SCMA 470 : Risk Analysis and Credibility

Pairote Satiracoo

2021-09-23

Chapter 4

Collective Risk Model

Mathematical models of the total amount of claims from a portfolio of policies over a short period of time will be presented in this chapter. The models are referred to as short term risk models. Two main sources of uncertainty including the claim numbers and claim sizes will be taken into consideration. We will begin with the model for aggregate (total) claims or collective risk models.

We define the following random variables:

- S denotes total amount of claims from a portfolio of policies in a fixed time interval, for e.g. one year,
- N represents the number of claims, and
- X_i denotes the amount of the ith claim.

Then the total claims S is given by

$$S = X_1 + ... + X_N.$$

The following assumptions are made for deriving the collective risk model:

1. $\{X_i\}_{i=1}^{\infty}$ are independent and identically distributed with distribution function F_X.
2. N is independent of $\{X_i\}_{i=1}^{\infty}$.

The distribution of the total claim S is said to be a compound distribution. The properties of the compound distribution will be given in the Section 2.

Note The distribution of S can be derived by using convolution technique. In general, the closed form expressions for the compound distribution do not exist so we will mainly concern with the moments of S. For more details about convolution, see Gray and Pitts (2012).

4.1 Conditional expectation and variance formulas

Some useful properties of conditional expectation and conditional variance are given. The conditional expectaion formula is

$$E[E[X|Y]] = E[X].$$

The conditional variance of X given Y is defined to be

$$\begin{aligned} Var[X|Y] &= Var[Z] \text{ where } Z = X|Y \\ &= E[(Z - E[Z])^2] = E[Z^2] - (E[Z])^2 \\ &= E[(X - E[X|Y])^2|Y] \\ &= E[X^2|Y] - (E[X|Y])^2. \end{aligned}$$

The conditional variance formula is

$$Var[X] = E[Var[X|Y]] + Var[E[X|Y]]. \tag{4.1}$$

Example 4.1. Show that

$$Var[X] = E[Var[X|Y]] + Var[E[X|Y]].$$

Solution:

Consider the terms on the right-hand side of (4.1). We have

$$\begin{aligned} E[Var[X|Y]] &= E\left[E[X^2|Y] - (E[X|Y])^2\right] \\ &= E[X^2] - E\left[(E[X|Y])^2\right], \end{aligned}$$

and

$$\begin{aligned} Var[E[X|Y]] &= Var[Z] \text{ where } Z = E[X|Y] \\ &= E[(E[X|Y])^2] - (E[E[X|Y]])^2 \\ &= E[(E[X|Y])^2] - (E[X])^2 \end{aligned}$$

Adding both terms gives the required result.

Example 4.2. In three coloured boxes - Red, Green and Blue, each box has two bags. The bags of Red box contain 1 and 2 (in units of THB) respectively, those of Green box contain 1 and 5, and those of Blue contain 1 and 10 . A box is chosen at random in such a way that Pr(Red) = Pr(Green) = Pr(Blue) = 1/3. A fair coin is tossed to determined which bag to be chosen from the chosen box. Let X be the value of the contents of the chosen bag.

1. Find the distribution of X.
2. Find $E[X]$ and $Var[X]$.
3. Use the conditional expectation and conditional variance formulas to verify your results.

Solution: 1. The distribution of X can be obtained by using the law of total probability: for example

$$\begin{aligned} P(X=1) &= P(X=1,R)+P(X=1,G)+P(X=1,B) \\ &= P(X=1|R)\cdot P(R)+P(X=1|G)\cdot P(G)+P(X=1|B)\cdot P(B) \\ &= \frac{1}{2}\cdot\frac{1}{3}+\frac{1}{2}\cdot\frac{1}{3}+\frac{1}{2}\cdot\frac{1}{3}=\frac{1}{2}. \end{aligned}$$

Similarly, we have

$$P(X=1)=\frac{1}{2}, \quad P(X=2)=P(X=5)=P(X=10)=\frac{1}{6}.$$

2. It follows that

$$E[X]=\frac{10}{3}, \quad Var[X]=\frac{98}{9}.$$

3. We first calculate

$$\begin{aligned} E[X|R] &= \frac{1}{2}\cdot(1+2)=\frac{3}{2} \\ E[X|G] &= \frac{1}{2}\cdot(1+5)=3 \\ E[X|B] &= \frac{1}{2}\cdot(1+10)=\frac{11}{2}. \end{aligned}$$

We have

$$\begin{aligned} E[X] &= E[X|R]\cdot P(R)+E[X|G]\cdot P(G)+E[X|B]\cdot P(B) \\ &= \frac{1}{3}\cdot(\frac{3}{2}+3+\frac{11}{2})=\frac{10}{3}. \end{aligned}$$

4.2 The moments of a compound distribution S

The moments and moment generating function of S can be easily derived from the conditional expectation formula.

4.2.1 The mean of S

Let m_k be the kth moment of X_1, i.e. $E[X_1^k] = m_k$. Conditional on $N = n$, we have

$$E[S|N = n] = E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i] = nE[X_i] = n \cdot m_1.$$

Hence, $E[S|N] = Nm_1$ and

$$E[S] = E[E[S|N]] = E[Nm_1] = E[N]m_1 = E[N] \cdot E[X_1].$$

It is no surprise that the mean of the total claims is the product of the means of the number of claims and the mean of claim sizes.

4.2.2 The variance of S

Using the fact that $\{X_i\}_{i=1}^{\infty}$ are independent, we have

$$Var[S|N=n] = Var[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} Var[X_i] = nVar[X_i] = n(m_2 - m_1^2),$$

and $Var[S|N] = N(m_2 - m_1^2)$. It follows that

$$\begin{aligned} Var[S] &= E[Var[S|N]] + Var[E[S|N]] \\ &= E[N(m_2 - m_1^2)] + Var[Nm_1] \\ &= E[N](m_2 - m_1^2) + Var[N]m_1^2. \end{aligned}$$

Example 4.3. Show that $M_S(t) = M_N(\log(M_X(t)))$.

Solution:

First, consider the following conditional expectation:

$$\begin{aligned} E\left[e^{tS}|N=n\right] &= E\left[e^{t(X_1+X_2+\cdots X_n)}\right] \\ &= E\left[e^{tX_1}\right]\cdot E\left[e^{tX_2}\right]\cdots E\left[e^{tX_n}\right], \text{ since } X_1, X_2 \ldots, X_n \text{ are independen} \\ &= (M_X(t))^n. \end{aligned}$$

Hence $E\left[e^{tS}|N\right] = (M_X(t))^N$.

From the definition of the moment generating function,

$$\begin{aligned} M_S(t) &= E[e^{tS}] \\ &= E\left[E[e^{tS}|N]\right] \\ &= E\left[(M_X(t))^N\right] \\ &= E\left[Exp(N\cdot\log(M_X(t))\right] \\ &= M_N(\log(M_X(t)))(\text{ since } M_X(t) = E[e^{tX}]). \end{aligned}$$

4.3 Special compound distributions

4.3.1 Compound Poisson distributions

Let N be a Poisson distribution with the parameter λ, i.e. $N \sim Poisson(\lambda)$ and $\{X_i\}_{i=1}^{\infty}$ are independent and identically distributed with distribution function F_X. Then $S = X_1 + ... + X_N$ is said to have a compound Poisson distribution and denote by $\mathcal{CP}(\lambda, F_X)$.

Note The same terminology can be defined similarly for other distributions, for e.g. if N has a negative binomial distribution, then S is said to have a compound negative binomial distribution.

Example 4.4. Let $S \sim \mathcal{CP}(\lambda, F_X)$. Show that

1. $E[S] = \lambda m_1$,
2. $Var[S] = \lambda m_2$,
3. $M_S(t) = Exp(\lambda(M_X(t) - 1))$.
4. The third central moment $E[(S - E[S])^3] = \lambda m_3$, and hence
$$Sk[S] = \frac{\lambda m_3}{(\lambda m_2)^{3/2}},$$

where m_k be the kth moment of X_1

Solution: 1. $E[S] = E[N] \cdot E[X] = \lambda m_1$,

2. $Var[S] = E[N](m_2 - m_1^2) + Var[N]m_1^2 = \lambda(m_2 - m_1^2) + \lambda m_1^2 = \lambda m_2$,

3. From

$$\begin{aligned}
M_S(t) &= M_N(\log(M_X(t))) \\
&= Exp\left(\lambda\left(e^{\log(M_X(t))} - 1\right)\right), \text{ since } M_N(t) = Exp(\lambda(e^t - 1)) \\
&= Exp(\lambda(M_X(t) - 1)).
\end{aligned}$$

4. The third central moment $E[(S - E[S])^3] = \lambda m_3$, and hence

$$Sk[S] = \frac{\lambda m_3}{(\lambda m_2)^{3/2}}.$$

In particular, we have

$$\begin{aligned}
E[(N - E[N])^3] &= E\left[N^3 - 3N^2 \cdot E[N] + 3N \cdot (E[N])^2 - (E[N])^3\right] \\
&= E[N^3] - 3E[N^2] \cdot E[N] + 2(E[N])^3 \\
&= M_N'''(0) - 3M_N''(0) \cdot M_N'(0) + 2(M_N'(0))^3
\end{aligned}$$

For $N \sim Poisson(\lambda)$, $M_N(t) = Exp(\lambda(e^t - 1))$. By differentiating $M_N(t)$ and evaluating at $t = 0$, we can show that

$$M'(0) = \lambda, \quad M''(0) = \lambda(1 + \lambda), \quad M'''(0) = \lambda(1 + 3\lambda + \lambda^2).$$

Hence, $E[(N - E[N])^3] = \lambda$.

Similarly,

$$E[(S - E[S])^3] = E[S^3] - 3E[S^2] \cdot E[S] + 2(E[S])^3$$

In addition, $M_S(t) = Exp(\lambda(M_X(t) - 1))$. By differentiating $M_S(t)$ we can show that

$$M_S'''(t) = \lambda M_X'''(t) M_S(t) + 2\lambda M_X''(t) M_S'(t) + \lambda M_X'(t) M_S''(t).$$

Evaluating $M_S'''(t)$ at $t = 0$ results in

$$M_S'''(0) = E[S^3] = \lambda m_3 + 3E[S] \cdot E[S^2] - 2(E[S])^3,$$

which gives

$$\begin{aligned} E[(S - E[S])^3] &= E[S^3] - 3E[S^2] \cdot E[S] + 2(E[S])^3 \\ &= \lambda m_3. \end{aligned}$$

Example 4.5. Let S be the aggregate annual claims for a risk where $S \sim \mathcal{CP}(10, F_X)$ and the individual claim amounts have a $Pa(4, 1)$ distribution. Calculate $E[S], Var[S]$ and $Sk[S]$.

Solution: Since $X \sim Pa(4, 1)$ with $\alpha = 4$ and $\lambda = 1$, we have

$$
\begin{aligned}
E[X^r] &= \frac{\Gamma(\alpha - r) \cdot \Gamma(1 + r) \cdot \lambda^r}{\Gamma(\alpha)} \\
E[X] &= \frac{\lambda}{\alpha - 1} = \frac{1}{4 - 1} = \frac{1}{3} \\
E[X^2] &= \frac{\Gamma(2) \cdot \Gamma(3) \cdot \lambda^2}{\Gamma(4)} = \frac{1}{3} \\
E[X^3] &= \frac{\Gamma(1) \cdot \Gamma(4) \cdot \lambda^3}{\Gamma(4)} = 1.
\end{aligned}
$$

We have

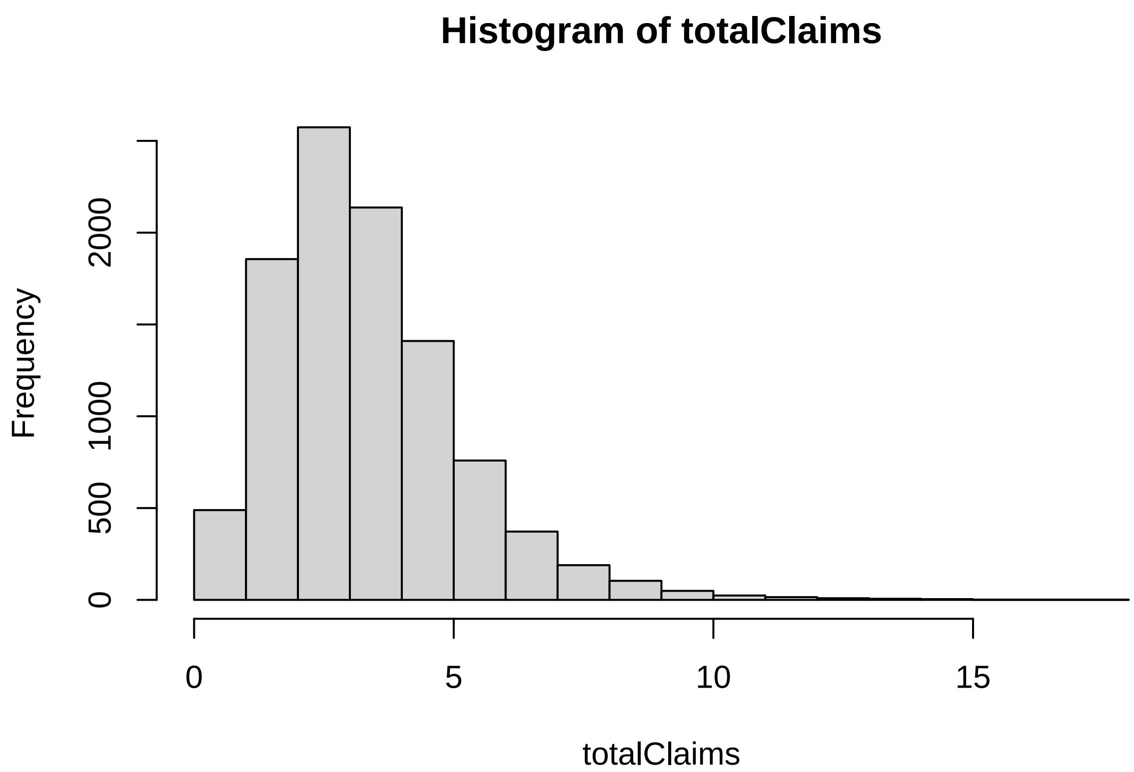
$$E[S] = \lambda E[X] = \frac{10}{3}$$
$$Var[S] = \lambda E[X^2] = \frac{10}{3}$$
$$Sk[S] = \frac{\lambda E[X^3]}{\left(\lambda E[X^2]\right)^{3/2}} = \frac{10}{(10/3)^{3/2}} = 1.6432.$$

In what follows, we will use R to simulate n observations from a compound Poisson distribution, where the Poisson parameter is λ and where the claims are exponentially distributed with mean μ, i.e. $CP(\lambda, Exp(1/\mu))$

```
# Simulation n observations from a  CP(lambda,FX) distribution
# Assumptions:
# N ~ Poisson(lambad)
# X ~ Pa(alpha,beta)
library(actuar)
```

```r
n <- 10000
lambda <-  10
alpha <- 4
beta <- 1

totalClaims <- rep(0,n)
numclaims <- rpois(n,lambda)
for (i in 1:n)
  totalClaims[i] <- sum(rpareto(numclaims[i], shape =alpha, scale =
hist(totalClaims)
```




```
mean(totalClaims)
```

```
## [1] 3.338976
```

```
var(totalClaims)
```

```
## [1] 3.29011
```

```r
library(moments)
skewness(totalClaims)
```

```
## [1] 1.345385
```

Note An important property of independent, but not necessarily identically distributed, compound Poisson random variables is that the sum of a fixed number of them is also a compound Poisson random variable.

Example 4.6. Let $S_1, \dots, S_n$ be independent compound Poisson random variables, with parameters λ_i and F_i. Then $S = \sum_{i=1}^{n} S_i$ has a compound distribution with parameter

$$\lambda = \sum_{i=1}^{n} \lambda_i,$$

and

$$F = \frac{1}{\lambda} \sum_{i=1}^{n} \lambda_i F_i.$$

Solution: Exercise.

Note The compound Poisson distribution is the most often used in practice. It possesses the additivity of independent compound Poisson distributions (as shown in Example 4.6, and the expressions of the first three moments are very simple.

4.3.2 Compound negative binomial distributions

A useful discrete random variable that can be used for modelling the distributions of claim numbers is a negative binomial distribution. A random variable N has a negative distribution with parameters k and p, denoted by $N \sim NB(k, p)$ if its probability mass function is given by

$$f_N(n) = \Pr(N = n) = \frac{\Gamma(k+n)}{\Gamma(n+1)\Gamma(k)} p^k (1-p)^n \quad n = 0, 1, 2, \ldots .$$

It can be interpreted as the probability of getting n failures before the kth success occurs in a sequence of independent Bernoulli trials with probability of success p.

Example 4.7. Let $N \sim NB(k,p)$. Show that the mean, variance and moment generating function of the compound negative binomial distribution, denoted by $\mathcal{CNB}(k,p,F_X)$, are as follows:

1. $E[S] = \frac{kq}{p} m_1$,
2. $Var[S] = \frac{kq}{p^2}(pm_2 + qm_1^2)$,
3. $M_S(t) = \left(\frac{p}{1-qM_X(t)}\right)^k$,

where m_k be the kth moment of X_1 and $q = 1-p$.

Solution: The results follows from the properties of the negative binomial distribution $N \sim NB(k,p)$:

$$E[N] = \frac{kq}{p}, \quad Var[N] = \frac{kq}{p^2},$$

and the moments of a compound distribution S derived in Section 4.2.

Notes 1. The negative binomial distribution is an alternative to the Poisson distribution for N, in the sense that it allows for any value of $N = 0, 1, 2, ...$, unlike the binomial distribution which has an upper limit.

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.

2. The compound negative binomial distribution is an appropriate to model the heterogeneity of the numbers of claims occurring for different risks. In particular, suppose that for each policy, the number of claims in a year has a Poisson distribution $N|\lambda \sim Poisson(\lambda)$, and that the variation in λ across the portfolio can be modelled using a Gamma distribution $\mathcal{G}(\alpha, \lambda)$. Then the number of claims in the year for a policy chosen at random from the portfolio has a negative binomial distribution.

4.3.2.1 Misture distributions

Suppose we model a policyholder's claim number N using a conditional distribution $N|\lambda$, where λ can be thought of as a "risk parameter" for that policyholder.

Policyholders represent a variety of risks and have different risk parameters, and we model the variation across policyholders by regarding the various λs as being independent realisations of a random variable with known probability distribution. This gives the joint density, which we can write as $f_{N,\lambda}(k, \lambda) = f_\lambda(\lambda) f_{N|\lambda}(k|\lambda)$.

This enables us to allow for variability in the risks across a portfolio; that is, to model the heterogeneity of the numbers of claims occurring for different risks.

Example 4.8. A portfolio consists of a large number of individual policies. For each policy, the number of claims in a year has a poisson distribution $N|\lambda \sim Poisson(\lambda)$. Let us suppose that the variation in λ across the portfolio of risks can be modelled using a gamma $\mathcal{G}(\alpha, \beta)$ distribution with known parameters, and let us use this to average across the risks.

We are considering a **mixture** of Poissons where the mixing distribution if gamma. This is also known as a mixture **distribution**. Derive the probability mass function of the mixture distribution.

Solution:

For $k = 0, 1, 2, ...,$ we have

$$
\begin{aligned}
\Pr(N=k) &= \int f_\lambda(\lambda)\Pr(N=k|\lambda)\,d\lambda \\
&= \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{\beta\lambda}e^{-\lambda}\frac{\lambda^k}{k!}\,d\lambda \\
&= \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)\Gamma(1+k)}\frac{\beta^\alpha}{(\beta+1)^{\alpha+k}} \times \int_0^\infty h(\lambda)\,d\lambda
\end{aligned}
$$

where $h(\lambda)$ is the probability density function of $\lambda \sim \mathcal{G}(\alpha+k, \beta+1)$.

Hence,

$$
\Pr(N=k) = \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)\Gamma(1+k)}\left(\frac{\beta}{\beta+1}\right)^\alpha\left(\frac{1}{\beta+1}\right)^k, k=0,1,2,\dots,
$$

which is the probability mass function of a $\mathcal{NB}(\alpha, \beta/(\beta+1))$ distribution.

This provides an illuminating view of the negative binomial distribution – it arises as a mixture of Poissons where the mixing distribution is gamma.

4.3.2.2 an Example in R

This R Markdown introduces the concept of mixture distributions which applies to models for claim numbers.

Suppose we model a policyholder's claim numbers N using a conditional distribution $N|\lambda$, where λ can be thought of as a **risk parameter** for that policyholder. Policyholders represent a variety of risks and have different risk parameters, and we model the variation across policyholders by regarding the various λs as being independent realisations of a random variable with known probability distribution.

The following R code produces the required n simulated values from this mixture distribution, where $N|\lambda \sim Poisson(\lambda)$ with mixing distribution $\mathcal{G}(\alpha, \beta)$, i.e. $\lambda \sim \mathcal{G}(\alpha, \beta)$.

eyJsYW5ndWFnZSI6InIiLCJzYW1wbGUiOiJzZXQuc2VlZCg1MzUzKVxubiA8LSA1M

4.3.3 Compound binomial distributions

A compound binomial distribution can be used to model a portfolio of policies, each of which can give rise to at most one claim.

Example 4.9. Consider a portfolio of n independent and identical policies where there is at most one claim on each policy in a year (for e.g. life insurance). Let p be the probability that a claim occurs. Explain that the aggregate sum S in this portfolio has a compound binomial distribution, denoted by $\mathcal{CB}(n, p, F_X)$. Derive the mean, variance and moment generating function of S.

Solution: Since n policies (lives) are independent with the probability p that a claim occurs, the number N of claims on the portfolio in one year has a binomial distribution i.e. $N \sim \mathrm{bi}(n, p)$. If the sizes of the claims are i.i.d. random variables, independent of N, then the total amount S claimed on this policy in one year has a compound binomial distribution.

The mean, variance and the moment generating function of S are as follows:

$$\begin{aligned} E[S] &= npm_1, \\ Var[S] &= npm_2 - np^2m_1^2, \\ M_S(t) &= (q + pM_X(t))^n, \end{aligned}$$

where m_k be the kth moment of X_1 and $q = 1 - p$.

4.4 The effect of reinsurance

The effect of reinsurance arrangements on an aggregate claims distribution will be presented. Let S denotes the total aggregate claims from a risk in a given time, S_I and S_R denote the insurance and reinsurance aggregate claims, respectively. It follows that

$$S = S_I + S_R.$$

4.4.1 Proportional reinsurance

Recall that under proportional reinsurance arrangement, a fixed proportion α is paid by the direct insurer and the remainder of the claim is paid by the reinsurer. It follows that

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

and

$$S_R = \sum_{i=1}^{N} (1-\alpha) X_i = (1-\alpha) S,$$

where X_i is the amount of the ith claim.

Notes

1. Both direct insurer and the reinsurer are involved in paying each claim.
2. Both have unlimited liability unless a cap on the claim amount is arranged.

Example 4.10. Aggregate claims from a risk in a given time have a compound Poisson distribution with Poisson parameter $\lambda = 10$ and an individual claim amount distribution that is a Pareto distribution, $Pa(4, 1)$. The insurer has effected proportional reinsurance with proportion retained $\alpha = 0.8$.

1. Find the distribution of S_I and S_R and their means and variances.
2. Compare the variances $Var[S_I] + Var[S_R]$ and $Var[S]$. Comment on the results obtained.

Solution: 1. We have

$$S_I = \sum_{i=1}^{N} (\alpha X_i) = \alpha \sum_{i=1}^{N} X_i = \alpha \cdot S,$$

$$S_R = \sum_{i=1}^{N} ((1-\alpha) X_i) = (1-\alpha) \sum_{i=1}^{N} X_i = (1-\alpha) \cdot S,$$

since both insurer and reinsurer are involved in paying each claim, i.e. $Y_i = \alpha X_i$ and $Z_i = (1-\alpha)X_i$. It follows that $S_I \sim \mathcal{CP}(10, F_Y)$ and $S_R \sim \mathcal{CP}(10, F_Z)$.

2. As we can show that if $X \sim Pa(\beta, \lambda)$, then $W = kX \sim Pa(\beta, k\lambda)$.

For $X \sim Pa(4,1)$ with $\beta = 4$ and $\lambda = 1$, we have $Y_i \sim Pa(\beta, \alpha \cdot \lambda) = Pa(4, 0.8)$ and

$$\begin{aligned}
E[S_I] &= 10 \cdot E[Y_i] = 10 \cdot \frac{\alpha \cdot \lambda}{\beta - 1} \\
&= \frac{8}{3}, \\
Var[S_I] &= 10 \cdot E[Y_i^2] = 10 \cdot \frac{\Gamma(\beta - 2) \cdot \Gamma(1+2) \cdot (\alpha \cdot \lambda)^2}{\Gamma(\beta)} \\
&= 10 \cdot \frac{\Gamma(2) \cdot \Gamma(3) \cdot (\alpha \cdot \lambda)^2}{\Gamma(4)} = 10 \cdot \frac{2!}{3!} \cdot (0.8)^2 \\
&= \frac{32}{15}
\end{aligned}$$

Alternatively, we can calculate by using the properties of the expectation and variance as follows:

$$\begin{aligned} E[S_I] &= E[\alpha S] = \alpha \cdot E[S] = \alpha \cdot \lambda \cdot E[X] = 10 \cdot 0.8 \cdot \frac{1}{3} = \frac{8}{3}, \\ Var[S_I] &= Var[\alpha S] = \alpha^2 \cdot Var[S] \\ &= \alpha^2 \cdot \lambda \cdot E[X^2] = \frac{32}{15}. \end{aligned}$$

Similarly,

$$\begin{aligned} E[S_R] &= E[(1-\alpha)S] = (1-\alpha) \cdot E[S] = \frac{2}{3}, \\ Var[S_R] &= Var[(1-\alpha)S] = (1-\alpha)^2 \cdot Var[S] = \frac{2}{15}. \end{aligned}$$

Note that $E[S_I]+E[S_R] = E[S]$, while $Var[S_I]+Var[S_R] = \frac{34}{15} < Var[S] = \frac{10}{3}$.

4.4.2 Excess of loss reinsurance

Recall that under excess of loss reinsurance arrangement, the direct insurer has effected excess of loss reinsurance with retention level $M > 0$. For a claim X,

- the insurance company pays any claim in full if $X \leq M$; and
- the reinsurer (or reinsurance company) pays the remaining amount of $X - M$ if $X > M$.

It follows that

$$S_I = \sum_{i=1}^{N} Y_1 + Y_2 + ... + Y_N = \sum_{i=1}^{N} \min(X_i, M) \tag{4.2}$$

and

$$S_R = \sum_{i=1}^{N} Z_1 + Z_2 + ... + Z_N = \sum_{i=1}^{N} \max(0, X_i - M), \tag{4.3}$$

where X_i is the amount of the ith claim. When $N = 0$, we set $S_I = 0$ and $S_R = 0$.

Note S_R can equal 0 even if $N > 0$. This occurs when all claims do not exceed M and hence the insurer pays the full amounts of claims.

As discussed in the previous section, the reinsurer is involved only claims which exceed the retention limit (a claim such that $X > M$). Such claims are called **reinsurance claims**. Taking in account of counting only non-zero claims, we can rewrite S_R as follows. Let N_R be the number of insurance (non-zero) claims for the reinsurer and W_i be the amount of the ith non-zero payment by the reinsurer. The aggregate claim amount paid by the reinsurer can be written as

$$S_R = \sum_{i=1}^{N_R} W_i.$$

Example 4.11. By using the probability generating function, show that if $N \sim Poisson(\lambda)$, then the distribution $N_R \sim Poisson(\lambda\pi_M)$ where $\pi_M = \Pr(X_j > M)$.

Solution: Define the indicator random variable $\{I_j\}_{j=1}^{\infty}$, where

$$I_j = \begin{cases} 1 & \text{if } X_j > M \\ 0 & \text{if } X_j \le M. \end{cases}$$

Therefore,

$$N_R = \sum_{j=1}^{N} I_j.$$

The variable N_R has a compound distribution with its probability generating function

$$P_{N_R}(r) = P_N[P_I(r)],$$

where P_I is the probability generating function of the indicator random variable. It can be shown that

$$P_I(r) = 1 - \pi_M + \pi_M r,$$

where $\pi_M = \Pr(I_j = 1) = \Pr(X_j > M) = 1 - F(M)$.

Note In the above example, one can derive the distribution of N_R by using the moment generating function:

$$M_{N_R}(t) = M_N(\log M_I(t)),$$

where M_N and M_I are the moment generating functions of N and I. Note also that

$$M_I(t) = 1 - \pi_M + \pi_M Exp(t).$$

4.4.3 Compound Poisson distributions under excess of loss reinsurance

Assume that aggregate claim amount $S \sim \mathcal{CP}(\lambda, F_X)$ has a compound Poisson distribution. Under excess of loss reinsurance with retention level M, it follows from (4.2) and (4.3) that

1. $S_I \sim \mathcal{CP}(\lambda, F_Y)$,

   where $f_Y(x) = f_X(x)$ for $0 < x < M$ and $\Pr(Y = M) = 1 - F_X(M)$.

2. $S_R \sim \mathcal{CP}(\lambda, F_Z)$,

   where $F_Z(0) = F_X(M)$ and $f_Z(x) = f_X(x + M), x > 0$.

3. Excluding zero claims, $S_R \sim \mathcal{CP}(\lambda\,(1 - F_X(M)), F_W)$,

   where $f_W(x) = \dfrac{f_X(x + M)}{1 - F_X(M)}, x > 0$.

Example 4.12. Suppose that S has a compound Poisson distribution with Poisson parameters $\lambda = 10$ and the claim sizes have the following distribution

x	1	2	5	10
$\Pr(X = x)$	0.4	0.3	0.2	0.1

The insurer enters into an excess of loss reinsurance contract with retention level $M = 4$.

1. Show that $S_I \sim \mathcal{CP}(\lambda, F_Y)$.
2. Show that $S_R \sim \mathcal{CP}(\lambda, F_Z)$.
3. By excluding zero claims, show that the S_R can also be expressed as $S_R \sim \mathcal{CP}(\lambda p, F_W)$ where $p = \Pr(X > M)$.
4. Find the mean and variance of the aggregate claim amount for both insurer and reinsurer.

Solution:

1. Recall that $S_I = \sum_{i=1}^{N} \min\{X_i, 4\}$. The number of claim remains the same, and hence $N \sim Poisson(10)$. The distribution of claim amount paid by the insurer, $F_Y(x)$ is given by

x	1	2	4
$\Pr(Y = x)$	0.4	0.3	0.3

Therefore, $S_I \sim \mathcal{CP}(10, F_Y)$ and

$$\begin{aligned} \mathrm{E}[S_I] &= 10E[Y] = 10(1(0.4) + 2(0.3) + 4(0.3)) = 22, \\ \mathrm{Var}[S_I] &= 10E[Y] = 10(1^2(0.4) + 2^2(0.3) + 4^2(0.3)) = 64. \end{aligned}$$

2. We have $S_R = \sum_{i=1}^{N} \max\{0, X_i - 4\}$. When zero claims are included, $N_R = N \sim Poisson(10)$. The distribution of claim amount paid by the reinsurer, $F_Z(x)$ is given by

x	0	1	6
$\Pr(Z = x)$	0.7	0.2	0.1

Therefore, $S_R \sim \mathcal{CP}(10, F_Z)$ and

$$\begin{aligned} \mathrm{E}[S_R] &= 10E[Z] = 0, \\ \mathrm{Var}[S_R] &= 10E[Z^2] = 38. \end{aligned}$$

Notes

a. $\mathrm{E}[S] = 10\mathrm{E}[X] = 30$ and $\mathrm{Var}[S] = 10\mathrm{E}[X^2] = 166$.

b. $\mathrm{E}[S_I + S_R] = \mathrm{E}[S]$, and

$$\mathrm{Var}[S_I + S_R] = 64 + 38 < 166 = \mathrm{Var}[S].$$

3. Consider the reinsurer's position when zero claims are excluded. We define

$$W = Z|Z > 0 = X - 4|X > 4.$$

We first compute π_M, the proportion of claims which involve the reinsurer, from

$$\pi_M = \Pr(X > 4) = \Pr(X = 5) + \Pr(X = 10) = 0.3.$$

Recall that $S_R = \sum_{i=1}^{N_R} W_i$. We have $S_R \sim \mathcal{CP}(0.3 \times 10, F_W)$ and the distribution of W, $F_W(x)$ is given by

$$\Pr(W = 1) = \Pr(X = 5|X > 4) = \frac{\Pr(X = 5, X > 4)}{\Pr(X > 4)} = \frac{\Pr(X = 5)}{\Pr(X > 4)} = \frac{2}{3}$$
$$\Pr(W = 6) = 1 - \Pr(W = 1) = \frac{1}{3}.$$

Hence,

$$\mathrm{E}[S_R] = (10 \times 0.3)(1(2/3) + 6(1/3)) = 8,$$
$$\mathrm{Var}[S_R] = (10 \times 0.3)(1^2(2/3) + 6^2(1/3)) = 38.$$

Example 4.13. Suppose that S has a compound Poisson distribution with Poisson parameters $\lambda = 40$ and the claim sizes have a Pareto distribution $Pa(3, 4)$. The insurer has an excess of loss reinsurance contract in place with retention level $M = 2$. Find the mean and variance of the aggregate claim amount for both insurer and reinsurer.

Solution: 1. **Zero claims included**. Recall that that for $X \sim \mathcal{P}a(\alpha, \lambda)$, its density function is

$$f_X(x) = \frac{\alpha\lambda^\alpha}{(x+\lambda)^{\alpha+1}}.$$

We know that $S_R \sim \mathcal{CP}(40, F_Z)$. Moreover,

$$
\begin{aligned}
\mathrm{E}[Z] &= 0F_Z(0) + \int_0^\infty x f_Z(x)\,dx \\
&= \int_0^\infty x f_X(x+M)\,dx \\
&= \int_0^\infty \frac{x \cdot 3 \cdot 4^3}{(x+2+4)^{3+1}}\,dx \\
&= \frac{4^3}{6^3}\int_0^\infty \frac{x \cdot 3 \cdot 6^3}{(x+6)^{3+1}}\,dx \\
&= \frac{4^3}{6^3}\frac{6}{3-1} = 0.88\ldots.
\end{aligned}
$$

Note that the last integral above is the mean of $\mathcal{P}a(3,6)$, which is equal to $6/(3-1)$.

For $\mathrm{E}[Z^2]$, we proceed as follows:

$$\begin{aligned}
\mathrm{E}[Z^2] &= 0^2 F_Z(0) + \int_0^\infty x^2 f_Z(x)\,dx \\
&= \int_0^\infty x^2 f_X(x+M)\,dx \\
&= \int_0^\infty \frac{x^2 \cdot 3 \cdot 4^3}{(x+2+4)^{3+1}}\,dx \\
&= \frac{4^3}{6^3}\int_0^\infty \frac{x^2 \cdot 3 \cdot 6^3}{(x+6)^{3+1}}\,dx \\
&= \frac{4^3}{6^3} 6^2 = 10.66\ldots.
\end{aligned}$$

Note that the last integral above is the second moment about the origin of $\mathcal{P}a(3,6)$, which is equal to $[6^2 \cdot \Gamma(3-2)\Gamma(1+2)]/\Gamma(3) = 6^2$.

Therefore,

$$\mathrm{E}[S_R] = \lambda E[Z] = 320/9,$$
$$\mathrm{Var}[S_R] = \lambda E[Z^2] = 1280/3.$$

2. **Zero claims excluded** We define

$$W = Z|Z > 0 = X - M|X > M.$$

$$\begin{aligned}
\mathrm{E}[W] &= \int_0^\infty x f_W(x)\, dx \\
&= \int_0^\infty \frac{x f_X(x+2)}{(1 - F_X(2))}\, dx \\
&= \frac{1}{(1 - F_X(2))} \int_0^\infty x f_X(x+2)\, dx \\
&= \frac{1}{(1 - F_X(2))} \cdot \mathrm{E}[Z].
\end{aligned}$$

It follows that

$$\mathrm{E}[S_R] = \lambda \cdot \Pr(X > M) \cdot \mathrm{E}[W] = 40(1 - F_X(2))\mathrm{E}[W] = 40\mathrm{E}[Z] = 320/9.$$

Similarly, one can show that

$$\begin{aligned}
\mathrm{E}[W^2] &= \int_0^\infty x^2 f_W(x)\,dx \\
&= \frac{1}{(1 - F_X(2))} \cdot \mathrm{E}[Z^2].
\end{aligned}$$

This results in

$$\mathrm{Var}[S_R] = \lambda \cdot \Pr(X > M) \cdot \mathrm{E}[W^2] = 40(1 - F_X(2))\mathrm{E}[W^2] = 40\mathrm{E}[Z^2] = 1280/3.$$

3. Note that $S = S_I + S_R$ and

$$\mathrm{E}[S] = \lambda \mathrm{E}[X] = 40\frac{4}{3-1} = 80.$$

Therefore,

$$\mathrm{E}[S_I] = 80 - \frac{320}{9} = \frac{400}{9}.$$

4.5 Approximation of the collective risk model

4.5.1 The normal approximation

According to the Central Limit Theorem, if the mean number of claims is large, then the distribution of aggregate claims S can be approximated by a normal distribution, i.e. $S \sim \mathcal{N}(\mathrm{E}[S], \mathrm{Var}[S])$.

Notes

1. The normal approximation may not provide a good approximation to the distribution of S because the true distribution of S is skew. However, the normal approximation is symmetric.

2. The normal approximation is likely to underestimate tail probabilities which are the most interest quantities of insurers.

Example 4.14. Aggregate claims from a risk in a given time have a compound Poisson distribution with Poisson parameter λ and an individual claim amount distribution that is a lognormal distribution with mean 1 and variance 2.5.

1. Approximate the distribution of S using the normal distribution when (a) $\lambda = 10$ and (b) $\lambda = 100$.
2. Find x such that $\Pr(S \leq x) = 0.95$ in both cases.
3. Comment on the obtained results.

4.5.2 The translated gamma approximation

The translated gamma approximation makes use of the first three moments of S and provides an improvement of the approximation over the normal approximation. We assume that S can be approximated by $Y + k$ where $Y \sim \mathcal{G}(\alpha, \lambda)$ and k is a constant. This distribution $Y + k$ is said to have a translated gamma distribution. By matching the moments of the two distribution, the parameters

α, λ and k can be found from

$$
\begin{aligned}
Sk[S] &= \frac{2}{\sqrt{\alpha}}, \\
Var[S] &= \frac{\alpha}{\lambda^2}, \\
E[S] &= \frac{\alpha}{\lambda} + k.
\end{aligned}
$$

Example 4.15. Show that the parameters α, λ and k satisfy

$$\alpha = \frac{4}{Sk[S]^2},$$
$$\lambda = \sqrt{\frac{\alpha}{Var[S]}}$$
$$k = E[S] - \frac{\alpha}{\lambda}.$$

Example 4.16. The aggregate claims S have the compound Poisson distribution as given in Example 4.14.

1. Use the translated gamma approximation to find x such that $\Pr(S \leq x) = 0.95$ when (a) $\lambda = 10$ and (b) $\lambda = 100$.
2. Comment on the obtained results.

4.6 Recursive calculation of the collective risk model

The Panjer recursion formula provides recursive calculation of the collective risk model. The algorithm can be numerically computed on a computer provided that distribution of claim numbers N satisfy Panjer's recursion formula,

$$p_n = \left(a + \frac{b}{n}\right) p_{n-1}, \quad n = 1, 2, \dots,$$

where a and b are constants.

Example 4.17. Show that a Poisson distribution $N \sim Poisson(\lambda)$ satisfies Panjer's recursion formula, i.e. find the constants a and b.

Assume that the claim size variable X takes only **positive integers** and the distribution of claim numbers satisfies the Panjer's recursion formula. We define

- $f_k = \Pr(X = k), \quad k = 1, 2, \dots,$
- $g_r = \Pr(S = r), \quad r = 0, 1, 2, \dots.$

Then the unknown g_r can be recursively calculated by

1. $g_0 = p_0$,
2. $g_r = \sum_{j=1}^{r} \left(a + \frac{bj}{r}\right) f_j g_{r-j}, \quad r = 1, 2 \dots.$

Note If X is not a discrete random variable, then we first approximate it by a discrete distribution and then apply the Panjer's recursion algorithm.

Example 4.18. Aggregate claims S have a compound Poisson distribution $\mathcal{CP}(\lambda, F_X)$ where $\lambda = 1$ and an individual claim amount X is either 1 or 2 with probability 3/4 and 1/4, respectively. Calculate g_r for $r = 0, 1, 2, 3, 4, 5$.

4.7 Premium calculation

In this section, rules for setting premium to be charged to cover a risk S (aggregate claims) are presented. The expected (mean) risk $E[S]$ is referred to as the **pure premium**. In practice, the premium must be set to cover the expected risk, i.e. $P > E[S]$. Some premium calculation rules are as follows:

1. **The expected value principle (EVP)** The premium is given by a simple formula:

$$P = E[S] + \theta\, E[S] = (1 + \theta)E[S],$$

   for some $\theta > 0$, which is called the **relative security loading** on the pure premium $E[S]$. The premium is increased by a percentage of the mean of the risk.

2. **The standard deviation principle (SVP)** The premium is increased by a percentage of the standard deviation of the risk.

$$P = E[S] + \theta\,\mathrm{SD}[S].$$

3. **The variance principle (VP)** The premium is increased by a percentage of the variance of the risk.

$$P = E[S] + \theta\,Var[S].$$

Example 4.19. Suppose that S has a compound Poisson distribution with Poisson parameters $\lambda = 10$ and the claim sizes have a Pareto distribution $Pa(4, 3)$.

1. Use the normal approximation and the translated gamma approximation to calculate the relative security loading such that the probability of a profit in the year is 0.95.

2. Repeat the same question as above assumed that the SVP is applied.

Solution:

The loss distribution has Pareto distribution $X \sim Pa(4, 3)$. Therefore, $\mathrm{E}[X] = 1$ and $\mathrm{E}[X^2] = 3$.

The mean and variance of the aggregate claim amounts are $\mathrm{E}[S] = \lambda \mathrm{E}[X] = 10$ and $\mathrm{Var}[S] = \lambda \mathrm{E}[X^2] = 30$.

1. Use the normal approximation $S \sim \mathcal{N}(10, 30)$. Let P denote the premium charged. We need to find the security loading θ such that

$$\Pr(P - S > 0) = 0.95.$$

Assume that the premium charged to cover the risk follows the expected value principle (EVP). So we set

$$P = E[S] + \theta\, E[S] = (1+\theta)E[S] = (1+\theta)(10).$$

$$\begin{aligned}
0.95 &= \Pr(P - S > 0) &\quad (4.4)\\
&= \Pr((1+\theta)(10) - S > 0) &\quad (4.5)\\
&= \Pr(S < (1+\theta)(10)) &\quad (4.6)\\
&= \Pr(Z < \frac{(1+\theta)(10) - 10}{\sqrt{30}}) &\quad (4.7)\\
&= \Pr(Z < \frac{(\theta)(10)}{\sqrt{30}}). &\quad (4.8)
\end{aligned}$$

This gives $\frac{(\theta)(10)}{\sqrt{30}} = 1.6448536$ and

$$\theta = 0.9009234, \quad P = 19.009.$$

Instead of using the normal approximation to aggregate claims S, we now assume that S is approximated by $Y + k$ where

$Y \sim \mathcal{G}(1.4814815, 0.2222222)$ and $k = \mathbf{3.3333333}$ is a constant. It follows that

$$
\begin{aligned}
0.95 &= \Pr(P - S > 0) && (4.9)\\
&= \Pr((1+\theta)(10) - S > 0) && (4.10)\\
&= \Pr(S < (1+\theta)(10)) && (4.11)\\
&= \Pr(Y < (1+\theta)(10) - \mathbf{3.3333333}) && (4.12)\\
& && (4.13)
\end{aligned}
$$

Therefore, $(1+\theta)(10)$ - 3.3333333 = qgamma(0.95,shape = 1.481481, rate = 0.222222) = 17.43845, which results in

$$\theta = 1.0771784, \quad P = 20.772.$$

2. Using the standard deviation principle (SVP), we set

$$P = E[S] + \theta \, \mathrm{SD}[S] = 10 + \theta \cdot sqrt30$$

For the normal approximation of the aggregate claims, we obtain

$$
\begin{aligned}
0.95 &= \Pr(P - S > 0) && (4.14)\\
&= \Pr(10 + \theta \cdot sqrt30 - S > 0) && (4.15)\\
&= \Pr(S < 10 + \theta \cdot sqrt30) && (4.16)\\
&= \Pr(Z < \frac{10 + \theta \cdot sqrt30 - 10}{\sqrt{30}}) && (4.17)\\
&= \Pr(Z < \theta). && (4.18)
\end{aligned}
$$

Hence, $\theta = 1.644854$ and $P = 19.009$.

Comment The loading factors (for normal approximation) in both two cases are different because they are applied to different quantities (i.e. $\mathrm{E}[S]$ and $\mathrm{Var}[S]$). But they give the same premium.

Chapter 5

Ruin Theory

5.1 The classical risk process

Short term risk models for a fixed time period have been studied in the previous sections. In this section, risk models that evolve over time will be presented. Suppose that an insurer

- begins with an initial capital u, called an initial surplus,

- collects premiums at a constant rate c per unit time,
- and pays claims when losses occur.

The insurer is in ruin if the insurer's capital becomes negative at some point in time, i.e. the insurer's surplus falls to zero or below.

Note A surplus is an excess of income or assets over expenditure or liabilities in a given period, typically a financial year:

Example 5.1. *An insurer has initial surplus u of 1 (in suitable units) and receives premium payments at a rate of 1 per year. Suppose claims from a portfolio of insurance over the first two years are as follows:*

Time (years)	*0.4*	*0.9*	*1.5*
Amount	*0.8*	*0.7*	*1.2*

Plot a surplus process and determine whether ruin occurs within the first three years.

Solution: The insurer's surplus (or cash flow) at any future time t (> 0) is a random variable, since its value depends on the claims experience up to time t. The insurer's surplus at time t is a random variable. The insurer's surplus at time t is denoted $U(t)$. The following formula for $U(t)$ can be written as

$$U(t) = u + ct - S(t), \tag{5.1}$$

where the **aggregate claim amount up to time** t, $S(t)$ is

$$S(t) = \sum_{i=1}^{N(t)} X_i. \tag{5.2}$$

The following table summarises the values of the surplus function at the time when claims occurs.

Time	Surplus (before claim)	Surplus (after claim)
0	1	1
0.4	1.4	0.6
0.9	1.1	0.4
1.5	1	-0.2

The surplus function increases at a constant rate c until there is a claim and the surplus drops by the amount of the claim. The surplus then increases again at the same rate c and drops are repeated when claims occur. In this example, ruin occurs at time 1.5. The plot of the surplus process is given in the following figure.

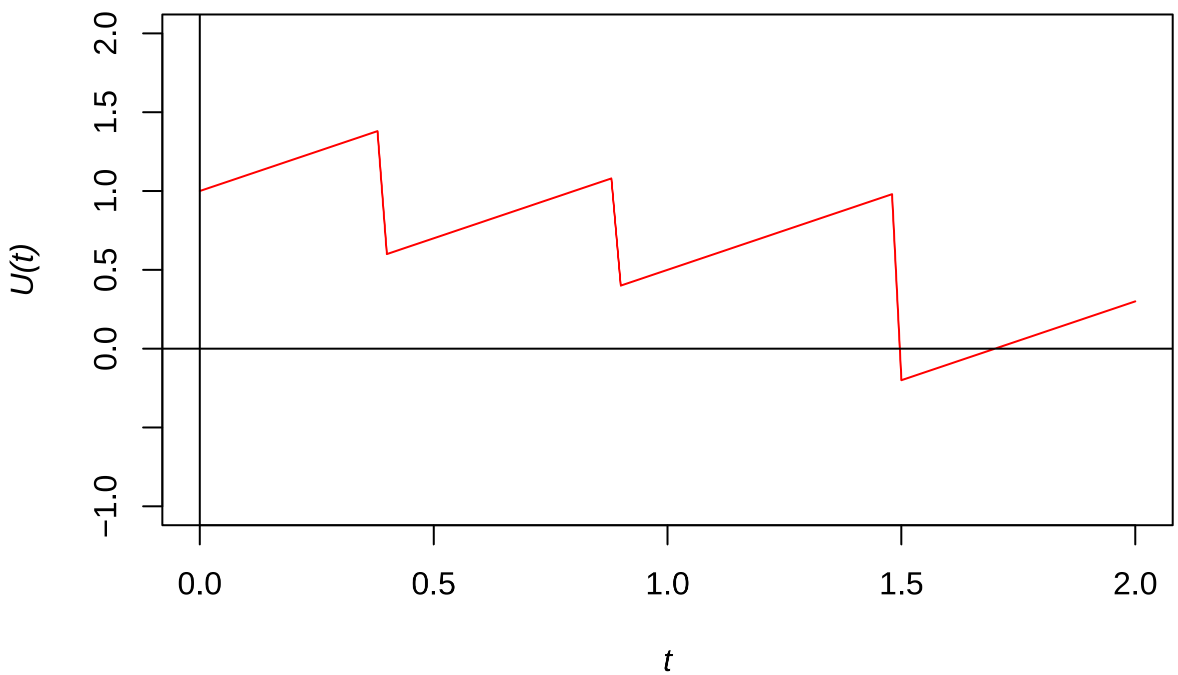


Figure 5.1: The surplus process befor reinsurance arrangment.

Example 5.2. *As given in Example 5.1, suppose that the insurer has effected proportional reinsurance with retained proportion of 0.7. The reinsurance premium is 0.4 per year to be paid continuously. Plot a surplus process and determine whether ruin occurs within the first three years. Comment on the results.*

Solution: The insurer's net premium income is 0.6 per year. The insurer's cash flow or surplus process is now given by

$$U_I(t) = u + (c - c_r)t - \alpha \cdot S(t), \tag{5.3}$$

where c_r is the reinsurance premium rate and α is the retained proportion.

The following table summarises the values of the surplus function at the time when claims occurs.

Time	Surplus (before claim)	Surplus (after claim)
0	1	1
0.4	1.24	0.68
0.9	0.98	0.49

Time	Surplus (before claim)	Surplus (after claim)
1.5	0.85	0.01

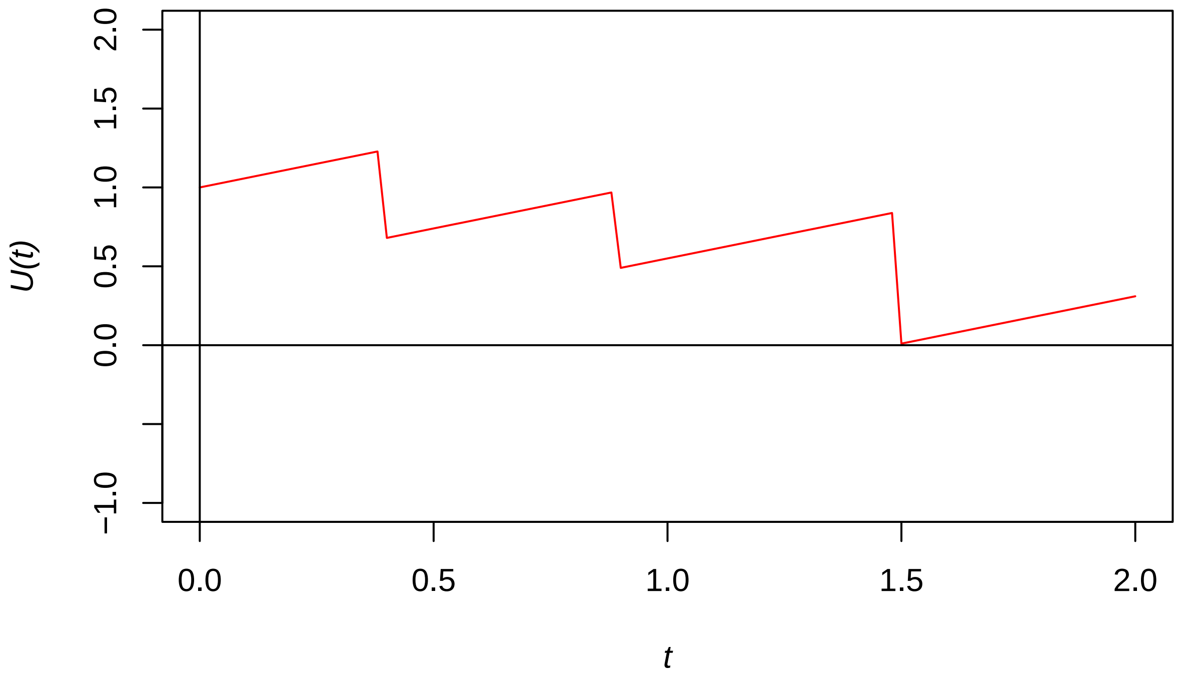


Figure 5.2: The surplus process under a proportional reinsurance arrangement.

It should be emphasised that under this proportional reinsurance arrangement, ruin does not occur within 2 years.

5.1.1 Classical risk process

The following assumptions are assumed for the study of the evolution of insurer's surplus over time.

1. The insurer's initial capital is u.
2. The premium rate per unit of time received continuously is c, i.e. the total amount of premiums received by time t is ct.
3. The counting process $\{N(t)\}_{t\geq 0}$ for the number of claims occurred in the time interval $[0, t]$ is a Poisson process with parameter λ.
4. The claim sizes (or individual claim amounts) $X_1, X_2, \ldots$ are independent and identically distributed random variables.
5. The claim sizes $X_1, X_2, \ldots$ are independent of the counting process $N(t)$.

The **surplus process** $\{U(t)\}_{t\geq 0}$ is then given by

$$U(t) = u + ct - S(t), \tag{5.4}$$

where the **aggregate claim amount up to time** t, $S(t)$ is

$$S(t) = \sum_{i=1}^{N(t)} X_i. \tag{5.5}$$

The evolution of insurer's surplus defined in (5.4) is also known as the **classical risk process**. The only random and uncertain quantity in (5.4) is the aggregate claims $S(t)$.

Notes The classical risk model contains many simplification.

1. The claim-arrival rate λ remains constant over time.
2. No interest is pain on the surplus.
3. There is no inflation.
4. The premium income is received continuously in time.
5. Claims are paid out **immediately**.
6. there are assumptions of independence.

5.1.2 Poisson processes

A **Poisson process** is a special type of counting process. It can be represented by a continuous time stochastic process $\{N(t)\}_{t\geq 0}$ which takes values in the non-negative integers. It can be used to model the occurrence or arrival of events over a continuous time interval. The state space is discrete but the time set is continuous. Here $N(t)$ represents the number of events in the interval $(0, t]$.

The following examples can also be modelled by a Poisson process:

1. Claims arrivals at an insurance company,

2. Accidents occurring on the highway, and

3. Telephone calls to a call centre.

Counting Process

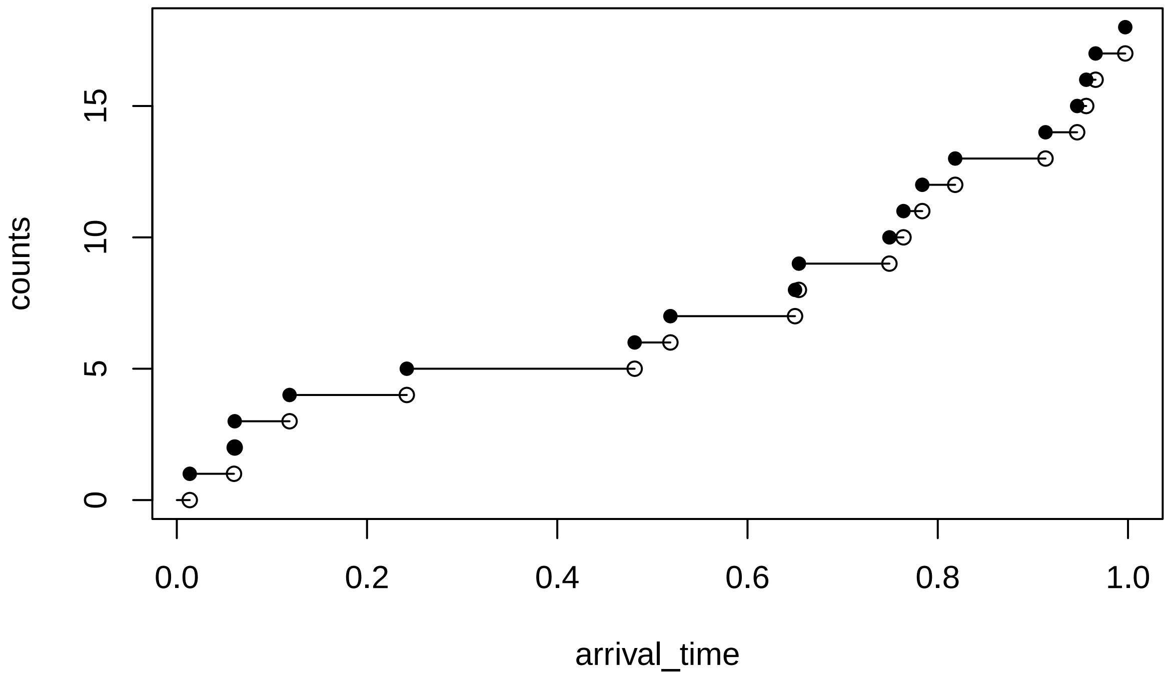
A counting process $\{N_t\}_{t\geq 0}$ is a collection of non-negative, integer-valued random variables such that if $0 \leq s \leq t$, then $N(s) \leq N(t)$.

The following figure illustrates a trajectory of the Poisson process. The sample path of a Poisson process is a right-continuous step function. There are jumps occurring at time $t_1, t_2, t_3, \ldots$.

```
lambda <- 17
# the length of time horizon for the simulation T_length <- 31
last_arrival <- 0
arrival_time <- c()
inter_arrival <- rexp(1, rate = lambda)
T_length <- 1
while (inter_arrival + last_arrival < T_length) {
  last_arrival <- inter_arrival + last_arrival
  arrival_time <- c(arrival_time,last_arrival)
  inter_arrival <- rexp(1, rate = lambda)
}


n <- length(arrival_time)
counts <- 1:n
```

```r
plot(arrival_time, counts, pch=16, ylim=c(0, n))
points(arrival_time, c(0, counts[-n]))
segments(
  x0 = c(0, arrival_time[-n]),
  y0 = c(0, counts[-n]),
  x1 = arrival_time,
  y1 = c(0, counts[-n])
)
```



Recall that a stochastic process $\{N(t)\}_{t\geq 0}$ is a Poisson process with parameter λ if the process satisfies the three properties:

1. $N(0) = 0$.

2. **Independent increments** For $0 < s < t \leq u < v$, the increment $N(t) - N(s)$ is independent of the increment $N(v) - N(u)$, i.e. the number of events in $(s, t]$ is independent of the number of events in $(u, v]$.

3. **Stationary increments** For $0 < s < t$, the distribution $N(t) - N(s)$ depends only on $t - s$ and not on the values s and t, i.e. the increments of the process over time has a distribution that only depend on the time difference $t - s$, the lenght of the time interval.

4. **Poisson distribution** For $t \geq 0$, the random variable $N(t)$ has a Poisson distribution with mean λt.

It follows from conditions the **Stationary Increments** and **Poisson Distribution** properties that

$$\Pr(N(t) - N(s) = n) = \Pr(N(t-s) - N(0) = n) = \frac{(\lambda(t-s))^n e^{-\lambda(t-s)}}{n!}, \qquad s < t, n$$

Notes

1. The sample paths of $\{N(t)\}_{t \geq 0}$ are non-decreasing step functions, or the process is referred to be as a counting process.

2. A process with stationary and independent increments can be thought of as **starting over** at any point in time in a probabilistic sense. The 'starting over' property follows from the fact that

- the exponential distribution has the memoryless property, and
- the times between successive events (or interarrival times) are independent and identically distributed exponential random variables with mean $1/\lambda$.

3. For more details about Poisson processes, please refer to the contents of the course "SCMA 469 Actuarial Statistics"

5.1.3 Compound Poisson processes

The aggregate claims process $S(t)$ defined in (5.5) of the classical risk process is said to be a **compound Poisson process** with Poisson parameter λ. The compound Poisson process has the following important properties:

1. For each t, the random variable $S(t)$ has a compound Poisson distribution with parameter λt, i.e.

$$S(t) \sim \mathcal{CP}(\lambda t, F_X(x)).$$

Thus, the mean and variance of the compound Poisson distribution are

$$\mathrm{E}[S(t)] = \lambda t \mathrm{E}[X], \quad \mathrm{Var}[S(t)] = \lambda t \mathrm{E}[X^2].$$

The moment generating function of $S(t)$ is

$$M_{S(t)}(r) = \exp(\lambda t(M_X(r) - 1)).$$

2. It has stationary and independent increments, i.e. for disjoint time intervals $0 < s < t \leq u < v$, the random variables $N(t) - N(s)$ and $N(v) - N(u)$ are independent and $N(t) - N(s)$ depends only on $t - s$ and not on the values s and t. Hence, the random variables $S(t) - S(s)$ and $S(v) - S(u)$ are **independent** and have $\mathcal{CP}(\lambda(t-s), F_X(x))$ and $\mathcal{CP}(\lambda(v-u), F_X(x))$ distributions, respectively.

Notes Various properties of the aggregate claims process $S(t)$ can be summarised as follows:

1. $S(1) \sim \mathcal{CP}(\lambda, F_X(x))$ is the aggregate claims in the first year.
2. $S(n) - S(n-1) \sim \mathcal{CP}(\lambda, F_X(x))$ is the aggregate claims in the nth year, for $n = 1, 2, ...$

3. The process $\{S(n) - S(n-1)\}_{n=1}^{\infty}$ is a sequence of **independent and identically distributed** random variables representing the aggregate claims in successive years.

5.1.4 The relative safety loading

According to the expected value principle, the premium rate c per unit time is defined by

$$c = (1+\theta)\mathrm{E}[S(1)] = (1+\theta)\lambda\mu_X.$$

Hence the **relative safety loading** (or **premium loading factor** or **relative security loading**) θ is given by

$$\theta = \frac{c - \lambda\mu_X}{\lambda\mu_X}.$$

In addition, the insurer should load the premium for profit so that $c > \lambda\mu_X$. This finding follows from the following example.

Let μ_X and σ_X^2 denote the mean and the variance of claim sizes X_i (in one period).

Example 5.3. Consider the following questions.

1. Calculate the expected surplus and the variance surplus at time t.
2. Calculate the expected profit per unit time in $(0, t]$.

Solution:

1. From $U(t) = u + ct - S(t)$, the expected surplus at time t is

$$\begin{aligned} \mathrm{E}[U(t)] &= u + ct - \mathrm{E}[S(t)] & (5.6)\\ &= u + ct - (\lambda t)\mathrm{E}[X] & (5.7)\\ &= u + ct - (\lambda t)\mu_X & (5.8)\\ &= u + (c - \lambda \mu_X) \cdot t, & (5.9) \end{aligned}$$

   and

$$\mathrm{Var}[U(t)] = \mathrm{Var}[S(t)] = (\lambda t)\mathrm{E}[X^2].$$

2. The expected profit per unit time in $(0, t]$ can be calculated from

$$\frac{\mathrm{E}[U(t) - U(0)]}{t} = c - \lambda \mu_X.$$

This motivates the **net profit condition**:

$$c > \lambda \mu_X.$$

Given λ and μ_X, we aim to set the premium rate c that satisfies the net profit condition.

Notes

1. The insurer can make a profit provided that $c > \lambda \mu_X$ or the relative safety loading θ is positive. In this case, the surplus will drift to ∞, but ruin could still occur. The rate at which premium income comes in is greater than the rate at which claims are paid out.

2. On the other hand, if $c < \lambda \mu_X$, then the surplus will drift to $-\infty$, but ruin is certain.

3. If $c = \lambda \mu_X$, the surplus will drift to ∞ and $-\infty$, but ruin is certain (eventually).

5.1.5 Ruin probabilities

Various definitions of ruin probabilities are given.

1. The **probability of ruin in infinite time** (or the **ultimate ruin probability**) is defined by

$$\psi(u) = \Pr(U(t) < 0 \quad \text{for some } t > 0).$$

2. The **finite-time ruin probability** (or the **probability of ruin by time** t) is defined by

$$\psi(u, t) = \Pr(U(s) < 0 \quad \text{for some } s \in (0, t]).$$

3. The **discrete time ultimate ruin probability** is defined by

$$\psi_h(u) = \Pr(U(t) < 0 \quad \text{for some } t \in \{h, 2h, 3h, ...\}).$$

4. The **discrete time ruin probability in finite time** is defined by

$$\psi_h(u, t) = \Pr(U(s) < 0 \quad \text{for some } s \in \{h, 2h, 3h, \ldots, t\}).$$

Notes

1. For $0 \leq u_1 \leq u_2$,
$$\psi(u_1) \geq \psi(u_2),$$
and
$$\psi(u_1, t) \geq \psi(u_2, t),$$
i.e. the ultimate ruin probability and finite-time ruin probability are non-increasing in u. Intuitively, the larger the initial surplus, the less likely it is that ruin will occur either in a finite time period or an unlimited time period.

2. If ruin occurs under the discrete time, it must occur under the continuous time, i.e.
$$\psi_h(u) < \psi(u).$$
Similarly,
$$\psi_h(u, t) < \psi(u, t).$$

3. For a given initial surplus u and $0 < t_1 < t_2$,
$$\psi(u, t_1) < \psi(u, t_2).$$

Intuitively, the longer the period considered when checking for ruin, the more likely it is that ruin will occur.

4. The discrete time ultimate ruin probability $\psi_h(u)$ could be used as an approximation of $\psi(u)$ provided h is sufficiently small.

5. The discrete time ruin probability in finite time $\psi_h(u,t)$ could be used as an approximation of $\psi(u,t)$ provided h is sufficiently small.

Example 5.4. *Suppose the annual aggregate claims for a portfolio of policies is approximately normal.*

- *The insurer's initial surplus is 1000 (in suitable units) and the premium rate is 1500 per year.*
- *The number of claims per year has a Poisson distribution with parameter 50.*
- *The distribution of claim sizes is lognormal with parameters* $\mu = 3$ *and* $\sigma^2 = 0.9$.

Calculate the probability that the insurer's surplus at time 2 will be negative.

Solution: Using the normal approximation, the total claims S can be approximated by $S \sim \mathcal{N}(\mathrm{E}[S], \mathrm{Var}[S])$. We have

$$\mathrm{E}[X] = e^{\mu+\sigma^2/2} = 31.500392 \tag{5.10}$$

$$\mathrm{E}[X^2] = e^{2\mu+2\sigma^2} = 2440.601978. \tag{5.11}$$

Therefore,

$$\mathrm{E}[S(2)] = 2(50)\mathrm{E}[X] = 3150.039231 \tag{5.12}$$
$$\mathrm{Var}[S(2)] = 2(50)\mathrm{E}[X2] = 2.440602 \times 10^5. \tag{5.13}$$

Hence, ruin will occur if $S(2)$ is greater than the initial surplus plus premiums received. Therefore, the probability of ruin is

$$\Pr(S(2) > u + 2c) = \Pr(S(2) > 1000 + 2(1500)) \tag{5.14}$$
$$= \Pr(Z > \frac{1000 + 2(1500) - 3150.0392309}{\sqrt{2.440602 \times 10^5}}) \tag{5.15}$$
$$= \Pr(Z > 1.720483) = 0.04267. \tag{5.16}$$

the probability of ruin is approximately 4.267%.

5.2 Simulation of ruin probabilities

In this section, we will use simulation to numerically estimate the probability of ruin. First, we introduce the **inverse transform method**, which is a method for generating random numbers from any probability distribution by using its inverse cumulative distribution.

Example 5.5. *Let $F(x)$ be a continuous cumulative density function. Let Y be a random variable with a $U(0,1)$ distribution. Define the random variable X by*

$$X = F^{-1}(Y).$$

Show that the cumulative density function of X, $F_X(x)$ is $F(x)$.

Solution: We need to show that $Pr(X \leq x) = F(x)$ for all x, i.e. $F_X(x) = F(x)$ as defined above.

It follows from the monotonicity of F and the definition

$$
\begin{aligned}
F_X(x) &= \Pr(X \leq x) && (5.17)\\
&= \Pr(F^{-1}(Y) \leq x) && (5.18)\\
&= \Pr(F(F^{-1}(Y)) \leq F(x) && (5.19)\\
&= \Pr(Y \leq F(x)) && (5.20)\\
& && (5.21)
\end{aligned}
$$

Since $Y \sim U(0,1)$, we have $\Pr(Y \leq t) = t$ for any $t \in [0,1]$. Therefore,

$$F_X(x) = \Pr(Y \leq F(x)) = F(x).$$

Note We can use this result to generate values from the required probability distribution (which will be useful in Excel). In order to generate $X_1, X_2, X_3, \dots, X_n$ from $\mathcal{G}(\alpha, \lambda)$ (or any other distributions) in Excel, we use `GAMMAINV(RAND(), alpha, 1/lambda)`. However, in R, we can simply use `rgamma(n, alpha, lambda)` to generate n random numbers from the $\mathcal{G}(\alpha, \lambda)$ distribution.

Example 5.6. *The aggregate claims process for a risk is compound Poisson with Poisson parameter $\lambda = 100$ per year. Individual claim amounts have $Pa(4,3)$. The premium income per year is $c = 110$ (in suitable units), received continuously.*

Using either Excel or R to simulate 1000 values of aggregate claims S, assuming that S is approximated by a translated gamma approximation,

1. *Estimate $\widehat{\psi}_1(50,5)$, an estimate of $\psi_1(50,5)$.*

2. *Estimate the standard error of $\widehat{\psi}_1(50,5)$.*

3. *Calculate a 95% confidence interval for your estimate in 1.*

4. *Estimate $\psi_{0.5}(50,5)$.*

Solution:

1. An estimate of $\psi_1(u,5)$ with $u = 50$ and $c = 110$ can be obtained as follows:

From the properties of $S(t)$,

- The aggregate claims in the first year have $S(1) \sim \mathcal{CP}(\lambda, F_X(x))$ distribution with $\lambda = 100$ and $X \sim \mathrm{Pa}(4, 3)$
- The aggregate claims in the jth year, for $j = 1, 2, \dots, 5$ have $S(j) - S(j-1) \sim \mathcal{CP}(\lambda, F_X(x))$ distribution.

It follows that

$$\begin{aligned}\psi_1(u, 5) &= \Pr(U(j) < 0 \quad \text{for at least one of } j \in \{1, 2, \dots, 5\}) &(5.22)\\ &= \Pr(u + cj - S(j) < 0), \text{ for at least one of } j = 1, 2, \dots, 5. &(5.23)\end{aligned}$$

When

- $j = 1, U(1) = 50 + 110 - S(1)$
- $j = 2, U(2) = 50 + 2 - S(2) = U(1) + c - (S(2) - S(1))$
- $j = 3, U(3) = 50 + 3 - S(3) = U(2) + c - (S(3) - S(2))$
- $j = 4, U(4) = 50 + 4 - S(4) = U(3) + c - (S(4) - S(3))$
- $j = 5, U(5) = 50 + 5 - S(5) = U(4) + c - (S(5) - S(4))$

The algorithm to estimate the finite time ruin in discrete time can be described as follows:

Step 1. Simulate values of $S(1), S(2) - S(1), ..., , S(5) - S(4)$ from $\mathcal{CP}(\lambda, F_X(x))$ distribution. Then compute $U(1), U(2), ... U(5)$.

Step 2. Check if one of $U(1), U(2), ... U(5)$ are negative.

Step 3. Repeat the simulations (1 and 2) 1000 times.

Step 4. Le M be the number of simulations out of 1000 where ruin occurs. Then $\hat{\psi}_1(50, 5) = \frac{M}{1000}$.

From the results, there are $M = 21$ simulations that ruin occurs, and hence

$$\hat{\psi}_1(50, 5) = \frac{M}{1000} = \frac{21}{1000}.$$

2. The estimation of the standard error of $\hat{\psi}_1(50, 5)$ can be obtained as follows. We know that $M \sim \mathcal{B}(1000, p)$ where $p = \hat{\psi}_1(50, 5)$. Then,

$$\text{Var}\left[\frac{M}{1000}\right] = \frac{1}{1000^2}\text{Var}[M] = \frac{1000(p)(1-p)}{1000^2}$$

and

$$\mathrm{SD}[\hat{\psi}_1(50,5)] = \frac{1000(0.021)(1-0.021)}{1000^2} = 0.004534$$

3. The 95% confidence interval of the estimate is

$$(\hat{\psi}_1(50,5) - z_{\alpha/2}\mathrm{SD}[\hat{\psi}_1(50,5)], \hat{\psi}_1(50,5) + z_{\alpha/2}\mathrm{SD}[\hat{\psi}_1(50,5)]) = (0.012113, 0.0$$

4. For the estimation of discrete time probability of ruin where the surplus process is checked at time intervals of length 0.5, we proceed as follows. First we note that

$$S(1/2), S(1) - S(1/2), S(3/2) - S(1), \ldots, S(5) - S(9/2) \sim \mathcal{CP}((1/2)\lambda, F_X(x))$$
distribution.

In addition,

- U(1/2) = U(0) + c (1/2) - S(1/2)$
- U(1) = U(1/2) + c (1/2) - (S(1) - S(1/2))$
- U(3/2) = U(1) + c (1/2) - (S(3/2) - S(1))$, ...,

- U(5) = U(9/2) + ‘c (1/2) - (S(5) - S(9/2))$.

It follows that

$$\mathrm{E}[S(1/2)] = (1/2)(100)\mathrm{E}[X] = (1/2)\mathrm{E}[S(1)] = 50 \tag{5.24}$$
$$\mathrm{Var}[S(1/2)] = (1/2)(100)\mathrm{E}[X^2] = (1/2)\mathrm{Var}[S(1)] = 150 \tag{5.25}$$
$$\mathrm{Sk}[S(1/2)] = \sqrt{2}\mathrm{Sk}[S(1)] = 0.734847. \tag{5.26}$$

Now we assume that $S(j) - S(j-1)$ for $j = 1/2, 1, 3/2, \dots, 5$ can be approximated by $Y + k$ where $Y \sim \mathcal{G}(\alpha, \lambda)$ and k is a constant. It follows that

$$\hat{\alpha} = 7.407407, \quad \hat{\beta} = 0.222222, \quad \hat{k} = 16.666667.$$

The simulations can be obtained in the same way.