INSTITUTE OF ACTUARIAL
& QUANTITATIVE STUDIES

Class: SY BSc

Subject : Statistical & Risk Modelling - 2

Chapter: Unit 1 Chapter 2

Chapter Name: Application of compound distribution in risk modelling

Agenda

1. Models for short-term insurance contracts

2. Collective Risk Model
   1. Distribution functions
   2. Finding convolutions
   3. Moments of Compound Distribution – Mean
   4. Moments of Compound Distribution – Variance
   5. Moment Generating Function for Compound Distribution

3. Types of Compound Distributions
   1. Compound Poisson Distribution
   2. Compound Binomial Distribution
   3. Compound Negative Binomial Distribution

4. Approximation of Aggregate Distribution

5. The Individual Risk Model

6. Parameter Variability/ Uncertainty

1 Models for short term insurance contracts

The basic model

Many forms of non-life insurance can be regarded as short-term contracts, for example motor insurance. Some forms of life insurance also fall into this category, for example group life and one-year term assurance policies. A short-term insurance contract can be defined as having the following attributes:

- The policy lasts for a fixed, and relatively short, period of time, typically one year.
- The insurance company receives from the policyholder(s) a premium.
- In return, the insurer pays claims that arise during the term of the policy.

At the end of the policy's term the policyholder may or may not renew the policy. If it is renewed, the premium payable by the policyholder may or may not be the same as in the previous period.

1 Models for short term insurance contracts

An important feature of **a short-term insurance contract is that the premium is set at a level to cover claims** arising during the (short) term of the policy only.
This contrasts with life assurance policies where mortality rates increasing with age mean that the (level) annual premium in the early years would be more than sufficient to cover the expected claims in the early years and insufficient may be in the later years.

Developing the model

A short-term insurance contract covering a risk will be considered. A risk includes either a single policy or a specified group of policies. For ease of terminology the term of the contract is assumed to be one year, but it could equally well be any other short period, for example six months.

The random variable S denotes the aggregate claims paid by the insurer in the year in respect of this risk. Models will be constructed for this random variable S – **The Collective Risk Model.**

Later, in the next chapter, the idea of a collective risk model is extended to an individual risk model.

2 The Collective Risk Model

The collective risk model

Recall from earlier that S is represented as the sum of N random variables X_i, where X_i denotes the amount of the i th claim.

Thus:

$$\boldsymbol{S = X_1 + X_2 + \cdots + X_N}$$

and $S = 0$ if $N = 0$.

S is said to have a compound distribution.

Note that it is the number of claims, N, from the risk as a collective (as opposed to counting the number of claims from individual policies) that is being considered, and this gives the name 'collective risk model'.

Within this framework, expressions in general terms for the distribution function, mean, variance and MGF of S can be developed.

2.1 The Collective Risk Model

Distribution functions

An expression for $G(x)$, the distribution function of S, can be derived by considering the event $\{S \leq x\}$. Note that if this event occurs, then one, and only one, of the following events must occur:

$\{S \leq x \text{ and } N = 0\}$ (ie no claims)

or $\{S \leq x \text{ and } N = 1\}$ (ie one claim of amount $\leq x$)
and so on. These events are mutually exclusive and exhaustive.

Hence:

$$P(S \leq x) = \sum_{n=0}^{\infty} P(S \leq x \text{ and } N = n)$$

$$= \sum_{n=0}^{\infty} P(N = n) P(S \leq x \mid N = n)$$

2.2 The Collective Risk Model

Finding a convolution

$$f_Z(z) = \sum_x f_X(x) f_Y(z-x)$$ for discrete random variables

$$f_Z(z) = \int f_X(x) f_Y(z-x)\,dx$$ for continuous random variables

'Sum or integrate over all values of x that could lead to a total of z.'

2.3 The Collective Risk Model

Moments of compound distributions – Expectation

To calculate the moments of s, conditional expectation results are used, conditioning on the number of claims, N. To find $E[S]$, apply the identity:

$$E[S] = E[E[S \mid N]]$$

Here we are using the conditional expectation formula, which is given on page 16 of the Tables.
Now $E[S \mid N = n] = \sum_{i=1}^{n} E[X_i] = nm_1$. Hence:

$$E[S \mid N] = Nm_1$$

and:

$$\boldsymbol{E[S] = E[Nm_1] = E[N]m_1}$$

2.4 The Collective Risk Model

Moments of compound distributions – Variance

To find an expression for var[S], apply the identity:

$$\text{var}[S] = E[\text{var}[S \mid N]] + \text{var}[E[S \mid N]]$$

Since $E(S \mid N) = Nm_1$, we have:

$$\text{var}[S] = E[\text{var}[S \mid N]] + \text{var}[Nm_1]$$

$\text{var}[S \mid N]$ can be found by using the fact that individual claim amounts are independent.
So $\text{var}[S \mid N] = N(m_2 - m_1^2)$.

Hence: $\text{var}[S] = E[N(m_2 - m_1^2)] + \text{var}[Nm_1]$

ie: $\text{var}[S] = E[N](m_2 - m_1^2) + \text{var}[N]m_1^2$

Alternatively, writing this solely in terms of means and variances:

$$\mathbf{var}(\boldsymbol{S}) = \boldsymbol{E}(\boldsymbol{N})\mathbf{var}(\boldsymbol{X}) + \mathbf{var}(\boldsymbol{N})[\boldsymbol{E}(\boldsymbol{X})]^2$$

2.5 The Collective Risk Model

Moment generating function

The MGF of S is also found using conditional expectation. By definition, $M_S(t) = E[\exp(tS)]$, so:

$$M_S(t) = E[E[\exp(tS) \mid N]]$$

Again, we are conditioning on the number of claims, exactly as we did before.

$$E[\exp(tX_1 + tX_2 + \cdots + tX_n)] = \prod_{i=1}^{n} E[\exp(tX_i)]$$

Also, since $\{X_i\}_{i=1}^{n}$ are identically distributed, they have common MGF, $M_X(t)$, so that:

$$\prod_{i=1}^{n} E[\exp(tX_i)] = \prod_{i=1}^{n} M_X(t) = [M_X(t)]^n$$

Hence:

$$E[\exp(tS) \mid N] = [M_X(t)]^N$$

$$\boldsymbol{M_S(t) = E[M_X(t)^N] = E[\exp(N\log M_X(t))] = M_N(\log M_X(t))}$$

3.1 Types of Compound Distributions

Compound Poisson Distribution

First consider aggregate claims when N has a Poisson distribution with mean λ denoted $N \sim \mathrm{Poi}(\lambda)$.

S then has a compound Poisson distribution with parameter λ, and $F(x)$ is the CDF of the individual claim amount random variable.
S is sometimes referred to as a compound Poisson random variable.

Mean, variance and skewness of a compound Poisson random variable
If $N \sim \mathrm{Poisson}(\lambda)$, then S is a compound Poisson random variable and:

$$\boldsymbol{E(S) = \lambda E(X) = \lambda m_1}$$
$$\boldsymbol{\mathrm{var}(S) = \lambda E\left(X^2\right) = \lambda m_2}$$
$$\boldsymbol{\mathrm{skew}(S) = \lambda E\left(X^3\right) = \lambda m_3}$$

3.1 Types of Compound Distributions

Sums of independent compound Poisson random variables

A very useful property of the compound Poisson distribution is that the sum of independent compound Poisson random variables is itself a compound Poisson random variable.

Let $S_1, S_2, \dots, S_n$ be independent random variables. Suppose that each S_i has a compound Poisson distribution with parameter λ_i, and

Define $A = S_1 + S_2 + \cdots + S_n$. Then A has a compound Poisson distribution with parameter Λ, and $F(x)$ is the CDF of the individual claim amount random variable for A, where:

$$\Lambda = \sum_{i=1}^{n} \lambda_i \quad \text{and} \quad F(x) = \frac{1}{\Lambda} \sum_{i=1}^{n} \lambda_i F_i(x)$$

Recall that Λ is the capital form of the Greek letter λ.

This is a very important result.

3.2 Types of Compound Distributions

Compound Binomial Distribution

Under certain circumstances, the binomial distribution is a natural choice for $\boldsymbol{N}$. For example, under a group life insurance policy covering $\boldsymbol{n}$ lives, the distribution of the number of deaths in a year is binomial if it is assumed that each insured life is subject to the same mortality rate, and that lives are independent with respect to mortality.

The notation $N \sim \mathrm{Bin}(n, p)$ is used to denote the binomial distribution for N.
When N has a binomial distribution, S has a compound binomial distribution.

Mean, Variance and MGF are as follows:

$$\boldsymbol{E[S] = npm_1}$$
$$\mathbf{var}[\boldsymbol{S}] = \boldsymbol{npm_2 - np^2m_1^2}$$

$$M_S(t) = M_N(\log M_X(t))$$
$$\Rightarrow \boldsymbol{M_S(t) = (pM_X(t) + 1 - p)^n}$$

3.3 Types of Compound Distributions

Compound Negative Binomial Distribution

The choice of distribution for $\boldsymbol{N}$ is the negative binomial distribution, which has probability function:

$$P(N = n) = \binom{k + n - 1}{n} p^k q^n \text{ for } n = 0,1,2, \dots$$

This is the Type 2 negative binomial distribution.

When N has a negative binomial distribution, S has a compound negative binomial distribution.

The negative binomial distribution is an alternative to the Poisson distribution for N .
One advantage that the negative binomial distribution has over the Poisson distribution is that its variance exceeds its mean. These two quantities are equal for the Poisson distribution. Thus, the negative binomial distribution may give a better fit to a data set which has a sample variance in excess of the sample mean. This is often the case in practice.

3.3 Types of Compound Distributions

Compound Negative Binomial Distribution

Mean, Variance and MGF are as follows:

$$E[S] = \frac{kq}{p} m_1$$

$$\mathrm{var}[S] = \frac{kq}{p} m_2 + \frac{kq^2}{p^2} m_1^2$$

$$M_S(t) = M_N(\log M_X(t)) \Rightarrow M_S(t) = \frac{p^k}{(1 - qM_X(t))^k}$$

Question

A compound random variable $S = X_1 + X_2 + \cdots + X_N$ has claim number distribution:

$$P(N=n) = 9(n+1)4^{-n-2}, \quad n = 0,1,2,\ldots$$

The individual claim size random variable, X , is exponentially distributed with mean 2.

Calculate E(S) and var(S).

Solution

The probability function of N can be written as:

$$P(N=n)=9(n+1)4^{-n-2}=\binom{n+1}{n}(3/4)^2(1/4)^n$$

We can see from this formula that N has a Type 2 negative binomial distribution with parameters $k=2$ and $p=3/4$.

Hence:

$$E(N)=\frac{kq}{p}=\frac{2\times 1/4}{3/4}=2/3$$

and: $$\text{var}(N)=\frac{kq}{p^2}=\frac{2\times 1/4}{(3/4)^2}=8/9$$

Solution

The individual claim amounts have an exponential distribution with $\lambda = ½$. So the mean and variance of the individual claims are:

$$E(X) = \frac{1}{\lambda} = 2 \quad \text{and} \quad \mathrm{var}(X) = \frac{1}{\lambda^2} = 4$$

Hence:

$$E(S) = E(N)\ E(X) = \frac{2}{3} \times 2 = \frac{4}{3}$$

and: $$\mathrm{var}(S) = E(N)\ \mathrm{var}(X) + \mathrm{var}(N)[E(X)]^2 = \frac{2}{3} \times 4 + \frac{8}{9} \times 2^2 = \frac{56}{9}$$

4 Approximation of Aggregate Distribution

In case of large portfolios or large N, we use an aggregate distribution.

By CLT we approximate the distribution to Normal.

Sometimes, a Lognormal approximation is also applied due to its advantages over normal in loss distribution modelling and insurance modelling.

Question

CT6 April 2015 Q8

The number of claims, N, in a given year on a particular type of insurance policy is given by:

$$P(N = n) = 0.8 \times 0.2^n \qquad n = 0, 1, 2, \ldots$$

Individual claim amounts are independent from claim to claim and follow a Pareto distribution with parameters $\alpha = 5$ and $\lambda = 1{,}000$.

(i) Calculate the mean and variance of the aggregate annual claims per policy. [4]

(ii) Calculate the probability that aggregate annual claims exceed 400 using:

(a) a Normal approximation.

(b) a Lognormal approximation. [6]

(iii) Explain which approximation in part (ii) you believe is more reliable. [2] [Total 12]

Solution

(i) First note that N has a type 2 negative binomial distribution with parameters $p = 0.8$ and $k = 1$. Hence

$$E(N) = \frac{0.2}{0.8} = 0.25$$

$$\text{Var}(N) = \frac{0.2}{0.8^2} = 0.3125$$

Let X denote the distribution of an individual claim. Then

$$E(X) = \frac{\lambda}{\alpha - 1} = \frac{1000}{4} = 250$$

$$\text{Var}(X) = 250^2 \times \frac{5}{3} = 104{,}166.666 = (322.75)^2$$

Now let S denote aggregate annual claims. Then

$$E(S) = E(N)E(X) = 0.25 \times 250 = 62.5$$

$$\begin{aligned} \text{Var}(S) &= E(N)\,\text{Var}(X) + \text{Var}(N)\,E(X)^2 \\ &= 0.3125 \times 250^2 + 0.25 \times 104{,}166.666 \\ &= 45{,}572.92 = 213.478^2 \end{aligned}$$

Solution

(ii) (a) $P(S > 400) = P(N\,(62.5, 213.478^2) > 400)$

$$= P\left(N(0,1) > \frac{400 - 62.5}{213.478}\right)$$

$$= P(N(0,1) > 1.581)$$

$$= 1 - [0.94295 \times 0.9 + 0.1 \times 0.94408]$$

$$= \underline{0.0569}$$

Solution

(b) Let μ and σ be the parameters of the underlying Normal distribution. Then

$$e^{\mu+\frac{\sigma^2}{2}} = 62.5 \qquad \text{(A)}$$

$$e^{2\mu+\sigma^2}(e^{\sigma^2}-1) = 213.478^2 \qquad \text{(B)}$$

$$\text{(B)} \div \text{(A)}^2 \Rightarrow e^{\sigma^2} - 1 = \frac{213.478^2}{62.5^2} = 11.66665$$

$$\sigma^2 = \log 12.66665 = 2.53897 = 1.5934^2$$

substituting into (A) $\mu + \dfrac{2.53897}{2} = \log 62.5$

so $\mu = \log 62.5 - \dfrac{2.53897}{2} = 2.8657$

and so $P(S > 400) = P(N(2.8657,\ 1.5934^2) > \log 400)$

$$= P\left(N(0,1) > \frac{\log 400 - 2.8657}{1.5934}\right)$$

$$= P(N(0,1) > 1.9617)$$

$$= 1 - (0.17 \times 0.97558 + 0.83 \times 0.97500)$$

$$= 0.0249$$

Solution

(iii) The Pareto distribution is significantly skewed and the Normal approximation is not. The Normal approximation in (ii)(b) has variance 213.48^2 and mean 62.5, so negative values of S (which are impossible in reality) are less than 1 standard deviation from the mean.

The approximation in (ii)(b) will therefore be more reliable.

5 The Individual Risk Model

Under this model a portfolio consisting of a fixed number of risks is considered. It will be assumed that:

- these risks are independent
- claim amounts from these risks are not (necessarily) identically distributed random variables
- the number of risks does not change over the period of insurance cover.

As before, aggregate claims from this portfolio are denoted by S. So:

$$\boldsymbol{S = Y_1 + Y_2 + \cdots + Y_n}$$

where Y_j denotes the claim amount under the $\boldsymbol{j}_{\text{th}}$ risk and $\boldsymbol{n}$ denotes the number of risks. It is possible that some risks will not give rise to claims. Thus, some of the observed values of $\{Y_j\}_{j=1}^{n}$ may be 0.

This approach is **referred to as an individual risk model because it is considering the claims arising from each individual policy.**

5 The Individual Risk Model

Assumptions

For each risk, the following assumptions are made:

- the number of claims from the j th risk, $\boldsymbol{N_j}$, is either 0 or 1
- the probability of a claim from the j th risk is q_j.

Above assumptions say that $N_j \sim \text{Bin}(1, q_j)$.

Thus, the distribution of Y_j is compound binomial, with individual claim amount random variable X_j.

5 The Individual Risk Model

Moments and MGF

Suppose that q_j is the probability of a claim from the j th risk. If a claim arises from the j th risk, suppose that the claim amount random variable is X_j. Then:

$$E(S) = \sum_{j=1}^{n} q_j \mu_j$$

$$\text{var}(S) = \sum_{j=1}^{n} \left[q_j \sigma_j^2 + q_j (1 - q_j) \mu_j^2 \right]$$

$$M_S(t) = \prod_{j=1}^{n} \left[q_j M_{X_j}(t) + (1 - q_j) \right]$$

where $\mu_j = E(X_j)$ and $\sigma_j^2 = \text{var}(X_j)$.

If, for a group of n risks, the probability of a claim is fixed and the claim amounts are IID random variables, then the individual risk model is equivalent to a collective risk model where S has a compound binomial distribution with $N \sim \text{Bin}(n, q)$.

6 Parameter variability / uncertainty

So far risk models have been studied assuming that the parameters, that is the moments and, in some cases, even the distributions, of the number of claims and of the amount of individual claims are known with certainty. In general, these parameters would not be known but would have to be estimated from appropriate sets of data.

In this section it will be seen how the models introduced earlier can be extended to allow for parameter uncertainty / variability.

6 Parameter variability / uncertainty

Variability in heterogeneous portfolio

It may be helpful to think of this as a model of part of a motor insurance portfolio. The policies in the whole portfolio have been subdivided according to their values for rating factors such as ‘age of driver’, ‘type of car’ and even ‘past claims experience’. The policies in the part of the portfolio being considered have identical values for these rating factors.

However, there are some factors, such as ‘driving ability’, that cannot easily be measured and so they cannot be taken explicitly into account. It is supposed that some of the policyholders in this part of the portfolio are ‘good’ drivers and the remainder are ‘bad’ drivers.

The individual claim amount distribution is the same for all drivers but ‘good’ drivers make fewer claims (0.1 pa on average) than ‘bad’ drivers (0.3 pa on average). It is assumed that it is known, possibly from national data, that a policyholder in this part of the portfolio is equally likely to be a ‘good’ driver or a ‘bad’ driver but that it cannot be known whether a particular policyholder is a ‘good’ driver or a ‘bad’ driver.

6

Parameter variability / uncertainty

Variability in homogeneous portfolio

Now a different example is considered.
Suppose, as before, there is a portfolio of n policies. The aggregate claims from a single policy have a compound Poisson distribution with parameters λ, and the CDF of the individual claim amounts random variable is $F(x)$. The Poisson parameters are the same for all policies in the portfolio.

If the value of λ were known, the aggregate claims from different policies would be independent of each other. It is assumed that the value of λ is not known, possibly because it changes from year to year, but that there is some indication of the probability that λ will be in any given range of values.

The uncertainty about the value of λ can be modelled by regarding λ as a random variable (with a known distribution).