{(r)}]$ *through Equation 46.*
6. *Repeat steps 3 to 5 for all the* $N_{Q,\,M}$ *elements of the collection* $\mathbb{Q}_M$*.*
7. *The ME solution is the one minimizing the value of* $H^{(r)}$ *(Equation 47).*

6 | KDMEM FOR GIVEN SAMPLE DATA

In practical problems, a sample of data is available, where the known quantities are not the population moments μ_k, but the sample moments m_k. This provides further challenges to the application of the ME, as discussed above. To this aim, it is convenient to interpret the KDME PDF (Equations 27 and 32-36) as a probabilistic model, whose parameters for fractional moments are obtained from $\boldsymbol{\alpha}$. The KL divergence between the target PDF $f_Z(z)$ and its KDME approximation $f_{KDME}(z, \boldsymbol{\alpha})$ is as follows:

$$D(f_Z, f_{KDME}) = \int_{\Omega_z} f_Z(z) log\left(\frac{f_Z(z)}{f_{KDME}(z, \boldsymbol{\alpha})}\right) dz = C - \frac{1}{n_s}\sum_{j=1}^{n_s} log\left[f_{KDME}\left(z_j, \boldsymbol{\alpha}\right)\right], \tag{48}$$

where $C = \int_{\Omega_z} f_Z \, log\, f_Z \, dz$, while n_s is the number of sample data. The minimum divergence between the 2 distributions coincides with the maximum logarithmic likelihood of $f_{KDME}(z, \boldsymbol{\alpha}) = \sum_{i=1}^{N} p_i(\boldsymbol{\alpha}) f_Z^K(z; x_i, h)$. The concept behind the maximum likelihood estimation is to estimate the parameters of the distribution that maximize the probability of occurrence of the observed data. However, the maximum likelihood estimation produces biased estimates of the true parameters. Thus, Akaike[16] suggested to maximize an unbiased estimation of the likelihood function, ie,

$$L(M, \boldsymbol{\alpha}) = \frac{1}{n_s}\sum_{j=1}^{n_s} log\left[\sum_{i=1}^{N} p_i(\boldsymbol{\alpha}) f_Z^K\left(z_j; x_i, h\right)\right] - \frac{M}{n_s}, \tag{49}$$

where the "bias" term M/n_s is a penalty term discouraging the model complexity. Therefore, for a given M, the optimal KDME PDF is as follows:

$$D(M) = \min_{\boldsymbol{\alpha}} D(f_Z, f_{KDME}) = \max_{\boldsymbol{\alpha}} L(M, \boldsymbol{\alpha}). \tag{50}$$

In this way, an optimal divergence $D(M)$ can be determined from Equation 50 while the optimal number of moments for the given sample data minimizes $D(M)$. Thus, in the presence of a sample of size n_s, the choice of the optimal number of moments is obtained from the following algorithm.

Algorithm 3.

1. *For a given number* M *of fractional moments, choose* $\Delta\alpha$ *and* α_{max}*, to define a set of* Q *fractional powers* α_1*,* α_2*, ...,* α_Q*, with* $Q \geq M$.
2. *Construct the collection* $\mathbb{Q}_M$ *of* M *real powers taken from the extended set* α_1*,* α_2*, ...,* α_Q *where*
   a. *The* r*th element* $\boldsymbol{\alpha}^{(r)}$ *of the collection* $\mathbb{Q}_M$*,* $r = 1, 2, ..., N_{Q,\,M}$*, is selected;*
   b. *The solution of the linear system in Equation 44 gives the Lagrange multipliers* $\boldsymbol{\lambda}^{(r)} = \boldsymbol{\lambda}(\boldsymbol{\alpha}^{(r)})$*;*
   c. *The* r*th KDME distribution* $p_{i^{(r)}} \equiv p_i^{ME}\left[\boldsymbol{\alpha}^{(r)}\right]$ *is determined through Equation 40;*
   d. *The* r*th unbiased likelihood* $L^{(r)}(M) \equiv L[M, \boldsymbol{\alpha}^{(r)}]$ *is determined through Equation 49;*
   e. *Repeat steps 2a to 2d for all the* $N_{Q,\,M}$ *elements of the collection* $\mathbb{Q}_M$*;*
   f. *The optimal KDME PDF for fixed* M *maximizes* $L^{(r)}(M)$ *(Equation 50);*
   g. *The divergence* $D(M)$ *(Equation 50) between the target PDF* $f_Z(z)$ *and the optimal KDME PDF is determined. It is noted that the coefficients* $\boldsymbol{\alpha}^{ME}$ *giving rise to the minimum entropy* $H[\mathbf{p}^{ME}(\boldsymbol{\alpha})]$ *of Equation 47 represent the optimal parameters for the divergence* $D(M)$ *of Equation 50.*
3. *Repeat steps 1 to 2 for different values of* M.
4. *The optimal KDME PDF for the given sample minimizes* $D(M)$.

7 | CREDIBLE BOUNDS FOR GIVEN SAMPLE DATA

In the presence of a given sample data, especially if it is based on a limited amount of data, it is of interest to investigate the error incurred by the KDME PDF. In the presence of one sample of data of small size, an effective tool is represented by the bootstrap resampling.[40] Assume a known original sample of n_s data $\widehat{\boldsymbol{z}} = \{z_1,\ z_2,\ \ldots,\ z_{n_s}\}$, drawn from an unknown cumulative distribution $F_Z(z)$. Determine the empirical estimate $\widehat{F}_Z(z)$ of $F_Z(z)$ from the sample $\widehat{\boldsymbol{z}}$. $\widehat{F}_Z(z)$ is defined as the discrete distribution with probability $1/n_s$ on each value z_i. A bootstrap sample $\widehat{\boldsymbol{z}}_B$ is a random sample of size n_s drawn from $\widehat{F}_Z(z)$ with replacement. The elements of $\widehat{\boldsymbol{z}}_B$ are the same as those of the original data set, but repetitions may occur, ie, in $\widehat{\boldsymbol{z}}_B$, some elements may appear only once, some may appear 2 or more times, and other may not appear. For illustration, if the original sample is $\widehat{\boldsymbol{z}} \equiv \{1,\ 2,\ 5,\ 8,\ 4\}$, some possible bootstrap samples are $\widehat{\boldsymbol{z}}_B^{(1)} \equiv \{8,\ 2,\ 2,\ 1,\ 4\}$, $\boldsymbol{z}_B^{(2)} \equiv \{5,\ 5,\ 1,\ 4,\ 4\}$, and so on. Following Algorithm 3 described above, for the sth bootstrap sample $\widehat{\boldsymbol{z}}_B^{(s)}$, $s = 1,\ 2,\ \ldots,\ S$, the KDME PDF is evaluated as $f_{KDME}\left[z, \boldsymbol{\alpha}^{(s)}\right] = \sum_{i=1}^{N} p_i^{(s)}(\boldsymbol{\alpha}) f_Z^K(z; x_i, h)$, where $p_i^{(s)}(\boldsymbol{\alpha}) = p_i\left[\boldsymbol{\alpha}^{(s)}\right]$. In this way, S values of p_i are determined and it is possible to define for each p_i the corresponding bootstrap distribution. Let us define $\boldsymbol{p}_{LB}(\boldsymbol{\alpha})$ and $\boldsymbol{p}_{UB}(\boldsymbol{\alpha})$ as credible lower and upper bounds, obtained by respectively considering 2 percentiles q_{LB} and q_{UB} of the boostrap distributions of p_i. Accordingly, the credible bounds of p_i determine the credible bounds of the KDME PDF as follows:

$$\sum_{i=1}^{N} p_{i,LB}(\boldsymbol{\alpha}) f_Z^K(z; x_i, h) \leq \sum_{i=1}^{N} p_i(\boldsymbol{\alpha}) f_Z^K(z; x_i, h) \leq \sum_{i=1}^{N} p_{i,UB}(\boldsymbol{\alpha}) f_Z^K(z; x_i, h). \tag{51}$$

8 | NUMERICAL APPLICATIONS

8.1 | Duffing oscillator

Consider a Duffing oscillator defined by the following nonlinear differential equation:

$$m\ddot{X}(t) + c\dot{X}(t) + k\left[aX(t) + \varepsilon X^3(t)\right] = W(t), \tag{52}$$

where $m = 1000$ kg, $c = 100(2\pi)$ N. s/m , and $k = 100(2\pi)^2$ N/m are the mass, damping, and stiffness coefficients, corresponding to a natural frequency $\omega_0 = (2\pi)$ rad/s and damping ratio $\zeta_0 = 5\%$. The coefficients of linear and cubic nonlinearity are both taken as a (different dimensionless values) and $\epsilon = 10\ \mathrm{m}^{-2}$. The stochastic input is defined as a white noise process $W(t)$ of intensity S_W. The PDF of the response in a stationary state is known in a closed form.[41] After transformation toward the z space, the target PDF has the following form:

$$f_Z(z) = N_0 \exp\left(-\delta_1 z - \delta_2 z^2 - \delta_3 z^3 - \delta_4 z^4\right), \tag{53}$$

where N_0 is a normalization constant. The above form is used as a benchmark to compare with the approximation given by the GKDMEM. In the following examples, the parameters of the fractional moments, as discussed above, are $\alpha_{\max} = 4$, with $\Delta\alpha = 0.25$, ie, $Q = 16$, and $M = 1, 2, 3, 4$ where it is necessary to respectively solve $\binom{16}{1} = 16$, $\binom{16}{2} = 120$, $\binom{16}{3} = 560$, and $\binom{16}{4} = 1820$ linear systems in the different M unknowns, with a total of 2516 linear systems of small size.

8.1.1 | Population moments

As a first example, we choose $a = 1$ and $S_W = 1\mathrm{m}^2/\mathrm{s}^3$. In this theoretical example, the fractional moments are evaluated from the knowledge of the target PDF, known in closed form. A window of analysis $\widehat{\Omega}_x \equiv [-1, 1]$ is chosen. The ME solution (see Equation 40) has the powers $\alpha_1 = 1$, $\alpha_2 = 2$, $\alpha_3 = 3$, and $\alpha_4 = 4$, which implies

$$p_i^{ME} = N_0 \exp\left(-\lambda_1 z_i - \lambda_2 z_i^2 - \lambda_3 z_i^3 - \lambda_4 z_i^4\right), \quad i = 1, ..., N. \tag{54}$$

This has the same mathematical structure of the exact PDF (Equation 53). In Figure 1, the target distribution and the KDME approximation are shown together with the Gaussian distribution to underline the degree of non-Gaussianity of the target distribution. In Figure 1A, the PDFs f are represented, while in Figure 1B, we show the corresponding probability of exceeding P, which is equal to the complementary CDF, ie, $F = 1 - P$. As expected, the GKDMEM gives the exact solution because the classical ME distribution with the first 4 power moments gives the exact PDF. It is noted that (1) the GKDMEM only requires the solution of a set of linear system of equations without the need for any numerical evaluation of integrals, (2) in the proposed procedure, the fractional moments converge to the integer ones to give the exact solution, and (3) the classical ME with fractional moments cannot be used for distributions with unbounded domains.

To show the flexibility of the GKDME approach, we consider another Duffing oscillator with $a = -1$ and $S_W = 0.1\mathrm{m}^2/\mathrm{s}^3$. In this case, the response of the oscillator is markedly bimodal. The ME solution is again given by Equation 54 giving an excellent fit, as shown in Figure 2.

8.1.2 | Sample moments

The analysis with population moments in the previous subsection is adopted to benchmark the KDME PDF. In general, we have samples of data, from which sample moments are derived. In this subsection, we consider the unimodal Duffing oscillator, characterized by $a = 1$ and $S_W = 1\mathrm{m}^2/\mathrm{s}^3$, and sample moments, relative to the response samples of different sizes, determined by modeling the stochastic input in terms of normal standard random variables,[42-44] ie,

$$F(t, \boldsymbol{u}) = \sum_{k=1}^{n_\omega} \sqrt{G(\omega_k)\Delta\omega}\left[\cos(\omega_k t)u_k^c + \sin(\omega_k t)u_k^s\right] = \sum_{k=1}^{n_\omega} s_k^c(t)u_k^c + s_k^s(t)u_k^s = \boldsymbol{s}(t)\cdot\boldsymbol{u}, \tag{55}$$

where $G(\omega_k)$ is the one-sided power spectral density (PSD) of the input, n_ω is the number of harmonic components, u_k^c and u_k^s, $k = 1, 2, ..., n_\omega$ are standard normal random variables, collected in the vector $\boldsymbol{u} = \{\boldsymbol{u}^c \quad \boldsymbol{u}^s\}^T$ of order $2n_\omega$, and the vector $\boldsymbol{s}(t) = \{\boldsymbol{s}^c(t) \quad \boldsymbol{s}^s(t)\}^T$ collects the corresponding deterministic shape functions $s_k^c(t) = \sigma_k \cos(\omega_k t)$ and $s_k^s(t) = \sigma_k \sin(\omega_k t)$, where $\sigma_k = \sqrt{G(\omega_k)\Delta\omega}$. In the case of the considered stationary white

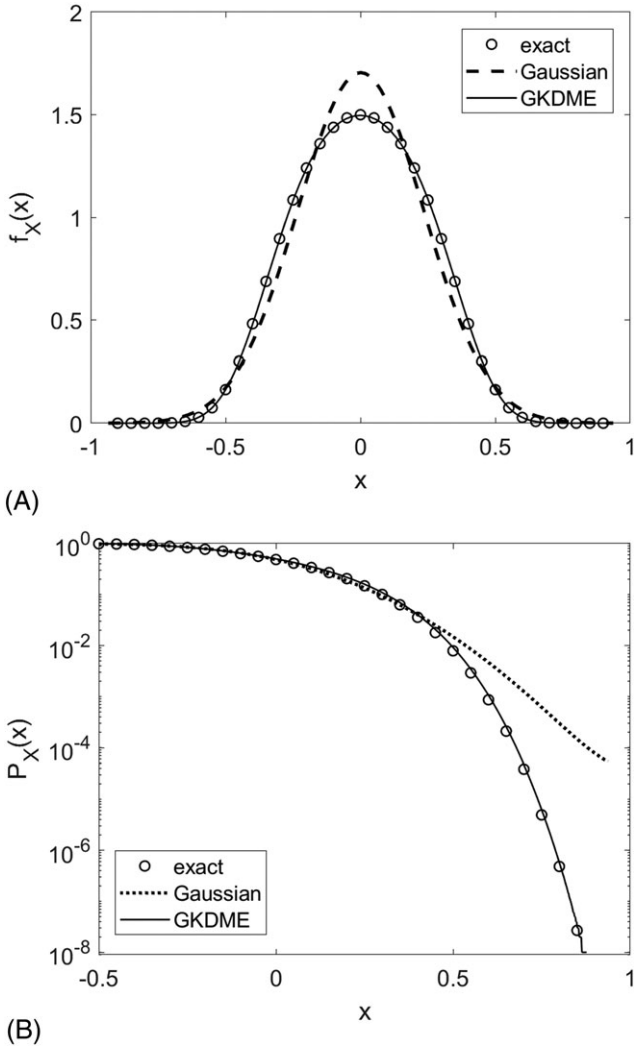

FIGURE 1 Response of a Duffing oscillator. Comparison with generalized kernel density maximum entropy method (GKDME) and Gaussian approximations using population moments. A, Probability density function. B, Probability of exceeding in semi-logarithmic scale

noise, let $G(\omega_k) = 2S_W = 2\text{m}^2/\text{s}^3$. The oscillator reaches the stationarity at time $t = 10$ seconds, and the frequency step in Equation 55 is taken as $\Delta\omega = (2\pi)/t = 0.628\text{rad/s}$.

Several sample sizes have been considered where the fractional moments of some of these samples are summarized in Table 1, while the corresponding results are represented in Figure 3. As expected, when the number of sample size increases, the approximation of the KDME PDF improves. Figure 3B shows that with only 2000 samples, the GKDMEM gives an excellent approximation of small probabilities, $Prob[X \geq 0.86] = 10^{-8}$. A crude Monte Carlo simulation (MCS) would requires more than 10^{10} samples for such low probability. From inspection of Table 1, it is noted that the number and powers of the fractional moments are related to the size of the sample. This property shows that the GKDMEM gives solution with a good trade-off between bias and variance.

8.2 | Seismic fragility functions of a reinforced concrete frame

As a practical example of the GKDMEM, we consider an application of seismic reliability analysis for a one-bay 4-story reinforced concrete frame modeled using OpenSees.[45] The column height and beam length are $h = 3$ m and $L = 5$ m, respectively. All the sections are rectangular 300 × 500 mm, modeled using displacement-based beam-column elements, with fiber-discretized sections. Column sections contain 8 ϕ20-mm reinforcing bars, while the beam sections are symmetrically reinforced with 4 ϕ20-mm bars at the top and bottom sides. Thus, the longitudinal reinforcement of all beams

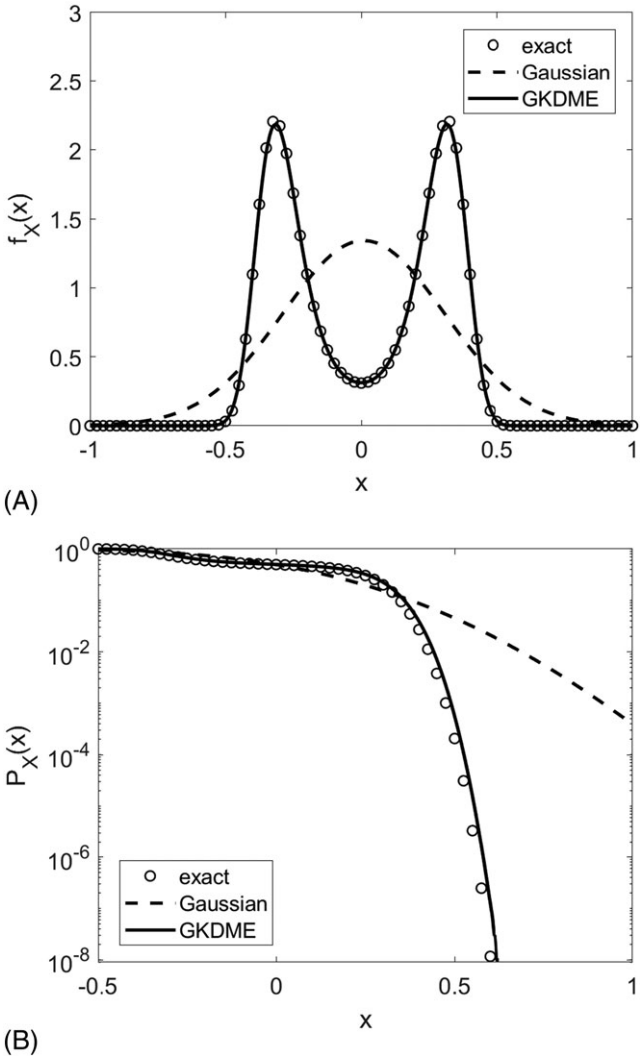

FIGURE 2 Response of a bimodal Duffing oscillator. Comparison with generalized kernel density maximum entropy method (GKDME) and Gaussian approximations using population moments. A, Probability density function. B, Probability of exceeding in semi-logarithmic scale

TABLE 1 Fractional moments for a Duffing oscillator, by using samples of different sizes

n_s	α_1	α_2	α_3	α_4
50	0.25	1.75	...	...
100	0.50	0.75	1	2.75
1000	0.75	3.25	3.75	4
$\rightarrow\infty$	1	2	3	4

and columns is 1.67%. The compressive strength of the concrete is 35 MPa, and it is modeled using *Concrete01* model in OpenSees.[45] The reinforcing bars are modeled with *Steel01* model in OpenSees,[45] whereas the yield strength of the steel is 420 MPa, the elastic modulus is 200 000 MPa, and the strain hardening ratio is $b = 0.01$. The mass of the structure includes the self-weight of the building and additional live load distributed on the beams in accordance with Eurocode 1[46] for a total of 15 kN/m. The frame is shown in Figure 4.

The seismic hazard is modeled through the response spectrum of Eurocode 8 (EC8),[47] whose parameters take into account the site conditions, while the intensity of the seismic action is given in terms of the peak ground acceleration

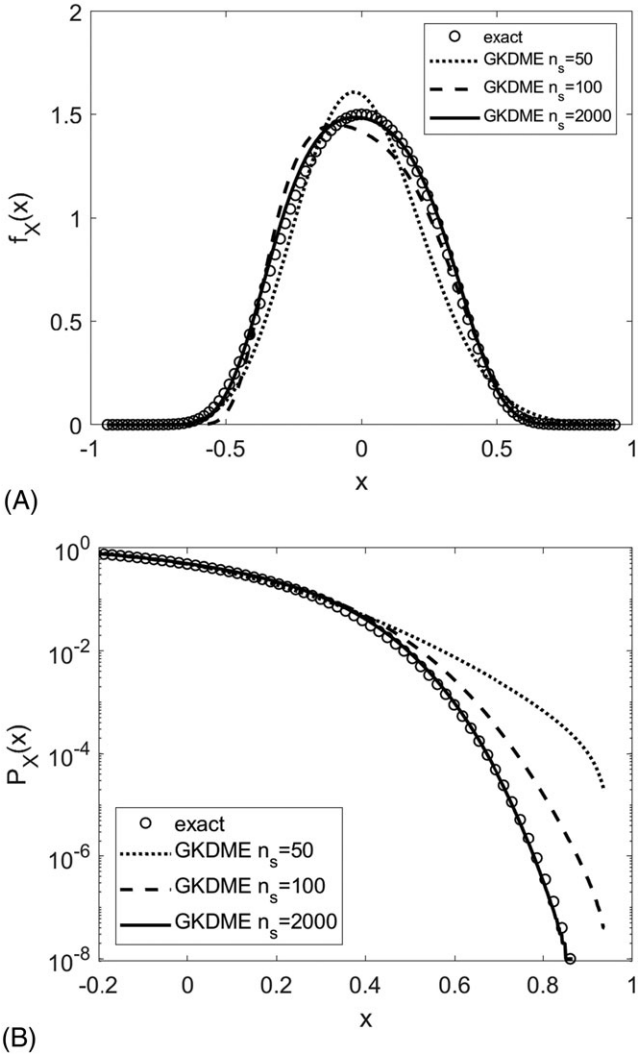

FIGURE 3 Response of a Duffing oscillator. Comparison with generalized kernel density maximum entropy method (GKDME) approximations using sample moments of different size. A, Probability density function. B, Probability of exceeding in semi-logarithmic scale

(*PGA*), A_g, defined from the seismic codes for each site and for different limit states:

$$RSA(T,\zeta_0) = \begin{cases} a_g S\left[1+\dfrac{T}{T_B}(2.5\eta-1)\right] & 0 \le T \le T_B \\ 2.5a_g S\eta & T_B \le T \le T_C \\ 2.5a_g S\eta\left(\dfrac{T_C}{T}\right) & T_C \le T \le T_D \\ 2.5a_g S\eta\left(\dfrac{T_C T_D}{T}\right) & T_D \le T \le T_E \end{cases}, \tag{56}$$

where $RSA(T,\zeta_0)$ is the elastic response spectrum, T is the natural period of the oscillator, ζ_0 is the damping ratio, $\eta = [10/(5+\zeta_0)]^{1/2}$, S is the soil factor, T_B and T_C are respectively the lower and upper limits of the period of the constant spectral acceleration branch, and T_D and T_E are respectively the values defining the beginning and end of the constant displacement response range of the spectrum. In the considered example, we choose type B soil profile,[47] $\zeta_0 = 5\%$, and $A_g = 0.3g$, where g is the acceleration of gravity. It follows that $T_B = 0.15$ seconds, $T_C = 0.50$ sec onds, $T_D = 2.00$ sec onds, and $T_E = 4.00$ seconds while $S = 1.20$. Figure 5A shows the corresponding response spectrum to these chosen values.

The seismic codes allow time history representations of the seismic action for analyzing the nonlinear behavior of the structures whereas the response spectrum technique might not provide accurate results. The codes do not suggest a

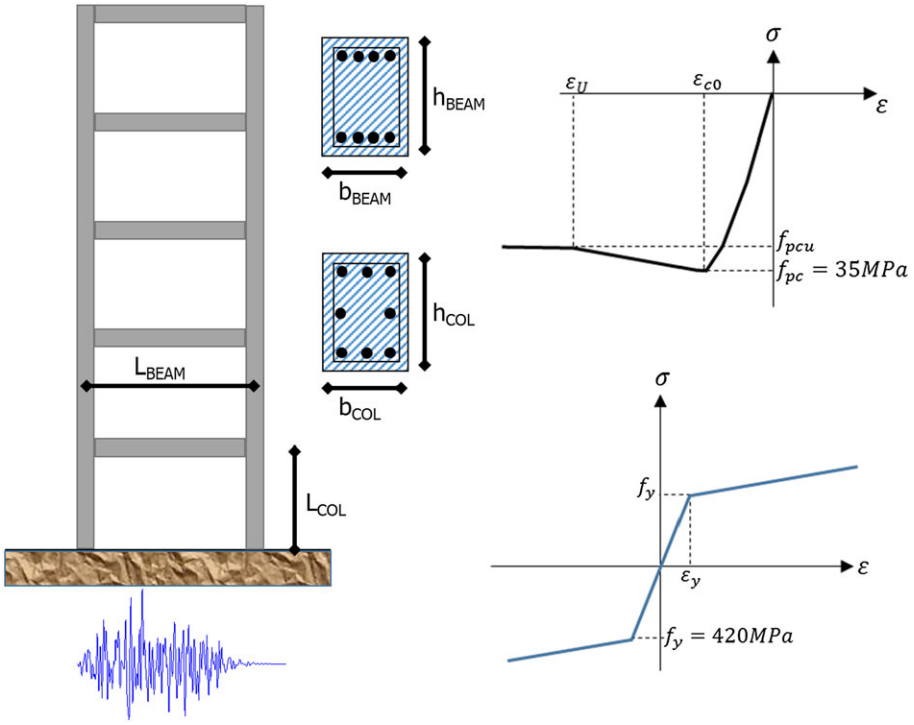

FIGURE 4 Schematic of the reinforced concrete frame subjected to spectrum-compatible ground acceleration

method for generating the earthquake time histories but only recommend that they are response spectrum compatible. Many common approaches model the earthquake-induced ground acceleration as realizations of a quasi-stationary Gaussian stochastic process. This is defined through evolutionary one-sided PSD:

$$G_{A_g}(\omega, t) = |\varphi(t)|^2 G_{A_g}(\omega), \tag{57}$$

where $\varphi(t)$ is a time-dependent modulating function modeling the temporal nonstationarity, while $G_{A_g}(\omega)$ is the one-sided PSD of the stationary counterpart. This representation is adopted in this paper using the model proposed in Cacciola et al,[48,49] which provides the following recursive expression to determine the one-sided PSD compatible with the chosen response spectrum:

$$\begin{cases} G_{A_g}(\omega_i) = 0 & 0 \le \omega \le \omega_a \\ G_{A_g}(\omega_i) = \dfrac{4\zeta_0}{\pi\omega_i - 4\zeta_0\omega_{i-1}}\left(\dfrac{RSA(\omega_i,\zeta_0)^2}{\overline{\eta}_U^2(\omega_i,\zeta_0)} - \Delta\omega\sum_{k=1}^{i-1} G_{A_g}(\omega_k)\right) & \omega > \omega_a \end{cases}, \tag{58}$$

where $RSA(\omega_i, \zeta_0)$ is the pseudo-acceleration response spectrum obtained from Equation 56 for given damping ratio ζ_0 and circular frequency $\omega_i = (2\pi)/T_i$, and $\overline{\eta}_U(\omega_i, \zeta_0)$ is the peak factor under the assumption of the following barrier outcrossing in clumps:

$$\overline{\eta}_U(\omega_i, \zeta_0) = \sqrt{2\ln\left\{2N_U\left[1 - \exp\left[-\delta_U^{1.2}\sqrt{\pi\ln(2N_U)}\right]\right]\right\}}, \tag{59}$$

$$\begin{cases} N_U = \dfrac{t_s}{2\pi}\omega_i(-\ln(0.5))^{-1} \\ \delta_U = \left[1 - \dfrac{1}{1-\zeta_0^2}\left(1 - \dfrac{2}{\pi}\arctan\dfrac{\zeta_0}{\sqrt{1-\zeta_0^2}}\right)\right]^{1/2} \end{cases}. \tag{60}$$

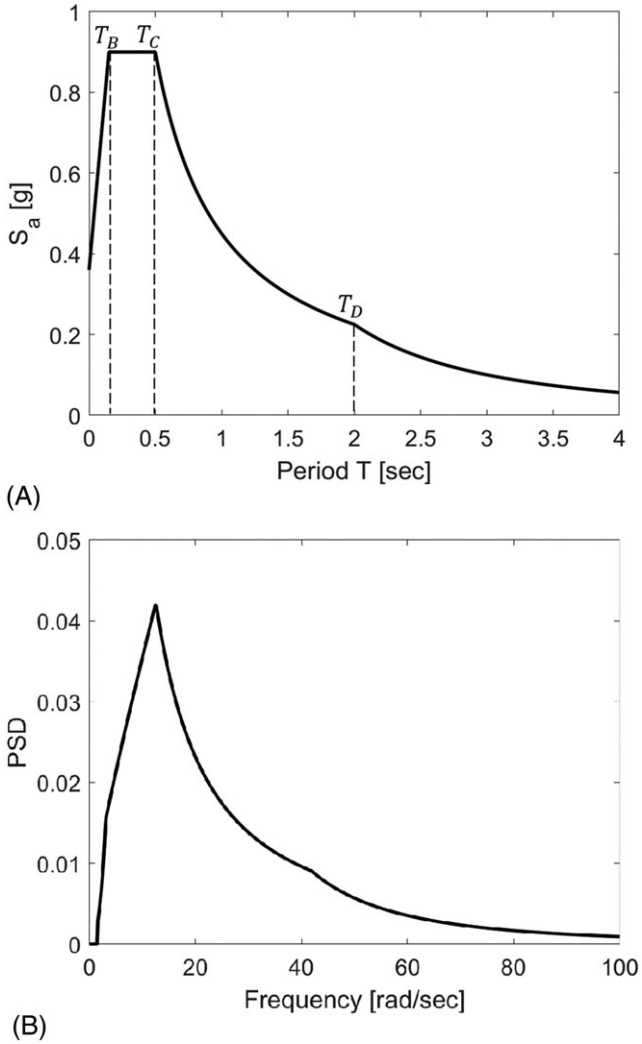

FIGURE 5 A, Response spectrum (Eurocode). B, Spectrum-compatible power spectral density (PSD)

In Equation 60, it has been assumed that the input PSD is smooth and $\zeta_0 \ll 1$. Moreover, t_s is the time observing window, assumed equal to the stationary part of the accelerogram. Here, $t_s = 10$ seconds has been chosen, coinciding with the mimimum duration suggested by the Eurocodes when site-specific data are not available. Finally, $\omega_a \cong 1$rad/s is the lowest bound of the existence domain of $\overline{\eta}_U$. Figure 5B represents the one-sided PSD $G_{A_g}(\omega)$ compatible with the response spectrum for a given site and *PGA*.[50-52]

The temporal nonstationarity is modeled through the modulating function proposed by Jennings et al[53]:

$$\varphi(t) = \begin{cases} \left(\frac{t}{t_1}\right)^2 & t < t_1 \\ 1 & t_1 \le t \le t_2 \\ \exp[-\beta(t-t_2)] & t > t_2 \end{cases}, \tag{61}$$

whose parameters are evaluated by imposing that the energy of the stochastic ground motion reaches the values of the 5% and 95%, respectively, in t_1 and t_2, providing

$$\beta = \frac{9}{t_s}, \quad t_1 = \frac{2.5}{\beta}, \quad t_2 = \frac{11.5}{\beta}. \tag{62}$$

The samples of spectrum-compatible ground acceleration are determined through Equation 55, where the PSD $G(\omega)$ is replaced by the evolutionary PSD $G_{A_g}(\omega, t)$ defined by Equation 57. For the sake of illustration, Figure 6 presents some samples.

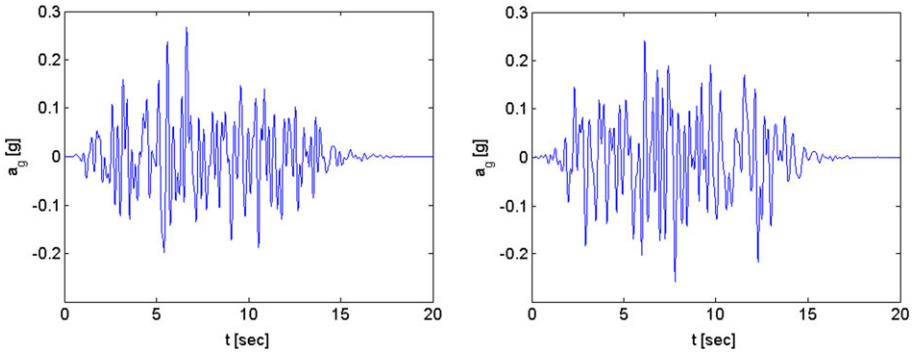

FIGURE 6 Two samples of simulated spectrum-compatible ground motions

For an informed decision-making process on the design of the structure, a powerful tool is represented by the performance-based earthquake engineering (*PBEE*) approach.[54,55] An important stage of the *PBEE* approach is represented by the fragility functions, defined as the probability that the engineering demand parameter (*EDP*) is greater than a chosen

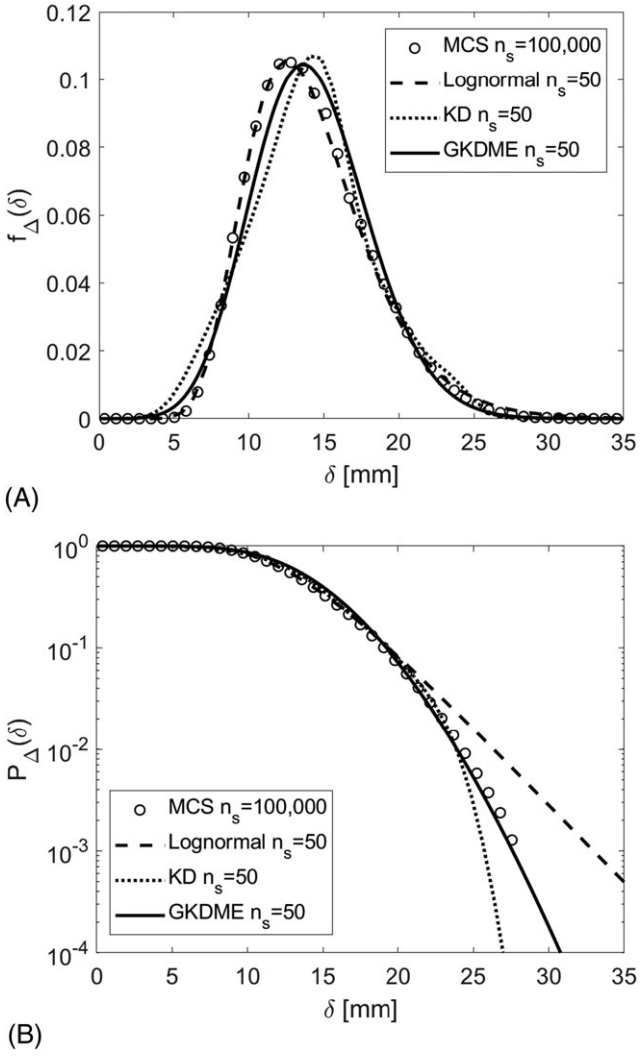

FIGURE 7 Fragility function in terms of the maximum interstory drift of a frame subjected to seismic excitation. Comparison of Monte Carlo simulation (MCS) (100 000 samples) with kernel density (KD), kernel density maximum entropy (KDME), and lognormal approximations (50 samples). A, Probability density function. B, Probability of exceeding in semi-logarithmic scale

threshold (edp) for a given intensity measure (IM) of the seismic input (im), ie, $P_{EDP}(edp) = Prob[EDP \geq edp|IM = im]$. In the considered case, the IM is the PGA with value $A_g = 0.3g$, while a typically adopted EDP in $PBEE$ is the maximum interstory drift Δ. Thus, the fragility function is $P_\Delta(\delta) = Prob[\Delta \geq \delta|A_g = 0.3g]$. Since all the simulated accelerograms have the same chosen PGA, the corresponding fragility function is determined from the distribution $f_\Delta(\delta)$ or $F_\Delta(\delta)$ of the maximum interstory drift.

Typically, the fragility functions are assumed unanimously to follow a lognormal distribution.[56,57] However, some researchers have recently questioned the validity of such assumption and proposed nonparametric methods for determining their distribution.[58] To this aim, in this numerical example, we compare the fits given by the lognormal distribution, the classical KD, and the newly developed GKDME in this paper with the reference solution, here represented by a MCS with $n_s = 100\,000$ samples. The approximate distributions are evaluated by considering a reduced set of dynamic computations, namely, $n_{s^{(1)}} = 50$ and $n_{s^{(2)}} = 2000$ samples. The corresponding results are shown in Figures 7 and 8. The target solution, represented by MCS, is represented with circle markers; only probability values $P_\Delta \geq 10^{-3}$ are shown. This is because, from reliability theory,[59,60] it is known that, for given number of samples n_s, the probability of exceeding P_Δ is estimated with a coefficient of variation ν_P only for

$$P_\Delta \geq \frac{1}{1 + n_s \upsilon_P^2}. \tag{63}$$

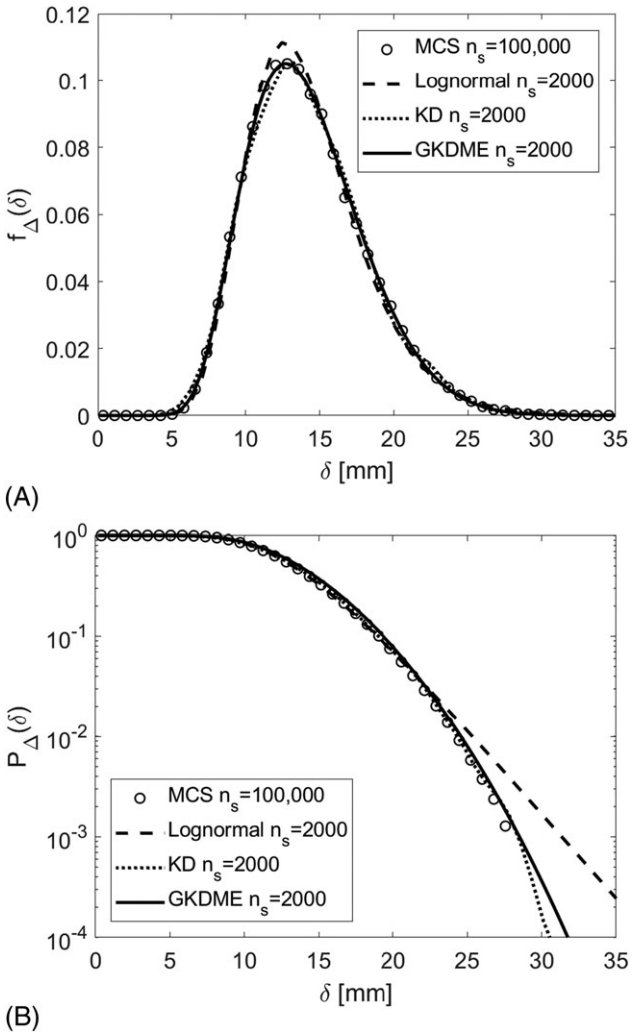

FIGURE 8 Fragility function in terms of the maximum interstory drift of a frame subjected to seismic excitation. Comparison of Monte Carlo simulation (MCS) (100 000 samples) with kernel density (KD), kernel density maximum entropy (KDME), and lognormal approximations (2000 samples). A, Probability density function. B, Probability of exceeding in semi-logarithmic scale

From Equation 63, it is seen that $P_\Delta \geq 10^{-3}$, for $n_s = 100\,000$ and a chosen accuracy $\nu_P = 10\%$. For a reduced number, $n_s^{(1)} = 50$ of simulated ground motions, the lognormal distribution fits very well the central part of $f_\Delta(\delta)$. However, the approximation of the tails is not as good, and this may represent a crucial issue for decision-making process with respect to events with low probability and high consequences. The KD and the GKDME are not in good agreement with the central part of the distribution, because with the KD estimations, no assumption is made about the functional form, but they are driven by the data. However, it is noted that only the GKDME predicts with good accuracy the tails of the distribution. In Figure 8, we show that the increase of the number of samples $n_s^{(2)} = 2000$ leads to an excellent fit of GKDME to the target distribution, including its tails. With 2000 samples, also, the KD gives a good approximation, but it is expected that for the lowest probability levels, ie, $P_\Delta \leq 10^{-3}$, GKDME outperforms KD, because of its prediction capabilities.

It is also noted that although the lognormal PDF predicts well the central part of the "true" $f_\Delta(\delta)$ for this example, some numerical applications could exist, where this is not the case. Conversely, the GKDMEM determines a data-driven distribution whose free parameters satisfy the principle of ME, and therefore, the GKDME PDF is the most "honest" distribution fitting the data but keeping good prediction capabilities with respect to the unseen data.

Finally, the bootstrap resampling technique described previously is used to assess the credible bounds. The percentiles $q = 5\%$ and 99% are used, while the bounds are obtained with 100 bootstrap samples. It is noted that the GKDME PDF with the bootstrap mean values of the probabilities p_i does not differ from the GKDME PDF estimated with the original set of observations. In Figure 9, we present the credible bounds of the GKDME distribution for the 2 samples of size $n_s^{(1)} = 50$ and $n_s^{(2)} = 2000$. As expected, the bounds are narrower for the larger $n_s^{(2)}$ case.

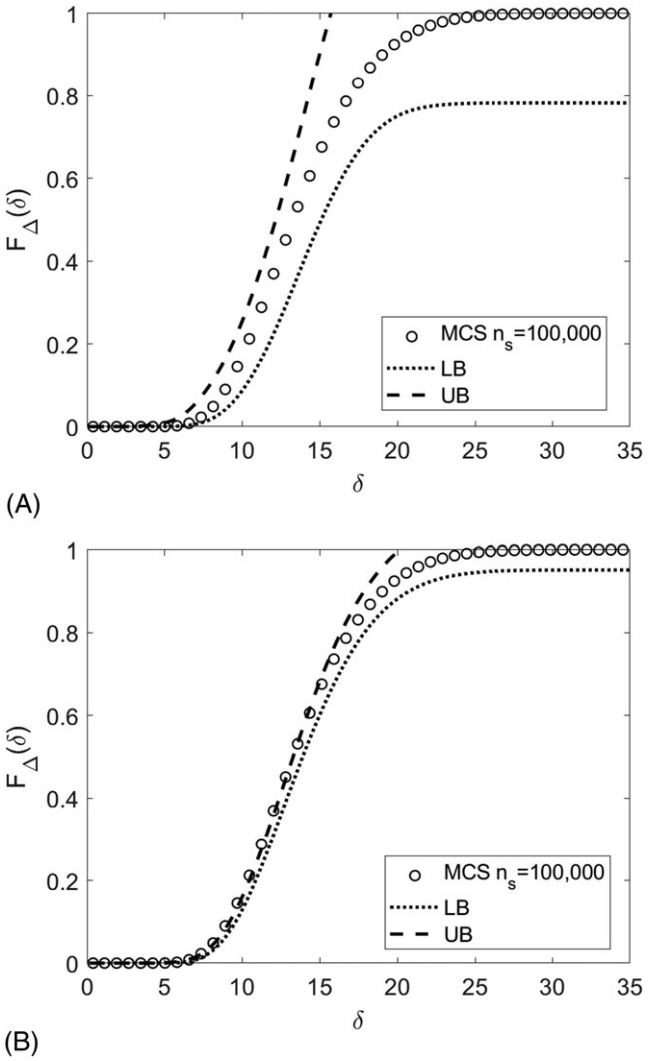

FIGURE 9 Credible bounds of the maximum interstory drift of a frame subjected to seismic excitation, in the case of data samples of size: A, 50. B, 2000

9 | CONCLUDING REMARKS

The paper presented a novel approach for determining the distribution of a random variable from a sample of data. The method is called KDME with generalized moments because it applies the principle of ME to the discretized moment problem, obtained through a KDR of the target distribution. The proposed approach has several advantages with respect to the existing literature. First, it is based on the principle of ME, which gives the least biased distribution with respect to the available information. Therefore, the obtained distribution is the most honest and least subjective choice. Second, the KDR guarantees high flexibility, providing a close fit to the data. From the other side, the proposed approach does not share the drawbacks of the existing KD methods: (1) it gives good approximations of the tails of the distributions also in the presence of samples of small size, (2) a good trade-off between bias and variance is obtained through the application of the Akaike information criterion, which penalizes the model complexity, and (3) the accuracy of the approximation is not significantly affected by the choice of the KDFs or by the evaluation of their bandwidth, since the free parameters are the probabilities determined through the principle of ME. Moreover, the parameters of the KDME are obtained through an algorithm with good properties of robustness and computational efficiency. The involved nested optimization procedure is simply reduced to the solution of a set of systems of linear equations with a reduced number of unknowns (typically not more than 4), which also implies the uniqueness of the solution for given sample of data.

From the considered numerical examples, the method can be applied by adopting good approximations of the population moments[13] or also through sample moments derived from samples of data. In the latter case, a determination of credible bounds can be developed through a bootstrap resampling technique. The main features of the method have been shown through the application to a Duffing oscillator, whose response distribution is known in a closed form. The application to a practical reinforced concrete example frame subjected to seismic excitation has shown the great potential of the proposed procedure to obtain an optimal distribution for the fragility functions, given the available information, represented from numerical analyses, or experimental data acquired through laboratory tests or a network of sensors in the field.

ACKNOWLEDGEMENT

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. Khalid M. Mosalam is a principal investigator of Tsinghua-Berkeley Shenzhen Institute (TBSI). The authors acknowledge the funding support from SinBerBEST and the partial support from TBSI.

The authors thank two anonymous referees whose comments have contributed to the enhancement of the paper.

ORCID

Umberto Alibrandi http://orcid.org/0000-0001-9729-9536
Khalid M. Mosalam http://orcid.org/0000-0003-2988-2361

REFERENCES

1. Pearson E, Johnson N, Burr I. Comparison of the percentage points of distributions with the same first four moments, chosen from eight different systems of frequency curves. *Commun Stat B*. 1979;8:191-229.
2. Low YM. A new distribution for fitting four moments and its applications to reliability analysis. *Struct Saf*. 2013;42:12-25. https://doi.org/10.1016/j.strusafe.2013.01.007
3. Charlier C. A new form of the frequency function. *Medd Lund Astron Obs Ser*. 1928;51
4. Edgeworth FY. The law of error. *Cambridge Ohilos Soc*. 1905;20:36-66. 113–41
5. Winterstein SR. Nonlinear vibration models for extremes and fatigue. *J Eng Mech*. 1988;114:1772-1790. https://doi.org/10.1061/(ASCE)0733-9399(1988)114:10(1772)
6. Jaynes ET. Information theory and statistical mechanics. *Phys Rev*. 1957;106:620-630. https://doi.org/10.1103/PhysRev.106.620
7. Jaynes ET. Where do we stand on maximum entropy? *The Maximum Entropy Formalism*. 1978;15-118. https://doi.org/10.1007/BF01008275

8. Kapur J, Kesavan H. *Entropy Optimization Principles with Applications*. San Diego, NY: Academic Press; 1992.
9. Vapnik VN. *The Nature of Statistical Learning Theory*. vol. 8 1995. doi:https://doi.org/10.1109/TNN.1997.641482.
10. Winterstein SR, Mackenzie CA. Extremes of nonlinear vibration: comparing models based on moments, L-moments, and maximum entropy. *J Offshore Mech Arct Eng*. 2015;135:1-7. https://doi.org/10.1115/1.4007050
11. Novi Inverardi PL, Tagliani A. Maximum entropy density estimation from fractional moments. *Commun Stat - Theory Methods*. 2003;32:327-345. https://doi.org/10.1081/STA-120018189
12. Taufer E, Bose S, Tagliani A. Optimal predictive densities and fractional moments. *Appl Stoch Model Bus Ind*. 2009;25:57-71.
13. Zhang X, Pandey MD. Structural reliability analysis based on the concepts of entropy, fractional moment and dimensional reduction method. *Struct Saf*. 2013;43:28-40. https://doi.org/10.1016/j.strusafe.2013.03.001
14. Athanassoulis GA, Gavriliadis PN. The truncated Hausdorff moment problem solved by using kernel density functions. *Probabilistic Eng Mech*. 2002;17:273-291. https://doi.org/10.1016/S0266-8920(02)00012-7
15. Alibrandi U, Ricciardi G. Efficient evaluation of the pdf of a random variable through the kernel density maximum entropy approach. *Int J Numer Methods Eng*. 2008;75:1511-1548. https://doi.org/10.1002/nme.2300
16. Akaike H. Information theory and an extension of the maximum likelihood principle. *2nd Int Symp Inf Theory*. 1973;267-281. https://doi.org/10.1016/j.econlet.2011.12.027
17. Jaynes ET. Information theory and statistical mechanics. *Phys Rev*. 1957;106:181-218. https://doi.org/10.1103/PhysRev.106.620
18. Jaynes ET. Prior probabilities. *IEEE Trans Syst Sci Cybern*. 1968;4:227-241. https://doi.org/10.1109/TSSC.1968.300117
19. Jaynes ET. On the rationale of maximum entropy methods. *Proc IEEE*. 1982;70:939-952.
20. Tagliani A. Discrete probability distributions in the generalized moment problem. *Appl Math Comput*. 2000;112:333-343.
21. Baker R. Probability estimation and information principles. *Struct Saf*. 1990;9:97-116. https://doi.org/10.1016/0167-4730(90)90002-7
22. Yuan C, Fox RO. Conditional quadrature method of moments for kinetic equations. *J Comput Phys*. 2011;230:8216-8246. https://doi.org/10.1016/j.jcp.2011.07.020
23. Silverman BW. *Density Estimation for Statistics and Data Analysis*. vol. 37 1986. doi:https://doi.org/10.2307/2347507.
24. Scott DW, Sain SR. Multidimensional density estimation. *Handb Stat*. 2004;24:229-261. https://doi.org/10.1016/S0169-7161(04)24009-3
25. Lin Y, Cai GQ. *Probabilistic Structural Dynamics*. New York: McGraw-Hill; 2004.
26. Alibrandi U, Ricciardi G. Stochastic method in nonlinear structural dynamics. In: Elishakoff I, Soize C, eds. *Nondeterministic Mechanics*. Udine, Italy: Springer; 2011:3-60.
27. Alibrandi U, Di Paola M, Ricciardi G. Path integral solution solved by the kernel density maximum entropy approach. Int. Symp. Recent Adv. Mech. Dyn. Syst. Probab. Theory, MDP – 2007, Palermo, June 3-6, 2007, 2007.
28. McGraw R. Description of aerosol dynamics by the quadrature method of moments. *Aerosol Sci Tech*. 1997;27:255-265. https://doi.org/10.1080/02786829708965471
29. Gautschi W. *Orthogonal Polynomials: Computation and Approximation*. USA: Oxford University Press; 2004.
30. Chalons C, Fox RO, Massot M. A multi-Gaussian quadrature method of moments for gas-particle flows in a LES framework. 2010.
31. Yuan C, Laurent F, Fox RO. An extended quadrature method of moments for population balance equations. *J Aerosol Sci*. 2012;51:1-23. https://doi.org/10.1016/j.jaerosci.2012.04.003
32. Wheeler JC. Modified moments and Gaussian quadratures. *Rocky Mt J Math*. 1974;4:287-296. https://doi.org/10.1216/RMJ-1974-4-2-287
33. Yuan C, Fox RO. Conditional quadrature method of moments for kinetic equations. *J Comput Phys*. 2011;230:8216-8246. https://doi.org/10.1016/j.jcp.2011.07.020
34. Lin GD. Characterization of distributions via moments. *Sankhya Indian J Stat Ser A*. 1992;54:128-132.
35. Tagliani A. Inverse z transform and moment problem. *Probab Eng Inf Sci*. 2000;14:393-404.
36. Novi Inverardi PL, Tagliani A. Discrete distributions from moment generating function. *Appl Math Comput*. 2006;182-200–9. https://doi.org/10.1016/j.amc.2006.03.048
37. Talenti G. Recovering a function from a finite number of moments. *Inverse Probl*. 1987;3:501. https://doi.org/10.1088/0266-5611/3/3/016
38. Tagliani A. Hausdorff moment problem and fractional moments: a simplified procedure. *Appl Math Comput*. 2011;218:4423-4432. https://doi.org/10.1016/j.amc.2011.10.019
39. Gzyl H, Novi-Inverardi PL, Tagliani A. Determination of the probability of ultimate ruin by maximum entropy applied to fractional moments. *Insur Math Econ*. 2013;53:457-463. https://doi.org/10.1016/j.insmatheco.2013.07.011
40. Efron B. The jackknife, the bootstrap and other resampling plans. *Complexity*. 1982;103. https://doi.org/10.1137/1.9781611970319
41. Lutes LD, Sarkani S. *Random Vibrations: Analysis of Structural and Mechanical Systems*. Burlington (MA): Elsevier; 2004.
42. Rice SO. Mathematical analysis of random noise. *Sel Pap Noise Stoch Process*. 1954;133-294.

43. Shinozuka M, Jan C-M. Digital simulation of random processes and its applications. *J Sound Vib*. 1972;25:111-128. https://doi.org/10.1016/0022-460X(72)90600-1

44. Shinozuka M, Deodatis G. Simulation of stochastic processes by spectral representation. *Appl Mech Rev*. 1991;44:191-204. https://doi.org/10.1115/1.3119501

45. McKenna F. Opensees user's manual. 2010.

46. CEN. Actions on structures—part 1-1: general actions—densities, self-weight, imposed loads for buildings. 2002;2.

47. CEN. Eurocode—basis of structural design. *En*. 2002;3:89. https://doi.org/10.1680/cien.144.6.8.40609

48. Cacciola P, Colajanni P, Muscolino G. Combination of modal responses consistent with seismic input representation. *J Struct Eng*. 2004;130:47. https://doi.org/10.1061/(ASCE)0733-9445(2004)130:1(47)

49. Cacciola P, D'Amico L, Zentner I. New insights in the analysis of the structural response to response-spectrum-compatible accelerograms. *Eng Struct*. 2014;78:3-16. https://doi.org/10.1016/j.engstruct.2014.07.015

50. Alibrandi U. Tail equivalent linearization methods for seismic response spectrum analysis. 1st ECCOMAS Themat Conf Int Conf Uncertain Quantif Comput Sci Eng (UNCECOMP 2015) 2015:368–77.

51. Alibrandi U, Mosalam KM. Stochastic dynamic analysis for code-conforming performance based engineering. Symp. Reliab. Eng. Syst. (SRES 2015), 2015.

52. Alibrandi U, Mosalam KM. Lifecycle optimal design using performance based engineering. Second Int. Conf. Performance-based life-cycle Struct. Eng. (PLSE 2015), Brisbane: 2015.

53. Jennings P, Housner G, Tsai C. Simulated earthquake motions for design purposes. 4th World Conf. Eart. Eng. Santiago, 1969, p. 145–60.

54. Cornell CA, Krawinkler H. Progress and challenges in seismic performance assessment. *PEER Cent News*. 2000;3:1-4.

55. Günay S, Mosalam KM. PEER performance-based earthquake engineering methodology, revisited. *J Earthq Eng*. 2013;17:829-858. https://doi.org/10.1080/13632469.2013.787377

56. Shinozuka M, Feng MQ, Lee J, Naganuma T. Statistical analysis of fragility curves. *J Eng Mech*. 2000;126:1224-1231. https://doi.org/10.1061/(ASCE)0733-9399(2000)126:12(1224)

57. Ellingwood BR. Earthquake risk assessment of building structures. *Reliab Eng Syst Saf*. 2001;74:251-262. https://doi.org/10.1016/S0951-8320(01)00105-3

58. Sudret B, Mai CV, Konakli K. Computing seismic fragility curves using non-parametric representations. *Earthq Eng Struct Dyn*. 2014;0:1-17. https://doi.org/10.1002/eqe

59. Ditlevsen O, Madsen HO. *Structural Reliability Methods*. Chichester: Wiley; 1996.

60. Melchers RE. *Structural Reliability, Analysis and Prediction*. New York: Wiley & Sons; 1999.

How to cite this article: Alibrandi U, Mosalam KM. Kernel density maximum entropy method with generalized moments for evaluating probability distributions, including tails, from a small sample of data. *Int J Numer Methods Eng*. 2018;113:1904–1928. https://doi.org/10.1002/nme.5725