IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received November 18, 2017, accepted December 20, 2017, date of publication January 5, 2018, date of current version February 28, 2018.

Digital Object Identifier 10.1109/ACCESS.2017.2787736

Moment-Constrained Maximum Entropy Method for Expanded Uncertainty Evaluation

ARVIND RAJAN[1], (Student Member, IEEE), YE CHOW KUANG[1], (Senior Member, IEEE), MELANIE PO-LEEN OOI[1,2,3], (Senior Member, IEEE), SERGE N. DEMIDENKO[4,5], (Fellow, IEEE), AND HERMAN CARSTENS[6]

[1]Department of Electrical and Computer Systems Engineering, School of Engineering, Monash University Malaysia, Bandar Sunway 47500, Malaysia
[2]Unitec Institute of Technology, Auckland 1025, New Zealand
[3]School of Engineering and Physical Sciences, Heriot-Watt University Malaysia, Putrajaya 62200, Malaysia
[4]School of Engineering and Advanced Technology, Massey University, Auckland 0745, New Zealand
[5]Department of Electrical and Computer Systems Engineering, Monash University, Clayton, Victoria 3800, Australia
[6]Centre for New Energy Systems, Department of Electrical, Electronic, and Computer Engineering, University of Pretoria, Pretoria 0002, South Africa

Corresponding author: Arvind Rajan (arvind.rajan@monash.edu)

This work was supported in part by the IEEE Instrumentation and Measurement Society through the 2017 Graduate Fellowship Grant, Malaysia Ministry of Science, Technology and Innovation, under Grant 03-02-10-SF0284, and in part by Monash University Malaysia through the Higher Degree Research Scholarship.

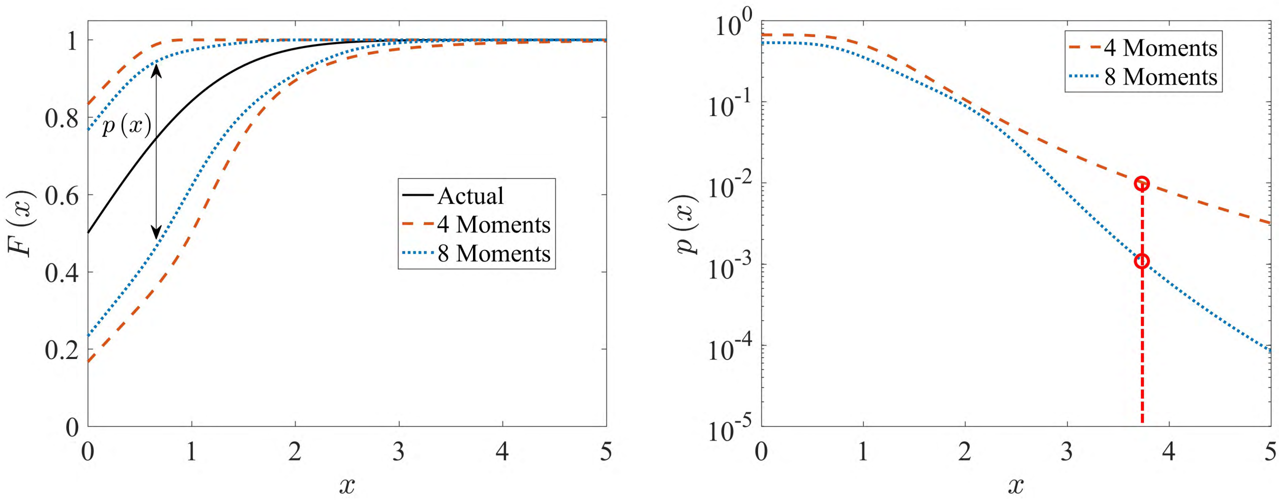
ABSTRACT The probability distribution is often sought in engineering for the purpose of expanded uncertainty evaluation and reliability analysis. Although there are various methods available to approximate the distribution, one of the commonly used ones is the method based on statistical moments (or cumulants). Given these parameters, the corresponding solution can be reliably approximated using various algorithms. However, the commonly used algorithms are limited by only four moments and assume that the corresponding distribution is unimodal. Therefore, this paper analyzes the performance of a relatively new and an improved parametric distribution fitting technique known as the moment-constrained maximum entropy method, which overcomes these shortcomings. It is shown that the uncertainty (or reliability) estimation quality of the proposed method improves with the number of moments regardless of the distribution modality. Finally, the paper uses case studies from a lighting retrofit project and an electromagnetic sensor design problem to substantiate the computational efficiency and numerical stability of the moment method in design optimization problems. The results and discussions presented in the paper could guide engineers in employing the maximum entropy method in a manner that best suits their respective systems.

INDEX TERMS Moments, probability distribution, confidence interval, uncertainty, maximum entropy, design optimization.

I. INTRODUCTION

Finding the probability distribution of a quantity-of-interest has always been a classical problem in the field of engineering. Depending on the type of information available, there are different methods to estimate the probability distribution. For example, it is common to employ Bayesian methodology [1] when the users intend to impose prior assumptions on the distribution estimation. Conversely, there are also many cases where direct observation of the output probability distribution is not available, but the statistics associated with the distribution are available, e.g., the moments (or cumulants). In such cases, the construction of the probability distribution must be based solely on moments. In engineering science and technology, moment-based probability distribution estimation methods are often employed for instrument calibrations [2], iterative procedures such as control design [3] and probabilistic design optimizations [4], solid state physics [5] and more.

The problem of finding the *cumulative distribution function* (CDF) or the *probability density function* (PDF) given a full knowledge of moments is known as the *moment problem*. Prominent mathematicians like Stieltjes, Chebyshev, Markov and many others laid a very solid theoretical foundation on the work [6]. The earlier works focus on the existence and uniqueness of solution (or distribution). However, finding the complete moment sequence is not practical in most cases and thus a finite moment sequence is often used in practice. The problem of finding CDF (or PDF) given a finite sequence of moments is commonly known as the *truncated moment problem*. It has been long known that the solution



(a) (b)

FIGURE 1. **(a) Bounds of the permissible CDF $F(x)$, whereby the solid line represents the CDF of a standard normal distribution while the dotted lines represent the bound/envelope of theoretically most extreme deviation from the normal distribution CDF given the identical first $n = 4, 8$ moments; and (b) the bound gap $p(x)$ in a logarithmic scale.**

to the truncated moment problem is not unique and is highly sensitive to the moment values [7]. These findings are very valuable in recognizing the limitation of the moment method (which will be clarified below), however, they also lead to the widespread misconception that the truncated moment approach is generally an unreliable and inferior method.

One of the most common uses of probability distributions approximated from truncated moments is for expanded uncertainty evaluation. According to the *Guide to the Expression of Uncertainty in Measurement* [8] and its supplement [9], which are one of the most widely cited documents for uncertainty evaluation, the coverage interval of a quantity at a prescribed confidence level will be referred to as the expanded uncertainty. Accordingly, the evaluation of expanded uncertainty only concerns the upper and lower "tails" of a probability distribution.

Study [10]–[12] presented the theoretical developments as well as the justification for using truncated moment sequence to find the expanded uncertainty. It went on to use the *local distribution bounds* [11], [12] to show that inferring the "tails" of a distribution from a finite sequence of moments is both robust and reliable, and that even the uncertainty within the functional space of all possible distributions is quantifiable. To illustrate this point, Fig. 1(a) shows the upper and lower bounds of the CDFs that have the first four and eight moments identical to those of the standard normal distribution. In other words, it illustrates the theoretical worst-case deviations of the possible CDFs that can be obtained from these moments in comparison to the actual CDF of the standard normal distribution. It is important to emphasize that the bounds are independent of any distribution fitting technique. Also, in agreement with the conclusion made in [10], Fig. 1(b) shows that the local distribution bounds would improve (or become narrower) asymptotically when increasing: 1) the number n of known moments; and 2) the distance of x from the mean regardless of the distribution (provided that it exists). The truncated moment method is therefore effective, easy to use, at lower risk of subjective judgement from parametric models and well-suited for the task of evaluating expanded uncertainty and reliability estimation.

There are many methods (analytical [13] and numerical [14]) that has been used to calculate the high-order moments of the output variables using information about the input variables. There are also various algorithms reported in literature to subsequently approximate the output probability distribution provided that the corresponding statistical moments are available. The algorithmic estimations of parametric distributions using the first four moments have become a standard practice in applied statistics and engineering. Study [15] summarizes various mainstream four-moment distribution fitting techniques developed over the last century and compares their performances across a wide variety of distributions. The results of the study showed that the Pearson distribution is the most reliable general-purpose algorithm for the evaluation of the expanded uncertainty. At the same time, it was found that the distribution fitting algorithms studied in [15] have the following limitations:

1) They support the use of only up to $n = 4$ moments. Consequently, the importance of the moments of higher orders ($n > 4$) in characterizing the distribution "tails" was not studied;
2) The studied algorithms cover only unimodal distributions and the effect of such an assumption on the quality of the expanded uncertainty evaluation for multimodal distributions is unknown.

Addressing these shortcomings, the objective of this study is to develop on the *principle of maximum entropy* [16] that overcomes the limitations of commonly used moments-based distribution fitting techniques and present an improved *maximum entropy* (MaxEnt) algorithm in Section II. The effectiveness of the algorithm is studied for the evaluation of expanded uncertainty in Section III. Consequently, the limitations of the MaxEnt method and the conditions in which the method performs the best have been reported. Section IV then goes on to highlight the advantages of using the moment method as a viable alternative to the state-of-the-art *Monte Carlo* (MC) method [9] in settings where the expanded uncertainty results are iteratively used in engineering design optimization. Here, case studies from a lighting retrofit project and an electromagnetic sesor design is employed for a more comprehensive discussion from computational efficiency and optimization convergence standpoint.

II. MOMENT-CONSTRAINED MAXIMUM ENTROPY METHOD

According to the principle of *maximum entropy* (MaxEnt) [17], all likelihoods are to be taken into account based on the information available in the data, which in this case is the set of truncated moments associated with the distribution-of-interest. The MaxEnt distribution would be the one with the largest overall uncertainty, measured by Shannons entropy, out of all possible distributions with the same moment sequence. Also, the MaxEnt distribution only commits to maximum entropy principle and it is sufficiently flexible to model various distribution characteristics such as multimodality, long-tail, etc., without needing the users to make prior assumptions.

The following section introduces a numerically stable MaxEnt algorithm that uses an arbitrary order n moments for the distribution fitting. An overview of the maximum entropy principle is given in Section II-A. Section II-B, however, presents a stable automatic numerical implementation of the MaxEnt algorithm, which is the main technical contribution of this paper. The subsequent Section III goes on to benchmark the performance of this approach for the expanded uncertainty against other distribution fitting methods [15].

A. FORMULATION OF THE MAXIMUM ENTROPY PROBLEM

For a random variable X, whereby its realization x takes all values over an interval of real numbers with unique PDF $f(x)$, the Shannon entropy S [18] is defined as:

$$S(x) = -\int_{\mathbb{R}} f(x) \ln f(x) \, \mathrm{d}x. \tag{1}$$

In the moment-based MaxEnt method, the information entropy S is maximized subject to:

$$\int_{\mathbb{R}} x^i f(x) \, \mathrm{d}x = m_i, \tag{2}$$

whereby $i = 0, \ldots, n$.

Using the method of Lagrange multipliers [19], the optimization problem with $n + 1$ constraints (considering 0-th moment) is then reduced to the optimization of the unconstrained function:

$$\mathcal{L}(\lambda) = \int_{\mathbb{R}} \exp\left(\sum_{i=0}^{n} \lambda_i x^i\right) \mathrm{d}x - \sum_{i=0}^{n} \lambda_i x^i, \tag{3}$$

whereby $\boldsymbol{\lambda} = \{\lambda_i\}$ for $i = \{0, \ldots, n\}$ is the Lagrange multiplier for its corresponding $\mathbf{x} = \{x^i\}$. Expression (3) has a closed form solution for $f(x)$:

$$f(x) = \exp\left(\sum_{i=0}^{n} \lambda_i x^i\right). \tag{4}$$

All the optimal MaxEnt distributions are achieved when $\partial\mathcal{L}/\partial\boldsymbol{\lambda} = 0$, which automatically satisfies the moment constraints in (2), and takes the general form $f(x)$ in (4). Note that since $x^0 = 1$ (for $i = 0$) in (4), λ_0 can be found by using the explicit function:

$$\lambda_0 = -\ln \int_{\mathbb{R}} \exp\left(\sum_{i=1}^{n} \lambda_i x^i\right) \mathrm{d}x. \tag{5}$$

The gradient $\partial\mathcal{L}/\partial\boldsymbol{\lambda}$ and the elements H_{ij} of the Hessian matrix $\mathbf{H}$ of the Lagrangian function in (3) are given, respectively, as:

$$(\nabla\mathcal{L})_i = \frac{\partial\mathcal{L}}{\partial\lambda_i} = \int_{\mathbb{R}} x^i f(x) \, \mathrm{d}x - x^i, \tag{6}$$

$$H_{ij} = \int_{\mathbb{R}} x^{i+j} f(x) \, \mathrm{d}x. \tag{7}$$

B. IMPLEMENTATION OF THE MOMENT-CONSTRAINED MAXIMUM ENTROPY METHOD

The MaxEnt problem described in Section II-A is an optimization problem to find the Lagrange multipliers $\boldsymbol{\lambda}$ in (4) such that the gradient $\nabla\mathcal{L} \cong 0$. The first algorithm to do it was developed in 1984 [16]. It uses monomials (i.e., x^i) as the basis functions in the expressions (2)-(7). However, as the number of moments increases (especially when $n > 5$), the procedure becomes highly sensitive to numerical imbalances in the moments, ill-conditioned gradient and Hessian matrix, and insufficient arithmetic precision [5], [20]. Addressing these shortcomings, advanced and numerically stable algorithms were developed in the past decade using different basis functions, such as the shifted Chebyshev polynomials [5] and the Fup functions [20]. Most recently, [21] extended the improvements to multidimensional problems using generalized orthogonal polynomials (GOPoly). However, these algorithms assume that the range of the distribution is known. Unfortunately, convergence is not guaranteed in an automatic application of the procedures [5], [20], [21] if the limit of integration is not known (at least approximately) beforehand.

This section therefore proposes a methodology to approximate the integral limits of the GOPoly method using the available set of moments. The improvement simplifies and enhances the numerical stability of the GOPoly algorithm. To keep this section succinct, the supporting expressions are provided in Appendix A whereas the improved MaxEnt

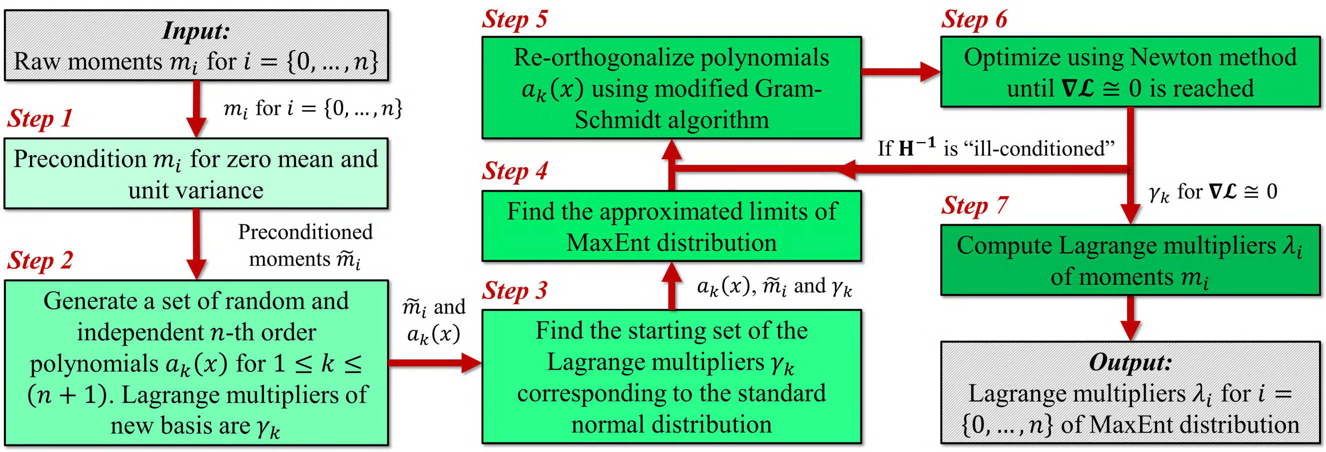


FIGURE 2. Framework of the proposed moment-constrained MaxEnt algorithm with enhanced numerical stability.

algorithm is illustrated in Fig. 2. Accordingly, the Lagrange multipliers $\boldsymbol{\lambda}$ in (4) can be found using the steps below:

1) Precondition the moments m_i for $i = \{0, \ldots, n\}$ by setting zero mean and unit variance (from Appendix A-A). The preconditioned moments are denoted by $\tilde{m}_i$;
2) Generate a set of random linearly independent n-th order polynomials $a_k(x)$, whereby $1 \le k \le (n+1)$. The Lagrange multipliers of the new basis $a_k(x)$ are denoted by γ_k. Mathematically, $\{x^i, \lambda_i\}$, for $0 \le i \le n \rightarrow \{a_k(x), \gamma_k\}$, for $1 \le k \le (n+1)$;
3) Using the preconditioned moments $\tilde{m}_i$ and the new polynomial basis $a_k(x)$, calculate the starting set of the Lagrange multipliers γ_k corresponding to the moments of the standard normal distribution;
4) Find the approximated integral limits (from Appendix A-B). The estimation considers the "tail" characteristics of the distribution based on the information provided by the moment values;
5) Re-orthogonalize the set of polynomials $a_k(x)$ using the modified Gram-Schmidt algorithm [22] (from Appendix A-C) for the current iteration. The orthogonal polynomials are denoted by $p_k(x)$. Re-compute its corresponding set of Lagrange multipliers γ_k;
6) Perform Newton method [19] until gradient $\nabla\mathcal{L} \cong 0$ is reached, or the inverse Hessian matrix $\mathbf{H}^{-1}$ becomes too "ill-conditioned" [21]. The inverse Hessian is deemed ill-conditioned if the condition number of the matrix $\kappa_{\mathbf{H}}$ exceeds a threshold value (in this paper it is 20). Compute the $\nabla\mathcal{L}$ and $\mathbf{H}$ from expressions (6) and (7) respectively;
7) If the gradient $\nabla\mathcal{L} \cong 0$ is reached, compute the Lagrange multipliers $\tilde{\lambda}_i$ of the preconditioned moments $\tilde{m}_i$ from γ_k and $p_k(x)$. Refer Appendix A-A to compute a standard set of Lagrange multipliers λ_i for the original moments m_i from $\tilde{\lambda}_i$. On the other hand, if $\kappa_{\mathbf{H}} \ge 20$, return to step 5. The polynomials are re-orthogonalized again to prevent the loss of their orthogonality, which could negatively affect the convergence of the Newton method.

Note that the optimization problem given in Section II-A is convex with a unique solution (if such a solution exists). Nevertheless, the convergence of the numerical algorithm provided above greatly depends on the accuracy of the calculation of the integrals. The paper employs the Gauss-Hermite quadrature rule [21] to compute the integrals effectively. The speed of the computation, on the other hand, depends on the computational speed of the computer hardware as well as the platform used to develop the algorithm. This study has developed a version of the improved algorithm using MATLAB [23] and it is employed in the subsequent Section III. The developed algorithm is available for download at: http://tc32.ieee-ims.org/content/maxent-distribution-fitting.

III. PERFORMANCE OF THE MOMENT-CONSTRAINED MAXIMUM ENTROPY METHOD

Prior to recommending the moment-constrained MaxEnt method discussed above for the expanded uncertainty estimation, validation against other distribution fitting techniques is required. It will also be the basis for recommending the conditions in which the MaxEnt method performs the best. Here, the expanded uncertainty estimation results from the MaxEnt method are compared against that of a set of analytical benchmark distributions, whereby the high-order moments of the analytical distributions are the inputs to the MaxEnt algorithm presented in Section II. The performance analysis is given in the following subsections which are divided into two parts corresponding to: 1) unimodal; and 2) multimodal distributions.

A. UNIMODAL DISTRIBUTIONS

This subsection utilizes the set of 124 analytically derived benchmark test distributions as well as the comprehensive procedural framework for the performance analysis of the expanded uncertainty evaluation techniques provided in [15].

First, the exact high-order moments up to $n = 12$ are calculated from the analytical expressions of the 124 distributions. Then, five sets of 124 MaxEnt distributions, using $n = \{4, 6, 8, 10, 12\}$ moments respectively, are obtained from the algorithm presented in Section II. The distributions are then used to estimate the values of x at the following percentile values (as per [15]): $\{0.001, 0.006, 0.01, 0.1, 0.27, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\} \times 10^{-2}$ for the lower "tail" and $\{90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.73, 99.9, 99.99, 99.994, 99.999\} \times 10^{-2}$ for the upper "tail" of the distributions. The accuracy of the estimation is then calculated using the actual values of x computed from the analytical distributions, and the performance of the MaxEnt method is analyzed by benchmarking against the most reliable algorithm from [15], which is the Pearson distribution. Although there are other variations of MaxEnt algorithms proposed in [5], [20], and [21], they cannot be used for performance comparison as those algorithms require that the limits of the MaxEnt distribution to be known, which is not the case in this study.

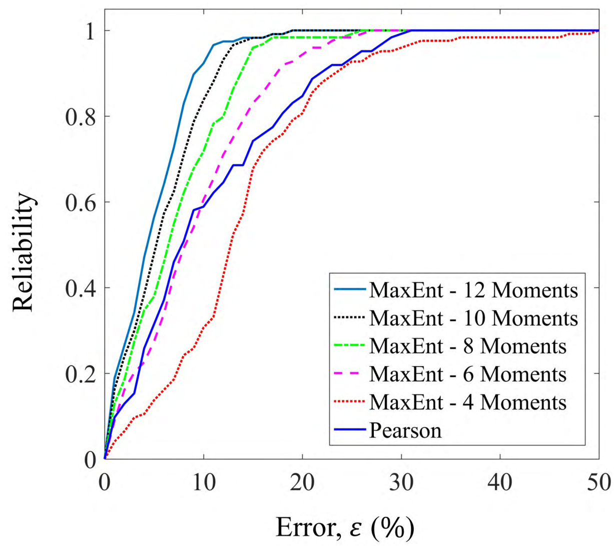


FIGURE 3. Expanded uncertainty estimation reliability plot for Pearson and MaxEnt methods with $n = \{4, 6, 8, 10, 12\}$ moments benchmarked against the analytical results of 124 test distributions [15]. The fastest technique to reach the maximum reliability of 1 is the most reliable algorithm since all the distributions fall within the smallest error interval, e.g., 90% of the 124 test distributions is within 25% expanded uncertainty estimation error when Pearson system is used.

1) PERFORMANCE METRIC AND RESULTS

The expanded uncertainty estimation error ε, which is the discrepancy in estimating x for a given percentile level $F(x)$, is used as the performance metric. The procedure to calculate ε, defined by

$$\varepsilon = \frac{|x - x^*|}{x - m_1}, \tag{8}$$

is outlined in [15], whereby x is the actual value that corresponds to a percentile value and x^* is the estimated value of x from the fitted distribution. Here, the estimation error ε is calculated for every percentile levels reported above and for all 124 analytical test distributions reported in [15]. In addition, the average computation time of each distribution is also reported.

Fig. 3 presents ε in the form of a reliability plot. The plot is determined by the number of distributions (out of 124) falling within the intervals of ε. Simply, the faster the reliability plot reaches the maximum value of 1, the more reliable the corresponding technique is. Table 1, on the other hand, presents the mean and standard deviation of ε for the 124 test distributions given in [15]. In addition, the table also reports the average computation time for each distribution (except for the Pearson method as it does not require an optimization algorithm for its deployment). Note that all computations were done on a computer with 64-bit Intel Core i5-3470 CPU and 8 GB RAM.

2) DISCUSSION

Based on Fig. 3, it can be deduced that for 90% of the 124 test distributions, the expanded uncertainty estimation error is within 25% when Pearson distribution is used. The same way of interpretation deduces that the performance of the MaxEnt method with $n = 4$ moments is marginally worse compared to the Pearson distribution. However, as anticipated from studies [10]–[12], increasing the number of moments improves the performance of the expanded uncertainty estimation. Table 1 further validates this outcome. The mean and standard deviation of ε for the MaxEnt method using only 4 moments are higher than those of the Pearson distribution method. On the contrary, impressively, they are lower than those of the Pearson distribution when 6 moments are used. The improvement is getting more significant with the higher number of moments. However, these performance enhancements come at the expense of additional computational time that can be reduced with more efficient numerical integration methods for higher n, better computer hardware, pre-compiled codes, parallel processing and different programming platforms (e.g., multi-dimensional MaxEnt algorithm [21] was developed on C programming language).

In summary, the results in Fig. 3 and Table 1 show that the MaxEnt method is capable of accommodating more than $n = 4$ moments while significantly improving the quality of the expanded uncertainty estimation. Although the longer moment sequence incurs the extended computational time, it provides considerably more information about the distribution "tails".

B. MULTIMODAL DISTRIBUTIONS

Even though the unimodal distributions constitute the most common class in applications, the multimodal distributions should not be ignored. In fact, the multimodal distribution type corresponds well to the operation of many real-world systems. For example, the power loss of a V6 gasoline engine [24] and diffusion concentration from hazardous releases [20] are characterized by bimodal distributions.

TABLE 1. Mean and standard deviation of the expanded uncertainty estimation error and average computation time of the distribution fitting algorithms based on 124 unimodal test distributions.

Method	Mean	Standard deviation	Average computation time (s)
Pearson	7.22%	11.30%	-
MaxEnt — 4 moments	10.99%	12.27%	0.4157
MaxEnt — 6 moments	7.03%	8.84%	1.2676
MaxEnt — 8 moments	5.63%	7.32%	2.8269
MaxEnt — 10 moments	4.43%	5.73%	43.6925
MaxEnt — 12 moments	3.72%	4.98%	63.6472

TABLE 2. Mean and standard deviation of the expanded uncertainty estimation error and average computation time of the distribution fitting algorithms based on 6 multimodal test distributions.

Method	Mean	Standard deviation	Average computation time (s)
Pearson	11.46%	10.77%	-
MaxEnt — 4 moments	3.55%	4.20%	1.3768
MaxEnt — 8 moments	1.00%	1.73%	3.3671
MaxEnt — 12 moments	0.34%	0.58%	72.8133

Unlike the unimodal distributions, extensive set of multimodal test distributions are not available in the literature. Therefore, six multimodal distributions from [20], [24]–[27] are used in this work as a test set to analyze the expanded uncertainty estimation performance of the MaxEnt method. The PDFs of the distributions are listed in Appendix B, whereby their high-order moments have been calculated analytically.

1) PERFORMANCE METRIC AND RESULTS

Following the same procedures as in Section III-A, the performance analysis here is performed by calculating the expanded uncertainty estimation error ε. Table 2 presents the mean and standard deviation of ε, as well as the average computation time for the six test distributions in Appendix B. To keep the section succinct, the MaxEnt method is employed using $n = 4, 8, 12$.

2) DISCUSSION

It can be seen from the mean and standard deviation values of ε (Table 2) that the Pearson distribution approach (which is the most reliable technique for unimodal distributions as shown in [15]), is less reliable compared to the MaxEnt method (regardless of n). The MaxEnt method demonstrates its superiority over the Pearson one for the multimodal distributions given in Appendix B. Furthermore, as observed in Section III-A, increasing the number of moments (e.g., from $n = 4$ to $n = 8$, and further to $n = 12$) reduces the uncertainty estimation error. However, it is worthwhile to note that the mean and standard deviation values of ε for multimodal distributions (Table 2) are lower than those of the unimodal distributions (Table 1). This is due to the inclusion of challenging distributions with extreme skewness/kurtosis in the unimodal benchmark set, while no such distributions are present in the multimodal distribution set employed in this study. The computation time also increases with the number of moments. Additionally, the results in Table 1 and Table 2 suggest that the average computation time of the proposed MaxEnt algorithm is longer for the multimodal distributions compared to that of the unimodal distributions.

The above results suggest that the MaxEnt method should always be preferred over the Pearson distribution unless $n \leq 4$ and the output distribution is known to be unimodal (modality testing is outside the scope of this paper).

IV. CASE STUDY: MOMENT-BASED EXPANDED UNCERTAINTY EVALUATION IN ENGINEERING DESIGN OPTIMIZATION

In agreement with the deductions made in [10], Section III has evidently shown that the expanded uncertainty evaluation performance of the proposed moment-constrained MaxEnt algorithm in Section II improves with the number of moments n. At the same time, the algorithm also addresses the two shortcomings of the commonly used distribution fitting techniques. The aim of this section, on the other hand, is to highlight the strength of moment methods for repeated uncertainty evaluation (or reliability analysis) in engineering design and optimization.

As the cost of computing power has decreased over the last few decades, the MC simulation has become the most popular technique to estimate non-standard PDFs. The MC method generates many independent realizations of the input quantities using known probability distributions of inputs and then propagates through a system to obtain the output probability distribution, which can be used to infer the expanded uncertainty. Not only can it model any output distribution, but its accuracy can also be increased reliably simply by increasing the number of MC sample size. However, the use of MC to find expanded uncertainty within an iterative procedure will lead to two significant problems.

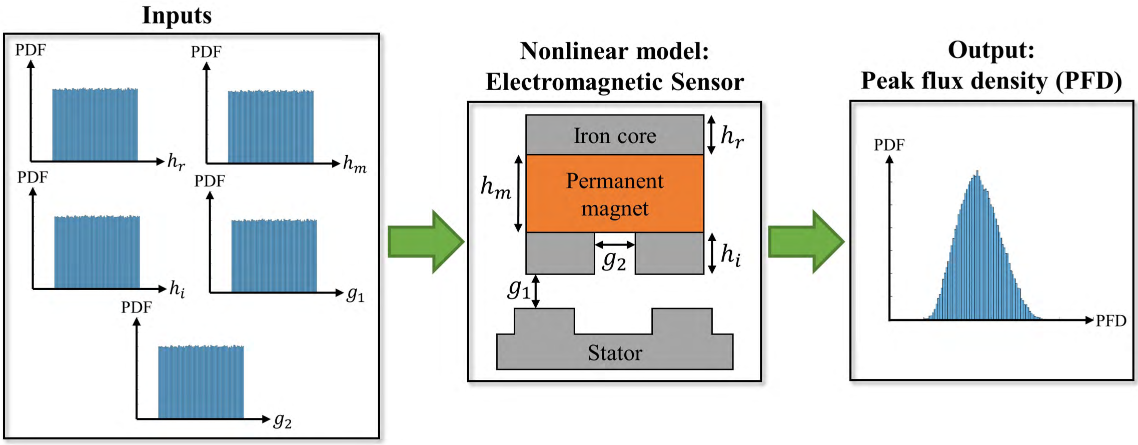


FIGURE 4. Illustration of the probability distributions of the input variables (design parameters), a highly nonlinear model (electromagnetic sensor [31]) and the probability distribution of the output quantity-of-interest (sensor peak flux density).

The first problem is the long computational time. This issue is very well-documented in engineering design optimization literature [28]. Various numerical methods to circumvent the use of MC in an optimization setting have been developed over last three decades. [29], [30] has shown that the moment method is faster and more accurate than these numerical algorithms. The moment methods have also been used to significantly speed-up the design and reliability assessment of magnetic sensors [31] and fuel cell [32]. The second problem, however, is a subtler one. The simulation noise from MC will affect the stability or convergence of the iterative search/optimization procedure. A recently reported lighting retrofit *Measurement and Verification* (M&V) project [33] and a three-dimensional electromagnetic sensor design problem [31] will be employed to illustrate the afore-mentioned shortcomings of MC method and the advantages of using the moment method in an iterative procedure.

In the M&V study, a large quantity of inefficient lamps is being replaced by energy efficient ones. M&V methodology has been used to quantify the savings realized over a number of years. It employs measurement instruments such as surveys and on-site energy meters. The expanded uncertainty of the reported savings must adhere to a given set of bounds in order to be eligible for incentive programs from the funding body. Increasing the sample size would reduce estimation uncertainty at the expense of a higher cost. Therefore, the goal is to design an optimal annual sampling/measurement plan, in a manner which satisfies the uncertainty requirements at the minimal cost. In the design stage, the energy saving Y is computed at each time step using the equation (9), whereby a is the number of lamps retrofitted. One of the uncertainty sources, the lamp survival X_1, is beta-distributed while the others, such as mean annual energy use X_2 and ratio of power consumption X_3, are normally distributed.

$$Y = aX_1X_2(X_3 - 1) \tag{9}$$

In the study, the *genetic algorithm* (GA) in conjunction with the integer linear programming [33] was used to design an appropriate sampling plan so that the M&V project results could satisfy the uncertainty requirement at the minimum sample size (cost). On a 64-bit Intel Core i7 quad-core CPU with 8 GB of memory running Linux Ubuntu 16.04 and Python 3.3 (Numpy), the application of the MC trial size of 10^7 on the genetic algorithm with 50 individuals, evolved over 50 generations taking 320 minutes. However, when the moment-based method was employed, the same problem converged within a minute, showing an impressive gain in the computational speed by three orders of a magnitude.

The computational speed gain becomes even more critical in engineering design optimization problems whereby the nonlinear model output (9) is replaced with the output from a finite element model. Take, for example, an electromagnetic sensor [31] (as illustrated in Fig. 4) that is optimized considering the manufacturing tolerance in its design parameters. Similar to the M&V problem described above, the PDFs of the design parameters (inputs) are known whereas the PDF of the peak flux density (output) of the sensor must be computed within the GA optimization framework. Here, one execution of the finite element model takes 65 minutes and even for a relatively small sample size of 10^6, reliable computation of the expanded uncertainty using MC simulation could take more than a century. However, with the use of response surface moment-based uncertainty analysis, the computational time can be reduced to several minutes, which in turn makes the GA-based optimization of the sensor practically feasible. This is also the case in various other design optimization problems [4], [25], [28], [31], [32], e.g., structural, mechanical, etc. Besides, as the solutions of these design problems were found using the traditional Pearson fitting technique, the results from the previous section show that the quality of the solution can be further improved with higher order moments in combination with the proposed MaxEnt algorithm.

The second problem (the optimizer stability), however, is a more serious one. Optimizers such as the genetic algorithm can find solutions that are close to (while not violating)

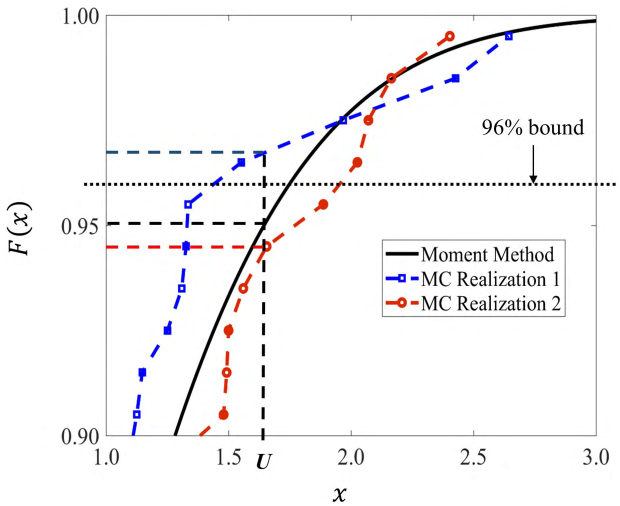


FIGURE 5. Assume that point *U* should cover less than 96% percentile. The moment method solution (solid line) satisfies this requirement consistently. At the same time, MC sometimes does not satisfy (in realization 1) and sometimes does satisfy (in realization 2) due to the simulation noise.

a specified coverage interval. The inter-simulation variation between the MC realizations would create a design that adheres to the constraints most of the time, but occasionally generate outliers that violate the constraints. Fig. 5 illustrates this phenomenon, whereby the solution from the moment method is consistent, whereas those of two independent MC simulations are different. Note that the CDFs representing MC simulation are shown only for the sake of discussion completeness and do not represent the actual results.

These 'false positives (or negatives)' happen rarely. However, when (if) they happen, the logical consistency of the optimization search is broken by the MC noise. Often, the optimizer terminates failing to determine the reason of the constraint violation that occurs randomly. Unless the complete MC realization is kept for a post-mortem inspection, the user will be left with no indication as to why the optimization algorithm fails randomly. The chance of these occurrences increases with the size and complexity of the sampling plan. In the worst-case scenario, some of the non-viable solutions can get through the optimization as viable ones due to the MC noise. The probability of occurrence may be reduced through the use of a larger MC trial size. However, that would significantly increase the computation load without fundamentally addressing the root cause. Furthermore, only a marginal benefit may be derived from an increased trial size, since most of the additional trials are located in the high-mass regions of the PDF, and not in the ''tails'', which are of the main interest for the uncertainty and reliability evaluation.

Therefore, in view of the problems outlined above, for cases where repeated uncertainty evaluation is required such as engineering design optimization and reliability estimation, using the method of moments with MaxEnt algorithm represents an attractive alternative due to its computational efficiency and the analytical representation that maintains its constraint-adherence consistency throughout the optimization process.

V. CONCLUSION

Methods using statistical information such as the moments (or cumulants) are one of the commonly used ones for finding the probability distribution of a desired output in various fields of engineering. Previous studies have shown that the moment method is useful for the purpose of expanded uncertainty evaluation done based on the resultant distribution, especially for systems with stringent reliability requirements. This is due to the increasing reliability of employing high-order moments along the ''tails'' of the distribution-of-interest, and with the increase in the number of available (truncated) moments. However, the commonly used distribution fitting methods are limited to supporting only four moments and assumes that the output distributions are unimodal.

This paper, therefore, studied the use the principle of *maximum entropy* (MaxEnt) to overcome these shortcomings and went on to propose an improved and a simplified moment-constrained MaxEnt algorithm. The improvement to the algorithm was made by adding a sub-algorithm to approximate the integral limits of the MaxEnt distribution for a better overall numerical stability. The obtained results suggest that the MaxEnt method is a highly reliable technique for the evaluation of the expanded uncertainty. At the same time, if only four moments are available and the distribution is known to be unimodal, Pearson distribution may also be deployed. Furthermore, the results also suggest that users employing the MaxEnt method with a long (known) moment sequence (typically more than 6), should take into account the computation time that increases nonlinearly with respect to the number of moments. However, the trade-off between the accuracy and computation time largely depends on the computer hardware and platform used to develop and implement the algorithm.

Lastly, the paper used an energy measurement and verification study and an electromagnetic sensor design optimization problem to draw attention to the advantages of using the moment method as an alternative to the Monte Carlo method when repetitive, consistent (or ''noiseless'') and computationally efficient uncertainty evaluation is sought.

APPENDIX A
EXPRESSIONS TO THE STEPS REPORTED IN SECTION II-B

A. SETTING ZERO MEAN AND UNIT VARIANCE

The transformation of moments m_i, for $i = 0, \ldots, n$, when conditioning them to those of zero-mean and unit variance distribution is given by (10):

$$\tilde{m_i} = \frac{1}{\left(\sqrt{m_2}\right)^i} \sum_{j=0}^{i} \binom{i}{j} (-1)^j \, m_{i-j} {m_0}^j, \tag{10}$$

whereby $\tilde{m_i}$ denotes the transformed moment values. The inverse transformation of the Lagrange multipliers $\tilde{\lambda_i}$ of the preconditioned moments $\tilde{m_i}$ to the Lagrange multipliers λ_i of

Algorithm 1 Algorithm for Modified Gram-Schmidt Orthogonalization

Input: Linearly independent n-th order polynomials $a_k(x)$ and their Lagrange multipliers γ_k for all k
Output: Orthogonalized n-th order polynomials $p_k(x)$ and their corresponding γ_k for all k

```
Initialize the square matrix a of size (n + 1), whereby each column vector corresponds to the coefficients of polynomial a_k(x); vector γ of size (n + 1) corresponding to all values of γ_k; γ^initial = γ; and a^initial = a.
for k = 1 to n + 1 do
    for l = 1 to 2 do
        for m = 1 to k = 1 do
            a_k(x) = a_k(x) − Q(a_k(x)p_m(x))p_m(x)
            Update matrix a
            γ = a^{-1}(a^initial γ^initial)
        end for
    end for
    p_k(x) = Q([a_k(x)]^2)^{-1/2} a_k(x)
end for
Initialize the square matrix p of size (n + 1), whereby each column vector corresponds to the coefficients of polynomial p_k(x)
γ = p^{-1}(p^initial γ^initial)
return p_k(x) and γ_k for all k
```

the original moments m_i is given by (11):

$$\lambda_i = \sum_{j=0}^{n} \binom{j}{i} (-1)^{j-i} \tilde{\lambda}_j m_1. \tag{11}$$

B. FINDING THE INTEGRAL LIMITS

This section presents the procedures to approximate the integral limits in a succinct manner based on the theoretical findings reported in [19]. First, shift the original moments m_i, for $i = 0, \ldots, n$, to point-of-interest C using the linear transformation (12):

$$\acute{m}_i = \sum_{j=0}^{i} \binom{i}{j} (-C)^{i-j} m_j, \tag{12}$$

whereby $\acute{m}_i$ is the linearly shifted moment. Then, find the maximal mass τ at 0 using (13):

$$\tau = \frac{\begin{vmatrix} \acute{m}_0 & \acute{m}_1 & \cdots & \acute{m}_{\acute{n}} \\ \acute{m}_1 & \acute{m}_2 & \cdots & \acute{m}_{\acute{n}+1} \\ \vdots & \vdots & \ddots & \vdots \\ \acute{m}_{\acute{n}} & \acute{m}_{\acute{n}+1} & \cdots & \acute{m}_{2\acute{n}} \end{vmatrix}}{\begin{vmatrix} \acute{m}_2 & \acute{m}_3 & \cdots & \acute{m}_{\acute{n}+1} \\ \acute{m}_3 & \acute{m}_4 & \cdots & \acute{m}_{\acute{n}+2} \\ \vdots & \vdots & \ddots & \vdots \\ \acute{m}_{\acute{n}+1} & \acute{m}_{\acute{n}+2} & \cdots & \acute{m}_{2\acute{n}} \end{vmatrix}}, \tag{13}$$

whereby $\acute{n} = \lfloor \frac{n}{2} \rfloor$. This paper uses $\tau = 10^{-(\acute{n}+1)}$ and solves expression (13) in the real coordinate space for C. The solutions are used as the upper and lower limits of the integrals in Section II-B.

C. MODIFIED GRAM-SCHMIDT ALGORITHM FOR POLYNOMIAL ORTHOGONALIZATION

The maximum entropy method proposed in [21] uses the modified Gram-Schmidt orthogonalization method [22] tailored for the polynomial basis $p_k(x)$, for $1 \le k \le (n+1)$. The algorithm is presented below, whereby $Q(\cdot)$ is given by (14), where g is a polynomial.

$$Q(g) = \int_{\mathbb{R}} g \exp\left(\sum_{i=1}^{K} \gamma_i p_i(x)\right) \mathrm{d}x \tag{14}$$

APPENDIX B
PROBABILITY DENSITY FUNCTIONS OF MULTIMODAL DISTRIBUTIONS

PDFs of the multimodal distributions from [20] and [24]–[27] are given by (15)-(20), whereby $\beta(\cdot)$ denotes the Beta function.

$$f(x) = \frac{x^7}{3\beta(8,1)} + \frac{x^{15}}{3\beta(16,1)} + \frac{x^{63}(1-x)^{-\frac{1}{2}}}{3\beta\left(64, \frac{1}{2}\right)} \tag{15}$$

$$f(x) = \frac{0.5\exp\left(-\frac{(x+4)^2}{32}\right)}{\sqrt{32\pi}} + \frac{0.5\exp\left(-\frac{(x-4)^2}{18}\right)}{\sqrt{18\pi}} \tag{16}$$

$$f(x) = \frac{0.4\exp\left(-\frac{(x+1)^2}{0.32}\right) + 0.6\exp\left(-\frac{(x-1)^2}{0.32}\right)}{\sqrt{0.32\pi}} \tag{17}$$

$$f(x) = \frac{0.4\exp\left(-\frac{x^2}{2}\right) + 0.6\exp\left(-\frac{(x-3)^2}{2}\right)}{\sqrt{2\pi}} \tag{18}$$

$$f(x) = \frac{0.4\exp\left(-\frac{x^2}{2}\right) + 0.6\exp\left(-\frac{(x-25)^2}{2}\right)}{\sqrt{2\pi}} \tag{19}$$

$$f(x) = \frac{(x+1)^{23}(1-x)^{11} + (x+1)^{11}(1-x)^{23}}{2^{36}\beta\left(64, \frac{1}{2}\right)} \tag{20}$$

REFERENCES

[1] I. Lira and D. Grientschnig, "Bayesian assessment of uncertainty in metrology: A tutorial," *Metrologia*, vol. 47, no. 3, p. R1, 2010.
[2] D. A. Lampasi, "An alternative approach to measurement based on quantile functions," *Measurement*, vol. 41, no. 9, pp. 994–1013, Nov. 2008.
[3] G. D'Antona, A. Monti, F. Ponci, and L. Rocca, "Maximum entropy multivariate analysis of uncertain dynamical systems based on the Wiener–Askey polynomial chaos," *IEEE Trans. Instrum. Meas.*, vol. 56, no. 3, pp. 689–695, Jun. 2007.
[4] B. H. Ju and B. C. Lee, "Reliability-based design optimization using a moment method and a kriging metamodel," *Eng. Optim.*, vol. 40, no. 5, pp. 421–438, 2008.
[5] K. Bandyopadhyay, A. K. Bhattacharya, P. Biswas, and D. A. Drabold, "Maximum entropy and the problem of moments: A stable algorithm," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 71, no. 5, p. 057701, 2005.
[6] N. I. Aheizer and N. Kemmer, *The Classical Moment Problem and Some Related Questions in Analysis*. London, U.K.: Oliver & Boyd, 1965.
[7] P. McCullagh, "Does the moment-generating function characterize a distribution?" *Amer. Statistician*, vol. 48, no. 3, p. 208, 1994.

[8] *Guide to the Expression of Uncertainty in Measurement*, document JCGM 100, Joint Committee for Guides in Metrology, 2008.
[9] *Evaluation of Measurement Data—Supplement 1 to the 'Guide to the Expression of Uncertainty in Measurement'—Propagation of Distributions Using a Monte Carlo Method*, document JCGM 101:2008, Joint Committee for Guides in Metrology, Sèvres, France, 2008.
[10] A. Rajan, Y. C. Kuang, M. P.-L. Ooi, and S. N. Demidenko, ''Moments and maximum entropy method for expanded uncertainty estimation in measurements,'' in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, May 2017, pp. 1–6.
[11] B. G. Lindsay and P. Basak, ''Moments determine the tail of a distribution (but not much else),'' *Amer. Statistician*, vol. 54, no. 4, pp. 248–251, 2000.
[12] S. Rácz, Á. Tari, and M. Telek, ''A moments based distribution bounding method,'' *Math. Comput. Model.*, vol. 43, no. 11, pp. 1367–1382, 2006.
[13] R. Willink, ''A procedure for the evaluation of measurement uncertainty based on moments,'' *Metrologia*, vol. 42, no. 5, p. 329, 2005.
[14] E. Isaacson and H. B. Keller, *Analysis of Numerical Methods*. Chelmsford, MA, USA: Courier Corporation, 1994.
[15] A. Rajan, Y. C. Kuang, M. P.-L. Ooi, and S. N. Demidenko, ''Benchmark test distributions for expanded uncertainty evaluation algorithms,'' *IEEE Trans. Instrum. Meas.*, vol. 65, no. 5, pp. 1022–1034, May 2016.
[16] L. R. Mead and N. Papanicolaou, ''Maximum entropy in the problem of moments,'' *J. Math. Phys.*, vol. 25, no. 8, pp. 2404–2417, 1984.
[17] E. T. Jaynes, ''Information theory and statistical mechanics,'' *Phys. Rev.*, vol. 106, no. 4, p. 620, 1957.
[18] C. E. Shannon, ''A mathematical theory of communication,'' *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 5, no. 1, pp. 3–55, 2001.
[19] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA, USA: Athena Scientific, 1999.
[20] H. Gotovac and B. Gotovac, ''Maximum entropy algorithm with inexact upper entropy bound based on Fup basis functions with compact support,'' *J. Comput. Phys.*, vol. 228, no. 24, pp. 9079–9091, 2009.
[21] R. V. Abramov, ''The multidimensional maximum entropy moment problem: A review of numerical methods,'' *Commun. Math. Sci.*, vol. 8, no. 2, pp. 377–392, 2010.
[22] L. Giraud, J. Langou, and M. Rozloznik, ''The loss of orthogonality in the Gram–Schmidt orthogonalization process,'' *Comput. Math. Appl.*, vol. 50, no. 7, pp. 1069–1075, 2005.
[23] MathWorks Inc. (2016). *MATLAB*. [Online]. Available: http://www.mathworks.com/products/matlab/
[24] Z. Xi, C. Hu, and B. D. Youn, ''A comparative study of probability estimation methods for reliability analysis,'' *Struct. Multidiscipl. Optim.*, vol. 45, no. 1, pp. 33–52, 2012.
[25] H. Y. Kang and B. M. Kwak, ''Application of maximum entropy principle for reliability-based design optimization,'' *Struct. Multidiscipl. Optim.*, vol. 38, no. 4, pp. 331–346, 2009.
[26] G. Li and K. Zhang, ''A combined reliability analysis approach with dimension reduction method and maximum entropy method,'' *Struct. Multidiscipl. Optim.*, vol. 43, no. 1, pp. 121–134, 2011.
[27] S. B. Provost, ''Moment-based density approximants,'' *Math. J.*, vol. 9, pp. 727–756, May 2005.
[28] M. A. Valdebenito and G. I. Schuëller, ''A survey on approaches for reliability-based optimization,'' *Struct. Multidiscipl. Optim.*, vol. 42, no. 5, pp. 645–663, 2010.
[29] A. Rajan, M. P.-L. Ooi, Y. C. Kuang, and S. N. Demidenko, ''Efficient analytical moments for the robustness analysis in design optimisation,'' *J. Eng.*, vol. 1, p. 8, Oct. 2016.
[30] A. Rajan, M. P.-L. Ooi, Y. C. Kuang, and S. N. Demidenko, ''Reliability-based design optimisation of technical systems: Analytical response surface moments method,'' *J. Eng.*, vol. 1, p. 11, Jan. 2017.
[31] S. Paul, A. Rajan, J. Chang, Y. C. Kuang, and M. P.-L. Ooi, ''Parametric design analysis of magnetic sensor based on model order reduction and reliability-based design optimization,'' *IEEE Trans. Magn.*, vol. 54, no. 3, pp. 1–4, 2017.
[32] A. Rajan, A. Garg, V. Vijayaraghavan, Y. C. Kuang, and M. P.-L. Ooi, ''Parameter optimization of polymer electrolyte membrane fuel cell using moment-based uncertainty evaluation technique,'' *J. Energy Storage*, vol. 15, pp. 8–16, Feb. 2018.
[33] H. Carstens, X. Xia, S. Yadavalli, and A. Rajan, ''Efficient longitudinal population survival survey sampling for the measurement and verification of lighting retrofit projects,'' *Energy Build.*, vol. 150, pp. 163–176, Sep. 2017.

ARVIND RAJAN (S'15) received the B.Eng. degree (Hons.) from Monash University Malaysia in 2014, where he is currently pursuing the Ph.D. degree in uncertainty and reliability analysis in engineering applications. He is a Student Member of the Institute of Electrical and Electronic Engineers (IEEE), the IEEE Instrumentation and Measurement Society, and the Institute of Engineering and Technology. He was a recipient of the IEEE IMS 2017 Graduate Fellowship Award.

YE CHOW KUANG (M'05–SM'15) received the B.Eng. degree (Hons.) in electromechanical engineering and the Ph.D. degree in non-invasive diagnostic techniques from the University of Southampton. He joined Monash University Malaysia in 2005, where he was involved in machine intelligence, machine vision, uncertainty modeling in engineering design, and a mass-production automation technology. He is currently an Associate Professor with the Department of Electrical and Computer Systems Engineering, Monash University Malaysia. He is also leading the Fault Tolerant Measurement Systems Committee of the IEEE Instrumentation and Measurement Society.

MELANIE PO-LEEN OOI (M'04–SM'12) received the Ph.D. degree from Monash University in 2011. She is currently the Head of Engineering, Unitec Institute of Technology, New Zealand, as well as an Associate Professor with the School of Engineering and Physical Sciences, Heriot-Watt University. Her research in electronics test technologies has led to several proposals of new testing standards and methodologies that have been adopted by the electronics industry, such as Freescale Semiconductor, Western Digital, and Texas Instruments. In addition to electronics testing, she has also proposed new measurement uncertainty theories and frameworks with application to medical measurement, and structural and mechanical systems. She is currently a member of the IEEE Instrumentation and Measurement Technical Committee on Fault Tolerant Measurement Systems and a Chartered Engineer conferred by U.K.s Institution of Engineering and Technology.

SERGE N. DEMIDENKO (M'92–SM'95–F'04) received the M.E. degree from the Belarusian State University of Informatics and Radioelectronics and the Ph.D. degree from the Institute of Engineering Cybernetics, Belarusian Academy of Sciences. He is currently a Professor, an Associate Head of the School of Engineering and Advanced Technology, and the Cluster Leader of the Electronics, Information and Communication Systems, Massey University, New Zealand. He is also holds an adjunct researcher position at Monash University Australia and a U.K. Chartered Engineer. His research interests include electronic design and test, instrumentation and measurements, and signal processing. He is a fellow of IET.

HERMAN CARSTENS received the B.Eng. degree in mechanical and aeronautical, the B.Eng. degree (Hons.) in industrial and systems, and the M.Eng. degree in electrical from the University of Pretoria (UP), in 2009, 2011, and 2014, respectively, and the Ph.D. degree in electrical engineering with the Centre for New Energy Systems, UP. He is currently involved in energy measurement and verification, and data science.