Two Models to Assess Fuzzy Risk of Natural Disaster in China

Huang Chongfu*

※Project Supported by China Natural Science Foundation

ABSTRACT

China is one of the few countries where natural disaster strike frequently and cause heavy damage. In this paper, we mathematically develop two models to assess fuzzy risk of natural disaster in China. One is to assess the risk based on database of historical disaster effects by using information diffusion method relevant in fuzzy information analysis. In another model, we give an overview over advanced method to calculate the risk of release, exposure and consequence assessment, where information distribution technique is used to calculate basic fuzzy relationships showing historical experience of natural disasters, and fuzzy approximate inference is employed to study loss risk based on these basic relationships. We also present an examples to show how to use the first model. Results show that the model is effective for natural disaster risk assessment.

Keywords: natural disaster, fuzzy risk, information distribution, information diffusion, fuzzy approximate inference.

I. Introduction

Over the past 40 more years, Chinese government has all along paid close attention to natural disaster reduction and obtained remarkable achievement. However, before probability methods were used commonly, in many cases, disaster reduction strategy is blind. For example, the buildings in Tangshan, an industry city in Hebei province of China, were not designed for resisting earthquake because there is no strong earthquake before Tangshan Great Earthquake in 1976. After that, in China seismic intensity zone is formulated by using probability approaches. For that, a great amount of information is necessary to assess natural disaster risk, otherwise, unreliable results must be obtained, and it would lead us to a wrong way to make mistakes again.

In fact, for many kinds of disaster as floods, earthquake, winds, droughts, etc., when the region studied is not large, the available data are insufficient to permit estimating reliably the frequencies of release of risk agents or other characteristics of concern. The data supporting natural disaster risk assessment is particularly poor. For example, earthquake events is rare for a city, adequate data for meaningful statistical inference simply may not exist. Another limitation of probabilistic methods[1-3], which are used to assess risk, is that the researchers have to face many physical hypotheses. For example, Poisson hypothesis is us-

*Department of Resources and Environmental Science, Beijing Normal University

ually used to study randomness of earthquake. Another example is that, many researchers calculate area S_I of earthquake damage in intensity I by linear formula as the following

$$log_{10}(S_I) = a + bM \quad (1.1)$$

where M is magnitude, a, b are constants related to seismic intensity and the district. In fact, earthquake is not a strict Poisson process, and in the earthquake region with complex tectonic system, linear assumption (1.1) is no longer good. In many cases, it is difficult to judge if the hypotheses are suitable. Statistical risk assessment is a idealized method.

Early on there is not fuzzy sets theory, there were economists[4, 5] using the concept of ambiguity to define a random variable as risky if its probability distribution were known, and uncertain if its distribution were unknown. Since the publication of the first paper on fuzzy sets, fuzzy risk analysis has become more and more attractive for risk management. The primary approaches[6, 7] were based on the premise that one can provide the input of natural language estimate as probability regarded as the fuzzy risk of subsystem and used to combine the fuzzy risk of the entire system. In the past few years, researchers studying fuzzy risk gradually focus on the concept of fuzzy risk itself. Fuzzy risk interval was introduced to analyse linguistic evidence[8]. Fuzzy risk criterion was used to select an optimal threshold for automatic target recognition[9]. Fuzzy risk index[10] presents wind speed and frequency of occurrence. Simultaneously, the subjective assessment[11] of risks and combining[12] the components of risk were studied. The basic problems of these models of fuzzy risk are that: (a) Nobody has mathematically given a strict definition of fuzzy risk; (b) Only linguistic evidence can be described in fuzzy methods for fuzzy risk as the probability of fuzzy events; (c) These fit simple systems rather than complicated ones as natural disaster systems.

The first of all, to solve these problems, we would all-around analyse the reason why in many cases conventional risk as the probability is unreliable and people tell more than they known.

Ⅱ. Natural Disaster Risk Assessment

Risk means many things to many people. The dictionary[13] defines risk as "exposure to the chance of injury or loss." In terms of insurance it defines risk as "the hazard or chance of loss."

Natural disaster risk can be defined as a probability distribution or similar quantification that describes uncertainty about the disaster magnitudes, timing, and so on.

A major challenge in risk assessment of natural disaster is the scientific approach to probability distribution estimation. Ignoring this problem is to assume implicitly that either no disaster will occur or that, if they do occur, they will be strictly determined in accordance with mathematical laws.

Because the level of risk of natural disaster depends on the specific nature and characteristics of the *risk source* might be floods, earthquake, winds, droughts, etc., the *exposure process* must exist by which people or the things they value may be exposed to the released risk agent of the risk source, and the *consequence process* must exist by which exposures produce adverse health or environmental consequences, a comprehensive natural disaster risk assessment must address each of these components comprehensively. Natural disaster assessment must determine, characterize, and quantify the following factors: (1) the potential of the source to release a risk agent; (2) the intensity, frequency, and duration of exposure, and nature of the populations and other valued entities that might be exposed; and (3) the relationship between exposure and the resulting health or environmental consequences. Finally, the combined influence of these factors of risk must be determined, characterized, and quantified. The final outputs of this process are estimates of the magnitudes of possible adverse heath or environmental consequences, including always a characterization of the probabilities, uncertainties, or degree of confidence associated with these estimates.

Generally, studying risk source is a physical analysing procedure, or one of historical data analysis process. Sometimes, the work involves both of them. It may be given by experts in relevant special departments. In many cases, exposure assessment is the most difficult task. The reason for this is that exposure assessment often depends on factors that are hard to estimate and for which there are few data. Critical information on the conditions of exposure is often lacking. Hundreds of exposure models have been developed for a diversity of risk source, risk agents, and routes of exposures. For example, models have been developed to represent the transmission of ground motion from the source of an earthquake to a given site, taking into account the magnitude of the earthquake, local soil conditions, and the distance from the epicenter to urban areas at risk. Although modeling methods provide a means for estimating exposures in the absence of comprehensive monitoring data, most uncertainties in modeling exposures are not caused by inherent deficiencies in modeling techniques. Instead, the uncertainties arise from lack of understanding and lack of data. The primary purpose of a consequence assessment model is to translate exposure to a specified risk agent into damage consequences. The principal type of consequence assessment model is the dose-response model. A dose-response model is a functional relationship between the dose (i.e., measure of exposure) and a adverse entity response (i.e., the measure of damage). Most dose-response models are derived from statistical data such as that from monitoring or testing. Examples are the linear dose-response models used to estimate human health effects and materials damage of buildings. Alternatively, dose-response models may be derived from theoretical considerations with little or no basis in empirical data. Dose-respones models have many limitations, including the availability of the data or the knowledge and understanding needed to set their parameters and verify their accuracy. For the overwhelming majority of risk agents, knowledge is insufficient to permit confidence in the selection of a dose-response function.

It is useful to recognize the existence of two different types of natural disaster risk:

1) Statistical risk – determined by currently available data, typically calculated by using the database of historical disaster effects.

2) Predicted risk – predicted analytically from system models structured from historical studies, typically calculated through the risk of release, exposure and consequence assessment.

Ⅲ. Fuzziness of Statistical Risk

The main difficulty to analyse statistical risk of natural disaster in real world is that the data to use for statistical analysis is small in amount.

We have a sample $\{x_1, x_2, \cdots, x_n\}$ of independent, identically distributed observations from a continuous distribution with probability density function $p(x)$, $x \in R$, which we are trying to estimate. Where n is called sample size. When n is small, the sample is called a small sample.

Before there were fuzzy theory, people used to regard the sample as random information employed to estimate $p(x)$ by classical statistical methods. In fact, when it is a small sample, it is not only provide random information but also fuzzy information.

The reason is that, if we want to know $p(x)$ by using a small sample, the recognition must be imprecise, and the explanation of the physical laws concerning $p(x)$ will be vague, unclear and ambiguous. If we increase the observations in the sample, the estimator will be more near $p(x)$.

Obviously, from small to large, the sample possesses the transition tendency. When a sample is small, the tendency makes each observation in the sample have a character which requests of an observation is that one would become several and more if we want estimation to become perfect. The character makes a given observation can be a representative of the others which will be near it but have not appeared. That means, the observation in a small sample can act spokesman for the implied observations around itself. Therefore, x_i can be regarded a representative of its "around". This implies x_i to provide information not only to the point where the observation value was obtained, but also to other points around that one. The bounds of the concept of "around" is unclear, fuzzy and soft. So, the information provided by x_i to help us to recognize $p(x)$ around x_i is fuzzy. In other words, x_i is a fuzzy representative that acts spokesman for a fuzzy group and provides fuzzy information.

As everyone knowns, in order to deal with small-sample problem, the models employed interval estimation and empirical Bayes methods were developed long ago.

When the size is small, the error between the estimator and $p(x)$ would be large. The aim of interval estimation is to select a region in the parameter space and specify the probability that the estimated values of a set of parameters will lie within the selected region.

In fact, interval estimation method is bounded in that the observations would be random variables normally distributed with mean λ and variance σ^2. It is well known that

$$t=\frac{\bar{x}-\lambda}{\sigma/\sqrt{n-1}} \tag{3.1}$$

is the Student's t-distribution, with $n-1$ degrees of freedom. Where $\bar{x}$ is estimator about λ.

Hence,

$$[\bar{x}-t_\alpha\frac{\sigma}{\sqrt{n-1}}, \bar{x}+t_\alpha\frac{\sigma}{\sqrt{n-1}}] \tag{3.2}$$

is a $1-\alpha$ confidence interval estimator for λ. $\forall\alpha\in[0,1]$, t_α can be obtained from the table of t-distribution.

In this fashion, a statistician is able to label his estimators with a measure of confidence.

However, in many cases, engineers are interested in looking for a more accurate estimator instead of a label to the estimator. Empirical Bayes and related techniques come into play a leading role when data are generated by repeated execution of the same type of random experiment. Empirical Bayes methods provide a way in which such historical data can be used in the assessment of the current results. A fundamental problem in the pure Bayes approach is the specification of the prior distribution. When we assess risk of natural disaster in the current situation, there is not availability of previous data for estimation of the prior distribution.

Beginning in the 1950's, the use of "fuzzy sets" (not by that name) for function approximation and for probability density estimation was worked out in great detail by Parzen, and others. The object is to produce a continuous function, given only finitely many data points. Unfortunately, they can't explain the reason why that it is more effective for probability density estimation after change a datum point to a kernel function[14, 15]. The limit is that they only can discuss the property of the estimation in the case where $n\to\infty$. On this condition, classical statistics approaches is effective too.

When we know that a small sample carries fuzzy information, it is easy to explain and develop the kernel technique by using information diffusion principle[16] which is a novel fuzzy method, in this paper, it will be employed for risk assessment based on natural disaster database.

Ⅳ. Fuzziness of Predicted risk

When we are very knowledgeable about the physical process of the natural disaster, we can use stochastic models for representing the random laws of the natural disaster. And predicted risk can be calculated. For example, in earthquake engineering, the frequency distribution of earthquake magnitudes, especially in the middle range ($0<M<7$), is reasonably well approximated by the exponential distribution[17]:

$$f(M)=\beta e^{-\beta M} \quad (4.1)$$

where $f(M)$ is the probability density function of M in a given volume of the earthquake's crust. Parameter β is a regional constant. Depending on the region, the focal depth, and the stress level, the parameter may fluctuate between 0.3 and 1.5.

Current situation is that researchers have to collect data in a large region for calculating β. And the predicted risk is no distinction in the district where the area may be as large as a province of China, no different location, eg. near fault-zones. The so-called "predicted risk" calculated by using stochastic models is only to tell us some ambiguous information for making decision. When we limit the area studied to a city or a county, the stochastic models no longer have any effect, we will face small-sample problem again.

If we want to predict the risk of loss, the fuzziness can't be avoided obviously. For example, there are many kinds of buildings, their mechanism to be destroyed by earthquakes is unknown[18], the damage risk would been studied by using the damage materials of historical earthquakes. The relationship between earthquake and damage is fuzzy.

V. Risk Assessment Based on Database

In this section, a model would be discussed for studying statistical risk in one year based on a database consists of historical disaster records.

Let l be a disaster record which might be damage or loss in one year. Firstly, we would change l into disaster index x that is a scale in which time factor and dimensions are removed. Then, the major task of risk assessment is to estimate the frequency distribution about the disaster index based on the observations. Suppose we have n records corresponding to n years, these are $l_1, l_2, \cdots, l_n$ and which can be changed into $x_1, x_2, \cdots, x_n$.

The simplest method to estimate the frequency of the index is the histogram. Given an *origin* x_0 and a *bin width* h, we define the *bins* of the histogram to be the intervals $[x_0+mh, x_0+(m+1)h]$ for positive and negative integers m. The intervals have been chosen closed on the left and open on the right for definiteness. The histogram is then defined by

$$\hat{p}(x)=\frac{1}{nh}\ (\text{no. of } x_i \text{ in same bin as } x) \quad (5.1)$$

Note that, to construct the histogram, we have to choose both an origin and a bin width; it is the choice of bin width which, primarily, controls the amount of smoothing inherent in procedure.

Obviously, the histogram is an useful tool merely for a sample which size is more large.

The most common method is parametric density estimation according to maximum likelihood. For a set of independent and identically distributed observations $x_1, x_2, \cdots, x_n$ depending on a parameter θ, the likelihood function is

$$L(x\mid\theta)=\Pi_{i=1}^{n}\ p(x_i\mid\theta) \quad (5.2)$$

where $p(\cdot\mid\theta)$ refers to the density of a single observation.

Fisher[19] introduced the concept of estimating θ from the observations by finding $\hat{\theta}$ such that for $\theta\in\Theta$ (the parameter space)

$$L(x\mid\hat{\theta})\geq L(x\mid\theta) \quad (5.3)$$

The likelihood function has no finite maximum over the class of all densities. It is not possible to use maximum likelihood directly for density estimation without placing restrictions on the class of densities over which the likelihood is to be maximized. Generally, we have to suppose that the distribution model is known. In other words, the parameter cannot be estimated by using maximum likelihood method unless one has specific knowledge of the probability distribution of the estimates as well as the parent population.

The kernel estimation[14] is a popular method in non-parameteric approach. The kernel estimator with kernel K is defined by

$$\hat{p}(x)=\frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{x-x_i}{h}\right) \tag{5.4}$$

Where h is the window width, also called the smoothing parameter or bandwidth. Where the kernel function K which satisfies the condition

$$\int_{-\infty}^{\infty}K(x)\,dx=1.$$

Usually, K will be a symmetric probability density function.

However, the problem of choosing K and h still is of crucial importance in density estimation. In actual fact, h is invalid when the population density $p(x)$ is unknown.

In the fuzzy point of view for a small sample, the risk assessment based on database can be done by using information diffusion method[16] relevant in fuzzy information analysis.

Definition 5.1 Let $x_i(i=1, \ldots, n)$ be observations drawn from a population with density $p(x)$, $x\in R$. Suppose $\mu(y)$ is a Borel measurable function in $(-\infty, \infty)$, $\Delta_n>0$ is a constant,

$$\tilde{f}_n(x)=\frac{1}{n\Delta_n}\sum_{i=1}^{n}\mu\left(\frac{x-x_i}{\Delta_n}\right) \tag{5.5}$$

is called a information diffusion estimator about $p(x)$, and $\tilde{f}_n(x)$ is called diffusion estimator for short. μ is called diffusion function. Δ_n is called a diffusion coefficient.

With the help of the molecules diffusion theory, through researching the similarities of the information and molecules and solving the partial differential equation, we obtain normal diffusion function in (5.6) which is more practical.

$$\tilde{f}_n(x)=\frac{1}{nh\sqrt{2\pi}}\sum_{i=1}^{n}\exp\left[-\frac{(x-x_i)^2}{2h^2}\right] \tag{5.6}$$

where h is normal diffusion coefficient which can be obtained by using formula (5.7)

$$h=\begin{cases}1.6987(b-a)/(n-1), & \text{for } 1<n\le 5;\\ 1.4456(b-a)/(n-1), & \text{for } 6\le n\le 7;\\ 1.4230(b-a)/(n-1), & \text{for } 8\le n\le 9;\\ 1.4208(b-a)/(n-1), & \text{for } 10\le n.\end{cases} \tag{5.7}$$

where $a=\min\{x_1, x_2, \cdots, x_n\}$, $b=\max\{x_1, x_2, \cdots, x_n\}$.

Let $U=\{u_j \mid j=1, \cdots, m\}$ be the discrete universe of the disaster index. x_i can be diffused on U according to the revised formula of normal information diffusion as the following:

$$f_i(u_j)=\exp\left[-\frac{(x_i-u_j)^2}{2h^2}\right] \tag{5.8}$$

Let:

$$C_i=\sum_{j=1}^{m}f_i(u_j) \quad\text{and}\quad \mu_i'(u_j)=\frac{f_i(u_j)}{C_i}$$

$\mu_i'(u_j)$ is called a normal information distribution. Let:

$$q(u_j)=\sum_{i=1}^{n}\mu_i'(u_j) \quad\text{and}\quad Q=\sum_{j=1}^{m}q(u_j)$$

Obviously,

$$p(u_j)=\frac{q(u_j)}{Q} \tag{5.9}$$

is the frequency of disaster by index u_j in one year.

The results[16,20] of simulations and mathematical discussion shows that information distribution estimation is more better than other methods, especially when n is small. The benefit is provided by fuzzifying observations as

$$\underset{\sim}{x_i}\triangleq\mu_{x_i}(u_j)=\frac{\exp\left[-\frac{(x_i-u_j)^2}{2h^2}\right]}{\Sigma_{j=1}^{m}\exp\left[-\frac{(x_i-u_j)^2}{2h^2}\right]}, \quad i=1, 2, \cdots, n \tag{5.10}$$

Ⅵ. Fuzzy Risk Model of Natural Disaster

In this section, in the concept of fuzzy risk, we will discuss how to calculate the risk of release, exposure, and consequence assessment, where information distribution technique[21] is used to calculate basic fuzzy relationships showing historical experience of natural disasters, and fuzzy approximate inference is employed to study loss risk based on these basic relationships.

Definition 6.1 Let $Y=\{y\}$ be the universe of discourse on natural disaster, and $\pi(y, p)$ be the possibility distribution of that disaster magnitude exceeds y in probability p. $\Pi=\{\pi(y, p)\,|\,y\in Y,\ p\in[0, 1]\}$ is called fuzzy risk.

For convenience sake, $\pi(y, p)$ is also called fuzzy risk, and $p(y)$ means $p(\xi\geq y)$.

Suppose probability risk $p(y)$ is known, it can be turned to fuzzy risk $\pi(y, p)$. In fact,

$$\pi(y, p)=\begin{cases} 1, & \text{when } p=p(y) \\ 0, & \text{others} \end{cases}$$

That is to say, probability risk is a special case of the fuzzy risk.

Let z be a risk agent which may be released by one or several risk sources. For example, constructive earthquake is the agent of active faults, and flood waters may be the agent of rainstorms.

Let m be measure of agent z. For example, when z is earthquake, m may be the Richter magnitude. If z is flood water, m may be the water level of a dam or a river.

Definition 6.3 Let $M=\{m\}$ be the universe of discourse on z, and $\pi_z(m, p)$ be the possibility distribution of that agent z exceeds m in probability p. We call

$$\Pi_z=\{\pi_z(m, p)\,|\,m\in M,\ p\in[0, 1]\}$$

fuzzy risk of agent z.

Generally, in engineering, $\pi_z(m, p)$ would be presented as a discrete form. Let the discrete universe of M be

$$M=\{m_j\,|\,j=1, \cdots, J\}$$

and the discrete universe of probability p be

$$P=\{p_k\,|\,k=1, 2, \cdots, K\}$$

Now, fuzzy risk of release assessment is $\pi_z(m_j, p_k)$. When we have a sample which consists of release records in history such as earthquake magnitude, the fuzzy risk can be obtained by using information distribution technique[21].

Suppose the records is $\{x_1, x_2, \cdots, x_n\}$ which means there were n times of occurrence, then, every x_i can be distributed with information gain q_{ij} into controlling point m_j in formula (6.1).

$$q_{ij}=\begin{cases} 1-\dfrac{|m_j-x_i|}{\Delta}, & \text{If } m_j\leq x_i\leq m_{j+1} \\ 0, & \text{otherwise} \end{cases} \tag{6.1}$$

where $\Delta=m_{j+1}-m_j$ called step length of discrete point.

After all observations have been treated with this simple process and information gains at each controlling point have been summed up, a distribution of information gains will turn out. That is $Q=\{Q_1, Q_2, \cdots, Q_J\}$, where $Q_j=\Sigma_{i=1}^{n} q_{ij}$.

In fact, Q_j means that there are Q_j disasters have occurred in the years studied and whose magnitude is about m_j. Namely, Q_j is the number of disaster with magnitude m_j. Using Q, the number of disaster with magnitude greater than or equal to m_j can be obtained as: $N_j=\Sigma_{i=j}^{J} Q_i$. They constitute a number distribution of exceeding magnitude as: $N=\{N_1, N_2, \cdots, N_J\}$.

Obviously, the frequency exceeding m_j is $f_j=\dfrac{N_j}{n}$.

We obtain an exceeding frequency distribution as

$$F=\{f_1(M\geq m_1), f_2(M\geq m_2), \cdots, f_J(M\geq m_J)\}$$

When n is small, it must be unreliable to regard f_j as the exceeding probability distribution, the reason is that knowledge provided by the observations is incomplete which carries fuzzy information. Using two dimensions information distribution method, F can be optimally changed to fuzzy risk.

We use the simplest formula (6.2) to distribute (m_j, f_j) to discrete point (m_i, p_k), mamely,

$$\tilde{f}_j(m_i, p_k)=\begin{cases}(1-\frac{|m_i-m_j|}{\Delta_1})(1-\frac{|p_k-f_j|}{\Delta_2}) \\ \quad \text{when } |m_i-m_j|\le\Delta_1 \text{ and } |p_k-f_j|\le\Delta_2 \\ 0, \quad \text{others}\end{cases} \tag{6.2}$$

where $\Delta_1=|m_{i+1}-m_i|$ and $\Delta_2=|p_{k+1}-p_k|$. Let

$$\tilde{f}(m_i, p_k)=\sum_{j=1}^{J}\tilde{f}_j(m_i, p_k) \text{ and}$$

$$g_i=\max\{\tilde{f}(m_i, p_k)|k=1, 2, \cdots, K\}$$

If $g_i=0$, let $g_i=1$. Then,

$$\pi_z(m_i, p_k)=\frac{\tilde{f}(m_i, p_k)}{g_i} \tag{6.3}$$

is fuzzy risk of agent z.

Assume that the attenuation relationship among risk sources and sites is

$$w=f(m, d) \tag{6.4}$$

where w is the site intensity, m is the magnitude at the source, and d is the shortest distance of the site from the source. The relationship can be improved by using a fuzzy relationship of M, D and W:

$$R_1=R_{M,D,W}=\{r^{(1)}(m, d, w)\} \tag{6.5}$$

which can be obtained by the experts. Where M, D and W is the universe of discourse on m, d and w respectively.

The site fuzzy intensity $\underset{\sim}{W}$ can be got by using:

$$\mu_W(w)=\sup_{m\in M, d\in D}\{r^{(1)}(m, d, w)\wedge\mu_{M,D}(m, d)\} \tag{6.6}$$

where $\mu_{M,D}(m, d)$ is the membership function of fuzzy magnitude and distance.

Let d_1, d_2 be the nearest and farthest distance from the site to the source respectively, the fuzzy distance $\underset{\sim}{D}$ can be expressed simply by using a bell function:

$$\mu_D(d)=\exp[-\frac{(\frac{d_2+d_1}{2}-d)^2}{\frac{(d_2-d_1)^2}{6}}$$

$$=exp[-1.5(\frac{d_2+d_1-2d}{d_2-d_1})^2] \tag{6.7}$$

If

$$\mu_M(m)=\pi_z(m, p) \tag{6.8}$$

we have

$$\mu_{M,D}(m, d)=\pi_z(m, p)\wedge\mu_D(d) \tag{6.9}$$

therefore, we can get the fuzzy risk of the site intensity as the following:

$$\pi_W(w, x)=\sup_{m\in M, d\in D}\{r^{(1)}(m, d, w)\wedge\pi_z(m, p)\wedge\mu_D(d)\} \tag{6.10}$$

Assume that the functional relationship between the dose and an adverse entity response can be expressed as:

$$y=g(w) \tag{6.11}$$

where w is the site intensity, y is the measure of damage. The relationship can be improved by using a fuzzy relationship of W, and Y

$$R_2=R_{W,Y}=\{r^{(2)}(w, y)\} \tag{6.12}$$

which can be obtained by the experts. Where W, and

Y is the universe of discourse on w, and y respectively.

The entity fuzzy response $\underset{\sim}{Y}$ can be got by using:

$$\mu_Y(y) = \sup_{w \in W} \{ r^{(2)}(w, y) \wedge \mu_W(w) \} \quad (6.13)$$

where $\mu_W(w)$ is the membership function of site fuzzy intensity as in (6.6).

If $\underset{\sim}{W}$ is an estimator of fuzzy risk, namely,

$$\mu_W(w) = \mu_W(w, x) = \sup_{m \in M,\, d \in D} \{ r^{(1)}(m, d, w) \wedge \pi_z(m, p) \wedge \mu_D(d) \}$$

we can get the fuzzy risk of the entity response as the following:

$$\pi_Y(y, p) = \sup_{w \in W} \left\{ r^{(2)}(w, y) \wedge \sup_{m \in M,\, d \in D} \{ r^{(1)}(m, d, w) \wedge \pi_z(m, p) \wedge \mu_D(d) \} \right\} \quad (6.14)$$

Ⅶ. Case Calculation

In a county of China, recently years, flood disaster often occurs, and an insurance want to know the loss risk of grain in one year. Firstly, we would collect disaster records represents the loss and prediction output, then, disaster index can obtained which are in Table 7.1.

Table 7.1 Loss, prediction output, and disaster index

Year	1985	1986	1987	1988
Loss(ton)	308200	361500	4162700	5779500
Prediction output(ton)	848070	933190	4752082	6370895
Disaster index	0.363	0.387	0.876	0.907

The normal diffusion coefficient is

$$h = 1.6987(b-a)/(n-1) = 1.6987(0.907-0.363)/(4-1) = 0.308$$

Let discrete universe of the disaster index be

$$U = \{u_j \mid j = 1, \cdots, 11\} = \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$$

$\forall x_i \in \{0.363, 0.387, 0.876, 0.907\}$ can be diffused on U according to formula (5.8) and obtained as

$u_j=$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$f_1(u_j)$:	.499	.694	.869	.979	.993	.906	.745	.550	.366	.219	.118
$f_2(u_j)$:	.453	.647	.831	.961	.999	.935	.788	.597	.408	.250	.138
$f_3(u_j)$:	.018	.042	.090	.174	.303	.475	.669	.849	.970	.997	.922
$f_4(u_j)$:	.013	.032	.072	.143	.258	.417	.608	.798	.941	1	.956

Therefore, the normal information distribution $\mu_i'(u_j)$ is

$u_j=$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$\mu_1'(u_j)$:	.072	.100	.125	.141	.143	.131	.107	.079	.053	.032	.017
$\mu_2'(u_j)$:	.065	.092	.119	.137	.143	.133	.112	.085	.058	.036	.020
$\mu_3'(u_j)$:	.003	.008	.016	.032	.055	.086	.122	.154	.176	.181	.167
$\mu_4'(u_j)$:	.002	.006	.014	.027	.049	.080	.116	.152	.180	.191	.182

Hence, we know that the primary information structure of the records is

$u_j=$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$q(u_j)$:	.142	.206	.274	.337	.390	.430	.457	.471	.467	.439	.387

Using formula (5.9), we obtain the frequency of disaster by index u_j in one year as

$u_j=$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$p(u_j)$:	.036	.052	.068	.084	.097	.107	.114	.118	.117	.110	.097

Ⅷ. Conclusion and Discussion

Comparing the fuzzy estimation with ones calculated by histogram, maximum likelihood, and kernel, we can know that fuzzy approach is more better.

There is an asymptotical optimal formula[20] to choose the number k of bins of histogram, the formula is in (8.1).

$$k = 1.87(n-1)^{2/5} \quad (8.1)$$

where n is the size of the sample. In our case $n=4$, and $k\approx 3$. Therefore, the histogram can be structured on three intervals which are [0.363, 0.544), [0.544, 0.726), [0.726, 0.907]. Because

$$x_1=0.363 \text{ and } x_2=0.387\in[0.363, 0.544),$$
$$x_3=0.876 \text{ and } x_4=0.907\in[0.726, 0.907]$$

Hence, the frequency f of disaster on the intervals is

$$f_{[0.363,\ 0.544)}=0.5, f_{[0.544,\ 0.726)}=0, f_{[0.726,\ 0.907]}=0.5$$

Generally, we use an exponential distribution function $p(x)=\lambda e^{-\lambda x}$ to estimate disaster risk by maximum likelihood principle. With the material in our case, the estimator of probability density is $p_{expon}(x)=1.579e^{-1.579x}$.

With Gaussian kernel and window width $h=1.06\sigma n^{-1/5}$, for our case, by using kernel method, estimation probability density is

$$p_{kernel}(x)=1.544(e^{-11.56(0.363-x)^2}+e^{-11.56(0.387-x)^2}+e^{-11.56(0.876-x)^2}+e^{-11.56(0.907-x)^2})$$

We know that the disaster index always be in interval [0, 1], namely, probability of loss over 1 must be 0. But, $p_{expon}(1.1)=1.579e^{-1.579\times 1.1}=0.278>0$, and $p_{kernel}(1.1)=1.873>0$. Hnce, $p_{expon}(x)$ and $p_{kernel}(x)$ can't be used for risk assessment. Obviously, histogram estimation is too rough, and it is not good for our case.

In paper [16], we show that information diffusion estimation is more better than histogram estimation. Therefore, the estimator as in section 7 is more reliable than ones calculated by histogram, maximum likelihood, and kernel.

The problem which we would study in the future is how to choose more reliable diffusion functions and concerning coefficients.

Ⅸ. Acknowledgment

This research was supported by the National Natural Science Foundation of China. The author would like to thank Dr. Da Ruan for his kind help and Professor Yee Leung for his support in completing this paper. The author is indebted to anonymous referees for their valuable comments for the revision of this paper.

References

1. C.A. Cornell, "Engineering seismic risk analysis", Bull. Seism. Soc. Am. Vol.58, pp. 1583-1606, 1968.
2. B.M. Douglas and A. Ryall, "Seismic risk in linear source regions, with application to the San Andreas fault", Bull. Seism. Soc. Am. Vol.67, pp. 233-241, 1977.
3. J.P. Willian, A.A. Arthur, Natural Hazard Risk Assessment and Public Policy-Anticipating the Unexpected, Springer-verlag, New York, 1982.
4. F.H. Knight, Risk Uncertainty and Profit, Houghton Mifflin, Boston, 1921.
5. D. Ellsberg, "Risk, ambiguity, and the savage axioms", Quarterly Journal of Economics, Vol.75, pp. 643-669, 1961.
6. L.J. Hoffman, E.H. Michelmen, and D.P. Clements, "SECURATE-Security evaluation and analysis using fuzzy metrics," Proc. of the 1978 National Computer Conference, AFIPS Press, Montvale, New Jersey, Vol.47, pp. 531-540, 1978.
7. K.J. Schmucker, Fuzzy Sets, Natural Language Computations, and Risk Analysis, Computer Science Press, Rockvill, Maryland, 1984.
8. M. Delgado, J.L. Verdegay, and M.A. Vila, "A model for linguistic partial information in decision-making problems," International Journal of Intelligent Systems, Vol.9, pp. 395-378, 1994.
9. Ning Fang, Hsiu-Ping Wang, and Ming-Chieh Cheng, "A fuzzy target recognition system with homomorphic invariant feature extraction", International Journal of Systems Science, Vol.24, pp. 1955-1971, 1993.

10. A.V. Machias and G.D. Skikos, "Fuzzy risk index of wind sites", IEEE Transactions on Energy Conversion, Vol.7, pp. 638-643, 1992.
11. J.H.M. Tah, A. Thorpe, and R. McCaffer, "Contractor project risks contingency allocation using linguistic approximation", Computing Systems in Engineering, Vol.4, pp. 281-293, 1993.
12. M. Jablonowski, "Fuzzy risk analysis: using AI systems", AI Expert, Vol.9, No.12, pp. 34-37, 1994.
13. Anon, Webster's Encyclopedic Unabridged Dictionary of the English Language, Gramercy Books, New York, 1989.
14. E. Parzen, "On estimation of a probability density function and mode", Ann. Math. Statist., Vol.33, pp. 1065-1076, 1962.
15. B.W. Silverman, Density Estimation for Statistics and Data Analysis, Chapman and Hall, London, 1986.
16. Huang Chongfu, "Fuzziness of incompleteness and information diffusion principle", Proceedings of FUZZ-IEEE/IFES'95, Yokoham, Japan, pp. 1605-1612, 1995.
17. C. Lomnitz and E. Rosenblueth, Seismic risk and engineering decisions, Elsevier Scientific Publishing Company, Amsterdam, 1976.
18. Xiu Xiangwen and Huang Chongfu, "Fuzzy identification between dynamic response of structure and structural earthquake damage", Earthquake Engineering and Engineering Vibration, Vol.9, No.2, pp. 57-66, 1989.
19. R.A. Fisher, "On the mathematical foundations of theoretical statistics", Phil. Trans., A, Vol.222, pp. 308-367, 1921.
20. Huang Chongfu and Wand Jiading, Technology of Fuzzy Information Optimization Processing and Applications, Beijing University of Aeronautics and Astronautics Press, 1995.
21. Liu Zhenrong and Huang Chongfu, "Information distribution method relevant in fuzzy information analysis", Fuzzy Sets and Systems, Vol.36, pp. 67-76, 1990.

Huang Chongfu

Huang Chongfu was born on September 9, 1958, in Chine. He received the B.S. Sc. degree in mathematics in 1982 from Yunnan University, Kunmin, China, and M.S. degree in earthquake engineering in 1985 from Institute of Engineering Mechanics, Harbin, China, and Ph.D. degree in applied mathematics in 1993 for Beijing Normal University, Beijing, China, respectively.

He is currently an Associate Professor of Natural Disaster Systems at Beijing Normal University, China. He was a Postdoctoral Research Fellow in Management Systems Simulation at Beijing University of Aeronautics and Astronautics from 1993 to 1995.

His current research interests includes fuzzy information processing, neural networks, computer simulation, earthquake engineering and natural disaster. He have received several awards from Chinese Government for studying information distribution concept and information diffusion principle.