UNIT 9 COLLECTIVE INSURANCE AND COMPANY OPERATIONS

Structure

9.0 OBJECTIVES

After going through this unit, you will be able to:

- understand individual and collective risk models of insurance; and
- apply important techniques used in risk modelling.

9.1 INTRODUCTION

Insurance operations relies on premiums to meet the costs (claims). The claim arrivals are random and the insurance portfolios need to be designed to cover the risks. The following discussion gives the operational tools on which the insurance companies rely to take decision.

9.2 ACCOUNTING PRACTICES OF INSURANCE COMPANY

9.2.1 Insurance Company Operation

The goal of insurance modelling is to develop a probability distribution for the total amount paid in return as benefits. Such a strategy helps manage company's capital account and honour its commitments. It becomes necessary, therefore, to understand their overall approach to achieve the basic objective. For that purpose we will take up three aspects, viz.,

1) Accounting practices of an insurance company and reserve
2) Behaviour of the loss characteristics of a groups of similar polices and
3) Other methods of premium setting for a policy.

9.2.2 Net Premium Reserve

To understand the concept of net premium reserves, we need to start with some basic terms used in accounting. Suffice it will be to remember five of these:

1) Asset
2) Liability
3) Equity
4) Reserve and
5) Surplus

Broadly speaking the terms are used to mean the following:

- Asset: Tangible and intangible properties owned by the business
- Liability: Payments owed by the business
- Equity: (Assets – liabilities).
- Reserve: Insurance name of liability
- Surplus: Insurance name of equity

A major source of liabilities in case of insurance as opposed to other business is the payment in return for premiums received from the insured. Therefore an insurance company's strategy to run its business successfully always aims at measuring the liabilities. In general, it is an accepted principle for a company to have

Reserve at time t = (Actuarial present value at time t of future benefits) –(Actuarial present value at time to of future premiums).

In the book of record, the accountant assumes that the premium charged for a policy is net level premium and makes provision for reserves accordingly. For this reason, the liabilities are measured as the net level premium reserve.

To understand the idea of reserve adopted in accounting consider the following example.

Let the income of a company for the year ending in March 2006 is given as follows:

S.No	Premiums	200
1.	Investment income	100
2.	Expenses	105
3.	Claims and maturities	70
4.	Increases in reserve	--
5.	Net income	--

In addition, let the balance sheet of the company is presented as

S.No	Balance sheet as on		
		March 31, 2005	March 31, 2006
1.	Assets	2000	--
2.	Reserve	--	700
3.	Surplus	1400	--

With the information given above, it will be possible for you to fill up the missing entries (--) in the tables. We know that net income is,

total income – total expenditure

So, (200+100) – (105 + 70) = 300 – 175 = 125

From the balance sheet data reserve at the end of 2005

2000 – 1400= 600

The net increase in reserve

= 700 – 600=100.

The net income = 125 – 100 =25.

In 2006, surplus = 1400 +25 = 1425 and asset = 700+1425 =2125.

In case of prospective reserve formula, the expected size of future loss at time K is considered which is based on the evaluation of future performance of the insurance portfolio. To appreciate the underlying idea, consider the example of a fully discrete whole life policy issued to (x). Remember that premium under this scheme is payable annually and is equal to net premium. To compute the reserve at time k, where k is an integer, you need to note the following: Given the underlying accounting assumption, the amount of reserve at any time t can be computed with the help of two methods, viz., ***prospective reserve formula*** and ***retrospective reserve formula***.

Prospective Reserve Formula = (APV of future benefits) – (APV of future assessments), where APV = actuarial present value. A scenario-based on stochastic prospective reserve formula is similar, except that the above formula is calculated for each scenario on stochastically generated variable.

Retrospective Reserve Formula = accumulated value of the net premiums received based on realised return – accumulated value of completed insurance payments.

Since an insurance company operates with a large number of policies at any given time, it has to manage the financial risk involved in such operations. Keeping such a possibility in view, we study the problem of risk analysis in terms of individual and collective risk models. One of our main tasks in the following is to estimate the amount of risk both in terms of its quantification and time of occurrence.

9.3 INDIVIDUAL RISK MODEL

We consider a certain portfolio of insurance policies and the total amount of claims arising from it during a given period (usually a year). The joint premium for the whole portfolio that will cover the accumulated risk connected with all policies will be determined at the beginning.

Let us assume that the portfolio consists of n insurance and the claim (or loss) made in respect of the policy K is denoted by X_k. Then the total, or aggregate, amount of claims (or total risk) is expressed in the form $S = X_1 + X_2 + ... + X_n$, where $X_1, X_2, ..., X_n$ are independent random variables and X_i = loss to the insurer on insured unit i. There are two types of models - closed and open - through which we analyse the probable losses. In case of closed model, it is assumed that the number of insured units n is known and fixed. On the other hand, the insurance system contains provision for inflow and outflow of the insured units in case if open model.

The claim amount variable X_k for each policy is usually presented as $X_k = I_k B_k$ where random variables $I_1, ..., I_n, B_1, ..., B_n$ are independent. The random variable I_k indicates whether or not k^{th} policy produced a payment. If the claim has occurred, then $I_k = 1$, if not, $I_k = 0$. We denote $q_k = P(I_k = 1)$ and $1 - q_k = P(I_k = 0)$. The random variable B_k can have an arbitrary distribution and represents the amount of the payment in respect of the K^{th} policy given that the payment was made.

9.3.1 General Premium Formulae

In order to determine the formulae for premiums, let us assume that the expectations and variances of B_k's exist and denote $\mu_k = E(B_k)$ and $\sigma^2{}_k = Var(B_k), k = 1, 2, ..., n.$. Then

$$E(X_k) = \mu_k q_k, \tag{9.1}$$

and the mean of the total loss in the individual risk model is given by

$$E(S) \sum_{k=1}^{n} \mu_k q_k \tag{9.2}$$

The variance of X_k is

$$Var(X_k) = Var\{E(X_k|I_k)\} + E\{Var(X_k|I_k)\} \tag{9.3}$$

$$= Var\{I_k E(B_k)\} + E\{I_k Var(B_k)\}$$

$$= \{E(B_k)\}^2 Var(I_k) + Var(B_k)E(I_k)$$

$$= \mu^2{}_k q_k(1-q_k) + \sigma^2{}_k q_k.$$

Since it is assumed that X_k's are independent, the variance of S is

$$Var(S) = \sum_{k=1}^{n} \{\mu^2{}_k q_k(1-q_k) + \sigma^2{}_k q_k\}. \tag{9.4}$$

With these distributional information, premiums for the individual risk model are as follows:

Pure risk premium: $P = \sum_{k=1}^{n} \mu_k q_k,$

Premium with safety loading: $P_{SL}(\theta) = (1+\theta)\sum_{k=1}^{n} \mu_k q_k, \ \theta \geq 0,$

Premium with variance loading:

$$P_V(a) = \sum_{k=1}^{n} \mu_k q_k + a\sum_{k=1}^{n} \{\mu^2{}_k q_k(1-q_k) + \sigma^2{}_k q_k\}, \ a \geq 0,$$

Premium with standard deviation loading:

$$P_{SD}(b) = \sum_{k=1}^{n} \mu_k q_k + b\sqrt{\sum_{k=1}^{n} \{\mu^2{}_k q_k(1-q_k) + \sigma^2{}_k q_k\}}, \ b \geq 0.$$

If we assume that for each $k = 1,2,...,n$ the moment generating function $M_{X_k}(t)$ exists, then

$$M_{X_k}(t) = 1 - q_k + q_k M_{B_k}(t), \tag{9.5}$$

and hence $$M_S(t) = \prod_{k=1}^{n} \{1 - q_k + q_k M_{B_k}(t)\}. \tag{9.6}$$

which gives the exponential premium as

$$P_E(c) = \frac{1}{c}\sum_{k=1}^{n} \ln\{1 - q_k + q_k M_{B_k}(c)\}, \ c > 0. \tag{9.7}$$

In the individual risk model, claims of an insurance company are modelled as a sum of the claims of many insured individuals. Therefore, to find the quantile premium given by

$$P_Q(\varepsilon) = F_S^{-1}(1-\varepsilon), \ \varepsilon \in (0,1) \tag{9.8}$$

the distribution of the sum of independent random variable has to be determined.

9.3.2 Premiums in the Case of the Normal Approximation

The distribution of the total claim in the individual risk model can be approximated by means of the central limit theorem. In such a case, it is sufficient to evaluate means and variances of the individual loss random variables, sum them to obtain the mean and variance of the aggregate loss of the insurer and apply the normal approximation. However, it is important to remember that the quality of the approximation depends not only on the size of the portfolio, but also on its homogeneity.

The approximation of the distribution of the total loss S in the individual risk model can be applied to find a simple expression for the quantile premium. If the distribution of S is approximated by a normal distribution with mean E (S) and variance Var (S), the quantile premium can be written as

$$P_Q(\varepsilon) = \sum_{k=1}^{n} \mu_k q_k + \Phi^{-1}(1-\varepsilon)\sqrt{\sum_{k=1}^{n}\left\{\mu^2{}_k q_k (1-q_K) + \sigma^2{}_k q_k\right\}}, \qquad (9.9)$$

where $\varepsilon \in (0,1)$ and $\Phi(.)$ denote the standard normal distribution function. Moreover, in the case of this approximation, it is possible to express the exponential premium as

$$P_E(c) = \sum_{k=1}^{n} \mu_k q_k + \frac{c}{2}\sum_{k=1}^{n}\left\{\mu^2{}_k q_k (1-q_k) + \sigma^2{}_k q_k\right\}, \quad c > 0, \qquad (9.10)$$

and you may see that such a premium is equal to the one with variance loading of $d = \frac{c}{2}$.

As the distribution of S is approximately normal with the same mean and variance, premiums defined in the expected value of the aggregate claims are given by the same formulae as given above.

To derive the probabilistic properties of the random variable S, we take the help of conditioning and obtain the formula for the distribution function of a sum of independent random variables. For that purpose, look at the following example.

Le there be two independent random variables X and Y with exponential distribution and with parameter 1. By conditioning

$$P[X+Y \le t] = \int_{-\infty}^{\infty} P[X+Y \le t = y] f_Y(y)dy$$
$$= \int P[X \le t-y] f_Y(y)dy$$
$$= \int_0^t \left(1-e^{-(t-y)}\right) f_Y(y)dy$$
$$= 1-e^{-t}-te^{-t} \text{ for } t \ge 0$$

Thus, if X and Y are independent and absolutely continuous, then

$$F_{x+y}(t) = \int_{-\infty}^{\infty} Fx(t-y)\, fy(y)dy$$

The integral above is called the convolution of the two distribution function and is often denoted $F_X * F_Y$

Another approach is to appeal to the Central Limit Theorem we have discussed above in the context of individual loss model premium. As you may recall, the

theorem states that if $X_1,.....,X_n$ are independent random variables, then the distribution of $\sum_{i=1}^{n} X_i$ is approximately the normal distribution with mean $\sum_{i=1}^{n} E[X_i]$ and variance $\sum_{i=1}^{n} Var(X_i)$. Thus, you have a case where an approximation of normal distribution does not depend on the detailed nature of the original distribution but only on the first two moments.

Note the insight gained from above theorem. If an insurance company has sold insurance at the pure premium, it will not achieve break-even and also face bankruptcy due to amount of claims arrivals. Therefore, insurance companies must charge an amount greater than the pure premium. A common methodology for the company is to charge $(1+\theta)$ times the pure premium. When this scheme is followed, θ is called the *relative security* loading and the amount $\theta \times$ (pure premium) is called the *security loading.*

9.4 THE COLLECTIVE RISK MODEL

Let us consider an alternative model describing the total claim amount in a fixed period in a portfolio of insurance contracts.

Let N be the number of claims arising from policies in a given period; X_1 be the amount of the first claim, X_2 be the amount of the second claim and so on. In the collective risk model, the random sum

$$S = X_1+X_2+... +X_n \tag{9.11}$$

represents the aggregate claims generated by the portfolio for the period under study. The number of claims N is a random variable and is associated with the frequency of claim. The individual claims X_1, X_2...are identically distributed random variable and the random variables N, X_1, X_2...are mutually independent.

9.4.1 General Premium Formulae

In order to find formulae for premiums based on the expected value of the total claim, let us assume that *E (X), E (N), Var (X)* and *Var (N)* exist. For the collective risk model, expected value of aggregate claims is the product of the expected individual claim amount and the expected number of claims,

$$E(S)=E(N)\,E(X), \tag{9.12}$$

while the variance of aggregate claims is the sum of two components where the first is attributed to the variability of individual claims amounts and the other to the variability of the number of claims:

$$Var(S) = E(N)\,Var(X) + \{E(x)\}^2 + Var(N). \tag{9.13}$$

Thus, it is easy to obtain the following premium formulae in the collective risk model:

Pure risk premium: $P = E(N)E(X)$,

Premium with safety loading: $P_{SL}(\theta) = (1+\theta)E(N)E(X)$, $\theta \geq 0$,

Premium with variance loading:

$$P_v(a)=E(N)E(X)+a\left[E(N)Var(X)+\{E(X)\}^2 Var(N)\right],\quad a\geq 0$$

Premium with standard deviation loading:

$$P_{SD}(b)=E(N)E(X)+b\sqrt{E(N)Var(X)+\{E(X)\}^2 Var(N)},\quad b\geq 0$$

If we assume that $M_N(t)$ and $M_X(t)$ exist, the moment generating function of S can be derived as

$$M_S(t)=M_N\{\ln M_X(t)\}$$

and thus the exponential premium is of the form

$$P_E(c)=\frac{\ln\left[M_N\{\ln M_X(c)\}\right]}{c},\quad c>0$$

It is often difficult to determine the distribution of the aggregate claims and this fact causes problems with calculating the quantile premium given by

$$P_Q(\varepsilon)=F_s^{-1}(1-\varepsilon),\ \varepsilon\in(0,1)$$

Although the distribution function of S can be expressed by means of the distribution of N and the convolution of the claim amount distribution, this is too complicated in practical applications. Therefore, approximations for the distribution of the aggregate claims are usually considered.

9.4.2 Normal and Gamma Approximations

In case of individual risk model above we have employed the normal approximation as an approximation for the distribution of aggregate claims. The same technique can be extended to the collective model when the expected number of claims is large.

As was observed earlier the normal approximation simplifies the calculations. If the distribution of S can be approximated by a normal distribution with mean $E(S)$ and variance $Var(N)$, the quantile premium is given by the formula

$$P_Q(\varepsilon)=E(N)E(X)+\Phi^{-1}(1-\varepsilon)\sqrt{E(N)Var(X)+\{E(X)\}^2 Var(N)},$$

where $\varepsilon\in(0,1)$ and $\phi(.)$ denote the standard normal distribution function. It is easy to notice, that this premium is equal to the standard deviation-loaded premium with $b=\Phi^{-1}(1-\varepsilon)$.

Moreover, in the case of the normal approximation, it is possible to express the exponential premium as

$$P_E(c)=E(N)E(X)+\frac{c}{2}[N]Var(X)+\{E(X)\}^2 Var(N)\ ],\quad c>0,$$

which is the same premium as resulting from the variance principle with $a=\frac{c}{2}$.

Let us also mention that since the mean and variance in the case of the normal approximation are the same as for the distribution of S, the premiums based on the expected value are given by the general formula.

Unfortunately, the normal approximation is not usually sufficiently accurate. The disadvantage of this approximation lies in the fact that the skewness of the normal distribution is always zero, as it has a symmetric probability density function. Since the distribution of aggregate claims is often skewed, another approximation of the distribution of aggregate claims that accommodates skewness is required.

The distribution function of the translated (shifted) gamma distribution is given by

$$G^{tr}(x;a,\beta,x_0) = F(x-x_0;a,\beta),\ x,a,\beta>0$$

where $F(x,\alpha,\beta)$ denotes the distribution function of the gamma distribution with parameters α and β

$$F(x;a,\beta) = \int_0^x \frac{\beta^a}{\Gamma(a)} t^{a-1}e^{-\beta t}dt,\ x,a,\beta>0$$

To apply the approximation, the parameters α, β and x_0 have to be selected so that the first, second, and third central moments of S equal the corresponding items for the translated gamma distribution. This procedure leads to the following result:

$$\alpha = 4\frac{\{Var(S)\}^3}{\left(E\left[\{S-E(S)\}^3\right]\right)^2}$$

$$\beta = 2\frac{Var(S)}{E\left[\{S-E(S)\}^3\right]}$$

$$x_0 = \beta = -2\frac{\{Var(S)\}^2}{E\left[\{S-E(S)\}^3\right]}$$

In the case of the translated gamma distribution, it is impossible to give a simple analytical formula for the quantile premium. Therefore, in order to find this premium a numerical approximation must be used. However, it is worth noticing that the exponential premium can be presented as

$$P_E(c) = x_0 + \frac{\alpha}{c}\ln\left(\frac{\beta}{\beta-c}\right), c>0$$

9.4.3 Compound Poisson Distribution

In many applications, the number of claims N is assumed to be described by the Poisson distribution with the probability function given by

$$P(N=n) = \frac{\lambda^n e^{-\lambda}}{n!}, n=0,1,2,.....,$$

where $\lambda > 0$. With this choice of the distribution of N, the distribution of S is called a compound Poisson distribution.

The compound Poisson distribution has a number of useful properties. Formulae for the exponential premium and for the premiums based on the expectation of the aggregate claims simplify because

$$E(N) = Var(N) = \lambda \text{ and } M_N(t) = exp\left\{\left(\lambda e^t - 1\right)\right\}.$$

Moreover, for large λ, the distribution of the compound Poisson can be approximated by a normal distribution with mean $\lambda E(X)$ and variance $\lambda E\left(X^2\right)$, and the quantile premium is given by

$$P_Q(\varepsilon) = \lambda E(X) + \Phi^{-1}(1-\varepsilon)\sqrt{\lambda E\left(X^2\right)} \varepsilon \in (0,1),$$

and the exponential premium is of the form

$$P_E(c) = \lambda E(X) + \frac{c}{2}\lambda E\left(X^2\right), \quad c > 0$$

If the first three central moments of the individual claim distribution exist, the compound Poisson distribution can be approximated by the translated gamma distribution with the following parameters

$$\alpha = 4\lambda \frac{\left\{E\left(X^2\right)\right\}^3}{\left\{E\left(X^3\right)\right\}^2},$$

$$\beta = 2\frac{E\left(X^2\right)}{E\left(X^3\right)},$$

$$x_0 = \lambda E(X) - 2\lambda \frac{\left\{E\left(X^2\right)\right\}^2}{E\left(X^3\right)}$$

With these parameters one can obtain the formula for the exponential premium.

It is worth mentioning that the compound Poisson distribution has many attractive features. Modelling for example, the combination of a number of portfolios, each of which has a compound Poisson distribution of aggregate claims, also has a compound Poisson distribution of aggregate claims. Moreover, this distribution can be used to approximate the distribution of total claims in the individual model. Although the compound Poisson distribution is normally appropriate in life insurance modelling, it sometimes does not provide an adequate fit to insurance data in other coverage.

9.4.4 Compound Negative Binomial Distribution

When the variance of the number of claims exceeds its mean, the Poisson distribution is not appropriate. In this situation, the use of the negative binomial distribution with the probability function given by

$$P(N=n)=\binom{r+n-1}{n}p^r q^n,\ n=0,1,2,.....,$$

where $r>0, 0<p<1$, and $q=1-p$, is suggested. In many cases it provides a significantly improved fit to that of the Poisson distribution. When a negative binomial distribution is selected for N, the distribution of S is called a compound negative binomial distribution.

Since for the negative binomial distribution we have

$$E(N)=\frac{rq}{p},\quad Var(N)=\frac{rq}{p^2},$$

and

$$M_N(t)=\left(\frac{p}{1-qe^t}\right)^r$$

the formulae for the exponential premium and for the premiums based on the expectation of the aggregate claims simplify.

For large r, the distribution of the compound negative binomial can be approximated by a normal distribution with the mean $\frac{rq}{p}\mathrm{E}(X)$ and variance $\frac{rq}{p}\mathrm{Var}(X)+\frac{rq}{p^2}\{\mathrm{E}(X)\}^2$. In this case the quantile premium is given by

$$P_Q(\varepsilon)=\frac{rq}{p}E(X)+\Phi^{-1}(1-\varepsilon)\sqrt{\frac{rq}{p}Var(X)+\frac{rq}{p^2}\{E(X)\}^2},\ \varepsilon\in(0,1),$$

and the exponential premium is of the form

$$P_E(c)=\frac{rq}{p}E(X)+\frac{c}{2}\left[\frac{rq}{p}Var(X)+\frac{rq}{p^2}\{E(X)\}^2\right],\ c>0$$

It is worth mentioning that the negative binomial distribution arises as a mixed Poisson variate. More precisely, various distributions for the number of claims can be generated by assuming that the Poisson parameter Λ is a random variable with probability distribution function $u(\lambda)$, $\lambda>0$, and that the conditional distribution of N, given $\Lambda=\lambda$, is Poisson with parameter λ. In such a case the distribution of S is called a compound mixed Poisson distribution. This choice might be useful. For example, when we consider a population of insured where various classes of insured within the population generate number of claims according to the Poisson distribution. But the Poisson parameters may be different for the various classes. The negative binomial distribution can be derived in this fashion when $u(\lambda)$ is the gamma probability density function.

9.4.5 Importance of Surplus at Time t

As seen above in the collective risk model, the time at which claims are made is taken into account. Here the aggregate claims up to time t *is* assumed to be given by $\sum_{k=1}^{N(t)}X_k$, where $X_1, X_2, ...$ are independent identically distributed

random variables representing the sizes of the respective claims, $N(t)$ is stochastic process representing the number of claims up to time t, and N and the $X's$ are independent. It is important to note the insurer's surplus at time t, denoted by $U(t)$, which is assumed to be of the form $U(t) = u + ct - \sum_{k=1}^{N(t)} X_k$, where u is the surplus at time $t = 0$, and c represents the rate of premium income.

Special attention will have to be paid to the problem of estimating the probability that the insurance company has negative surplus at some time t since this would mean that the company is ruined. To gain familiarity with some of the ideas involved, the classical gambler's ruin problem is studied below.

Check Your Progress 1

1) Write the meaning of net premium reserve in life insurance.

.....

.....

.....

.....

.....

.....

2) Differentiate between prospective reserve and reserve formulae.

.....

.....

.....

.....

.....

.....

3) What is the meaning of individual risk model in life insurance?

.....

.....

.....

.....

.....

4) How collective risk model is different from individual risk model in life insurance?

.....

.....

.....

.....

5) How will you compute the formula of compound Poisson distribution in life insurance premium?

.....

.....

.....

.....

.....

9.5 CLASSICAL RISK MODEL

The Classical Model

Suppose we model a risk process as follows: The occurrence of the claims is described by a point process and the amount of money to be paid by the company at each claim by a sequence of random variables. The company receives a certain amount of premium of cover is liability and the difference between the premium income and the (average) cost for the claims is the 'safety loading' r.

In the classical Cramer-Lundberg-model the claim sizes $Y_1, Y_2, \ldots$ are independent identically distributed (iid) positive random variables with distribution function F and finite mean $\mu = E(Y_1)$. The claims Y_k arrive at time $T_k = D_1 + \ldots + D_k$ according to a homogenous Poisson process, where the inter-occurrence times $D_1, D_2, \ldots$ are independent exponentially distributed random variables with intensity α. If the corresponding claim number is $N(t) = \sup\{k \geq 1 : T_k \leq t\}, t \geq 0$, then the capital at time t is

$$U(t) = u + ct - V(t), \quad t \geq 0,$$

where u is initial capital, $V(t) = \sum_{n=1}^{N(t)} Y_n$ the total claim amount until time t and $c > 0$ is the premium income rate. The ruin probability in infinite time is then defined by

$$\psi(u) = P\left(\inf_{0 \leq t \leq \infty} U(t) < 0\right).$$

For the case of exponentially distributed claim sizes Y_k there exists an explicit formula for the ruin probability first established by [Lundberg 1903]:

$$\psi(u) = \frac{1}{1+r} e^{-\frac{ru}{\mu(1+r)}}, \qquad (1)$$

where $r = \frac{c}{\alpha\mu} - 1$ is the safety loading, which is positive for a reasonable insurance policy. Using an empirical value for μ and α, a desired value for the ruin probability can in practice (under the assumption of independence between the claims) be approximated by choosing a suitable risk premium

rate C (this procedure relies on the law of large numbers which only holds for independent events).

In the more general case of $Y_1, Y_2, \ldots$ being identically distributed according to an arbitrary (but exponentially bounded) distribution function F, there exists the famous Cramer-Lundberg approximation

$$\lim_{u\to\infty} e^{Ru}\psi(u) = C, \tag{2}$$

where C is a constant and the so-called Lundberg exponent R depends on F and is given by a functional equation.

This asymptotic result (2) is only proved for exponentially distributed D_i. However, the theorem shows below that it holds for a still wider class of distributions.

For further purposes it is convenient to rewrite the process in the form of a random walk

$$S_i = \sum_{j=1}^{i} X_j \tag{3}$$

where $X_i = Y_i - cD_i$, so the process direction is reversed (starting at $U(t) = 0$ and causing ruin, if $U(t) = u$). In the classical model, $X_1, X_2, \ldots$ is an iid sequence of random variables. Note that for a reasonable choice of the premium rate $c, E(X_i) < 0$, so the random walk has negative drift. Using this notation the ruin probability can equivalently be defined by

$$\psi(u) = P\left(\max(0, S_1, S_2, \ldots) \geq u\right) \tag{4}$$

If the distribution function F is normal with mean μ and variance σ^2, then the random walk S_n can be seen as discrete equidistant points taken from its continuous analog to the well-known Wiener process $W(t)$. But for a Wiener process $W(t)$, it is known that the ruin probability is

$$\psi(u) = P\left(\max_{t\geq 0}\left[\sigma W(t) - \mu t\right] \geq U\right) = e^{-\frac{2\mu u}{\sigma^2}},$$

so there exists a Lundberg exponent $R = 2\mu/\sigma^2$ for the continuous case. For our discrete random walk with normal distribution, the following theorem applies and hence shows the existence of a Lundberg exponent R:

Theorem 1 Let F be the common (compound) distribution function of the *iid* X_i. If F is such that there exists a constant V with

$$\int_{-\infty}^{\infty}\left(1 - F(u)\right)e^{Vu}du = 1, \tag{5}$$

then the Cramer-Lundberg approximation

$$\lim_{u\to\infty} e^{Ru}\psi(u) = C$$

holds for some constant C.

9.6 STOPPING TIMES AND MARTINGALES

Let us see the problem as given in Veeh (2001)

Discrete time collective risk model: We start with the gambler's ruin problem in which a gambler enters a casino with z dollars and plays a game of chance. She wins with probability p and her loss is \$1with probability $q=1-p$. Suppose also that the gambler will quit playing if her fortune ever reaches $a>z$ and will be forced to quit by being ruined if her fortune reaches 0. Note that in this problem, we are interested to find the probability that the gambler is ultimately ruined.

Consider a simple case $p=q=\frac{1}{2}$. Let X_j be the amount won or lost on the j^{th} play of the game. When there is no restriction on quitting this game, the fortune of the gambler after k plays of the game is given by

$$z+\sum_{j=1}^{k} X_j$$

When the actual game is played the gambler either reaches her goal or is ruined. Let T be the game on which this occurs. So we write

$$T=\inf\left(k: z+\sum_{j=1}^{k} X_j=0 \ \ or \ \ a\right)$$

where $t=$random time. Note that for this event $[T\le k]$ depends only on the random variables $X_1,.....X_k$. and in order to decide at time k whether or not necessary to look into the future. Such special random times are called stopping times. Thus $X_1, X_2,...$ are random variables and T is a nonnegative integer valued random variable with the property that for each integer k the event $[T\le k]$depends only on $X_1,.....X_k$. Then T is said to be a *stopping time* (relative to the sequence $X_1 X_2...$)

The random variable $z+\sum_{j=1}^{T} X_j$ gives the gambler's fortune when she leaves the casino, which is either a *or* 0. Let $\Psi(z)$ be probability that the gambler leaves the casino with 0. Then by direct computation

$$E\left[z+\sum_{j=1}^{T} X_j\right]=a(1-\Psi(z)).$$

The ruin probability $\Psi(z)$ can be obtained by computing the expectation in another way. Since each of the random variables X_j takes values 1 and -1 with equal probability, $E[X_j]=0$. Hence for any integer $k, E[\sum_{j=1}^{k} X_j]=0$. So it is at least plausible that $E[\sum_{j=1}^{k} X_j]=0$ and $E\left[z+\sum_{j=1}^{k} X_j\right]=z$. Using the above result, we get $z=a(1-\Psi(z))$. Thus $\Psi(z)=1-\frac{z}{a}$ for $0\le z\le a$ are the ruin probabilities.

To compute the ruin probabilities we rely on stopping time T and fair game feature of gambling which gives $p=q=\frac{1}{2}$. Close to the notion of a fair game

is the idea behind a martingale. To appreciate the idea you need to recall that a sequence of random variable $M_0, M_1, \ldots\ldots M_n$ is a martingale if

$E[M_k | M_{k-1}, \ldots\ldots M_0] = M_{k-1}$ for all $k \geq 1$. In the context of the gambling, if M_k is the gambler's fortune after k plays of a fair game, then given M_{k-1} the expected fortune after one more play is still M_{k-1}.

Example 1.

$$M_k = z + \sum_{i=1}^{k} X_j$$

With $M_0 = Z$ is a martingale. Because $X_1, \ldots\ldots X_{K-1}$ give the same information as that of $M_0, M_1, M_2 \ldots$ we have

$$E[M_k | M_0, \ldots\ldots M_{k-1}] = E[M_k | X_0, \ldots\ldots X_{k-1}]$$
$$= t + \sum_{j=1}^{k-1} X_j + E[X_k | X_0, \ldots\ldots X_{k-1}] = M_{k-1}$$

as $X_0, \ldots\ldots X_{k-1}$ are independent variable the expectation on these is zero.

Example 2.

The sequence $M_0 = z^2$ and $M_k = (z + \sum_{j=1}^{k} X_j)^2 - k$ for $k \geq 1$ is a martingale, if knowing $M_0, \ldots M_{k-1}$ is the same as knowing $X_1, \ldots, X_{k-1}$ and $X's$ are independent. To see why, write

$$M_k = (z + \sum_{j=1}^{k} X_j + X_k)^2 - k = \left(z + \sum_{j=1}^{k-1} X_j \right)^2 + 2X_k \left(z + \sum_{j=1}^{k-1} X_j \right) + X_k^2 - k$$

On taking conditional expectation and recalling that X_K is independent of other $X's$, we get $E(X_k) = 0$, $E(X_k^2) = 0$ and above results follow.

Optional Stopping Theorem:

It states that if $\{M_k\}$ is martingale and T is a stopping time, then $E[M_T] = E[M_o]$. In the gambling context, this concept shows gambling strategy T can make a fair game biased. We can obtain the duration of gambler stay in the casino using the martingale $M_0 = z^2$ and $M_K = \left(z + \sum_{j=1}^{k} X_j\right)^2 - k$ for $k \geq 1$ along with the same stopping time T. The expected value of the random variable $M_T = \left(z + \sum_{j=1}^{T} X_j\right)^2 - T$ can be computed directly as $E[M_T] = a^2(1 - \Psi(z) - E(T)$. By the optional stopping theorem, $E[M_T] = E[M_0] = z^2$. When these two expressions are compared we get, $E[T] = a^2(1 - \Psi(z)) - z^2 = az - z^2$.

This is the expected duration of the game.

To derive the appropriate martingale, we often take the help of Wald's martingale. This formulation is also known as exponential martingale. To get this, take $X_1, X_2 \ldots\ldots$ as independent and identically distributed random variables to define

$$W_k = \frac{e^{t\sum_{j=1}^{k} X_j}}{E\left[e^{t\sum_{j=1}^{k} X_j}\right]}$$

Note that in the denominator above we have the moment generating function of the sum evaluated at t. For each fixed t the sequence W_k is a martingale (here $W_0 = 1$). This follows if X and Y are independent and that $E\left[e^{t(X+Y)}\right] = E\left[e^{tX}\right]E\left[e^{tY}\right]$ which is Wald's martingale (or the exponential martingale) for the X sequence. The independence assumption result in

$$E\left[e^{t\left(\sum_{j=1}^{k} X_j\right)}\right] = E\left[e^{t\left(\sum_{j=1}^{k-1} X_j\right)}\right] \times E\left[e^{tX_k}\right]$$

and $\{W_k : k \geq 0\}$ is a martingale irrespective of any fixed value of t.

A non-zero value of t can be found in many cases so that the denominator of the Wald martingale is 1. Use of this particular value of t makes application of the Optional Stopping Theorem easy.

We use the above technique for analysing the collective risk model. Our objective is to derive ruin probability $\Psi(t)$. Let the insurer has initial reserve z and that premium income is collected at the rate of c per unit time. Let X_K be the claims that are payable at time k, and the X's are independent and identically distributed random variables. Then the insurers reserve at time k is $z + ck - \sum_{j=1}^{k} X_j = z + \sum_{j=1}^{k}(c - X_j)$ if t denotes the time of ruin, we can write

$$T = \min\left\{k : z + ck - \sum_{j=1}^{k} X_j 0\right\}.$$

In the above formulation if $E[c - X_j] \leq 0$, then ruin is guaranteed since premium income in each period is not adequate to balance the average amount of claims in the period. Thus, to arrive at meaningful result let us assume that $E[c - X_j] > 0$.

If we suppose there is a number τ such that $E\left[e^{\tau(c - X_j)}\right] = 1$, we will have Wald's martingale with the denominator 1. Moreover, we will get

$$M_R = e^{t(t + (k - \sum_{j=1}^{k} X_j))}$$

which is martingale. Taking expectation of M_T and using the Optional Stopping Theorem, we have $E[M_T] = E[M_0] = e^{Tt}$. Direct computation of it is, therefore,

$$E[M_T] = E\left[e^{\tau(z+cT-\sum_{j=1}^{T} X_j)} | T < \infty\right] \Psi(z).$$

$$\Psi(z) = e^{\tau z} / E\left[e^{\tau(z+cT-\sum_{j=1}^{T} X_j)} | T < \infty\right]$$

Note that if $\tau < 0$, the denominator of above fraction is greater than 1. Hence $\Psi(z) \leq e^{\tau z}$. Consequently, the ruin probability decays exponentially as the initial reserve increases. If we put $R = -\tau$ as an adjustment coefficient, then we have $\psi(z) \leq e^{-Rz}$ and a large adjustment coefficient implies that the ruin probability declines rapidly as the initial reserve increases.

9.7 COLLECTIVE RISK MODEL AND RUIN PROBLEM

We will use the gambler's ruin problem discussed above to compute the ruin probability in collective risk model. Instead of discrete time being used so far, we will take the help of continuous time t and calculate the ruin probability. The discussion that follows takes into account development of the collective risk model using martingale framework and adopts the notations as given by Veeh (2001).

To understand technical derivation there, we have discussed these separately. Let us take the claim process as $\sum_{k=1}^{N(t)} X_k$ where $X_1, X_2, \ldots$ are independent identically distributed random variables representing the sizes of the formulation $N(t)$ is number of claims up to time t which is a stock size process. Let N and X be independent and take insurance surplus as $u(t) = u + ct - \sum_{k=1}^{N(t)} X_K$ where $u > 0$ is the surplus at time $t = 0$; $c > 0 =$ rate of premium arrival per unit of time. We denote the probability of ruin with initial surplus by u. As has been done above in case of discrete time, the Wald martingale will be used with Optional Stopping Theorem to derive the information on ruin probability.

As in discrete time setting the Wald martingale will be used together with the Optional Stopping Theorem in order to obtain information about the ruin probability. See that the denominator of the Wald martingale is $E[e^{vU(t)}]$, so first step is to find a $v \neq 0$ such that

$$E\left[e^{v(ct-\sum_{k=1}^{N(t)} X_k)}\right] = 1 \text{ irrespective of the value of } t.$$

Consider the Wald martingale as the random sum $\sum_{k=1}^{N(t)} X_k$. For each moment generating function, this sum can be easily computed by conditioning on the value of the discrete random variable $N(t)$.

$$E\left[e^{v\sum_{k=1}^{N(t)}X_k}\right]=E\left[E\left[e^{v\sum_{k=1}^{N(t)}X_k}N(t)\right]\right]$$

$$=\sum_{j=0}^{\infty}E\left[e^{v\sum_{k=1}^{N(t)}X_k}N(t)=j\right]P[N(t)=j]$$

$$=\sum_{j=0}^{\infty}E\left[e^{v\sum_{k=1}^{j}X_k}\right]P[N(t)=j]$$

$$=\sum_{j=0}^{\infty}\left(E\left[e^{vX}\right]\right)^{j}P[N(t)=j]$$

$$=\sum_{j=0}^{\infty}e^{j\ln(E[e^{vX}])}P[N(t)=j]$$

$$=M_{N(t)}\ln(M_X(v))).$$

- Hence there exists a $v\neq 0$ so that $E\left[e^{v(ct-\sum_{k=1}^{N(t)}X_k}\right]=1$ if and only if $e^{vct}M_{N(t)}\ln(M_X(-v)))=1$ for all t.

If there is number $R>0$ so that $e^{-Rct}M_{N(t)}(\ln(M_X(R)))=1$ for all t, R is called the adjustment coefficient. Using $-R$ as the value of v in Wald's martingale we have

$W_t=e^{-R\left(u+ct-\sum_{k=1}^{N(t)}X_k\right)}$ is a martingale.

Define a stopping time T_a and $T_a=\inf\{s:u+cs-\sum_{k=1}^{N(s)}X_k\leq 0 \ \ or\geq a\}$ where a is an arbitrary fixed positive number. Intuitively, T_a is stopping time. By the Optional Stopping Theorem, $E[W_{T_a}]=e^{-Ru}$. which gives

$$E[W_{Ta}]=E[e^{-R(u+cT_a-\sum_{k=1}^{N(T_a)}X_k)}\,|\,u+cT_a-\sum_{k=1}^{N(T_a)}X_k\leq 0]P[u+cT_a-\sum_{k=1}^{N(T_a)}X_k\leq 0]$$

$$+E[e^{-R(u+ct-\sum_{k=1}^{N(T)}X_k)}\,|\,u+cT_a-\sum_{k=1}^{N(T_a)}X_K\geq a]P[u+cT_a-\sum_{k=1}^{N(T_a)}X_k\geq a]$$

Since this equation is valid for any fixed positive a, and as $R>0$, we can set limits as $a\to\infty$. Since $\lim_{a\to\infty}P[u+cT_a-\sum_{k=1}^{N(T_a)}X_k)\leq 0=\Psi(u)$ and $\lim_{a\to\infty}e^{-Ra}=0$ the following result is obtained.

Theorem.

Suppose that in the collective risk model the adjustment coefficient $R>0$ satisfies $e^{-Rct}M_{N(t)}(\ln(M_x(R)))=1$ for all t. Let $T\inf\{s:u+cs-\sum_{k=1}^{N(s)}X_k\leq 0\}$ be the random time at which ruin occurs. Then

$$\Psi(u)=\frac{l^{Ru}}{E[e^{R(u+cT-\sum_{k=1}^{N(T)}X_k}\,|\,T<\infty]}\leq e^{-Ru}.$$

As in the case of discrete time model, the existence of an adjustment coefficient guarantees that the ruin probability decreases exponentially as the initial surplus u increases. To see how it works, we consider a Poisson process

with constant intensity $\lambda > 0$ and suppose $W_1, W_2, \ldots$ are independent identically distributed exponential random variables with mean $\frac{1}{\lambda}$ and common density $\lambda e^{-\lambda x} 1_{(0,\infty)}(x)$. Let $W's$ be the waiting times between claims. The Poisson process can then be viewed as the number of claims that arrive up to time t. This means that $N(t) = \inf\{k : \sum_{j=1}^{k+1} W_j > t\}$ and it can be shown that for any fixed t the random variable, $N(t)$ has independent increments. Thus, whenever $t_1 < t_2 < \ldots < t_n$ are fixed real numbers, the random variables $N(t_2) - N(t_1) \ldots N(t_n) - N(t_{n-1})$ are independent. Using this, we can get

$$
\begin{aligned}
&E[e^{vN(t)}] \\
&= \sum_{j=0}^{\infty} e^{vj} P\left[N(t) = j\right] \\
&= \sum_{j=0}^{\infty} h^{e^{vj}} \; e^{-\lambda t} (\lambda t)^j / j! \\
&= e^{-\lambda t} \sum_{j=0}^{\infty} \left(e^{v} \lambda t\right)^j / j! \\
&= e^{\lambda t} e^{v} - 1
\end{aligned}
$$

From the moment generating function of $N(t)$, we derive a simple formula for the adjustment coefficient in this case. The general equation for the adjustment coefficient was earlier found to be $e^{-Rct} \mathrm{M}_{N(t)}(\ln(M_X(R))) = 1$. Taking logarithms and using the form of the moment generating function of $N(t)$ shows that the adjustment coefficient is the positive solution of the equation

$$\lambda + cR = \lambda M_X(R).$$

If we take $N(t)$ as a Poisson process, more specific results can be generated. To do this define a stopping time $T_{u\lambda} = \inf\{s : U(s) < u\}$ to be the first time that the surplus falls below its initial level and denote by $L_1 = u - U(T_u)$, the amount by which the surplus falls below its initial level. Then

$$P[T_u < \infty, L_1 \geq y] = \frac{1}{(1+\theta)E[X]} \int_y^{\infty} (1 - F_x(x))dx.$$

There are two useful consequences of the above formula. First, by taking $y = 0$, the probability that the surplus ever drops below its initial level is $1/(1+\theta)$. Second, an explicit formula for the size of the drop below the initial level is obtained as

$$P[L_1 \leq y \mid T_u < \infty] = \frac{1}{E[X]} \int_0^{y} (1 - F_X(x))dx$$

which can be evaluated in specific cases.

Consider a random variable L, which represents the maximum aggregate loss, and is defined by

$$L = \max_{t\geq 0}\left\{\sum_{k=1}^{N(t)} X_k - ct\right\}.$$

Note that $P[L \leq u] = 1 - \Psi(u)$ from which the distribution of L has discontinuity at the origin of size $1 - \Psi(0) = \theta/(1+\theta)$, and is continuous otherwise. In fact a reasonably explicit formula for the moment generating function of L can be obtained.

If $N(t)$ is a Poisson process and $L = \max_{t\geq 0}\left\{\sum_{k=1}^{N(t)} X_k - ct\right\}$, then

$$M_L(v) = \frac{\theta E[X]v}{1 + (1+\theta)E[X]v - M_X(v)}.$$

This formula for the moment generating function of L can sometimes be used to find an explicit formula for the distribution function of L, and hence $\Psi(u) = 1 - P[L \leq u]$.

Example. Suppose X is exponential with mean 1 and that $\theta = 1$. Then substitution gives

$$M_L(t) = \frac{1}{2} + \frac{1}{2}\frac{1}{1-2t},$$

after using long division (or partial fractions). The second term is half of the moment generating function of an exponential random variable with mean $\frac{1}{2}$, while the first term is half the moment generating function of a random variable that is degenerate at zero. The ruin probability is therefore $\Psi(u) = P[L > u] = e^{-2u}/2$ for $u > 0$. Notice that $\Psi(0) = \frac{1}{2}$.

The analysis of random sums of the form $\sum_{j=1}^{N} X_j$ in which the $X's$ are independent and identically distributed has played a key role in the preceding analysis. The distribution of such a sum is called a compound distribution, with N as the compounding variable and X as the compounded variable. The conditioning method used earlier shows that the moment generating function of such a sum is $M_N(\ln M_x(t))$.

The first two moments are $E\left[\sum_{j=1}^{N} X_j\right] = E[N]E[X]$ and

$$Var\left(\sum_{j=1}^{N} X_j\right) = E[N]Var(X) + (E[X])^2 Var(N).$$

A random variable S has the compound Poisson distribution with Poisson parameter λ and mixing distribution $F(x)$, denoted $S = CP(\lambda, F)$, if S has the same distribution as $\sum_{j=1}^{N} X_j$ where $X_1, X_2, \ldots$ are independent identically distributed random variables with common distribution function F and N is random variable which is independent of the $X's$ and has a Poisson distribution with parameter λ.

Example. For each fixed t, the aggregate claims process $CP(\lambda_t, F_x)$.

Example. If S has the $CP(\lambda, F)$ distribution then the moment generating function of S is

$$M_S(v) = \exp\left\{\lambda \int_{-\infty}^{\infty} (e^{uv} - 1) dF(u)\right\}$$

This follows from the earlier general derivation of the moment generating function of a random sum.

Suppose that S has the $CP(\lambda, F)$ distribution and T has the $CP(\delta, G)$ distribution and that S and T are independent. Then $S+T$ has the

$$CP(\lambda+\delta, \frac{\lambda}{\lambda+\delta}F + \frac{\delta}{\lambda+\delta}G) \text{ distribution.}$$

This property is very useful in the insurance and results of the analysis of different policy types can be easily combined into one grand analysis of the company's prospects as a whole. A compound Poisson distribution can also be decomposed.

Example. Suppose each claim is either for \$1 or \$2, each event having probability 0.5. If the number of claims is Poisson with parameter λ then the amount of total claims, S, is compound Poisson distributed with moment generating function

$$M_S(v) = \exp\{0.5\lambda(e^v - 1) + 0.5\lambda(e^{2v} - 1)\}.$$

Hence S has the same distribution as $Y_1 + 2Y_2$ where Y_1 and Y_2 are independent Poisson random variable with mean $\lambda/2$. Thus the number of claims of each size are independent.

Example. The collective risk model can be used as an approximation to the individual risk model. In the individual risk a product often represents model the claims amount $B_j X_j$ in which B is a Bernoulli random variable which represent whether a claim is paid or not and X is the amount of the claim. Then

$$BX = \sum_{j=1}^{B} X_j \approx \sum_{j=1}^{N} X_j$$

where N has a Poisson distribution with parameter $P[B=1]$ and $X_1, X_2, \ldots$ are independent random variable each having the same distribution as X. Thus the distribution of BX may be approximated by the $CP(P[B=1], F_X)$ distribution.

The analysis of the aggregate loss random variable L introduced the idea of mixing. A mixture of distributions often arises when the outcome of an experiment is the result of a two-step process.

9.7.1 Collective Model of Risk Theory: A Discussion

The following discussion is taken from Mackova, Monika (2005).

The collective risk model is a stochastic model that uses losses as random variables in certain portfolio of insurance policies.

The basic mathematical assumptions of the model are

- The number of individual claims (losses) within a certain portfolio of insurance policies during a certain period of time, denoted as N, is a non-negative, integer valued random variable, called frequency.
- The individual claims (losses) occurring during this period, denoted as X_i, for $i = 1,\ldots, N$, are positive random variables, stochastically independent, identically distributed,
- The individual claims (losses) $X_1,\ldots,X_N$ are independent also from frequency N.
- If $N = 0$ then $X_1 + X_2 + \ldots + X_N = 0$.

Then the random variable

$$S = \sum_{i=1}^{N} X_i .$$

is called compound variable and describes the aggregate claim or aggregate loss for the some fixed period $(0,T]$ under consideration (mostly $T = 1$ year) for a (collective) insurance contract.

We do not restrict assumption only to the discrete random variable X. By extending to continuous variable, we can make a rescaling. Thus, we work with $E[S] = E[E[S/N]] = E[p_1 N] = p_1 E[N]$, *for* $p_1 = E[X]$, and

$$Var[S] = E[Var(S/N)] + Var(E[S/N]) = E[NVarX] + Var(p_1 N)$$
$$= E[N]VarX + p_1^2 VarN$$

That is, the expected value of aggregate claim depends on expected value of individual claim and the expected number of individual claims; a similar logic applies to the variance.

It may be useful to note that the period under consideration is mostly one year. We suppose that distribution of individual claims is not changing within this period of time. Therefore, the model is much more used for non-life insurance contract.

A frequently used distribution of X is Gamma, log-Gamma or log-normal distribution. In case of N, the distribution are assumed to be N Poisson, binomial, negative binomial.

Let us look at the notations used.

- If the probability distributions for the claims (losses) $X_i, i = 1,..,\ N$ are continuous with a density function (df) then f gives cumulative distribution function $(cdf)F$, which is given by $F(x) = \int_0^x f(u)du, x \geq 0$.

  The corresponding survival function (sf) is given by $\bar{F}(x) = 1 - F(x) = \int_x^{\infty} f(u)du, x \geq 0$.

- If the probability distribution for the claims (losses) $X_i, i = 1,..,N$ are discrete with a probability function $\{f_j, j \in N s.t. f_j = f(j) : p(X_i = j)\}$, then, a cumulative distribution function $(cdf)F$ is given by $F(x) = P(X_i \le x)$ *for* $x \ge 0$, and 0 otherwise. The corresponding survival function (sf) is given by $\bar{F}(x) = 1 - F(x)$.
- The cdf of the aggregate claim (loss) G is given by $G(x) = P(S \le x)$ with a density function g or probability function $\{g_k, k \in N \ .s.t. g_k = g(k) = P(S = k)\}$.
- The compounding probability function of N is denoted as $\{p_n, n \in N_0 .s.t. p_n = p(N = n)\}$.

Discretisation or Scaling

When the distribution function of random variable X is continuous we often use discretisation for numerical simulation (convolution, Panjer recursion, fast Fourier transform, etc.).

Let us write

$$X_\Delta = \left\lceil \frac{X}{\Delta} \right\rceil = \min\{k \in N \mid k\Delta \ge X\}$$

with $\Delta > 0$ and with probabilities

$$P(X_\Delta = k) = P\left(\left\lceil \frac{X}{\Delta} \right\rceil = k\right) = P\left(k - 1 < \frac{X}{\Delta} < k\right) = F\left(k\Delta - F(k-1)\Delta\right), k \in N$$

Then the discretised aggregate claim loss S_Δ has the probability generating function $\varphi_{S_\Delta}(s) = \varphi_N\left(\varphi_{S_\Delta}(s)\right), |s| \le 1$.

Convolution Method

To compute the aggregate claim distribution in the collective risk model we introduce recursive methods – the convolution principle.

As in the collective risk model we are looking for the compound distribution function $G(x)$ of random variable S, the distribution of aggregate claim losses, S for any discrete distribution of individual losses, r.v. X can be written as:

$$G(x) = P\left(\bigcup_{n=0}^{\infty}(S \le x \wedge N = n)\right) = \sum_{n=0}^{\infty} P(S \le x \wedge N = n)$$

$$= \sum_{n=0}^{\infty} P(S \le x / N = n).P(N = n)$$

$$= \sum_{n=0}^{\infty} P(X_1 + X_2 + ... + X_n \le x).P(N = n) = \sum_{n=0}^{\infty} P(N = n).F^{*n}(x),$$

where F^{*n} denotes the n-fold convolution of F. So,

$$F^{*n}(x) = P(S \le x) = P(X_1 + ... + X_n \le x).$$

Denote: $f^{*n}(x) = F^{*n}(x) - F^{*n}(x-1)$.

For the continuous random variable X, we use integral version of convolution.

Step 1:

In practice we often assume that the individual claims have positive integer value. So, for the discrete *r.v.* X we evaluate following expressions:

$$g_{k=0} = P(S=0) = P(N=0) = p_0$$

$$g_k = P(S=k) = \sum_{n=1}^{\infty} P(N=n).f^{*n}(k) \quad \text{k=1, 2,...}$$

If the value of each claim is greater then 1, we decrease the upper bound of the sum operator to level k:

$$g_{k=0} = P(S=0) = p(N=0) = p_0$$

$$g_k = P(S=k) = \sum_{n=1}^{k} p(N=n).f^{*n}(k) \quad \text{k=1, 2,...}$$

The computation difficulty of the modified formula is confined to only k^2.

Step 2:

Consider the discrete *r.v.* X_i, i=1,..., N, which reaches the values x_i with probability function $f_j = P(X_j = x_j)$, as probability generating function

$$\varphi(Z) = \sum_{j=1}^{\infty} P(X_j = x_j).z^{x_j} = f_1 z^{x_1} + f_2 z^{x_2} + ...$$

Then it holds for $\varphi^n(z) = (f_1 z^{x_1} + f_2 z^{x_2} + ...)^n$

If *r.v.* X_i are *i.i.d.* then the polynomial term $(f^j.z^{xj})^n$ of probability generating function implies that by the occurrence of n losses the aggregate claim value will be $(z^{xj})^n$ with the probability $(f^j)^n$. Then the probability of aggregate claims value reaches x_j which we derive in following steps:

1) exponentiate the probability generating function (with power from1 to x_j),
2) sum the exponent coefficient of term x_j, and
3) obtain following formula:

$$g_{x_j=0} = P(S=0) = P(N=0) = p_0$$

$$g_{x_j} = P(S=x_j) = \sum_{n=1}^{x_j} f_j^n \qquad x_j = 1,\ 2,...$$

Panjer Recursion

Panjer presented results for some specific distributions, when the claim amount is discrete and the recursive definition can be used to compute the aggregate claims distribution without the use of convolution.

Panjer Theorem

Consider the family of claim number distribution satisfying the recursion:

$$p_n = P(N=n) = (a+b/n)p_{n-1} \quad n=1,2,\ldots \quad a,b \in R$$

where p_n denotes the probability that exactly n claim occurs in the fixed time interval. Consider the compound distribution function

$$G(x) = \sum_{n=1}^{\infty} p_n F^{*n}(x) \quad x>0$$
$$= p_0 \qquad\qquad x=0$$

for arbitrary claim amount distribution $F(x)$, $x>0$. Consider the continuous function (corresponding results will be given for the discrete case as well). Then the density of total claim is:

$$g_k = \sum_{n=1}^{\infty} p_n f_k^{*n} \quad k>0$$
$$g_{k=0} = p_0 \qquad k=0$$

when $f(x)$ is the density associated with $F(x)$. Then the following recursion holds:

$$g_k = p_1 f_k + \int_0^k (a+by/k) f_y \cdot g_{k-y} dy \quad k>0$$

If the claim amount distribution is discrete and defined on the positive integers, the corresponding recursive definition of total claims is:

$$g_0 = P(S=0) = P(N=0) = p_0 \qquad k=0$$
$$g_k = P(S=k) = \sum_{j=1}^{k} \left(a+\frac{bj}{k}\right) f_j \cdot g_{k-j} \qquad k=1,2,\ldots$$

The formula for computing g_k is only k. In comparison to convolution, the computation the modified convolution formula is equal to k^2. Therefore the Panjer recursion formula is much more efficient.

Members of the family of claim number distribution satisfying the recursion:

$$p_n = P(N=n) = (a+b/n)p_{n-1} \quad n=1,2,\ldots \quad a,b \in R$$

are Poisson distribution, binomial distribution, negative binomial distribution and geometric distribution (as a special type of negative binomial).

Let us drive the recursive definition of the density of total claims:

- If random variable N is a Poisson distribution, then:

$$g_0 = e^{-\lambda}, g_k = \frac{\lambda}{k}\sum_{j=1}^{k} j.f_j.g_{k-j} \qquad k = 1,2,...$$

- If random variable N is a binomial distribution, then

$$g_0 = (1-p)^N, g_k = \frac{p}{1-p}\sum_{j=1}^{k}\left[(N+1)\frac{j}{k}-1\right].f_j.g_{k-j} \qquad k = 1,2,...$$

- If random variable N is a negative binomial distribution, then

$$g_0 = (1-p)^{\alpha}, g_k = p\sum_{j=1}^{k}\left[(\alpha-1)\frac{j}{k}+1\right].f_j.g_{k-j} \qquad k = 1,2,..$$

- If random variable N is a geometric distribution, then

$$g_0 = 1-p, g_k = p\sum_{j=1}^{k} f_j.g_{k-j} \qquad k = 1,2,...$$

Probability of Ruin and Panjer Recursion

In the following we will evaluate upper and lower bound of the probability of ruin under assumption of the classical risk model.

The following discussion illustrate the typical behavior of the insurer's surplus process:

The Surplus Process

Our interest in ruin theory is in the ruin probability for the insurer, i.e., the probability that at some point in time the claims exceed the premium income.

We formulate this by considering the following three elements:

- the initial surplus of the insurer – a fixed amount,
- the income flow of the premium – in the simplest case we consider here, aggregate premium income is a linear function of time t,
- the outflow of aggregate claims amount.

The insurer's surplus at time t is:

$$R(t) = R_0 + P(t) - S(t)$$

where $R(t)$ denotes the insurer's surplus at time t, R_0 represents the initial insurer's surplus at time $t = 0$; $P(t)$ represents insurer's premium income. Assume deterministic premium income $P(t) = Pt$, where P is a constant at time t.

Hence, we have $R(t) = R_0 + Pt - S(t)$

Let $N(t)$ be the number of claim that has occurred up to time t. Then $S(t)$ denotes the aggregate claims up to time t, $S(t) = X_1 + X_2 + ... + X_{N_t}$. Random variable X_i denotes the individual claim (loss) at the time t, $t_i = T_1 + T_2 + ... + T_i$. Individual claims amounts have a continuous distribution with distribution function $P(x)$ and the mean (μ).

Since $\{N(t): t \geq 0\}$ and $\{S(t): t \geq 0\}$ are stochastic processes, i.e., families of random variables, for fixed t, N(t) and S(t) we have exactly the set-up of a collective risk model.

Note that the process $\{S(t): t \geq 0\}$ is called the aggregate claims process. It is in fact a compound Poisson process with parameter λ. When we have fixed t, S(t) has a compound Poisson distribution.

The insurer's premium income per unit time to cover the risk is $P = (1+\theta)\lambda\mu$, where θ is a relative security loading factor, λ is Poisson parameter and μ is the mean of individual claims. The classical risk model uses equality $\psi(R_0) = 1/(1+\theta)$.

The probability of ruin is:

$$\psi(R_0) = p\left[\min R(t) < 0, t \geq 0\right]$$

When the surplus become negative for the first time it does not mean that the company becomes bankrupt. The initial capital is interpreted as the capital the company is willing to risk. If ruin occurs at this point, the company has to take action so that the business becomes profitable.

We may also define the surplus process in terms of the aggregate loss process $\{L(t)\}$, where

$$L = \max_{t \geq 0}\{S(t) - Pt\}$$

In this formulation, for a fixed value of t, $L(t)$ represents the accumulated loss up to time t.

For $R_0 \geq 0$ to holds,

$$\begin{aligned} 1-\psi(R_0) &= P\left[R(t) \geq 0, \forall t\right] \\ &= P\left[R_0 + Pt - S(t) \geq 0, \forall t\right] \\ &= P\left[S(t) - Pt \leq R_0, \forall t\right] \end{aligned}$$

Hence left side of the equation is equivalent to $P[L \leq R_0]$ and we get

$$1-\psi(R_0) = p[L \leq R_0]$$

The term on the left side $1-\psi(R_0)$ can be interpreted as a distribution function of random variable L.

Ruin occurs on the first occasion so that the aggregate loss exceeds R_0. Hence, the variable N denotes the number of occasion that the aggregate loss process reaches a new record value. In this context N has geometric distribution with probability function:

$$\begin{aligned} p(N=n) &= \left[1-\psi(0)\right]\left[\psi(0)\right]^n \\ &= \theta\left(\frac{1}{1+\theta}\right)^{n+1} \qquad n = 0,1,2,\ldots \end{aligned}$$

The distribution function of i.i.d. random variables X_i is independent of N has the distribution function:

$$F(x) = \int_0^x \frac{1-p(y)}{\mu} dy$$

Next, we can use Panjer recursion, by replacing $F(x)$ by a discrete function,

$$p[X_1 + X_2 + ... + X_N \le x] = F^{*n},$$

$$F^{0*} = 1 \text{ for } x \ge 0, \text{ otherwise } 0.$$

Hence N has geometric distribution and we get the relationship for $\psi(R_0)$:

$$1-\psi(R_0) = \sum_{n=0}^{\infty} \psi(0)^n (1-\psi(0)).\ F^{n*}(R_0).$$

Now define two discrete distribution which have the probability functions $l(x)$ *and* $u(x)$. Taking the relevant distribution function $L(x)$ and $U(x)$ as:

$$u(x) F(x+h) - F(x) \quad x = 0, h, 2h, ...$$
$$l(x) = F(x) - F(x-h) \quad x = h, 2h, 3h, ...$$

for a very small h, near zero.

From these definitions we have $L(x) \le F(x) \le U(x)$

Then $L^{n*}(x) \le F^{n*}(x) \le U^{n*}(x)$,

where L^{*n} and U^{*n} denote the n- fold convolution. Hence,

$$1-\sum_{n=0}^{\infty} \psi(0)^n (1-\psi(0)) U^{n*} \le \psi(R_0) \le 1-\sum_{n=0}^{\infty} \psi(0)^n (1-\psi(0)) L^{n*}$$

Since $L(x)$ and $U(x)$ are discrete function (discretisation of $F(x)$), we apply Panjer recursion and derive the lower and the upper bounds for the ruin probability ψ, i.e., evaluate values for $\psi(jh)$, $j = 0, 1, ...$

Check Your Progress 2

1) What is Cramer-Lundberg approximation in classical risk theory?

.....

.....

.....

.....

.....

.....

2) How do you explain the individual risk in insurance with gambler's ruin problem?

.....

.....

.....

.....

.....

.....

3) Why martingale is used while modelling individual risk in life insurance?

.....

.....

.....

.....

.....

4) What framework you will use to model the collective risk in life insurance?

.....

.....

.....

.....

.....

5) How convolution method is helpful in collective risk modelling?

.....

.....

.....

.....

.....

6) At which point of collective risk modelling you will make use of Panjer recursion.

.....

.....

.....

.....

9.8 LET US SUM UP

In this unit we have discussed the insurance companies' operational tools for dealing with individual and collective risks. The probability distribution of claims has been developed, which help company's decision-making process to honour its commitments. The individual risk model considers a specific portfolio of insurance policies. The premium collection under such a model is discussed by introducing normal distribution. Necessary formulae for premium setting under expected value of total claims have been highlighted. This line of discussion has been extended to normal and gamma approximations as well as to the class of distributions such as compound Poisson and negative binomial. Classical gambler's ruin probability problem is discussed and risk modelling with martingales has been included. From individual risk model, the analysis is broadened for collective risk model. The collective risk model uses losses as random variables in certain portfolio of insurance polices. Under this theme, we have covered the probability distribution of aggregate claims and compound the ruin probability.

9.9 KEY WORDS

Conditioning: A measure on problem's amenability to digital computation. For example, the condition number associated with the linear equation $Ax = b$ gives a bound indicating degree of the inaccurate the solution given by x. Also, you may think of the condition number as an approximate rate at which the solution, x, will change with respect to a change in b. That is, if the condition number is large, a small error in b may cause a large error in x. On the contrary, with a small condition number, the error in x will not be much bigger than the error in b.

Convolution: A mathematical operator which takes two functions f and g and produces a third function that in a sense represents the amount of develop between f and a reversed and translated version of g. A convolution is a kind of very general moving average as one can see by taking one of the functions to be an indicator function of an interval.

Discretisation: The process of transferring continuous models and equations into discrete counterparts. This process usually carried out as a first step toward making them suitable for numerical evaluation.

Net Level Premium Reserves (or benefit reserves): A premium value of the reserve such that at time zero is it is equal to zero. The net level premium reserve is found by taking the expected value of the loss random variable.

Recursion: A method of defining functions in which the function being defined may be used within its own definition.

Scaling: A linear transformation that enlarges or diminishes objects.

9.10 SOME USEFUL BOOKS

Bowers, Gerber, Hickman, Jones Nesbit (1986), Actuarial Mathematics, The Society of Actuaries, Illinois

Mathias Winkel (2003), Actuarial Science University of Oxford (see Internet)

Mackova, Monika (2005), Selected Recursive Methods of Collective Risk Insurance – Convolution vs. Panjer Recursion, University of Vienna (see Internet)

Trowbridge, C.L (1989), Fundamental concepts of Actuarial science, Actuarial Education and Research Fund (see internet)

Veeh, Jerry Alan (2001), Lecture Notes on Actuarial Mathematics (see Internet)

9.11 ANSWER OR HINTS TO CHECK YOUR PROGRESS

Check Your Progress 1

1) See Section 9.2 and answer.

2) See Section 9.2 and answer.

3) See Section 9.3 and answer.

4) See Section 9.4 and answer.

5) See Section 9.4.3 and answer.

Check Your Progress 2

1) See Section 9.5 and answer.

2) See Section 9.6 and answer.

3) See Section 9.6 and answer.

4) See Section 9.7 and answer.

5) See Section 9.7 and answer.

6) See Section 9.7 and answer.