

Design Optimization with Imprecise Random Variables

Jeffrey W. Herrmann
University of Maryland

Copyright © 2009 SAE International

ABSTRACT

Design optimization is an important engineering design activity. Performing design optimization in the presence of uncertainty has been an active area of research. The approaches used require modeling the random variables using precise probability distributions or representing uncertain quantities as fuzzy sets. This work, however, considers problems in which the random variables are described with imprecise probability distributions, which are highly relevant when there is limited information about the distribution of a random variable. In particular, this paper formulates the imprecise probability design optimization problem and presents an approach for solving it. We present examples for illustrating the approach.

INTRODUCTION

Design optimization is an important engineering design activity in automotive, aerospace, and other development processes. In general, design optimization determines values for design variables such that an objective function is optimized while performance and other constraints are satisfied [1, 2, 3]. The use of design optimization in engineering design continues to increase, driven by more powerful software packages and the formulation of new design optimization problems motivated by the decision-based design (DBD) framework [4, 5] and the corresponding idea of design for market systems [6].

Because many engineering problems must be solved in the presence of uncertainty, developing approaches for solving design optimization problems that have uncertain variables has been an active area of research. The approaches used require modeling the random variables

using precise probability distributions or representing uncertain quantities as fuzzy sets. Haldar and Mahadevan [7] give a general introduction to reliability-based design optimization, and many different solution techniques have been developed [8, 9, 10]. Other approaches include evidence-based design optimization [11], possibility-based design optimization [12], and approaches that combine possibilities and probabilities [13]. Zhou and Mourelatos [11] discussed an evidence theory-based design optimization (EBDO) problem. They used a hybrid approach that first solves a RBDO to get close to the optimal solution and then generates response surfaces for the active constraints and uses a derivative-free optimizer to find a solution.

The amount of information available and the outlook of the decision-maker (design engineer) determines the appropriateness of different models of uncertainty. No single model should be considered universally valid. In this paper, we consider situations in which there is insufficient information about the random variables to model them with precise probability distributions. Instead, imprecise probability distributions (described in more detail below) are used to capture the limited information or knowledge. In the extreme case, the imprecise probability distribution may be a simple interval. This paper presents an approach for solving design optimization problems in which the random variables are described with imprecise probability distributions because there exists limited information about the uncertainties.

IMPRECISE PROBABILITIES

In traditional probability theory, the probability of an event is defined by a single number between in the range [0, 1]. However, because this may be inappropriate in cases of incomplete or conflicting

Information, researchers have proposed theories of imprecise probabilities. For these situations, probabilities can be intervals or sets, rather than precise numbers [14, 15, 16]. The theory of imprecise probabilities, formalized by Walley [15], uses the same fundamental notion of rationality as the work of de Finetti [17, 18]. However, the theory allows a range of indeterminacy—prices at which a decision-maker will not enter a gamble as either a buyer or a seller. These in turn correspond to ranges of probabilities.

Imprecise probabilities have previously been considered in reliability analysis [19, 20, 21] and engineering design [22, 23, 24]. Aughenbaugh and Herrmann [25, 26, 27] have compared techniques using imprecise probabilities to other statistical approaches for making reliability-based design decisions. The work described in this paper builds upon these previous results.

In many engineering applications, the relevant random variables (e.g., parameters or manufacturing errors) are continuous variables. One common way to represent the imprecision in the probability distribution of such a random variable is a probability box (“p-box”) that is a set of cumulative probability distributions bounded by an upper distribution $\bar{F}$ and a lower distribution $\underline{F}$. These bounds model the epistemic uncertainty about the probability distribution for the random variable. Of course, a traditional precise probability distribution is a special case of a p-box, in which the upper and lower bounds are equal.

There are multiple ways to construct a p-box for a random variable [28]. In some cases, the type of distribution is known (or assumed) but its parameters are imprecise (such as an interval for a mean). In other cases, the distribution is constructed from sample data. Additionally, one can create a p-box from a Dempster-Shafer structure, in which intervals (not points) within the range of the random variable are assigned probabilities. For more about p-boxes and the link between p-box representation and Dempster-Shafer structures, see Ferson *et al.* [29].

Functions of random variables that have imprecise probability distributions also have imprecise probability distributions. Methods exist for calculating these convolutions [30, 31, 32, 33]. Wang [34] proposes a new interval arithmetic that could be used as well.

Therefore, p-boxes are a very general way to represent uncertainty. For computational purposes, in the approach below, we will convert a p-box into a “canonical” Dempster-Shafer structure [28], which will necessarily be bounded.

DESIGN OPTIMIZATION WITH IMPRECISE PROBABILITIES

In the imprecise probability design optimization (IPDO) problem, there is a set of deterministic design variables for which the designer chooses values and a set of random variables, which may be manufacturing errors, uncertain engineering parameters, or other sources of uncertainty. Unlike other work, this formulation does not include in the model “random design variables.” Such variables are typically those in which the designer chooses the mean, but the actual value is random. In the IPDO formulation presented here, each such quantity is modeled with two quantities: a deterministic design variable and a random parameter that represents the error of that variable. This does not limit the scope of the model. For instance, suppose we have a “random design variable” X that is a dimension of a part. The mean of X , denoted μ_X , is chosen by the designer, but the dimension is a normally distributed random variable with a standard deviation of σ . Examples of this type of variable have been considered in Zhou and Mourelatos [12] and elsewhere. In this formulation, we replace the variable X with $X = d_X + Z_X$, where the first term, which corresponds to the mean, is a deterministic design variable, and the second term is a random variable that is normally distributed with a mean of 0 and a standard deviation of σ .

The general IPDO is formulated as follows:

$$\begin{aligned} & \min_d V[f(d, Z)] \\ \text{s.t. } & \bar{P}\{g_i(d, Z) \leq 0\} \leq p_i \quad i = 1, \dots, n \\ & d^L \leq d \leq d^U \end{aligned} \quad (1)$$

In this formulation, $d \in R^k$ is the vector of deterministic design variables, and $Z \in R^r$ is the vector of random variables that have imprecise probability distributions. The probabilistic constraints are functions of the deterministic design variables and the random variables. We want $g_i(d, Z) \geq 0$ (which is the “safe region”) but will be satisfied if the upper bound on the failure probability is less than the target p_i . We choose the upper probability in order to be conservative.

The function f is the system performance, which may be random, in which case the function V is a moment of that random performance, such as the upper limit for the mean; thus V is a deterministic function of d . In many cases, the objective is specified as a function of only the deterministic design variables, in which case we get the following formulation:

$$\begin{aligned} & \min_{\mathbf{d}} f(\mathbf{d}) \\ \text{s.t. } & P\{g_i(\mathbf{d}, \mathbf{Z}) \leq 0\} \leq p_i \quad i = 1, \dots, n \\ & \mathbf{d}^L \leq \mathbf{d} \leq \mathbf{d}^U \end{aligned} \quad (2)$$

First, this formulation has the usual difficulty of computing the failure probability for each constraint. Analytically evaluating the failure probability is possible only in special cases. An additional complication is the imprecision of the random variables, which makes applying standard RBDO techniques difficult.

Now, $\mathbf{Z}$ has an imprecise joint probability distribution, which can be considered as a set H of precise joint probability distributions. For any precise joint probability distribution $F_j \in H$, let $P_j\{g_i(\mathbf{d}, \mathbf{Z}) \leq 0\}$ be the probability of violating constraint i when $\mathbf{Z}$ has that precise joint probability distribution. Then, we could reformulate the IPDO as the following RBDO:

$$\begin{aligned} & \min_{\mathbf{d}} f(\mathbf{d}) \\ \text{s.t. } & P_j\{g_i(\mathbf{d}, \mathbf{Z}) \leq 0\} \leq p_i \quad i = 1, \dots, n, F_j \in H \\ & \mathbf{d}^L \leq \mathbf{d} \leq \mathbf{d}^U \end{aligned} \quad (3)$$

Unfortunately, because of the large number of constraints, this reformulation is not helpful unless the set H is limited to a reasonable number of “extreme” distributions that can be used as surrogates for the entire set. Research on this topic is ongoing and may provide a way to increase the computational efficiency of IPDO in the future.

Due to these difficulties, we will pursue a numerical approach. To do this, we will first partition the constraints $g_i(\mathbf{d}, \mathbf{Z}) \geq 0$ into two sets. Set S_1 includes any constraint that can be rearranged so that $g_i(\mathbf{d}, \mathbf{Z}) = a_i(h_i(\mathbf{Z}) - b_i(\mathbf{d}))$, where a_i is a positive scalar. Note $g_i(\mathbf{d}, \mathbf{Z}) \leq 0$ if and only if $h_i(\mathbf{Z}) \leq b_i(\mathbf{d})$. The constraints that cannot be rearranged in this way are placed in set S_2 .

For each constraint in S_1 , we will perform the convolution needed to get the imprecise distribution of $h_i(\mathbf{Z})$ by combining the Dempster-Shafer structures for the relevant random variables. Because the upper cumulative probability distribution will be a discontinuous function, we will approximate it with $F_i(x) \approx \bar{P}\{h_i(\mathbf{Z}) \leq x\}$. Therefore, we can replace each of the constraints in S_1 by $F_i(b_i(\mathbf{d})) \leq p_i$.

SOLUTION APPROACH

To solve the IPDO, we will use a sequential approach similar to that of Du and Chen [35] and Zhou and Mourelatos [12].

A key part of the approach is to solve the following deterministic optimization problem P given values for the random variables in each constraint in S_2 . Let $\mathbf{Z}^{i(k)}$ be specific values for the random variables in constraint $i \in S_2$ in iteration k .

$$\begin{aligned} & \min_{\mathbf{d}} f(\mathbf{d}) \\ \text{s.t. } & F_i(b_i(\mathbf{d})) \leq p_i \quad i \in S_1 \\ & g_i(\mathbf{d}, \mathbf{Z}^{i(k)}) \geq 0 \quad i \in S_2 \\ & \mathbf{d}^L \leq \mathbf{d} \leq \mathbf{d}^U \end{aligned} \quad (4)$$

In the space of the design variables, the $g_i(\mathbf{d}, \mathbf{Z}^{i(k)}) \geq 0$ constraints move the boundaries of the “safe region” (by making it smaller) in order to reduce the probability of failure. However, it is still necessary to determine the probability of failure and compare it to the target. If it is too large, then we have to move that constraint some more.

Given these preliminaries, the complete approach follows:

1. Let $k = 0$. For $i \in S_2$, let each component of $\mathbf{Z}^{i(k)}$ equal a value within the range of its expected value.
2. Solve P to get the solution $\mathbf{d}^{(k+1)}$.
3. For $i \in S_2$, evaluate $\bar{P}\{g_i(\mathbf{d}^{(k+1)}, \mathbf{Z}) \leq 0\}$. If $\bar{P}\{g_i(\mathbf{d}^{(k+1)}, \mathbf{Z}) \leq 0\} \leq p_i$ for all $i \in S_2$, then the design point is feasible; stop. Otherwise, for all constraints $i \in S_2$ where $\bar{P}\{g_i(\mathbf{d}^{(k+1)}, \mathbf{Z}) \leq 0\} \leq p_i$, set $\mathbf{Z}^{i(k+1)} = \mathbf{Z}^{i(k)}$.

For the others, find $\mathbf{Z}^{i(k+1)}$ by solving the following problem:

$$\begin{aligned} & \min_{\mathbf{Z}^{i(k+1)}} g_i(\mathbf{d}^{(k+1)}, \mathbf{Z}^{i(k+1)}) \\ \text{s.t. } & \bar{P}\{g_i(\mathbf{d}^{(k+1)}, \mathbf{Z}) \leq g_i(\mathbf{d}^{(k+1)}, \mathbf{Z}^{i(k+1)})\} = p_i \end{aligned} \quad (5)$$

This yields a “very bad” (but not worst-case) value of those random variables used in that constraint. (A technique for solving this problem is described below.)

4. $k = k + 1$. Repeat steps 2 and 3 until a feasible design point is found.