

Chapter 10

Methods in Risk Analysis



Eiki Yamakawa and Toshihiro Kawaguchi

Abstract To protect our lives and properties from disasters and accidents, we need to identify sources of risks, acknowledge weaknesses in our societies and ourselves, and be readily prepared in both hardware and software. We frequently express the odds of events taking place with probabilistic numbers; thus, we must have at least the minimum knowledge in mathematical statistics. This chapter first introduces quantitative evaluations of probabilities of events to cause damages to us and the magnitudes of the damages. We will then learn how to analyze and estimate risks using the evaluations. The chapter closes with decision-making methods in finding the best measures that minimize the risks.

Keywords Confidence interval · Linear model · Mathematical programming · Random variables · Relative risk

10.1 Evaluation and Probabilities of Risks

When natural disasters, like flooding or earthquakes, or accidents caused by physical or chemical phenomena take place, their formation or outbreak greatly differs with environmental factors and chances. For evaluating uncertain events like disasters or accidents, we generally apply the function called measure that maps the set of results of interest to a nonnegative real number to express its “magnitude.” When we express the set for the entire results that can take place with Ω , and its subsets with E_1, E_2 , the function μ that satisfies the following conditions is called the fuzzy measure.

E. Yamakawa (✉) · T. Kawaguchi
Faculty of Societal Safety Sciences, Kansai University, Takatsuki, Osaka, Japan
e-mail: eiki@kansai-u.ac.jp

$$\mu(\emptyset) = 0 \quad (10.1)$$

$$\mu(\Omega) = 1 \quad (10.2)$$

$$E_1 \subseteq E_2 \Rightarrow \mu(E_1) \leq \mu(E_2). \quad (10.3)$$

The first two conditions show that when we set the size of the set of all possible results to 1, fuzzy measure is a function that expresses the ratio of the results of interest occupy. The third condition is called monotonicity, and it shows that the evaluation goes up when more results are included.

The characteristics of measures depend on how measures of unions of sets are defined. When parameter λ is a real number larger than -1 , a fuzzy measure that satisfies the next condition is called λ -fuzzy measure.

$$E_1 \cap E_2 = \emptyset \Rightarrow \mu(E_1 \cup E_2) = \mu(E_1) + \mu(E_2) + \lambda\mu(E_1)\mu(E_2). \quad (10.4)$$

Probabilities that we are most familiar with are nothing but λ -fuzzy measures with $\lambda = 0$. Probabilities have complete additivity expressed with the following equation:

$$E_1 \cap E_2 = \emptyset \Rightarrow \mu(E_1 \cup E_2) = \mu(E_1) + \mu(E_2). \quad (10.5)$$

Complete additivity is a basic property to derive many characteristics about probability and is important for easily analyzing and estimating risks; however, it often conflicts with human psychology that wants to avoid uncertainty. The λ -fuzzy measure is a measure that generalizes complete additivity required to probability. In fact, when the parameter λ takes a positive number, the measure shows the property called super-additivity with the expression:

$$E_1 \cap E_2 = \emptyset \Rightarrow \mu(E_1 \cup E_2) > \mu(E_1) + \mu(E_2), \quad (10.6)$$

and the property is sub-additive when the parameter λ takes a negative value expressed:

$$E_1 \cap E_2 = \emptyset \Rightarrow \mu(E_1 \cup E_2) < \mu(E_1) + \mu(E_2). \quad (10.7)$$

In classic probability theory, when a trial results in N equally probable cases, and the number of events of interest is r , the occurrence probability of the events is defined r/N . For example, rolling a cubic die and recording the result have six possible outcomes. With a proper die, all results have equal chances; thus, the possibility of getting 1 is $1/6$. The occurrence probability of an event equals the relative frequency of the event occurrence with an infinite number of independent trials. Natural phenomena, however, like flood or earthquake often break out with some events in combination, and their occurrence usually depends on past records; thus, we need to proceed carefully in applying probabilistic models to their risks.

We can easily calculate the probability of accidents of a person that flies on an airplane everyday by applying the method of “probability of the complement of an event” that we learn in high school. Setting the probability of encountering an accident with a single flight to p , and the number of flights to n , the probability of not being involved in an accident with a single flight is $1 - p$. The probability, thus, of not encountering an accident at all with n flights is $(1 - p)^n$. Subtracting this probability from the entire probability of 1, we find the probability of being involved with at least one accident is $1 - (1 - p)^n$. The frequency today of airplane accidents is about 0.3 times for each million flights; thus, if a person flies on an airplane everyday for 80 years, the probability we are looking for is:

$$1 - \left(1 - \frac{0.3}{1,000,000}\right)^{365 \times 80} \approx 0.0087, \quad (10.8)$$

that is about 0.87%.

A variable that changes its values probabilistically is called a random variable, and a random variable that changes its chances with time is called a stochastic process. When the set of values that a random variable can take, like in the case of amount of rainfall or earthquake magnitude, forms a section of continuous real values. In this case, the distribution function that defines the probability that the value is less than a threshold or its derivative, the probabilistic density function, gives an idea of what the probability distribution is like. The sum of many mutually independent random variables forms a normal distribution. Normal distributions are often used to model probabilistic random phenomena like measurement errors. The probabilistic density function for a normal distribution is bilaterally symmetric around the average value. A random variable with normal distribution is most likely to take a value near the average, and the probability drops as the value shifts away from the average.

On the other hand, for stochastic processes with increasing occurrence probability with time, like for the case of machine failure, the random variable that expresses the machine life or time to failure follows the Weibull distribution. If the distribution function or the probabilistic density function is known, we can find the confidence interval about characteristic values of the average or variance without having to observe the phenomena for an infinite number of times.

Odds is a measure that we often use in comparing the likelihood of events. We can calculate the odds by dividing the number of cases that the event of interest is taking place by the number of cases it is not. If the occurrence probability of event is known, the odds for the event is the occurrence probability divided by the probability the event does not occur. For example, assume there are ten test procedures for a certain illness and A was positive with five tests, whereas B had eight positive results. A’s odds is $5/(10-5) = 1$ and B’s odds $8/(10-8) = 4$. B’s odds is 4 times that of A and the possibility of B having the illness is 4 times that of A. In general, the

odds of an event of interest divided by that of a reference event is called the odds ratio. As we will discuss in the next section, the odds ratio is often used in statistically estimating the magnitude of risk for an event.

In quantitatively evaluating risk, we also need to decide how to calculate the magnitude of damage. When calculating casualty deduction for income tax filing, the following equation rationally determines the property damage from a disaster based on current value:

$$[\text{damage amount}] = ([\text{acquisition cost}] - [\text{depreciation from acquisition to damage}]) \times [\text{damage ratio}].$$

On the other hand, damage to facilities and buildings is often calculated with

$$[\text{damage amount}] = [\text{replacement cost}] \times [\text{damage ratio}],$$

because the cost for reconstruction has to enter the equation. In case of a major disaster that caused damage to transportation systems, the indirect cost of opportunity loss for being unable to use the systems sometimes enters the calculation in addition to the direct cost of reconstruction.

Among objects lost with tsunami or fires, there are things like “photo album of memories” that are difficult to give monetary evaluation, i.e., things that cause big psychological pain when lost. In case it is difficult to directly calculate the absolute value of a property, comparing the relative value to evaluations of other properties can lead to absolute values. As we can easily verify, when there are n pieces of property, $1, \dots, n$, with values w_1, \dots, w_n , the $n \times n$ matrix with (i, j) -th entry w_i/w_j has the maximum eigenvalue n , and the corresponding eigenvector is $(w_1, \dots, w_n)^T$. Then for each pair of (i, j) from the set $i, j \in \{1, \dots, n\}, i \neq j$, having the owner answer how many times property i is worth property j and set the answer to a_{ij} . The diagonal components of the matrix are 1, and the eigenvector corresponding to the maximum eigenvalue of this matrix is a multiple of vector $(w_1, \dots, w_n)^T$ of absolute values of properties $1, \dots, n$.

10.2 Analysis and Forecast Models of Risks

We can estimate occurrence probability of an independent event by counting the number of occurrences of the event of interest during a large number of trials. In fact, if the event of interest took place X times while repeating the trial N times, with a large enough N and X/N at a reasonable value, the occurrence probability p of the event of interest is within the following range with a 95% confidence level:

$$\frac{X}{N} - 1.96s \leq p \leq \frac{X}{N} + 1.96s. \quad (10.9)$$

In this equation, s is standard error expressed with the following equation:

$$s = \sqrt{\frac{X(N-X)}{N^3}} = \sqrt{\frac{\frac{X}{N}(1-\frac{X}{N})}{N}}. \quad (10.10)$$

In case an event of interest took place 38 times out of 380 independent trials, the 95% confidence interval for the occurrence probability is evaluated at 0.07–0.13. From the definition of standard error, quadrupling the number of trials N will halve the confidence interval for the occurrence probability p .

When the occurrence probability is extremely small, like in the case of disasters or accidents, the number of occurrences of the event of interest during a large number of repeated trials is of interest. In general, during n independent trials of an event with occurrence probability p , binominal distribution gives the probability of the event of interest to occur k times:

$$B_{n,p}(k) = {}_n C^n k p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}. \quad (10.11)$$

For example, if we set the probability p of getting 1 with one throw of die at $p = 1/6$, we expect to get 1 once with six throws of the die. The probability of getting 1 exactly once, however, is only $B_{6,1/6}(1) \simeq 0.402$, twice $B_{6,1/6}(2) \simeq 0.201$, and never $B_{6,1/6}(0) \simeq 0.335$.

On the other hand, when n is sufficiently large and p is sufficiently small, the Poisson distribution approximates the probability of the event of interest occurring k times at

$$P_\lambda(k) = \frac{\lambda^k}{k!} e^{-\lambda}. \quad (10.12)$$

The parameter λ is the expectation for the number of times the event of interest takes place and is calculated by $\lambda = np$. The constant e is Napier's constant, and it approximately equals 2.72. For example, with a machine that produces 1 defective product for every 500 pieces, the probability of finding at least 1 defective product in 1000 pieces produced by this machine is:

$$1 - P_{1000/500}(0) = 1 - \frac{1}{e^2} \simeq 0.865. \quad (10.13)$$

In general, there are multiple factors that cause a damage, and we often find different risk sizes for these factors. If you have two groups, one with factor A and the other without, continuous observation of the two counting the number of cases

Table 10.1 Cross table of causal correlation

		Result B		Total
		Occurred	Did not occur	
Factor A	With	a	b	l
	Without	c	d	$n-l$
Total		m	$n-m$	n

with damaging result B can lead to statistical evaluation of how factor A affects the occurrence of result B. With observation results shown in Table 10.1, the risk of result B occurring with factor A is a/l , whereas that without factor A is $c/(n-l)$; thus, the risk of facing result B when factor A is present is:

$$q = \frac{a}{l} \div \frac{c}{n-l} = \frac{a(c+d)}{c(a+b)} \quad (10.14)$$

times the risk without factor A. The value q , in general, is called relative risk or risk ratio.

When the time to see whether result B takes place or not takes too long, there are cases that the risk ratio can be estimated by comparing the number of objects with and without factor A for objects with and without result B. With the observation results in Table 10.1, the odds of factor A in the group with result B is a/c , and that for the group without result B is b/d . Thus, the odds ratio of the former to the latter is:

$$r = \frac{a}{c} \div \frac{b}{d} = \frac{ad}{bc}. \quad (10.15)$$

When the probability of occurrence of result B is extremely low, i.e., when $a \ll b$, $c \ll d$, the odds ratio r is a good approximation of risk ratio q .

When Table 10.1 is not the observation results of the entire investigation objects, but of n samples, randomly picked out from the parent population, the odds ratio is within the range:

$$\frac{ad}{bc} \div e^{1.96s} \leq r \leq \frac{ad}{bc} \times e^{1.96s} \quad (10.16)$$

with a probability of 95%. The symbol s is standard deviation defined with the following equation:

$$s = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}. \quad (10.17)$$

For example, if $a = b = c = d = 460$, $e^{1.96s} \simeq 1.2$ and accurate estimation of the odds ratio will require a fairly large sample size.

When the factors that cause damaging results include quantitative data, we can analyze the risk with generalized linear model. Generalized linear model estimates the occurrence probability p of result B based on observations x_1, \dots, x_n for factors A_1, \dots, A_n , with the following equation:

$$p = \frac{1}{1 + e^{-z}}, \quad z = a_1x_1 + \dots + a_nx_n + h. \quad (10.18)$$

The calculation of determining the coefficients a_1, \dots, a_n and constant h from observations of factors A_1, \dots, A_n with result B occurrence and those without result B occurrence is called logistic regression analysis. For generalized linear models, the odds $p/(1 - p)$ of result B is e^{-z} ; thus, when factor A_i increases by 1 unit, the risk of occurrence of result B is e^{-z} times in terms of odds ratio.

10.3 Decision-Making for Risk Minimization

Mathematical programming is one of the methodologies applied in rationally solving decision-making problems that we encounter in various fields of natural sciences and social sciences. Mathematical programming formulates the decision-making problem into a mathematical optimization problem to maximize or minimize the value of the objective function with variables subject to some constraints. Thus, by setting the policies for disaster management and accident prevention to decision variables, the physical and social conditions that govern the policies to constraints, and the sizes of possible risks under the policies to the objective function, we can solve the decision-making problem within the framework of mathematical programming.

When the objective function is linear and the constraints are a system of linear equations or inequalities, the optimization problem is called linear programming and is expressed in the following manner:

$$\begin{aligned} &\text{minimize : } \mathbf{c}^T \mathbf{x} \\ &\text{subject to : } \mathbf{Ax} = \mathbf{b}, \mathbf{x} \geq 0. \end{aligned}$$

For this set of formulae, \mathbf{x} is the vector of decision variables, \mathbf{A} is the parameter matrix, and \mathbf{b} and \mathbf{c} are parameter vectors. Setting the sizes of risks with accidents or disasters is difficult, and parameters \mathbf{A} , \mathbf{b} , and \mathbf{c} have uncertainties. Especially when parameters \mathbf{A} and \mathbf{b} have uncertainties, both sides of the equation $\mathbf{Ax} = \mathbf{b}$ are uncertain; thus, it takes clarifying the constraints that require the two sides are equal for solving the linear programming problem. We then turn the problem into a “Chance Constraining Problem” that looks for the decision variables \mathbf{x} that minimize the objective function $\mathbf{c}^T \mathbf{x}$ under the constraints that the probability of the equation $\mathbf{Ax} = \mathbf{b}$ or its fuzzy measure is not less than a certain threshold or the “recourse problem” that adds the magnitude of the residual error $\mathbf{Ax} - \mathbf{b}$ to the objective function instead of equation $\mathbf{Ax} = \mathbf{b}$.

If the range of parameters A and b are known, the method to find decision variables x that minimize the objective function $c^T x$ while the equation is satisfied whatever values A and b take is another effective approach (“robust optimization”). In assessing the effectiveness of risk management, “worst case analysis” is also well practiced that identifies the case that maximizes the objective function $c^T x$ among optimized results of linear programming problems for all possible combinations of parameters A and b . Worst case analysis takes solving the two-level mathematical programming problem that have the original linear programming problem in its lower level.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits any noncommercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if you modified the licensed material. You do not have permission under this license to share adapted material derived from this chapter or parts of it.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

