

# Efficient Propagation of Imprecise Probabilities

Jiaxin Zhang and Michael D. Shields

*Department of Civil Engineering, Johns Hopkins University, Baltimore, MD 21212, USA,  
michael.shields@jhu.edu*

**Abstract:** It is often the case that sparse statistical data prohibits the assignment of a precise probability distribution to a given uncertain variable. In these cases, conventional statistical approaches such as Bayesian inference provide uncertain probability distributions - or imprecise probabilities. For convenience, many probabilistic approaches assign a specific distribution based on Maximum Likelihood Estimation (MLE) or some other criteria. Yet propagation of this single MLE distribution ignores potentially importance uncertainty/variability in these variables. Meanwhile, some quasi-probabilistic approaches treat the uncertainties in the distribution using intervals (e.g. mean value of a distribution may be assigned an interval) and propagate a family of distributions to construct a probability-box (or p-box) to bound the probabilities (Beer et al., 2013). This approach is generally computationally very expensive requiring several Monte Carlo analyses to propagate a large number of distributions.

In this work, we propose an approach based on importance sampling to propagate imprecise probability distributions with a single Monte Carlo analysis. The approach uses Bayesian inference to quantify the imprecise probabilities - determining a set of possible candidate distributions (that may come from different families - e.g. normal, lognormal, gamma, etc.) weighted according to their probability of occurrence. We then identify an optimal sampling distribution that best represents all possible candidate distributions. Samples from this optimal sampling distribution are propagated using Monte Carlo simulation and are re-weighted according to the different candidate distributions. Hence, we achieve the propagation of many probability distributions with a single Monte Carlo simulation. A further advantage of the methodology is that the underlying probability models can be updated using Bayesian updating and propagation of these updated distributions does not require additional simulations. Instead, the weights associated with the existing simulations are updated directly to update the probability model of the solution.

**Keywords:** imprecise probability, Bayesian inference, importance sampling

## 1. Introduction

Modern engineering structures and systems, and the models used to represent them, are often characterized by a high degree of complexity. Modeling the influence of various physical phenomena in these systems presents a significant challenge. With the rapid growth of deterministic modeling capabilities, it is becoming widely acknowledged that predictive modeling capabilities can only be achieved by accounting for uncertainties in the modeling process. Therefore, uncertainty quantification plays an important role in computational modeling from design to reliability analysis and risk assessment of complex systems.

A prominent obstacle in uncertainty quantification is the discrepancy between the required information and the available information. This discrepancy is often caused by limitations in the ability to collect sufficient information/data to accurately assign values or probability models to all variables in the problem (i.e. the presence of epistemic uncertainty). For instance, naval ship structures are incredibly complex engineering systems that are subject to highly stochastic marine environments. Further complicating the analysis of these structures is that design specifications provide only nominal values for material and geometric variables and maintenance/inspections are sparse, qualitative, and cover only a small fraction of critical structural components. Although some powerful statistical tools can be used to enable probabilistic modeling given sufficient data of suitable quality, the available data in engineering practice are often limited and of poor quality, which prevents the identification of a precise probability model and its associated parameters.

Imprecise probability approaches have been widely employed and have proven to be effective tools to overcome the issues caused by vague, limited and equivocal information. A series of influential developments on generalized methods has been discussed in the literature (Helton and Oberkampf, 2004; Beer et al., 2013) from various perspectives, including Bayesian approaches (Der Kiureghian et al., 2009), interval methods (Moore and Bierbaum, 1979; Ferson and Hajagos, 2004; Ferson and Ginzburg, 1996), Dempster-Shafer evidence theory (Nelsen et al., 2004) and fuzzy theory (Dubois and Prade, 2005; Stein et al., 2013) among other approaches. More generally, attempts have been made to bring all of these components together under a unifying theory of imprecise probability, which is mainly part of the generalized framework of information theory (Walley, 1991; Walley, 2000). In this framework, probability theory is employed on behalf of stochastic variations in the system variables while epistemic uncertainties associated with the probability laws and their parameters are described using the concepts discussed above.

Traditional statistical inference is based on an assumed probability model (i.e. distribution type) and its parameters (e.g. mean and standard deviation) are estimated using observed data. When sufficient data is provided, these conventional approaches assign a specific distribution with parameters estimated using e.g. maximum likelihood estimation or method of moments. However, it may be impossible to identify a precise distribution type and its parameters given uncertainty due to limited data. This results in many possible/viable probability models with interval or probabilistic model parameters such that uncertainty propagation becomes a major challenge. Existing methods for propagation of imprecise probabilities are usually computationally intractable for all but the simplest problems because they typically require a large number of individual probability studies spanning the range of possible probability distributions and parameters.

This paper proposes an efficient methodology to propagate imprecise probability distributions using Bayesian inference and importance sampling with a single Monte Carlo study. An optimal sampling distribution is identified to represent all candidate distributions by minimizing the overall differences between the sampling distribution and the ensemble of candidate variable distributions characterized by the Hellinger distance and the total variation distance. Samples are generated by this optimal sampling distribution and re-weighted according to importance sampling to simultaneously propagate all candidate distributions. The method has the further advantage that the probability study can be updated directly as additional data is collected. This is enabled through Bayesian updating of the joint model parameter distribution and re-weighting of the samples according to the updated probability laws.

## 2. Review of Important Concepts

### 2.1. BAYESIAN INFERENCE

Bayesian inference holds a central position in data-driven uncertainty quantification and uncertainty propagation in engineering science. As a statistical method, Bayesian inference is used to obtain a posterior PDF  $p(\phi|\mathbf{d}, M)$  for the parameters  $\phi$  of a model class  $M$  using experimental observations  $\mathbf{d}$ . This is achieved through Bayes' rule

$$p(\phi|\mathbf{d}, M) = \frac{p(\mathbf{d}|\phi, M)p(\phi, M)}{p(\mathbf{d}, M)} \quad (1)$$

where  $p(\phi, M)$  is the prior PDF that expresses existing knowledge (or lack thereof) about the parameters,  $p(\mathbf{d}|\phi, M)$  is the likelihood of observing the data  $\mathbf{d}$  from the model class  $M$  with parameters  $\phi$ , and  $p(\mathbf{d}, M)$  is the evidence of model  $M$  and is equal to

$$p(\mathbf{d}, M) = \int p(\mathbf{d}|\phi, M)p(\phi, M)d\phi \quad (2)$$

Evaluating  $p(\mathbf{d}, M)$  is often a non-trivial task because the integration in Eq. (2) is usually analytical intractable for nonlinear and high-dimensional models. To overcome this issue, either Markov Chain Monte Carlo (MCMC) approaches are employed or conjugate distributions are utilized.

The posterior model parameters determined from Eq. (1) are stochastic and are presented in the form of a PDF. From this PDF, the model parameters are frequently assigned using maximum likelihood estimation (MLE) but this ignores the essential variability in distribution parameters, which may have an important role in the results of a probability study - especially when data are sparse, imprecise or limited. Furthermore, the selection of the model  $M$  itself is subject to question unless abundant data are provided or some rationale is used to supply convincing evidence that a specific model form applies (e.g. Central Limit Theorem applies so the distribution should be normal). In the coming sections, we discuss issues of non-unique model selection and the simultaneous propagation of many models having uncertain parameters.

### 2.2. MODEL SELECTION

In the case of sparse data, the selection of a probability model is a substantial challenge. Existing model selection procedures often rank the candidate models based on some widely accepted criteria. Most approaches of model selection can be categorized into three classes: (a) frequentist methods (Guyon et al., 2010), (b) Bayes factor method (Berger and Pericchi, 1996), and (c) methods based on information theory such as the Akaike information criterion (AIC) (Akaike, 1974) and Bayesian Information Criterion (BIC) (Schwartz, 1978). The AIC and BIC, defined as follows, are used here

$$\text{AIC}(M) = 2p - 2L(\hat{\phi}) \quad (3)$$

$$\text{BIC}(M) = p \log n - 2L(\hat{\phi}) \quad (4)$$

where  $L$  is the log-likelihood for model  $M$ ,  $\hat{\phi}$  are the parameter values that maximize the likelihood function,  $p$  is the dimensionality of the parameter space, and  $n$  is the number of observations.

Minimizing the AIC and BIC corresponds to maximizing the posterior model probability for a large amount of data. Again, the AIC and BIC cannot provide an exact means of model selection (especially under limited data), but instead provide evidence that certain types may be valid while others are not.

### 2.3. IMPORTANCE SAMPLING

To improve the accuracy and reduce the cost of stochastic simulation approaches, diverse variance reduction techniques exist. One of the most widely-used approaches is Importance Sampling (IS). Based on the introduction of a proposal sampling density (referred to as the IS density), IS concentrates the computational effort in regions of the uncertain model parameter space that play a more important role in the overall probabilistic performance. Consider a system that involves some design variables  $\mathbf{x}$  and stochastic model parameters  $\boldsymbol{\theta}$ . If the performance function of the system model is given by  $f(\mathbf{x}, \boldsymbol{\theta})$ , then the expected performance is computed as

$$\mathcal{L} = \int f(\mathbf{x}, \boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta} = E_p[f(\mathbf{x}, \boldsymbol{\theta})] \approx \frac{1}{N} \sum_{i=1}^N f(\mathbf{x}, \boldsymbol{\theta}_i) \quad (5)$$

where  $\boldsymbol{\theta}_i$  are independent identically distributed (*i.i.d.*) samples drawn from  $p(\boldsymbol{\theta})$ ,  $E_p[\cdot]$  is the expectation under target distribution  $p(\boldsymbol{\theta})$ , and  $N$  is the number of samples. Importance sampling associated with proposal sampling density  $q(\boldsymbol{\theta})$  transforms the integral to

$$\hat{\mathcal{L}} = \int f(\mathbf{x}, \boldsymbol{\theta})\frac{p(\boldsymbol{\theta})}{q(\boldsymbol{\theta})}q(\boldsymbol{\theta})d\boldsymbol{\theta} = E_q[f(\mathbf{x}, \boldsymbol{\theta})\frac{p(\boldsymbol{\theta})}{q(\boldsymbol{\theta})}] \approx \frac{1}{N} \sum_{i=1}^N f(\mathbf{x}, \boldsymbol{\theta}_i)\frac{p(\boldsymbol{\theta}_i)}{q(\boldsymbol{\theta}_i)} \quad (6)$$

where  $E_q[\cdot]$  is the expectation with respect to  $q(\boldsymbol{\theta})$ . The ratio  $\omega(\boldsymbol{\theta}) = p(\boldsymbol{\theta})/q(\boldsymbol{\theta})$  can be interpreted as the weights of samples generated from  $q(\boldsymbol{\theta})$ . Selection of an efficient proposal sampling density will lead to significant improvement in accuracy (i.e. variance reduction). Conversely, poor IS density selection can increase variance. The optimal choice for the proposal sampling density is derived by minimizing the variance of the estimator  $\hat{\mathcal{L}}$  as

$$q^*(\boldsymbol{\theta}) = \frac{|f(\mathbf{x}, \boldsymbol{\theta})p(\boldsymbol{\theta})|}{\int |f(\mathbf{x}, \boldsymbol{\theta})p(\boldsymbol{\theta})| d\boldsymbol{\theta}} \quad (7)$$

where  $|\cdot|$  means the absolute value. In practice, this choice is intractable since it needs simulation of samples from a PDF that is proportional to the integrand of the expectation and the expectation itself.

### 2.4. HELLINGER DISTANCE AND TOTAL VARIATION DISTANCE

As the choice of optimal sampling distribution is not feasible using Eq. (7), one way is to find a proposal sampling density that has a minimal difference with the optimal one. The Hellinger distance, as a special case of the Csiszár's  $\phi$ -divergence (Csiszar, 1975), can be used to quantify the

difference between probability distributions  $p(\boldsymbol{\theta})$  and  $q(\boldsymbol{\theta})$  as

$$H(P, Q) = \left( \frac{1}{2} \int \left( \sqrt{p(\boldsymbol{\theta})} - \sqrt{q(\boldsymbol{\theta})} \right)^2 d\boldsymbol{\theta} \right)^{\frac{1}{2}} \quad (8)$$

One way to minimize the difference between two distributions is to find the distribution that produces the minimum Hellinger distance estimator (MHDE) (Beran, 1977). In order to obtain the solution of

$$\hat{h} = \arg \min \frac{1}{\sqrt{2}} \left\| \sqrt{p(\boldsymbol{\theta})} - \sqrt{q(\boldsymbol{\theta})} \right\|_2 \quad (9)$$

where  $\|p - q\|_2 = \left( \int \left( \sqrt{p(\boldsymbol{\theta})} - \sqrt{q(\boldsymbol{\theta})} \right)^2 d\boldsymbol{\theta} \right)^{\frac{1}{2}}$  denotes the  $\ell_2$  norm, we instead use the square MHDE which is convenient for expressing the derivative of the optimal sampling distribution in the following section

$$\hat{h}^2 = \arg \min \frac{1}{2} \left\| \sqrt{p(\boldsymbol{\theta})} - \sqrt{q(\boldsymbol{\theta})} \right\|_2^2 \quad (10)$$

where  $\|p - q\|_2^2 = \int \left( \sqrt{p(\boldsymbol{\theta})} - \sqrt{q(\boldsymbol{\theta})} \right)^2 d\boldsymbol{\theta}$  denotes the square  $\ell_2$  norm. This is achieved by constructing a nonparametric proposal density  $q(\boldsymbol{\theta})$  and conducting an optimization with Eq. (10) as the objective.

### 3. Optimal Importance Sampling for Multiple Distributions

To date, importance sampling has always been used with a precisely specified target distribution  $p(\boldsymbol{\theta})$ . However, in the imprecise probability case caused by lack of data, the target distribution cannot be assigned precisely and therefore, the estimation of the optimal sampling distribution becomes a significant challenge. This work first utilizes a Bayesian inference approach to quantify the imprecise probabilities by identifying several viable probability models  $M$  through AIC/BIC and retaining the distributions of their model parameters  $\boldsymbol{\phi}$ . From this, a series of possible candidate target distributions are assigned on the basis of their probability of occurrence. The fundamental difference with classical importance sampling is that a set of probabilistically weighted target distributions are taken into consideration. The crucial issue addressed in this study is how to select an optimal importance sampling density to effectively represent and propagate all candidate target distributions. To do so, we need to identify a sampling distribution that is as close as possible to the ensemble of target distributions.

Our object of maximizing the representativeness of the importance sampling distribution for multiple target distributions can be restated as minimizing the total difference between the importance sampling distribution and the ensemble of target distributions. This is the equivalent of minimizing the distance in the probability space. Utilizing the Hellinger distance, an overall square MHDE with uncertain distribution parameters and distribution types is proposed as

$$\hat{h}^2 = \arg \min \sum_{i=1}^{N_d} \int_{\boldsymbol{\phi}} \frac{1}{2} \left\| \sqrt{p_i(\boldsymbol{\theta}|\boldsymbol{\phi})} - \sqrt{q(\boldsymbol{\theta})} \right\|_2^2 d\boldsymbol{\phi} \quad (11)$$

where  $N_d$  is the number of candidate probability models defined through the target densities  $p_i(\boldsymbol{\theta}|\boldsymbol{\phi})$  having parameters  $\boldsymbol{\phi}$ . An advantage of this Hellinger distance metric is that we can obtain the analytical solution for the optimal proposal sampling. If we set the derivative of the overall square MHDE  $\hat{h}^2$  with respect to  $q(\boldsymbol{\theta})$  as

$$\frac{\partial \hat{h}^2}{\partial q(\boldsymbol{\theta})} = \frac{\partial \sum_{i=1}^{N_d} \int_{\boldsymbol{\phi}} \frac{1}{2} \left( \sqrt{p_i(\boldsymbol{\theta}|\boldsymbol{\phi})} - \sqrt{q(\boldsymbol{\theta})} \right)^2 d\boldsymbol{\phi}}{\partial q(\boldsymbol{\theta})} \quad (12)$$

$$\frac{\partial \hat{h}^2}{\partial q(\boldsymbol{\theta})} = \frac{1}{2} \sum_{i=1}^{N_d} \left( \frac{\partial \int_{\boldsymbol{\phi}} \left( \sqrt{p_i(\boldsymbol{\theta}|\boldsymbol{\phi})} - \sqrt{q(\boldsymbol{\theta})} \right)^2 d\boldsymbol{\phi}}{\partial q(\boldsymbol{\theta})} \right) \quad (13)$$

$$\frac{\partial \hat{h}^2}{\partial q(\boldsymbol{\theta})} = \frac{1}{2} \sum_{i=1}^{N_d} \left( \frac{\partial \int_{\boldsymbol{\phi}} p_i(\boldsymbol{\theta}|\boldsymbol{\phi}) - 2\sqrt{p_i(\boldsymbol{\theta}|\boldsymbol{\phi})q(\boldsymbol{\theta})} + q(\boldsymbol{\theta}) d\boldsymbol{\phi}}{\partial q(\boldsymbol{\theta})} \right) \quad (14)$$

$$\frac{\partial \hat{h}^2}{\partial q(\boldsymbol{\theta})} = \frac{1}{2} \sum_{i=1}^{N_d} \left( \int_{\boldsymbol{\phi}} \left( 1 - \sqrt{\frac{p_i(\boldsymbol{\theta}|\boldsymbol{\phi})}{q(\boldsymbol{\theta})}} \right) d\boldsymbol{\phi} \right) \quad (15)$$

For the special case of two parameter distributions with  $\boldsymbol{\phi} = (\mu, \sigma)$  where  $\mu$  and  $\sigma$  are the mean and standard deviation respectively, we have

$$\frac{\partial \hat{h}^2}{\partial q(\boldsymbol{\theta})} = \frac{1}{2} \sum_{i=1}^{N_d} \left( \int_{\mu} \int_{\sigma} \left( 1 - \sqrt{\frac{p_i(\boldsymbol{\theta}|\mu, \sigma)}{q(\boldsymbol{\theta})}} \right) d\mu d\sigma \right) \quad (16)$$

By partitioning the joint probability density of model parameters  $p_i(\boldsymbol{\theta})$  for each distribution type, we extract a finite but representative number of candidate target distributions from the infinite set. The derivative can then be approximated using discretized means  $\mu_j (j = 1, 2, \dots, N_{mi})$  and standard deviations  $\sigma_k (k = 1, 2, \dots, N_{si})$  as

$$\frac{\partial \hat{h}^2}{\partial q(\boldsymbol{\theta})} \approx \frac{1}{2} \sum_{i=1}^{N_d} \left( \sum_{j=1}^{N_{mi}} \sum_{k=1}^{N_{si}} \left( 1 - \sqrt{\frac{p_i(\boldsymbol{\theta}|\mu_j, \sigma_k)}{q(\boldsymbol{\theta})}} \right) \right) \quad (17)$$

where  $N_{mi} \times N_{si}$  represents the total number of uncertain distribution parameters for model class  $M_i$  (e.g. mean and standard deviation for two-parameter distributions). Setting  $\frac{\partial \hat{h}^2}{\partial q(\boldsymbol{\theta})} = 0$  and (for simplicity) assuming  $N_{mi} = N_m \forall i$  and  $N_{si} = N_s \forall i$ , we have

$$\sum_{i=1}^{N_d} \sum_{j=1}^{N_m} \sum_{k=1}^{N_s} 1 - \sum_{i=1}^{N_d} \sum_{j=1}^{N_m} \sum_{k=1}^{N_s} \sqrt{\frac{p_i(\boldsymbol{\theta}|\mu_j, \sigma_k)}{q(\boldsymbol{\theta})}} = N_d \cdot N_m \cdot N_s - \frac{\sum_{i=1}^{N_d} \sum_{j=1}^{N_m} \sum_{k=1}^{N_s} \sqrt{p_i(\boldsymbol{\theta}|\mu_j, \sigma_k)}}{\sqrt{q(\boldsymbol{\theta})}} = 0 \quad (18)$$

Therefore, the optimal proposal sampling density will be

$$q^*(\boldsymbol{\theta}) = \left( \frac{\sum_{i=1}^{N_d} \sum_{j=1}^{N_m} \sum_{k=1}^{N_s} \sqrt{p_i(\boldsymbol{\theta}|\mu_j, \sigma_k)}}{N_d \cdot N_m \cdot N_s} \right)^2 \quad (19)$$

However, as a nonparametric model,  $q^*$  is not always a PDF that satisfies  $\int q^*(\boldsymbol{\theta})d\boldsymbol{\theta} = 1$  unless all candidate target distributions are the same. Actually, all target distributions are different here, so it will be improper to choose  $q^*(\boldsymbol{\theta})$  as the proposal sampling density. In order to ensure  $q^*(\boldsymbol{\theta})$  is a valid PDF, we may alternatively consider the optimization according to the total variation distance  $\Delta(P, Q)$  defined through the following  $\ell_1$  norm

$$\Delta(P, Q) = \max \|P - Q\|_1 = \frac{1}{2} \sum_{\boldsymbol{\theta}} |p(\boldsymbol{\theta}) - q(\boldsymbol{\theta})| \quad (20)$$

thanks to the following property stated by (Lindsay, 1994) relating it to the Hellinger distance

$$H^2(P, Q) \leq \Delta(P, Q) \leq \sqrt{2}H(P, Q) \quad (21)$$

Thus, an equivalent way to obtain the optimal proposal sampling distribution is to minimize the total variation distance in Eq. (20), which yields

$$\hat{q}^*(\boldsymbol{\theta}) = \frac{1}{N_d \cdot N_m \cdot N_s} \sum_{i=1}^{N_d} \sum_{j=1}^{N_m} \sum_{k=1}^{N_s} p_i(\boldsymbol{\theta}|\mu_j, \sigma_k) \quad (22)$$

where  $\hat{q}^*(\boldsymbol{\theta})$  is a nonparametric density estimator that consists of multiple mixture distributions. If the weights of different distribution types and parameters are identical, the coefficient for each candidate target will be equal. More generally, the optimal proposal density can be written with different weighting coefficients

$$\hat{q}^*(\boldsymbol{\theta}) = \sum_{i=1}^{N_d} \sum_{j=1}^{N_m} \sum_{k=1}^{N_s} \lambda_{jk}^i \cdot p_i(\boldsymbol{\theta}|\mu_j, \sigma_k) \quad (23)$$

where  $\lambda_{jk}^i$  is the weighting coefficient for the  $i^{th}$  distribution type with parameters of the  $j^{th}$  mean and the  $k^{th}$  standard deviation, and satisfies

$$\sum_{i,j,k} \lambda_{jk}^i = 1 \quad (24)$$

These coefficients may be derived from probabilistic considerations that account for the probability of occurrence of distribution parameters and model classes from the provided data.

#### 4. Propagation of Imprecise Probability

By the discussion above, there are theoretically infinite candidate target distributions. Even after discretization, thousands of candidate distributions remain. Existing algorithms are often very computationally intensive since individual probability studies (Monte Carlo analyses) are needed to propagate each distribution separately. The method proposed here utilizes importance sampling to propagate these thousands of distributions simultaneously through a single Monte Carlo study

by generating samples from the optimal proposal sampling density in Eq. (22), and reweighting the samples according to each target. Specifically, after sampling from  $q^*(\boldsymbol{\theta})$ , importance sampling estimators are constructed for each candidate target  $p_i(\boldsymbol{\theta}|\mu_j, \sigma_k)$  according to Eq. (6) and denoted  $\hat{\mathcal{L}}_{jk}^i, i = 1, \dots, N_d, j = 1, \dots, N_m, k = 1, \dots, N_s$ . Note that all estimates  $\hat{\mathcal{L}}_{jk}^i$  are constructed from the same set of samples drawn from  $q^*(\boldsymbol{\theta})$ , greatly reducing the computational cost over existing methods.

This procedure generates a set of probability distributions for the system performance that can be further assessed probabilistically. Given the Bayesian nature of the model construction, each of the probability distributions has an associated probability of occurrence derived from the joint density of its model parameters  $\boldsymbol{\phi}$ . Consequently, these distributions provide a rich source of information that goes beyond, for example, probability bounds that are typically available (in the form of a p-box for instance) to give a fully probabilistic measure of the output. This can be used to evaluate quantities like the PDF/CDF for statistical quantities or probability of failure that account for uncertainties associated with insufficient data.

A final advantage of the proposed method is that it can be readily updated to incorporate new data as it is collected. Using Bayesian updating, the joint parameter distribution  $p(\boldsymbol{\phi})$  is updated and the associated candidate target distributions are updated. However, it is not strictly necessary to identify a new optimal sampling density and the existing samples can be used for uncertainty propagation by re-weighting according to the importance sampling weights with the new target densities. Note that this does not ensure a low-variance estimator. In fact, the sampling density after updating is no longer optimal and may yield an increase in variance. If necessary, this variance can be reduced by either adding samples from  $q^*(\boldsymbol{\theta})$  or solving for a new optimal PDF and resampling.

## 5. Numerical Example

The proposed methodology and its application are illustrated by an example considering the probabilistic assessment of plate buckling strength with uncertain geometric and material parameters. Carlsen (Carlsen, 1977) derived an analytical expression to evaluate the normalized buckling strength  $\psi$  of a simply supported plate in uniaxial compression considering the effects of non-dimensional initial deflections  $\delta_0$  and residual stress  $\eta t$  resulting from welding as

$$\psi = \frac{\sigma_b}{\sigma_0} = \left( \frac{2.1}{\lambda} - \frac{0.9}{\lambda^2} \right) \left( 1 - \frac{0.75\delta_0}{\lambda} \right) \left( 1 - \frac{2\eta t}{b} \right) \quad (25)$$

where  $\sigma_b$  is the stress at which buckling occurs,  $\lambda = b/t\sqrt{\sigma_0/E}$  is referred to as the slenderness of the plate with width  $b$ , thickness  $t$ , yield stress  $\sigma_0$ , and elastic modulus  $E$ . These material and geometric variabilities are estimated from the data presented by Hess et al. (Hess et al., 2002) and Soares (Soares, 1988) as presented in Table 1. According to a global sensitivity analysis (Saltelli, 2008), the yield stress shows the highest influence on the buckling strength accounting for nearly half of its variance, and therefore will be the main concern of this work.

## Efficient Propagation of Imprecise Probabilities

Table I. Statistics for plate material, geometry and imperfection variables from Hess et al. (Hess et al., 2002) and Soares (Soares, 1988).

| Variables  | Physical Meaning   | Nominal Value | Mean        | COV    | Sensitivity Index |
|------------|--------------------|---------------|-------------|--------|-------------------|
| $b$        | Width              | 24            | 0.992*24    | 0.028  | 0.017             |
| $t$        | Thickness          | 0.5           | 1.05*0.5    | 0.044  | 0.045             |
| $\sigma_0$ | Yield Strength     | 34            | 1.3*34      | 0.1235 | 0.482             |
| $E$        | Elastic Modulus    | 29000         | 0.987*29000 | 0.076  | 0.194             |
| $\delta_0$ | Initial Deflection | 0.35          | 1.0*0.35    | 0.05   | 0.043             |
| $\eta$     | Residual Stress    | 5.25          | 1.0*5.25    | 0.07   | 0.233             |

We start with a collection of 25 yield stress values synthetically generated according to a normal distribution with parameters given in Table I. All other variables are assumed deterministic and taking their mean values. Next, we use the AIC (Eq. (3)) and BIC (Eq. (4)) to select viable distribution types based on the limited data. Table II presents the AIC and BIC values for several candidate distribution types. Note that only the Rayleigh and Exponential distributions stand out as having particularly high AIC/BIC and therefore we do not consider these two types as representative of our data. This leaves 11 distribution types that we consider as viable.

Table II. The rank sequence of distribution type using AIC and BIC.

| Rank sequence | Distribution type | Akaike Information Criterion (AIC) | Bayesian Information Criterion (BIC) |
|---------------|-------------------|------------------------------------|--------------------------------------|
| 1             | Inverse Gaussian  | 160.9097                           | 163.3474                             |
| 2             | Birnbaum-Saunders | 160.9122                           | 163.3499                             |
| 3             | Lognormal         | 160.9450                           | 163.3828                             |
| 4             | Gamma             | 161.3044                           | 163.7422                             |
| 5             | Loglogistic       | 161.7626                           | 164.2004                             |
| 6             | Nakagami          | 161.7886                           | 164.2264                             |
| 7             | Rician            | 162.3961                           | 164.8338                             |
| 8             | Normal            | 162.4284                           | 164.8662                             |
| 9             | Logistic          | 162.8088                           | 165.2465                             |
| 10            | Weibull           | 166.3174                           | 168.7552                             |
| 11            | Extreme Value     | 170.0299                           | 172.4676                             |
| 12            | Rayleigh          | 208.0995                           | 209.3183                             |
| 13            | Exponential       | 241.5110                           | 242.7299                             |

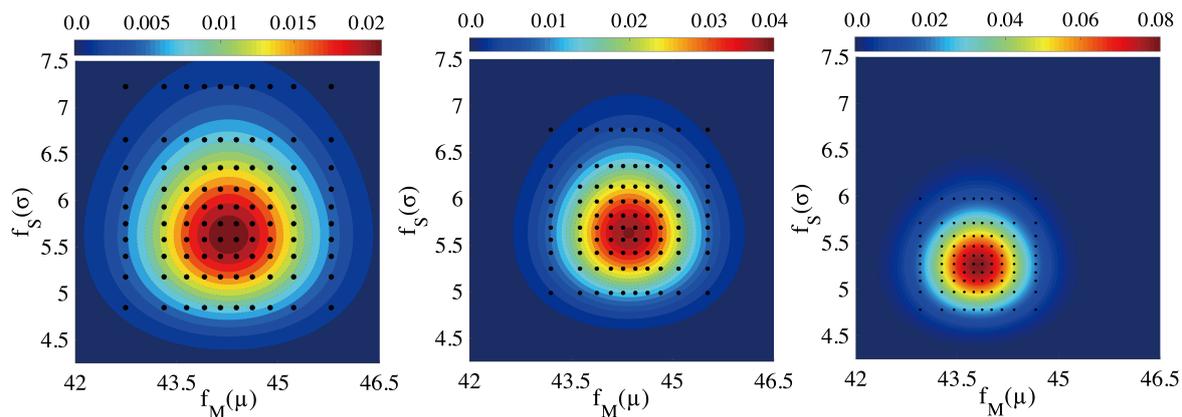


Figure 1. Joint posterior distribution for mean and standard deviation from Bayesian inference using datasets with 25 (left), 50 (middle) and 100 (right) yield stress values.

Using Bayesian inference, the joint posterior distribution of the mean and standard deviation are obtained from the initial 25 data. An equal partition of the joint probability distribution into a  $10 \times 10$  grid yields 100 distinct values of the model parameters shown by the black dots in Figure 1. These pairs of mean and standard deviation values are then used with each of the 11 distribution types to produce a representative set of target distributions shown by the gray densities in Figure 2. When more data are collected, the joint posterior distribution of  $(\mu, \sigma)$  is updated using Bayesian updating, as shown by the contours in Figure 1 for 25, 50, and 100 yield stress values. Again, the joint density is discretized through an equal partitioning of the probability space shown by the black dots.

The optimal sampling distribution is constructed analytically based on these 11 candidate target distributions and 100 possible parameter values (i.e. from a total of 1100 candidate target distributions - each equally probable). The optimal sampling density is shown by the black curve in Figure 2. The suite of distributions in Figure 2 are propagated through Eq. (25) by generating 5000 samples from the optimal sampling density and applying the importance sampling weights for each distribution. This propagation results in the gray band of CDFs for the compressive strength shown in Figure 3. Note that we do not update the optimal sampling density as data is gathered. This can be observed in Figure 2 where it is seen that the gray band of candidate target densities narrows but the black optimal sampling density remains the same. Hence, the optimal sampling density is no longer optimal in the right and middle images but remains sufficient for propagation and no new samples need to be generated. All of the CDFs shown in Figure 3 (from left to right) have been generated from the same 5000 random samples. Notice also that the band of CDFs in Figure 3 narrows considerably as additional yield stress values are collected. This is expected as the additional data reduces the epistemic uncertainty.

Next, consider that failure occurs when  $\psi < 0.56$  - shown by the dashed vertical line in Figure 3. Given the equal partitioning of the probability space, each of the CDFs resulting from a given input distribution in Figure 3 is equally probable (note that the distribution types are not necessarily equally probable). We can therefore determine the empirical CDF of  $P(\psi < 0.56)$  conditioned on the input distribution type as shown in Figure 4. Each of these CDFs has 100 values corresponding

## Efficient Propagation of Imprecise Probabilities

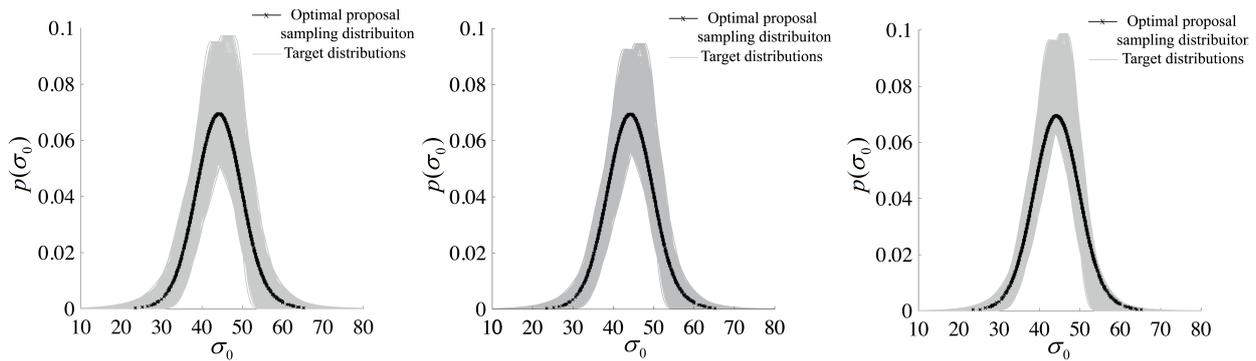


Figure 2. Ensemble of candidate target densities and the optimal proposal sampling density for yield stress from datasets with 25 (left), 50 (middle) 100 (right) yield stress values.

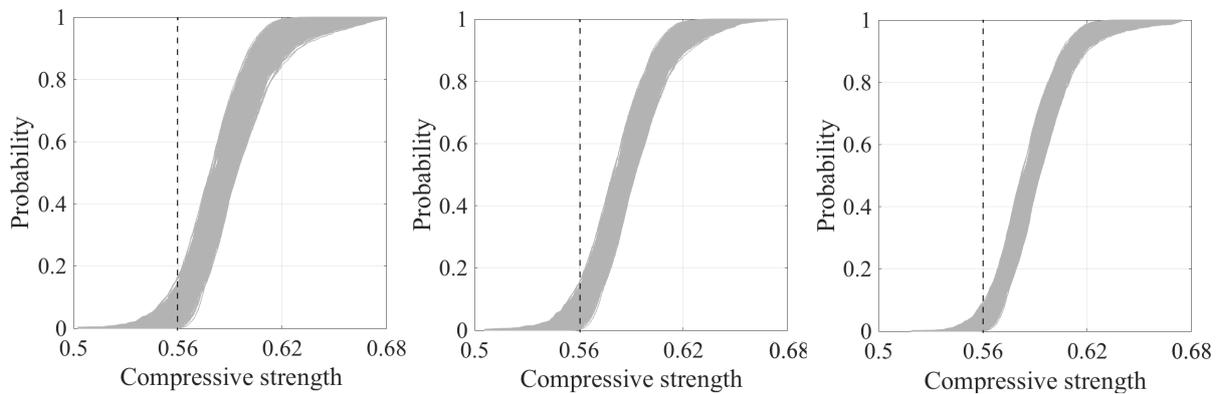


Figure 3. Ensemble of cumulative distribution functions with uncertain yield stress derived from propagation of uncertain probability models constructed using 25 (left), 50 (middle), and 100 (right) yield stress values.

to the 100 points in the  $10 \times 10$  grid for the partitioning of the joint parameter probability space. Notice again that when there is little data, the CDFs are wide with  $P(\psi < 0.56)$  in the range  $[0, 0.18]$ . However, as data is collected the range narrows considerably such that  $P(\psi < 0.56)$  in the range  $[0, 0.09]$ . Theoretically, as a very large amount of data is collected these CDFs will converge toward a step function such that  $P(\psi < 0.56)$  is known with certainty for each assumed input distribution.

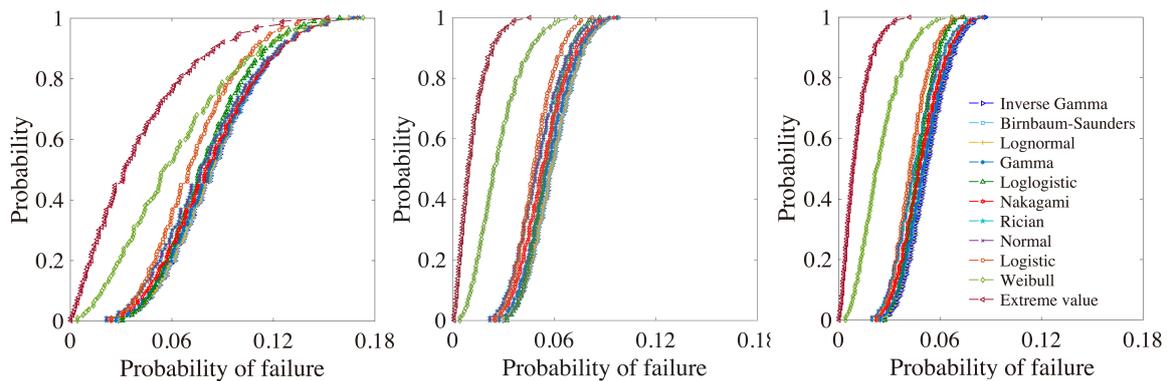


Figure 4. Empirical CDF of the probability of plate buckling failure  $P(\psi < 0.56)$  with uncertain yield stress derived from propagation of uncertain probability models constructed using 25 (left), 50 (middle), and 100 (right) yield stress values..

## 6. Conclusion

Bayesian inference is used to quantify the uncertainty associated with a distribution derived from data. Yet, in the almost universal case of lack of complete data, the derived distribution cannot be precisely specified. We propose an efficient approach, based on Importance Sampling, for propagating uncertain probability distributions. The method identifies an optimal sampling distribution that is representative of the possible range of distributions and adaptively reweights the samples to simultaneously propagate the full range. An advantage of this approach is that the underlying probability models can be simply updated using Bayesian updating and we don't need additional simulations for propagation of the updated distributions. Instead, the existing simulations are reweighted and updated directly to update the probability model.

## Acknowledgements

The work presented herein has been supported by the Office of Naval Research under Award Number N00014-15-1-2754 with Dr. Paul Hess as program officer.

## References

- Beer, M., S. Ferson and V. Kreinovich. Imprecise probabilities in engineering analyse. *Mechanical Systems and Signal Processing*, 37(1-2):4–29, 2013.
- Helton, J. C. and W. L. Oberkampf. Alternative representations of epistemic uncertainty. *Reliability Engineering and System Safety*, 85(1):1–10, 2004.
- Der Kiureghian, A. and O. Ditlevsen. Aleatory or epistemic? Does it matter?. *Structural Safety*, 31(2):105–112, 2009.
- Moore, R. E. and F. Bierbaum. *Methods and applications of interval analysis*. Vol. 2. Philadelphia: Siam, 1979.

- Ferson, S. and J. G. Hajagos. Arithmetic with uncertain numbers: rigorous and (often) best possible answers. *Reliability Engineering and System Safety* , 85(1):135–152, 2004.
- Ferson, S. and L. R. Ginzburg. Different methods are needed to propagate ignorance and variability. *Reliability Engineering and System Safety* , 54(2):133–144, 1996.
- Nelsen, R. B., J. Hajagos, D. J. Berleant, J. Zhang, W. T. Tucker, L. R. Ginzburg and W. L. Oberkampf. *Dependence in probabilistic modeling, Dempster-Shafer theory, and probability bounds analysis*. Sandia National Laboratories, 2004.
- Ferson, S., V. Kreinovich, L. Ginzburg, D. S. Myers and K. Sentz. *Constructing probability boxes and Dempster-Shafer structures*. Vol. 835. Sandia National Laboratories, 2002.
- Dubois, D. and H. Prade. *Interval-valued Fuzzy Sets, Possibility Theory and Imprecise Probability*. In EUSFLAT Conf, pages 314–319, 2005.
- Stein, M., M. Beer and V. Kreinovich. Bayesian approach for inconsistent information. *Information sciences*, 245:96–111, 2013.
- Walley P. Statistical reasoning with imprecise probabilities. London: Chapman and Hall, 1991.
- Walley P. Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning*, 24(2):125–148, 2000.
- Guyon, I., A. Saffari, G. Dror and G. Cawley. Model selection: Beyond the bayesian/frequentist divide. *The Journal of Machine Learning Research*, 11:61–87, 2010.
- Berger, J. O. and L. R. Pericchi. The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association*, 91(433):109–122, 1996.
- Akaike, H. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6):716–723, 1974.
- Schwarz, G. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.
- Csiszr, I. I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, pages 146–158, 1975.
- Beran, R. Minimum Hellinger distance estimates for parametric models. *The annals of statistics*, pages 445–463, 1977.
- Lindsay, B. G. Efficiency versus robustness: the case for minimum Hellinger distance and related methods. *The annals of statistics*, pages 1081–1114, 1994.
- Carlsen, C. Simplified collapse analysis of stiffened plates. *Norwegian Maritime Research*, 5(4):20–36, 1977.
- Hess, P., D. Bruchman, I. Assakkaf and B. Ayyub. Uncertainties in material and geometric strength and load variables. *Naval Engineers Journal* , 114(2):139–166, 2002.
- Soares, C. G. Uncertainty modelling in plate buckling. *Structural Safety* , 5(1):17–34, 1988.
- Saltelli, A., F. Campolongo, J. Caliboni, D. Gatelli, F. Pennoni, M. Ratto, M. Saisana and S. Tarantola. *Global Sensitivity Analysis: The Primer*. John Wiley and Sons, Inc., 2008.

