Principle of Maximum Entropy and Risk Analysis of Disaster Loss

Lihua FENG*, Xingcai ZHANG
Department of Geography, Zhejiang Normal University, Jinhua 321004, China
(*Corresponding author: fenglh@zjnu.cn)

ABSTRACT

Disasters occur everywhere in the most disordered way, indicating that disaster entropy has reached the maximum value. Under given constraint conditions, when disaster entropy is the maximum value, the disaster loss series should follow $P$-III distribution. The occurrence interval of disaster loss refers to the average time interval that disaster loss of certain degree happens in the future. We could, according to the field disaster data and using $P$-III distribution function, calculate the value of future disaster loss with certain recurrence interval. Explicit in concept and easy to use, such a method has significant meaning in practice.

Key words: Principle of Maximum Entropy, P-III Distribution Function, Recurrence Interval, Disaster Loss, Risk Analysis

1. INTRODUCTION

In 1854, Clausius referred to the ratio of heat absorbed by working substances to temperature in reversible process as “entropy”. In the following 100 years and more, the concept of entropy had been widely used in various science fields [1, 11, 19]. As a state function, entropy has profound meanings: in thermodynamics, it is the measurement of unavailable energy; in statistical physics, it is the measurement of the number of microscopic states of a system; in informationism, it is the measurement of the uncertainty of random events [12, 14].

Major disasters in the world include rainstorms, floods, droughts, hurricanes/typhoons, earthquakes, and debris. Many scholars have been exploring how to use the concept and theory of entropy to analyze disasters for a long period of time [2, 5]. For instances, the application of entropy removal in the theory of dissipative structures [4] improved the accuracy of forecasting storms; an analysis of China’s Huaihe River basin has predicted flood change times nicely by calculating Kolmogorov entropy and Lyapunov index [17]; a study of the concept of entropy in seismic exploration was done [15]; an evaluating model of debris disaster based on the theory of entropy was constructed [16]; and the concept of disaster entropy and its formula were proposed by adopting information entropy [13]. However, nothing has been applied to the theory of disaster entropy in risk analysis of disaster loss.

Disasters are everywhere in the development process of society and economy. One disaster, huge disaster especially, could pull the social and economic development of the disaster area back to two or three years, or even seven to eight years, ago. As a result, risk analysis of future disaster losses is playing an important role for the sustainable development of certain region’s society and economy. By far, some researches have been done in the field of disaster losses analysis [3, 9]. In this article, we will, according to the features of disaster losses and by use of principle of maximum entropy, conduct risk analysis of disaster losses.

2. PRINCIPLE OF MAXIMUM ENTROPY

In 1948, Shannon used entropy to describe quantitatively the uncertainty or information content of a random event:

$$H = -C \sum_{i=1}^{n} p_i \ln p_i \quad (1)$$

Where: $H$ is known as information entropy; $p_i$ is the occurrence probability of a random event; $C$ is a constant. From Eq. (1) we know that entropy $H$ is a function of $p_i$. Thus, under given experimental conditions, there exists a distribution, which makes $H$ the maximum value. This distribution has dominant probability and is the most common distribution, so it is called the “Most Probable Distribution”. To sum up, the principle of maximum entropy is to choose, under given restrictions, the distribution when entropy is the maximum value in all possible compatible distributions.

Based on the principle of maximum entropy and by use of Lagrange undetermined multiplicator method, we could get the distribution when entropy is maximum [6, 18]. Let random variable $x$ be $x_1, x_2, \cdots, x_n$, and corresponding probabilities are...
They could satisfy:

\[ \sum_{i=1}^{n} p_i = 1 \quad p_i \geq 0 \]  

(2)

And the mean value \( F_k \) of several known functions \( f_k(x_i) \) is given:

\[ F_k = \sum_{i=1}^{n} f_k(x_i)p_i \quad k = 1,2,\ldots,m(m < n) \]  

(3)

In order to find out the distribution when entropy is the maximum value under constraint condition Eq. (2) and (3), we introduce undetermined multipliers \( \alpha \) and \( \beta_k \) to form a new function: \( H = -\alpha - \beta_1 F_1 - \beta_2 F_2 - \cdots - \beta_m F_m \). From Eq. (1), (2) and (3), we could get:

\[ H = -\alpha - \sum_{k=1}^{m} \beta_k F_k \]

\[ = -\sum_{i=1}^{n} p_i \ln p_i - \alpha \sum_{i=1}^{n} p_i - \sum_{k=1}^{m} \beta_k \sum_{i=1}^{n} f_k(x_i)p_i \]

\[ = \sum_{i=1}^{n} p_i \ln \left( \frac{1}{p_i} \exp \left[ -\alpha - \sum_{k=1}^{m} \beta_k f_k(x_i) \right] \right) \]

Using the inequality \( \ln x \leq x - 1 \), the above equation is changed into:

\[ H \leq \sum_{i=1}^{n} p_i \left( \frac{1}{p_i} \exp \left[ -\alpha - \sum_{k=1}^{m} \beta_k f_k(x_i) \right] - 1 \right) - \alpha + \sum_{k=1}^{m} \beta_k F_k \]

If we want \( H \) to be the maximum value, the above formula must be an equation, then \( p_i \) should satisfy the following equation:

\[ p_i = \exp \left[ -\alpha - \sum_{k=1}^{m} \beta_k f_k(x_i) \right] \quad i = 1,2,\ldots,n \]  

(4)

Using (2), the Eq. (4) could be written as:

\[ \alpha = \ln \left( \sum_{i=1}^{n} \exp \left[ -\sum_{k=1}^{m} \beta_k f_k(x_i) \right] \right) \]

If let \( Z = e^{\alpha} \), then it could be changed into:

\[ Z = \sum_{i=1}^{n} \exp \left[ -\sum_{k=1}^{m} \beta_k f_k(x_i) \right] \]  

\( Z \) is called Partition Function). Thus Eq. (4) would become:

\[ p_i = \frac{\exp \left[ -\sum_{k=1}^{m} \beta_k f_k(x_i) \right]}{Z} \]

(5)

In order to get the value of \( \beta_k \), we substitute (5) into constraint Eq. (3) and get:

\[ F_k = \sum_{i=1}^{n} f_k(x_i) \exp \left[ -\sum_{k=1}^{m} \beta_k f_k(x_i) \right] / Z \]

(6)

In Eq. (6), both \( F_k \) and \( f_k(x_i) \) are known, while the real unknowns are \( m \) values of \( \beta = (\beta_1, \beta_2, \cdots, \beta_m) \). \( M \) equations could get \( m \beta \) values, thus we could have the value of \( p_i \) when entropy is the maximum value.

The above calculation formula obtained from discrete conditions could also be used in the calculation process in continuous conditions.

### 3. DISTRIBUTION OF DISASTER LOSS SERIES

OBTAINED BY THE PRINCIPLE OF MAXIMUM ENTROPY

The principle of entropy increase shows that, under isolated or adiathermal conditions, the spontaneous development of the system from non-equilibrium states to equilibrium states is a process of entropy increase, in which the equilibrium state is corresponding to the maximum entropy and the status of equilibrium state system is the most confused and disordered one.

As a random event, occurrence of disaster is uncertain and could be described by entropy. Disasters occur everywhere in the most disordered way, indicating that entropy has reached the maximum value, so we could use the principle of maximum entropy to determine the distribution of disaster losses series (that is on-the-spot disaster losses data) of a certain area in a certain period.

From disaster loss series we know that disasters with small losses have more opportunities to occur than the disasters with large losses, and all disaster losses are larger than \( a_0 \). Then constraint Eq. (2) and (3) would become:

\[ \int_{a_0}^{\infty} p(x) dx = 1 \]

(7)

\[ \int_{a_0}^{\infty} (x - a_0) p(x) dx = E(x - a_0) \]

(8)

\[ \int_{a_0}^{\infty} \ln(x - a_0) p(x) dx = E(\ln(x - a_0)) \]

(9)

Values on the right of Eq. (7)-(9) are given. If these constraint conditions are satisfied, then the distribution function when entropy is the maximum value is:

\[ p(x) = \exp(-\beta_0 - \beta_1(x - a_0) + \beta_2 \ln(x - a_0)) \]

(10)

From Eq. (7) and Eq. (10) we could get:

\[ \exp(\beta_0) = \int_{a_0}^{\infty} \exp(-\beta_1(x - a_0) + \beta_2 \ln(x - a_0)) dx \]
\[ p(x) = \frac{\beta_1^{\beta_2-1}}{\Gamma(\beta_2+1)} (x-a_0)^{\beta_2}\exp(-\beta_1(x-a_0)) \]  

(12)

It could be seen that the Eq. (12) is a P-III distribution function.

4. CALCULATION METHOD

If we express disaster loss series by \( S \), then when disaster entropy is the maximum value, \( S \) should follows P-III distribution (after being arranged):

\[ p(S) = \frac{\alpha^\beta}{\Gamma(\alpha)} (S-a_0)^{\alpha-1}\exp(-\beta(S-a_0)) \]

(13)

In which \( \alpha, \beta, a_0 \) are three undetermined and unknown parameters. Relations between them and the three statistic parameters of disaster loss series \( \bar{S}, C_v \) and \( C_s \) are:

\[ \alpha = 4/C_v ; \quad \beta = 2/\bar{S}C_vC_s ; \quad a_0 = \bar{S}(1-2C_v/C_s) \]

In which \( \bar{S} \) — mean value of disaster loss series; \( C_v \) — variation coefficient of disaster loss series; \( C_s \) — skewness coefficient of disaster loss series.

In risk analysis, we need to find out the value of disaster loss \( S_P \) which is corresponding to the specified probability \( P \),

\[ P = P(S \geq S_P) \]

\[ = \frac{\alpha^\beta}{\Gamma(\alpha)} \int_{S_P}^{\infty} (S-a_0)^{\alpha-1}\exp(-\beta(S-a_0))dS \]

(14)

In this equation, \( P \) is determined by decision-making department according to the risk level or the disaster-bearing capability. Once \( P \) is determined, \( S_P \) would be the estimated disaster loss with certain probability (or recurrence interval \( T \)).

Under this condition, \( T = 1/P \). For example, if we want to get disaster loss \( S_{1\%} \) when \( P = 1\% \), then recurrence interval \( T = 1/0.01 = 100 \) years, and then \( S_{1\%} \) should be the loss of maximum disaster over 100 years. Therefore, the “recurrence interval” refers to the average time interval that disaster loss of certain degree happens in the future.

In order to calculate \( S_P \) according to distribution function Eq. (14), we could, after derivation, let \( \Phi = (S_P - \bar{S})/\bar{S}C_v \); in which \( \Phi \) is referred to as deviation coefficient from average, which could be checked in a list [8]. Then

\[ S_P = (\Phi C_v + 1)\bar{S} \]  

(15)

So when \( \bar{S}, C_v, C_s \) and \( P \) are known, we could get the value of \( S_P \).

In this case, \( \bar{S}, C_v \) and \( C_s \) could be calculated according to

\[ \bar{S} = \frac{1}{n} \sum_{i=1}^{n} S_i \]; \quad C_v = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (S_i - \bar{S})^2} \]

while \( C_s \) could be estimated by \( C_s = mC_v \) (m is multiple).

Due to the relatively large error in calculation of \( \bar{S}, C_v \) and \( C_s \) with the above mentioned method, so we often use the diagram curve-fitting method to determine \( \bar{S}, C_v \) and \( C_s \), with the value of \( \bar{S}, C_v \) and \( C_s \) obtained from the previous calculation as the initial-value coefficients \( \bar{S}_0, C_{v0} \) and \( C_{s0} \) of the diagram curve-fitting process. When we use diagram curve-fitting method, firstly we arrange the disaster loss series by the order from large to small: \( S_1 \geq S_2 \geq \cdots \geq S_n \), then the probabilities of each items should be [10]:

\[ P_i = i/(n+1) \]  

(16)

Then we draw the disaster loss series and their corresponding \( P_i \) points on probability grid paper and get discrete dots of disaster loss. Then we calculate \( S_P \) according to \( \bar{S}_0, C_{v0} \) \( C_{s0} \) and Eq. (15), and draw \( S_P \) and \( P_i \) on the probability grid paper too, and get a continuous theoretical curve (Figure 1). If the theoretical curve does not match well with disaster loss dots, then change \( \bar{S}, C_v \) and \( C_s \) and fit the curve over again until the two match well. The \( \bar{S}, C_v \) and \( C_s \) are then the statistic coefficients we want to get. In order to avoid man-made errors when estimating curve-fitting by eye, we could determine the three statistical coefficients by using on computer the method of multi-coefficient searching and the curve-fitting principle of minimum sum of absolute value of deviation. Finally, according to the specified probability \( P \) and the best-fit theoretical curve, we could get the disaster loss \( S_P \) with certain recurrence interval.

5. APPLICATION CASES

Flood and draught are the most serious calamities in natural disasters, losses of which account for over 55% of the total disaster losses. Now we take flood and draught disaster areas of
Shandong Province and Zhejiang Province in China as examples to show the application of principle of maximum entropy in risk analysis of disaster loss. Information about flood disaster area of Shandong Province is obtained from reference [7], period of which is 1949-1994. Using $P$-III distribution function to calculate, we could get the curve-fitting of flood disaster area of Shandong Province (Figure 1), which shows that the curve matches well.

![Figure 1 Curve-fitting of flood disaster area of Shandong Province](image)

In order to reduce the influence of the change of farmland area, we adopt the data of draught disaster area of Zhejiang Province in recent period (since 1991) in calculation. Table 1 listed the draught disaster areas and their calculation results from 1991 to 2005, with a sequence from large to small; then we could get $S = 345.93$ (hm$^2$), $C_r = 1.0$ and $C_s = 0.9$ when the theoretic curve fits best with dots (figure omitted).

Finally, we could get, by calculation of Eq. (15), that the extreme future draught disaster area over 5 years in Zhejiang Province is 612(hm$^2$); over 10 years is 809(hm$^2$); over 20 years is 989(hm$^2$).

### 6. CONCLUSION

As a random event, disaster can occur in the most disordered way everywhere, indicating that disaster entropy has reached the maximum value, so we could use the principle of maximum entropy to determine the distribution of disaster losses series of certain region in a certain period of time.

The above analysis shows that the disaster loss determined by principle of maximum entropy is explicit in concept and easy to be implemented. Due to the reiteration of future disasters in certain area, the future disaster loss estimation by the historical disaster loss series of such area is based on substantial data. Meanwhile, both time and size of the future disaster are undetermined, and our current knowledge and forecast technologies are still to be improved. As a result, using the principle of maximum entropy to estimate, according to existing data and as above shown, the future disaster loss is feasible. With the above method, once risk level or anti-disaster ability is determined, we could get the disaster loss with certain recurrence interval.

Loss of future disaster is an important factor in the frame of regional plan and development strategy by governments at different level. Through risk analysis of disaster loss, we could understand the disaster loss of each area in an all-round way, which will not only provide important information to leaders at different levels in scientific management and macro decision-making, but also provided reliable bases for theoreticians to draw up policies, which is of great practical significance.

### ACKNOWLEDGMENTS

This work was supported by National Natural Science Foundation of China (No. 40771044) and Zhejiang Provincial Science and Technology Foundation of China (No. 2006C23066).
REFERENCES