African Journal of Basic & Applied Sciences 8 (4): 193-197, 2016
ISSN 2079-2034

DOI: 10.5829/idosi.ajbas.2016.8.4.23311

An Extended Hybrid Approach to Model Variability and Uncertainty: Application in Risk Assessment

Palash Dutta

Department of Mathematics, Dibrugarh University, Dibrugarh,-786004, India

Abstract: In any risk assessment model, the input parameters are usually tainted with one or other type of uncertainty. When the uncertainty of random nature probability distribution is used for representing the input parameters but when the uncertainty arises due to lack of information possibility distribution is appropriate for its representation. In this paper, an effort has been made to propose a method for joint propagation of probabilistic and possibilistic in any risk model where parameters of probability distributions are available in the form of interval or fuzzy number.

Key words: Variability and Uncertainty · Hybrid Approach · Risk Assessment

INTRODUCTION

The aspect of uncertainty is an important and integral to any risk assessment process. For any decision making process involving risk the modelling and quantification of the uncertainties is required. The uncertainties are basically two types viz., aleatory and epistemic. When some parameters are affected by aleatory uncertainty and other by epistemic uncertainty, how far computation of the risk, then one can either transform all the uncertainties to one type or use some methods for propagation of the uncertainties. Many researchers have studied the issue of combining probabilistic and possibilistic representation of variability and uncertainty respectively within the same computation of risk. For example, the hybrid method proposed in [1] combines the random sampling of probability distribution functions (PDFs) with fuzzy interval analysis on the á-cuts. In order to compare random fuzzy set to a tolerance threshold authors performed a post-processing of this result. Authors [2] laid bare a shortcoming of this post-processing method where it was showed how the theory of evidence, also called theory of Dempster-Shafer (or theory of belief functions; Shafer, [3]) could provide a simple and rigorous answer to the problem of summarizing the results of the hybrid computation for comparison with a tolerance threshold. In the hybrid approach proposed [4] combined utilization of fuzzy and random variables produces membership functions of risk to individuals at different fractiles of risk as well as probability distributions of risk for various alpha-cut levels of the membership function. Different hybrid method for joint handling of probability and possibility distributions were presented in [5]. A hybrid method to deal with both variability and uncertainty within the same framework of computation of risk was proposed in [6]. Dutta ([7-10]) also studied probability-possibility distributions and their applications in risk assessments.

In some situations parameters of probability distributions (i.e., mean and standard deviation) are obtained in the form of interval or fuzzy numbers. In such situations probability bounds approach is often used which combines probability theory and interval arithmetic to produce probability-boxes (p-boxes). In particular, probability bounds analysis provides solution to the problems involving unknown dependencies between variables and uncertainties in the exact nature of distributions. Different authors proposed different methods to compute probability bounds. Chebyshev [11] described bounds on a distribution only when mean and standard deviation of the variable are known. Markov [12] found similar bounds for a positive variable when only its mean is known. Frechet [13] proposed how to compute bounds on probabilities of logical conjunctions and disjunctions without making independence assumptions. Yager [14] described the elementary procedures by which bounds on convolution can be computed an assumption of independence. At the same time, Authors in [15] solved

Corresponding Author: Palash Dutta, Department of Mathematics, Dibrugarh University, Dibrugarh,-786004, India.

a question posted by A. N. Kolmogorov about how to finds bounds on distributions of sum of random variables when no information about their interdependency was available. Extending the approach of [15] Williamson and Downs in [16] develop a semi analytical approach that computes rigorous bounds on the cumulative distribution functions of convolution without necessarily assuming independents between the operands. Ferson and Hajagos [17] gave method to compute probability bounds. Dutta and Ali [18] proposed a method to construct probability bounds when parameters of probability distribution viz., mean and standard deviation (variance) are available in the form of interval or fuzzy number.

In this paper, we propose an extended hybrid approach to deal with both variability and uncertainty where parameters of probability distributions are available in the form of interval or fuzzy number.

Possibility Theory: Possibility theory normally associated with some fuzziness, either in the background knowledge on which possibility is based or in the set for which possibility is asserted. This constitute a method of formalizing non-probabilistic uncertainties on events i.e., a mean of assessing to what extent the occurrence of an event is possible and to what extent we are certain of its occurrence, without knowing the evaluation of the possibility of its occurrence.

A possibility distribution [19] denoted by π, here is a mapping from the real line to the unit interval, unimodal and upper semicontinuous. A possibility distribution describe the more or less plausible values of some uncertain variable *X*. Possibility theory provides two evaluations of the likelihood of an event, for instance that the value of a real variable *X* should lie within a certain interval: possibility Π and the necessity N Possibility measure Π and necessity measure *N* are define

$$\left.\begin{aligned}\Pi(A) &= \sup_{x\in A} \pi(x)\\ \mathrm{N}(A) &= 1-\pi(A^c)\end{aligned}\right\}$$

where A^c is the complement of *A*.

Possibility theory can also be considered as special branch of evidence theory that deals only with bodies of evidence whose focal elements are nested. Events are called as nested if $A_1 \subset A_2 \subset A_3 \subset \ldots \subset A_n$. terms belief and plausibility measures in context of possibility theory are called as necessity measure and possibility measure respectively.

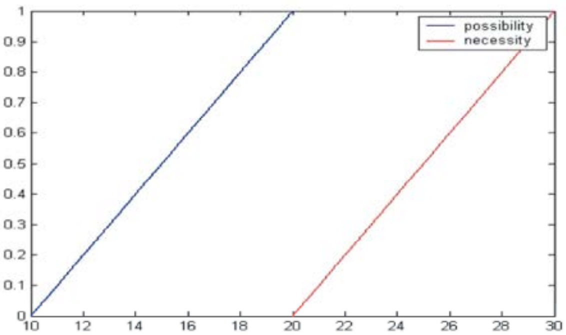

Fig. 1: Possibility and necessity measure of the fuzzy number A

If focal elements are nested then Π satisfies the following conditions:

$$\left.\begin{aligned}\Pi(A\cup B) &= \max(\Pi(A),\Pi(B)),\ \forall\ A,B\subseteq R\\ \Pi(A\cap B) &= \min(\Pi(A),\Pi(B)),\ \forall\ A,B\subseteq R\end{aligned}\right\}$$

For triangular (trapezoidal) fuzzy numbers, possibility and necessity measures are straight lines. For example, if a continuous possibility distribution is a triangular fuzzy number say, [a, b, c] then possibility measure is given by $\frac{x-a}{b-a}, a \le x \le b$ and necessity measure is given by $\frac{x-b}{c-b}, b \le x \le c$.

In particular, consider a fuzzy number A = [10, 20, 30]. Then the possibility measure of the fuzzy number A is $\frac{x-10}{10}, 10 \le x \le 20$ and necessity measure of the fuzzy number A is $\frac{x-20}{10}, 20 \le x \le 30$ and which are depicted below in Figure 1.

Sampling Technique for Possibility Theory: Here first uniformly distributed random numbers between 0 and 1 are generated. Random variables of given uncertainty are generated by equating these numbers to necessity function and possibility function. Two numbers are generated in this process, one corresponding to necessity function and the other corresponding to the possibility function. This process is repeated for all the uncertainty variables present in the model.

For a uniformly distributed random number u the uncertain variable x_n having necessity function $Nec(x_n)$ and uncertainty variable x_p having possibility function $Pos(x_p)$ are obtained as

$x_n = Nec^{-1}(u)$ and $x_p = Pos^{-1}(u)$.

For example, for the fuzzy number A= [10, 20, 30] the possibility measure and necessity measure are depicted in Figure 1. Now, for the uniformly generated random number say 0.6, the value of the random variable is 26 for the necessity measure and 16 for the possibility measure.

Proposal for Construction of P-box: Lower and upper probability can be constructed when parameters of probability distributions (Normal distribution, lognormal distribution, triangular distribution, uniform distribution) are not precisely known. Dutta and Ali [18] proposed a method to obtain lower and upper for symmetric probability distributions where parameters of probability distributions are available as closed intervals or fuzzy numbers. If parameters are fuzzy numbers then they considered the support of the fuzzy number using 0-cut as range of the variable.

Suppose *prodist* indicates one of the probability distribution i.e., normal distribution, lognormal distribution, triangular distribution, uniform distribution. Let A be a *prodist* whose parameters are available in the form of intervals, mean [*a, b*] and standard deviation (variance) [*c, d*]. The p-box for A has to be calculated by taking all the combination such as (*a, c*), (*a, d*), (*b, c*) and (*b, d*). Then the envelope over four distributions *prodist*(*a, c*), *prodist*(*a, d*), *prodist*(*b, c*) and *prodist*(*b, d*), gives the resulting p-box for *A*. To obtain lower and upper probability, uniformly distributed random numbers in between 0 and 0.5 and in between 0.5 and 1 are generated separately, as cumulative probability distributions of symmetric probability meet at mean and 0.5 (i.e., at (mean, 0.5).

Algorithm for P-box:
Step 1: Generate *N* number of uniformly distributed random numbers in between 0 and 0.5 and *N* numbers of uniformly distributed random numbers in between 0.5 and 1.

i.e., say $r = unifrnd(0,0.\ 5,1,N)$ and $s = unifrnd(0.5,1,1,N)$

Step 2: Take the inverse of the cumulative distributions function. That is,

$$x1 = prodistinv(r,a,d), x2 = prodistinv(s,a,c)$$
$$y1 = prodistinv(r,b,c), y2 = prodistinv(s,b,d)$$

Step 3: Consider $x = [x1, x2]$ and $y = [y1, y2]$

Step 4: Plotting of cumulative distribution function (*cdf*) for *x* and *y* will give the resulting p-box for *A*.

Proposed Extended Hybrid Approach: When models parameters are tainted with variability and uncertainty then hybrid approach comes into picture. Dutta and Ali [6] proposed a hybrid approach for combining probability and possibility distribution functions within the same framework of computation of risk. They used both Monte Carlo simulation and possibility theory in their method and also independency between the parameters has been assumed. Here, we extend their [6] hybrid method by incorporating that some parameters of probability distributions are available in the form of interval or fuzzy numbers and also we have consider vertex method to perform interval operations.

Consider a Model:

$$M = g\ (P_1, P_2 . . . P_{m,} Q_1, Q_2,...,Qr, F1, F_2 . . . , F_n)$$

which is a function of parameters where representations of some parameters are probabilistic and some parameters are possibilistic (Fuzzy number) and some parameters of probability distribution are available in interval or fuzzy numbers. Suppose $P_1, P_2 . . . P_m$ are *m* parameters presented by probabilistic distributions while $Q_1, Q_2,...,Q_r$ are *r* parameters presented by probabilistic distributions where mean and standard deviation are interval or fuzzy numbers and $F_1, F_2 . . . , F_n$ are *n* parameters presented by possibilistic distributions (Fuzzy numbers).

The approach is explained below:

- Generate *m* number of uniformly distributed random numbers from [0, 1] and perform Monte Carlo simulation to obtain random numbers by sampling probability distribution.
- Consider the *r* probability distributions where mean and standard deviation are available as interval or fuzzy numbers. Then, calculate lower and upper probability using the technique given is section 4. Generate *r* number of uniformly distributed random numbers from [0, 1] and perform Monte Carlo simulation to obtain random numbers by sampling probability distribution. Here we will get n numbers of close intervals i.e., 2*r* numbers of random numbers will be generated (*r* for lower probability and *r* for upper probability).
- Consider the possibility distribution *f*: X? [0, 1] (i.e., fuzzy numbers). Then, we use possibility measure and necessity measure defined as

$$pos(A) = \sup_{x \in A} f(x) \text{ and } Nec(A) = 1 - pos(A^c)$$

Table 1: Parameter values used in the risk assessment

Parameter	Units	Type of Variable	Value/distribution
Concentration (C)	mg/L	Probabilistic	Normal([0.14, 0.15, 0.16], [0.0005, 0.0006])
Intake rate(IR)	L/day	Probabilistic	Normal(5, 0.001)
Exposure frequency (EF)	Days/year	Constant	350
Exposure Duration (ED)	Years	Constant	30
Average Time (AT)	Days	Constant	25550
Body Weight (BW)	Kg	Fuzzy	[65, 70, 75]
Cancer slope factor (CSF)	$(mg/kg\text{-}day)^{-1}$	constant	0.15

to obtain upper and lower probability.

- Possibilistic Sampling: Generate *n* numbers of uniformly distributed random numbers from [0, 1] and perform Monte Carlo simulation to obtain random numbers by sampling possibility distribution. Here we will get n numbers of close intervals i.e., 2*n* numbers of random numbers will be generated (*n* for possibility measure and *n* for necessity measure).
- Assign all *m* random numbers, *r* closed intervals and *n* closed intervals in the model *M*. Then perform arithmetic operation between the close intervals using Vertex method. Output will be a single closed interval.
- Repeat step 1 to step 4 *N* times. So, we will have *N* numbers of close intervals.
- Consider M_1 and M_2, the collections of all initial and end points of the resulting intervals respectively.
- Cdf plotting of M_1 and M_2 will give the upper probability and lower probability respectively.

In the following section we consider a synthetic example to illustrate the use of our proposed method.

Case Study: To demonstrate and make use of the extended proposed hybrid method a hypothetical case study for cancer risk assessment is presented here. Suppose water became contaminated due to the release of radionuclide to the water. Need to calculate cancer risk for the ingestion pathway.

The risk assessment model due to the ingestion of radionuclides in water as provided by EPA, 2001 [22] is follows

$$\text{Risk} = \frac{C \times IR \times EF \times ED}{BW \times AT} \times CSF \qquad (1)$$

where *C* is concentration (mg/L), *IR* is the ingestion rate (L/day), *EF* is the exposure frequency (days/year), *ED* is exposure duration (years), *BW* is the body weight (kg), *AT* is averaging time (equal to 70 years x 365 days/year) and *CSF* is the cancer slope or potency factor associated with ingestion (mg/kg-day)-1.

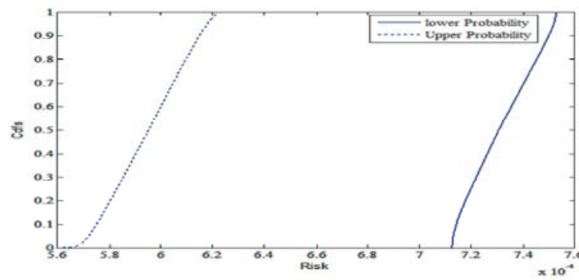

Fig. 2: Cancer Risk

Values of the parameters for the calculation of cancer risk are given in the Table 1.

The result of the cancer risk assessment due to the ingestion of radionuclides in water using equation (1) is depicted in Figure (2).

In this study, representation the parameter concentration (*C*) is considered as probabilistic distribution having with mean [0.14, 0.15, 0.16] and standard deviation [0.0005, 0.0008], intake rate (*IR*) is taken as probabilistic distribution with mean 5 and standard distribution 0.001while representation of the parameter body weight (*BW*) is considered as triangular fuzzy number and other parameters are considered as constant. Using our proposed method to deal with both type uncertainty natures in the risk assessment we have the result in the form of lower and upper probability. From these lower and upper probabilities, risk at different fractiles ([4], [20] & [21]) can be calculated and which are obtained in the form of close intervals. For instance, at 95th fractile cancer risk value lies in [6.174e-04, 7.519e-04]. Similarly, at 85th and 80th fractiles risk values lie in [6.121e-04, 7.476e-04] and [6.095e-04, 7.452e-04] respectively.

CONCLUSION

In risk assessment, it is most important to know the nature of all available information, data or model parameters. More often, it is seen that available information is interpreted in probabilistic sense because probability theory is a very strong and well established mathematical tool to deal with variability. However, not all available information, data or model parameters are affected by variability (i.e., nature of the data, information or parameters are random) and can be handled by traditional probability theory. Imprecision may occur due to scarce or incomplete information or data, measurement error or data obtain from expert judgment or subjective interpretation of available data or information. Thus, model parameters, data may be affected by epistemic uncertainty. Fuzzy set theory or possibility theory can be explored to handle this type of uncertainty. Sometimes, it is also seen that some model parameters are affected by uncertainty and some parameters are affected by variability then there is the need for joint propagation of uncertainties. Dutta and Ali [6] proposed a method for this. Sometimes the parameters of the probability distribution may be imprecise and they may be in some interval. In this paper we have proposed a method to deal with such situation. This method is an extension of [6].

REFERENCES

1. Guyonnet, D., B. Bourgine, D. Dubois, H. Fargier, B. Côme and J.P. Chilès, 2003. Hybrid approach for addressing uncertainty in risk assessments, Journal of Environmental Engineering, 126: 68-78.
2. Baudrit, C., D. Dubois, D. Guyonnet and H. Fargier, 2004. Joint Treatment of imprecision and Randomness in Uncertainty Propagation, Proc. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Perugia, pp: 873-880.
3. Shafer, G., 1976. A Mathematical Theory of Evidence, Princeton University Press.
4. Kentel, E. and M.M. Aral, 2004. Probalistic-fuzzy health risk modeling, Stoch Envir Res and Risk Ass 18: 324-338.
5. Li, J., G.H. Huang, G.M. Zeng I. Maqsood and Y.F. Huang, 2007. An integrated fuzzy-stochastic modeling approach for risk assessment of groundwater contamination, Journal of Environmental Management, 82: 173-188.
6. Dutta, P. and T. Ali, 2012. A Hybrid Method to deal with aleatory and epistemic uncertainty in risk assessment, International Journal of Computer Applications, 42: 37-44.
7. Dutta, P., 2015. Uncertainty Modeling in Risk Assessment based on Dempster-Shafer Theory of Evidence with Generalized Fuzzy Focal Elements, Fuzzy Information and Engineering, Elsevier (1616-8658) 7: 15-30.
8. Dutta, P., 2016. Dempster Shafer Structure-Fuzzy Number Based Uncertainty Modeling in Human Health Risk Assessment, International Journal of Fuzzy System Applications, 5: 96-117.
9. [#] Dutta, P., 2013. Combined Approach to Propagate Aleatory and Epistemic Uncertainty in Risk Assessment, International Journal of Mathematics and Computer Applications Research, 3: 2249-6955.
10. Dutta, P., 2013. An approach to deal with aleatory and epistemic uncertainty within the same framework: case study in risk assessment, International Journal of Computer Applications, 80(12): 40-45.
11. Chebyshev [Tchebichef] P., 1874. Sur les valeurs limites des integrales. J. Math Pure Appl., 19: 157-160.
12. Markov [Markoff], A., 1886. Sur une question de maximum et de minimum propose´e par M Tchebycheff. Acta Math, 9: 57-70.
13. Fre´chet, M., 1935. Ge´ne´ralisations du the´ore`me des probabilite´s totales. Fundam Math, 25: 379-387.
14. Yager, R.R., 1986. Arithmetic and other operations on Dempster-Shafer structures, International Journal of Man-machine Studies, 25: 357-366.
15. Frank, M.J., R.B. Nelsen and B. Schweizer, 1987. Bestpossible bounds for the distribution of a sum—a problem of Kolmogorov. Probab. Theory Related Fields, 74: 199-211.
16. Williamson, R.C. and T. Downs, 1990. Probabilistic Arithmetic I: numerical methods for calculating convolutions and dependency bounds, International journal of approximate reasoning, 4:89-158.
17. Ferson, S. and J.G. Hajagos, 2004. Arithmetic with uncertain numbers: rigorous and (often) best possible answers, Reliability Engineering and System Safety, 85: 135-152.
18. Dutta, P. and T. Ali, 2011. A proposal for construction of P-box, International Journal of Advanced Research in Computer Science, 2: 221-225.
19. Baudrit, C., D. Dubois and D. Guyonnet, 2006. Joint Propagation and Exploitation of Probabilistic and Possibilistic Information in Risk Assessment, IEEE Transaction on Fuzzy Systems, 14: 593-608.
20. Maxwell, R.M., S.D. Pelmulder, A.F.B. Tompson and W.E. Kastenberg, 1998. On the development of a new methodology for groundwater-driven health risk assessment, Water Resources Res., 34: 833-847.
21. Dutta, P. and T. Ali, 2012. Uncertainty Modeling in Risk Analysis: A Fuzzy Set Approach, International Journal of Computer Applications, 43: 35-39.
22. U.S. EPA. 2001. Risk Assessment Guidance for Superfund, Volume I: Human Health Evaluation Manual (Part E, Supplemental Guidance for Dermal Risk Assessment). Office of Emergency and Remedial Response, EPA/540/R/99/005, Interim, Review Draft. United States Environmental Protection Agency.