Bodily Injury Liability
Loss Model & Credibility
(In Korea insurance Case)

Heung ki, Jun

Chief Actuary, Hanwha Non-Life Insurance

Index

Study of Loss Model and Credibility about Bodily Injury Liability

Chapter 1. Introduction

1.1 Introduction and Purpose

Korean casualty industry has been for the last 80 years. During the period, we have evaluated the operating profit of insurance companies with risk management system, premium, and paid benefit estimating loss ratio by the method of Incurred to Earned Basis.

However, now more scientific and technical methods of estimating risk are required to be prepared to the dramatic growth of finance of insurance in stable economic growth, specially globalization of the world economy, FTA, and Cash Flow Pricing according to the introduction of international GAAP (Generally Accepted Accounting Principles)

In a view of actuarial science, liability has a special character of Long Tail Business because of factors such as compensation for damage based on the civil law and so on. Therefore, it is hard to estimate loss of liability.

That is, compensation periods after accidents are very different according to the item of insurance. There are few delay of payments in life insurance because the amount is usually fixed before. However, delay of payment occurs often in casualty because the amount is actual loss as the rule of insurable interests. For property and injury insurance, benefits are paid right after the report of accident because it is not difficult to estimate the loss. However, in liability, sometimes the loss is not reported for a long time or it may take more than 5 years to estimate and compensate the loss even though the it is reported.

Estimating correct loss is not easy because it is a prospect of the future trend of loss. To solve this problem, actuarial method is used significantly because it has many potential difficulties to estimate loss individually that will be settled for the next few years.

1.2 Study of models

The aim of this study is making various probability models with accidents and loss of insurance first. After then, we will see that those models follow distributions of population parameters. We will also calculate estimations of population parameters.

We will find a method to estimate loss properly and reasonably by applying actual data of frequency and severity distribution to loss models. Through this process, we make it possible to forecast the operating achievements and the prospect of casualty covering Outstanding Reserve, Loss Ratio, Pricing Method.

To do this, we will set up models of loss and probability distribution function by studying the situation of loss occurrence and characters of liability.

We will find out how accurate estimations of loss are, and also how close they are to unknown parameters of future loss by modeling and finding frequency, severity, and other related variables.

Thus, we applied and tested statistical models as Credibility to verify the consistency.

Chapter 2. Computing the amount of loss of liability

Liability has its unique standards about guarantee. Those are Occurrence-Basis Insurance and Claims-Made Basis Insurance.

Some factors such as reserve and hazard ratio are calculated considering the character of risk and credibility of data by a statistical method that can take the trend of loss, change of income, the trend of economic performance, the innovation of technology and so on into consideration because the risk of concerned year is affected much by previous experience of accidents.

The insurer can protect the insured and strive for the stability and growth of insurance industry by appropriating the amount that will be paid for occurred accidents as reserve.

Actual results of Casualty, Liability, Car Liability I in Korea

(unit: number, ₩1,000)

Year	Casualty		Liability			Car Liability I		
	Number	Premium	Number	Premium	Portion	Number	Premium	Portion
2000	40,446,507	16,018,733,834	222,086	314,436,150	1.96%	11,628,810	2,162,885,155	13.50%
2001	43,445,558	17,853,696,818	240,965	339,918,979	1.90%	12,391,010	2,609,539,997	14.62%
2002	49,573,410	19,985,172,215	272,085	438,486,852	2.19%	13,313,951	2,640,984,337	13.21%
2003	51,077,501	20,716,061,311	296,696	466,303,394	2.25%	13,849,882	2,621,329,019	12.65%
2004	59,618,155	22,566,311,406	303,646	486,307,540	2.16%	14,109,602	2,852,138,646	12.64%
2005	61,825,125	24,865,683,891	313,540	451,360,889	1.82%	14,503,169	2,950,587,886	11.87%

*Portion: Portion of premium of Liability and Car Liability I compared to Casualty

(Korean Insurance Development Institute 2006)

Chapter 3. The bodily Injury Liability Loss Model

3.1 Summary of Loss analysis

3.1.1 Observation period

Observation period is not settled by law or any stipulated regulation but it can be settled properly by actuaries considering the stability of loss ratio of each insurance item. For the bodily liability, the period needs to be as long as possible to level off the change between years. On the other hand, for the property liability, a short period can be better to reflect recent trend of frequency or loss ratio because it is simple to decide the loss amount.

The table below shows us that we need a observation period more than 10 years in Japan, more than 5 years in Korea for the bodily liability.

Liability in Japan

(unit: persons, $)

From Incurred date to Claim Settlement	Number of claims	Portion of claims	Portion of paid amount	Average paid amount
~ 1 year	33	6.8%	3.6%	253,942
1 year ~ 2 years	100	20.6%	20.3%	467,379
2 years ~ 3 years	122	25.1%	24.4%	461,896
3 years ~ 4 years	81	16.7%	19.0%	541,123
4 years ~ 5 years	73	15.0%	12.9%	408,964
5 years ~ 10 years	72	14.8%	19.7%	630,158
10 years ~	5	1.0%	0.1%	32,790
Total	486	100.0%	100.0%	

* Source: Liability Insurance Report No.1 (Nippon fire: 1987)

Liability in Korea (FY2000~FY2006)

(unit: persons, ₩1,000,000)

From Incurred date to Claim Settlement	Number of claims	Portion of claims	Portion of paid amount	Average paid amount
~ 1 year	753,283	83.2%	63.7%	498,051
1 year ~ 2 years	126,252	13.9%	25.3%	197,929
2 years ~ 3 years	17,652	1.9%	6.6%	51,257
3 years ~ 4 years	5,699	0.6%	3.2%	25,325
4 years ~ 5 years	1,788	0.2%	0.7%	5,336
5 years ~	842	0.1%	0.5%	4,060
Total	905,516	100.0%	100.0%	

3.1.2 Trend of frequency and severity of accident

We use trend factors to manipulate the past statistical data to reflect changes of loss that are regarded to increase until new rates are applied. The existing trend factors were made with the idea that trend of frequency or severity for previous years will continue in the predictable future.

For example, we can expect that the annual average loss, a trend factor, will increase by 159,000 Won if the average annual increase of loss was 159,000 Won for the last few years.

We also can expect that the frequency will increase by 2/1000 if the frequency increased by 2/1000 for the last few years.

If the average loss or frequency increases or decreases monotonously such as trend factors above, we call it linear trend factor.

Thus, the future average loss or frequency of accident is estimated with a formula below.

$$Y_t = a + b \cdot t$$

a and b are constant, Y is the average loss of frequency of accident, t is a future year of prediction.

a and b are computed by linear regression.

Also, exponential trend factor is used when the average loss or frequency of accident increases regularly by the same rate every year.

Thus, the change of average amount of loss (or number of accident with 1000) is constant when the factors are linear trend factor but those factors increase as time goes by when the factors are exponential trend factor.

The future average loss or frequency of accident is estimated with a formula below.

$$Y_t = a \cdot b^t$$

a and b are constant, Y and t are the same as those in the formula of linear trend factors. If we take natural logarithms at both sides of the formula,

$$\ln Y_t = \ln a + \ln b \cdot t$$

then, a and b can be estimated by linear regression.

We can decide from the trend of actual data which trend curve is better between linear curve and exponential curve. We also can use other curves though those curves are not used often currently.

3.1.3 Unit for the measurement of loss

Premium rates are set to apply to the concerned year because the amount of loss is used to calculate premium. Thus, we need to examine factors that consist of individual risk(Loss Exposure).

The amount of loss per victim has a tendency to increase automatically interlocking with inflation.

Increase rate of liability in Japan

Comparing year	Increase rate		
	Consumer Price Index	Average benefit per claimant	Average benefit per accident
1985 vs 1979	46.3%	52.3%	52.0%
1985 vs 1975	97.6%	150.5%	99.4%

Source: Liability Insurance Report No.1 (Nippon fire: 1987)

3.2 Loss Model Flow

3.2.1 Sampling and Modeling

I tested a data of the loss of bodily injury liability occurred during past five years in Korea by investigating 251,030 victims with the Stratified random sampling.

Number of victims of sample

Accident year	2002	2003	2004	2005	2006
Number of victims	43,787	52,559	51,908	54,673	48,103

Modeling process is executed through 6 steps below.

Step 1 - Choose one or more models by the analyst's knowledge, experience or the character and type of available data. For example, in a study about the amount of loss, a model includes various related factors such as incidence, the distribution of accidents, some kinds of life related variables. In case of a study about loss ratio of insurance, every kind of statistical distributions(for instance: Log distribution, Gamma distribution, Weibull distribution) can be chosen. (MODEL Choice)

Step 2 - Models are manipulated based on available data. For example, when we study death rate, information about life insurance policy is available. For a study about the conflict of ownership, information about actual loss of insurance is available data. (MODEL Calibration)

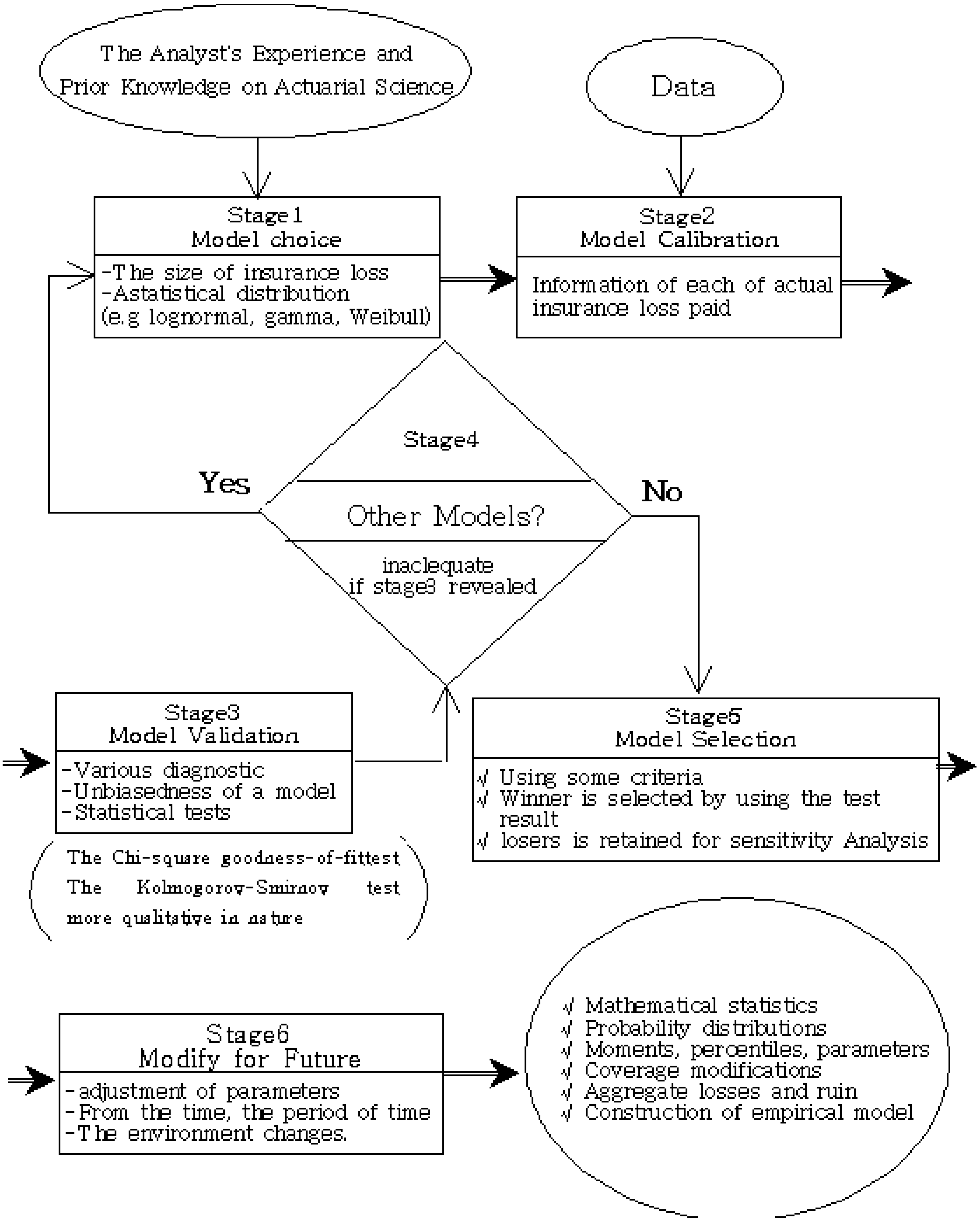
Step 3 - Suitable models are effective to verify the consistency of data. Various diagnosis tests are used. There are a few well-known tests such as : Chi-Square Goodness-of-Fit test, Kolmogorov-Smirnov test. We can choose Credibility as a more qualitative method. Test choice is related with the ultimate purpose of modeling directly. Suitable models are required to reflect the total loss acquired from data occurred actually for a study of insurance. It is called the fairness of model in insurance. (MODEL Validation)

Step 4 - We should consider any possibility of existing of other possible models. If we concluded in the step3 that every model is not suitable, this step4 is more useful specially. In this step, we should consider the possibility that there are available models more than one.

Step 5 - MODEL Selection: Compare every available models considered through step1 to step4 to use standards chosen between those models. Do step5 using preacquired test results or another standard. First, select good methods and keep other methods for the sensitivity analysis.

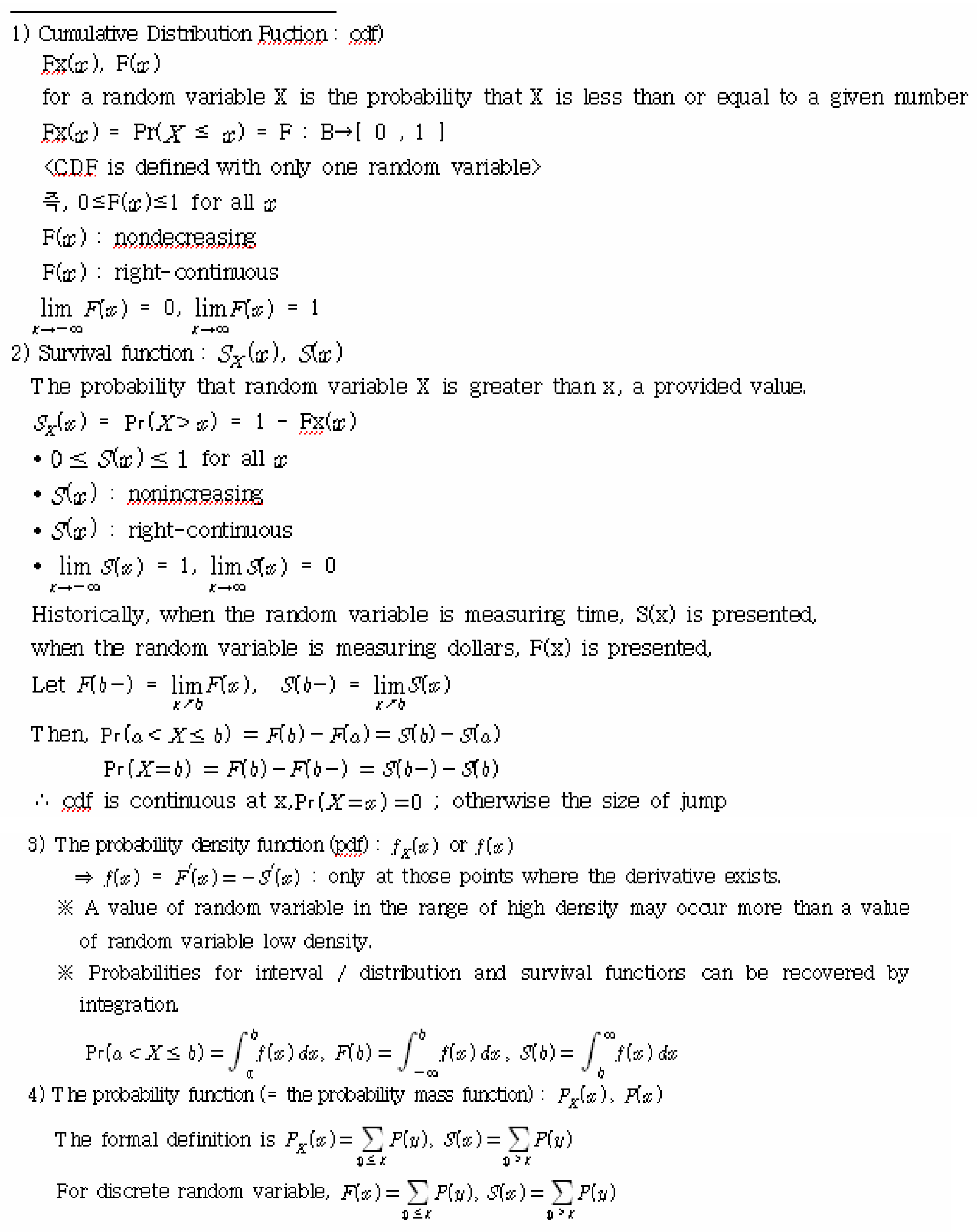
Lastly, apply chosen models to the future. This step includes adjustment of related variables by reflecting the expected inflation to apply to the future point where models are applied to.

We need to do step6 repeatedly to improve models if we collect new data or the circumstance change.



2. Design of Modeling Function

We can design a model of frequency and severity of accident by measuring X and x($X=x$, X:random variable, x:value of X) with cumulative distribution function $F(x)$[1], Survival function $S_X(x)$[2], probability density function $f(x)$[3], probability mass function $P(x)$[4].



1) Cumulative Distribution Fuction : cdf)

$Fx(x)$, $F(x)$

for a random variable X is the probability that X is less than or equal to a given number

$Fx(x) = \Pr(X \le x) = F : B \rightarrow [\,0\,,1\,]$

<CDF is defined with only one random variable>

즉, $0 \le F(x) \le 1$ for all x

$F(x)$: nondecreasing

$F(x)$: right-continuous

$\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$

2) Survival function : $S_X(x)$, $S(x)$

The probability that random variable X is greater than x, a provided value.

$S_X(x) = \Pr(X > x) = 1 - Fx(x)$

- $0 \le S(x) \le 1$ for all x
- $S(x)$: nonincreasing
- $S(x)$: right-continuous
- $\lim_{x \to -\infty} S(x) = 1$, $\lim_{x \to \infty} S(x) = 0$

Historically, when the random variable is measuring time, S(x) is presented, when the random variable is measuring dollars, F(x) is presented.

Let $F(b-) = \lim_{x \nearrow b} F(x)$, $S(b-) = \lim_{x \nearrow b} S(x)$

Then, $\Pr(a < X \le b) = F(b) - F(a) = S(b) - S(a)$

$\Pr(X = b) = F(b) - F(b-) = S(b-) - S(b)$

∴ cdf is continuous at x, $\Pr(X = x) = 0$; otherwise the size of jump

3) The probability density function (pdf) : $f_X(x)$ or $f(x)$

$\Rightarrow f(x) = F'(x) = -S'(x)$: only at those points where the derivative exists.

※ A value of random variable in the range of high density may occur more than a value of random variable low density.

※ Probabilities for interval / distribution and survival functions can be recovered by integration.

$$\Pr(a < X \le b) = \int_a^b f(x)\,dx,\ F(b) = \int_{-\infty}^b f(x)\,dx,\ S(b) = \int_b^{\infty} f(x)\,dx$$

4) The probability function (= the probability mass function) : $P_X(x)$, $P(x)$

The formal definition is $P_X(x) = \sum_{y \le x} P(y)$, $S(x) = \sum_{y > x} P(y)$

For discrete random variable, $F(x) = \sum_{y \le x} P(y)$, $S(x) = \sum_{y > x} P(y)$

3.2.3 Application of actual data

We will find out that a probability model follow a distribution of population parameters by applying the probability model to Loss Model and applying actual loss data of bodily injury liability to the probability model. After then, we will estimate the distribution of the amount of loss and the total loss about population parameters.

Chapter 4. Loss Model and Testing of Actual Data

The first table shows us actual number of victims and amount of paid benefit of bodily injury liability per year in Korea. The second table is about numbers of victims per grade of accident.

Number of victims and paid benefit per accident year

(unit: persons, ₩1,000)

Accident year	Victims	Paid benefit per development year					O/S at the end of 2006
		1 year	2 years	3years	4 years	5 years	
2002	43,787	112,482,110	33,919,350	6,660,378	3,223,866	1,170,494	291,926
2003	52,559	129,919,382	39,493,238	7,514,018	3,678,064		1,436,986
2004	51,908	122,233,004	34,345,902	8,264,166			3,022,294
2005	54,673	134,733,508	49,163,578				10,166,996
2006	48,103	134,431,854					27,990,598

Number of victims per accident year

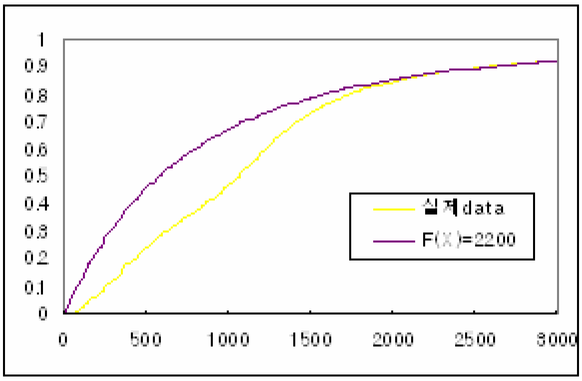
(unit: persons)

Grade of injury	Accident year			The average
	2004	2005	2006	
Death	170	175	92	146
Grade 1	243	208	81	177
Grade 2	190	165	95	150
Grade 3	137	108	60	102
Grade 4	184	153	104	147
Grade 5	472	463	274	403
Grade 6	253	198	136	196
Grade 7	520	438	302	420
Grade 8	12,914	13,098	9,524	11,845
Grade 9	28,226	28,802	23,136	26,721
Grade 10	360	363	291	338
Grade 11	593	537	349	493
Grade 12	550	502	369	474
Grade 13	1,553	2,932	3,874	2,786
Grade 14	5,543	6,531	9,416	7,163
Total	51,908	54,673	48,103	51,561

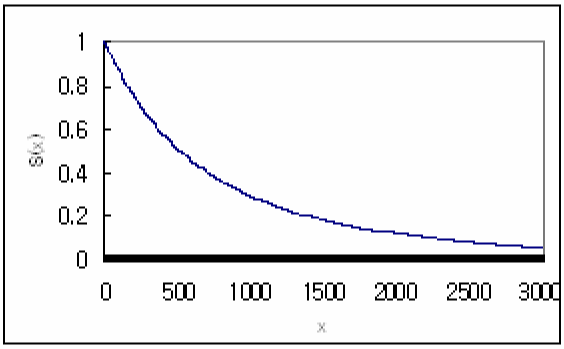
4.1 Probability model

Following graphs are probability models of the loss amount of bodily injury liability from the data above.

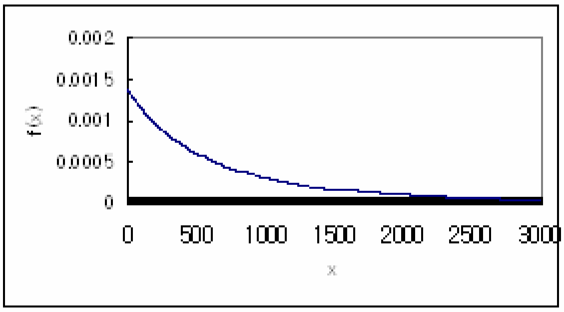
$$F_1(x) = \begin{cases} 0 & , x < 0 \\ 1 - \left(\dfrac{2200}{x+2200}\right)^3 & , x \geq 0 \end{cases}$$



$$S_1(x) = \left(\frac{2200}{x+2200}\right)^3 \quad , x \geq 0$$

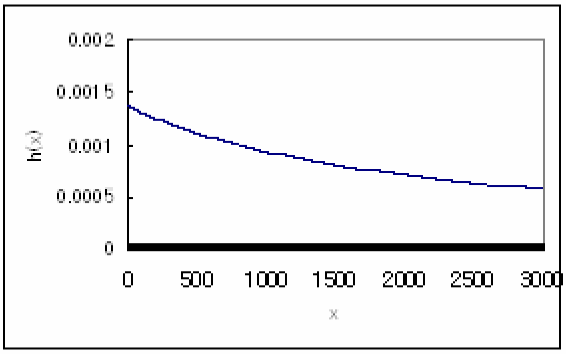


$$f_1(x) = \frac{3(2200)^3}{(x+2200)^4} \quad , x > 0$$



$p_1(x)$ is not defined

$$h_1(x) = \frac{3}{x + 2200} \quad , x > 0$$



Distributions of the insured of bodily injury liability are as below.

Discrete
(The random variable places probability only at 0,1,2,3,4(the support))

$$F_2(x) = \begin{cases} 0 & , x < 0 \\ 0.004 & , 0 \le x < 1 \\ 0.009 & , 1 \le x < 2 \\ 0.013 & , 2 \le x < 3 \\ 0.016 & , 3 \le x < 4 \\ 0.019 & , 0 \le x < 5 \\ 0.029 & , 1 \le x < 6 \\ 0.033 & , 2 \le x < 7 \\ 0.042 & , 3 \le x < 8 \\ 0.284 & , 0 \le x < 9 \\ 0.816 & , 1 \le x < 10 \\ 0.824 & , 2 \le x < 11 \\ 0.833 & , 3 \le x < 12 \\ 0.843 & , 2 \le x < 13 \\ 0.889 & , 3 \le x < 14 \\ 1 & , x \ge 14 \end{cases} , \quad S_2(x) = \begin{cases} 0 & , x < 0 \\ 0.996 & , 0 \le x < 1 \\ 0.991 & , 1 \le x < 2 \\ 0.987 & , 2 \le x < 3 \\ 0.984 & , 3 \le x < 4 \\ 0.981 & , 0 \le x < 5 \\ 0.971 & , 1 \le x < 6 \\ 0.967 & , 2 \le x < 7 \\ 0.958 & , 3 \le x < 8 \\ 0.716 & , 0 \le x < 9 \\ 0.284 & , 1 \le x < 10 \\ 0.176 & , 2 \le x < 11 \\ 0.167 & , 3 \le x < 12 \\ 0.157 & , 2 \le x < 13 \\ 0.111 & , 3 \le x < 14 \\ 1 & , x \ge 14 \end{cases} , \quad p_2(x) = \begin{cases} 0.004 & , x = 0 \\ 0.005 & , x = 1 \\ 0.004 & , x = 2 \\ 0.003 & , x = 3 \\ 0.003 & , x = 4 \\ 0.010 & , x = 5 \\ 0.004 & , x = 6 \\ 0.009 & , x = 7 \\ 0.242 & , x = 8 \\ 0.532 & , x = 9 \\ 0.008 & , x = 10 \\ 0.009 & , x = 11 \\ 0.010 & , x = 12 \\ 0.046 & , x = 13 \\ 0.111 & , x = 14 \end{cases}$$

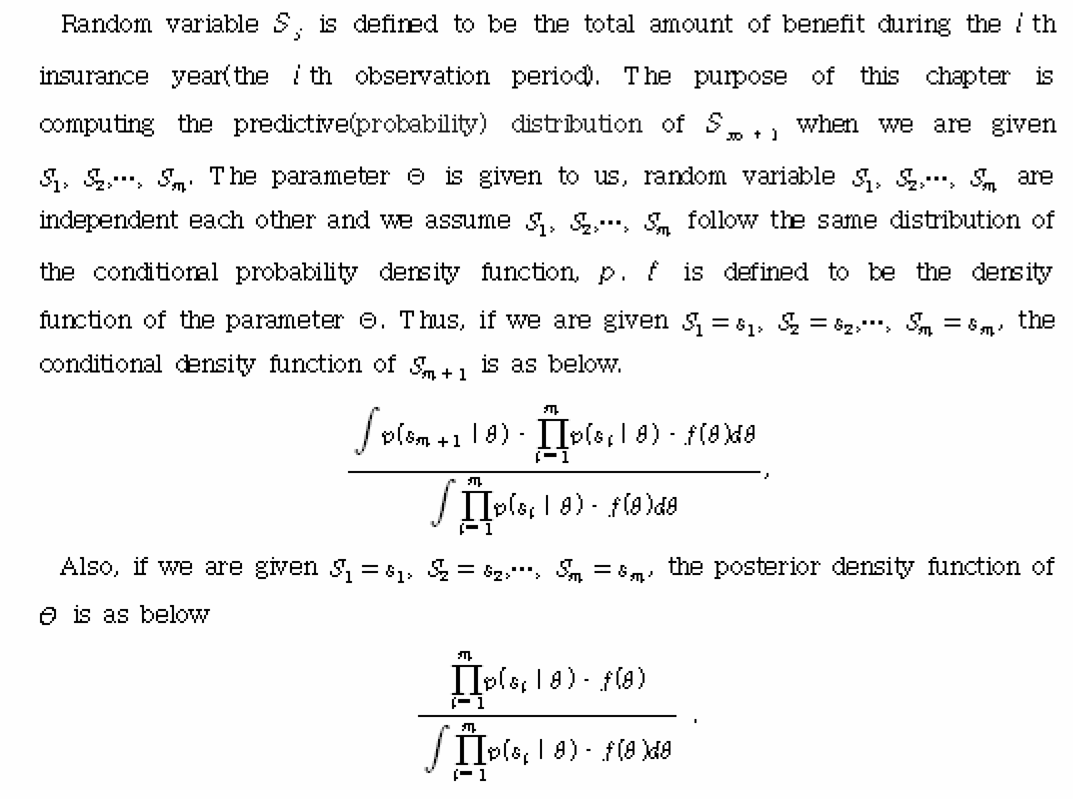
$f_2(x)$ is not defined

$h_2(x)$ is not defined

Chapter 5. Credibility of Bodily injury liability

In this chapter, we will see another model of 2 steps about loss of bodily injury liability. The analysis of frequency of benefit payment is based on Poisson distribution. The analysis of benefit amount is based on Exponential distribution. Let's assume that the number of benefit payments and individual benefit amount are independent. After seeing predictive distributions and posterior distributions, We will make models from distributions of the number of benefit payments and the benefit amount shown in chapter 3.

5.1 Predictive distribution



Random variable S_i is defined to be the total amount of benefit during the i th insurance year(the i th observation period). The purpose of this chapter is computing the predictive(probability) distribution of S_{m+1} when we are given $S_1, S_2, \cdots, S_m$. The parameter Θ is given to us, random variable $S_1, S_2, \cdots, S_m$ are independent each other and we assume $S_1, S_2, \cdots, S_m$ follow the same distribution of the conditional probability density function, p. f is defined to be the density function of the parameter Θ. Thus, if we are given $S_1 = s_1, S_2 = s_2, \cdots, S_m = s_m$, the conditional density function of S_{m+1} is as below.

$$\frac{\int p(s_{m+1} \mid \theta) \cdot \prod_{i=1}^{m} p(s_i \mid \theta) \cdot f(\theta) d\theta}{\int \prod_{i=1}^{m} p(s_i \mid \theta) \cdot f(\theta) d\theta},$$

Also, if we are given $S_1 = s_1, S_2 = s_2, \cdots, S_m = s_m$, the posterior density function of Θ is as below

$$\frac{\prod_{i=1}^{m} p(s_i \mid \theta) \cdot f(\theta)}{\int \prod_{i=1}^{m} p(s_i \mid \theta) \cdot f(\theta) d\theta}.$$

5.2 Analysis of number of claims

Random variable N_i $(i = 1, 2, \cdots)$ is defined to be the number of benefit payment during the i th insurance period. Let's assume that N_i follows Poisson distribution with parameter(mean) Λ. If we are given m values of observation, the posterior distribution of Λ is. Parameter α and β decide the prior Gamma distribution. Parameter m and $m\bar{n} = \sum_{i=1}^{m} n_i$ are summarized by data. $g(\lambda)$ is defined to be the density function of $G(\alpha + m\bar{n}, \beta + m)$. Therefore, if we are given $N_1 = n_1, N_2 = n_2, \cdots, N_m = n_m$, the conditional probability distribution of $N_{m+1} = n$ is as below. if $n = 0, 1, \cdots$

$$\frac{\int_0^\infty \frac{e^{-\lambda}\lambda^m}{n!} \cdot g(\lambda)d\lambda}{\int_0^\infty g(\lambda)d\lambda} = \frac{\int_0^\infty \frac{e^{-\lambda}\lambda^m}{n!} \cdot e^{-(\beta+m)\lambda}\lambda^{\alpha+m\bar{n}-1}d\lambda}{\int_0^\infty e^{-(\beta+m)\lambda}\lambda^{\alpha+m\bar{n}-1}d\lambda}$$

$$= \frac{\frac{1}{n!}\int_0^\infty e^{-(\beta+m+1)\lambda}\lambda^{\alpha+m\bar{n}-1+n}d\lambda}{\Gamma(\alpha+m\bar{n}) \cdot (\beta+m)^{-(\alpha+m\bar{n})}}$$

$$= \frac{\Gamma(\alpha+m\bar{n}+n) \cdot (\beta+m+1)^{-(\alpha+m\bar{n}+n)}}{n! \cdot \Gamma(\alpha+m\bar{n}) \cdot (\beta+m)^{-(\alpha+m\bar{n})}}$$

$$= \binom{\alpha+m\bar{n}+n-1}{n}\left(\frac{1}{\beta+m+1}\right)^{n}\left(\frac{\beta+m}{\beta+m+1}\right)^{\alpha+m\bar{n}}$$

The formula above has the form of the negative binomial density function.

The mean of a negative binomial density function is $\frac{\alpha+m\bar{n}}{\beta+m}$ which is equal to the estimation of Buhlmann's credibility about the number of benefit payment. The density function of a negative binomial distribution shows the predictive density function of the number of benefit payment during the $(m+1)$ th insurance period, if we are given $N_1 = n_1, N_2 = n_2, \cdots N_m = n_m$. $n = n_{m+1}$ has values of 0, 1, 2, ...

This predictive density function provides (predictive) probability for each available benefit payment and more information than the predicted mean gives.

Now, let's apply the predictive density function to actual data

The number of claims of bodily injury liability per year

(unit : claim)

Year	2003	2004	2005	2006
Claims	51,518	50,909	49,552	48,775

Applying actual data to a negative binomial distribution, the probability that claims are requested more than 50,000 in 2007 is

☞ $p[N_5 \geqq 50{,}000 \mid N_1 = 51{,}518,\ N_2 = 50{,}909,\ N_3 = 49{,}552,\ N_4 = 48{,}775]$

$m = 4$ and $m\bar{n} = \sum_{i=1}^{4} n_i$ = 51,518 + 50,909 + 49,552 + 48,775 = 200,754

Using a negative binomial conditional probability, then we have

$$\sum_{n=5}^{\infty} \binom{\alpha + 200{,}754 + n - 1}{n} \left(\frac{1}{\beta+4+1}\right)^{n} \left(\frac{\beta+4}{\beta+4+1}\right)^{\alpha+200{,}754}$$

$$= \sum_{n=5}^{\infty} \binom{\alpha + n + 200{,}753}{n} \left(\frac{1}{\beta+5}\right)^{n} \left(\frac{\beta+4}{\beta+5}\right)^{\alpha+200{,}754}$$

$$= 1 - \sum_{n=0}^{5} \binom{\alpha + n + 200{,}753}{n} \left(\frac{1}{\beta+5}\right)^{n} \left(\frac{\beta+4}{\beta+5}\right)^{\alpha+200{,}754}$$

5.3 Analysis of benefit

If X, each amount of individual payment, is a exponentially distributed random variable with Δ(mean) and $x > 0$, $\Delta > 0$, then we assume that we have a density function as below.

$$p(x \mid \Delta) = \frac{e^{-x/\Delta}}{\Delta}$$

If X is a exponentially distributed random variable with parameter Δ, $E_X[X \mid \Delta] = \Delta$ and $Var_X(X \mid \Delta) = \Delta^2$. The mean '$\Delta$' follows a joint prior distribution. If $y' > 0$, $m' > 3$ and $\sigma > 0$, then the probability density function is proportioned to $\frac{e^{-y/\sigma}}{\delta^{m}}$. This density function is called a inverse gamma density function and its mean is $\frac{y'}{m' - 2}$. In this part, the main interest is the process that n_i, the number of benefit payment, is claimed during the i th insurance period for the m insurance period. The total amount of benefit payment about observed values during m insurance period is

$$y = \sum_{i=1}^{m\bar{n}} x_i$$

Therefore, the posterior density function of Δ, $f(\delta \mid m', y', m\bar{n}, y)$, is proportioned to

$$\left(\prod_{i=1}^{mn} \frac{e^{-x_i/\delta}}{\delta}\right)\left(\frac{e^{-y/\sigma}}{\delta^{m}}\right) = \frac{e^{-(y+y)/\delta}}{\delta^{m+mn}}$$

This function is also a inverse gamma density function. The mean of of Δ, Z, can be expressed as below.

$$\frac{y'+y}{m'+m\bar{n}-2} = \left(\frac{m'-2}{m'+m\bar{n}-2}\right)\left(\frac{y'}{m'-2}\right) + \left(\frac{m\bar{n}}{m'+m\bar{n}-2}\right)\left(\frac{y}{m\bar{n}}\right)$$

$$= (1-Z)(prior\,mean) + Z(sample\,mean)$$

We can understand that the mean of the predictive distribution of X is equal to the mean of the posterior distribution of Δ because $E_X[X \mid \Delta] = \Delta$.

The 예측밀도함수 of X reflects the ambiguousness of benefit amount as well as estimation of parameters. The function is

$$p(x \mid m', y', m\bar{n}, y) = \int_0^{\infty} p(x \mid \delta) \cdot f(\delta \mid m', y', m\bar{n}, y)\,d\delta$$

$$= C\int_0^{\infty} \left(\frac{e^{-x/\delta}}{\delta}\right)\left(\frac{e^{-(y+y)/\delta}}{\delta^{m+mn}}\right) d\delta$$

$$p(x \mid m', y', m\bar{n}, y) = C\int_0^\infty e^{-(x+y+y)/\delta}\delta^{-m-mn-1}d\delta$$

This function is a density function belong to the family of Pareto.

During the first four periods, 200,754 claims are requested and the total amount of payments is 401,508,000,000 won. The probability that a claim amount is more that the average benefit amount(2,000,000 won) is

$$p[x_{200,755} \geqq 200 \mid \sum_{i=1}^{200,754} X_i = 40,150,800]$$

If the value of the parameter Δ is uncertain, comparatively a gentle exponential distribution is transformed to a thick-tailed Pareto distribution.

Chapter 6. Summary and Conclusion

We found out that the loss amount of bodily injury liability is not linear or discrete but exponentially distributed in probability models. The distribution function of the loss size is a constant function of sections.

Credibility is a conditional density function. We also saw that the number of claim is negative binomially distributed and the benefit is Poissonly distributed. Both are under thick-tailed Pareto family.

Thus, any approach based on models should be treated according to the object of study. In actuarial statistics(science), many problems are about to make scientific models which will be used to predict insurance cost in the future.

Models are clear scientific description which is obtained from actuaries' experience and knowledge based on past data. Data is an index when actuaries decide unknown quantities called parameters and select a type of model. Models balance between the clearness and the fitness of available data.

The clearness is measured by numbers of unknown parameters. The fitness for data is measured by the relation of discordance between models and data. Model selection is done based on the balance of two conditions, the fitness and the clearness.

References

Herzog, T.N. An Introduction to Stochastic Simulation by Actex Publications, Inc. 1999

Hewitt, CC. Credibility for Severity. CAS

Klugman, S.A., Panjer.H.H., Willmot, G.E., Loss Models From Data to Decisions by John Wiley. 2004

Patrik, G., Estimating Casualty Insurance Loss Amount Distributions, Proceeding of the Casualty Actuarial Sociery, LXVⅡ, 57-109

Ross, S. Introduction to Probability Models, 8th ed., by Academic Press, 2003

Sundt, B. An Introduction to Non-Life Insurance Mathematics, 5th ed., by University of Manngeim Press, 1999

Credibility Theory Jun Heung Ki, FreedomAcademics, 2006

The Mathematical Non-life Insurance, Japan Actuary Society