A Study of the hierarchical structure of risk models and proposal of a new approach

Yoshitaka Hoshino
Tomohiro Uezono
Eiji Oshima
Planning Group, Commercial Lines Underwriting Department
Tokio Marine & Nichido Fire Insurance Co., Ltd.

The content of this paper represents the personal views of the authors, and does not represent the official opinions of either the Tokio Marine & Nichido Fire Insurance Co., Ltd. or the Institute of Actuaries of Japan.

Introduction

1. Risk Models

Unlike that of another business, the direct objective of the insurance business is to take risks in order to gain profit. It therefore goes without saying that an understanding of the characteristics and details of risk form the foundation of corporate strategy in an insurance company.

Since risks are uncertain not only for the insured, but also for the insurance company (although the asymmetric nature of information must also be taken into consideration), those risks must be expressed quantitatively through some judgment. This expression, which is referred to as risk modeling, is indispensable in order for insurance companies to accept an appropriate level of risk under appropriate conditions, and to manage the outcomes. Due to the factors mentioned below, the importance of risk modeling is regarded to be increasing.

a. Due to the deregulation of the financial administration policies in Japan, the demand for accountability and internal governance at insurance companies has increased. In other words, the need to account for management and responsibilities using economic indicators has been increased.
b. Insurance companies have turned to the sophistication and optimization of risk measurements, in order to maintain profits under the heightened competitive environment resulting from deregulation.
c. Due to the international trend to place priority on the protection of investor and market profitability, there has been an increased demand for making risks transparent, and recording appropriate profits considering the risks.

It is under such circumstances that individual actuaries work with the measuring and modeling of risks. It should be noted that different models are constructed for the same risk, depending on the intended use of the risk model from the following viewpoints:

a. What level of risk segmentation is required;

b. On which risk measure and which type of the estimation of parameter should more emphasis be placed;
c. What conservative factors should be built into the model according to the answer to b. above.

In this paper, observations are made with respect to question a. above.

2. Segmentation of a risk model

As stated above, the required level of segmentation of a risk model differs greatly depending on its intended use. For example, models used for capital management are not segmented, and only require that risk evaluation be performed per business line. On the other hand, such non-segmented models are, of course, unsuitable for underwriting and even for tariff-making as well.

It goes without saying that it is desirable for business management that the models used for individual purposes are quantitatively consistent with each other. One common solution to this problem is to utilize the segmented risk models (hereafter referred to as the "lower model") prepared for tariff-making as a model for capital management (hereafter referred to as the "upper model") by aggregating the lower models. This means that an attempt is made to evaluate the whole by aggregating the segmented risk evaluations.

However, there are three disadvantages to this method, as mentioned below. For these reasons, the approach mentioned above may not lead to a total optimization of risk evaluation and corporate strategies.

a. Placing emphasis on individual risk measures may lead to errors in whole risk measures.
b. This procedure prevents top-down decision making.
c. Since the responsibilities for the construction of lower models are decentralized, it becomes more difficult for management or administrative departments to govern and check on risk evaluations in a practical fashion.

3. Proposition for a hierarchical structure of risk models

We introduce a concept called the "hierarchical risk model" constructed by the following two factors as a concrete approach to address the issues stated above.

a. In this approach, we perform a "segmentation of the upper model" (i.e., generate individual risk models) while maintaining the risk measure of the optimized "upper model" as much as possible, and reflecting the risk characteristics of each "lower model."
b. We construct an algorithm for deciding how to segment the optimized "upper model" into "lower models." In other words, given an aggregation of insurance policies, we select the risk factors which cause differentiations in the risk characteristics, and determine the appropriate level of segmentation.

By utilizing the "hierarchical risk model" approach, it becomes possible to attain consistency between the "upper model" and "lower model", while taking advantage of the stable analytical result from the “upper model.” The actual techniques for processes a. and b. in this approach are demonstrated in Chapters 1 and 2, respectively.

Chapter 1 Estimation of lower models utilizing the upper model

1. Introduction

Our objective is to estimate the distributions of the random variables (e.g., severity) in a given risk segment. However, it is practically difficult to obtain a sample set which is sufficiently large to capture the shape of the probability distribution. Therefore, an approach is often used whereby the shape of the distribution is defined using a small number of parameters, and these parameters are estimated. The maximum likelihood estimation is an example of this approach. This is effective if sufficient information exists regarding the shape of the probability distribution, and the shape can be determined using a small number of parameters. However, random variables comprise various factors, and it is difficult to prescribe these factors using only a small number of parameters. In addition, it is realistically improbable that sufficient information on the shape of a probability distribution can be obtained when the sample size is insufficient.

In considering the hierarchical risk model, an approach using Bayes' Theorem is utilized in order to address this issue. The assumption for proceeding with the discussions in this paper is as follows.

Assumption

While the number of samples in a given risk segment is small, a sufficient number of samples exists if restriction is not made to a particular risk segment (i.e., the upper model has been established).

2. Bayes' Theorem

First of all, a summary of Bayes' Theorem is given below. Bayes' Theorem relates the joint, marginal, and conditional distributions of two random variables. If the conditional distribution of a random variable X, which is dependent on parameter θ is modelled by $P(x \mid \theta)$, and the distribution of θ is also modeled by $P(\theta)$, then Bayes' Theorem states that the joint distribution, $P(x,\theta)$ of X and θ is given by:

$$P(x,\theta) \equiv P(x \mid \theta)P(\theta).$$

The following set of relational equations is obtained by transforming the joint distribution:

$$\begin{aligned} P(x) &\equiv \int P(x \mid \theta)\, d\theta \\ P(x \mid \theta) &\equiv \frac{P(x,\theta)}{P(x)} \\ P(\theta \mid x)P(x) &= P(x,\theta) \\ P(x \mid \theta)P(\theta) &= P(\theta \mid x)P(x). \end{aligned}$$

In the above equations, x is regarded as sample (i.e., data). However, to clarify this assignment, the symbol d is used rather of x, hereafter. In addition, θ becomes the parameter for the distribution of d, and $P(\theta)$ represents the distribution of θ, independent of the sample ($P(\theta)$ is thus, hereafter referred to as the "*prior* distribution of θ"). In order to emphasize the difference of such roles, the distribution of θ is denoted by $\pi(\theta)$ instead $P(\theta)$. Thus, the last relational equation given above becomes:

$$P(d \mid \theta)\pi(\theta) = \pi(\theta \mid d)P(d). \tag{1}$$

While we are mainly interested in the parameter θ, we are particularly interested in θ under a given sample d, and not in θ itself, in general. This can be obtained from $\pi(\theta \mid d)$ in the above equation. Rearranging the equation (1) with respect to $\pi(\theta \mid d)$ yields:

$$\pi(\theta \mid d) = \frac{P(d \mid \theta)\pi(\theta)}{P(d)}.$$

If $P(d)$ is taken as a normalization constant, this becomes:

$$\pi(\theta \mid d) \propto P(d \mid \theta)\pi(\theta).$$

This means that by choosing $P(d \mid \theta)$ and $\pi(\theta)$ appropriately, the above equation can yield $\pi(\theta \mid d)$. Lastly, we introduce a hyper-parameter ω for $\pi(\theta)$. While $\pi(\theta)$ was introduced as the *prior* distribution of θ, this can alternatively be viewed as a (soft) constraint to the *prior* information θ, where the constraint is represented by $\pi(\theta)$. The $\pi(\theta)$ would contain the extent of constraint against the *prior* information and other parameters, which is denoted by ω. In addition, using the hyper-parameter ω, $\pi(\theta)$ is written as $\pi(\theta;\omega)$, and $\pi(\theta \mid d)$ as $\pi(\theta \mid d;\omega)$. Accordingly, $\pi(\theta \mid d;\omega)$ becomes:

$$\pi(\theta \mid d;\omega) \propto P(d \mid \theta)\pi(\theta;\omega). \tag{2}$$

3. Proposed Approach

We apply $\pi(\theta \mid d;\omega)$ to our problem. First, let the θ represent many parameters, which differs from the maximum likelihood estimation which only utilizes a small number of parameters. In this paper, we discretize the distribution, and consider the probability in each segment as a parameter. In other words, θ is represented as follows:

$$\theta \equiv \vec{\theta} = (p_1, p_2, p_3, \cdots, p_k).$$

In the above equation, p_i denotes the probability in segment i and $\pi(\vec{\theta};\omega)$ can be regarded as a (soft) constraint for the *prior* information $\vec{\theta}$, as mentioned above. Then, what can be adopted for the *prior* information for $\vec{\theta}$? As stated in the underlying assumption, there is sufficient information for the upper model. In addition, since a given risk segment is a part of the upper model, we believe that it is logical to take the upper model as the *prior* information for $\vec{\theta}$. Thus, a determination of $\pi(\vec{\theta};\omega)$ is interpreted as a determination of the constraining condition for $\vec{\theta}$ using the upper model. Let the functional form of $\pi\left(\vec{\theta};\omega\right)$ be assumed to be expressed via the following equation:

$$\pi\left(\vec{\theta};\omega\right)\propto \frac{1}{\sqrt{2\pi\omega^2}^{k}}\exp\left(-\frac{\left\|\vec{\theta}-\vec{\varphi}\right\|_2^2}{2\omega^2}\right). \tag{3}$$

In the equation (3), $\vec{\varphi}$ can be expressed as the probabilities q_i in segment i of the upper model using the formula below, and $\left\|\vec{\theta}-\vec{\varphi}\right\|_2$ is the 2- norm.

$$\vec{\varphi}=\left(q_1,q_2,q_3,\cdots,q_k\right)$$

The term $\pi\left(\vec{\theta};\omega\right)$ represents the fact that the distance between $\vec{\theta}$ and $\vec{\varphi}$ will be constrained by ω. In addition, the proportionality sign ($\propto$) is used because the right hand side of the equation (3) is inadequate as a probability expression. By expanding the right hand side of the equation (3), the following is obtained:

$$\prod_i \frac{1}{\sqrt{2\pi\omega^2}}\exp\left(-\frac{\left(p_i-q_i\right)^2}{2\omega^2}\right).$$

Since p_i is the probability in segment i, its domain is bounded, and the expression does not satisfy the criteria for a probability (i.e., that the sum equals 1). This is the reason for using the proportionality sign. Finally, we define $P\left(d\mid\vec{\theta}\right)$ as follows. First, data d can be expressed as follows by denoting the number of data in segment i as n_i:

$$d\equiv\vec{d}=\left(n_1,n_2,n_3,\cdots,n_k\right).$$

Since $P\left(\vec{d}\mid\vec{\theta}\right)$ is the probability that data n_i occur when the probability in each segment is determined to be p_i, it can be expressed as a likelihood function:

$$P\left(\vec{d}\mid\vec{\theta}\right)=\frac{N!}{\prod_i n_i!}\prod_i p_i^{n_i}, \tag{4}$$

where $N=\sum_i n_i$.

By summarizing the arguments to this point, $\pi\left(\theta\mid\vec{d};\omega\right)$ can be expressed using the following equation:

$$\pi\left(\vec{\theta}\mid\vec{d};\omega\right)\propto\frac{N!}{\prod_i n_i!}\prod_i p_i^{n_i}\times\frac{1}{\sqrt{2\pi\omega^2}^{k}}\exp\left(-\frac{\left\|\vec{\theta}-\vec{\varphi}\right\|_2^2}{2\omega^2}\right).$$

Expanding $\left\|\vec{\theta}-\vec{\varphi}\right\|_2^2$ yields the following:

$$\pi\left(\vec{\theta}\mid\vec{d};\omega\right)\propto\frac{N!}{\prod_i n_i!}\prod_i\frac{p_i^{\;n_i}}{\sqrt{2\pi\omega^2}}\exp\left(-\frac{(p_i-q_i)^2}{2\omega^2}\right) \tag{5}$$
.

Having defined $\pi\left(\vec{\theta}\mid\vec{d};\omega\right)$, this can be used to estimate $\vec{\theta}$. While the expected value of $\vec{\theta}$ could also be used to estimate $\vec{\theta}$, in this paper, we adopt the $\vec{\theta}$ value which maximizes $\pi\left(\vec{\theta}\mid\vec{d};\omega\right)$. This is referred to as the maximum *a posteriori* (MAP) estimation. We wish to derive $\vec{\theta}$ that maximizes $\pi\left(\vec{\theta}\mid\vec{d};\omega\right)$ according to the MAP estimation. Since $\vec{\theta}$ is the probability in each segment, it must satisfy the following two conditions:

$$\begin{aligned}&p_i\geq 0\quad i=1,2,3,\cdots,k\\&\sum_k p_i=1\end{aligned}$$
.

In this paper, the maximization of $\pi\left(\vec{\theta}\mid\vec{d};\omega\right)$ under the above conditions is expressed using Lagrange Multiplier Method. Since $p_i\geq 0$, let us assume $p_i=\exp(r_i)$. This is not a rigorous expression, but is used to simplify the expansion of the formula. If we choose to maximize $\ln\pi\left(\vec{\theta}\mid\vec{d};\omega\right)$ instead of $\pi\left(\vec{\theta}\mid\vec{d};\omega\right)$, then applying Lagrange Multiplier Method yields the following formulation:

$$\begin{cases}\dfrac{\partial}{\partial r_i}T=0\quad i=1,2,3,\cdots k\\[2ex]\dfrac{\partial}{\partial\lambda}T=0\end{cases}$$

$$where\quad T=\ln\pi\left(\vec{\theta}\mid\vec{d}\right)-\lambda\left(\sum_i p_i-1\right).$$

In addition, expanding T yields the following:

$$T=C+\sum_i\left(n_i r_i-\frac{(\exp(r_i)-q_i)^2}{2\omega^2}\right)-\lambda\left(\sum_i\exp(r_i)-1\right),$$

where C is a constant for r_i and λ. Substituting T into the above equation yields the following simultaneous equation:

$$\begin{cases} n_i - \dfrac{\exp(r_i)(\exp(r_i)-q_i)}{\omega^2} - \lambda \exp(r_i) = 0 \quad i = 1,2,3,\cdots k \\ \sum_i \exp(r_i) = 1 \end{cases}.$$

By inserting $\exp(r_i)$ back into p_i , this becomes:

$$\begin{cases} p_i^{\,2} - (q_i - \lambda\omega^2)p_i - n_i\omega^2 = 0 \quad i = 1,2,3,\cdots k \\ \sum_i p_i = 1 \end{cases}.$$

The first equation, above is a quadratic equation in p_i , so the solution is given by the following:

$$p_i = \frac{q_i - \lambda\omega^2 + \sqrt{(q_i - \lambda\omega^2)^2 + 4n_i\omega^2}}{2} \quad i = 1,2,3,\cdots k \qquad (6)$$

In addition, p_i must satisfy $\sum_i p_i = 1$. Although this condition determines λ, it cannot be expressed analytically. Thus, λ is left as a constant which satisfies $\sum_i p_i = 1$.

We have thus far handled ω as a given constant, but it can be seen from the equation for p_i that the value of ω greatly affects the estimation of p_i . For determining the value of ω, we use the method which adopts the ω that minimizes the AIC (Akaike's Information Criteria). In this paper, we define the AIC as below (in this case, the dimension of ω is 1 so that we ignore the term):

$$-2\ln \pi(\vec{d};\omega).$$

Here $\pi(\vec{d};\omega)$ is obtained by integrating $\vec{\theta}$ out of $\pi(\vec{\theta} \mid \vec{d};\omega)$. In this paper, a simplified method (not rigor) is used to evaluate ω.
As given above, $\pi(\vec{\theta} \mid \vec{d};\omega)$ is expressed as follows:

$$\pi(\vec{\theta} \mid \vec{d};\omega) \propto \frac{N!}{\prod_i n_i!} \prod_i \frac{p_i^{\,n_i}}{\sqrt{2\pi\omega^2}} \exp\left(-\frac{(p_i - q_i)^2}{2\omega^2} \right).$$

The term $\prod_i p_i^{\,n_i}$ becomes an issue when trying to integrate $\vec{\theta}$ out. Thus, we first approximate $\prod_i p_i^{\,n_i}$. Since $\prod_i p_i^{\,n_i}$ is maximum when $p_i = \dfrac{n_i}{N}$ $(i = 1,2,3,\cdots,n)$, expanding around this maximum value (i.e., let $p_i = \dfrac{n_i}{N} + \Delta p_i$ where $\Delta p_i << 1$) allows us to transform $p_i^{n_i}$ as follows (note that the assumption $\Delta p_i << \frac{n_i}{N}$ does not hold when $n_i = 0$, but we consider this to hold for the sake of simplicity):

$$\begin{aligned} p_i^{n_i} &= \left(\frac{n_i}{N} + \Delta p_i \right)^{n_i} \\ &= \left(\frac{n_i}{N} \right)^{n_i} \left(1 + \frac{N}{n_i} \Delta p_i \right)^{n_i} \\ &\approx \left(\frac{n_i}{N} \right)^{n_i} \left(1 + N \Delta p_i + \frac{N(n_i - 1)}{2} \Delta p_i^2 \right) . \end{aligned}$$

Therefore, $\prod_i p_i^{n_i}$ becomes as follows:

$$\begin{aligned} \prod_i p_i^{n_i} &\approx \prod_i \left(\frac{n_i}{N} \right)^{n_i} \left(1 + N \Delta p_i + \frac{N(n_i - 1)}{2} \Delta p_i^2 \right) \\ &\approx \prod_i \left(\frac{n_i}{N} \right)^{n_i} \times \prod_i \left(1 + N \Delta p_i + \frac{N(n_i - 1)}{2} \Delta p_i^2 \right) \\ &\approx \prod_i \left(\frac{n_i}{N} \right)^{n_i} \times \left(1 + N \sum_i \Delta p_i + N^2 \sum_{i \neq j} \Delta p_i \Delta p_j + \sum_i \frac{N(n_i - 1)}{2} \Delta p_i^2 \right) . \end{aligned}$$

In transforming the second equation to the third above, all terms beyond the third term of Δp_i were ignored. By using the relationship $\sum_i \Delta p_i = 0$, $\prod_i p_i^{n_i}$ can be expressed as follows (assuming $\frac{n_i - 1}{2N} << 1$).

$$\begin{aligned} \prod_i p_i^{n_i} &\approx \prod_i \left(\frac{n_i}{N} \right)^{n_i} \times \left(1 - N^2 \sum_i \Delta p_i^2 + \sum_i \frac{N(n_i - 1)}{2} \Delta p_i^2 \right) \\ &\approx \prod_i \left(\frac{n_i}{N} \right)^{n_i} \times \prod_i \exp\left(- N^2 \Delta p_i^2 \left(1 - \frac{n_i - 1}{2N} \right) \right) \approx \prod_i \left(\frac{n_i}{N} \right)^{n_i} \times \prod_i \exp\left(- N^2 \left(p_i - \frac{n_i}{N} \right)^2 \right) . \end{aligned}$$

From the above equation, $\pi\left(\vec{\theta} \mid \vec{d}; \omega\right)$ becomes:

$$\pi\left(\vec{\theta} \mid \vec{d}; \omega\right) \propto N! \prod_i \frac{1}{n_i!} \left(\frac{n_i}{N} \right)^{n_i} \times \prod_i \frac{\exp\left(- N^2 \left(p_i - \frac{n_i}{N} \right)^2 - \frac{1}{2\omega^2} (p_i - q_i)^2 \right)}{\sqrt{2\pi\omega^2}} .$$

In order to integrate $\vec{\theta}$ out of $\pi\left(\vec{\theta} \mid \vec{d}; \omega\right)$, the following must be computed:

$$\pi\left(\vec{d};\omega\right) \propto N!\prod_i \frac{1}{n_i!}\left(\frac{n_i}{N}\right)^{n_i} \times \int d\vec{\theta} \prod_i \frac{\exp\left(-N^2\left(p_i - \frac{n_i}{N}\right)^2 - \frac{1}{2\omega^2}\left(p_i - q_i\right)^2\right)}{\sqrt{2\pi\omega^2}}.$$

However, in order to calculate $\int d\vec{\theta}$, the integration range must be taken over all combinations of $\{p_i\}$ which satisfy the conditions. Since performing this integration analytically is difficult, the following simplification is considered. The term inside the exponential for each i is quadratic in p_i, which is maximized when

$$p_i = \frac{N^2 \times \frac{n_i}{N} + \frac{1}{2\omega^2} \times q_i}{N^2 + \frac{1}{2\omega^2}}.$$

Since this satisfies the condition $\sum_i p_i = 1$ as well, a value of p_i that maximizes $\pi\left(\vec{\theta} \mid \vec{d};\omega\right)$ exists regardless of ω. With respect to $\int d\vec{\theta}$, while in principal, the sum must be taken over all combinations of $\{p_i\}$ that satisfy the conditions, we ignore conditions $\sum_i p_i = 1$ and $p_i \geq 0$ for simplicity, and consider p_i to be independent and unbounded. Note, however, that the existence of a p_i value which maximizes $\pi\left(\vec{\theta} \mid \vec{d};\omega\right)$ under the assumption $\sum_i p_i = 1$ is assured. Since the computation of $\int d\vec{\theta}$ under such an assumption for p_i is a Gaussian integration, this does not lead to a large error. Since the Gaussian integration is:

$$\int \exp\left(a(x-b)^2 + c\right)dx = \sqrt{\frac{2\pi}{a}}\exp(c).$$

$\pi\left(\vec{d};\omega\right)$ can be approximated as follows:

$$\pi\left(\vec{d};\omega\right) \propto \frac{\exp\left(-\frac{N^2}{2\omega^2 N^2 + 1} \times M\right)}{\sqrt{\omega^2 N^2 + \frac{1}{2}}^{\,k}}.$$

Here, let $M = \sum_i \left(\frac{n_i}{N} - q_i\right)^2$. The value of ω which maximizes $\pi\left(\vec{d};\omega\right)$ becomes:

$$\omega \approx \sqrt{\frac{2N^2 M - k}{2N^2 k}}. \tag{7}$$

The above expression for ω can be interpreted as follows. Whereas M is a measure that represents the gap in the sample, ω becomes greater when the gap is large, and the constraint to the upper model becomes weaker. Thus, ω self-adjusts in correspondence with the gap between the upper model and the sample.

4. Some actual examples

Finally, we introduce some actual examples that utilize the method proposed in this paper (hereafter referred to as "the proposed method"). The assumptions for the actual examples are given as follows.

- The number of accidents per insured person (lower model) follows a Poisson distribution with a mean of λ.
- The distribution of λ is a normal distribution with a mean of 2 and a standard deviation of 0.5 (assume that λ takes on discrete values in whole 0.5 units, and λ is chosen to be 0 when negative).

A loss simulation was performed for a portfolio of 1,000 contracts using the above assumptions to yield the following accident count table (we take this to be the upper model[1]):

Number of Accidents	0	1	2	3	4	5	6	7	8	Total
Number of Samples	153	256	264	179	88	47	8	2	3	1000
Constituent Ratio	0.153	0.256	0.264	0.179	0.088	0.047	0.008	0.002	0.003	1.000

Further, the data belonging to a risk segment (λ=1.5) in the same simulation was as follows:

Number of Accidents	0	1	2	3	4	5	6	7	8	Total
Number of Samples	10	10	11	6	4	0	0	0	0	41
Constituent Ratio	0.244	0.244	0.268	0.146	0.098	0.000	0.000	0.000	0.000	1.000

Application of the above result, using the method described in this paper to estimate the lower model of the same risk segment yields the following:

Number of Accidents	0	1	2	3	4	5	6	7	8	Total
Constituent Ratio	0.177	0.261	0.272	0.180	0.097	0.013	0.000	0.000	0.000	1.000

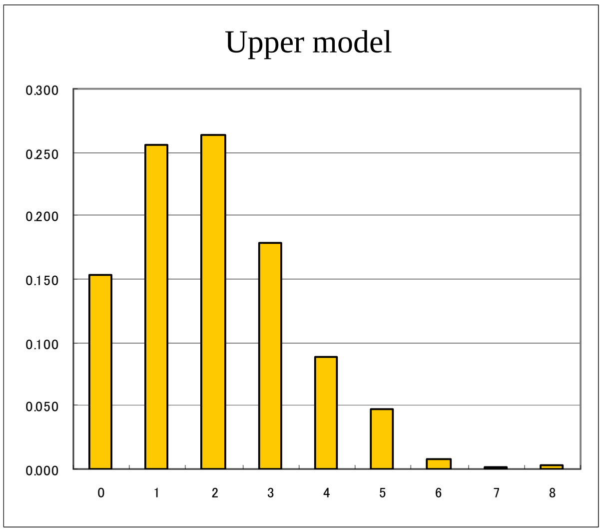
It can be seen from the graphs on the next page that the lower model is estimated based on the "upper model" and the "samples in the risk segment".

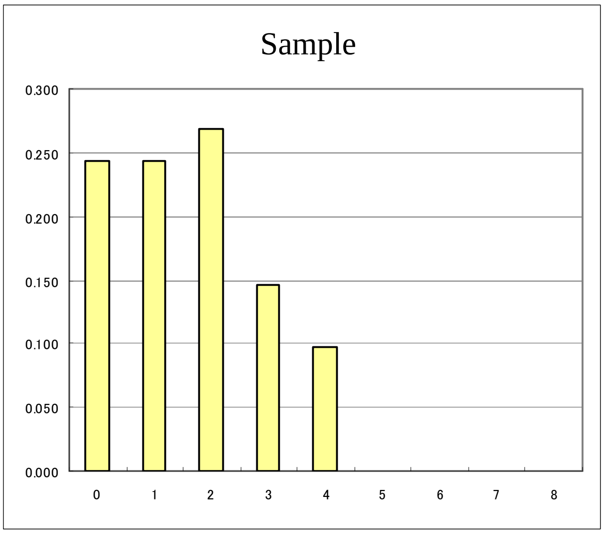
While in reality, the estimation of the probability distribution of a risk segment with a small number of data is difficult (or has associated dangers[2]), this becomes possible using the proposed method.

---

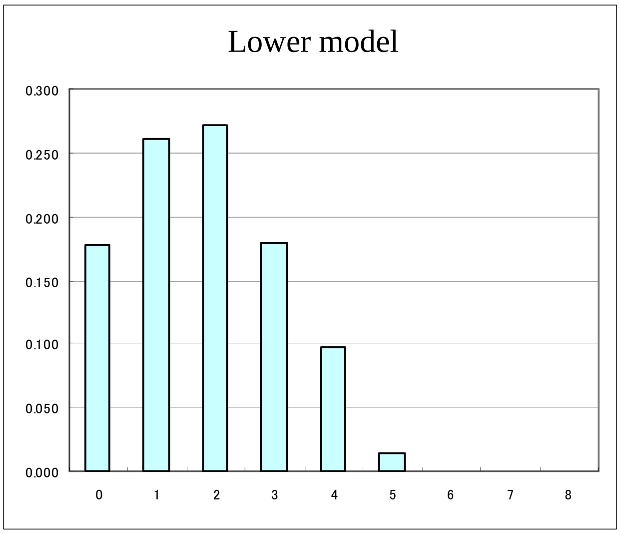
[1] In actuality, the results from the simulation are samples for the upper model, and it is necessary to estimate the upper model parametrically using these results (whereas here, we have used the result of the simulation as the optimized upper model).

[2] We assumed a Poisson distribution for generating the simulated samples, but such information is not available for real phenomena. Thus, determining the probability distribution (shape and parameters) for a risk segment with limited data involves associated dangers.

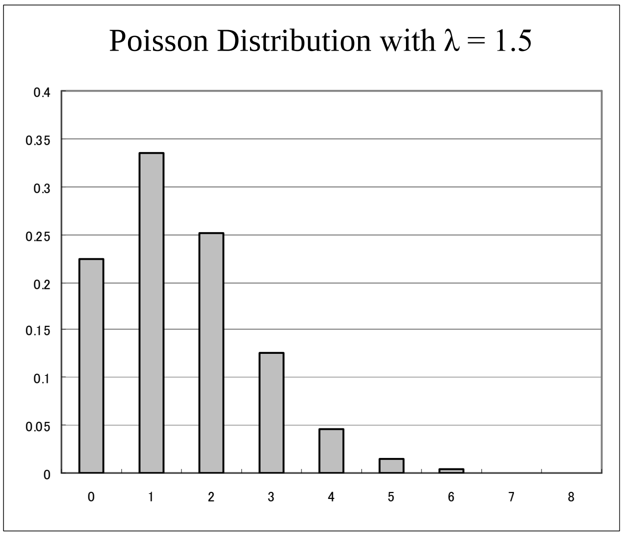

Upper model


Sample

Proposed Methodology


Lower model

Reference: Poisson distribution with λ=1.5


Poisson Distribution with λ = 1.5

Chapter 2

1. Problem statement

In Chapter 1, we examined a method for estimating the distribution of random variables (e.g., damage rates) in a given risk category using the upper model. In other words, we have proceeded with the discussion based on the assumption that a risk category already exists. In reality, however, risk categories must be evaluated and set by ourselves. In this chapter, we examine the question, "what are the appropriate risk categories for estimating distribution?" for the purpose of expanding upon the previous discussion.

[Algorithm for generating risk categories]
We examine a method for generating risk categories quantitatively using a model case. In the model case, it is assumed that there are four risk factors, "geographic region", "vehicle type", "gender", and "age" that affect the loss ration in automobile insurance.

Risk Factor	Risk Factor Elements
Geographic Region	Northern Japan, Eastern Japan, Western Japan, Southern Japan
Vehicle Type	Sedan, Sports Car, RV
Gender	Male, Female
Age	Less than 30, 30 to 50, more than 50

In order to generate risk categories based on a quantitative evaluation, the model case is defined as follows:

Region R_1 = {Northern Japan, Eastern Japan, Western Japan, Southern Japan}
Vehicle Type: R_2 = { Sedan, Sports Car, RV}
Gender: R_3 = {Male, Female}
Age: R_4 = { Less than 30, 30 to 50, more than 50}

The set Θ of the most segmented (and the most easily created) risk categories is given as the direct product of the risk factors using the following equation:

$$\Theta = \prod_{i=1}^{4} R_i$$

However, segmentation will result in issues later, such as the amount of available data in each risk category being too small, making subsequent work problematic. Thus, the desirable risk categories are generated via the following steps:

Step 1
Extract the elements in risk factor R_1 that affect the damage rate.

Step 2
If such elements exist, categorize them according to the extent and trend of the effect (e.g., if Eastern Japan exhibited a different effect compared to other regions in Japan, then categorize the elements into Eastern Japan and the remaining regions), and define it as a new risk factor R_1'. If there no such elements exist, then let $R_1' = \phi$.

Step 3
Configure R_2', R_3', R_4' in a similar fashion, and obtain a new set of risk categories using their direct product:

$$\Theta' = \prod_{i=1}^{4} R_i'$$

Hereafter, we shall proceed by examining the method for generating risk factor $R_i^{'}$.

2. General approach

As a method for generating risk factor R_i', we first introduce "multiple regression analysis" and "analysis of variance," which are considered to be the most general approaches, followed by an introduction to the method that we propose.

(1) Multiple regression analysis

The first approach to this type of problem is multiple regression analysis. Using the damage rate as the explained variable, and other risk factors, including geographic region and age as the explanatory variables, the maximum likelihood estimation is used to determine a set of regression coefficients. The extent that each risk factor affects the damage rate can then be seen as follows:

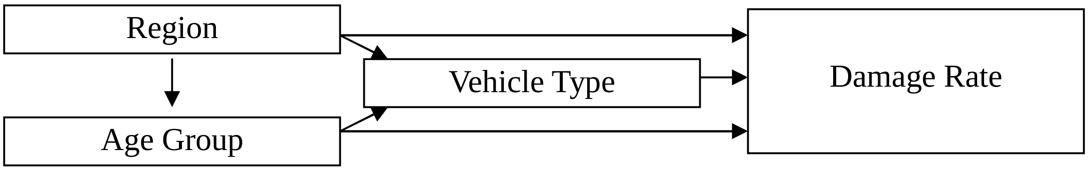
$$y_i = a + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_m x_{im} + e_i \quad (1 \le i \le n)$$

The above equation can be expressed equivalently as $\mathbf{y} = \mathbf{Xb} + \mathbf{e}$.

Problem

One issue that we must be aware of when performing regression analysis is multi-colinearity. In other words, when there are correlations between risk factors such as geographic region and vehicle type, then the regression coefficient cannot be computed correctly. For example, if younger people live in the cities, and if younger people tend to prefer sports cars, then the region affects not only the damage rate but also the age group and vehicle type.

<Conceptual Diagram>



When an explanatory variable has influence not only on the explained variable but also on other explanatory variables, then some method must be used to exclude that influence. Generally, this would mean extracting some variable that is not affected by another explanatory variable by computing a correlation matrix and partial correlation coefficients. However, this would imply completely ignoring the effect that the excluded explanatory variables exert on the explained variable.

In addition, the risk factors themselves cannot be used as explanatory variables directly, thus each element comprising the risk factors must be expressed as individual dummy variables. This is another issue, in which the number of explanatory variable becomes too numerous.

(2) Generalized linear models (GLM)

In addition to the above mentioned problem related to regression analysis, other issues exist, including the fact that an assumption is made regarding the existence of a linear relationship between the explained variable and the explanatory, and that a normal distribution is assumed for the error. In order to respond to these issues, the use of a method called generalized linear models (GLM) has become common. A GLM is a model that expands on the normal regression model in the following manner:

1) The error distribution may not necessarily need to be normal.
2) The relationship between the explanatory and explained variables is not restricted to be linear.

In other words, the model has been expanded to a generalized linear model, where the observed variable y follows an exponential family of distributions.

$$\mathbf{y} = \mathbf{Xb} + \mathbf{e} \ \Rightarrow \ f[E(\mathbf{y})] = \mathbf{Xb}$$

However, while a GLM is more flexible compared to general regression model, multi-colinearity remains an issue, and various solutions to this issue are being considered at each insurance company.

(3) Analysis of variance

An overview of the analysis of variance is next presented below, which is a test in which the differences between independent groups are assessed.

Below, we assume that damage rate x_{ij} is determined by two risk factors A and B, each of which has three elements.[3]

		A		
		A_1	A_2	A_3
B	B_1	x_{11}	x_{12}	x_{13}
	B_2	x_{21}	x_{22}	x_{23}
	B_3	x_{31}	x_{32}	x_{33}

---

[3] In standard terms used in analysis of variance, this is referred to as a two-factor three-way model. There are places where we intentionally do not use standard analysis of variance terms, including these above.

In the above table, the following relationship, called the structural formula is established:

$$x_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ij}$$
$$\sum_i \alpha_i = \sum_j \beta_j = \sum_j \gamma_{ij} = \sum_i \gamma_{ij} = 0$$

The symbols used in the above equation have the following meanings:

μ : Average
α_i : Effect of element A_i of risk factor A
β_j : Effect of element B_j of risk factor B
γ_{ij} : Effects of both elements A_i and B_j (interaction)
ε_{ij} : Error for x_{ij}

Under these conditions, the null hypotheses $H_0 : \alpha_1 = \alpha_2 = \cdots = 0$ and $H_0' : \beta_1 = \beta_2 = \cdots = 0$, are established and tested as follows.[4] In order to utilize general symbols, let l and m denote the number of elements in risk factors A and B, respectively. If H_0 is rejected, then we conclude that there are no differences in the elements of risk factor A (i.e., there is no need to consider risk factor A). A similar argument is used for H_0'.

Null Hyp.	Statistical Test (Condition for Rejecting Hypothesis)	Formulae
H_0	$F_0 = \dfrac{S_A/(l-1)}{S_E/(l-1)(m-1)} > F_{(l-1)(m-1)}^{l-1}(\varepsilon)$	$S_A = m\sum_{i=1}^{l}(\bar{x}_{i.} - \bar{x}_{..})^2$ $S_E = \sum_{i=1}^{l}\sum_{j=1}^{m}(x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})^2$ $\bar{x}_{i.} = \frac{1}{m}\sum_{j=1}^{m} x_{ij}$, $\bar{x}_{..} = \frac{1}{l}\sum_{i=1}^{l}\bar{x}_{i.}$
H_0'	$F_0 = \dfrac{S_B/(m-1)}{S_E/(l-1)(m-1)} > F_{(l-1)(m-1)}^{m-1}(\varepsilon)$	$S_B = l\sum_{j=1}^{m}(\bar{x}_{.j} - \bar{x}_{..})^2$ $\bar{x}_{.j} = \frac{1}{l}\sum_{i=1}^{l} x_{ij}$

Calculations involved in the analysis of variance can become tedious when the number of risk factors and their elements increase, but it is a suitable method for examining the difference in and the extent of the effect that each risk factor has on the damage rate.

---

[4] Since the we are assuming that repeated observations are not made in this case, we cannot test for $\gamma_{ij} = 0$; it can be tested if there are repeated observations or if there are three factors or more.

Below, we expand on the discussion using a type of analysis of variance called the Friedman's test.

3. Proposed approach

We assumed that there are four risk factors in the model case, but the number of variables including interactions becomes massive if a four-dimensional analysis of variance is to be performed. Thus, we utilize one of the risk factors, and assign blocks to it that combine the remaining risk factors. Further, we repeat the analysis for one risk factor[5] as many times as there are risk factors (i.e., we perform the Friedman's test repeatedly). Using this method, it becomes possible to exclude effects other than the selected risk factor within a single block.

Step 1

Extract a risk factor R_1 (geographic region) and combine the remaining risk factors into blocks to generate the table below. There are r blocks, and c elements for risk factor R_1 (in the model case, $r = 18, c = 4$, and a block contains a combination such as {sports car, male, less than 30 })

	1	2	……	c
Block 1	O_{11}	O_{12}	:	O_{1c}
Block 2	O_{21}	O_{22}	:	O_{2c}
:	:	:	:	:
Block r	O_{r1}	O_{r2}	:	O_{rc}

Step 2

Consider the data with a single risk factor arranged in blocks, and perform Friedman's test. In other words, sort the blocks in ascending order of damage rate of the geographic region for each block (let O_{ij} denote the ordinal in each block), and use the statistics value χ_0^2 to test the null hypothesis that asserts that there are no differences between elements of a risk factor.

$$\chi_0^2 = \frac{12}{rc(c+1)}\sum_{j=1}^{c} K_j^2 - 3r(c+1)$$

In the above equation, $K_j = \sum_{i=1}^{r} O_{ij}$ and χ_0^2 has a χ^2 distribution with $c-1$ degrees of freedom. It is also clear that $\sum_{j=1}^{c} K_j = rc(c+1)/2$.

---

[5] Perform a single factor analysis using randomized blocks

Step 3
If the null hypothesis is rejected in Step 2, Scheffe's method can be used inspect whether the differences between the elements are significant.[6] In particular, test statistics S_{ij} can be obtained using the following equation:

$$S_{ij} = \frac{r^2(c-1)(\overline{K}_i - \overline{K}_j)^2}{2V}$$

where $\overline{K}_j = \sum_{i=1}^{r} \frac{O_{ij}}{r} = \frac{K_j}{r}$ and $V = \sum_{i=1}^{r}\sum_{j=1}^{c}\left(O_{ij} - \frac{c+1}{2}\right)^2$

In this case S_{ij} follows a χ^2 distribution with $c-1$ degree of freedom.

A new risk factor R_1' is generated based on the difference between elements in this manner.

Step 4
By performing steps 1 to 3 repeatedly, we generate additional risk factors R_2', R_3', R_4', and then let Θ' denote the group of risk categories given by the direct product of the risk factors.

$$\Theta' = \prod_{i=1}^{4} R_i'$$

The characteristics of the proposed method can be summarized as follows:

- For an insurance product with multiple risk factors, it is possible to identify which risk factors, and in particular which elements, affect the damage rate, without taking the correlations between the risk factors into consideration.
- Since the analysis is performed using aggregated statistics such as the damage rate of each category, computation is easy. Thus, the required amount of computation can be relatively small compared to other methods such as multiple regression analysis.

4. Numerical experiment

Two risk factors in the model case, R_1 = {Northern Japan, Eastern Japan, Western Japan, Southern Japan} and R_2 = {Sedan, Sports Car, RV} are extracted to generate risk categories using the numeric values in the table below[7]. The percentages in the table show the damage rate in each category.

	N. Japan	E. Japan	W. Japan	S. Japan
Sedan	3%	18%	5%	6%
Sports Car	6%	25%	4%	8%
RV	4%	10%	1%	2%

---

[6] Normally when multiple comparison technique such as Scheffe's method is used in an analysis of variance, we must provide proof that the test statistics used in the analysis of variance is included in the procedure for performing multiple comparison; otherwise, there may be multiplicity in the test.

[7] Numerical values in the table have been prepared for this experiment, and do not reflect the portfolio of our company.

Computations are performed according to the procedure presented in the previous section.

Step 1
Letting Block 1 = Sedan, Block 2 = Sports Car, Block 3 = RV, and assigning the order of the damage rate per block yields the following table:

	N. Japan	E. Japan	W. Japan	S. Japan
Sedan	1	4	2	3
Sports Car	2	4	1	3
RV	3	4	1	2

Step 2
A statistical analysis using Friedman's test yields $\chi_0^2 = 7$. If the significance threshold is set to 90%, then the null hypothesis is rejected, since $\chi_3^2(0.10) = 6.25 < 7$.

Step 3
Using Scheffe's method, a comparison is made to determine whether there are differences between the geographic regions. Letting Northern Japan = 1, Eastern Japan = 2, Western Japan = 3, and Southern Japan = 4 (e.g., S_{12} represents the difference between Northern Japan and Eastern Japan), a statistical analysis yields $S_{12} = 3.6, S_{13} = S_{14} = 0.4, S_{23} = 6.4, S_{24} = S_{34} = 1.6$. Using the same rejection criteria as above, we conclude that there is a significant difference between Eastern and Western Japan, but not between the others.

From the above, the following two observations can be made[8] (since it cannot be determined whether Northern and Southern Japan should be included as separate risk categories, this decision should be made via judgment):

1) Geographic regions should be considered as a risk factor
2) Eastern Japan and Western Japan should be considered as separate risk categories

Next, perform a similar test using vehicle type.

	Sedan	Sports Car	RV
N. Japan	1	3	2
E. Japan	2	3	1
W. Japan	3	2	1
S. Japan	2	3	1

As a result, the null hypothesis cannot be rejected since $\chi_2^2(0.10) = 4.61 > \chi_0^2 = 4.5$, but there is no significant difference between the vehicle types. Thus, vehicle type does not need to be considered as a risk factor in this example.

---

[8] Strictly speaking, to conclude that "(a) and (b) should be actively supported" is an overstatement.

Step 4
The conclusion is that only geographic region needs to be considered as a risk factor, and the following final risk categories can be utilized[9]:

$$\Theta' = \{\{\text{Northern/Southern Japan}\}, \{\text{Eastern Japan}\}, \{\text{Western Japan}\}\}$$

5. Proposition for a test that expands on the use of an ordinal scale

We generated risk categories using Friedman's test and Scheffe's method, but there are issues with this method since it utilizes an ordinal scale as the measure:

- Since only an ordinal scale is used, the actual differences in damage rate between the elements are not reflected in the calculation of statistics. In other words, a difference in damage rate is reflected only as a difference on the ordinal scale.
- The number of data (exposure frequency) is not reflected in the outcome. Reliabilities should differ for the statements, "this category showed the lowest damage rate as a result of a statistical analysis based on 100 samples" and, "this category showed the lowest damage rate as a result of a statistical analysis based on 10 samples."

(1) Expansion of Friedman's test

The problem with the method described in the previous section occurred because only the ordinal scale was used for the tests. Accordingly, the concept of an ordinal scale was expanded as follows. In the table below, l_{ij} represents the damage rate calculated from numerical data n_{ij}.

	1	2	……	c
Block 1	l_{11}, n_{11}	l_{12}, n_{12}	:	l_{1c}, n_{1c}
Block 2	l_{21}, n_{21}	l_{22}, n_{22}	:	l_{2c}, n_{2c}
:	:	:	:	:
Block r	l_{r1}, n_{r1}	l_{r2}, n_{r2}	:	l_{rc}, n_{rc}

Whereas in the method of the previous section, $l_{11}, l_{12}, \cdots, l_{1c}$ were sorted in ascending order and ranked from 1 to c, this scheme is now replaced with the expected value of the ordinal scale. Here, let us assume[10] that the damage rate is a random variable X_{ij} which has a distribution characterized by an expected value of l_{ij}, a variance of n_{ij}, and let the expected value of order $\hat{O}_{ij}$ be its final order O_{ij}. Thus:

$$(O_{11}, O_{12}, \cdots, O_{1c}) = \sum_A P(\hat{O}_{11} = o_1, \hat{O}_{12} = o_2, \cdots, \hat{O}_{1c} = o_c)(o_1, o_2, \cdots, o_c)$$

---

[9] All that is strictly required is that Eastern Japan and Western Japan are in different risk categories, and the above Θ' is not the sole possibility.

[10] For example, a model with normal distribution $X_{ij} \sim N(l_{ij}, \sigma / \sqrt{n_{ij}})$ may be conceivable; however, realistically speaking, calculating the expected value is difficult even if normal distribution is assumed.

In the above equation, A represents the set of all ordinals that $\hat{O}_{ic}$ could take. However, computation is difficult even if a normal distribution is assumed for X_{ij}, and additional refinement must be made regarding the assumptions and computation, in order to use this method in practice.

(2) Test statistics

Test statistics for Friedman's test were calculated by assuming that O_{ij} adopted an integer value between 1 and c, uniformly. By changing this assumption to $O_{ij} \sim U(1,c)$, χ_0^2 can be obtained using the following equation:

$$\chi_0^2 = \frac{12r}{c(c-1)} \sum_{j=1}^{r} \left(K_j - \frac{c+1}{2} \right)^2 = \frac{12}{rc(c-1)} \sum_{j=1}^{c} K_j^2 - \frac{3r(c+1)^2}{c-1}$$

where χ_0^2 follows a χ^2 distribution with $c-1$ degrees of freedom. In addition, the statistics from Scheffe's method do not change, even if we let $O_{ij} \sim U(1,c)$.

Chapter 3 Usefulness of the hierarchical risk model and future challenges

We can summarize the usefulness of the proposed “hierarchical risk model” described in the above chapters as follows:

- The evaluation of a portfolio can be segmented to lower models, while performing evaluation using the upper model which is the most stable outcome of analysis. Unlike Buhlmann’s model, where segmentation can only be performed based on the expected value, the hierarchical risk model can segment probability distribution; therefore, it can be widely applied,
- The hierarchical risk model will facilitate a top-down corporate strategy, and it is easy to control the risk management process.

In addition, while determining risk categories is a common critical task in the non-life insurance industry, few publications have appeared which summarize the methodologies for it. Under such circumstances, this paper is meaningful in having introduced and overviewed the commonly used methodologies and their limitations, introduced Friedman’s test which is not often employed in the non-life insurance industry, and suggested the future development of the methodology.

On the other hand, we recognize the following challenges in order to further expand this methodology.

- There is a room for further improvement so that the risk measure of the upper model is well maintained when lower models derived by segmenting the upper model are again aggregated considering the portfolio.
- In some cases, the method we propose in the chapter 1 underestimates risk measures such as VaR and T-VaR rather than the method of the maximum likelihood estimation. So, when we use in actual business, some adjustments are required.
- Since the lower models estimated from the upper model will have discrete distributions, further studies are required in order to optimize the definition of the segments.
- The method of setting preconditions and calculating in the expanded Friedman’s test must be further enhanced to stand up to actual business use.

In this paper, we have adopted an approach that is in opposition to ordinary methods. We hope that this paper will not only help to develop the methodology, but also to assist in stimulating discussions.

References

「階層ベイズモデルとその周辺　時系列・画像・認知への応用」石黒真木夫　松本隆　乾敏郎　田邉國士　著　岩波書店
「統計的多重比較法の基礎」　永田靖　吉田道弘　著　サイエンティスト社
「多変量解析の展開　隠れた構造と因果を推理する」　甘利俊一　狩野裕　佐藤俊哉　松山裕　竹内啓　石黒真木夫　著　岩波書店
「確率統計演習 2　統計」　国沢清典編　培風館
「モデリング」　社団法人　日本アクチュアリー会
「損保数理」　社団法人　日本アクチュアリー会
Myles Hollander and Douglas A. Wolfe, “Nonparametric Statistical Methods Second Edition”,

Wiley-Interscience

Stuart A. Klugman, Harry H. Panjer and Gordon E. Willmot, "Loss Models From Data to Decision" John Wiley & Sons Inc.

“http://aoki2.si.gunma-u.ac.jp/lecture/TwoWayANOVA/friedman.html”