Applying Fuzzy Bayesian Maximum Entropy to Extrapolating Deterioration in Repairable Systems

Chi-Chang Chang
Chung Shan Medical University
Taiwan

Ruey-Shin Chen
National Kinmen Institute of Technology
Taiwan

Pei-Ran Sun
Chung Shan Medical University Hospital
Taiwan

1. Introduction

In general, most complex systems, which are constructed by collecting more than one part to perform either single or multiple functions, are usually repaired rather than replaced after failures, since such systems can be restored to fully implement the required functions by methods other than replacing the entire system (Ascher & Feingold, 1984). However, the successive times between failures are not necessarily identically distributed, as in renewal processes. More generally, they can become smaller (an indication of deterioration), or conversely larger and larger (an indication of reliability growth) (Barlow & Proschan, 1965). If deterioration is detected, then the decision of when to overhaul or discard the system, given the costs of repairs and failures, is of fundamental importance. At the time of the decision, the degree of future deterioration, which is likely to be uncertain, is of primary interest for the decision maker. Decision analysis seems to be able to provide methods to deal with such uncertain situation. However, the decision structure is usually formulated in such a way as to imposingly quantify particular qualitative characteristics on human being as decision rules (Freeling, 1984). The most important of these characteristics is that of human ability to precisely specify numerical values of ends and means in a decision process. These non-fuzzy decision rules are useful; however, they are limited in their applicability to real world situations where nearly all real human decision problems are imprecise, ill-definedness and vagueness (Dompere, 1982). In such situations, decision-making depends on numerous factors which limit human ability and increase difficulties to deal with (Asai & Okuda, 1975). Therefore, the use of fuzzy method may be very helpful in solving the decision making problems of deteriorating repairable systems (Bellman & Zadeh, 1970). In section 2, describes some related researches about modeling deteriorating
issue. In addition, the nonhomogeneous Poisson process (NHPP) was introduced. In section 3, delineates the implementation procedure for fuzzy Bayesian decision process. For the sake of information value analysis, we will discuss Bayesian decision process when the collected information is assumed to be fuzzy. Section 4, we will describes some fuzzy aggregation operations researches method and process. Further, section 5 a case study to illustrate the use of the models developed in the previous sections. Finally, section 6 discusses the work items and contributions of this study.

2. Deteriorating Repairable Systems

In order to model deterioration in repairable systems, the Non-Homogeneous Poisson Process (NHPP) was introduced since it seems more plausible for systems consisting of many components (Härtler, 1989). The system failure process is time-dependent and its intensity function of the failure model, which is given by

$$\lambda(x) = \lambda_0 \beta x^{\beta-1}, \lambda_0 > 0, \beta > 0$$

(1)

The likelihood function of the first \(N=n\) failure times for the case of time-truncated data is

$$L(x_1, x_2, \cdots, x_n \mid \lambda_0, \beta) = \lambda_0^n \beta^n (\prod_{i=1}^{n} x_i)^{\beta-1} \exp(-\lambda_0 x_n^\beta)$$

(2)

Huang and Bier (1998) proposed a natural conjugate prior distribution for the power law intensity function of the form (Ascher & Feingold, 1984):

$$f(\lambda_0, \beta) = K \lambda_0^{m-1} \beta^{m-1} \exp(-\alpha y_m^m) \beta^{1-1} \exp(-\lambda_0 cy_m^\beta)$$

(3)

Comparing with other approaches, this natural conjugate prior distribution has many desirable properties, which are summarized as follows:

1. The marginal distribution of \(\beta\) is a gamma distribution with parameters \(m\) and \(\alpha\), expectation \(\mu_\beta = m/\alpha\), and coefficient of variation \(CV_{\beta} = 1/\sqrt{m}\).
2. The conditional distribution of \(\lambda_0\) given \(\beta\) is a gamma distribution with parameters \(m\) and \(cy_m^\beta\).
3. The expectation of \(\lambda_0\) is given by \(\mu_{\lambda_0} = c m \left(\frac{\alpha}{\alpha + z_m}\right)^n\), where \(z_m = \ln(y_m)\).
4. The coefficient of variation of \(\lambda_0\) is given by \(CV_{\lambda_0} = \sqrt{\frac{n^m (m+1)}{m} - 1}\), where
The likelihood function of the first $N=n$ failure times for the case of time-truncated data is given by the function of the form (Ascher & Feingold, 1984):

$$
\ell(y_m; \eta, \alpha, \beta, \mu, \lambda, \nu) = \left(1 + \frac{z_m^2}{\alpha^2 + 2\alpha z_m}ight)^{-1},
$$

where \(z_m = \ln(y_m)\).

These properties provide guidance on how to choose the parameters $\alpha$, $m$, $y_m$ and $c$ to achieve joint distributions with the desired prior moments. For example, $m$ can be chosen to give the desired value of $CV_{\beta}$, $\alpha$ can be selected to give the desired value for $\mu_0$, $y_m$ can be selected to give the desired value for $CV_{\lambda_0}$ and $c$ can be selected to give the desired value for $\mu_{\alpha_0}$.

### 3. Fuzzy Bayesian Decision Process

Generally, the information is expressed by means of numerical values in a quantitative setting (Crow, 1974). However, in a qualitative setting, which is filled with vague or imprecise knowledge, the information cannot be estimated with an exact numerical value (Herrera & Herrera, 2000). In such cases, a more realistic approach may be used for linguistic assessments instead of numerical values, that is, to presume that the variables involved in the problem are assessed by means of linguistic terms (Hisdal, 1984; Tong & Bonissone, 1984). This approach is appropriate since it allows a representation of information in a more realistic and adequate form when precision is not achievable.

In real Bayesian decision problems there are two different types of vagueness: states and information (Fruhwirth-Schnatter, 1993). In this section we will discuss decision models for the cases in which only the states are fuzzified and both the states and additional information are fuzzified (Roubens, 1997; Stephen & Donnell, 1979). Finally, we will also present the decision flowchart and the decision analysis process.

As mentioned in the previous section, Huang and Bier (1998) developed the joint natural conjugate prior distributions for $\lambda_0$ and $\beta$. This proposed conjugate prior distribution provides guidance on how to choose the parameters $\alpha$, $m$, $y_m$ and $c$ by collecting the experts’ knowledge and observed data about the prior moments (i.e., $\mu_0$, $CV_{\beta}$, $\mu_{\alpha_0}$, $CV_{\lambda_0}$). In other words, in order to apply the joint natural conjugate prior distribution into the decision process, the experts need to specify a specific value for each of these four moments, respectively. However, this is not necessarily realistic for real world cases, since it is usually a tough task for the experts to specify a value to a prior moment with sufficient confidence. Instead, the experts often think of a prior moment as a fuzzy number, that is, a range of numbers and each number within the range has a different membership value. Furthermore, since the prior moments are not exactly provided by the experts, the parameters (i.e., $\alpha$, $m$, $c$, and $y_m$) are therefore formed as the functions of these fuzzy numbers. From the properties of the joint natural conjugate prior in the previous section, we can have (Huang & Chang, 2004)

$$
\alpha \equiv \alpha(\mu_0, CV_{\beta}) = \frac{1}{\mu_0 CV_{\beta}^2},
$$

$$
m \equiv m(CV_{\beta}) = \frac{1}{CV_{\beta}^2}.
$$
Based on the above discussion and Equations (4) to (7), we start the discrimination problem with fuzzy states space $F=\{\mu_\beta, CV_\beta, \mu_{\lambda_0}, CV_{\lambda_0}\}$ and exact observation space $X=\{x\}$ (Dompère, 1982). Consider a sequence of observations $x(n)=(x_1, x_2, \ldots, x_n)$, therefore the fuzzy prior distribution of the fuzzy state $F_j$ for the deteriorating repairable system is given by

$$f(F_j) = K\int_{\mu_\beta CV_\beta \mu_{\lambda_0} CV_{\lambda_0}} \int \lambda_0^{-m-1}\beta^{n-1}\{\exp(-\alpha)y_m^{-m}\}^{\beta-1}\exp(-\lambda_0 cy_m^{\beta})\lambda_{F_j}(\mu_\beta)\lambda_{F_j}(CV_\beta)\lambda_{F_j}(\mu_{\lambda_0})\lambda_{F_j}(CV_{\lambda_0})d\mu_\beta dCV_\beta d\mu_{\lambda_0} dCV_{\lambda_0}$$

where $\lambda_{F_j}(r) (r = \mu_\beta, CV_\beta, \mu_{\lambda_0},$ or $CV_{\lambda_0})$ denotes the membership functions for the four fuzzy numbers, respectively, and $f(F_j)$ is the fuzzy integral over the fuzzy hyper space of four dimensions. Similar argument can be applied to the fuzzy posterior distribution derived from the fuzzy state $F_j$ along with the additional exact data for the deterioring repairable system, and which is given by

$$f(F_j \mid x_1, x_2, \ldots, x_n) = K'\int_{\mu_\beta CV_\beta \mu_{\lambda_0} CV_{\lambda_0}} \int \lambda_0^{-m-n-1}\beta^{n-1}\{\exp(-\alpha)y_m^{-m}\}^{\beta-1}\exp(-\lambda_0 cy_m^{\beta} + x_{n}^{\beta})\lambda_{F_j}(\mu_\beta)\lambda_{F_j}(CV_\beta)\lambda_{F_j}(\mu_{\lambda_0})\lambda_{F_j}(CV_{\lambda_0})d\mu_\beta dCV_\beta d\mu_{\lambda_0} dCV_{\lambda_0}$$

where $(x_1, x_2, \ldots, x_n)$ means we observe the additional exact data until the $n$th failures. Furthermore, in the part of observe additional fuzzy data, the fuzzy likelihood function of observing the $n$th failure times is given by

$$Lik(x_1, x_2, \ldots, x_n \mid \lambda_0, \beta) = \int \cdots \int \lambda_0^{-n}\beta^{n-1}\{\exp(-\lambda_0^nx_n^{\beta})\lambda_1^1(x_1)\cdots\lambda_n^1(x_n)d\lambda_1 \cdots d\lambda_n$$
where \((x'_{1}, x'_{2}, \ldots, x'_{n})\) means we observe additional fuzzy data until the 2nd failures, and 
\(\chi_{i}(x'_{i}) (i=1,2,\ldots,n)\) denotes the membership functions for the 2 fuzzy failure times, 
respectively. If we observe fuzzy state \(F_{j}\) and additional fuzzy data \((x'_{1}, x'_{2}, \ldots, x'_{n})\), then the 
posterior distribution will satisfy

\[
f(F_{j} | x'_{1}, x'_{2}, \ldots, x'_{n}) = K'' \int_{\nu_{j}} \int_{C_{j}} \int_{F_{j}} \int_{R_{j}} \ldots \int_{x_{n}} \lambda_{0}^{m+n-1} \beta^{n-1} \{\exp(-\alpha) y_{n}^{m} \prod_{i=1}^{n} x_{i} \}^{\beta-1} \exp[-\lambda_{0} (cy_{n}^{\beta} + x_{n}^{\beta})] \]

Note that \(K, K'\) and \(K''\) in Equations (8), (9) and (11) are normalizing factors. Figure 1 shows 
a fuzzy Bayesian analysis in deteriorating repairable systems with experts’ prior knowledge. 
The same decision elements proposed by Huang (2001) of a Bayesian decision analysis for a 
deteriorating repairable system are as follows:

(a) Parameter space \(\Theta: \{(\lambda_{0}, \beta) | \lambda_{0} > 0\}\).

(b) Action space \(A: \{a_{1}, a_{2}\}\), where \(a_{1}\) is the status quo, and \(a_{2}\) is the risk reduction action.

(c) Loss function \(L: \) a real function defined on \(\Theta \times A\). If we decide to keep the system 
operating, then the loss we face is \(L(0, a_{1})\); if we decide to take the risk reduction 
action, then the loss we face is \(L(0, a_{2})\).

(d) Sample space \(X: \) The additional information available to be collected (e.g., 
successive failure times). The cost of collecting this additional data or information 
should also be reflected in the decision process.

We assume (i) that the status of system after a repair is essentially the same as it was 
immediately before failure occurred (as good as old); and (ii) that the repair times can be 
eglected. The following terminology will be used in the decision analysis process:

\(C_{F}: \) the cost of a failure if it occurs.

\(C_{R}: \) the cost of the proposed risk reduction action.

\(C_{I}: \) the cost of collecting additional information.

\(\rho: \) the reduction in failure rate that would result from the proposed risk reduction 
action \((0 < \rho < 1)\).

\(T: \) the time horizon under consideration.

\(t: \) the time at which the decision is being made.

\(\Psi: \) the expected number of failures during the time period \([t, T]\) under the status quo.
Suppose that the repairable system has a planned lifetime (i.e., time horizon) \( T \), and the decision of whether to maintain the status quo or perform some risk reduction action must be made at time \( t \). The decision variable we are dealing with is then the expected number of failures in \([t, T]\). Since the system failure times are assumed to be drawn from a non-homogeneous Poisson process with power law intensity function, the expected number of failures in \([t, T]\) under the status quo is given by

\[
\varphi = \varphi(T, t, \lambda_0, \beta) = \int_t^T \lambda(s) \, ds = \int_t^T \lambda_0 \beta s^{\beta - 1} \, ds = \lambda_0 (T^\beta - t^\beta)
\] (12)

Suppose that the risk reduction action will reduce the failure intensity by a fraction \( \rho \), where \( 0 < \rho < 1 \). Then the expected number of failures in \([t, T]\) if the risk reduction action is performed is given by

\[
\int_t^T \lambda(s)(1 - \rho) \, ds = (1 - \rho)\varphi
\] (13)

Fig. 1. Flowchart for Fuzzy Bayesian Decision Process
On the basis of the assumptions given above, we therefore have a two-action problem with a linear loss function, where the loss for taking action \( a_1 \) (i.e., continuing with the status quo) is \( C_F \Psi \) and the loss for taking action \( a_2 \) (i.e., undertaking the risk reduction action) is \( C_F (1-p) \Psi + C_R \). The expected loss for the status quo is simply \( C_F E\{\Psi\} \), and the expected loss for the risk reduction action is \( C_F (1-p) E\{\Psi\} + C_R \). Since the fuzzy prior and posterior density functions for \( \Psi \) are available by using defuzzified techniques and bivariate transformation for Equations (8), (9) and (11), respectively, the prior and posterior mean values of \( \Psi \) can be evaluated. Therefore, Fuzzy Bayesian decision analyses can be performed by comparing the prior and posterior mean values of \( \Psi \) with the cutoff value \( \Psi_0 = C_R / (C_F p) \). If the relevant mean is smaller than \( \Psi_0 \), then we should keep the system operating as in the status quo; if not, then we should perform the risk reduction action.

4. Fuzzy Aggregation Operation Methods

Generally, changing one’s beliefs when new information becomes available is a common mode of human reasoning. It is observed in the deliberate gathering of pertinent evidence during industrial troubleshooting, or medical diagnosis and so on. In another word, if one can make independence assumptions, many of the problems disappear, and in fact, this is often the method of choice even when it is obviously incorrect. There are different methodologies for dealing with this problem, e.g., maximal entropy and Dempster-Shafer Theory (Oberkampf et al., 2004). However, it still left some the Challenge Problems to solve, these questions were (Ferson et al., 2004; Fetz & Oberguggenberger, 2004): (1) How should epistemic uncertainty about a quantity be represented? (2) How can epistemic and aleatory uncertainty about a quantity be combined and propagated in calculations? (3) How should multiple estimates of uncertain quantities be aggregated before calculation? (4) How should the technical issue of repeated uncertain parameters be handled in practical calculations? (5) How might various approaches be adapted for use in practical calculations based on sampling strategies?

This section reviews the (1) to (3) of five technical issues addressed by the Challenge Problems that are commonly involved in computational problems involving epistemic uncertainty. In a sense, this is a problem of too much information because it means the analyst must decide how to combine this information before proceeding with the analysis. In point of problem, we will examine the fuzzy entropy aggregation operators in two ways: through the fuzziness of the prior moments \( \mu_\beta, CV_\beta, \mu_\alpha, CV_\alpha \) and through the fuzziness of failure data set. A fuzzy number is not a measurement. In other word, a fuzzy number is a subjective valuation assigned by one or more human operators. In addition, defuzzification methods have been widely studied for several years and were applied to fuzzy arithmetic (Kandel, 1986; Kim et al., 1998; Ma et al., 2002). The major idea behind these methods was to obtain a typical value from a given fuzzy set according to some specified characters, such as central gravity, median, etc. In other words, each defuzzification method provides a correspondence from the set of all fuzzy sets into the set of real numbers (Roychowdhury & Pedrycz, 2001). Therefore, in order to transfer the subjective valuation into real valuation, we have to use the fuzzy concept of entropy measure method (Chang, 2008). It should lend itself to probabilistic updating formulas by allowing heuristic estimation of the degree of
independence. After describing these formulae, one example illustrates how fuzzy entropy might be applied.

4.1 Uncertainty measure of Entropy

Entropy is a measure of the amount of uncertainty in the outcome of a random experiment, or equivalently, a measure of the information obtained when the outcome is observed. This concept has been defined in various ways (Shannon, 1948; Renyi, 1961; Kosko, 1986; Pal & Chakraborty, 1986) and generalized in different applied fields, such as communication theory, mathematics, statistical thermodynamics, and economics (Belahut, 1987; Cover & Thomas, 1992; Ching et al., 1995). Of these various definitions, Shannon contributed the broadest and the most fundamental definition of the entropy measure in information theory. In Shannon’s entropy, entropy can be considered as a measure of the uncertainty of a random variable \( X \). Let \( X \) be a discrete random variable with a finite alphabet set containing \( N \) symbols given by \( \{x_0, x_1, \ldots, x_{N-1}\} \). If an output \( x_j \) occurs with probability \( p(x_j) \), then the amount of information associated with the known occurrence of output \( x_j \) is defined as

\[
I(x_j) = -\log_2 p(x_j)
\]  
(14)

That is, for a discrete source, the information generated in selecting symbol \( x_j \) is \( -\log_2 p(x_j) \) bits. On average, the symbol \( x_j \) will be selected \( n \cdot p(x_j) \) times in a total of \( N \) selections, so the average amount of information obtained from \( n \) source outputs is

\[
-n \cdot p(x_0)\log_2 p(x_0) - n \cdot p(x_1)\log_2 p(x_1) - \cdots - n \cdot p(x_{N-1})\log_2 p(x_{N-1})
\]  
(15)

Dividing (15) by \( n \), we obtain the average amount of information per source output symbol. This is known as the average information, the uncertainty, or the entropy, and is defined: The entropy \( H(X) \) of a discrete random variable \( x \) is defined as (Shannon, 1948)

\[
H(X) = -\sum_{j=0}^{N-1} p(x_j)\log_2 p(x_j)
\]  
(16)

or \( H(X) = -\sum_{j=0}^{N-1} p_j \log_2 p_j \)  
(17)

where \( p_j \) denotes \( p(x_j) \).

Hence, entropy is a function of the distribution of \( X \). Further, this amount of information is estimated by the average weighted information provided by the expected probabilities of occurrence of the events as follows (Kosko, 1986):

\[
-\sum_{i=1}^{n} p_i \cdot \ln p_i
\]  
(18)

where \( p_i \) is the probability of occurrence of event \( i \) and \( n \) is the number of events.
Equation (18) is also called Entropy as its form suggests, one can realize that when maximizing entropy, not only the probabilities of the events will affect the quantity of information, but also the number of events will cause certain impacts. Since these events will provide information, they are the factors concerning the information of the system and thus, their probabilities of occurrence are the weights of importance of these factors.

Several important properties regarding this Entropy model (Kosko, 1990; Kosko, 1997) which will be quoted by as below:

1. The objective function is a continuous function of \( p_1, p_2, \ldots, p_n \). Therefore, small changes in \( p_1, p_2, \ldots, p_n \) will causes small changes in \( H_n \). This means that information provided by factor \( i \) will be changed when the probability of occurrence of factor \( i \) changes.

2. \( H_n \) is a symmetric function of its arguments. Therefore, the amounts of information will not be changed by the different orders of factors.

3. \( H_{n+1}(p_1, p_2, \ldots, p_n, 0) = H_n(p_1, p_2, \ldots, p_n) \). Thus, the amount of information is not changed if an impossible outcome is added to the probability scheme. That is, if a factor \( i \) with probability of occurrence equal 0, it will not give any contribution to the expected information and thus it can be deleted.

4. \( H_n \) will be reach the maximum when \( p_1 = p_2 = \ldots = p_n = 1/n \) and the maximum information by giving the outcomes equal probabilities of occurrence when the maximum uncertainty is faced.

5. The maximum value of \( H_n \) equals \( \ln(n) \). The maximum value of \( H_n \) increase as \( n \) increased. So, when we investigate more factors of a system, the expected information about the system will increase.

In what follows, we will propose the fuzzy entropy measure which is an extension of Shannon’s definition.

### 4.2 Fuzzy Entropy

In general, the membership function of a fuzzy set is determined by the users subjectively, which means that the membership function specified for the same concept by different persons may vary considerably. The shapes of the membership functions always present the knowledge grade of the elements in the fuzzy sets for the users (Zadeh, 1983). In other words, every membership function also presents the fuzziness of the corresponding fuzzy set in the idea of users. Therefore, it is necessary for us to have some measurements to measure the fuzziness of fuzzy sets. According to Szmidt and Kacprzyk (2000), fuzziness, a feature of imperfect information, results from the lack of crisp distinction between the elements belonging and not belonging to a set (i.e. the boundaries of the set under consideration are not sharply defined). A measure of fuzziness often used and cited in the literature is entropy first mentioned in 1965 by Zadeh (1965). The name entropy was chosen due to an intrinsic similarity of equations to the ones in the Shannon entropy. However, the two functions measure fundamentally different types of uncertainty. Basically, the Shannon entropy measures the average uncertainty in bits associated with the prediction of outcomes in a random experiment. Until now, there are several typical methods to be used to measure the fuzziness of fuzzy sets. In 1972, De Luca and Termini (1972) introduced some requirements which capture human intuitive comprehension of the degree of fuzziness. Kaufmann (1975) proposed that the fuzziness of a fuzzy set can be measured through the
distance between the fuzzy set and its nearest non-fuzzy set. Another way given by Yager (1979) suggested the measure of fuzziness can be expressed by the distances between the fuzzy set and its complement. De Luca and Termini (1972) utilized the conception of the entropy to indicate the fuzziness of a fuzzy set. Kosko (1997) investigated the fuzzy entropy in relation to a measure of subsethood.

In order to elicit expert’s knowledge, and advances in numerical methods and computation have made it possible to implement fuzzy Bayesian analysis in ways previously research (Chang & Cheng, 2007). The fuzzy mutual entropy and explores the information theoretic structure of fuzzy cubes was be applied (we will explore it more in detail at following context). The fuzzy mutual entropy of a fuzzy set $F$ acts as a type of distance measure between $F$ and its set complement $F^c$. The logistic map equates the sum of a real vector’s $n$ components with the mutual entropy of some fuzzy set $F$ and its complement $F^c$. This cube geometry motivates the ratio measure of fuzziness $E(F) = \alpha / \beta$ (Kosko, 1986), where $\alpha$ is the distance $\ell^1(F, F_{\text{near}})$ from $F$ to the nearest vertex $F_{\text{near}}$ and $\beta$ is the distance $\ell^1(F, F_{\text{far}})$ from $F$ to the farthest vertex $F_{\text{far}}$. The fuzzy entropy theorem reduces this ratio of distances to a ratio of counts in Equation (19) and Figure 2 shows the fuzzy entropy theorem in the unit square.

$$E(F) = \frac{c(F \cap F^c)}{c(F \cup F^c)} \quad (19)$$

Fig. 2. Geometry of fuzzy entropy theorem (Data Source: Kosko, 1986)

Fuzzy cubes map smooth onto extended real spaces of the same dimension and vice versa. The $2^n$ infinite limits of extended real space $[-\infty, \infty]^n$ map to the $2^n$ binary corners of the fuzzy cube $I^n$. The real origin 0 maps to the cube midpoint. Each real point $\mathbf{x}$ will mapping to a unique fuzzy set $F$ as Figure 3 shows.
Fuzzy mutual entropy equals the negative of the divergence of Shannon entropy. Uncertainty descriptions define points in the fuzzy cube parameter space. Versions of both extended Shannon entropy and fuzzy mutual entropy define vector fields on the fuzzy cube. As to the methods of defuzzification, there have been widely studied for decades and effectively utilized to the applications of fuzzy arithmetic (Kandel, 1986; Kim et al., 1998; Ma et al., 2002). The foremost idea behind these methods was to obtain a typical value from a given fuzzy set according to some specified characters, such as central gravity, mean, or median, etc. In other words, each defuzzification method provides a correspondence from the set of all fuzzy sets into the set of real numbers (Roychowdhury & Pedrycz, 2001). Therefore, in order to transfer the subjective valuation into real valuation, we have planning to apply the intuitionistic fuzzy sets and discuss the extension of Luca-Termini Axioms for the measurement of entropy-based defuzzification method. The following will be explored both contents more in detail:

First, a geometric interpretation of intuitionistic fuzzy sets and fuzzy sets is presented in Figure 4 which summarizes considerations presented in (Szmidt & Kacprzyk, 2000). Basically, an intuitionistic fuzzy set \( X \) is mapped into the triangle \( ABD \) in that each element of \( X \) corresponds to an element of \( ABD \), as an example, a point \( x' \in ABD \) corresponding to \( x \in X \) is marked. In Figure 4, this condition is fulfilled only on the segment \( AB \). Segment \( AB \) may be therefore viewed to represent a fuzzy set. The orthogonal projection of the triangle \( ABD \) gives the representation of an intuitionistic fuzzy set on the plane. (The orthogonal projection transfers \( x' \in ABD \) into \( x' \in ABC \).) The interior of the triangle \( ABC = ABD' \) is the area where \( \pi >0 \). Segment \( AB \) represents a fuzzy set described by two parameters: \( \mu \) and \( \nu \). The orthogonal projection of the segment \( AB \) on the axis \( \mu \) (the segment \( [0; 1] \) is only considered) gives the fuzzy set represented by one parameter \( \mu \) only. (The orthogonal projection transfers \( x' \in ABC \) into \( x'' \in CA \).) As it was shown in Szmidt and Kacprzyk (2000), distances between intuitionistic fuzzy sets should be calculated taking into account three parameters describing an intuitionistic fuzzy set.
Second, in order to transfer the subjective valuation into real valuation, our proposed entropy-based defuzzification method is based on the Luca–Termini Axioms and Lee et al. (2001), and developed defined as follows:

1. Let $X = \{r_1, r_2, \ldots, r_n\}$ be a universal set with elements $r_i$ distributed in a pattern space, where $i = 1, 2, 3, \ldots, n$.

2. Let $\tilde{A}$ be a fuzzy set defined on an interval of pattern space which contains $k$ elements ($k < n$). The mapped membership degree of the element $r_i$ with the fuzzy set $\tilde{A}$ is denoted by $\mu_{\tilde{A}}(r_i)$.

3. Let $C_1, C_2, \ldots, C_m$ represent $m$ classes into which the $n$ elements are divided.

4. Let $S_{C_j}(r_n)$ denote a set of elements of class $j$ on the universal set $X$. It is a subset of the universal set $X$.

5. The match degree $D_j$ with the fuzzy set $\tilde{A}$ for the elements of class $j$ in an interval, where $j = 1, 2, \ldots, m$, is defined as
The fuzzy entropy \( FE_{C_j}(\sim A) \) of the elements of class \( j \) in an interval is defined as

\[
FE_{C_j}(\sim A) = -D_j \log_2 D_j
\]  

(21)

(6) The fuzzy entropy \( FE_{C_j}(\sim A) \) of the elements of class \( j \) in an interval is defined as

\[
FE_{C_j}(\sim A) = -D_j \log_2 D_j
\]  

In Equation (21), the fuzzy entropy \( FE_{C_j}(\sim A) \) is a non-probabilistic entropy and the match degree \( D_j \) in fuzzy entropy is measured via the membership values of occurring elements.

First of all, we will start our investigations with fuzzy states and consider the prior moments \( \mu_{\hat{\beta}}, CV_{\hat{\beta}}, \mu_{\hat{\beta}} \), \( CV_{\hat{\beta}} \), where be recorded with the interval \([i_1,i_2]\) of successive failure data \( \xi_r \) (i.e., about 0.3 within the interval \([i_1,i_2]\) for prior moment, respectively), the pattern space shown in the top of the Figure 5 (Hoffman et al., 1996). Subsequently, let us consider the importance weights of the prior parameters are assessed in linguistic terms represented by fuzzy numbers, such as “L” (Low), “M” (Medium) “H” (High) where provided by experts’ linguistic form, and the membership functions of the three linguistic weight terms with triangular membership function are shown the bottom of the Figure 5. In addition, the value of a membership function can be viewed as the degree to which a pattern belongs to a specified pattern space. Fuzzy entropy of the observed interval \([i_1,i_2]\) has shows in the bottom of the Figure 5.

In order to show the advanced method, let us consider about the both parts of (a) and (b) with the probability formula in the Figure 5, the probability of “star” is:

\[
p(star) = \frac{4}{5} = 0.8
\]

the probability of “circle” is:

\[
1 - p(star) = \frac{1}{5} = 0.2
\]

In such situations, decision making depends on numerous factors which limit human ability and increase difficulties to deal with. Since, the result of Shannon’s entropy is:

\[
H(X) = -\sum_{j=0}^{N-1} p(x_j) \log_2 p(x_j)
\]

\[
= -p(\star) \log_2 p(\star) - p(\bigcirc) \log_2 p(\bigcirc)
\]

\[
= -0.8 \cdot \log_2 0.8 - 0.2 \cdot \log_2 0.2
\]

\[
= -0.25756814 \cdot 0.464431896
\]

\[
= 0.72200036
\]
As mentioned above, the match degree $D_j$ in fuzzy entropy is based on mapped membership values of elements. Assume we begin by assigning three triangular membership functions with overlapped regions in the pattern space of $[0, 1]$, as shown in Figure 5. The value of a membership function can be viewed as the degree to which a pattern belongs to a specified pattern space. The fuzzy entropy of the interval $[i_1, i_2]$ for the degree of membership functions as follows:

(1). From the corresponding membership function $L$, the total membership degree of “★” is $0.38 + 0.22 + 0 + 0 = 0.7$. Total membership degree of “○” is 0.0.

The match degree of “★” is $D_1 = \frac{0.7}{0.7 + 0} = 1.0$

The match degree of “○” is $D_2 = \frac{0}{0 + 0} = 0.0$

The fuzzy entropy of $FE_{C_1}(L)$

$FE_{C_1}(L) = -1.0 \times \log_2(1.0) = 0$

$FE_{C_1}(L) = -0.0 \times \log_2(0.0) = 0$

Hence, the fuzzy entropy of $FE_{C_j}(L)$ for the patterns of the interval $[i_1, i_2]$ in the feature dimension $x_i$ is $FE(L) = FE_{C_1}(L) + FE_{C_2}(L) = 0$. 

Fig. 5. Two cases of pattern distribution with corresponding membership functions (The symbol of ★ and ○ stand for the degree of membership functions; $C_1, C_2, C_3$ denote the center of three triangular fuzzy sets, respectively)
(2). From the corresponding membership function \( \tilde{M} \), the total membership degree of “\( \star \)’ is \( 0.6 + 0.88 + 0.85 + 0.57 = 2.3 \). Total membership degree of “\( \circ \)” is 0.56.

- The match degree of “\( \star \)” is \( D_1 = \frac{2.3}{2.3 + 0.56} = 0.80 \)
- The match degree of “\( \circ \)” is \( D_2 = \frac{0.56}{2.3 + 0.56} = 0.20 \)

The fuzzy entropy of \( FE_{C_j}(\tilde{M}) \)

\[
FE_{C_1}(\tilde{M}) = -0.80 \times \log_2(0.80) = 0.258
\]

\[
FE_{C_2}(\tilde{M}) = -0.20 \times \log_2(0.20) = 0.464
\]

Hence, the fuzzy entropy of \( FE_{C_j}(\tilde{M}) \) for the patterns of the interval \([i_1, i_2]\) in the feature dimension \( x_1 \) is \( FE(\tilde{M}) = FE_{C_1}(\tilde{M}) + FE_{C_2}(\tilde{M}) = 0.722 \).

(3). From the corresponding membership function \( \tilde{H} \), the total membership degree of “\( \star \)” is \( 0 + 0.1 + 0.4 = 0.5 \). Total membership degree of “\( \circ \)” is 0.4.

- The match degree of “\( \star \)” is \( D_1 = \frac{0.5}{0.5 + 0.4} = 0.56 \)
- The match degree of “\( \circ \)” is \( D_2 = \frac{0.4}{0.5 + 0.4} = 0.44 \)

The fuzzy entropy of \( FE_{C_j}(\tilde{H}) \)

\[
FE_{C_1}(\tilde{H}) = -0.56 \times \log_2(0.56) = 0.468
\]

\[
FE_{C_2}(\tilde{H}) = -0.44 \times \log_2(0.44) = 0.521
\]

Hence, the fuzzy entropy of \( FE_{C_j}(\tilde{H}) \) for the patterns of the interval \([i_1, i_2]\) in the feature dimension \( x_1 \) is \( FE(\tilde{H}) = FE_{C_1}(\tilde{H}) + FE_{C_2}(\tilde{H}) = 0.989 \)

Further, we can obtain the whole fuzzy entropy via summation of all corresponding fuzzy entropies as follows:

\[
FE = FE_{C_1}(\tilde{L}) + FE(\tilde{M}) + FE(\tilde{H}) = 0 + 0.722 + 0.989 = 1.711
\]

From the above illustration, the entropy-based defuzzification method will be able to discriminate the actual distribution of patterns better. By employing membership functions for measure match degrees, the value of entropy not only considers the number of patterns but also takes the actual distribution of patterns into account.
5. Application

In this section we will show the discrimination problems with both cases of fuzzy states and exact information and exact states and fuzzy information, which are studied from the viewpoint of fuzzy arithmetic measures.

5.1 Fuzzy States and Exact Information

Suppose that the states spaces of the four fuzzy moments and the linguistic importance weight of each value assigned by experts are assessed and shown in Table 1.

<table>
<thead>
<tr>
<th>The Prior Moments of Interval Observations</th>
<th>Linguistic Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_{\lambda_1} = \xi_{\mu_{\lambda_1}}[0.4,0.7]$</td>
<td>L[0.1,0.3,0.5]</td>
</tr>
<tr>
<td>$CV_{\lambda_1} = \xi_{CV_{\lambda_1}}[0.3,0.9]$</td>
<td>M[0.3,0.5,0.7]</td>
</tr>
<tr>
<td>$\mu_{\beta} = \xi_{\mu_{\beta}}[0.6,0.8]$</td>
<td>M[0.3,0.5,0.7]</td>
</tr>
<tr>
<td>$CV_{\beta} = \xi_{CV_{\beta}}[0.4,0.6]$</td>
<td>H[0.5,0.7,0.9]</td>
</tr>
</tbody>
</table>

Table 1. The fuzzy prior moments provided by experts

Furthermore, by applying the fuzzy entropy method, the fuzzy number can be defuzzified into the crisp value. For example the defuzzified value of $\xi_{CV_{\lambda_1}}[0.3,0.9]$ is attainable as shown in Figure 6.

![Fig. 6. The patterns of the interval [0.3, 0.9] in the feature dimension $\xi_{CV_{\lambda_1}}$](image)

(1). From the corresponding membership function $VL$, the total membership degree of “appropriate” is $0 + 0 + 0 = 0.0$

The total membership degree of “excellent” is $0 + 0 + 0 = 0.0$

The total membership degree of “very high” is $0 + 0 + 0 = 0.0$
5. Application

In this section we will show the discrimination problems with both cases of fuzzy states and exact information and exact states and fuzzy information, which are studied from the viewpoint of fuzzy arithmetic measures.

5.1 Fuzzy States and Exact Information

Suppose that the states spaces of the four fuzzy moments and the linguistic importance weight of each value assigned by experts are assessed and shown in Table 1.

<table>
<thead>
<tr>
<th>Prior Moments of Interval Observations</th>
<th>Linguistic Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>7.0, 4.0</td>
</tr>
<tr>
<td>0</td>
<td>L[0.1,0.3,0.5]</td>
</tr>
<tr>
<td>0</td>
<td>CV[9.0,3.0]</td>
</tr>
<tr>
<td>0</td>
<td>M[0.3,0.5,0.7]</td>
</tr>
<tr>
<td>0</td>
<td>β[8.0,6.0]</td>
</tr>
<tr>
<td>0</td>
<td>M[0.3,0.5,0.7]</td>
</tr>
<tr>
<td>0</td>
<td>CV[6.0,4.0]</td>
</tr>
<tr>
<td>0</td>
<td>H[0.5,0.7,0.9]</td>
</tr>
</tbody>
</table>

Table 1. The fuzzy prior moments provided by experts

Furthermore, by applying the fuzzy entropy method, the fuzzy number can be defuzzified into the crisp value. For example the defuzzified value of $[0.9, 3.0]$ is attainable as shown in Figure 6.

![Figure 6](image_url)

(1). From the corresponding membership function $\tilde{L}$,

The total membership degree of ““ is $D_1 = \frac{0}{0 + 0 + 0} = 0.0$

The match degree of “” is $D_2 = \frac{0}{0 + 0 + 0} = 0.0$

The match degree of “” is $D_3 = \frac{0}{0 + 0 + 0} = 0.0$

The fuzzy entropy of $FE_{C_1}(\tilde{L})$

$FE_{C_1}(\tilde{L}) = -0.0 \times \log_2(0.0) = 0$

$FE_{C_1}(\tilde{L}) = -0.0 \times \log_2(0.0) = 0$

$FE_{C_1}(\tilde{L}) = -0.0 \times \log_2(0.0) = 0$

Hence, the fuzzy entropy of $FE_{C_1}(\tilde{L})$ for the patterns of the interval [0.3, 0.9] in the feature dimension $\xi_{ev_{10}}$ is $FE(\tilde{L}) = FE_{C_1}(\tilde{L}) + FE_{C_1}(\tilde{L}) + FE_{C_1}(\tilde{L}) = 0 + 0 + 0 = 0.0$

(2). From the corresponding membership function $\tilde{L}$,

The total membership degree of ““ is $0 + 0 + 0 = 0.0$

Total membership degree of “” is $0.6 + 0 = 0.6$

Total membership degree of “” is $0.5 + 0 + 0 = 0.5$

The match degree of ““ is $D_1 = \frac{0}{0 + 0 + 0} = 0.0$

The match degree of “” is $D_2 = \frac{0.6}{0 + 0 + 0.5} = 0.546$

The match degree of “” is $D_3 = \frac{0}{0 + 0 + 0} = 0.454$

The fuzzy entropy of $FE_{C_1}(\tilde{L})$

$FE_{C_1}(\tilde{L}) = -0.0 \times \log_2(0.0) = 0$

$FE_{C_2}(\tilde{L}) = -0.546 \times \log_2(0.546) = 0.4767$

$FE_{C_3}(\tilde{L}) = -0.454 \times \log_2(0.454) = 0.5173$

Hence, the fuzzy entropy of $FE_{C_1}(\tilde{L})$ for the patterns of the interval [0.3, 0.9] in the feature dimension $\xi_{ev_{10}}$ is $FE(\tilde{L}) = FE_{C_1}(\tilde{L}) + FE_{C_1}(\tilde{L}) + FE_{C_1}(\tilde{L}) = 0 + 0.4767 + 0.5173 = 0.994$

(3). From the corresponding membership function $\tilde{L}$,

The total membership degree of ““ is $0.75 + 0 + 0 = 0.75$

Total membership degree of “” is $0.35 + 0.50 = 0.85$
Total membership degree of “□” is $0.55 + 0.4 + 0 = 0.95$

The match degree of “□” is $D_1 = \frac{0.75}{0.75 + 0.85 + 0.95} = 0.294$

The match degree of “○” is $D_2 = \frac{0.85}{0.75 + 0.85 + 0.95} = 0.333$

The match degree of “□” is $D_3 = \frac{0.5}{0.75 + 0.85 + 0.95} = 0.373$

The fuzzy entropy of $FE_{c_1}(\tilde{M})$

$FE_{c_1}(\tilde{M}) = -0.294 \times \log_2(0.294) = 0.5192$

$FE_{c_2}(\tilde{M}) = -0.333 \times \log_2(0.333) = 0.5283$

$FE_{c_3}(\tilde{M}) = -0.373 \times \log_2(0.373) = 0.5307$

Hence, the fuzzy entropy of $FE_{c_1}(\tilde{M})$ for the patterns of the interval [0.3, 0.9] in the feature dimension $\xi_{cv_{140}}$ is

$FE(\tilde{M}) = FE_{c_1}(\tilde{M}) + FE_{c_2}(\tilde{M}) + FE_{c_3}(\tilde{M}) = 0.5192 + 0.5283 + 0.5307 = 1.5782$

(4). From the corresponding membership function $\tilde{H}$,

The total membership degree of “□” is $0.15 + 0.5 + 0.2 = 0.85$

Total membership degree of “○” is $0.0 + 0.4 = 0.40$

Total membership degree of “□” is $0.0 + 0.8 + 0.6 = 1.4$

The match degree of “□” is $D_1 = \frac{0.85}{0.85 + 0.4 + 1.4} = 0.32$

The match degree of “○” is $D_2 = \frac{0.4}{0.85 + 0.4 + 1.4} = 0.15$

The match degree of “□” is $D_3 = \frac{0.14}{0.85 + 0.4 + 1.4} = 0.53$

The fuzzy entropy of $FE_{c_1}(\tilde{H})$

$FE_{c_1}(\tilde{H}) = -0.32 \times \log_2(0.32) = 0.5260$

$FE_{c_2}(\tilde{H}) = -0.15 \times \log_2(0.15) = 0.4105$

$FE_{c_3}(\tilde{H}) = -0.53 \times \log_2(0.53) = 0.4854$

Hence, the fuzzy entropy of $FE_{c_1}(\tilde{H})$ for the patterns of the interval [0.3, 0.9] in the feature dimension $\xi_{cv_{140}}$ is

$FE(\tilde{H}) = FE_{c_1}(\tilde{H}) + FE_{c_2}(\tilde{H}) + FE_{c_3}(\tilde{H}) = 0.5260 + 0.4105 + 0.4854 = 1.4219$

(5). From the corresponding membership function $\tilde{H}$,
The total membership degree of “○” is 0.0 + 0.5 + 0.8 = 1.3
Total membership degree of “○” is 0.0 + 0.0 = 0.0
Total membership degree of “○” is 0.0 + 0.0 + 0.45 = 0.45

The match degree of “○” is \[ D_1 = \frac{1.3}{1.3 + 0.0 + 0.45} = 0.74 \]
The match degree of “○” is \[ D_2 = \frac{0.0}{1.3 + 0.0 + 0.45} = 0.0 \]
The match degree of “○” is \[ D_3 = \frac{0.45}{1.3 + 0.0 + 0.45} = 0.26 \]

The fuzzy entropy of \( FE_{C_1}(VH) \)

\[
FE_{C_1}(VH) = -0.74 \times \log_2(0.74) = 0.3214
\]
\[
FE_{C_2}(VH) = -0.0 \times \log_2(0.0) = 0.0
\]
\[
FE_{C_3}(VH) = -0.26 \times \log_2(0.26) = 0.5053
\]

Hence, the fuzzy entropy of \( FE_{C_i}(VH) \) for the patterns of the interval [0.3, 0.9] in the feature dimension \( \xi_{cv_{q0}} \) is

\[
FE(VH) = FE_{C_1}(VH) + FE_{C_2}(VH) + FE_{C_3}(VH) = 0.3214 + 0.0 + 0.5053 = 0.8267
\]

Finally, we can obtain the whole fuzzy entropy via summation of all corresponding fuzzy entropies as follows:

\[
FE = FE(VL) + FE(L) + FE(M) + FE(H) + FE(VH) = 0.0 + 0.994 + 1.5782 + 1.4219 + 0.8267 = 4.8208
\]

Finally, Equations (8) and (9) can be used to study the prior and posterior decision for the decision maker when dealing with the decision problem for deteriorating repairable systems.

5.2 Fuzzy States and Fuzzy Information

When the states and the addition information are both fuzzy, besides the work for the fuzzy prior as described in the previous case, we have to also deal with the problem of defuzzifying the failure data as shown in Table 2.

<table>
<thead>
<tr>
<th>Failure Data</th>
<th>Fuzzy interval ( Hour / 24)</th>
<th>Linguistic Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1989.07.06 (ADD₁)</td>
<td>[0.33,0.50]</td>
<td>L[0.1,0.3,0.5]</td>
</tr>
<tr>
<td>1989.10.23 (ADD₂)</td>
<td>[0.42,0.58]</td>
<td>M[0.3,0.5,0.7]</td>
</tr>
<tr>
<td>1990.01.12 (ADD₃)</td>
<td>[0.58,0.66]</td>
<td>M[0.3,0.5,0.7]</td>
</tr>
<tr>
<td>1990.09.08 (ADD₄)</td>
<td>[0.42,0.50]</td>
<td>H[0.5,0.7,0.9]</td>
</tr>
<tr>
<td>1990.11.14 (ADD₅)</td>
<td>[0.54,0.63]</td>
<td>H[0.5,0.7,0.9]</td>
</tr>
</tbody>
</table>

Table 2. The fuzziness of the added failure data
The fuzzy numbers of failure data can also be defuzzified into crisp values and form the likelihood function. Equations (8) and (11) can be used to study the prior and posterior decision for the decision maker when dealing with the decision problem for deteriorating repairable systems. The decision problems for both the previous two cases can be assessed; however, the computing problem in Huang (2001) for the posterior mean of the decision variable is also encountered. If the expected number of failures from the decision time until the system is discarded is used as the decision variable, the numerical integration is still needed for evaluating the values and therefore making the decision.

6. Conclusion

In this chapter, we have presented a method to solve the decision problem of deteriorating repairable systems and we also present an approach to illustrate the fuzzy entropy-based arithmetic approach for modeling experts’ epistemic uncertainty in deteriorating repairable systems. The decision process is useful in selecting the best alternative when the deteriorating repairable system associated with alternatives are known in terms of linguistic variables (Zadeh, 1975), in particular, when these linguistic variables can be modeled by fuzzy numbers. In real world situations, the deteriorating phenomena are usually expressed as some degrees of severity. In such case, the proposed decision process can provide more realistic solutions. In this chapter, we have assumed that the importance weights of different criteria are assessed in linguistic terms represented by triangular fuzzy numbers. However, there are still several limitations and further study may undergo by considering other kinds of fuzzy membership function, since it still leaves lots of space for extension.

7. References


Machine learning techniques have the potential of alleviating the complexity of knowledge acquisition. This book presents today’s state and development tendencies of machine learning. It is a multi-author book. Taking into account the large amount of knowledge about machine learning and practice presented in the book, it is divided into three major parts: Introduction, Machine Learning Theory and Applications. Part I focuses on the introduction to machine learning. The author also attempts to promote a new design of thinking machines and development philosophy. Considering the growing complexity and serious difficulties of information processing in machine learning, in Part II of the book, the theoretical foundations of machine learning are considered, and they mainly include self-organizing maps (SOMs), clustering, artificial neural networks, nonlinear control, fuzzy system and knowledge-based system (KBS). Part III contains selected applications of various machine learning approaches, from flight delays, network intrusion, immune system, ship design to CT and RNA target prediction. The book will be of interest to industrial engineers and scientists as well as academics who wish to pursue machine learning. The book is intended for both graduate and postgraduate students in fields such as computer science, cybernetics, system sciences, engineering, statistics, and social sciences, and as a reference for software professionals and practitioners.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:
