Fuzzy Data and Imprecise Probabilities in Engineering

Beer, Michael
University of Liverpool, School of Engineering, Centre for Engineering Sustainability
The Quadrangle
Liverpool L69 3GH, United Kingdom
E-mail: beer.micha@gmail.com

ABSTRACT

Predicting the behavior and reliability of engineering structures and systems is often plagued by uncertainty and imprecision caused by sparse data, poor measurements and subjective information. Accounting for such limitations complicates the mathematical modeling required to obtain realistic results in engineering analyses. The framework of imprecise probabilities provides a mathematical basis to deal with these problems which involve both probabilistic and non-probabilistic sources of uncertainty. A common feature of the various concepts of imprecise probabilities is the consideration of an entire set of probabilistic models in one analysis. But there are differences between the concepts in the mathematical description of this set and in the theoretical connection to the probabilistic models involved. This study is focused on fuzzy probabilities, which combine a probabilistic characterization of variability with a fuzzy characterization of imprecision. We discuss how fuzzy modeling can allow a more nuanced approach than interval-based concepts. The application in an engineering analysis is demonstrated by means of an example.

INTRODUCTION

The analysis and reliability assessment of engineering structures and systems involves uncertainty and imprecision in parameters and models of different type. In order to derive predictions regarding structural behavior and reliability, it is crucial to represent the uncertainty and imprecision appropriately according to the underlying real-world information which is available. To capture variation of structural parameters, established probabilistic models and powerful simulation techniques are available for engineers, which are widely applicable to real-world problems; for example, see (Schenk and Schuëller 2005). The required probabilistic modeling can be realized via classical mathematical statistics if data of a suitable quality are available to a sufficient extent.

In civil engineering practice, however, the available data are frequently quite limited and of poor quality. These limitations create epistemic uncertainty, which can sometimes be substantial. It is frequently argued that expert knowledge can compensate for the limitations through the use of Bayesian methods based on subjective probabilities. If a subjective perception regarding a probabilistic model exists and some data for a model update can be made available, a Bayesian approach can be very powerful, and meaningful results with maximal information content can be derived. Bayesian approaches have attracted increasing attention in the recent past and considerable advancements have been reported for the solution of various engineering problems (Papadimitriou et al. 2001, Igusa et al. 2002, Der Kiureghian and Ditlevsen 2009). An important feature of Bayesian updating is that the subjective influence in the model assumption decays quickly with growing amount of data. It is then reasonable practice to estimate probabilistic model parameters based on the posterior distribution, for example, as the expected value thereof.

When less information and experience are available, greater difficulties will be faced. If the available information is very scarce and is of an imprecise nature rather than of a stochastic nature, a subjective probabilistic model description may be quite arbitrary. For example, a distribution parameter may be known merely in the form of bounds. Any prior distribution which is limited to these bounds would then be an option for modeling. But the selection of a particular model would introduce unwarranted information that cannot be justified sufficiently. Even the assumption of a uniform distribution, which is commonly used in those cases, ascribes more information than is actually given by the bounds. This situation may become critical if no or only very limited data are available for a model update. The initial subjectivity is then dominant in the posterior distribution and in the final result. If these results, such as failure probabilities, determine critical decisions, one may wish to consider the problem from the following angle.

If several probabilistic models are plausible for the description of a problem, and no information is available to assess the suitability of the individual models or to relate their suitability with respect to one another, then it may be of interest to identify the worst case for the modeling rather than to average over all plausible model options with arbitrary weighting. The probabilistic analysis is carried out conditional on each of many particular probabilistic models out of the set of plausible models. In reliability assessment, this implies the calculation of an upper bound for the failure probability as the worst case. This perspective can be extended to explore the sensitivity of results with respect to the variety of plausible models, that is, with respect to a subjective model choice. A mathematical framework for an analysis of this type has been established with imprecise probabilities (see Walley 1991). Applications to reliability analysis (Kozine and Filimonov 2000, Möller et al. 2003, Utkin 2004) and to sensitivity analysis (Ferson and Tucker 2006, Hall 2006) have been reported. This intuitive view, however, is by far not the entire motivation for imprecise probabilities (see Klir 2006). Imprecise probabilities are not limited to a consideration of imprecise distribution parameters. They are also capable of dealing with imprecise conditions and dependencies between random variables and with imprecise structural parameters and model descriptions. They allow statistical estimations and tests with imprecise sample elements. Results from robust statistics in form of solution domains of statistical estimators can be considered directly and appropriately (Augustin and Hable 2010).

In this paper, the implementation of intervals and fuzzy sets as parameters of probabilistic models is discussed in the context of proposed concepts of imprecise probabilities. A structural reliability analysis is employed to illustrate the effects in an example.

PROBABILISTIC MODELS WITH IMPRECISE PARAMETERS

In engineering analyses, parameters of probabilistic models are frequently limited in precision and are only known in a coarse manner. This situation can be approached with different mathematical concepts. First, the parameter can be considered as uncertain with random characteristics, which complies with the Bayesian approach. Subjective probability distributions for the parameters are updated by means of objective information in form of data. The result is a mix of objective and subjective information – both expressed with probability. Second, the parameter can be considered as imprecise but bounded within a certain domain, where the domain is described as a set. In this manner, only the limitation to some domain and no further specific characteristics are ascribed to the parameter, which introduces significantly less information in comparison with a distribution function as used in the Bayesian approach. Imprecision in the form of a set for a parameter does not migrate into probabilities, but it is reflected in the result as a set of probabilities which contains the true probability. Intervals and fuzzy sets can thus be considered as models for parameters of probability distributions.

An interval is an appropriate model in cases where only a possible range between crisp bounds x_l and x_r is known for the parameter x, and no additional information concerning value frequencies, preference, etc. between interval bounds is available nor any clues on how to specify such information. Interval modeling of a parameter of a probabilistic model connotes the consideration of a set of probabilistic models, which are captured by the set of parameter values

$$X_I = \{ x \mid x \in [x_l, x_r] \} \ . \qquad (1)$$

This modeling corresponds to p-box approach (Ferson and Hajagos 2004) and to the theory of interval probabilities (Weichselberger 2000). Events E_i are assessed with a range of probability, $\left[P_l\left(E_i\right), P_r\left(E_i\right)\right] \subseteq [0,1]$, which is directly used for the definition of interval probability, denoted as *IP*, as follows,

$$IP: \; E_\Omega \rightarrow I \;\; with$$
$$E_\Omega = \mathfrak{P}(\Omega),\; I = \left\{[a,b] \,\forall a,b \,\middle|\, 0 \le a \le b \le 1\right\} \;. \tag{2}$$

In Eq. (2), $\mathfrak{P}(\Omega)$ is the power set on the set Ω of elementary events ω. This definition complies with traditional probability theory. Kolmogorov's axioms and the generation scheme of events are retained as defined in traditional probability theory (see also Yamauchi and Mukaidono 1999). Traditional mathematical statistics are applicable for quantification purposes. In reliability analysis with interval probabilities, the parameter interval X_I is mapped to an interval of the failure probability,

$$X_I \rightarrow P_{fI} = \left\{P_f \,\middle|\, P_f \in \left[P_{fl}, P_{fr}\right]\right\} \;. \tag{3}$$

Scrutinizing the modeling of parameters as intervals shows that an interval is a quite crude expression of imprecision. The specification of an interval for a parameter implies that, although a number’s value is not known exactly, exact bounds on the number can be provided. This may be criticized because the specification of precise numbers is just transferred to the bounds. Fuzzy set theory provides a suitable basis for relaxing the need for precise values or bounds. It allows the specification of a smooth transition for elements from belonging to a set to not belonging to a set. Fuzzy numbers are a generalization and refinement of intervals for representing imprecise parameters. The essence of an approach using fuzzy numbers that distinguishes it from more traditional approaches is that it does not require the analyst to circumscribe the imprecision all in one fell swoop with finite characterizations having known bounds. The analyst can now express the available information in form of a series of plausible intervals, the bounds of which may grow, including the case of infinite limits. This allows a more nuanced approach compared to interval modeling.

Fuzzy sets provide an extension to interval modeling that considers variants of interval models, in a nested fashion, in one analysis. A fuzzy set $\tilde{X}$ of parameter values can be represented as a set of intervals X_I,

$$\tilde{X} = \left\{ \left(X_\alpha, \mu\left(X_\alpha\right)\right) \,\middle|\, \begin{array}{l} X_\alpha = X_I, \\ \mu\left(X_\alpha\right) = \alpha \\ \qquad \forall \alpha \in (0,1] \end{array} \right\} , \tag{4}$$

which is referred to as α-discretization; see Figure 1 (Zimmermann 1992). In Eq. (4), X_α denotes an α-level set of the fuzzy set $\tilde{X}$, and $\mu(.)$ is the membership function. This modeling applied to parameters of a probabilistic model corresponds to the theory of fuzzy random variables and to fuzzy probability theory. Detailed discussions are provided, for example, in (Kruse and Meyer 1987, Li et al. 2002, Gil et al. 2006, Beer 2009, Viertl 1996, Viertl 2011). The definition of a fuzzy random variable refers to imprecise observations as outcome of a random experiment. A fuzzy random variable $\tilde{\mathrm{Y}}$ is the mapping

$$\tilde{\mathrm{Y}}: \; \Omega \rightarrow \mathfrak{F}(\mathbf{Y}) \tag{5}$$

with $\mathfrak{F}(\mathbf{Y})$ being the set of all fuzzy sets on the fundamental set **Y**, whereby the standard case is $\mathbf{Y} = \mathbb{R}^n$. The pre-images of the imprecise events described by $\mathfrak{F}(\mathbf{Y})$ are elements of a traditional probability space $\left[\Omega, \mathfrak{S}, P\right]$. This complies with traditional probability theory and allows statistics with imprecise data (Kruse and Meyer 1987, Bandemer and Näther 1992, Viertl 1996, Viertl 2011). As a consequence of Eq. (5), parameters of probabilistic models, including descriptions of the dependencies and distribution type, and probabilities are obtained as fuzzy sets. This builds the relationship to the p-box approach and to the theory of interval probabilities. A representation of a fuzzy probability distribution function of a fuzzy random variable $\tilde{Y}$ with aid of α-discretization leads to interval probabilities $\left[F_{\alpha l}(y), F_{\alpha r}(y)\right]$ for each α-level as one plausible model variant,

$$\tilde{F}(y) = \left\{ \left(F_\alpha(y), \mu\left(F_\alpha(y)\right)\right) \;\middle|\; \begin{array}{l} F_\alpha(y) = \left[F_{\alpha l}(y), F_{\alpha r}(y)\right] \\ \mu\left(F_\alpha(y)\right) = \alpha \forall \alpha \in (0,1] \end{array} \right\} . \quad (6)$$

As depicted in Figure 1, in a reliability analysis, the fuzzy set $\tilde{X}$ of parameter values is mapped to a fuzzy set of the failure probability,

$$\tilde{X} \rightarrow \tilde{P}_f \; . \quad (7)$$

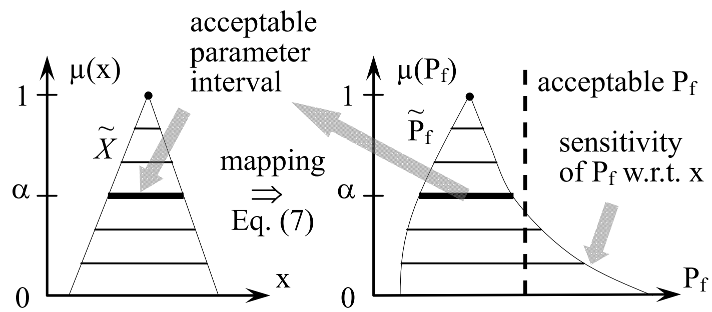

Figure 1. Relationship between fuzzy parameters and failure probability.

The membership function serves only instrumentally to summarize various plausible interval models in one embracing scheme. The interpretation of the membership value μ as epistemic possibility, which is sometimes proposed may be useful for ranking purposes, but not for making critical decisions. The importance of fuzzy modeling lies in the simultaneous consideration of various magnitudes of imprecision at once in the same analysis.

The features of a fuzzy probabilistic analysis can be utilized to identify sensitivities of the failure probability with respect to the imprecision in the probabilistic model specification; see Figure 1. Sensitivities of P_f are indicated when the interval size of $P_{f\alpha}$ grows strongly with a moderate increase of the interval size of X_α of the parameters. If this is the case, the membership function of $\tilde{P}_f$ shows outreaching or long and flat tails. An engineering consequence would be to pay particular attention to those model options X_α, which cause large intervals $P_{f\alpha}$ and to further investigate to verify the reasoning for these options and to possibly exclude these critical cases.

A fuzzy probabilistic analysis also provides interesting features for design purposes. The analysis can be performed with coarse specifications for design parameters and for probabilistic model parameters. From the results of this analysis, acceptable intervals for both design parameters and probabilistic model parameters can be determined directly without a repetition of the analysis; see Figure 1. Indications are provided in a quantitative manner to collect additional specific information or to apply certain design measures to reduce the input imprecision to an acceptable magnitude. This implies a limitation of imprecision only to those acceptable magnitudes and so also caters for an optimum economic effort. For example, a minimum sample size or a minimum measurement quality associated with the acceptable magnitude of imprecision can be directly identified. Further, revealed sensitivities may be taken as a trigger to change the design of the system under consideration to make it more robust. Beer and Liebscher (2008) describe a related method for designing robust structures in a pure fuzzy environment. These methods can also be used for the analysis of aged and damaged structures to generate a rough first picture of the structural integrity and to indicate further detailed investigations to an economically reasonable extent–expressed in form of an acceptable magnitude of input imprecision according to some α-level.

EXAMPLE

To illustrate this approach, we use an example reliability analysis for a reinforced concrete frame (Möller et al. 2003) shown in Figure 2. The structure is loaded by its dead weight, a small horizontal load P_H, and the vertical loads P_{V0} and p_0 which are increased with the factor ν until global structural failure is attained. For the purpose of demonstration, only the load factor ν is introduced as a random variable with an extreme value distribution of Ex-Max Type I with mean $\tilde{m}_\nu$ and standard deviation $\tilde{\sigma}_\nu$. Imprecision of the probabilistic model is modeled with triangular fuzzy numbers $\tilde{m}_\nu = \langle 5.7, 5.9, 6.0 \rangle$ and $\tilde{\sigma}_\nu = \langle 0.08, 0.11, 0.12 \rangle$. In addition, the rotational stiffness of the springs at the column bases is modeled as a triangular fuzzy number $\tilde{k}_\varphi = \langle 5, 9, 13 \rangle$ MNm/rad to take account of the only vaguely known soil properties. Based on this input information, the fuzzy reliability index $\tilde{\beta}$ shown in Figure 3 is calculated.

The result spreads over a large range of possible values for β. The shaded part of $\tilde{\beta}$ does not comply with the safety requirements. This means that a sufficient structural reliability is not ensured when the parameters are limited to the plausible ranges for $\alpha = 0$. In a traditional reliability analysis, using crisp assumptions for the parameters out of their plausible range such as the values associated with the membership $\mu = 1$, this critical situation is not revealed. So far, the results from p-box approach or from interval probabilities would lead to the same conclusions. As an additional feature of fuzzy probabilities, it can be observed that the left tail of the membership function of $\tilde{\beta}$ slightly tends to flatten towards small values. This indicates a slight sensitivity of β with respect to imprecision of the fuzzy input when this grows in magnitude. So one may wish to reduce the input imprecision to a magnitude which is associated with the steeper part of the membership function of β. In Figure 3, the part $\mu(\beta) \geq 0.4$ is a reasonable choice in this regard. Further, the result $\beta_{\alpha=0.4} = [3.935, 6.592]$ for $\mu(\beta) \geq 0.4 = \alpha$ (according to the definition of α-level sets) satisfies the safety requirement $\beta_{\alpha=0.4} \geq 3.8$. That is, a reduction of the imprecision of the fuzzy input parameters to the magnitude on α-level $\alpha = 0.4$ would lead to an acceptable reliability of the structure despite the remaining imprecision in the input. For example, a collection of additional information can be pursued to achieve the requirements $k_\varphi \in [6.6, 11.4]$ MNm/rad $= k_{\varphi,\alpha=0.4}$, $m_\nu \in [5.78, 5.96] = m_{\nu,\alpha=0.4}$, and $\sigma_\nu \in [0.092, 0.116] = \sigma_{\nu,\alpha=0.4}$. If this cannot be achieved for one or more parameters, the fuzzy analysis can be repeated with intervals for the parameters with non-reducible imprecision and with fuzzy sets for the parameters with reducible imprecision to separate the effects. The evaluation of the results then leads to a solution with proposed reduction of the imprecision only of those parameters for which this is possible. In this manner, it is also possible to explore sensitivities of the result β with respect to the imprecision of certain groups of input parameters or of individual input parameters. The repetition of the fuzzy analysis for these purposes can be avoided largely when a global optimization technique is used for the fuzzy analysis. This type of fuzzy analysis leads to a set of points distributed over the value ranges of the fuzzy input parameters and associated with results $\beta \in \tilde{\beta}$. For each construction of membership functions for the fuzzy input parameters, it is then immediately known which points belong to which α-level so that a discrete approximation of a result can be obtained directly without a repeated analysis. Repetition of the analysis is then only required for a detailed verification.

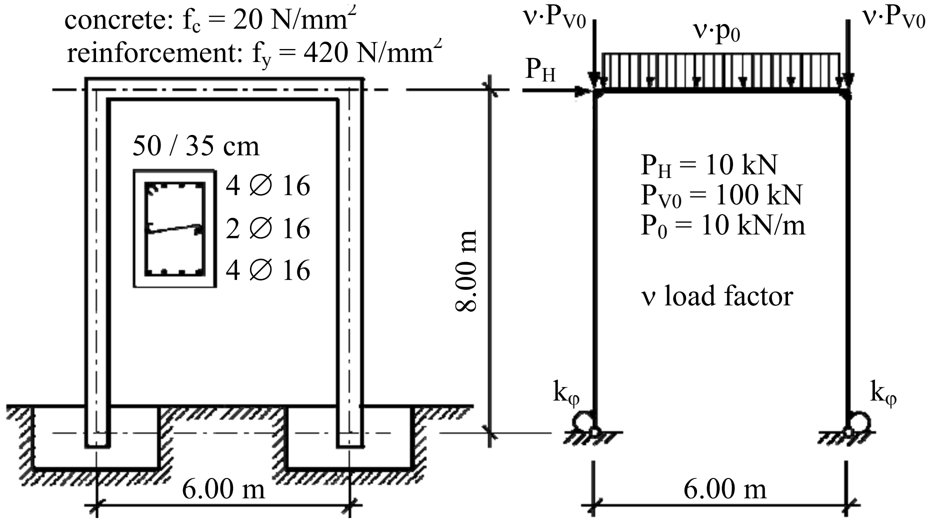

Figure 2. Reinforced concrete frame, structural model, and loading.

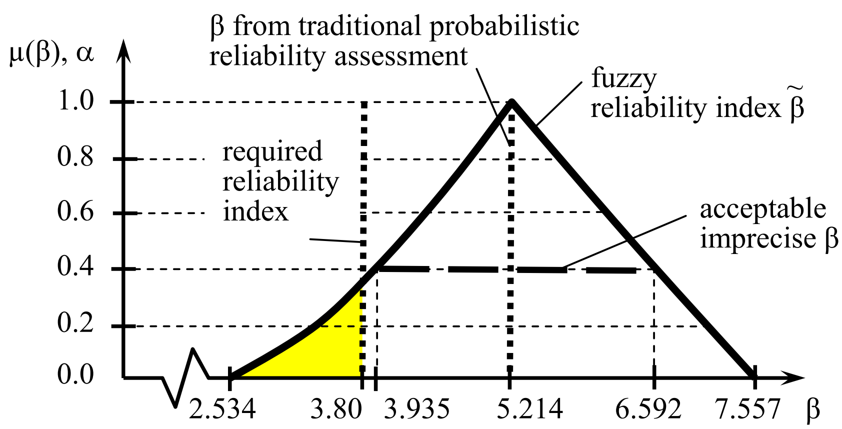

Figure 3. Fuzzy reliability index and evaluation against safety requirement.

REFERENCES

Augustin, Th. and Hable, R. (2010). On the impact of robust statistics on imprecise probability models: a review, *Structural Safety*, 32(6): 358–365.

Bandemer, H. and Näther, W. (1992). *Fuzzy Data Analysis*, Kluwer Academic Publishers, Dordrecht.

Beer, M. (2009). Fuzzy probability theory, in: R. Meyers (Ed.), *Encyclopedia of Complexity and Systems Science*, Vol 6, Springer, New York, 4047–4059.

Beer, M. and Liebscher, M. (2008). Designing robust structures – a nonlinear simulation based approach, *Computers and Structures*, 86(10), 1102–1122.

Der Kiureghian, A. and Ditlevsen, O. (2009). Aleatory or epistemic? Does it matter?, *Structural Safety*, 31(2), 105–112.

Ferson, S. and Tucker, W. T. (2006). Sensitivity analysis using probability bounding, *Reliability Engineering & System Safety*, 91(10–11), 1435–1442.

Ferson, S. and Hajagos, J. G. (2004). Arithmetic with uncertain numbers: rigorous and (often) best possible answers, *Reliability Engineering & System Safety,* 85(1–3), 135–152.

Gil, M. A., López-Díaz, M. and Ralescu, D. A. (2006). Overview on the development of fuzzy random variables, *Fuzzy Sets and Systems*, 157(19), 2546–2557.

Hall, J. W. (2006). Uncertainty-based sensitivity indices for imprecise probability distributions, *Reliability Engineering & System Safety*, 91(10–11), 1443–1451.

Igusa, T., Buonopane, S. G., and Ellingwood, B. R. (2002). Bayesian analysis of uncertainty for structural engineering applications, *Structural Safety,* 24(2–4), 165–186.

Klir, G.J. (2006). *Uncertainty and Information: Foundations of Generalized Information Theory*, Wiley-Interscience, Hoboken.

Kozine, I. and Filimonov, Y. (2000). Imprecise reliabilities: experiences and advances, *Reliability Engineering and System Safety*, 67, 75–83.

Kruse, R. and Meyer, K. (1987). *Statistics with Vague Data*, Reidel, Dordrecht.

Li, S., Ogura, Y. and Kreinovich, V. (2002). *Limit Theorems and Applications of Set Valued and Fuzzy Valued Random Variables*, Kluwer Academic Publishers, Dordrecht.

Möller, B., Graf, W. and Beer, M. (2003). Safety assessment of structures in view of fuzzy randomness, *Computers and Structures*, 81(15), 1567–1582.

Papadimitriou, C., Beck, J. L. and Katafygiotis, L. S. (2001). Updating robust reliability structural test data, *Probabilistic Engineering Mechanics*, 16, 103–113.

Schenk, C. A. and Schuëller, G. I. (2005). *Uncertainty Assessment of Large Finite Element Systems*, Springer, Berlin, Heidelberg.

Utkin, L. V. (2004). An uncertainty model of structural reliability with imprecise parameters of probability distributions, *ZAMM - Zeitschrift für Angewandte Mathematik und Mechanik,* 84(10–11), 688–699.

Viertl, R. (1996). *Statistical Methods for Non-Precise Data*, CRC Press, Boca Raton New York London Tokyo.

Viertl, R. (2011). *Statistical Methods for Fuzzy Data*, Wiley, Chichester.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*, Chapman & Hall, London.

Weichselberger, K. (2000). The theory of interval-probability as a unifying concept for uncertainty, *International Journal Approximate Reasoning*, 24(2–3), 149–170.

Yamauchi, Y. and Mukaidono, M. (1999). Interval and paired probabilities for treating uncertain events, *IEICE Transactions on Information and Systems,* E82_D(5), 955–961.

Zimmermann, H. J. (1992). *Fuzzy Set Theory and Its Applications*, Kluwer Academic Publishers, Boston London.