Struct Multidisc Optim (2012) 45:33–52
DOI 10.1007/s00158-011-0656-5

RESEARCH PAPER

A comparative study of probability estimation methods for reliability analysis

Zhimin Xi · Chao Hu · Byeng D. Youn

Received: 25 September 2010 / Revised: 23 February 2011 / Accepted: 26 February 2011 / Published online: 28 May 2011

Abstract In this paper we investigate the performance of probability estimation methods for reliability analysis. The probability estimation methods typically construct the probability density function (PDF) of a system response using estimated statistical moments, and then perform reliability analysis based on the approximate PDF. In recent years, a number of probability estimation methods have been proposed, such as the Pearson system, saddlepoint approximation, Maximum Entropy Principle (MEP), and Johnson system. However, no general guideline to suggest a most appropriate probability estimation method has yet been proposed. In this study, we carry out a comparative study of the four probability estimation methods so as to derive the general guidelines. Several comparison metrics are proposed to quantify the accuracy in the PDF approximation, cumulative density function (CDF) approximation and tail probability estimations (or reliability analysis). This comparative study gives an insightful guidance for selecting the most appropriate probability estimation method for reliability analysis. The four probability estimation methods are extensively tested with one mathematical and two engineering examples, each of which considers eight different combinations of the system response characteristics in terms of response boundness, skewness, and kurtosis.

Keywords Reliability analysis · Pearson system · Saddlepoint approximation · MEP · Johnson system

Z. Xi
Department of Industrial and Manufacturing Systems Engineering, University of Michigan, Dearborn, MI 48168, USA
e-mail: zxi@umd.umich.edu

C. Hu
Department of Mechanical Engineering, University of Maryland, College Park, MD 20742, USA
e-mail: huchaost@umd.edu

B. D. Youn (✉)
School of Mechanical and Aerospace Engineering, Seoul National University, Seoul, South Korea
e-mail: bdyoun@snu.ac.kr

1 Introduction

In the past few decades, uncertainty propagation and reliability analysis have been widely recognized as of great importance in engineering product development (Du and Chen 2004; Youn et al. 2005; Xie et al. 2007; McDonald and Mahadevan 2008). Hence, tremendous research advances have been made in quantifying the uncertainty of a system response originating from various uncertainty sources (e.g., material properties, loads, geometric tolerances) and analyzing the engineering reliability. In general, existing reliability analysis methods can be grouped into the following four categories: (1) the sampling-based methods such as the direct or smart Monte Carlo simulation (MCS) (Rubinstein 1981; Fu and Moses 1988), (2) the MPP-based methods such as the first- or second-order reliability method (FORM/SORM) (Hasofer and Lind 1974; Tvedt 1984); (3) the stochastic response surface methods such as the stochastic spectral method (Ghanem and Spanos 1991; Xiu and Karniadakis 2003) and stochastic collocation method (Smolyak 1963; Xiong et al. 2010), and (4) the numerical integration-based methods. Among the four categories of numerical and simulation methods, we are specifically interested in the last one in which the numerical integration is first employed to compute statistical moments, and then the probability estimation method is used to approximate the probability density functions (PDFs), probabilities and, more specifically, reliabilities of system responses based on the computed moments. Many methods have been proposed to estimate the statistical moments, such as the generalized method of moments (Hall 2005), high-dimensional model representation (Rabitz et al. 1999; Rabitz and Alis 1999), Dimension Reduction (DR) (Rahman and Xu 2004; Xu and Rahman 2004), and Eigenvector DR (EDR) (Youn et al. 2008; Youn and Xi 2009). The merits and drawbacks of these methods have also been thoroughly studied. However, the approaches for approximating the PDF of the system response from the statistical moments have not been thoroughly investigated. There is lack of a comparative study to reveal the drawbacks and merits of the probability estimation methods for reliability analysis.

It was well known how to fit parametric probability distribution functions (e.g., PDFs) to the first two statistical moments (the mean and variance). However, it is not straightforward how to construct the probability distribution functions when skewness (the 3rd statistical moments) and kurtosis (the 4th statistical moments) are additionally available. Pearson first attempted to develop a family of probability distribution types to fit the observed data with the first four arbitrary statistical moments, known as the Pearson system (Pearson 1901, 1916). Afterwards Johnson proposed the Johnson system (Johnson 1949; Johnson et al. 1994) to fit the observed data. The Johnson system involves a transformation of the raw variable into a standard normal variable (Johnson et al. 1994). The saddlepoint approximation (Daniels 1954) was introduced to statistics for approximation of a PDF. In a strictly mathematical viewpoint, the saddlepoint approximation is used to approximate the contour integral in the Fourier inversion theorem in order to recover the density or distribution function from the characteristic function (Huzurbazar 1999). In 1957 E.T. Jaynes proposed a rule to assign numerical values to probabilities by maximizing an information entropy subject to the constraints of the information, known as the maximum entropy principle (MEP) (Jaynes 1957).

In recent years, these methods have been employed to approximate the PDF, probabilities, and reliabilities of system responses from the known statistical information (Farnum 1997; Chen and Kamburowska 2001; Soize 2001; Huang and Du 2006; Straeten and Beck 2008; Youn and Xi 2009). However, no general guidelines to suggest a most appropriate probability estimation method have yet been developed. In this study, we carry out a comparative study of the four probability estimation methods so that the general guidelines can be extracted. Several comparison metrics are proposed to quantify the accuracy in the probability density function (PDF) approximation, cumulative density function (CDF) approximation and tail probability estimations (or reliability analysis). This comparative study gives the general guidelines for selecting the most appropriate probability estimation method for reliability analysis. The four probability estimation methods are extensively tested with one mathematical and two engineering examples, each of which considers eight different combinations of the system response characteristics in terms of response boundness, skewness, and kurtosis.

The rest of the paper is organized as follows. Section 2 presents a brief review of these probability estimation methods. In Section 3, four comparison metrics are proposed to quantify the accuracy of the probability estimation methods. An extensive comparative study is carried out in Section 4 to reveal the relative merits and drawbacks of these methods. The capability of approximating a bimodal PDF is also discussed. The comparison results are briefly summarized in Section 5.

2 Review of probability estimation methods

This section briefly reviews the four commonly used probability estimation methods, namely the Pearson system, saddlepoint approximation, Maximum Entropy Principle (MEP), and Johnson system.

2.1 Pearson system

Pearson system (Pearson 1901, 1916) can be used to construct the PDF of a random response y based on its first four central moments (mean, standard deviation, skewness and kurtosis). Detailed expressions of the PDF ($p(y)$) can be achieved by solving a differential equation as

$$\frac{1}{p(y)}\frac{dp(y)}{dy} = -\frac{a+y}{c_0 + c_1 y + c_2 y^2} \tag{1}$$

where a, c_0, c_1 and c_2 are the four coefficients determined by the first four moments of the random response y and expressed as

$$\begin{cases} c_0 = \left(4\beta_2 - 3\beta_1^2\right)\left(10\beta_2 - 12\beta_1^2 - 18\right)^{-1}\mu_2 \\ c_1 = \beta_1(\beta_2+3)\left(10\beta_2 - 12\beta_1^2 - 18\right)^{-1}\sqrt{\mu_2} \\ c_2 = \left(2\beta_2 - 3\beta_1^2 - 6\right)\left(10\beta_2 - 12\beta_1^2 - 18\right)^{-1} \end{cases} \tag{2}$$

where β_1 is the skewness, β_2 is the kurtosis, and μ_2 is the variance. The mean value is always treated as zero in the Pearson System since it can be easily shifted to the true mean value once the differential equation is solved. Basically, the differential equation can be solved based on different criteria of the three coefficients c_0, c_1, and c_2 as

Table 1 PDFs of Pearson system

Normal	Criteria	$c_1 = c_2 = 0$
	PDF	$p(y) = K \exp\left[-(y+a)^2/(2c_0)\right]$, where $y \in (-\infty, \infty)$
Type I and II	Criteria	$c_1^2 - 4c_0c_2 > 0$ and roots (a_1, a_2) of $c_0 + c_1y + c_2y^2 = 0$ satisfy $a_1 < 0 < a_2$
	PDF	$p(y) = K(y - a_1)^{m_1}(a_2 - y)^{m_2}$, where $m_1 = \frac{a+a_1}{c_2(a_2-a_1)}$, $m_2 = \frac{a+a_2}{c_2(a_1-a_2)}$, $y \in [a_1, a_2]$
Type III	Criteria	$c_2 = 0$, $c_1 \neq 0$
	PDF	$p(y) = K(c_0 + c_1y)^m \exp\left(\frac{-y}{c_1}\right)$, where $m = c_1^{-1}\left(c_0c_1^{-1} - a\right)$ If $c_1 > 0$, $y \in [-c_0/c_1, \infty)$; If $c_1 < 0$, $y \in (\infty, c_0/c_1]$
Type IV	Criteria	$c_1^2 - 4c_0c_2 < 0$
	PDF	$p(y) = K[C_0 + c_2(y + C_1)^2]^{-(2c_2)^{-1}} \exp\left[-\frac{a-C_1}{\sqrt{c_2C_0}} \tan^{-1} \frac{y+C_1}{\sqrt{C_0/c_2}}\right]$ where $C_0 = c_0 - c_1^2c_2^{-1}/4$, $C_1 = c_1c_2^{-1}/2$, $y \in (-\infty, \infty)$
Type V	Criteria	$c_1^2 = 4c_0c_2$
	PDF	$p(y) = K(y + C_2)^{-1/c_2} \exp\left[\frac{a-C_2}{c_2(y+C_2)}\right]$, where $C_2 = c_1/(2c_2)$ If $(a - C_2)/c_2 < 0$, $y \in [-C_2, \infty)$; If $(a - C_2)/c_2 > 0$, $y \in (\infty, C_2]$
Type VI	Criteria	roots (a_1, a_2) of $c_0 + c_1y + c_2y^2 = 0$ are real and of the same sign
	PDF	If $a_1 < a_2 < 0$, $p(y) = K(y - a_1)^{m_1}(y - a_2)^{m_2}$ where $y \in [a_2, \infty)$ If $a_1 > a_2 > 0$, $p(y) = K(a_1 - y)^{m_1}(a_2 - y)^{m_2}$ where $y \in (\infty, a_2]$
Type VII	Criteria	$c_1 = 0$, $c_0 > 0$, $c_2 > 0$
	PDF	$p(y) = K(c_0 + c_2y^2)^{-(2c_2)^{-1}}$, where $y \in (-\infty, \infty)$

shown in Table 1. The coefficient K in Table 1 is numerically calculated by satisfying the property that the integral of the PDF over the integration domain equals one.

In general, the PDF can be successfully constructed based on the first four moments of the random response. However, the Pearson system can fail to construct the PDF, especially when the statistical moments in the Pearson curve fall into the region that several distribution types merge, as shown in Fig. 1. The solid dots stand for the locations having an instability problem while constructing the PDF.

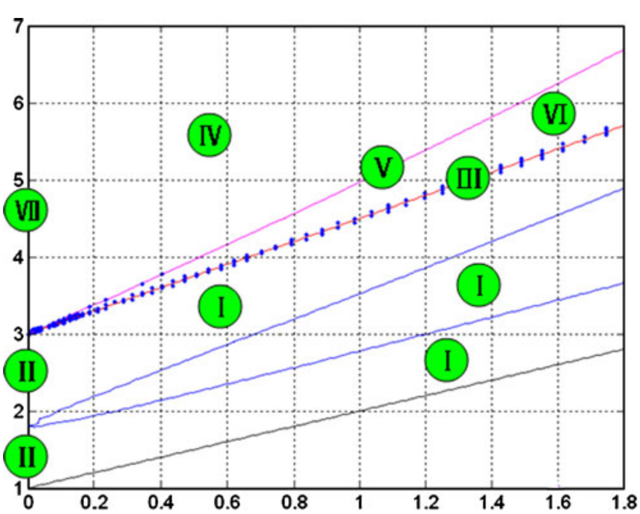


Fig. 1 Pearson curve (x-axis is the square of skewness, β_1^2, and y-axis is the kurtosis, β_2)

The numerical instability may occur while computing the coefficient K of a specific distribution type. In the EDR method (Youn et al. 2008; Youn and Xi 2009), a stabilized Pearson system was proposed to avoid the instability by generating two hyper-PDFs with slightly adjusted kurtosis values. These two hyper-PDFs were finally used to approximate the PDF with the original statistical moments. In this paper, the stabilized Pearson system is employed for the comparison study.

2.2 Saddlepoint approximation

For a PDF $p(y)$, the moment generation function is defined as

$$M(t) = \int_{\infty}^{+\infty} e^{ty} p(y) dy \tag{3}$$

The density function $p(y)$ can be obtained through the inverse Fourier transform (Daniels 1954) as

$$\begin{aligned} p(y) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} M(it) e^{-ity} dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{K(it) - ity} dt \end{aligned} \tag{4}$$

where $K(t)$ is the Cumulant Generating Function (CGF) and $K(t) = ln[M(t)]$. By replacing *it* with τ, the PDF $p(y)$ can be rewritten as

$$p(y) = \frac{1}{2\pi i} \int_{-i\infty}^{+i\infty} e^{K(\tau) - \tau y} d\tau \tag{5}$$

We then apply the Taylor series expansion to the exponent function $f(\tau) = K(\tau) - \tau y$ at the saddlepoint τ_s where $f'(\tau_s) = 0$ and obtain an approximation of $f(\tau)$ as

$$\begin{aligned} f(\tau) &\approx f(\tau_s) + \frac{(\tau - \tau_s)^2}{2} f''(\tau_s) \\ &= K(\tau_s) - \tau_s y + \frac{(\tau - \tau_s)^2}{2} K''(\tau_s) \end{aligned} \tag{6}$$

Substituting (6) into (5) gives the saddlepoint approximation of the PDF $p(y)$ as

$$p(y) \approx \left[\frac{1}{2\pi K''(\tau_s)}\right]^{1/2} \exp\left[K(\tau_s) - \tau_s y\right] \tag{7}$$

It has been proved that the expansion at the saddlepoint is a proper asymptotic expansion with the steepest descent method (Daniels 1954). An impressive accuracy of the saddlepoint approximation was observed from previous works (Goutis and Casella 1999; Huzurbazar 1999) under the condition that the CGF of the response y is exactly known. However, it is impossible to obtain the exact CGF of the system response y in general. Instead the CGF can be approximated by a few statistical moments as

$$K_Y(t) = \sum_{j=1}^{N} \kappa_j \frac{t^j}{j!} \tag{8}$$

where κ_j is the jth cumulant and the first four cumulants can be expressed in terms of the central moments as

$$\begin{cases} \kappa_1 = \mu \\ \kappa_2 = \mu_2 \\ \kappa_3 = \mu_3 \\ \kappa_4 = \mu_4 - 3\mu_2^2 \end{cases} \tag{9}$$

where μ is the mean of the system response y and μ_n is the nth central moment, defined as

$$\mu_n = \int_{-\infty}^{+\infty} (y - \mu)^n p(y) dy \tag{10}$$

It is noted that, in addition to using statistical moments, it is also possible to use the first-order Taylor expansion of the performance function at the most likelihood point to estimate the CGFs of the performance function (Du and Sudjianto 2004). For the purpose of consistency and thus fairness in this comparative study, the CGF is approximated using the first four statistical moments in this paper.

2.3 Maximum entropy principle (MEP)

In 1957 E.T. Jaynes proposed a rule to assign numerical values to probabilities in circumstances where certain partial information is available. Today this rule, known as the maximum entropy principle (MEP) (Jaynes 1957), has been used in many fields. The PDF of the system response y can be approximated by maximizing the entropy subject to the known information such as the statistical moments. The problem can be formulated as

$$\begin{aligned} &\text{Maximize} \quad f = -\int_{b_1}^{b_2} p(y) \log p(y) dy \\ &\text{Subject to} \quad p(y) \geq 0, \ \int_{b_1}^{b_2} p(y) dy = 1, \\ &\qquad\qquad \int_{b_1}^{b_2} y^j p(y) dy = \mu_j' \end{aligned} \tag{11}$$

where u_j' is the jth known raw moment of system response, and b_1 and b_2 are the lower and upper bounds of the system response y, respectively.

The optimization problem has the solution of the following form

$$p(y) = \exp\left[-\lambda_0 - \sum\nolimits_{j=1}^{N} \lambda_j y^j\right], \ \ y \in [a, b] \tag{12}$$

where λ_j is the Lagrange multiplier and N is the total number of known raw moments. The Lagrange multipliers can be obtained by solving a set of nonlinear equations defined in (11). Although this method is straightforward, it is relatively difficult to find the optimum numerical solution when more than four nonlinear equations are involved. A more efficient method is suggested by introducing a potential function (Mead and Papanicolaou 1984) defined as

$$\begin{aligned} Y &= \lambda_0 + \sum_{j=1}^{N} \mu_j' \lambda_j \\ &= \ln\left[\int_{b_1}^{b_2} \exp\left(-\sum_{j=1}^{N} \lambda_j y^j\right) dy\right] + \sum_{j=1}^{N} \mu_j' \lambda_j \end{aligned} \tag{13}$$

The stationary points of (13) satisfy the equality constraints defined in (11) and they can be solved efficiently due to the convexity and positive definiteness of the Hessian matrix of the potential function. One numerical algorithm that can be employed to find the stationary points is the so-called iterative Newton algorithm, of which the formulation at the kth iteration can be expressed as

$$\boldsymbol{\lambda}^{[k+1]} = \boldsymbol{\lambda}^{[k]} - (\mathbf{H}^{-1})^{[k]} \left[\boldsymbol{\mu}' - \langle \boldsymbol{\mu} \rangle^{[k]}\right] \tag{14}$$

where $\boldsymbol{\lambda}^{[k]}$ is the vector consisting of Q Lagrange multipliers at the kth iteration; $\boldsymbol{\mu}'$ is the vector consisting of known raw moments; $\langle \boldsymbol{\mu} \rangle^{[k]}$ is the vector of calculated raw moments with the kth Lagrange multiplier; and $\mathbf{H}$ is the Hessian matrix.

2.4 Johnson system

Johnson system (Johnson 1949; Johnson et al. 1994) was proposed to fit a set of data using a mathematical transformation function. The system contains three families of distributions: lognormal (S_L), bounded (S_B), and unbounded (S_U) distributions. Johnson proposed the following transformation functions for converting them into standard normal distributions as

$$\begin{cases} z=\gamma+\delta\ln\left(\dfrac{y-\upsilon}{\lambda}\right) & (y\geq\upsilon) \quad \textit{for } S_L \\ z=\gamma+\delta\ln\left(\dfrac{y-\upsilon}{\lambda+\upsilon-y}\right) & (\upsilon\leq y\leq\upsilon+\lambda) \quad \textit{for } S_B \\ z=\gamma+\delta\sinh^{-1}\left(\dfrac{y-\upsilon}{\lambda}\right) & (-\infty<y<+\infty) \quad \textit{for } S_U \end{cases} \tag{15}$$

where υ and λ are the location and scale parameters, respectively; γ and δ are the shape parameters; $\delta > 0$, $\lambda > 0$, $-\infty < \gamma < +\infty$, and $-\infty < \upsilon < +\infty$. For the sake of completeness, an identity transformation for the normal distribution is defined as

$$z=\gamma+\delta\left(\frac{y-\upsilon}{\lambda}\right) \quad (-\infty < y < +\infty) \quad \textit{for } S_N \tag{16}$$

In using the Johnson system, the first step is to determine the distribution family that should be used. With the known skewness and kurtosis of the system response, the distribution family can be directly decided from the Johnson curve (Johnson 1949). The PDF of the system response can then be formulated, depending on the determined distribution family, as

$$p(y)=\begin{cases} \dfrac{\delta}{(2\pi)^{1/2}}\dfrac{1}{y-\upsilon}\exp\left\{-\dfrac{1}{2}\left[\gamma+\delta\ln\left(\dfrac{y-\upsilon}{\lambda}\right)\right]^2\right\} & (y\geq\upsilon) \quad \textit{for } S_L \\ \dfrac{\delta}{(2\pi)^{1/2}}\dfrac{\lambda}{(y-\upsilon)(\lambda+\upsilon-y)}\exp\left\{-\dfrac{1}{2}\left[\gamma+\delta\ln\left(\dfrac{y-\upsilon}{\lambda+\upsilon-y}\right)\right]^2\right\} & (\upsilon\leq y\leq\upsilon+\lambda) \quad \textit{for } S_B \\ \dfrac{\delta}{(2\pi)^{1/2}}\dfrac{1}{\lambda\sqrt{\left(\dfrac{y-\upsilon}{\lambda}\right)^2+1}}\exp\left\{-\dfrac{1}{2}\left[\gamma+\delta\sinh^{-1}\left(\dfrac{y-\upsilon}{\lambda}\right)\right]^2\right\} & (-\infty<y<+\infty) \quad \textit{for } S_U \\ \dfrac{\delta}{(2\pi)^{1/2}}\dfrac{1}{\lambda}\exp\left\{-\dfrac{1}{2}\left[\gamma+\delta\dfrac{y-\upsilon}{\lambda}\right]^2\right\} & (-\infty<y<+\infty) \quad \textit{for } S_N \end{cases} \tag{17}$$

The unknown distribution parameters can be estimated using any of the four estimation methods, namely moment matching, percentile matching, least square estimation and minimum L_p norm estimation (DeBrota et al. 1988). To achieve consistency and thus fairness in this comparative study, we employed the moment matching method with the first four statistical moments which can typically be obtained with efficient numerical integration for reliability analysis.

2.5 Summary of probability estimation methods

Characteristics and limitations of four probability estimation methods are elaborated as following based on the above review. Both the Pearson and Johnson systems cover the entire (β_1, β_2) plane with multiple PDF formulations. Since each combination of skewness and kurtosis corresponds to a specific distribution in the systems, we would expect that the Pearson and Johnson systems consist of a wide variety of distribution types. However, they could encounter difficulties in approximating an unbounded bimodal PDF although the bounded family in Johnson system is capable of fitting a bounded bimodal PDF with a restricted shape. The saddlepoint approximation is able to approximate PDFs of system outputs using more than four statistical moments. However, this method suffers from the following two drawbacks: (1) numerical instability could be encountered in certain cases (e.g., with a negative cumulant κ_4); and (2) the inherent exponential form, while ensuring the smoothness in the approximate PDF, could lead to the impossibility to accurately estimate a PDF with a non-exponential behavior. Similar to the saddlepoint approximation, the MEP method is capable of using more than four statistical moments to increase the accuracy in representing a bimodal or multi-mode PDF. However, the inherent exponential form taken by the MEP method may limit its capability to accurately represent a PDF with a non-exponential behavior. In the subsequent sections, some of these relative merits and drawbacks will be verified using the predefined comparison metrics with numerical and engineering examples.

3 Comparison of probability estimation methods

In this section, four comparison metrics $\boldsymbol{\zeta}$, namely, the PDF-based metric ζ_1, CDF-based metrics ζ_2 and ζ_3, and reliability accuracy metric ζ_4, are proposed to quantitatively assess the performance of the probability estimation methods. The PDF-based metric ζ_1 is the cross entropy (Kullback and Leibler 1951), while the CDF-based metrics ζ_2 and ζ_3 are the U-pooling (Ferson et al. 2008; Xiong et al. 2008) and area metric, respectively. It is possible that a given probability estimation method may perform poorly under one comparison metric while outperforming many other methods under the other comparison metrics. Thus the unique characteristics of each comparison metric will be discussed to extract general guidelines for choosing appropriate methods for specific engineering applications. We note that, in the numerical integration-based methods, the estimation of statistical moments using the numerical integration (the 1st step) requires the evaluation of system responses and might be computationally expensive while, upon the completion of the moment estimation, the approximation of response PDFs employing any of the four probability estimation methods (the 2nd step) can be done in a very efficient manner. Typically, the CPU time for this approximation is less than 1 s because the computation is conducted through the evaluation of explicit mathematical PDF functions. Therefore, we intend not to investigate the efficiency of these four methods in this comparative study.

3.1 PDF-based metric: cross entropy ζ_1

The cross entropy (Kullback and Leibler 1951), also called Kullback–Leibler (KL) distance, was proposed to measure the similarity between a true PDF and an estimated PDF. The smaller the expected cross entropy, the higher degree of similarity is the approximate PDF to the true PDF. Let $p(y)$ and $\hat{p}(y)$ denote the true and approximate PDFs, respectively, with regard to a random system response variable y. The cross entropy or KL distance is defined as

$$\zeta_1(p, \hat{p}) = \int p(y) \cdot \ln\left[\frac{p(y)}{\hat{p}(y)}\right] dy \tag{18}$$

Based on the concept of Shannon's entropy, (18) computes the difference in the expected information between two distributions as

$$\zeta_1(p, \hat{p}) = E_p\left[\ln(p)\right] - E_p\left[\ln(\hat{p})\right] \tag{19}$$

where the random system response y is omitted for clarity. It should be noted that ζ_1 is not a physical distance between p and $\hat{p}$ in the common sense, since the KL distance is not associative in general, i.e., $\zeta_1(\hat{p}, p) \neq \zeta_1(p, \hat{p})$. Nevertheless, ζ_1 is an information-theoretic distance measure between two different distributions.

3.2 CDF-based metrics: U-pooling ζ_2 and area metric ζ_3

The U-pooling (Ferson et al. 2008; Xiong et al. 2008) value is defined as the area difference between the actual and approximate CDF in the normalized space as shown in Fig. 2. The smaller the shaded area, the higher the overall accuracy of the approximate CDF. The U-pooling value is calculated as

$$\zeta_2(\hat{P}, P) = \int_0^1 \left|\hat{P}(u) - P(u)\right| du \tag{20}$$

where $\hat{P}(u)$ and $P(u)$ are the approximate CDF and actual CDF, respectively.

Similar to the U-pooling, the area metric also measures the area between the approximate and true CDFs and can be expressed as (Ferson et al. 2008)

$$\zeta_3(\hat{P}, P) = \int_{-\infty}^{\infty} \left|\hat{P}(y) - P(y)\right| dy \tag{21}$$

However, the integration domain in the area metric is in the y space in contrast to the normalized space in the U-pooling.

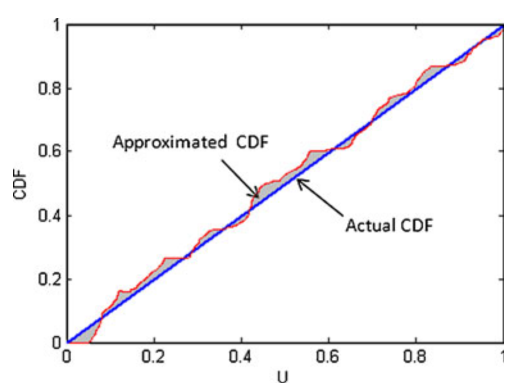


Fig. 2 Illustration of the U-pooling value

Due to the same intrinsic characteristic (area measurement), only the U-pooling ζ_2 will be utilized as a comparison criterion in this study. In contrast to the cross entropy ζ_1, which measures the similarity between the approximate and true PDFs, the U-pooling ζ_2 and area metric ζ_3 measure the area difference between the approximate and true CDFs.

3.3 Reliability accuracy metric ζ_4

While the above three comparison metrics assess the overall qualities of the PDF or CDF approximations, the accuracy of reliability estimation, as an important aspect in reliability analysis, measures the accuracy of pointwise CDF approximations and should also be considered as a comparison metric. Given a limit state value y_C, the reliability can be defined as

$$R(p|y_C) = \int_{-\infty}^{y_C} p(y)dy = P(y_C) \tag{22}$$

The reliability accuracy ζ_4 is defined as the absolute difference between the estimated and true reliability

$$\zeta_4(\hat{p}, p|y_C) = \left|R(\hat{p}|y_C) - R(p|y_C)\right| \tag{23}$$

The integration of the reliability accuracy in the normalized space is the U-pooling value.

3.4 Other comparison metrics

We note that, in addition to the comparison metrics detailed above, there are also several other metrics to compare the similarity of two PDFs in the literature. In what follows, we will briefly review two of them, namely the Kolmogorov–Smirnov test and the Bhattacharyya distance.

The Kolmogorov–Smirnov test uses the maximum vertical difference between two cumulative distribution functions (CDFs) as the statistic to compare the equality of two CDFs. This statistic is equal to the maximum value in the reliability accuracy metric ζ_4. The Bhattacharyya distance can be defined as

$$D = \int \sqrt{p(y)\hat{p}(y)}dy \tag{24}$$

where p and $\hat{p}$ are the true and approximate PDFs, respectively. D ranges from 0 to 1 where the value “1” stands for two identical PDFs. Both the cross entropy metric ζ_1 and Bhattacharya distance can be used to measure the similarity of two PDFs. Because of the log operator in cross entropy metric ζ_1, the former is more sensitive than the latter in identifying the error in the bound definition of the approximate PDF.

3.5 Summary of comparison metrics

The proposed comparison metrics offer different perspectives on the extent to which a probability estimation method can accurately represent a response PDF. In what follows, we intend to present three remarks regarding the differences and connections between these metrics. Firstly, as mentioned before, the PDF- and CDF-based metrics (i.e., cross entropy ζ_1, U-pooling ζ_2 and area metric ζ_3) measure the overall qualities of the PDF or CDF approximations, while the reliability accuracy metric ζ_4 measures the accuracy of pointwise CDF approximations, or reliability. If we integrate the pointwise reliability accuracy metric ζ_4 in the normalized and original spaces, we then obtain the U-pooling ζ_2 and area metric ζ_3, respectively. Secondly, although both the cross entropy ζ_1 and any CDF-based metric ζ_2 or ζ_3 reflect the overall accuracy in the probability estimation, the former is strongly associated with the accuracy in the reliability estimation over the tail region of a response PDF whereas the latter, as the integration of the reliability prediction error over the entire domain, typically stands for that over the entire region. This remark will later be verified in the case study. Lastly, among these metrics, only the reliability accuracy metric ζ_4 is directly associated with reliability analysis while the others are not. However, the very motivation for using the PDF- and CDF-metrics lies in the fact that the reliability accuracy metric ζ_4 falls short in the way of achieving a comprehensive assessment on the accuracy of a probability estimation method, precisely because it only gives an integration error at a specific point, typically near the tail region.

4 Case study

A number of examples are tested below to compare the performance of the probability estimation methods introduced in Section 2. Section 4.1 presents design principles and a brief overview of case studies. Detailed descriptions of examples along with test results are given in Sections 4.2 and 4.3 for unimodal and bimodal PDFs, respectively. Section 4.4 provides a discussion on the test results.

4.1 Design of case studies

We intend to investigate the performance of the representative probability estimation methods under various PDF settings in a number of examples. For that purpose, we first employ a two-level design of experiment to derive eight cases with different PDF characteristics for a unimodal PDF, the most encountered one in reliability analysis. Then, we study several mathematical and engineering examples for each case with a unimodal PDF and for a bimodal PDF.

4.1.1 PDF characteristics of system response

The PDF of a system response can be classified in terms of three properties: (1) bounded, (2) skewed, and (3) kurtosis.

1. Bounded property: The PDF of a system response can be unbounded, one or two-side bounded. Indeed, the bounded property is often found in many PDFs. For example, both the Pearson and Johnson systems clearly define the distribution types with bounded and unbounded properties, whereas the bounded property is not defined in the saddlepoint approximation and MEP method. Hence, this study considers both bounded and unbounded cases.
2. Skewed property: Skewness is a measure of lack of symmetry. The normal distribution has a skewness of zero. An asymmetric behavior of the PDF is clearly observed when the absolute skewness value is larger than 0.5. Hence, the value of 0.5 is used to distinguish the low and high skewness level in this study.
3. Kurtosis property: Kurtosis is a measure of whether the PDF is peaked or flat relative to a normal distribution which has a kurtosis value of 3. That is, a PDF with high kurtosis tends to have a distinct peak near the mean and decline rapidly. A PDF with low kurtosis tends to have a flat top near the mean rather than a sharp peak. A uniform distribution would be the extreme case. In this study, low and high kurtosis is defined with the kurtosis value of 3.

Since a probability estimation method represents the probability density function (PDF) of a response based on its statistical moments that are affected by the above PDF characteristics (e.g., bounded, highly skewed), we expect that the estimation accuracy would highly depend on the PDF characteristics. Therefore, we devise a two-level design of experiment to investigate how the above factors affect the accuracy of probability estimation. Specifically, eight cases need to be studied as shown in Table 2. We note, however, that, in numerical integration-based reliability analysis methods, the relationship between the statistical inputs ($\mathbf{x}$) and response (y) only directly affects the accuracy in moment estimation (the first step) rather than that in probability estimation (the second step). Therefore, we intend not to investigate the effect of this functional relationship in this comparative study.

Table 2 Eight cases from two-level design of experiment

Case	Bound	Skewness	Kurtosis
1	Unbounded	Low	Low
2	Bounded	Low	Low
3	Unbounded	High	Low
4	Bounded	High	Low
5	Unbounded	Low	High
6	Bounded	Low	High
7	Unbounded	High	High
8	Bounded	High	High

4.1.2 Overview of case studies

Five examples are employed to compare the performance of the representative probability estimation methods and to extract general guidelines for selecting the most appropriate probability estimation method for reliability analysis. Table 3 summarizes these examples of which the first three are used for unimodal PDFs and the rest are used for bimodal PDFs. For examples 1, 2 and 3, PDFs derived from MCSs with 1,000,000 samples are treated as true PDFs. For example 4, we do not need to compute the true PDF since it is already given. For example 5, 20,000 MCS samples are used to construct the true PDF. Since we aim at comparing the accuracy of the probability estimation methods in approximating the PDF of a system response with its statistical moments, we intend to minimize the errors in statistical moments, thereby minimizing their effects on the comparison results. To this end, the statistical moments used for all the probability estimation methods are obtained from MCS with a sufficiently large sample size. As aforementioned in Section 3, the computation is very fast for the four probability estimation methods under study and, therefore, we intend not to compare the efficiency of these four methods in the case studies.

Table 3 Five examples for comparative studies

PDF type	Example index	System response	Case
Unimodal	Example 1	$Y = x_1 + x_2 + x_3 + x_4 + x_5$	Eight cases
	Example 2	Maximum stress in an I-beam	Eight cases
	Example 3	First buckling mode in a plate	Eight cases
Bimodal	Example 4	Response with a predefined bimodal PDF	One case
	Example 5	Power loss of a V6 gasoline engine	One case

4.2 Case studies for unimodal PDF

4.2.1 Description of examples

Three examples are employed for a unimodal PDF associated with each case in Table 2. For all the three examples, the distributions of input random parameters are appropriately assigned to make the output response meet the criteria of each case shown in Table 2. The first example is a simple addition function defined as

$$Y = x_1 + x_2 + x_3 + x_4 + x_5 \tag{25}$$

where x_i can be assigned to arbitrary distributions. The second example is an I-beam problem as shown in Fig. 3 where the maximum stress can be expressed as

$$\sigma_{\max} = \frac{Fx(L-x)h}{2LI} \tag{26}$$

where

$$I = \frac{wh^3 - (w-t_2)(h-2t_1)^3}{12} \tag{27}$$

The third example is a buckling problem. As shown in Fig. 4, the buckling problem is considered with three shape design variable: the height ($h = 500 + x_1$) and width ($w = 500+x_2$) of the plate and the diameter ($d = 100+x_3$) of the middle hole. A morphing technique in the HyperWorks 8.0 software package is used to deal with the shape variables (h, w and d) in the FEA model. The plate is modeled using plane stress quad4 elements, consisting of 1,681 nodes, 1,571 elements, and 9,798 DOF. A unit load is applied along the top edge of the plate, while the bottom edge of the plate remains fixed in all six direction. The plate is made of Aluminum 6061, where E = 67.6 GPa and $\nu = 0.3$.

4.2.2 Results and remarks

This subsection presents the probability estimation results, based on which brief remarks are drawn regarding the performance of the probability estimation methods for all eight cases.

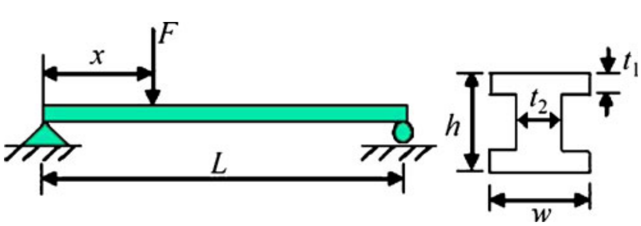


Fig. 3 Loading condition and structure of an I-beam

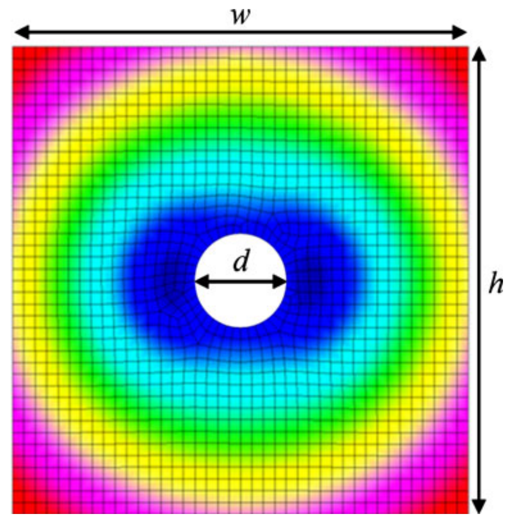


Fig. 4 Plate FE model

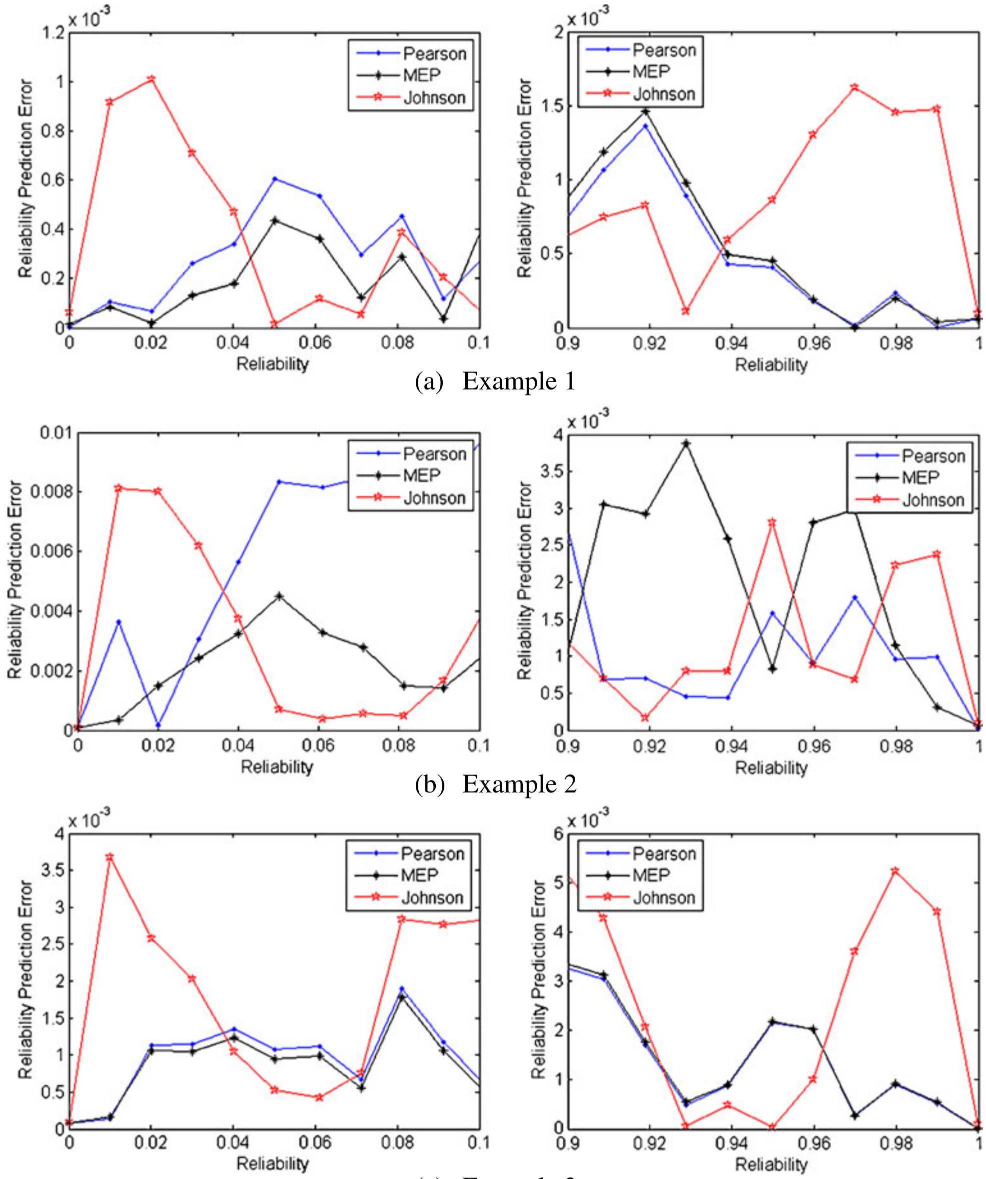
Case 1 (unbounded, low skewness and low kurtosis)

Given the distribution types and parameters shown in Table 12 in Appendix, the responses of the three examples meet the criteria of case 1. The cross entropy (ζ_1), U-pooling (ζ_2), and reliability accuracy (ζ_4) results are shown in Table 4 and Fig. 5, respectively. The bold items in Table 4 refer to the minimum cross entropy (ζ_1) or U-pooling value (ζ_2) produced by the probability estimation methods.

Johnson system presents the best overall CDF quality for all the three examples. However, it gives larger errors in reliability estimation than the Pearson and MEP for both high (>0.99) and low (<0.01) reliability levels as shown

Table 4 Cross entropy and U-pooling comparison of case 1

Example	Moments	Methods	ζ_1	ζ_2
1	$\mu = 9.7570$	Pearson	**3.58E-04**	5.08E-04
	$\sigma = 1.5417$	Saddlepoint	–	–
	$\beta_1 = 0.0417$	MEP	4.40E-04	4.96E-04
	$\beta_2 = 2.8871$	Johnson	8.81E-04	**2.34E-04**
2	$\mu = 182{,}716$	Pearson	3.24E-02	5.41E-03
	$\sigma = 27{,}072$	Saddlepoint	–	–
	$\beta_1 = 0.3887$	MEP	**5.90E-03**	3.43E-03
	$\beta_2 = 2.7114$	Johnson	4.88E-02	**3.31E-03**
3	$\mu = 3.3503$	Pearson	**4.45E-03**	2.61E-03
	$\sigma = 0.0024$	Saddlepoint	–	–
	$\beta_1 = 0.0026$	MEP	4.78E-03	2.57E-03
	$\beta_2 = 2.9079$	Johnson	1.47E-02	**2.36E-03**



(a) Example 1

(b) Example 2

(c) Example 3

Fig. 5 Reliability accuracy (ζ_4) of case 1

in Fig. 5. Based on the skewness and kurtosis of the system response, the bounded distribution (S_B) is decided from the Johnson curve (Johnson 1949) to model all three response distributions. Hence, if either a lower or upper bound is not properly defined, a relatively large error at the tail region is expected in the Johnson system.

Similar to the Johnson system, type I distribution is determined in the Pearson curve (Pearson 1916) based on the skewness and kurtosis of the system response and lower and upper bounds are defined. The three examples indicate that the Pearson system preserves better bound definition than the Johnson system since smaller reliability estimation errors are observed for both high (>0.99) and low (<0.01) reliability levels as shown in Fig. 5.

Unlike the Johnson and Pearson system, the MEP method does not have pre-defined lower and upper bounds. Hence, in the numerical implementation, it is applicable to employ a wide range, e.g., $\mu \pm 12\sigma$, where μ and σ are the response mean and standard deviation, respectively. Besides, the PDF from the MEP method is basically in the form of an exponential function which tends to behave smoothly in tail regions. These two may be the reasons that the MEP method produces accurate reliability estimations in tail regions as shown in Fig. 5.

The saddlepoint approximation fails to construct the PDF and CDF due to numerical instability in this case. This can be attributed to the small kurtosis value ($\beta_2 < 3$) which results in a negative cumulants κ_4. Hence, it is very likely that the 2nd derivative of the Cumulant Generating Function (CGF) at the saddlepoint t_s is less than zero, causing failures in the PDF and CDF constructions in (7).

Case 2 (bounded, low skewness and low kurtosis)

Given the distribution types and parameters shown in Table 13 in Appendix, the responses of three examples meet the criteria of case 2. The cross entropy (ζ_1) and U-pooling (ζ_2) comparison results are shown in Table 5. The reliability accuracy (ζ_4) results are not shown in this case because they are very similar as in case 1. The Johnson system shows very good accuracy in terms of the overall CDF approximation. The Pearson system preserves better bound definition than the Johnson system. It is also the main reason that the Pearson system shows better overall CDF quality than the Johnson system in the 1st example. The MEP method performs the best in all aspects in the 2nd example where the skewness ($\beta_1 = 0.4755$) is relatively high. Again, the saddlepoint approximation fails to construct the PDF and CDF due to the numerical instability problem explained in case 1.

Remarks on case 1 and case 2

As a summary for case 1 and case 2, the following facts were observed: (1) the Johnson and Pearson systems may produce inaccurate bound prediction of a PDF with a relatively steep tail region (high skewness), resulting in relatively large reliability errors; (2) the MEP method generally outperforms the other methods in all aspects for a PDF with relatively high skewness; (3) the Pearson system typically produces better reliability accuracy than the Johnson system for the high (>0.99) and low (<0.01) reliability levels; (4) the Johnson system generally produces better reliability accuracy than the other methods at all reliability levels except the high (>0.99) and low (<0.01) levels; and (5) the saddlepoint approximation fails to construct the PDF and CDF due to the numerical instability.

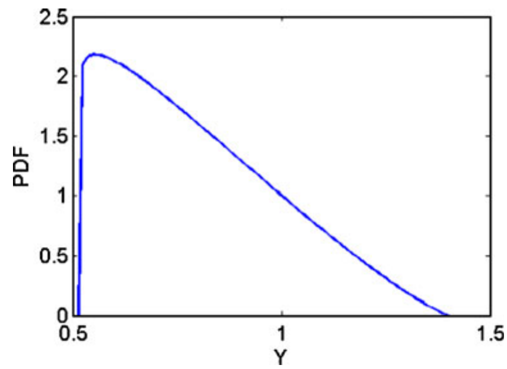


Fig. 6 PDF with high skewness and low kurtosis

Remarks on case 3 and case 4

It is difficult to build a response PDF to meet the criteria of case 3 and case 4. To meet the criteria of high skewness and low kurtosis, the PDF should have the form as shown in Fig. 6. Although it may be possible to have a system random input with such property, i.e. skewed triangular distribution in some engineering application, it is not likely that the system response would keep the same form after uncertainty propagation. Besides, this type of distribution is close to the impossible PDF region in the Pearson and Johnson system and has a very small probability of occurrence. Due to the above reasons, case 3 and case 4 are not studied.

Table 5 Cross entropy and U-pooling comparison of case 2

Example	Moments	Methods	ζ_1	ζ_2
1	$\mu = 3.9275$	Pearson	1.23E-03	**5.13E-04**
	$\sigma = 1.0970$	Saddlepoint	–	–
	$\beta_1 = -0.0028$	MEP	1.25E-03	1.99E-03
	$\beta_2 = 2.4348$	Johnson	**1.13E-03**	6.06E-04
2	$\mu = 165,752$	Pearson	2.20E-02	5.71E-03
	$\sigma = 35,384$	Saddlepoint	–	–
	$\beta_1 = 0.4755$	MEP	**1.37E-03**	**1.76E-03**
	$\beta_2 = 2.9206$	Johnson	5.10E-03	3.07E-03
3	$\mu = 3.3421$	Pearson	**9.37E-04**	1.00E-03
	$\sigma = 0.0027$	Saddlepoint	–	–
	$\beta_1 = 0.0310$	MEP	1.38E-03	1.51E-03
	$\beta_2 = 2.4757$	Johnson	1.25E-03	**5.36E-04**

Table 6 Cross entropy and U-pooling comparison of case 5

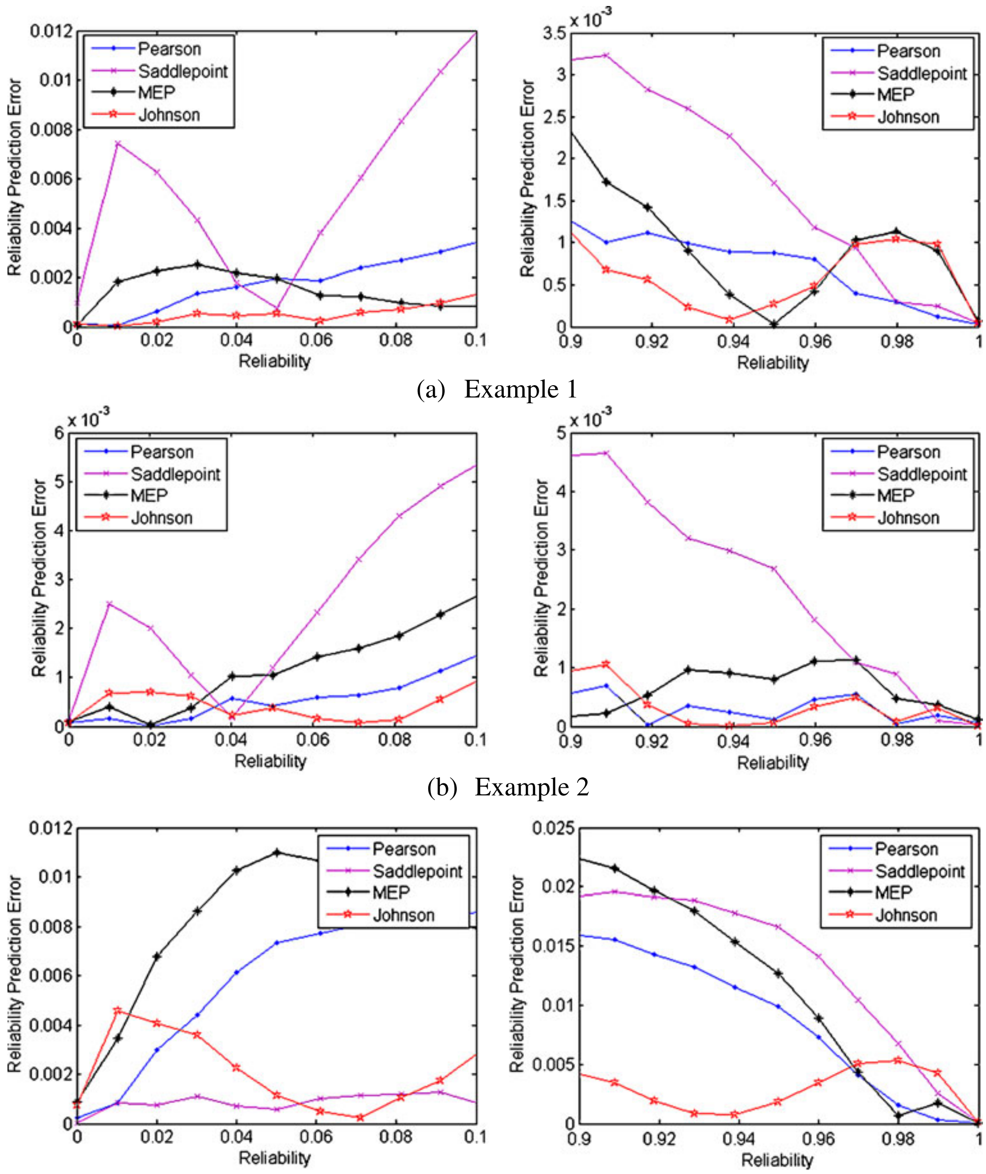
Example	Moments	Methods	ζ_1	ζ_2
1	$\mu = 5.3114$	Pearson	1.52E-03	1.64E-03
	$\sigma = 3.6903$	Saddlepoint	5.64E-02	7.32E-03
	$\beta_1 = -0.0178$	MEP	1.12E-03	2.31E-03
	$\beta_2 = 4.1201$	Johnson	**9.72E-04**	**6.69E-04**
2	$\mu = 165,592$	Pearson	1.12E-03	6.17E-04
	$\sigma = 19,878$	Saddlepoint	3.88E-02	3.29E-03
	$\beta_1 = 0.4573$	MEP	1.35E-03	1.40E-03
	$\beta_2 = 3.6591$	Johnson	**1.03E-03**	**5.73E-04**
3	$\mu = 3.3480$	Pearson	7.60E-03	1.27E-02
	$\sigma = 0.0102$	Saddlepoint	8.34E-02	1.21E-02
	$\beta_1 = -0.1775$	MEP	1.46E-02	1.69E-02
	$\beta_2 = 4.6057$	Johnson	**4.69E-03**	**4.42E-03**

Case 5 (unbounded, low skewness and high kurtosis)

Given the distribution types and parameters shown in Table 14 in Appendix, the responses of three examples meet the criteria of case 5. The cross-entropy (ζ_1), U-pooling (ζ_2), and reliability accuracy (ζ_4) results are shown in Table 6 and Fig. 7, respectively. The Johnson system shows the best overall PDF and CDF quality for all three examples. The Pearson system shows better accuracy than the Johnson system in reliability estimations for the high (>0.99) and low (<0.01) reliability levels in all three examples as shown in Fig. 7. It indicates that the Pearson system still preserves better bound definition than the Johnson system in this case. Although the saddlepoint approximation presents inaccurate overall PDF and CDF approximation, it has the best reliability accuracy at the left tail region of the 3rd example as shown in Fig. 7c. It is noticed that the 3rd example has the longest left tail among all three examples due to the properties of the skewness and kurtosis. The saddlepoint approximation presents the best reliability accuracy at the long tail region. The MEP method does not outperform the other methods in terms of both overall PDF and CDF quality and reliability accuracy at tail regions.

Case 6 (bounded, low skewness and high kurtosis)

Given the distribution types and parameters shown in Table 15 in Appendix, the responses of three examples meet



(a) Example 1

(b) Example 2

(c) Example 3

Fig. 7 Reliability accuracy (ζ_4) of case 5

Table 7 Cross entropy and U-pooling comparison of case 6

Example	Moments	Methods	ζ_1	ζ_2
1	$\mu = 2.2301$	Pearson	2.82E-02	2.19E-02
	$\sigma = 0.1946$	Saddlepoint	1.03E-01	1.79E-02
	$\beta_1 = -0.3711$	MEP	4.56E-02	2.79E-02
	$\beta_2 = 4.7417$	Johnson	**1.69E-02**	**6.51E-03**
2	$\mu = 168{,}945$	Pearson	**7.60E-04**	9.15E-04
	$\sigma = 33{,}429$	Saddlepoint	2.55E-02	2.98E-03
	$\beta_1 = 0.4594$	MEP	7.97E-04	1.14E-03
	$\beta_2 = 3.4032$	Johnson	8.45E-04	**4.90E-04**
3	$\mu = 3.3513$	Pearson	**1.28E-03**	2.93E-03
	$\sigma = 0.0006$	Saddlepoint	4.37E-02	2.37E-03
	$\beta_1 = -0.2066$	MEP	3.46E-03	5.75E-03
	$\beta_2 = 3.6638$	Johnson	1.32E-03	**9.65E-04**

Table 8 Cross entropy and U-pooling comparison of case 7

Example	Moments	Methods	ζ_1	ζ_2
1	$\mu = 19.9800$	Pearson	1.36E-01	1.29E-02
	$\sigma = 8.5225$	Saddlepoint	2.16E-01	3.56E-02
	$\beta_1 = 1.6336$	MEP	8.51E-02	1.75E-02
	$\beta_2 = 7.6582$	Johnson	**4.36E-03**	**3.78E-03**
2	$\mu = 189{,}591$	Pearson	**7.80E-04**	1.10E-03
	$\sigma = 35{,}342$	Saddlepoint	6.97E-02	9.50E-03
	$\beta_1 = 0.7341$	MEP	1.05E-03	3.25E-03
	$\beta_2 = 4.0098$	Johnson	9.28E-04	**9.76E-04**
3	$\mu = 3.3376$	Pearson	2.82E-04	5.56E-04
	$\sigma = 0.0045$	Saddlepoint	9.10E-02	1.17E-02
	$\beta_1 = -0.8785$	MEP	7.41E-03	3.70E-03
	$\beta_2 = 4.4598$	Johnson	**2.32E-04**	**4.72E-04**

the criteria of case 6. The cross-entropy (ζ_1) and U-pooling (ζ_2) comparison results are shown in Table 7. The reliability accuracy (ζ_4) results are not shown in this case because they are very similar as in case 5. The Johnson system shows the best overall CDF quality for all three examples. The Pearson system presents better accuracy in reliability estimation than the Johnson system for both high (>0.99) and low (<0.01) reliability levels. The same as in case 5, the saddlepoint approximation shows the best reliability accuracy in the long tail region in the 1st example. Again, the MEP method does not show advantages over other methods in this case.

Remarks on case 5 and case 6

As a summary for case 5 and case 6, the following facts were observed: (1) unbounded distribution types are employed for both the Pearson and Johnson systems, regardless of the bound property; (2) the Johnson system shows the best overall CDF quality for all six examples; (3) the Pearson system produces better reliability accuracy than the Johnson system for the high (>0.99) and low (<0.01) reliability levels; and (4) the saddlepoint approximation shows the best reliability accuracy in a long tail region where the kurtosis value is typically larger than 4 in a skewed PDF.

Case 7 (unbounded, high skewness and high kurtosis)

Given the distribution types and parameters as shown in Table 16 in Appendix, the responses of three examples meet the criteria of case 7. The cross-entropy (ζ_1), U-pooling (ζ_2), and reliability accuracy (ζ_4) results are shown in Table 8 and Fig. 8, respectively. The Johnson system shows the best overall CDF quality for all three examples. It also shows very good reliability accuracy in the tail region. The Pearson system presents comparable accuracy with the Johnson system except when the PDF is too much skewed as shown in the 1st example in Fig. 8. The MEP method and saddlepoint approximation do not present advantages over the Johnson and Pearson system in this case.

Case 8 (bounded, high skewness and high kurtosis)

Given the distribution types and parameters shown in Table 17 in Appendix, the responses of three examples meet the criteria of case 8. The cross entropy (ζ_1) and U-pooling (ζ_2) comparison results are shown in Table 9. The reliability accuracy (ζ_4) results are not shown in this case because they are very similar as in case 7. Again, the Johnson system shows the best overall CDF quality for all three examples. Similar to case 7, it also shows very good reliability accuracy in the tail region. Hence, it is recommended to use the Johnson system for reliability estimation in case 8.

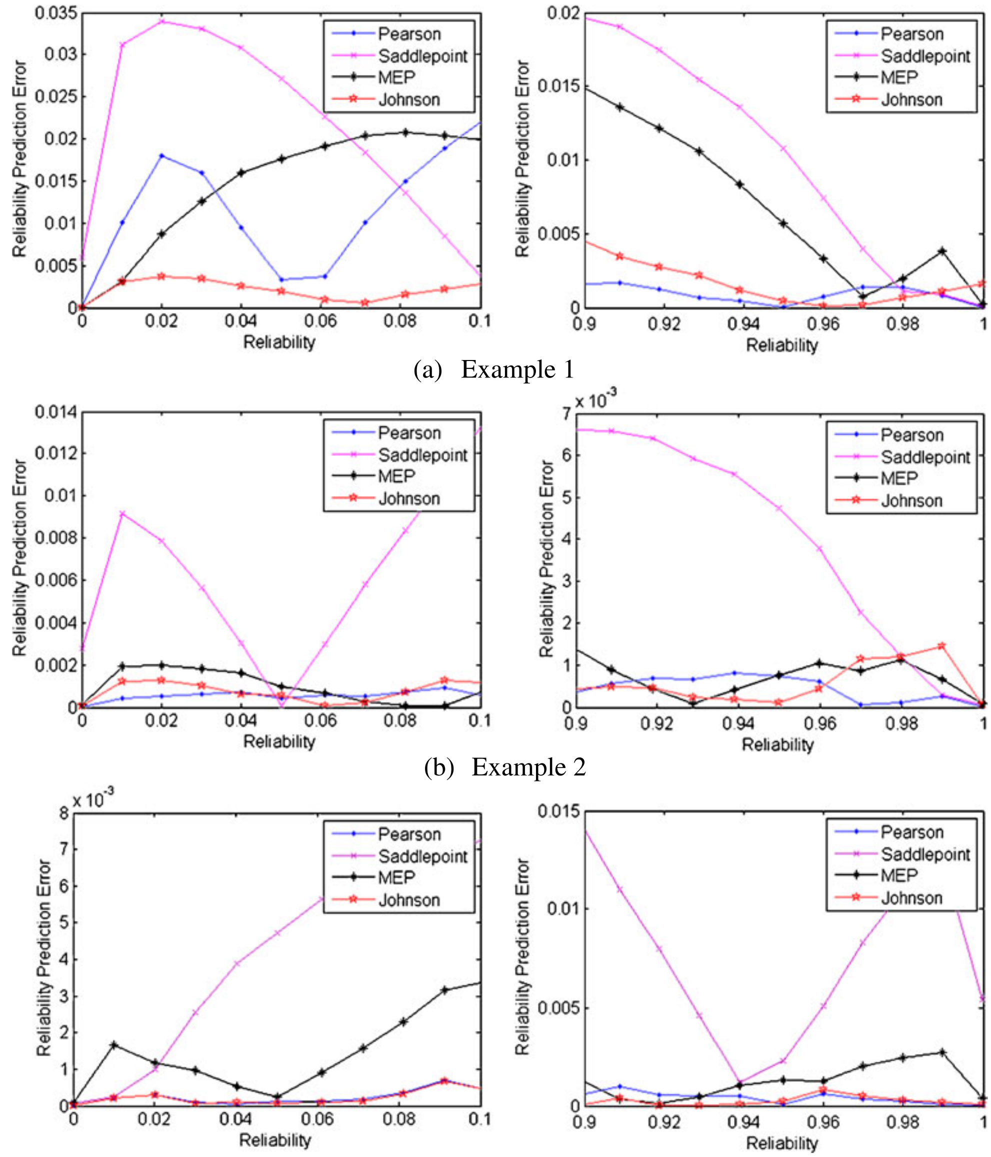
Remarks on case 7 and case 8

As a summary for case 7 and case 8, the following facts were observed: (1) the unbounded and bounded PDFs are employed in the Johnson and Pearson systems for the probability estimations, respectively, regardless of the bound property; (2) the Pearson system shows relatively large reliability errors when the PDF is highly skewed; (3) the Johnson system shows the best overall CDF quality for all six examples and also presents very good reliability accuracy in the tail regions.

4.3 Case studies for bimodal PDF

The multi-modal response PDF, especially the bimodal PDF, can also be observed in many engineering applications.

Fig. 8 Reliability accuracy (ζ_4) of case 7



(a) Example 1

(b) Example 2

(c) Example 3

Table 9 Cross entropy and U-pooling comparison of case 8

Example	Moments	Methods	ζ_1	ζ_2
1	$\mu = 2.5131$	Pearson	5.84E-01	5.87E-02
	$\sigma = 0.1872$	Saddlepoint	2.62E-01	2.84E-02
	$\beta_1 = -1.8446$	MEP	1.71E-01	4.24E-02
	$\beta_2 = 8.2679$	Johnson	**3.33E-02**	**1.34E-02**
2	$\mu = 184{,}821$	Pearson	8.25E-04	1.11E-03
	$\sigma = 39{,}965$	Saddlepoint	4.46E-02	4.92E-03
	$\beta_1 = 0.6179$	MEP	**8.17E-04**	6.74E-04
	$\beta_2 = 3.7742$	Johnson	1.17E-03	**6.05E-04**
3	$\mu = 3.3507$	Pearson	3.05E-03	3.71E-03
	$\sigma = 0.0008$	Saddlepoint	4.09E-02	4.41E-03
	$\beta_1 = -0.5383$	MEP	3.16E-03	3.81E-03
	$\beta_2 = 3.7229$	Johnson	**1.06E-03**	**8.85E-04**

In this section, the accuracy of the bimodal probability estimation using the four methods is studied.

4.3.1 Description of examples

Two examples are employed to create the bimodal PDF. The first is a mathematical example to create a bimodal PDF by mixing two normal PDFs as

$$f_X(x) = W f_{X_1}(x_1) + (1 - W) f_{X_2}(x_2) \tag{28}$$

where $X_1 \sim \text{Normal}(0, 1)$, $X_2 \sim \text{Normal}(3, 1)$, and the weight factor W is set as 0.4.

The second example is a V6 gasoline engine problem (Chan et al. 2004). The V6 gasoline engine can be

Table 10 Input parameters for V6 gasoline engine problem

Input parameters	Type	Mean	Std. dev.
Ring surface roughness, [μm]	Normal	4.000	1.000
Liner surface roughness, [μm]	Normal	6.199	1.000
Liner Young's modulus, [GPa]	Normal	0.800	0.040
Liner hardness, [BHV]	Normal	2.400	0.120

Table 11 Cross entropy and U-pooling comparison of bimodal PDFs

Example	Moments	Methods	ζ_1	ζ_2
1	$\mu = 1.8000$	Pearson	1.06E-00	1.57E-02
	$\sigma = 1.7781$	Saddlepoint	–	–
	$\beta_1 = -0.2276$	MEP	**6.34E-03**	**9.05E-03**
	$\beta_2 = 2.1421$	Johnson	9.05E-00	1.38E-02
2	$\mu = 0.3936$	Pearson	**2.99E-02**	1.75E-02
	$\sigma = 0.0314$	Saddlepoint	–	–
	$\beta_1 = -0.6024$	MEP	4.52E-02	2.20E-02
	$\beta_2 = 3.0761$	Johnson	1.27E+01	**1.45E-02**

decomposed into a subsystem that represents the piston-ring/cyliner-liner subassembly of a single cylinder. The ring/liner subassembly simulation computes the power loss due to friction, oil consumption, blow-by, and liner wear rate. In the simulation, four random parameters were considered and listed in Table 10. MCS with 20,000 samples was employed to run the simulations at the current design. As shown in see Fig. 9b, the response PDF exhibits a bimodal behavior.

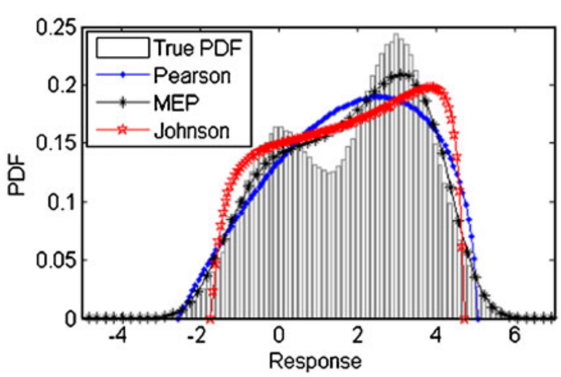
4.3.2 Results and remarks

The cross-entropy (ζ_1), U-pooling (ζ_2), and reliability accuracy (ζ_4) results are shown in Table 11 and Fig. 10, respectively. The MEP method shows the best accuracy for the 1st example. Although the MEP method cannot accurately reproduce the two modes of the PDF, it provides the best approximation among all the methods as shown in Fig. 9a. The Pearson and Johnson system produce relatively large cross-entropy (ζ_1) value because of incorrect representation of the PDF at both tail regions. The saddlepoint approximation fails to generate the PDF due to the numerical instability.

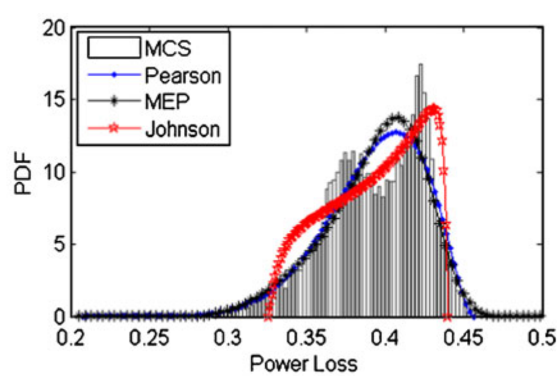
In the 2nd example, the Johnson system shows the best overall CDF quality according to the U-pooling (ζ_2) value. However, it has the largest cross-entropy (ζ_1) value due to inaccurate approximation of the PDF as shown in Fig. 9b. As a result, the Johnson system produces relatively large reliability errors in both tail regions as shown in Fig. 10b. Both the Pearson system and MEP method cannot reproduce the two-modes of the PDF as shown in Fig. 9b. The saddlepoint approximation fails to generate the PDF due to the numerical instability.

It is expected that more moments are required for the bimodal probability estimation. Theoretically, only the MEP method and saddlepoint approximation do not have any limitation on the number of statistical moments. Therefore, it may be feasible to use these two methods for the bimodal probability estimation by employing more statistical moments. Due to the numerical instability of the saddlepoint approximation for these two examples, only the MEP method was employed for this study.

It is clearly shown in Fig. 11 that the approximation accuracy is improved when more statistical moments are employed in the MEP method. Hence, the MEP method is able to better represent the bimodal PDF with more statistical moments. However, it would be a difficult task to obtain the high-order moments with sufficient accuracy in reality.



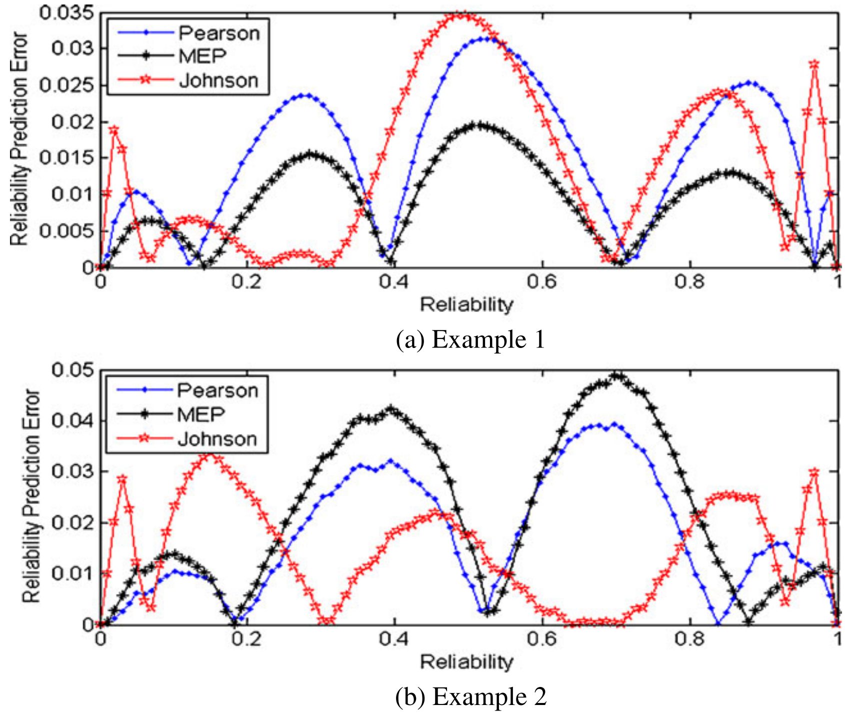
(a) PDF comparison of example 1



(b) PDF comparison of example 2

Fig. 9 Bimodal PDF comparison

Fig. 10 Reliability accuracy (ζ_4) of the bimodal PDF



This dilemma limits the feasibility of employing the probability estimation methods for reliability analysis when the system response has a multi-modal PDF.

4.4 Discussions

Figure 12 summarizes the probability estimation accuracy based on the U-pooling (see Fig. 12a) and cross entropy (see Fig. 12) under different settings of skewness and kurtosis considered in the case studies in Section 4.2. Each point represents the probability estimation method with the best accuracy under a specific setting of skewness and kurtosis. The limit curve separates the possible region (above the curve) in the Pearson and Johnson systems from the impossible region (below the curve).

Regarding the probability estimation accuracy based on the U-pooling (see Fig. 12a), three important remarks can be derived from the results. First of all, it is observed that the Johnson system presents the best accuracy in most case studies. Note that the U-pooling metric, as a measure of the area difference between the true and approximate CDFs, reflects the overall accuracy of a probability estimation method in the reliability estimation. The superb performance of the Johnson system in cases 1 and 2 can be attributed to the use of the logit-normal distribution.

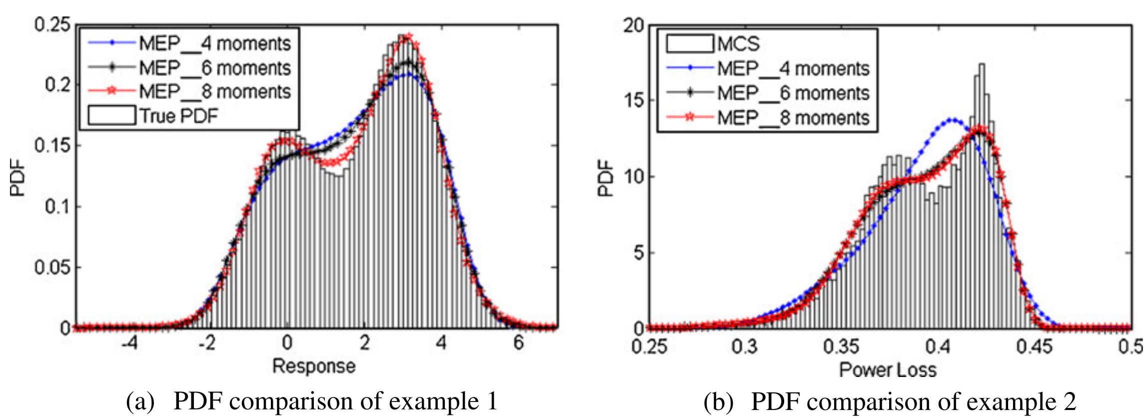


Fig. 11 Bimodal PDF comparison using the MEP method

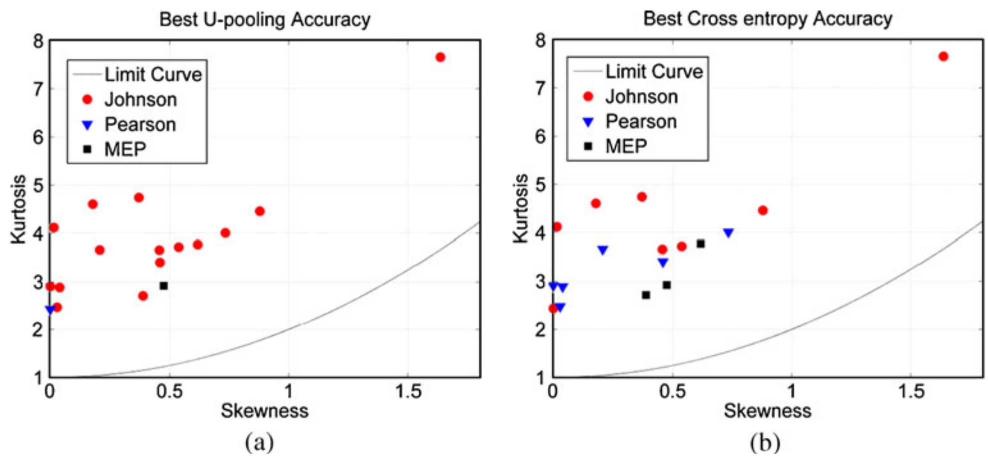


Fig. 12 The best accuracy for U-pooling and cross entropy

More specifically, although both the logit-normal distribution employed by the Johnson system and type I (beta) distribution by the Pearson system can reproduce a wide variety of distribution shapes with different parameter settings, the former was reported to be richer and possess higher flexibility than the later (Aitchison and Begg 1976). Thus, the Johnson system gives higher probability estimation accuracy than the Pearson system. Secondly, the Johnson and Pearson systems could yield relatively low accuracy for problems with high skewness (e.g., the 2nd example in cases 1 and 2) due to inaccurate bound definitions. In such cases, the MEP method delivers comparable or even better accuracy since it can capture the bound accurately. Thirdly, we observe that, for the 2nd example in cases 5 and 6 with low kurtosis, the Johnson and Pearson systems provide approximate response distributions that exhibit close agreement with each other (see Example 2 in Tables 6 and 7 and Fig. 7). This agreement becomes poorer as the kurtosis becomes higher, as can be seen from Examples 1 and 3 in Tables 6 and 7 and Fig. 7. This observation is consistent with that in the previous study (Johnson et al. 1994). We can make similar observation in cases 7 and 8 where both skewness and kurtosis are relatively high.

Regarding the probability estimation accuracy based on the cross entropy (see Fig. 12b), we observe that the probability estimation methods with the best U-pooling accuracy do not necessarily yield the best cross entropy accuracy. Note that the cross entropy measures the similarity between the true and approximate PDFs while the U-pooling measures the area difference between the true and approximate CDFs in the normalized space. We can distinguish these two comparison metrics using the results of the 2nd example in Section 4.3.2. As shown in Fig. 9b, the Johnson system gives an approximate PDF with the smallest similarity to the true one and thus has the largest cross entropy of 12.7 (see Table 11). However, the Johnson system surprisingly produces the best U-pooling accuracy. These results, though counterintuitive, suggest that the cross entropy is strongly associated with the accuracy in the reliability estimation over the tail region of a response PDF whereas the U-pooling, as the integration of the reliability prediction error over the entire reliability domain [0, 1], typically stands for that over the entire region.

5 Conclusion

In this work, a comparative study on four probability estimation methods was carried out to study their advantages and limitations for reliability analysis. The results from this comparative study give an insightful guidance for selecting the most appropriate probability estimation method for reliability analysis. Three general findings and limitations of this study are summarized as follows:

- Accuracy of the probability estimation: It is observed that none of the methods consistently shows the best accuracy. However, their relative merits are revealed through the extensive case studies. In most examples, the Johnson system shows better accuracy than the other methods in terms of the overall CDF quality. The Pearson system presents better accuracy for the high and low reliability levels than the Johnson system in case 1, case 2, case 5, and case 6. The MEP method is expected to produce better accuracy than the other methods for the high and low reliability levels when the skewness is relatively high in case 1 and case 2. The saddlepoint approximation shows better accuracy in a long tail region than the other methods in case 5 and case 6.
- Stability of the probability estimation: The stability is another important concern for reliability analysis. The stabilized Pearson system and MEP method are demonstrated to be very stable for the PDF construction. The

numerical instability issue of the saddlepoint approximation is observed when dealing with the negative value of $K''_Y(t_s)$. It is most likely to occur when the kurtosis value is less than 3 and the CGF is approximated from the first four statistical moments. Although the Johnson system presents high accuracy for the overall CDF approximation, the numerical instability is observed in searching for the optimal solution using the moment matching approach. Inappropriate initial values may lead to unconverged solutions for the four unknown parameters in the Johnson system.

- Capability of the bimodal probability estimation: All of the four probability estimation methods are capable of estimating the system reliability with high accuracy if the system response has a unimodal PDF. However, the accuracy will deteriorate significantly in cases when the system response has a bimodal or multi-mode PDF. Although it is possible to increase the accuracy using the MEP method by employing more statistical moments, it may not be a practically attractive approach due to the difficulty in estimating high-order statistical moments with sufficient accuracy.
- Limitations of this study: The comparative study focuses on a specific category of reliability analysis method, that is, the numerical integration-based method, where the numerical integration is first employed to compute statistical moments, and then the probability estimation method is used to approximate the probability density functions (PDFs), probabilities and, more specifically, reliabilities of system responses based on the computed moments. Two error sources exist in this category of method: (1) errors in estimating four statistical moments and (2) error in approximating the response PDF or CDF based on the four true moments. This paper only considers the 2nd error source by using accurate statistical moments from the MCS with a large number of random samples. Furthermore, with some special combinations of random variables, PDFs of system responses may exhibit rarely observed characteristics (e.g., more than two modes, extremely high skewness and kurtosis) that are beyond the case studies discussed in the paper. These situations are not considered in the paper.

Future works under consideration include: the employment of an ensemble of probability estimation methods for better reliability prediction, and possible ways to resolve the numerical instability in the saddlepoint approximation and Johnson system.

Acknowledgments Research is partially supported by the NSF GOALI 07294, the STAS contract (TCN-05122) sponsored by the U.S. Army TARDEC, by the New Faculty Development Program by Seoul National University, and by the SNU-IAMD.

Appendix

Table 12 Random input properties of three examples in case 1

Example	Variables	Dist. type	Para. #1	Para. #2
1	x_1	Normal	3	1
	x_2	Gamma	3	0.3
	x_3	Uniform	2	4
	x_4	Lognormal	0.3	0.1
	x_5	Uniform	0	3
2	P	Normal	6,070	200
	L	Normal	120	6
	a	Normal	65	5
	h	Normal	2.3	0.02
	w	Uniform	2.7415	3.1915
	t_1	Uniform	0.0976	0.1824
	t_2	Uniform	0.1976	0.3224
3	x_1	Normal	0.7	0.2
	x_2	Normal	0.3	0.1
	x_3	Uniform	0	1

Table 13 Random input properties of three examples in case 2

Example	Variables	Dist. type	Para. #1	Para. #2
1	x_1	Beta	2	4
	x_2	Beta	2	1
	x_3	Beta	3	4
	x_4	Uniform	−1	2
	x_5	Uniform	1	3
2	P	Uniform	5,470	6,670
	L	Uniform	102	138
	a	Uniform	50	80
	h	Uniform	2.24	2.36
	w	Uniform	2.7415	3.1915
	t_1	Uniform	0.0976	0.2224
	t_2	Uniform	0.1976	0.3224
3	x_1	Beta	3	30
	x_2	Beta	3	3
	x_3	Beta	3	1

Table 14 Random input properties of three examples in case 5

Example	Variables	Dist. type	Para. #1	Para. #2
1	x_1	Normal	0	1
	x_2	Beta	2	1
	x_3	Exponential	2	–
	x_4	Extreme	1	1
	x_5	Gamma	4	2
2	P	Normal	6,070	200
	L	Normal	120	6
	a	Normal	65	5
	h	Normal	2.3	0.02
	w	Normal	2.9665	0.075
	t_1	Normal	0.16	0.0208
	t_2	Normal	0.26	0.0208
3	x_1	Lognormal	0.01	0.6
	x_2	Weibull	1	1.5
	x_3	Normal	0	0.2

Table 16 Random input properties of three examples in case 7

Example	Variables	Dist. type	Para. #1	Para. #2
1	x_1	Normal	3	1
	x_2	Exponential	8	–
	x_3	Extreme value	6	2
	x_4	Lognormal	0.3	0.2
	x_5	Weibull	3	5
2	P	Normal	6,070	200
	L	Normal	120	6
	a	Normal	65	5
	h	Uniform	2.24	2.36
	w	Uniform	2.7415	3.1915
	t_1	Beta	16	100
	t_2	Beta	50	160
3	x_1	Normal	0.3	0.1
	x_2	Lognormal	0.01	0.3
	x_3	Beta	3	5

Table 15 Random input properties of three examples in case 6

Example	Variables	Dist. type	Para. #1	Para. #2
1	x_1	Beta	4	0.1
	x_2	Beta	4	0.5
	x_3	Beta	0.3	5
	x_4	Beta	0.5	8
	x_5	Uniform	0.2	0.3
2	P	Uniform	5,470	6,670
	L	Uniform	102	138
	a	Uniform	50	80
	h	Uniform	2.24	2.36
	w	Uniform	2.7415	3.1915
	t_1	Beta	16	100
	t_2	Beta	100	160
3	x_1	Beta	3	30
	x_2	Beta	3	50
	x_3	Beta	3	80

Table 17 Random input properties of three examples in case 8

Example	Variables	Dist. type	Para. #1	Para. #2
1	x_1	Beta	2	0.1
	x_2	Beta	2	0.1
	x_3	Beta	0.3	10
	x_4	Beta	0.3	10
	x_5	Uniform	0.5	0.6
2	P	Uniform	5,470	6,670
	L	Uniform	102	138
	a	Uniform	50	80
	h	Uniform	2.24	2.36
	w	Uniform	2.7415	3.1915
	t_1	Beta	16	100
	t_2	Beta	52	160
3	x_1	Beta	3	30
	x_2	Beta	3	30
	x_3	Beta	3	60

References

Aitchison J, Begg C (1976) Statistical diagnosis when cases are not classified with certainty. Biometrika 63(1):1–12

Chan KY, Kokkolara M, Papalambros PY, Skerlos SJ, Mourelatos Z (2004) Propagation of uncertainty in optimal design of multilevel systems: piston-ring/cyliner-liner case study. In: Proceedings of the 2004 SAE world congress, 8–11 March, Detroit, Michigan, Paper No 2004-01-1559

Chen H, Kamburowska G (2001) Fitting data to the Johnson system. J Stat Comput Simul 70(1):21–32

Daniels HE (1954) Saddlepoint approximations in statistics. Ann Math Stat 25(4):631–650

DeBrota DJ, Swain JJ, Roberts SD, Venkataraman S (1988) Input modeling with the Johnson system of distributions. In: Abrams MA, Haigh PL, Comfort JC (eds) Proceedings of the 1988 winter simulation conference. Institute of Electrical and Electronics Engineers, Piscataway, pp 165–179

Du X, Chen W (2004) Sequential optimization and reliability assessment method for efficient probabilistic design. ASME J Mech Des 126(2):225–233

Du X, Sudjianto A (2004) The first order saddlepoint approximation for reliability analysis. AIAA J 42(6):1199–1207

Farnum NR (1997) Using Johnson curves to describe nonnormal process data. Qual Eng 9(2):329–336

Ferson S, Oberkampf WL, Ginzburg L (2008) Model validation and predictive capability for the thermal challenge problem. Comput Methods Appl Mech Eng 197(29–32):2408–2430

Fu G, Moses F (1988) Importance sampling in structural system reliability. In: Proceedings of ASCE joint specialty conference on probabilistic methods. Blacksburg, VA, pp 340–343

Ghanem RG, Spanos PD (1991) Stochastic finite elements: a spectral approach. Springer, New York

Goutis C, Casella G (1999) Explaining the saddlepoint approximation. Am Stat 53(3):216–224

Hall AR (2005) Generalized methods of moments. Oxford University Press

Hasofer AM, Lind NC (1974) Exact and invariant second-moment code format. ASCE J Eng Mech 100(1):111–121

Huang B, Du X (2006) Uncertainty analysis by dimension reduction integration and saddlepoint approximations. J Mech Des 128:26–33

Huzurbazar S (1999) Practical saddlepoint approximations. Am Stat 53:225–232

Jaynes ET (1957) Information theory and statistical mechanics—I. Phys Rev 106(4):620–630

Johnson NL (1949) Systems of frequency curves generated by methods of translation. Biometrika 36:149–176

Johnson NL, Kotz S, Balakrishnan N (1994) Continuous univariate distributions. Wiley, New York

Kullback S, Leibler RA (1951) On information and sufficiency. Ann Math Stat 22(1):79–86

McDonald M, Mahadevan S (2008) Reliability-based optimization with discrete and continuous decision and random variables. ASME J Mech Des 130(6):061401

Mead LR, Papanicolaou N (1984) Maximum entropy in the problem of moments. J Math Phys 25:2404–2417

Pearson K (1901) Mathematical contributions to the theory of evolution, X: supplement to a memoir on skew variation. Philos Trans R Soc Lond, A Contain Pap Math Phys Character 197:443–459

Pearson K (1916) Mathematical contributions to the theory of evolution, XIX: second supplement to a memoir on skew variation. Philos Trans R Soc Lond A, Contain Pap Math Phys Character 216:429–457

Rabitz H, Alis OF (1999) General foundations of high-dimensional model representations. J Math Chem 25:197–233

Rabitz H, Alis OF, Shorter J, Shim K (1999) Efficient input–output model representations. Comput Phys Commun 117:11–20

Rahman S, Xu H (2004) A univariate dimension-reduction method for multi-dimensional integration in stochastic mechanics. Probab Eng Mech 19:393–408

Rubinstein RY (1981) Simulation and the Monte Carlo method. Wiley, New York

Smolyak SA (1963) Quadrature and interpolation formulas for tensor products of certain classes of functions. Sov Math Dokl 4:240–243

Soize C (2001) Maximum entropy approach for modeling random uncertainties in transient elastodynamics. J Acoust Soc Am 109(5):1979–1996

Straeten EVD, Beck C (2008) Superstatistical distributions from a maximum entropy principle. Phys Rev E 78(5):051101 (1–7)

Tvedt L (1984) Two second-order approximations to the failure probability. Section on structural reliability. A/S Vertas Research, Hovik

Xie K, Wells L, Camelio JA, Youn BD (2007) Variation propagation analysis on compliant assemblies considering contact interaction. ASME J Manuf Sci Eng 129(5):934–942

Xiong F, Greene MS, Chen W, Xiong Y, Yang X (2010) A new sparse grid based method for uncertainty propagation. Struct Multidisc Optim 41(3):335–349

Xiong Y, Chen W, Tsui K (2008) A better understanding of model updating strategies in validating engineering models. In: AIAA-2008–2155, 49th AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics, and materials conference, 7–10 April 2008, Schaumburg, IL

Xiu D, Karniadakis GE (2003) The Wiener–Askey polynomial chaos for stochastic differential equations. SIAM J Sci Comput 187:137–167

Xu H, Rahman S (2004) A generalized dimension-reduction method for multidimensional integration in stochastic mechanics. Int J Numer Methods Eng 61:1992–2019

Youn BD, Xi Z (2009) Reliability-based robust design optimization using the eigenvector dimension reduction (EDR) method. Struct Multidisc Optim 37(5):475–492

Youn BD, Choi KK, Yi K (2005) Performance moment integration (PMI) method for quality assessment in reliability-based robust design optimization. Mech Des Struct Mach 33:185–213

Youn BD, Xi Z, Wang P (2008) Eigenvector dimension-reduction (EDR) method for sensitivity-free probability analysis. Struct Multidisc Optim 37(1):13–28