

# Likelihood-based representation of epistemic uncertainty and its application in robustness-based design optimization

Kais Zaman<sup>1</sup> · Prithbey Raj Dey<sup>2</sup>

Received: 11 August 2015 / Revised: 15 March 2017 / Accepted: 18 March 2017 / Published online: 10 April 2017  
© Springer-Verlag Berlin Heidelberg 2017

**Abstract** In this paper, we propose a new likelihood-based methodology to represent epistemic uncertainty described by sparse point and/or interval data for input variables in uncertainty analysis and design optimization problems. A worst-case maximum likelihood-based approach is developed for the representation of epistemic uncertainty, which is able to estimate the distribution parameters of a random variable described by sparse point and/or interval data. This likelihood-based approach is general and is able to estimate the parameters of any known probability distributions. The likelihood-based representation of epistemic uncertainty is then used in the existing framework for robustness-based design optimization to achieve computational efficiency. The proposed uncertainty representation and design optimization methodologies are illustrated with two numerical examples including a mathematical problem and a real engineering problem.

**Keywords** Interval data · Maximum likelihood · Epistemic uncertainty · Robust design

## 1 Introduction

Robustness-based design optimization considers uncertainty in the objective function and constraints that arises from three types of sources: natural or physical variability (aleatory

uncertainty), data uncertainty (epistemic), and model uncertainty (epistemic). In recent years, many methods have been developed for robust design optimization. Some of these methods can only handle aleatory uncertainty (e.g., Parkinson et al. 1993; Du and Chen 2000; Doltsinis and Kang 2004; Huang and Du 2007), while others can handle both aleatory and epistemic uncertainty (e.g., Dai and Mourelatos 2003; Youn et al. 2007; Zaman et al. 2011a).

The efficiency of design optimization methods largely depends on how different sources of uncertainty are represented and included in the design optimization framework. An efficient robust design methodology should be able to treat all types of uncertainty in a unified manner, thus reducing the computational effort and simplifying the optimization problem. This paper focuses on the handling of sparse point and/or interval data in a manner that facilitates efficient algorithms for reliability analysis or design optimization. The developed uncertainty representation method is then used to propose a new and efficient approach for robustness-based design optimization under epistemic uncertainty.

Epistemic uncertainty regarding input or design variables can be viewed in two ways. It can be defined with reference to a stochastic quantity whose distribution type and/or distribution parameters are not precisely known (Baudrit and Dubois 2006), or with reference to a deterministic quantity whose value is not precisely known (Helton et al. 2004). This paper focuses on handling the first definition of epistemic uncertainty, i.e., epistemic uncertainty with reference to a stochastic quantity, whose distribution type and/or parameters are not precisely known due to sparse and/or imprecise data; thus the uncertainty in such quantities is a mixture of aleatory and epistemic uncertainty.

Most existing methods under epistemic uncertainty assume that only sparse point data or interval data are available for an input random variable. However, it is also possible that the

---

✉ Kais Zaman  
kaiszaman@yahoo.com

<sup>1</sup> Bangladesh University of Engineering and Technology,  
Dhaka, Bangladesh

<sup>2</sup> Dhaka University of Engineering and Technology,  
Gazipur, Bangladesh

information on an input random variable is available as a mixture of both sparse point data and interval data. In this paper, we propose an efficient likelihood-based approach for representation of epistemic uncertainty described by both sparse point and interval data.

Interval data are encountered frequently in practical engineering problems. Several such situations where interval data arise are discussed in Du et al. (2005), Ferson et al. (2007) and Zaman et al. (2011b), for example, a collection of expert opinions, which specify a range of possible values for a random variable. There exists an extensive volume of literature that presents efficient probabilistic as well as non-probabilistic methods to treat interval data in uncertainty analysis and design optimization problems.

Some studies within the context of probability theory have focused on representing interval uncertainty by a  $p$ -box (e.g., Ferson et al. 2007), which is the collection of all possible empirical distributions for the given set of intervals. Other research has focused on developing bounds on cumulative distribution functions (CDFs) (e.g., Hailperin 1986). Zaman et al. (2011b) proposed a probabilistic approach to represent interval data for input variables in reliability and uncertainty analysis problems, using flexible families of continuous Johnson distributions. Zaman et al. (2011c) developed both sampling and optimization-based approaches for uncertainty propagation in system analysis, when the information on the uncertain input variables and/or their distribution parameters may be available as either probability distributions or simply intervals (single or multiple). The Bayesian approach has also been used for reliability modeling with epistemic uncertainty (Zhang and Mahadevan 2000; Youn and Wang 2008). One non-probabilistic approach for representation of interval data is evidence theory (Shafer 1976). Evidence theory has been used with interval data for reliability-based design optimization (Mourelatos and Zhou 2006) and multidisciplinary systems design (Agarwal et al. 2004). Du (2008) proposed a unified uncertainty analysis method based on the first-order reliability method (FORM), where he modeled aleatory uncertainty using probability theory and epistemic uncertainty using evidence theory. Other approaches for epistemic uncertainty quantification based on evidence theory include Guo and Du (2007) and Guo and Du (2009). A discussion on different solution approaches (e.g., interval analysis, fuzzy analysis) to uncertainty quantification problems with imprecise data can be found in Zaman et al. (2011b, 2011c).

There is now an extensive volume of methods available for representation of epistemic uncertainty. However, several of these existing non-probabilistic methods, when used in the framework for uncertainty propagation and design optimization, can be computationally expensive. One reason is that for every combination of interval values, the probabilistic analysis for aleatory variables has to be repeated, which results in a computationally expensive *nested* analysis (Zaman et al.

2011c). Also, the current probabilistic methods that result in a family of distributions can be computationally expensive due to the use of a double-loop sampling strategy.

Sankararaman and Mahadevan (2011) developed a likelihood-based methodology for representation of epistemic uncertainty due to both sparse point data and interval data, where they used the full likelihood function to calculate the entire PDF of the distribution parameters, instead of maximizing the likelihood. The uncertainty in the distribution parameters is then integrated to calculate a single PDF for the random variable, which they referred to as the “averaged PDF” or the “weighted sum PDF”. In the context of uncertainty propagation, the single PDF approach is computationally efficient as it eliminates the need for a double-loop sampling strategy required by the family of distributions-based approach. However, the implementation of this single PDF approach is not straightforward as it involves Bayesian updating to account for uncertainty in distribution parameters and numerical integration to obtain unconditional PDF for the random variable of interest. Also, this single PDF approach cannot generate the same probability distribution that the random variable was previously assumed to follow; the resulting PDF is non-parametric. However, a parametric distribution is a convenient choice as it lends itself to easy transformation to a standard normal space, which then can be conveniently applied in well known reliability analysis and reliability-based design optimization methods. Therefore, methods to efficiently represent and propagate epistemic uncertainty (or a mixture of aleatory and epistemic uncertainty) in the context of design optimization and reliability analysis are yet to be developed.

This paper develops and illustrates a new approach for the representation of epistemic uncertainty available in the form of sparse point and/or multiple interval data based on the maximum likelihood principle. The proposed methodology solves a nested optimization formulation to find the worst-case maximum likelihood estimates of the distribution parameters of a random variable described by sparse point and/or multiple interval data. This worst-case maximum likelihood estimation (WMLE) approach is general and is able to estimate the parameters of any known probability distributions. The proposed likelihood-based representation of epistemic uncertainty is then used in the framework for robustness-based design optimization to achieve computational efficiency.

As mentioned earlier, there exist a few methods for robust design optimization that can handle both aleatory and epistemic uncertainty. Zaman et al. (2011a) proposed two formulations for robustness-based design optimization under both aleatory and epistemic uncertainty, namely nested and decoupled formulations. In this approach, the uncertainty analysis for the epistemic variables is carried out inside the design optimization framework, which makes these methods computationally expensive. In addition, this approach results in an overly conservative design. Lee and Park (2006)

developed a *nominal the best* type robust design optimization method using incomplete data (i.e., both fully and partially observed data) on a random variable to calculate the optimal operating conditions for the process based on a dual response approach. The efficiency of this surrogate based optimization method is highly dependent on the accuracy of surrogate modeling. In addition, this method has the following limitations. First, this method requires that the underlying random variable be normal. Second, this method is not applicable to all kinds of multiple interval data. In particular, this method is not suitable for the case, where there exists a common region of overlap among the intervals. Third, this approach underestimates input uncertainty.

Most existing methods of robust design are computationally expensive and result in an overly conservative or overly optimistic design. In this paper, a single-loop formulation is proposed to completely separate the epistemic analysis from the design optimization framework to achieve computational efficiency. In the context of interval uncertainty, the proposed likelihood-based single-loop approach generates realistic solutions as the resulting design is not too optimistic or too conservative. In order to investigate the efficiency of the proposed method, the robust optimization method (i.e., decoupled formulation) developed in Zaman et al. (2011a) is also studied in this paper. We have also studied and modified the robust optimization method developed in Lee and Park (2006) to include non-normal random variables. The proposed method is illustrated by using a mathematical example and an engineering example (the conceptual level design process of a two-stage-to-orbit (TSTO) vehicle).

The rest of the paper is organized as follows. Section 2 describes the proposed methodology for representation of epistemic uncertainty using maximum likelihood principle. Section 3 proposes a robustness-based design optimization framework that considers sparse point and/or interval data for the random variables. In Section 4, we illustrate the proposed methods for a mathematical example and an engineering example. Section 5 provides conclusions and suggestions for future work.

## 2 Likelihood-based approach to epistemic uncertainty representation

This section discusses the proposed likelihood-based approach that estimates the distribution parameters of random variables described by sparse point and/or multiple interval data. A brief background on maximum likelihood principle is provided first.

### 2.1 Maximum likelihood approach

Likelihood-based approach has been applied particularly to the estimation of the distribution parameters, when only point

data are available. Let  $f(x|p)$  be a conditional distribution (probability density or mass function) for the random variable  $X$  given the unknown parameters  $p$ . For the observed data,  $X=x$ , the function  $L(p) \propto f(x|p)$ , conditioned on the parameters  $p$ , is called the *likelihood function* (Edwards 1972; Pawitan 2001). For  $n$  independent and identically distributed observations of  $X$ , the likelihood function of the whole sample can be written as follows.

$$L(p) \propto \prod_{i=1}^n f(x_i|p) \quad (1)$$

The maximum likelihood estimate is then given by

$$\hat{p} = \arg \max_{p \in P} \prod_{i=1}^n f(x_i|p) \quad (2)$$

For computational convenience, instead of maximizing the likelihood function, we often maximize the log-likelihood function as given in Eq. (3).

$$\ln L(p) \propto \log \prod_{i=1}^n f(x_i|p) \quad (3)$$

Note that the likelihood function expressed in Eq. (1) assumes that point data are available for the random variable  $X$ . However, our focus in this paper is to estimate the unknown distribution parameters of a random variable  $X$ , which is described by sparse point and/or interval data. The definition of the likelihood function is not immediately obvious when the information on a random variable is available as intervals. There have been several attempts to extend likelihood beyond its usual use in inference with point data to inference with interval data, which include Gentleman and Geyer (1994), Meeker and Escobar (1995), and Sankararaman and Mahadevan (2011).

Gentleman and Geyer (Gentleman and Geyer 1994) constructed the likelihood function for interval censored data using the cumulative distribution function (CDF) of the random variable. A similar formulation is presented in Meeker and Escobar (1995) as given in Eq. (4).

$$L(p) \propto \prod_{i=1}^n \int_{a_i}^{b_i} f(x_i|p) = \prod_{i=1}^n [F(b_i|p) - F(a_i|p)] \quad (4)$$

where the random variable  $X$  is described by  $n$  intervals and  $a_i$  and  $b_i$  are the lower and upper bounds of the  $i$ th interval data, respectively.

Sankararaman and Mahadevan (2011) modified this formulation to include both point data and interval data in the likelihood function as follows.

$$L(p) \propto \left[ \prod_{i=1}^n f(x_i|p) \right] \left[ \prod_{i=1}^n [F(b_i|p) - F(a_i|p)] \right] \quad (5)$$

In Eq. (5), the random variable  $X$  is described by a combination of  $m$  point data and  $n$  intervals. Note that in Eq. (5), the

likelihood function is constructed as the joint probability density function (PDF) for sparse point data and the joint cumulative distribution function (CDF) for interval data. This multiplication may not be justified due to the fact that the PDF is a measure of relative probability, whereas the CDF is a measure of probability. Note that a similar approach can be found in Lee and Park (2006), where they used Eq. (4) to estimate the distribution parameters for normal random variables using partially observed observations. For both partially and fully observed observations, they used Eq. (5).

Sankararaman and Mahadevan (2011) suggests that the maximum likelihood estimates of the parameters  $\mathbf{p}$  can be obtained by maximizing the expression in Eq. (5) in the presence of both point data and interval data. However, instead of maximizing Eq. (5), they used a full likelihood estimate to construct the PDF of the distribution parameters  $\mathbf{p}$ . This likelihood-based approach has the following additional shortcomings:

First, their maximum likelihood-based approach is not applicable to all kinds of interval data. For a special case of overlapping interval data, where there is a region of overlap common to all intervals, the maximum likelihood estimate of the variance is zero (Sankararaman and Mahadevan 2011). Zaman et al. (2011b) proposed moment bounding algorithms to calculate the bounds on the first four moments for interval data and showed that the lower bound on the variance for overlapping interval data having a common region of overlap is zero. Therefore, it is evident that the maximum likelihood estimate of variance proposed by Sankararaman and Mahadevan (2011) is actually the lower bound variance proposed by Zaman et al. (2011b). This implies that this maximum likelihood-based approach underestimates uncertainty. Second, as mentioned earlier in Section 1, their approach cannot generate the same probability distribution that the random variable was previously assumed to follow; the resulting PDF is non-parametric.

Our focus in this paper is not to calculate the full likelihood estimate, however, to propose a new approach to estimate the distribution parameters in the presence of sparse point and/or interval data which, unlike existing approaches, does not underestimate or overestimate input uncertainty. In the following section, a maximum likelihood-based approach is proposed for variable  $X$  described by sparse point and/or multiple interval data.

## 2.2 Proposed worst-case maximum likelihood estimation (WMLE) approach

For a random variable  $X$  described by multiple interval data, the use of Eq. (2) to find the maximum likelihood estimates of the parameters,  $\mathbf{p}$  is not straightforward. Consider a set of intervals given as  $lb_i \leq x_i \leq ub_i$ ,  $i = \{1, \dots, n\}$  where  $n$  is

the number of intervals. Estimating the parameters  $\mathbf{p}$  involves identifying a configuration of scalar points ( $x_i$ ,  $i = \{1, \dots, n\}$ ), (where  $x_i$  indicates the true value of the observation within the interval) within the respective intervals and then solving Eq. (2) with this configuration of scalar points. Therefore, the presence of interval data requires that the maximum likelihood formulation in Eq. (2) be solved using every possible configuration of scalar points, with each configuration resulting in a different estimate for the same parameter. Theoretically, infinitely many different estimates of the parameters can be obtained from the given interval data.

In this paper, we solve this problem as a nested optimization problem with the objective of the outer optimization problem being maximization of the likelihood function (Eq. (1)) and the objective of the inner optimization problem being minimization of the likelihood with the data points constrained to fall within each of the respective intervals. The maximum likelihood estimation problem under interval uncertainty can now be formulated with the following generalized statement, where the objective is to maximize the worst case likelihood (i.e., lower bound of the likelihood, which is due to the epistemic uncertainty):

$$\max_{\mathbf{p}} \left( \min_x (f(\mathbf{x}|\mathbf{p}) = \log(L(\mathbf{p};\mathbf{x}))) \right) \quad (6)$$

$$s.t. lb_i \leq x_i \leq ub_i \quad \text{for } i = 1, 2, \dots, n$$

where the decision variables  $\mathbf{x}$  of the inner loop optimization problem are the configurations of multiple interval data ( $\mathbf{x} = [x_1 \ x_2 \ x_3 \ \dots \ x_n]$ ), which are constrained to fall within the respective intervals ( $[lb \ ub]$ ).

Note that in this formulation, the outer loop decision variables  $\mathbf{p}$  are the parameters of the distribution described by the PDF  $f(\mathbf{x}|\mathbf{p})$ . The outer loop optimization is a classical maximum likelihood estimation (MLE) problem, where an MLE is carried out for a fixed configuration of scalar points. The inner loop optimization is due to the presence of epistemic uncertainty, where the optimizer searches among the possible configurations of interval data to calculate the lower bound of the objective function value. The rationale behind this minimization problem is to obtain a conservative, i.e., worst-case estimate of parameters in the presence of epistemic uncertainty.

The proposed worst-case maximum likelihood approach presented in Eq. (6) can be extended to include both interval data and sparse point data on a random variable  $X$ . Given that a random variable  $X$  is described by  $m$  sparse point data and  $n$  intervals and that the data are obtained from independent sources, the likelihood function for this mixed data type can now be written as follows. For notational convenience, we order the observations so that the first  $m$  observations are the point data fixed at the respective point values ( $c$ ) and the next

$n$  observations are the interval data restricted to lie within the respective intervals ( $[lb\ ub]$ ).

$$L(\mathbf{p}) \propto \left[ \prod_{i=1}^m f(x_i = c_i | \mathbf{p}) \right] \left[ \prod_{i=m+1}^{m+n} f(x_i | \mathbf{p}) \right] \quad (7)$$

Therefore, the worst-case maximum likelihood formulation under this mixed data type can now be expressed as follows.

$$\begin{aligned} \max_{\mathbf{p}} \left( \min_x (f(\mathbf{x} | \mathbf{p}) = \log(L(\mathbf{p}; \mathbf{x}))) \right) \\ \text{s.t. } lb_i \leq x_i \leq ub_i \text{ for } i = m+1, m+2, \dots, m+n \end{aligned} \quad (8)$$

The optimization problem in Eq. (8) is solved with the objective of the outer optimization problem being maximization of the likelihood function in Eq. (7) and the objective of the inner optimization problem being minimization of the likelihood with the first  $m$  observations (sparse data points) being fixed at the respective point values ( $c$ ) and the next  $n$  observations (interval data) being constrained to fall within each of the respective intervals.

The proposed maximum likelihood approach is general and capable of handling different types of data (i.e., interval data or a mixture of sparse point and interval data) and any known probability distribution. However, in the following section, we introduce the Johnson family of distributions to fit interval data.

### 2.3 WMLE approach with Johnson family of distributions

The Johnson family is a generalized four-parameter family of distributions that can represent normal, lognormal, bounded, or unbounded distributions. Because of their flexibility, Johnson distributions can be used as a probabilistic representation of the interval data when the underlying probability distribution is not known. The Johnson family is a convenient choice among other four-parameter distributions, because it lends itself to easy transformation to a standard normal space, which can then be conveniently applied in well known reliability analyses and reliability-based design optimization methods.

Zaman et al. (2011c) argued that only bounded probability distributions are suitable to fit interval data. The use of a bounded distribution guarantees that the proposed approach does not estimate distribution parameters that yield realizations of random variables beyond the actual interval data set and thereby does not yield overly conservative design. In this paper, we have used the bounded Johnson distribution to fit interval data. For a complete description of interval uncertainty, it is required that an interval variable be described using higher order moments, in addition to the first two moments. Therefore, a four-parameter distribution (e.g., bounded Johnson) that uses four moments to estimate the distribution parameters is a rational choice, as it enables a rigorous, yet efficient implementation of epistemic uncertainty analysis.

If  $X$  follows bounded Johnson distribution and  $y = \left(\frac{X-\xi}{\lambda}\right)$ , then its PDF can be written as follows (Johnson 1949):

$$p(y) = \frac{\delta}{\sqrt{2\pi}} \times \frac{1}{y(1-y)} \times \exp\left\{-\frac{1}{2}\left[\gamma + \delta \log \frac{y}{1-y}\right]^2\right\}, \xi < X < +\xi + \lambda \quad (9)$$

In this paper, we estimate the parameters ( $\delta$ ,  $\gamma$ ,  $\xi$  and  $\lambda$ ) of the bounded Johnson distribution using the proposed WMLE approach as follows. Consider the likelihood function for  $n$  independent observations of the random variable  $X$  for bounded Johnson distribution,

$$\begin{aligned} L(\delta, \gamma, \xi, \lambda) = \prod_{i=1}^n \frac{\delta}{\sqrt{2\pi}} \times \frac{1}{\left(\frac{x_i-\xi}{\lambda}\right)\left(1-\frac{x_i-\xi}{\lambda}\right)} \\ \times \exp\left\{-\frac{1}{2}\left[\gamma + \delta \log \frac{x_i-\xi}{\lambda-x_i+\xi}\right]^2\right\} \end{aligned} \quad (10)$$

As discussed earlier, for computational convenience, instead of maximizing the likelihood function in Eq. (10), we often deal with the log-likelihood function as given in Eq. (11) below.

$$\begin{aligned} \log(L(\delta, \gamma, \xi, \lambda)) = n \log \delta - n \log \sqrt{2\pi} - \sum_{i=1}^n \log \left(\frac{x_i-\xi}{\lambda}\right) - \sum_{i=1}^n \log \left(\frac{\lambda-x_i+\xi}{\lambda}\right) \\ - \frac{1}{2} \sum_{i=1}^n \left[\gamma + \delta \log \left(\frac{x_i-\xi}{\lambda-x_i+\xi}\right)\right]^2 \end{aligned} \quad (11)$$

The log-likelihood function in Eq. (11) is used in the nested optimization formulations in Eqs. (6) and (8) to estimate the worst-case maximum likelihood estimates of the parameters ( $\delta$ ,  $\gamma$ ,  $\xi$  and  $\lambda$ ) of the bounded Johnson distribution for multiple interval data and mixed data, respectively.

Note that as opposed to commonly used family of distributions approach (Zaman et al. 2011b) to represent interval uncertainty, the proposed method generates a single PDF for a random variable in the presence of interval uncertainty, which then can be conveniently used in any existing algorithms for uncertainty propagation and design optimization.

In the following section, we propose a new methodology for robustness-based design optimization using the proposed likelihood-based representation of epistemic uncertainty.

## 3 Robustness-based design optimization under epistemic uncertainty

### 3.1 Existing methods

The robustness-based design optimization problem under aleatory uncertainty alone can be formulated as follows (Zaman et al. 2011a):

$$\begin{aligned}
& \min_{\mathbf{d}} \left( w \times \mu_f(d, \mu_z) + v \times \sigma_f(d, \sigma_x, \mu_z, \sigma_z) \right) \\
& \text{s.t. } LB + k\sigma_{g_i}(d, \sigma_x, \mu_z, \sigma_z) \leq \mu_{g_i}(d, x, z) \leq UB - k\sigma_{g_i} \\
& (d, \sigma_x, \mu_z, \sigma_z) \text{ for all } i \text{ } lb_i + k\sigma_{x_i} \leq d_i \leq ub_i - k\sigma_{x_i} \text{ for } i = 1, 2, \dots, \\
& \text{nr}dv \text{ } lb_i \leq d_i \leq ub_i \text{ for } i = 1, 2, \dots, \text{nd}dv
\end{aligned} \tag{12}$$

where  $\mu_f$  and  $\sigma_f$  are the mean value and standard deviation of the objective function, respectively;  $\mathbf{d}$  is the vector of deterministic design variables as well as the mean values of the uncertain design variables  $\mathbf{x}$ ;  $\text{nr}dv$  and  $\text{nd}dv$  are the numbers of the random design variables and deterministic design variables, respectively; and  $\mathbf{z}$  is the vector of non-design input random variables, whose values are kept fixed at their mean values  $\mu_z$  as a part of the design.  $w \geq 0$  and  $v \geq 0$  are the weighting coefficients that represent the relative importance of the objectives  $\mu_f$  and  $\sigma_f$  in Eq. (12);  $g_i(d, x, z)$  is the  $i$ th constraint;  $\mu_{g_i}(d, x, z)$  is the mean and  $\sigma_{g_i}(d, \sigma_x, \mu_z, \sigma_z)$  is the standard deviation of the  $i$ th constraint.  $LB$  and  $UB$  are the vectors of lower and upper bounds of constraints  $g_i$ 's;  $lb$  and  $ub$  are the vectors of lower and upper bounds of the design variables;  $\sigma_x$  is the vector of standard deviations of the random variables and  $k$  is some constant. The role of the constant  $k$  is to adjust the robustness of the method against the level of conservatism of the solution.

In the robustness-based design optimization formulation given in Eq. (12), the performance functions considered are in terms of the model outputs. The means and standard deviations of the objective and constraints are estimated by using a first-order Taylor series approximation (Haldar and Mahadevan 2000).

The implementation of Eq. (12) requires that variances of the random design variables  $\mathbf{X}$  and the means and variances of the random non-design variables  $\mathbf{Z}$  be precisely known, which is possible only when a large number of data points are available. However, in practice, only a small number of data points may be available for the non-design input variables  $\mathbf{Z}$ . In other cases, information about random input variables  $\mathbf{Z}$  may only be specified as intervals, as by expert opinion. Zaman et al. (2011a) proposed two formulations for robustness-based design optimization, namely nested and decoupled formulations, to take this data uncertainty into account. Their nested formulation does not often guarantee to converge and is computationally very expensive, because for every iteration of the epistemic analysis, the design optimization problem under aleatory uncertainty has to be repeated. Therefore, they proposed decoupled formulations that un-nest the design optimization problem from the epistemic analysis to achieve some computational efficiency. However, this is an iterative approach, where a design problem and an uncertainty analysis problem for epistemic variables are solved iteratively until

convergence. We can achieve further computational efficiency if the uncertainty analysis for the epistemic variable is carried out outside the design optimization framework. In the following section, we propose such an efficient single-loop approach for robustness-based design optimization under both aleatory and epistemic uncertainty.

### 3.2 Proposed likelihood-based robust design optimization

In the proposed robustness-based design optimization framework, the uncertainty analysis of the epistemic variables is done outside the design optimization framework using the proposed WMLE approach. The resulting single-loop formulation is equivalent to a design formulation under aleatory uncertainty alone, which completely eliminates the need for a nested analysis or an epistemic uncertainty analysis within the design optimization framework. Therefore, the proposed robustness-based design optimization methodology can solve the design problem with a marginally increased computational effort than a design formulation under aleatory uncertainty alone, where the increased computational cost is due to the worst-case maximum likelihood estimates of epistemic uncertainty.

The proposed single-loop formulation for the likelihood-based robust design optimization can be expressed as:

$$\begin{aligned}
& \mathbf{d}^* = \underset{\mathbf{d}}{\operatorname{argmin}} \left( w \times \mu_f(d, \mu_z^*) + v \times \sigma_f(d, \sigma_x, \mu_z^*, \sigma_z^*) \right) \\
& \text{s.t. } LB + k\sigma_{g_i}(d, \sigma_x, \mu_z^*, \sigma_z^*) \leq \mu_{g_i}(d, x, z) \leq UB - k\sigma_{g_i} \\
& (d, \sigma_x, \mu_z^*, \sigma_z^*) \text{ for all } i \text{ } lb + k\sigma_x \leq d \leq ub - k\sigma_x
\end{aligned} \tag{13}$$

In Eq. (13),  $\mu_z^*$  and  $\sigma_z^*$  are the worst-case maximum likelihood estimates of the mean values and standard deviations, respectively, of the non-design epistemic variables  $\mathbf{z}$  obtained through the proposed likelihood-based approach discussed in Section 2. Note that unlike the nested and decoupled formulations discussed in Section 3.1 where the mean values of the non-design epistemic variables are used as optimization decision variables, in the proposed robustness-based design optimization formulation in Eq. (13), the mean values of the epistemic variables are kept fixed at  $\mu_z^*$ .

Since the optimization formulation in Eq. (13) is solved with a fixed set of non-design epistemic variables, Eq. (13) is equivalent to a robust design formulation under aleatory uncertainty alone. Unlike the nested formulation, the proposed formulation does not suffer from any convergence issues. The proposed formulation also does not require any epistemic analysis within the design optimization framework.

Note that the decoupled approach was developed in Zaman et al. (2011a) to account for epistemic uncertainty in robustness-based design optimization. Since this method uses the upper bound variances of the epistemic variables and searches among the possible mean values of epistemic

variables to find optimal solution where the objective being maximization of the cost function, the resulting solution is overly conservative. This approach selects a distribution that has the largest variance from a family of distributions. However, it does not assert that any one CDF in the family is more or less likely to be the true CDF than the others. On the other hand, the robustness-based design optimization method proposed in this paper uses the worst-case maximum likelihood estimates of the mean and variance for the epistemic variable. Selecting the distribution parameters by minimizing the maximum likelihood ensures that the robust design problem is solved using the worst-case most probable estimate of the variance. However, unlike for the approach developed in Zaman et al. (2011a), the resulting design is not overly conservative as discussed later in Section 4.

As mentioned earlier, the mean values  $\mu_z^*$  and the standard deviations  $\sigma_z^*$  of the epistemic variables are obtained from the worst-case maximum likelihood estimates of the distribution parameters. For some known distributions (e.g., normal distribution), the moments and the distribution parameters are identical. Hence, the mean ( $\mu_z$ ) and standard deviation ( $\sigma_z$ ) of the non-design epistemic variables for the normal distribution can be directly calculated by the proposed maximum likelihood approach. For some other distributions (e.g., log-normal distribution), closed-form formulas are available to estimate moments from distribution parameters and vice-versa. However, in this paper, we use bounded Johnson distribution to model interval uncertainty for which no closed-form formulas for correspondence between moments and distribution parameters are available.

In the following discussion, we present a numerical method to calculate the first two moments (i.e., mean and variance) of bounded Johnson distribution from the parameters ( $\delta$ ,  $\gamma$ ,  $\xi$  and  $\lambda$ ) obtained through the proposed likelihood-based approach. The moments, thus calculated, are then used

in Eq. (13) to solve the robustness-based design optimization problem under epistemic uncertainty.

### 3.2.1 Estimation of moments from the bounded Johnson distribution parameters

The numerical integration-based procedure presented below is not a part of the likelihood-based parameter estimation methodology presented in Section 2. This procedure is only required for the design optimization. For uncertainty propagation or reliability analysis under epistemic uncertainty, it is not required to estimate moments from the distribution parameters.

Consider the general standard form of the bounded Johnson distribution (Johnson 1949),

$$Z = \gamma + \delta \log \left\{ \frac{y}{1-y} \right\} \quad (14)$$

where  $Z$  is a unit normal variable and  $y = \frac{x-\xi}{\lambda}$ ,  $\xi < x < \xi + \lambda$ .

The  $r$ th moment of the transformation variable  $y$  about zero can be expressed as the following generalized expression (Draper 1952):

$$\mu_r(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}z^2} \left( 1 + e^{-\frac{(z-\gamma)}{\delta}} \right)^{-r} dz \quad (15)$$

However, as discussed in Draper (1952), this integral is not easy to solve directly. In this paper, we have used numerical integration to evaluate Eq. (15). Once the first two central moments  $\mu_1$  and  $\mu_2$  are calculated using Eq. (15), the mean and the standard deviation of the input random variable  $X$  can be estimated as follows (Draper 1952):

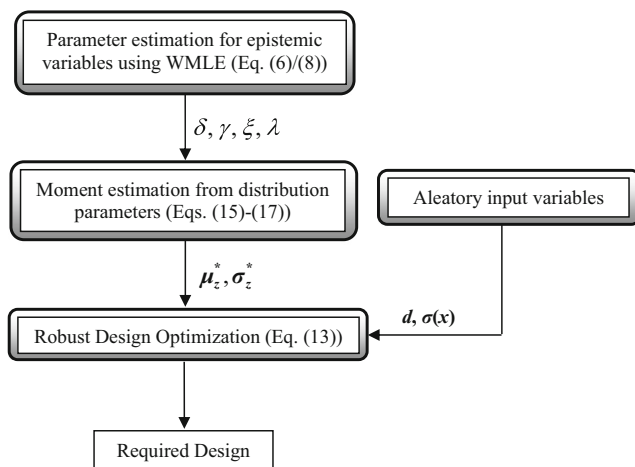
$$\mu(x) = \xi + \lambda \mu_1(y) \quad (16)$$

$$\sigma(x) = \lambda \sigma(y) \quad (17)$$

where  $\sigma(y) = \sqrt{\mu_2(y) - (\mu_1(y))^2}$ .

Once the parameters of the bounded Johnson distribution are estimated through the proposed WMLE approach, the mean ( $\mu_z$ ) and standard deviation ( $\sigma_z$ ) of the epistemic variables are calculated using Eqs. (15), (16) and (17), which are then used in the proposed robustness-based design optimization formulation given in Eq. (13). Figure 1 illustrates the proposed WMLE-based robustness-based design optimization under both aleatory and epistemic uncertainty.

Note that the robustness-based design optimization formulation developed in Zaman et al. (2011a) considers the worst case relatively to the robustness function (i.e., objective function of the robust design), whereas the proposed WMLE approach considers the worst case relatively to the likelihood function. Therefore, the worst-case likelihood function may not necessarily mean the worst-case responses or system



**Fig. 1** Likelihood-based approach for robustness-based design optimization

characteristics in the robust design. The relationship depends on the natures of the random variables. However, the advantage of the proposed method is to obtain a unique distribution for the random variables with interval data so that the double-loop procedure for robust design can be eliminated.

As mentioned earlier, the proposed maximum likelihood approach is capable of handling any known probability distribution. However, in this paper, we have used the bounded Johnson distribution to fit interval data for the sake of illustration only. Other bounded probability distributions can be easily substituted in the proposed methodology. For example, if we substituted Johnson distribution by a simpler probability distribution (e.g., beta distribution) whose moments are known analytically, we could easily avoid the calculations that were required to estimate the mean and variance from distribution parameters. However, since these calculations are not iterative and do not involve any performance function, i.e., a computational model (e.g., a finite element code), which is generally the more expensive analysis in a practical problem, the computational effort involved in this step is negligible.

In the proposed robustness-based design optimization framework, we have not used the complete probabilistic information of the epistemic variables in spite of their availability. However, it would be interesting to study how this information available as PDF could be utilized to solve robust design problem using more complex robust design optimization methods as discussed below.

First, in the proposed robustness-based design optimization framework, we have used variance as a measure of variation. However, there exists other probabilistic measures of robustness in the literature such as percentile difference, confidence interval, etc. that are able to provide more accurate evaluation of the variation of the performance function. Second, we have used the Taylor series expansion method to estimate the mean and variance of the performance function. Taylor series expansion is a simple approach. However, for a nonlinear performance function, if the variances of the random variables are large, this approximation may result in large errors (Du et al. 2004). Several other methods are available in the literature such as sampling-based methods, and point estimate methods. Sampling based methods, which require information on distributions of the random variables, can provide better precision at the expense of large computational effort. Point estimate methods are often a more practical alternative. However, they also require probabilistic information on the random variable. Chang et al. (1995) showed that in order to enhance the accuracy, additional statistical information in the form of skewness and kurtosis, other than the commonly used first two moments, should be used to estimate the mean and standard deviation of the performance function. Third, there are some methods available in the literature for robust design optimization (e.g., Congedo et al. 2014) that take into account also the

higher moments. The idea is obtaining an optimal solution that is not sensitive to the large variation in the skewness, which is achieved by adding another objective being minimization of the absolute value of the skewness. All these approaches can be easily incorporated in our proposed robustness-based design optimization framework as we have the complete PDF information on each epistemic variable.

In the following section, we illustrate our proposed methodologies for worst-case maximum likelihood estimation and robustness-based design optimization with sparse point and/or interval data.

### 4 Numerical examples

The proposed worst-case maximum likelihood approach and robustness-based design-optimization formulation are illustrated with two numerical examples: 1) a simple mathematical problem, and 2) an engineering problem (Two Stage To Orbit (TSTO) vehicle).

#### 4.1 Example 1: Mathematical example

Consider the following design problem:

$$\begin{aligned} \min_x f(x, z) &= x_1^2 - z_1 x_1 x_2 + x_1 x_2^2 + z_2 x_1 x_3 \\ \text{s.t. } g_1(x) &= -2x_1^2 - x_1 x_2 + x_3 + 5 \leq 0 \\ g_2(x) &= -x_1 x_2 + x_2^2 - x_1 x_3 + 5 \leq 0 \end{aligned} \tag{18}$$

In this example problem, the variables  $x_i$  (for  $i = 1, 2, 3$ ) are treated as random design variables and the variables  $z_i$  (for  $i = 1, 2$ ) are treated as non-design epistemic variables. Each random design variable has a lower bound of 1 and an upper bound of 10. It is assumed that each random design variable is normally distributed with a standard deviation of 0.5. The epistemic variable  $z_1$  is assumed to be described by 2 point estimates  $\{3.5, 4.5\}$  and 3 sets of intervals  $([2.8, 3.1], [4.7, 5.0], [5.2, 6.0])$ . The epistemic variable  $z_2$  is assumed to be described by a multiple interval data set  $([1.6, 2.2], [1.8, 2.3], [2.0, 2.5], [2.1, 2.6], [2.2, 2.7])$ , which has a common region of overlap among the intervals.

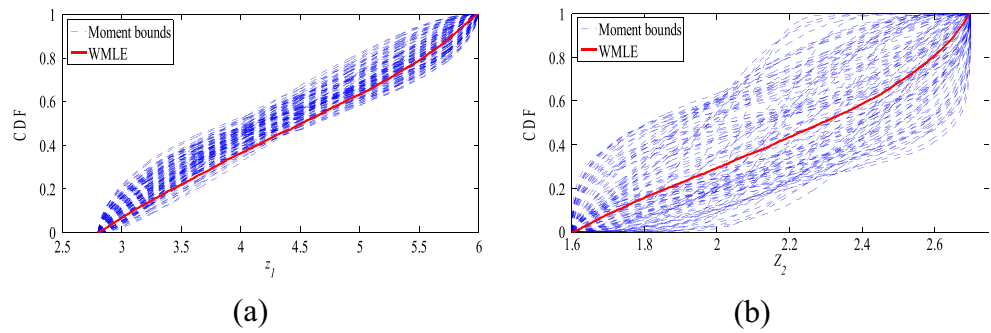
Note that since the probability distribution types for the random variables  $z_1$  and  $z_2$  are unknown; these variables are represented using the bounded Johnson distributions in the design optimization framework. The information on the

**Table 1** Distribution parameters for  $z_1$  and  $z_2$

	$\delta$	$\gamma$	$\xi$	$\lambda$
$z_1$	0.5392	-0.0182	2.8	3.2
$z_2$	0.4915	-0.2745	1.6	1.1



**Fig. 2** Comparison between WMLE-based CDF and family of CDFs for mathematical problem



epistemic variable  $z_1$  is available as both sparse point and interval data. Therefore, the distribution parameters ( $\delta, \gamma, \xi$  and  $\lambda$ ) of  $z_1$  are estimated by the proposed worst-case maximum likelihood method using the optimization formulation given in Eq. (8). Since the epistemic variable  $z_2$  is described by only multiple interval data, its distribution parameters are estimated using the optimization formulation given in Eq. (6). Both formulations are solved by the MATLAB solvers ‘fminunc’ and ‘fmincon’. The worst-case maximum likelihood estimates of the distribution parameters for the epistemic variables are given in Table 1.

We have generated a bounded Johnson CDF for each random epistemic variable using the distribution parameters estimated by the WMLE approach. The CDFs for the bounded Johnson distributions for input variables  $z_1$  and  $z_2$  are shown in Fig. 2. In order to verify the results of the proposed approach, we also use the moment bounding approach developed in Zaman et al. (2011b) to fit a family of bounded Johnson distributions. Several sample CDFs from the family of Johnson distributions for input variables  $z_1$  and  $z_2$  are also shown in Fig. 2.

Note that the proposed WMLE approach selects the worst-case distribution parameters from infinitely many possible maximum likelihood estimates, not from the family of distributions obtained using the moment bounding approach. However, we have compared the CDFs obtained by the WMLE approach with the family of CDFs obtained by the moment bounding approach, because the later is able to provide rigorous bounds on the distribution parameters. By rigorous, it is meant that the true interval of the possible values lies within the computed bounds. Therefore, a single CDF for the interval random variable  $X$  obtained by any probabilistic method must lie within the bounds of CDFs obtained using the moment bounding approach. It is seen in Fig. 2 that the single

CDF obtained using the proposed method is well within the bounds of CDFs obtained using the moment bounding approach for each epistemic variable.

Once the uncertainty in the epistemic variable is quantified as a single bounded Johnson distribution, the next step is to estimate the mean ( $\mu_z$ ) and standard deviation ( $\sigma_z$ ) of each epistemic variable from the bounded Johnson distribution parameters using Eqs. (15), (16) and (17), which are listed later in Table 2.

The likelihood-based single-loop robustness-based design formulation for the design problem given in Eq. (18) can now be written as:

$$d^* = \underset{d}{\operatorname{argmin}} \left( w \times \mu_f(d, \mu_z^*) + (1-w) \times \sigma_f(d, \sigma_x, \mu_z^*, \sigma_z^*) \right)$$

$$\text{s.t. } \mu_{g_1}(x) + k\sigma_{g_1}(d, \sigma_x) \leq 0$$

$$\mu_{g_2}(x) + k\sigma_{g_2}(d, \sigma_x) \leq 0$$

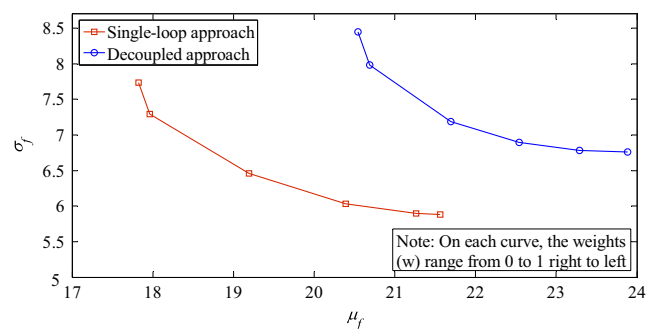
$$lb_i + k\sigma_{x_i} \leq d_i \leq ub_i - k\sigma_{x_i} \quad \text{for } i = 1, 2, 3 \quad (19)$$

The mean values and the standard deviations of the objective function and two functional constraints are obtained through the first-order Taylor series approximation assuming independence among the uncertain input variables.

In this paper,  $k$  is assumed to be unity. The weight parameter  $w$  is varied (from 0 to 1) and the optimization problem in Eq. (19) is solved by the Matlab solver ‘fmincon’. To investigate the efficiency of the proposed method, this example problem is also solved by the decoupled approach developed in Zaman et al. (2011a). The solutions from both the proposed and decoupled approaches are presented in Fig. 3. Note that

**Table 2** Mean values ( $\mu_z$ ) and standard deviations ( $\sigma_z$ ) of  $z_1$  and  $z_2$

Method	$w$	$z_1$		$z_2$	
		$\mu_z$	$\sigma_z$	$\mu_z$	$\sigma_z$
Single-loop	0–1	4.4171	0.9693	2.2414	0.3419
Decoupled	0–1	4.1400	1.1218	2.4600	0.4499



**Fig. 3** Robustness-based design optimization for example 1

this problem contains a multiple interval data set that has a common region of overlap among the intervals. Therefore, we have not solved this problem using the robust design optimization method developed in Lee and Park (2006).

In Table 2, we present the mean values and the standard deviations of the epistemic variables used in the single-loop formulation; for the sake of comparison the optimum values for the mean and the upper bound standard deviations of the epistemic variables used in the decoupled formulations are also included.

It is seen from Table 2 that the mean values and the standard deviations of the epistemic variables used in the single-loop approach differ significantly from the ones used in the decoupled approach. This is obvious due to the fact that their robustness is not computed from the same input distributions. However, the results are still comparable as their robustness is computed from the same multiple interval data.

Figure 3 shows the solutions of the robust design in the presence of epistemic uncertainty. It is seen in Fig. 3 that, as the weight ( $w$ ) increases, the mean of the performance function decreases, the standard deviation increases, and vice versa. This is a well known characteristic for any multi-objective optimization problem. A decrease in the standard deviation implies that some robustness is achieved in the design. Therefore, there is a tradeoff between the two objectives, minimizing the mean as well as the standard deviation of the performance function.

It is also seen from Fig. 3 that the WMLE-based single-loop formulation generates smaller values of mean ( $\mu_f$ ) and standard deviation ( $\sigma_f$ ) than the decoupled approach. This is

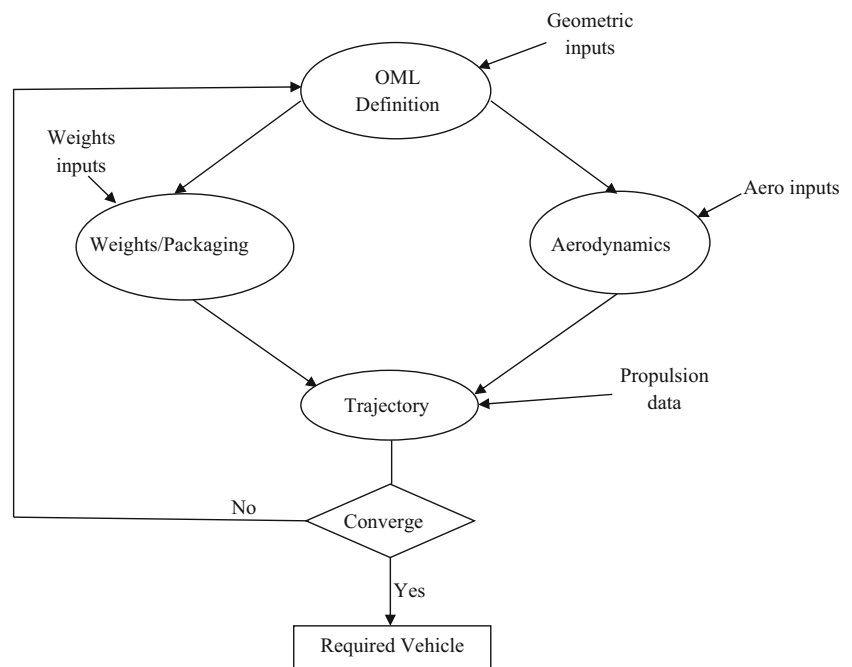
**Table 3** Design bounds and standard deviations of the design variables

Design variables	Lower bound (lb)	Upper bound (ub)	Standard deviation
ExpRatio	40	150	9.50
Payload	8000	40,000	2400
SepMach	7	12	0.95
SepQ	40	200	12

because the decoupled approach results in an overly conservative solution of robust design as it uses the upper bound variances of the epistemic variables and searches among the possible mean values of epistemic variables, which is a maximization problem, to find optimal solution as discussed in Section 3.

As mentioned earlier in Section 3, the decoupled approach developed in Zaman et al. (2011a) is an iterative approach, which required 2 iterations between the design problem and the uncertainty analysis for the non-design epistemic variables for convergence for all weights ( $w$ ) except for one which required 3 iterations. On the other hand, we just solved the WMLE-based formulation given in Eq. (19) once to obtain optimal solutions. It is seen that the proposed WMLE-based robust design methodology solved this design problem with only 185 function evaluations, whereas the decoupled approach required 853 function evaluations. Therefore, for this example problem, the proposed WMLE-based robustness-based design optimization is much more efficient than the decoupled formulation. However, the proposed robustness-based design optimization approach requires that the

**Fig. 4** TSTO vehicle concept (Zaman et al. 2011a)



uncertainty in epistemic input variables be quantified as a single PDF before the start of the design optimization algorithm. Therefore, the efficiency of the proposed design optimization approach partly depends on the efficiency of the proposed WMLE approach. However, since the proposed likelihood approach does not involve any performance function, the computational effort involved in the parameter estimation step is negligible.

Note that the proposed methodology generates optimal solutions that are not overly conservative and performs better than the existing method (e.g., decoupled approach) in terms of computational effort. Further, the proposed single-loop approach has some advantages over the existing method in terms of epistemic uncertainty modeling. First, unlike the decoupled approach where epistemic analysis is performed within the design optimization framework, the proposed single-loop approach completely separates the epistemic analysis from the design optimization framework and thereby achieves computational efficiency. Second, the proposed approach estimates the distribution parameters of the epistemic variables from the same configuration of multiple interval data as opposed to the decoupled approach where the mean and variance are estimated from different configurations of multiple interval data, which is not a feasible approach in the context of interval uncertainty.

## 4.2 Example 2: Engineering example

In this section, the proposed methods are illustrated for the conceptual level design process of a TSTO vehicle. This problem is adapted from Zaman et al. (2011a) and modified in this example to include both aleatory and epistemic uncertainty. In this example problem, the multidisciplinary system analysis consists of geometric modeling, aerodynamics, aerothermodynamics, engine performance analysis, trajectory analysis, mass property analysis and cost modeling (Stevenson et al. 2002). In this paper, a simplified version of the upper stage design process of a TSTO vehicle is used to illustrate the proposed methods. High fidelity codes of individual disciplinary analysis are replaced by inexpensive surrogate models. Figure 4 illustrates the analysis process of a TSTO vehicle.

The analysis outputs (performance functions) are Gross Weight (GW), Engine Weight (EW), Propellant Fraction Required (PFR), Vehicle Length (VL), Vehicle Volume (VV), and Body Wetted Area (BWA). Each of the analysis outputs is approximated by a second-order response surface

**Table 5** Distribution parameters for SepAngle and Fineness

	$\delta$	$\gamma$	$\xi$	$\lambda$
SepAngle	0.4902	-0.2411	8	1.5
Fineness	0.4923	-0.2008	4	2

and is a function of the random input variables Nozzle Expansion Ratio (ExpRatio), Payload Weight (Payload), Separation Mach (SepMach), Separation Dynamic Pressure (SepQ), Separation Flight Path Angle (SepAngle), and Body Fineness Ratio (Fineness).

In this paper, our objective is to optimize an individual analysis output (e.g., Gross Weight) while satisfying the constraints imposed by each of the design variables as well as the analysis outputs. It is assumed that four input variables (ExpRatio, Payload, SepMach, and SepQ) are random design variables and the remaining two variables (SepAngle, and Fineness) are the non-design epistemic variables described by multiple interval data. Each random design variable is assumed to be normally distributed. The numerical values of the design bounds and the standard deviations of the random design variables are given in Table 3. The multiple interval data for the epistemic variables are given in Table 4.

This example problem contains two epistemic variables (SepAngle and Fineness), for which the probability distribution types are unknown. In this example, it is assumed that both SepAngle and Fineness are characterized by bounded Johnson distributions. We follow the procedure described in Section 2 to obtain the distribution parameters of SepAngle and Fineness as given in Table 5.

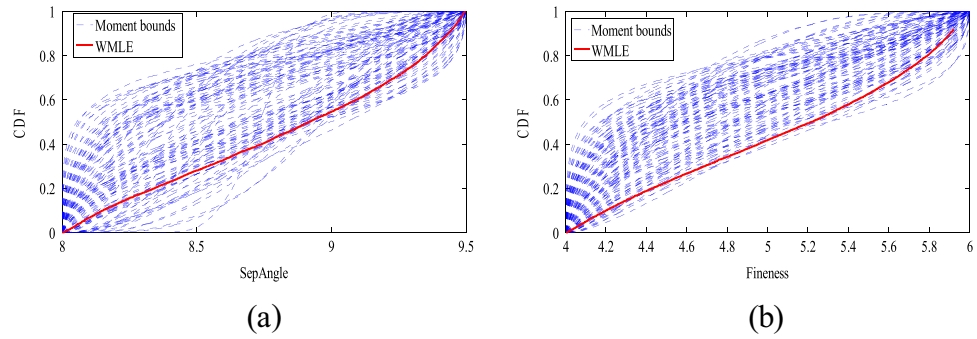
We use the distribution parameters estimated by the WMLE approach to fit a bounded Johnson distribution to each multiple interval data set. For the sake of comparison, we have also generated a family of bounded Johnson distributions for each epistemic variable using the moment bounding approach outlined in Zaman et al. (2011b). The results from both the approaches are shown in Fig. 5. It is seen in Fig. 5 that the single CDF obtained using the proposed method is well within the bounds of CDFs obtained using the moment bounding approach for each epistemic variable.

Once the uncertainty in the epistemic variable is quantified as a single bounded Johnson distribution, we then estimate the mean ( $\mu_z$ ) and standard deviation ( $\sigma_z$ ) of each epistemic variable from the bounded Johnson distribution parameters using Eqs. (15), (16) and (17), which are listed later in Table 6.

**Table 4** Multiple interval data for the epistemic input variables

SepAngle	[8.00, 8.50], [8.20, 8.60], [8.50, 9.00], [8.20, 9.20], [8.80, 9.50]
Fineness	[4.00, 4.50], [4.25, 4.75], [4.50, 5.25], [4.50, 5.50], [5.00, 6.00]

**Fig. 5** Comparison between WMLE-based CDF and family of CDFs for TSTO problem



The design problem can now be formulated as follows:

$$\begin{aligned}
 d^* &= \underset{d}{\operatorname{argmin}} (w \times \mu_{GW}(d, \mu_z^*) + (1-w) \times \sigma_{GW}(d, \sigma_x, \mu_z^*, \sigma_z^*)) & (20) \\
 \text{s.t. } & LB_1 + k\sigma_{GW}(d, \sigma_x, \mu_z^*, \sigma_z^*) \leq \mu_{GW}(x, z) \leq UB_1 - k\sigma_{GW}(d, \sigma_x, \mu_z^*, \sigma_z^*) \\
 & LB_2 + k\sigma_{EW}(d, \sigma_x, \mu_z^*, \sigma_z^*) \leq \mu_{EW}(x, z) \leq UB_2 - k\sigma_{EW}(d, \sigma_x, \mu_z^*, \sigma_z^*) \\
 & LB_3 + k\sigma_{PFR}(d, \sigma_x, \mu_z^*, \sigma_z^*) \leq \mu_{PFR}(x, z) \leq UB_3 - k\sigma_{PFR}(d, \sigma_x, \mu_z^*, \sigma_z^*) \\
 & LB_4 + k\sigma_{VL}(d, \sigma_x, \mu_z^*, \sigma_z^*) \leq \mu_{VL}(x, z) \leq UB_4 - k\sigma_{VL}(d, \sigma_x, \mu_z^*, \sigma_z^*) \\
 & LB_5 + k\sigma_{VV}(d, \sigma_x, \mu_z^*, \sigma_z^*) \leq \mu_{VV}(x, z) \leq UB_5 - k\sigma_{VV}(d, \sigma_x, \mu_z^*, \sigma_z^*) \\
 & LB_6 + k\sigma_{BWA}(d, \sigma_x, \mu_z^*, \sigma_z^*) \leq \mu_{BWA}(x, z) \leq UB_6 - k\sigma_{BWA}(d, \sigma_x, \mu_z^*, \sigma_z^*) \\
 & lb_i + k\sigma_{x_i} \leq d_i \leq ub_i - k\sigma_{x_i} \quad \text{for } i = 1, 2, 3, 4
 \end{aligned}$$

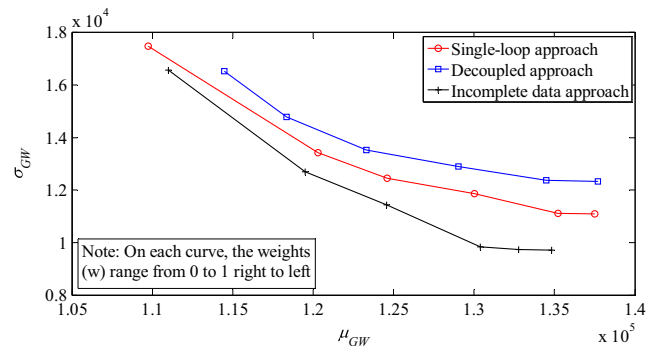
The mean values and the standard deviations of the performance functions are estimated by the first-order Taylor series approximation. The weight parameter  $w$  is varied (from 0 to 1), and the optimization formulation in Eq. (20) is solved by the Matlab solver “fmincon”. We also solve this problem using the decoupled approach developed in Zaman et al. (2011a). Since this problem does not contain a multiple interval data set that has a common region of overlap among the intervals, this problem is also solved by the robust design optimization method with incomplete data developed in Lee and Park (2006). Note that Lee and Park (2006) used a *nominal the best* type formulation, where they considered robustness in the objective function only; feasibility robustness, i.e., robustness in the constraint

functions was not considered. In this paper, we have used a similar formulation as in Eq. (20) for this incomplete data approach; however, the mean values and the standard deviations of the epistemic variables have been estimated using Eq. (4) by assuming that the underlying distribution is bounded Johnson. The solutions from all the approaches are presented in Fig. 6.

In Table 6, we present the mean values and the standard deviations of the epistemic variables used in the single-loop formulation; for the sake of comparison the mean values and the standard deviations of the epistemic variables used in the decoupled and incomplete data approaches are also included.

**Table 6** Mean values ( $\mu_z$ ) and standard deviations ( $\sigma_z$ ) of SepAngle and Fineness

Method	Weights ( $w$ )	SepAngle		Fineness	
		$\mu_z$	$\sigma_z$	$\mu_z$	$\sigma_z$
Single-loop	0–1	8.8597	0.4687	5.1219	0.6264
Decoupled	0–0.8	8.3400	0.5810	5.2000	0.7681
	1	8.9600	0.5810	4.4500	0.7681
Incomplete data	0–1	8.6166	0.2623	4.7640	0.3566



**Fig. 6** Robustness-based design optimization for Example 2

It is seen from Table 6 that the incomplete data approach results in overly optimistic estimates for the standard deviation of the epistemic variables (i.e., the standard deviations assume lower values). Therefore, this approach results in an overly optimistic design. The decoupled approach selects a distribution that has the largest variance and results in an overly conservative design. On the contrary, the single-loop approach proposed in this paper results in a design that is not overly conservative or overly optimistic as the estimated variance is less than the upper bound variance used in the decoupled approach and much larger than the one obtained by the incomplete data approach.

Figure 6 shows the solutions of the robust design for the TSTO problem in the presence of epistemic uncertainty. It is seen in Fig. 6 that for the same value of the mean ( $\mu_{GW}$ ) of the objective function, the likelihood-based single-loop approach generates smaller values of standard deviation ( $\sigma_{GW}$ ) than the decoupled approach. Similarly, for the same value of the standard deviation ( $\sigma_{GW}$ ), the optimal solutions obtained by the single-loop approach have smaller values of the mean ( $\mu_{GW}$ ) than the decoupled approach. However, the incomplete data approach developed in Lee and Park (2006) results in the smallest values for both the mean ( $\mu_{GW}$ ) and standard deviation ( $\sigma_{GW}$ ) of the objective function. This behavior is intuitive given the fact that the decoupled approach results in an overly conservative design, whereas the incomplete data approach underestimates input uncertainty and thereby results in an overly optimistic design.

From engineering and economic perspectives, a design must be realistic, not too optimistic or too conservative. An overly optimistic design is disappointing as it may cause a system to be designed for superoptimal performance, which may cause faulty operation, whereas an overly conservative design is likely to cause a system to be designed for suboptimal performance, which may leave the competition with a better design. Therefore, in the context of interval uncertainty, the proposed likelihood-based single-loop approach generates realistic solutions as the resulting design is not too optimistic or too conservative.

As mentioned earlier, the decoupled approach is an iterative approach, which required 2–3 iterations between the design problem and the uncertainty analysis for the non-design epistemic variables for convergence, depending on the weight parameter  $w$ . The proposed WMLE-based robust design methodology solved this design problem with only 300 function evaluations, whereas the decoupled approach required 832 function evaluations. Therefore, for this example problem, the proposed WMLE-based single-loop robustness-based design optimization is much more efficient than the decoupled formulations. However, the incomplete data approach resulted in a computational effort (350 function evaluations) comparable to the single-loop approach.

## 5 Conclusions

This paper proposes a worst-case maximum likelihood estimation (WMLE) methodology to estimate the distribution parameters of random variable described by sparse point and/or interval data. The distribution parameters of the epistemic variables thus estimated are then used to develop an efficient methodology for robustness-based design optimization under both aleatory and epistemic uncertainty. The proposed methodologies are illustrated for two numerical example problems – a general mathematical problem and the upper level conceptual design of a TSTO vehicle.

In this paper, we present a new methodology to convert sparse point and/or interval data to a probabilistic format. Unlike the existing family of distributions approach, the proposed approach results in a single PDF for a random variable described by sparse point and/or interval data. This single PDF uncertainty representation approach enables an efficient combined treatment of aleatory and epistemic input uncertainty from the perspective of uncertainty propagation and design optimization. Existing methods typically produce the maximum and minimum output quantities of interest and are computationally expensive because they often require a double loop approach – an uncertainty analysis/design optimization loop with respect to random variables and an epistemic analysis loop for extreme responses with respect to epistemic variables. However, in the context of uncertainty propagation, the proposed single PDF approach can achieve computational efficiency by eliminating the need for double loop analysis through the estimation of a single PDF for each epistemic variable and thereby treating the uncertainty propagation problem as a single-loop problem.

This single PDF approach also facilitates the implementation of design optimization under both aleatory and epistemic uncertainty by completely separating the epistemic analysis from the design optimization framework. In this paper, we propose a single-loop formulation for robustness-based design optimization based on the proposed worst-case maximum likelihood uncertainty representation method. Unlike the existing methods that either use a nested optimization formulation or a decoupled approach of optimization, the proposed single-loop formulation completely eliminates the epistemic analysis from the design optimization framework to achieve computational efficiency. The uncertainty analysis for the epistemic variables is performed outside the design optimization framework. In the context of interval uncertainty, the proposed likelihood-based single-loop approach generates realistic solutions as the resulting design is not too optimistic or too conservative.

The proposed likelihood-based approach is general and is able to estimate the parameters of any known probability distributions. However, in this paper, we have used bounded Johnson distribution to illustrate the proposed method.

The proposed worst-case maximum likelihood approach has the following advantages. First, the proposed methodology has the ability to convert epistemic uncertainty to probabilistic format as a single PDF, which eliminates the need for any nested or double loop analysis for both uncertainty propagation and design optimization problems. Second, the proposed approach can deal with mixed data, i.e., both sparse point and interval data on a random variable. Third, unlike existing approaches, the proposed methodology does not underestimate or overestimate input uncertainty and is able to retain the parametric form of the distribution. Fourth, the proposed approach is also valid for any type of multiple interval data, i.e., non-overlapping, overlapping, or mixed intervals (a combination of the former two). Since both types of uncertainty are treated in a unified manner using single PDF, the proposed methodologies can facilitate the efficient implementation of multidisciplinary uncertainty propagation and design optimization, which are computationally demanding problems in the presence of both aleatory and epistemic uncertainty.

**Acknowledgements** This study was supported by funds from Committee for Advanced Studies and Research (CASR), Bangladesh University of Engineering and Technology (BUET). The support is gratefully acknowledged.

## References

- Agarwal H, Renaud JE, Preston EL, Padmanabhan D (2004) Uncertainty quantification using evidence theory in multidisciplinary design optimization. *Reliab Eng Syst Saf* 85(1):281–294
- Baudrit C, Dubois D (2006) Practical representations of incomplete probabilistic knowledge. *Comput Stat Data Anal* 51:86–108
- Chang CH, Tung YK, Yang JC (1995) Evaluation of probability point estimate methods. *Appl Math Model* 19(2):95–105
- Congedo PM, Geraci G & Iaccarino G (2014) On the use of high-order statistics in robust design optimization. In 6th european conference on computational fluid dynamics (ECFD VI), July 20–25, Barcelona, Spain
- Dai Z and Mourelatos ZP (2003) Incorporating epistemic uncertainty in robust design. Proceedings of DETC, 2003 ASME Design Engineering Technical Conferences, Chicago
- Doltsinis I, Kang Z (2004) Robust design of structures using optimization methods. *Comput Methods Appl Mech Eng* 193(23):2221–2237
- Draper J (1952) Properties of distributions resulting from certain simple transformations of the normal distribution. *Biometrika* 39(3–4): 290–230
- Du X (2008) Unified uncertainty analysis by the first order reliability method. *J Mech Des* 130(9):091401(10)
- Du X, Chen W (2000) Towards a better understanding of modeling feasibility robustness in engineering. *ASME J Mech Des* 122(4):385–394
- Du X, Sudjianto A, Chen W (2004) An integrated framework for optimization under uncertainty using inverse reliability strategy. *J Mech Des* 126(4):562–570
- Du X, Sudjianto A, Huang B (2005) Reliability based design with mixture of random and interval variables. *ASME J Mech Des* 127(6): 1068–1076
- Edwards AWF (1972) *Likelihood*. Cambridge University Press, Cambridge First edition
- Ferson S, Kreinovich V, Hajagos J, Oberkampf W, Ginzburg L (2007) Experimental uncertainty estimation and statistics for data having interval uncertainty. Sandia National Laboratories Technical report SAND2007–0939, Albuquerque
- Gentleman R, Geyer CJ (1994) Maximum likelihood for interval censored data: consistency and computation. *Biometrika* 81(3):618–623
- Guo J, Du X (2007) Sensitivity analysis with mixture of epistemic and aleatory variables. *AIAA J* 45(9):2337–2349
- Guo J, Du X (2009) Reliability sensitivity analysis with random and interval variables. *Int J Numer Methods Eng* 78(13):1585–1617
- Hailperin T (1986) *Boole's logic and probability*. Elsevier, North-Holland
- Haldar A and Mahadevan S (2000) *Probability, reliability and statistical methods in engineering design*, John Wiley & Sons, Inc.
- Helton JC, Johnson JD, Oberkampf WL (2004) An exploration of alternative approaches to the representation of uncertainty in model predictions. *Reliab Eng Syst Saf* 85(1):39–71
- Huang B, Du X (2007) Analytical robustness assessment for robust design. *Struct Multidiscip Optim* 34(2):123–137
- Johnson NL (1949) Systems of frequency curves generated by methods of translation. *Biometrika* 36(1/2):149–176
- Lee SB, Park C (2006) Development of robust design optimization using incomplete data. *Comput Ind Eng* 50(3):345–356
- Meeker WQ, Escobar LA (1995) *Statistical methods for reliability data*. John Wiley and Sons, New York
- Mourelatos ZP, Zhou J (2006) A design optimization method using evidence theory. *J Mech Des* 128(4):901–908
- Parkinson A, Sorensen C, Pourhassan N (1993) A general approach for robust optimal design. *J Mech Des* 115(1):74–80
- Pawitan Y (2001) *In all likelihood: statistical modeling and inference using likelihood*. Oxford Science Publications, New York
- Rao SS, Annamdas KK (2009) An evidence-based fuzzy approach for the safety analysis of uncertain systems. 50th AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics, and materials conference, Palm Springs Paper number AIAA-2009-2263
- Sankararaman S, Mahadevan S (2011) Likelihood-based representation of epistemic uncertainty due to sparse point data and/or interval data. *Reliab Eng Syst Saf* 96(7):814–824
- Shafer G (1976) *A mathematical theory of evidence*. Press, Princeton University. isbn:0-608-02508-9
- Stevenson MD, Hartong AR, Zweber JV, Bhungalia AA and Grandhi RV (2002) Collaborative design environment for space launch vehicle design and optimization, RTO AVT symposium on reduction of military vehicle acquisition time and cost through advanced modeling and virtual simulation, Paris, France
- Youn BD, Wang P (2008) Bayesian reliability-based design optimization using eigenvector dimension reduction (EDR) method. *Struct Multidiscip Optim* 36(2):107–123
- Youn BD, Choi KK, Du L (2007) Integration of possibility-based optimization and robust design for epistemic uncertainty. *J Mech Des* 129(8):876–882
- Zaman K, McDonald M, Mahadevan S, Green L (2011a) Robustness-based design optimization under data uncertainty. *Struct Multidiscip Optim* 44(2):183–197
- Zaman K, Rangavajhala S, McDonald PM, Mahadevan S (2011b) A probabilistic approach for representation of interval uncertainty. *Reliab Eng Syst Saf* 96(1):117–130
- Zaman K, McDonald M, Mahadevan S (2011c) Probabilistic framework for uncertainty propagation with both probabilistic and interval variables. *J Mech Des* 133(2):0210101–02101014
- Zhang R, Mahadevan S (2000) Model uncertainty and Bayesian updating in reliability-based inspection. *Struct Saf* 22(2):145–160