Stochastic Processes in Non-Life Insurance (SkadeStok) 2023/2024

Lecture notes

Rasmus Frigaard Lemvig (rfl@math.ku.dk)

October 2023

Introduction

These lecture notes were written for the course *Stochastic Processes in Non-Life Insurance* (abbreviated *SkadeStok*) at the University of Copenhagen in the fall of 2023 and are based on the lectures given by Jeffrey F. Collamore. I have strived to keep the notes as faithful as possible to the lectures but I have made minor changes that I felt would improve the overall readability and presentation. These changes include presenting new notions as definitions and important remarks as results such as propositions and lemmata whenever possible. Thus the notes are written more in the style of a book. I have also added certain proofs that helped me absorb the different notions, but proofs belonging to results from other areas are usually omitted. Occasionally, an example not presented in the lecture has been added.

I want to stress that these notes are not meant as a replacement for the lectures but as a supplement. The notes can be used to prepare for the lectures or to revisit specific concepts and explanations (which is why I added an index) but they cannot provide the same intuition as the (physical) lectures. This intuition is also an indispensable tool for solving the exercises and developing problem solving skills (which is a key point of the course).

Lastly, I want to thank Rasmus Benn and Anders Lund Mortensen for pointing out some typos. There are likely still typos remaining and maybe a few mathematical errors. These are all due to me. If the reader spots any of these mistakes, they are more than welcome to contact me.

Rasmus Frigaard Lemvig
November 2023

Table of contents

Table of contents

Week 1 - Basic notions

1 The Cramér-Lundberg model

We start by giving a brief overview of the Cramér-Lundberg model. The model was first introduced informally in the doctoral thesis of Filip Lundberg in 1903. The model was republished in the 1930s by Harald Cramér, this time with a more rigorous mathematical foundation. The model is specified by two payment streams:

- **Premiums from policyholders:** The premiums from the policyholders are assumed to be constant over time. We denote the income at time t by $I_t = ct$ for some constant $c > 0$.
- **Claims losses:** Claims losses occur at random times. The idea of the model is that the probability of a claim happening in a small (infinitesimal) interval $[t, t + \Delta t]$ is approximately proportional to a positive constant λ times the length of the interval and only a single claim occurs at a given time. Somewhat informally,

$$\begin{aligned} P(\text{claim in } [t, t + \Delta t]) &= \Delta t \cdot \lambda + o(\Delta t) \\ P(\text{more than one claim in } [t, t + \Delta t]) &= o(\Delta t). \end{aligned}$$

  Furthermore, these probabilities should not depend on t. Let N_t denote the number of claims in the interval $[0, t]$. We shall soon see that the above probabilities imply that the stochastic process $\{N_t\}$ is a Poisson process. Let $\{Y_i\}$ denote the claim sizes (a discrete stochastic process of a.s. positive random variables). The stream of accumulated losses $\{L_t\}$ in $[0, t]$ is given by

$$L_t = \sum_{i=1}^{N_t} Y_i.$$

  We assume that $\{Y_i\}$ is an iid sequence.

From these two payment streams, we can formulate the fundamental object of study in this lecture, namely the *total capital process* $\{C_t\}$ given by

$$C_t = u + ct - \sum_{i=1}^{N_t} Y_i$$

where $u > 0$ is a constant called the *initial capital*.

We end this section by addressing some criticisms of the model. First of all, the model assumes homogeneous behaviour of the policy-holders. In reality, there are many different types of policy-holders for a company and they pay different premiums for different products, and the claims will have quite different distributions. It also assumes that the number of policy-holders is fixed, but in the real world, policy-holders tend to change company from time to time. There is also the obvious issue of the growth of the capital. If everything goes well (larger cash flows from premiums than from claims), the model dictates that the capital will grow indefinitely. Most real shareholders would prefer that a part of the surplus was paid out as dividends. Last but not least, there is the problem of seasonality. The number of claims in the model is assumed to not depend on time but only on a timespan. In reality, many claims happen more often in certain months of the year. Furthermore, some claims may increase in frequency over time. An example could be flooding or other extreme wheather phenomena. Nevertheless, the model strikes a good balance between mathematical tractability and usefulness.

2 The ruin problem

Ruin occurs if the accumulated losses surpass the initial capital plus the accumulated premiums. We are interested in the ruin probability in the long run, and therefore we let $\psi(u)$ denote the probability that ruin occurs at some point as a function of the initial capital u. Formally,

$$\psi(u) = P(C_t < 0 \text{ for some } t \geq 0) = P\left(\inf_{t\geq 0} C_t < 0\right).$$

Letting $X_t = I_t - L_t$, we can also write this probability as

$$\psi(u) = P(X_t < -u \text{ for some } t \geq 0).$$

In an applied context, we want to find an initial capital u so large that $\psi(u) \leq \beta$ for some fixed threshold $\beta \in (0, 1)$.

3 Stochastic processes

In this section we will introduce the basic concepts of stochastic processes that will be needed in the course.

Definition 3.1. Let $(\Omega,\mathcal{F}, P)$ denote a probability space i.e. Ω is a set called the *sample space*, $\mathcal{F}$ is a σ-algebra and P is a probability measure. A stochastic process $\{X_t\}$ is a map

$$X : \Omega \times [0, \infty) \to \mathbb{R}$$

where we will usually write $X_t(\omega) := X(\omega, t)$. For a given outcome $\omega \in \Omega$, we call the function $t \mapsto X(\omega, t)$ the *sample path* associated with ω.

In the rest of the section, we implicitly assume that a probability space $(\Omega, \mathcal{F}, P)$ is given. Often it is necessary to assume some continuity properties of the stochastic processes we work with. An especially important assumption is the property of being càdlàg.

Definition 3.2. A stochastic process $\{X_t\}$ is called càdlàg (French for *continue à droite, limite à gauche*) if the sample paths are right continuous and have left limits. In other words, for all t,

$$\lim_{s \downarrow t} X_s = X_t$$

and

$$\lim_{s \uparrow t} X_s$$

exists.

We also need to model a flow of information in time. This is done via the mathematical concept of a filtration.

Definition 3.3. A *filtration* $\{\mathcal{F}_t\}$ is a sequence of σ-algebras such that $\mathcal{F}_t \subseteq \mathcal{F}$ for all t and such that $\mathcal{F}_s \subseteq \mathcal{F}_t$ for all $s \leq t$.

In this course we have the convention $\mathcal{F}_0 = \{\emptyset, \Omega\}$ i.e. $\mathcal{F}_0$ is the trivial σ-algebra. Intuitively this means that we have no information available at time zero. It is necessary to have some measurability conditions on our stochastic processes. If the value of a stochastic process is known at time t, we call such a process adapted.

Definition 3.4.

(i) A stochastic process $\{X_t\}$ is called $\{\mathcal{F}_t\}$-*adapted* if X_t is $\mathcal{F}_t$-measurable for all t. We will often omit $\{\mathcal{F}_t\}$ and simply say that $\{X_t\}$ is adapted.

(ii) For a given stochastic process $\{X_t\}$, we call the filtration given by $\mathcal{F}_t^X = \sigma(X_s : s \leq t)$ the filtration generated by the process $\{X_t\}$. This is the smallest filtration such that $\{X_t\}$ is adapted.

A particularly important subclass of stochastic processes in this course is the following.

Definition 3.5. A stochastic process $\{X_t\}$ is called a Lévy process if

(i) $X_0 = 0$.

(ii) $\{X_t\}$ has *independent increments*. That is, for every finite partition $0 < t_1 < t_2 < \cdots < t_k$, the variables $\{X_{t_{i+1}} - X_{t_i}\}_{i=1}^{k-1}$ are independent.

(iii) $\{X_t\}$ is a *stationary* sequence. For every $s \leq t$, $X_t - X_s \overset{\mathrm{d}}{=} X_{t-s}$ where $\overset{\mathrm{d}}{=}$ denotes equality in distribution.

Let us finally have a look at two important examples of stochastic processes (that are also both Lévy processes).

Example 3.6. A stochastic process $\{W_t\}$ is called a *standard Brownian motion* if

(i) $W_0 = 0$,

(ii) $\{W_t\}$ has independent increments and

(iii) $W_t - W_s \sim \mathcal{N}(0, t - s)$.

In other words, a standard Brownian motion is a Lévy process with an increment over $[s,t]$ being normal distributed with mean zero and variance $t-s$. We write $\{W_t\} \sim \mathrm{BM}(0,1)$ for a standard Brownian motion.

A stochastic process $\{W_t\}$ is called a general Brownian motion $\mathrm{BM}(m,\eta^2)$ if

$$\left\{\frac{W_t - mt}{\eta\sqrt{t}}\right\} \sim \mathrm{BM}(0,1).$$

Example 3.7. A stochastic process $\{N_t\}$ is called a *Poisson process* with *intensity* $\lambda > 0$ if N_t takes values in $\mathbb{N}_0 = \{0,1,2,...\}$ and

(i) $P(N_h \geq 1) = \lambda h + o(h)$,

(ii) $P(N_h \geq 2) = o(h)$ and

(iii) $\{N_t\}$ has stationary and independent increments.

All limits are understood in the sense that $h \to 0$. The concept of a Poisson process can be extended to that of a *compound Poisson process* . This is a process of the form

$$V_t = \sum_{i=1}^{N_t} Y_i$$

with $\{N_t\}$ a Poisson process and $\{Y_i\}$ iid and independent of $\{N_t\}$.

Remark 3.8. Lévy processes seem like a very broad class of stochastic processes but it turns out that every Lévy process $\{X_t\}$ can be decomposed into a sum of the form

$$X_t = ct + \sigma W_t + J_t$$

with c a constant, W_t a standard Brownian motion and J_t a jump process. The interested reader can look up *Lévy-Itô decomposition.*

Another interesting type of process is a point process.

Definition 3.9. A stochastic process $\{N_t\}$ taking values in $\mathbb{N}_0$ is called a *point process* if $N_0 = 0$ and $N_t \geq N_s$ for $t \geq s$.

We end this lecture by establishing a useful property of Poisson processes.

Proposition 3.10. *Let $\{N_t^1\}$ and $\{N_t^2\}$ be independent Poisson processes with intensities λ_1 and λ_2, respectively. Then $\{N_t^1 + N_t^2\}$ is a Poisson process with intensity $\lambda_1 + \lambda_2$.*

Proof. $\{N_t^1 + N_t^2\}$ clearly takes non-negative integer values. Using independence, we have

$$\begin{aligned} P(N_h^1 + N_h^2 \geq 1) &= 1 - P(N_h^1 = 0, N_h^2 = 0) = 1 - P(N_h^1 = 0)P(N_h^2 = 0) \\ &= 1 - (1 - \lambda_1 h + o(h))(1 - \lambda_2 h + o(h)) \\ &= 1 - (1 - \lambda_2 h - \lambda_1 h + \lambda_1\lambda_2 h^2 + o(h)) \\ &= \lambda_1 h + \lambda_2 h + o(h) = (\lambda_1 + \lambda_2)h + o(h). \end{aligned}$$

Consider now the event that $N_h^1 + N_h^2 \geq 2$. If this is the case, then either $N_h^1 \geq 2$, $N_h^2 \geq 2$ or $N_h^1, N_h^2 \geq 1$. Hence

$$\begin{aligned} P(N_h^1 + N_h^2 \geq 2) &\leq P(N_h^1 \geq 2) + P(N_h^2 \geq 2) + P(N_h^1 \geq 1, N_h^2 \geq 1) \\ &= o(h) + o(h) + P(N_h^1 \geq 1)P(N_h^2 \geq 1) \\ &= o(h) + (\lambda_1 h + o(h))(\lambda_2 h + o(h)) \\ &= o(h) + (\lambda_1 \lambda_2 h^2 + o(h)) = o(h). \end{aligned}$$

This completes the proof. ■

A Poisson process can be identified with its *arrival* times

$$T_k = \inf\{t \geq 0 : N_t = k\}$$

i.e. the time that the process jumps from $k-1$ to k. Note that $T_0 = 0$. Given the arrival times, we have

$$N_t = \sum_{i=0}^{\infty} 1_{\{T_n \leq t\}}$$

which establishes the one-to-one correspondence between a Poisson process and its arrival times. In describing a Poisson process, the *interarrival times* $\tau_i = T_i - T_{i-1}$ for $i = 1, 2, ...$ are particularly important. There are many equivalent formulations of the definition of a Poisson process. Some of them are given in the following proposition.

Proposition 3.11. *The following are equivalent:*

(i) $\{N_t\}$ *is a Poisson process with intensity* $\lambda > 0$.

(ii) $\{N_t\}$ *is a point process with independent interarrival times* $\{\tau_i\}$ *with* $\tau_i \sim \text{Exp}(\lambda)$.

(iii) $\{N_t\}$ *has independent and stationary increments and* $N_t \sim \text{Poisson}(\lambda t)$ *for all* $t \geq 0$.

Proof. See Proposition 1.10 on page 3 in [6] or Theorem 5.3 on page 18 in [2]. ■

4 Martingales and optional sampling

In this section we go through the martingale machinery necessary for our purposes. We assume that we are given a probability space $(\Omega, \mathcal{F}, P)$.

Definition 4.1. A stochastic process $\{M_t\}$ is a martingale (relative to the filtation $\{\mathcal{F}_t\}$) if

(i) $\{M_t\}$ is $\{\mathcal{F}_t\}$-adapted.

(ii) $\mathbf{E}[|M_t|] < \infty$ for all $t \geq 0$.

(iii) $\mathbf{E}[M_{t+s} \mid \mathcal{F}_t] = M_t$ a.s. for all $s, t \geq 0$.

Example 4.2. If $\{N_t\}$ is a Poisson process with intensity $\lambda > 0$, $\{N_t - \lambda t\}$ is a martingale. The verification of this is left as an exercise.

Example 4.3. A standard Brownian motion is a martingale. Again the reader can verify this.

A useful concept in many situations is the notion of a stopping time.

Definition 4.4. A random variable $\tau : \Omega \to [0, \infty)$ is called a *stopping time* with respect to a filtration $\{\mathcal{F}_t\}$ if

$$\{\tau \leq t\} \in \mathcal{F}_t \quad \forall t \geq 0.$$

One should intuitively think of a stopping time τ as the time of an event where we know whether this event has occured given the information available at time t.

Example 4.5. Let $\{W_t\} \sim \text{BM}(0, 1)$ and $\tau = \inf\{t \geq 0 : |W_t| \geq a\}$ for some $a \geq 0$. τ is the first time that $\{W_t\}$ leaves the region $[-a, a]$. This is a stopping time since we know whether $|W_t| \geq a$ at time t (assuming that W_t is adapted).

A non-example of a stopping time could be: τ indicates the last time a standard Brownian motion surpasses 1. We can only determine τ if we know the entire future behaviour of the process and hence τ is not a stopping time. Before we can state the next essential result, we need to define the information available at a stopping time.

Definition 4.6. For a stopping time τ, define the σ-algebra

$$\mathcal{F}_\tau = \{A \in \mathcal{F} \mid A \cap \{\tau \leq t\} \in \mathcal{F}_t \ \forall t \geq 0\}.$$

We leave it as an exercise for the reader to verify that $\mathcal{F}_\tau$ is indeed a σ-algebra. The intuition behind $\mathcal{F}_\tau$ is that it contains all the information available at time τ. Note that if τ is constant and equal to t, then $\mathcal{F}_\tau = \mathcal{F}_t$. We can now state an essential tool from the theory of martingales, suitably formulated for our purposes.

Theorem 4.7 (**Optional sampling theorem**). *Assume $\{M_t\}$ is a martingale and let τ_1 and τ_2 be stopping times. Then*

$$\mathbf{E}[M_{t\wedge\tau_2} \mid \mathcal{F}_{\tau_1}] = M_{t\wedge\tau_1\wedge\tau_2} \ a.s.$$

with $\wedge$ denoting "minimum" i.e. $t \wedge s := \min\{t, s\}$.

Proof. See for example Theorem 1.2.6 in [7] or [4]. ■

For our purposes, a special case of the above theorem will often suffice.

Corollary 4.8. *Let $\{M_t\}$ be a martingale and τ a stopping time. Then*

$$\mathbf{E}[M_{t\wedge\tau} \mid \mathcal{F}_0] = M_0.$$

If τ is also a.s. bounded, we have

$$\mathbf{E}[M_\tau \mid \mathcal{F}_0] = M_0.$$

We now consider an example of how the technique of optional sampling can be applied.

Example 4.9. Consider a Poisson process $\{N_t\}$ with intensity λ. Let $a \in \mathbb{N}$ and $\tau = \inf\{t \geq 0 : N_t \geq a\}$ denote the first time N_t reaches a. Our goal is to compute $\mathbf{E}[\tau]$. We know that $M_t = N_t - \lambda t$ is a martingale. Using optional sampling, we have $\mathbf{E}[M_{t\wedge\tau} \mid \mathcal{F}_0] = M_0$. We have $\mathbf{E}[M_{t\wedge\tau}] = M_0 = 0$ by our convention that $\mathcal{F}_0 = \{\emptyset, \Omega\}$. We thus have

$$\mathbf{E}[N_{t\wedge\tau}] = \lambda \mathbf{E}[t \wedge \tau].$$

Let us rewrite the left hand side as follows

$$\mathbf{E}[N_{t\wedge\tau}] = \mathbf{E}[N_{t\wedge\tau}1_{\{t\geq\tau\}}] + \mathbf{E}[N_{t\wedge\tau}1_{\{t<\tau\}}] = \mathbf{E}[N_\tau 1_{\{t\geq\tau\}}] + \mathbf{E}[N_t 1_{\{t<\tau\}}] = a\mathbf{E}[1_{\{t\geq\tau\}}] + \mathbf{E}[N_t 1_{\{t<\tau\}}]$$

and so

$$a\mathbf{E}[1_{\{t\geq\tau\}}] + \mathbf{E}[N_t 1_{\{t<\tau\}}] = \lambda \mathbf{E}[t \wedge \tau].$$

Let us consider the limit $t \to \infty$ for each term. By dominated convergence,

$$\mathbf{E}[1_{\{t\geq\tau\}}] \to \mathbf{E}[1_{\{\tau<\infty\}}] = P(\tau < \infty) = 1$$

since $N_t \to \infty$ a.s. (use the strong law of large numbers on N_t/t). $N_t \to \infty$ also implies that $1_{\{t<\tau\}} \to 0$ so using dominated convergence,

$$\mathbf{E}[N_t 1_{\{t<\tau\}}] = \mathbf{E}[N_{t\wedge\tau} 1_{\{t<\tau\}}] \to 0.$$

Lastly, monotone convergence yields

$$\mathbf{E}[t \wedge \tau] \to \mathbf{E}[\tau].$$

Combining all these limits, we get

$$a + 0 = \lambda \mathbf{E}[\tau]$$

and we conclude that $\mathbf{E}[\tau] = a/\lambda$. This result also makes sense intuitively. A larger intensity means that the process is more likely to reach a early, and a larger a should take longer time to reach.

We end this section by stating and proving a theorem which tells us how to construct martingales using Lévy processes. First we need a short definition.

Definition 4.10. For a random variable X, the function

$$\kappa_X(\alpha) = \mathbf{E}[e^{\alpha X}]$$

is called the *moment-generating function* of X (when it exists in a neighbourhood around zero). When the moment-generating function exists, the function

$$\Lambda(\alpha) = \log \kappa_X(\alpha) = \log \mathbf{E}[e^{\alpha X}]$$

is called the *cumulant-generating function* of X.

The following lemma is purely technical and is used to construct two important examples of martingales.

Lemma 4.11. *If a real function f satisfies the Cauchy functional equation $f(x + y) = f(x) + f(y)$, then f restricted to $\mathbb{Q}$ is a linear function i.e. $f(q) = cq$ for some constant c for all q. If f is continuous, then f is a linear function (on its entire domain).*

Proof. Left as an exercise for the reader. ■

Theorem 4.12. *Assume $\{X_t\}$ is a Lévy process with respect to the natural filtration and $\mathbf{E}[|X_t|] < \infty$ for all $t \geq 0$. Then:*

(i) $\{X_t - \mu t\}$ with $\mu = \mathbf{E}[X_1]$ is a martingale.

(ii) If $\kappa_X(\alpha)$ exists, $e^{\alpha X_t - t\Lambda(\alpha)}$ is a martingale where Λ denotes the cumulant-generating function of X_1, $\Lambda(\alpha) = \log \mathbf{E}[e^{\alpha X_1}]$.

Proof. (i) is left as an exercise for the reader. To prove (ii), we first claim that

$$\mathbf{E}[e^{\alpha X_t}] = e^{t\Lambda(\alpha)}.$$

Define $f(t) = \mathbf{E}[e^{\alpha X_t}]$. The Lévy process property gives

$$\begin{aligned} f(t+s) &= \mathbf{E}[e^{\alpha(X_t + X_s)}] = \mathbf{E}[e^{\alpha X_t}]\mathbf{E}[e^{\alpha(X_{t+s} - X_t)}] \\ &= \mathbf{E}[e^{\alpha X_t}]\mathbf{E}[e^{\alpha X_s}] = f(t)f(s). \end{aligned}$$

Hence $\log f(t+s) = \log f(t) + \log f(s)$ and so $\log f(t)$ satisfies the Cauchy functional equation. Recall that the moment-generating function is continuous and hence $\log f(t) = ct$ for some constant c by the above lemma. To identify c, simply note that $f(1) = e^c$ so that $c = \log f(1) = \log \mathbf{E}[e^{cX_1}] = \Lambda(\alpha)$. Now let $M_t = e^{\alpha X_t - t\Lambda(\alpha)}$. The rest of the proof consists of the following calculation:

$$\begin{aligned} \mathbf{E}[M_{t+s} \mid \mathcal{F}_t] &= \mathbf{E}[e^{\alpha X_{t+s} - (t+s)\Lambda(\alpha)}] = e^{-(t+s)\Lambda(\alpha)}\mathbf{E}[e^{\alpha X_{t+s}} \mid \mathcal{F}_t] \\ &= e^{-(t+s)\Lambda(\alpha)}\mathbf{E}[e^{\alpha(X_{t+s} - X_t)}e^{\alpha X_t} \mid \mathcal{F}_t] \\ &= e^{-(t+s)\Lambda(\alpha)}e^{\alpha X_t}\mathbf{E}[e^{\alpha(X_{t+s} - X_t)}] = e^{-(t+s)\Lambda(\alpha)}e^{\alpha X_t}\mathbf{E}[e^{\alpha X_s}] \\ &= {}^{-(t+s)\Lambda(\alpha)}\ e^{\alpha X_t}e^{s\Lambda(\alpha)} = e^{\alpha X_t - t\Lambda(\alpha)} = M_t. \end{aligned}$$

In the third equality we used the fact that a Lévy process has independent increments. ■

The martingale in (ii) is important to remember. It will be used frequently throughout the course.

5 The net profit condition

We end this lecture by again considering the Cramér-Lundberg process,

$$C_t = u + ct - \sum_{i=1}^{N_t} Y_i, \quad X_t = ct - \sum_{i=1}^{N_t} Y_i.$$

$\{X_t\}$ is a Lévy process (see the exercises). By the previous theorem, $M_t = X_t - \mu t$ for $\mu = \mathbf{E}[X_1]$ is a martingale. A natural question to ask is when the ruin probability $\psi(u)$ is strictly less than one. Using the strong law of large numbers, it follows that $\psi(u) = 1$ when $\mu < 0$. A difficult case is when $\mu = 0$. One can show that in this case, C_t will cross zero infinitely often almost surely. Hence ruin is also inevitable in this case. We conclude that the ruin problem is only non-trivial for $\mu > 0$. We have

$$0 < \mu = c - \mathbf{E}[N_1]\mathbf{E}[Y_1] = c - \lambda\mu_Y$$

where $\mu_Y = \mathbf{E}[Y_1]$. It follows that $\mu > 0$ if and only if $c > \lambda\mu_Y$. This equality is called the *net profit condition* or simply *NPC* . This has a natural interpretation. Indeed, NPC simply says that the premium rate exceeds the rate of claims losses.

Week 2 - The Lundberg inequality and renewal theory

6 The Lundberg/adjustment coefficient

Consider the Cramér-Lundberg model

$$C_t = u + ct - \sum_{i=1}^{N_t} Y_i, \quad X_t = ct - \sum_{i=1}^{N_t} Y_i.$$

From the previous lecture, we know that the following are martingales

$$\begin{aligned}
&\tilde{M}_t = X_t - \mu t, \quad \mu = \mathbf{E}[X_1] = c - \lambda\mu_Y > 0 \text{ (by NPC)} \\
&M_t = e^{\alpha X_t - t\Lambda(\alpha)}, \quad \Lambda(\alpha) = \log \mathbf{E}[e^{\alpha X_1}]
\end{aligned}$$

where $\alpha \in \mathbb{R}$ and $\mu_Y = \mathbf{E}[Y_1]$. $\{M_t\}$ is called the exponential martingale or the Wald martingale. Later in this lecture we will use this martingale to establish the Lundberg inequality. First we do some computations. Define

$$\kappa(\alpha) := \mathbf{E}[e^{\alpha X_1}] = \mathbf{E}\left[e^{\alpha\left(c - \sum_{i=1}^{N_1} Y_i\right)}\right].$$

To compute this mean, we use the tower property of conditional expectations. Note that for deterministic n,

$$\mathbf{E}\left[e^{-\alpha\sum_{i=1}^{n} Y_i}\right] = \mathbf{E}[e^{-\alpha Y_1 - \cdots - \alpha Y_n}] = \mathbf{E}[e^{-\alpha Y_1}] \cdots \mathbf{E}[e^{-\alpha Y_n}] = \mathbf{E}[e^{-\alpha Y_1}]^n,$$

where we have used that $\{Y_i\}$ is an iid sequence. Letting $\kappa_Y(\alpha) = \mathbf{E}[e^{\alpha Y_1}]$, we get

$$\kappa(\alpha) = e^{\alpha c}\mathbf{E}\left[\mathbf{E}\left[e^{-\alpha\sum_{i=1}^{N_1} Y_i} \mid N_1\right]\right] = e^{\alpha c}\mathbf{E}\left[\mathbf{E}[e^{-\alpha Y_1}]^{N_1}\right] = e^{\alpha c}\mathbf{E}[\kappa_Y(-\alpha)^{N_1}].$$

As N_1 is Poisson distributed with rate $\lambda > 0$, we have

$$\begin{aligned}
\kappa(\alpha) &= e^{\alpha c}\sum_{n=0}^{\infty} \kappa_Y(-\alpha)^n P(N_1 = n) = e^{\alpha c}\sum_{n=0}^{\infty} \kappa_Y(-\alpha)^n \frac{\lambda^n}{n!} e^{-\alpha} \\
&= e^{\alpha c - \lambda}\sum_{n=0}^{\infty} \frac{(\kappa_Y(-\alpha)\lambda)^n}{n!} = e^{\alpha c - \lambda} e^{\kappa_Y(-\alpha)\lambda} = e^{\alpha c + \lambda(\kappa_Y(-\alpha) - 1)}.
\end{aligned}$$

It follows that the cumulant-generating function is

$$\Lambda(\alpha) = \alpha c + \lambda(\kappa_Y(-\alpha) - 1).$$

It is of interest to study the behaviour of Λ. We note that $\Lambda(0) = 0$ and that

$$\Lambda'(\alpha) = c - \lambda\kappa_Y'(-\alpha)$$

so $\Lambda'(0) = c - \lambda\kappa_Y'(0) = c - \lambda\mu_Y > 0$ by NPC. Hence the graph of Λ goes through and is increasing at the origin. In general, one can show that Λ is a convex continuous function on the interior of its domain (given that it exists in a neighbourhood of zero), see the supplementary part at the end. Also, since $P(X_1 < 0) > 0$, $\Lambda(\alpha) \to \infty$ as $\alpha \to -\infty$. From these facts we may conclude that if $\Lambda(\alpha)$ is finite for all $\alpha \leq 0$ or continuous on its entire domain, $\Lambda(-R) = 0$ for some $R > 0$. This R is unique and is called the *adjustment coefficient* .

7 The Lundberg inequality

From the above considerations, $M_t = e^{-RX_t}$ is a martingale. Let $T = \inf\{t \geq 0 : X_t < -u\}$ denote the time of ruin. This is a stopping time. Recall that we have $\mathcal{F}_0 = \{\emptyset, \Omega\}$ by convention and that $X_T < -u$. Optional sampling yields

$$1 = M_0 = \mathbf{E}[M_{t\wedge T}] = \mathbf{E}[M_T 1_{\{T\leq t\}}] + \mathbf{E}[M_t 1_{\{T>t\}}] \geq \mathbf{E}[M_T 1_{\{T\leq t\}}]$$

since $M_t \geq 0$. We have $1_{\{T\leq t\}} \to 1_{\{T<\infty\}}$ for $t \to \infty$. Using monotone convergence,

$$\begin{aligned} 1 &= \lim_{t\to\infty} \mathbf{E}[M_T 1_{\{T\leq t\}}] = \mathbf{E}[M_T \lim_{t\to\infty} 1_{\{T\leq t\}}] = \mathbf{E}[M_T 1_{\{T<\infty\}}] \\ &= \mathbf{E}[e^{-RX_T} 1_{\{T<\infty\}}] > \mathbf{E}[e^{Ru} 1_{\{T<\infty\}}] = e^{Ru} P(T < \infty) = e^{Ru}\psi(u). \end{aligned}$$

Rearranging, we have proved the following central result.

Theorem 7.1 (**The Lundberg inequality**). *Consider the Cramér-Lundberg model and assume that the adjustment coefficient R exists. Then*

$$\psi(u) < e^{-Ru}.$$

The Lundberg inequality says that if the moment-generating function of the claim sizes exist, then the ruin probability decays at an exponential rate as a function of the initial capital. In practice this means that claim sizes with exponentially decaying tails are not very dangerous for the company. The rest of this lecture concerns how strong the Lundberg inequality is. To study this problem in depth, we need tools from renewal theory.

8 Renewal theory

Definition 8.1. $\{N_t\}$ is a renewal process if

(i) $N_t \in \mathbb{N}_0$ i.e. the process attains non-negative integer values,

(ii) N_t is increasing i.e. $N_s \leq N_t$ for $s \leq t$ and

(iii) the interarrival times $\{\tau_i\}$ are iid.

Example 8.2. A Poisson process is a renewal process. In this case the interarrival times are exponentially distributed.

Example 8.3. Consider the number of times the lightbulb of a specific lamp is replaced. It would be unrealistic to assume that the time until the next change is independent of the time since the last replacement. Therefore a Poisson process is likely not suitable since the exponential distribution has the memorylessness property.

Let $\{N_t\}$ denote any renewal process and F the distribution function of τ_1. We will need the concept of a convolution.

Definition 8.4. Let f be a function and G a nondecreasing function. The *convolution* $f*G$ is defined as the function

$$f * G(x) = \int_0^x f(x-y)dG(y).$$

Let F be a non-decreasing function. The n-fold convolution F^{*n} is defined inductively by $F^{*2} = F * F$ and $F^{*n} = F^{*(n-1)} * F$. We define $F^{*0} = 1_{[0,\infty)}$.

The convolution naturally appears when considering sums of independent variables. If X and Y are non-negative independent variables with distribution functions F and G, respectively, then the distribution function of the sum is

$$P(X+Y \leq z) = \int_0^z F(z-y)dG(y) = F * G(z).$$

A conditioning argument yields this result easily. The argument is left as an exercise.

We are interested in computing the mean $\mathbf{E}[N_t]$. Recall that for any variable N taking values in $\mathbb{N}_0$, we have the following result

$$\mathbf{E}[N] = \sum_{n=1}^{\infty} P(N \geq n)$$

sometimes called the *tail sum formula*. The formula follows by Tonelli's theorem or more intuitively by counting the quantities

$$\begin{aligned} &P(N=1)+ \\ &P(N=2) + P(N=2)+ \\ &P(N=3) + P(N=3) + P(N=3)+ \\ &\qquad\vdots \end{aligned}$$

vertically instead of horisontally (as in the usual formula for the mean). We let the time of the n'th claim be denoted by $T_n = \tau_1 + \cdots \tau_n$. Then

$$\mathbf{E}[N_t] = \sum_{n=1}^{\infty} P(N_t \geq n) = \sum_{n=1}^{\infty} P(T_n \leq t) = \sum_{n=1}^{\infty} P(\tau_1 + \cdots + \tau_n \leq t) = \sum_{n=1}^{\infty} F^{*n}(t).$$

We are now ready to define the renewal function.

Definition 8.5. Let N_t be a Lévy process and define the process $\tilde{N}_t = 1 + N_t$. The function

$$U(t) = \mathbf{E}[\tilde{N}_t] = \sum_{n=0}^{\infty} F^{*n}(t)$$

is called the *renewal function*.

$U(t)$ counts the expected number of events in a renewal process. We want to show that $U(t)$ satisfies a special case of the so-called renewal equation. This equation is defined below.

Definition 8.6. Let F be a known distribution function and z a known function. The equation

$$Z(t) = z(t) + \int_0^t Z(t-s)dF(s), \quad t \geq 0$$

where Z is an unknown function is called the *renewal equation*.

Let us show that $\mathbf{E}[\tilde{N}_t]$ satisfies the renewal equation using a conditioning argument. Consider $\mathbf{E}[\tilde{N}_t \mid \tau_1 = s]$. If $s > t$, we know that only one event has occured, so in this case $\mathbf{E}[\tilde{N}_t \mid \tau_1 = s] = 1$. If $s \leq t$, one event has occured plus all the events that occur from time s to time t, hence $\mathbf{E}[\tilde{N}_t \mid \tau_1 = s] = 1 + \mathbf{E}[\tilde{N}_{t-s}]$. All in all,

$$\mathbf{E}[\tilde{N}_t \mid \tau_1 = s] = \begin{cases} 1, & s > t \\ 1 + \mathbf{E}[\tilde{N}_{t-s}], & s \leq t \end{cases}.$$

Using the tower property, we have

$$U(t) = \mathbf{E}[\tilde{N}_t] = \mathbf{E}[\mathbf{E}[\tilde{N}_t \mid \tau_1]] = 1 + \int_0^t \mathbf{E}[\tilde{N}_{t-s}]dF(s) = 1 + \int_0^t U(t-s)dF(s),$$

and we see that $U(t)$ satisfies the renewal equation with $z(t) = 1$ and F the distribution function of the interarrival times. We finish this section by determining the solution to the general renewal equation.

Proposition 8.7. *If the fixed function z in the renewal equation is bounded on bounded intervals, there exists a unique solution to the renewal equation which is bounded on bounded intervals. The solution is given by*

$$Z(t) = z * U(t) = \int_0^t z(t-x)dU(x).$$

Proof. We first verify that the function $Z(t) = z * U(x)$ is a solution. We have

$$\begin{aligned} Z(t) = z * \sum_{n=0}^{\infty} F^{*n}(t) &= z(t) + z * \left(\sum_{n=1}^{\infty} F^{*(n-1)} \right) * F(t) \\ &= z(t) + z * \left(\sum_{n=0}^{\infty} F^{*n} \right) * F(t) = z(t) + Z * F(t) \end{aligned}$$

so Z indeed solves the renewal equation. As for uniqueness, assume Z_1 is another solution which is bounded on bounded intervals. Then

$$|Z(t) - Z_1(t)| = \left| \int_0^t Z(t-x) - Z_1(t-x)dF(x) \right| \leq \int_0^t |Z(t-x) - Z_1(t-x)|dF(x).$$

We may continue this bound inductively and obtain

$$|Z(t) - Z_1(t)| \leq \int_0^t |Z_(t-x) - Z_1(t-x)| dF^{*n}(x) \leq \sup_{0 \leq x \leq t} |Z(x) - Z_1(x)| F^{*n}(t)$$

As Z and Z_1 are bounded on $[0,t]$ and F^{*n} tends to zero for $n \to \infty$, we get $Z(t) = Z_1(t)$ as desired. ■

9 The renewal theorems

The goal of this section is to study the solution

$$Z(t) = z * U(t) = \int_0^t z(t-s) dU(s).$$

further.

Definition 9.1. A distribution function F is called arithmetic if the support of F is $\{0, \gamma, 2\gamma, ...\}$ for some $\gamma \in \mathbb{N}$. The largest such γ is called the *span* of F. If F is not arithmetic, F is called nonarithmetic.

Recall that we write

$$f(t) \sim g(t) \text{ as } t \to \infty$$

if

$$\lim_{t \to \infty} \frac{f(t)}{g(t)} = 1.$$

We can now state the first form of the renewal theorem. For proofs of the results in this section, consult chapter XI in [3].

Theorem 9.2 (**Renewal theorem (first form)**). *If F is a nonarithmetic distribution function with $F(x) = 0$ for $x < 0$, then for $h > 0$,*

$$U(t+h) - U(t) \sim \frac{h}{\mu} \text{ as } t \to \infty$$

where $\mu = \int_0^\infty t dF(t)$ is the mean of F. If F is arithmetic then the result holds for h a multiple of the span γ.

Example 9.3. Let us consider the special case where our renewal process $\{N_t\}$ is Poisson process. In this case, the interarrival times are exponentially distributed with mean $\mu = 1/\lambda$. Thus

$$\begin{aligned} U(t+h) - U(t) &= \mathbf{E}[\tilde{N}_{t+h}] - \mathbf{E}[\tilde{N}_t] = 1 + \mathbf{E}[N_{t+h}] - (1 + \mathbf{E}[N_t]) \\ &= \lambda(t+h) - \lambda t = \lambda h = \frac{h}{\mu}. \end{aligned}$$

Hence the statement of the renewal theorem not only holds in the limit but at any given time.

We can now characterize the solution to the renewal equation using heuristic arguments. By the renewal theorem,

$$U(t+h) - U(t) \sim \frac{h}{\mu},$$

and we can think of this statement heuristically as $dU(t) \approx dt/\mu$. Hence

$$Z(t) = \int_0^t z(t-s)dU(s) \approx \int_0^t z(t-s)\frac{ds}{\mu} = \frac{1}{\mu}\int_0^t z(t-s)ds = \frac{1}{\mu}\int_0^t z(y)dy \to \frac{1}{\mu}\int_0^\infty z(y)dy$$

as $t \to \infty$. This statement can be stated in a formal manner which we will do shortly. We first need a new concept of integrability.

Definition 9.4. Let z be a function on $[0,\infty)$ and $h > 0$. Define

$$\overline{m}_k(h) = \sup\{z(t) : (k-1)h \le t < kh\}, \quad \underline{m}_k(h) = \inf\{z(t) : (k-1)h \le t < kh\}$$

i.e. $\overline{m}_k(h)$ and $\underline{m}_k(h)$ are the largest and smallest value in the interval $[(k-1)h, kh)$, respectively. Define the corresponding Riemann sums

$$\overline{\sigma}(h) = h\sum_{k=1}^\infty \overline{m}_k(h), \quad \underline{\sigma}(h) = h\sum_{k=1}^\infty \underline{m}_k(h).$$

z is called *directly Riemann integrable* if

$$-\infty < \lim_{h\downarrow 0}\underline{\sigma}(h) = \lim_{h\downarrow 0}\overline{\sigma}(h) < \infty.$$

The difference between ordinary Riemann integrability and direct Riemann integrability can be illustrated as follows. Whenever we want to compute an integral of the form

$$\int_0^\infty z(t)dt,$$

we start by computing

$$\int_0^K z(t)dt$$

and afterwards we let $K \to \infty$. This is an indirect way of evaluating the integral and it does not take into account how $z(t)$ fluctuates for large values of t. Direct Riemann integrability is a more strict notion of integrability that makes no distinction between finite and infinite intervals. Hence direct Riemann integrability makes fewer functions integrable. The difference between the two notions becomes more clear with an example.

Example 9.5. Consider the function $f : [0,\infty) \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} n, & x \in [n - 1/2^n, n + 1/2^n] \text{ for some } n \in \mathbb{N} \\ 0, & \text{else} \end{cases}.$$

Sketching the graph it is clear that f is well-defined. We claim that f is Riemann integrable but not directly Riemann integrable. To see that f is Riemann integrable, note that for $K \in \mathbb{N}$,

$$\int_0^K f(x)dx = \sum_{n=1}^{K-1}\frac{n}{2^{n-1}} + \frac{K}{2^{K+1}} \to \sum_{n=1}^\infty \frac{n}{2^{n-1}} < \infty.$$

To see that f is not directly Riemann integrable, let $h > 0$. As the length of each interval $[n-1/2^n, n+1/2^n]$ goes to zero, eventually each such interval will be contained in an interval of the form $[(k-1)h, kh)$ and thus $\overline{\sigma}(h) = \infty$.

Remark 9.6. It is not even sufficient for a function to be continuous and Riemann integrable and with limit zero for $t \to \infty$ for it to be directly Riemann integrable. See example (a) on page 363 in [3].

To establish that a function is directly Riemann integrable, the following lemma will be useful.

Lemma 9.7. *The following are sufficient conditions for a function $z(t)$ on $[0, \infty)$ to be directly Riemann integrable.*

(i) *$z(t)$ is monotone and Riemann integrable.*

(ii) *$z(t)$ is continuous a.e. and $a(t) \leq z(t) \leq b(t)$ for directly Riemann integrable $a(t)$ and $b(t)$.*

(iii) *$z(t) \geq 0$ is continuous a.e. and $\overline{\sigma}(h) < \infty$ for some $h > 0$.*

Proof. See page 69 of [1]. ■

We can now state the second form of the renewal theorem.

Theorem 9.8 (**Renewal theorem (second form)**). *If F is a nonarithmetic distribution function and z is directly Riemann integrable then*

$$Z(t) \to \frac{1}{\mu} \int_0^\infty z(y)dy$$

for $t \to \infty$.

10 Exact asymptotics and the perturbation argument

Let

$$C_t = u + ct - \sum_{i=1}^{N_t} Y_i$$

be the standard Cramér-Lundberg process where $\{N_t\}$ is a Poisson process independent of the iid sequence $\{Y_i\}$ of claim sizes with common distribution function G. Recall that the Lundberg inequality states that $\psi(u) < e^{-Ru}$ where R is the non-zero solution to $\Lambda(-\alpha) = 0$. We want to use our new tools from renewal theory to study the strength of this inequality. This will lead to the so-called Cramér-Lundberg estimate.

The preliminary work on establishing the Cramér-Lundberg estimate uses a so called *perturbation argument* . Let $\delta(u) = 1 - \psi(u)$ denote the survival probability. Consider a "small" $h > 0$. In the time interval $[0, h]$, there are three cases:

(1) $N_h = 0$ (no claims).

(2) $N_h = 1, Y_1 = y$ (one claim of size y).

(3) $N_h \geq 2$ (more than one claim).

We now consider what happens with the survival probability $\delta(u)$ when we condition on one of these events.

(1) $N_h = 0$: No claims in the interval $[0, h]$ corresponds to restarting the process at time zero with initial capital $u + ch$. Also, $P(N_h = 0) = 1 - \lambda h + o(h)$.

(2) $N_h = 1, Y_1 = y$: If $y \leq u$, we claim that the survival probability changes to $\delta(u - y) + o(1)$. To justify this, we note that ruin probabilities are continuous from the right due to continuity from above for measures. In particular, $\delta(u - y + ch) \to \delta(u - y)$ for $h \to 0^+$. Also, $P(N_h = 1) = \lambda h + o(h)$. If $y > u$, then $\delta(u + ch) \to 0$ for $h \to 0$ because ruin has already happened. Hence this case is only interesting if $y \leq u$.

(3) $N_h \geq 2$ (more than one claim). This case is negligible by the definition of a Poisson process. We have $P(N_h \geq 2) = o(h)$.

We can now collect these observations and make a partition/conditioning argument to obtain

$$\delta(u) = \delta(u + ch)(1 - \lambda h + o(h)) + \int_0^u \delta(u - y) + o(1) dG(y)(\lambda h + o(h)) + o(h).$$

Term one corresponds to case (1) and so forth. Elementary algebra yields

$$\delta(u + ch) - \delta(u) = \lambda h \left(\delta(u + ch) - \int_0^u \delta(u - y) + o(1) dG(y) \right) + o(h)$$

and thus

$$c \frac{\delta(u + ch) - \delta(u)}{ch} = \lambda \left(\delta(u + ch) - \int_0^u \delta(u - y) + o(1) dG(y) \right) + o(1).$$

Assuming that the derivative $\delta'(u)$ exists, we get

$$\frac{c}{\lambda} \delta'(u) = \delta(u) - \int_0^u \delta(u - y) dG(y)$$

which is a so-called "integral differential equation". In the next lecture, we will use this expression to derive a renewal equation which will lead to the Cramér-Lundberg estimate.

Supplementary: Convexity of the cumulant generating function

Proposition 10.1. *Let X be a random variable where the moment generating function $\kappa(\alpha)$ is finite in some neighbourhood of zero. The cumulant generating function $\Lambda(\alpha) = \log \kappa(\alpha)$ is convex.*

Proof. Let $\alpha = \lambda \alpha_1 + (1 - \lambda)\alpha_2$ with $0 < \lambda < 1$. Letting $p = 1/\lambda$ and $q = 1/(1 - \lambda)$, we have $1/p + 1/q = 1$. Then by Hölder's inequality,

$$\begin{aligned} \kappa(\alpha) = \mathbf{E}\left[e^{\alpha X}\right] &= \mathbf{E}[e^{\lambda \alpha_1 X} e^{(1-\lambda)\alpha_2 X}] \leq \mathbf{E}[e^{\lambda p \alpha_1 X}]^{1/p} \mathbf{E}[e^{(1-\lambda) q \alpha_2 X}]^{1/q} \\ &= \mathbf{E}[e^{\alpha_1 X}]^{\lambda} \mathbf{E}[e^{\alpha_2 X}]^{1-\lambda} = \kappa(\alpha_1)^{\lambda} \kappa(\alpha_2)^{1-\lambda} \end{aligned}$$

and taking logarithms yields

$$\Lambda(\alpha) \leq \lambda\Lambda(\alpha_1) + (1-\lambda)\Lambda(\alpha_2).$$

■

Corollary 10.2. *The set* $\{\alpha \in \mathbb{R} : \Lambda(\alpha) < \infty\}$ *is an interval.*

Week 3 - The Cramér-Lundberg estimate and subexponential distributions

11 The Cramér-Lundberg estimate

At the end of the previous lecture, we derived the following integral differential equation for the survival probability $\delta(u)$:

$$\frac{c}{\lambda}\delta'(u) = \delta(u) - \int_0^u \delta(u-y)dG(y).$$

In this section, we continue on our quest to derive a renewal equation involving the probability of ruin. We get the idea to integrate the left and right hand side. Explicitly, replace u with x and integrate from 0 to u:

$$\frac{c}{\lambda}(\delta(u) - \delta(0)) = \int_0^u \delta(x)dx - \int_0^u \int_0^x \delta(x-y)dG(y)dx.$$

For simplicity, write

$$I(u) = \int_0^u \delta(x)dx, \quad A(u) = \int_0^u \int_0^x \delta(x-y)dG(y)dx.$$

To compute the integral $A(u)$, we interchange the order of integration and apply integration by parts to get

$$\begin{aligned}
A(u) &= \int_0^u \int_0^x \delta(x-y)dG(y)dx = \int_0^u \int_y^u \delta(x-y)dxdG(y) \\
&= \int_0^u \int_0^{u-y} \delta(w)dwdG(y) = \int_0^u I(u-y)dG(y) = -\int_0^u I(u-y)d(1-G)(y) \\
&= -\left[I(u-y)(1-G(y))\right]_0^u - \int_0^u \delta(u-y)(1-G(y))dy \\
&= I(u) - \int_0^u \delta(u-y)(1-G(y))dy,
\end{aligned}$$

so that

$$\frac{c}{\lambda}(\delta(u) - \delta(0)) = \int_0^u \delta(u-y)(1-G(y))dy.$$

We rewrite this in terms of the ruin probability:

$$-\frac{c}{\lambda}(\psi(u)-\psi(0)) = \int_0^u (1-\psi(u-y))(1-G(y))dy.$$

Letting $u \to \infty$ and applying monotone convergence yields

$$\frac{c}{\lambda}\psi(0) = \int_0^\infty (1-G(y))dy$$

and thus

$$\psi(0) = \frac{\lambda}{c}\int_0^\infty (1-G(y))dy.$$

Substituting this into the equation

$$-\frac{c}{\lambda}(\psi(u)-\psi(0)) = \int_0^u (1-\psi(u-y))(1-G(y))dy$$

gives

$$\frac{c}{\lambda}\psi(u) = \int_u^\infty (1-G(y))dy + \int_0^u \psi(u-y)(1-G(y))dy$$

i.e.

$$\psi(u) = \frac{\lambda}{c}\int_u^\infty (1-G(y))dy + \int_0^u \psi(u-y)\frac{\lambda}{c}(1-G(y))dy.$$

This looks a lot like a renewal equation. The question is whether the function F with the dynamics

$$dF(y) = \frac{\lambda}{c}(1-G(y))dy$$

is a distribution function. This can be checked:

$$\int_0^\infty dF(y) = \frac{\lambda}{c}\int_0^\infty (1-G(y))dy = \frac{\lambda\mu_Y}{c} < 1$$

due to the net profit condition. Hence F is not a distribution function and we are not done yet. To remedy this issue, we multiply the expression for $\psi(u)$ above with $e^{\alpha u}$. This gives the equation

$$e^{\alpha u}\psi(u) = e^{\alpha u}\frac{\lambda}{c}\int_u^\infty (1-G(y))dy + \int_0^u e^{\alpha(u-y)}\psi(u-y)e^{\alpha y}dF(y).$$

Let us fix some notation:

$$Z_\alpha(u) = e^{\alpha u}\psi(u), \quad z_\alpha(u) = e^{\alpha u}\frac{\lambda}{c}\int_u^\infty (1-G(y))dy, \quad dF_\alpha(y) = e^{\alpha y}dF(y).$$

With this new notation, the equation above becomes

$$Z_\alpha(u) = z_\alpha(u) + \int_0^u Z_\alpha(u-y)dF_\alpha(y)$$

and the goal now is to determine α such that F_α is a distribution function. We compute

$$\begin{aligned}\int_0^\infty dF_\alpha(y) &= \int_0^\infty e^{\alpha y}\frac{\lambda}{c}(1-G(y))dy = \int_0^\infty e^{\alpha y}\frac{\lambda}{c}\int_y^\infty dG(x)dy \\ &= \frac{\lambda}{c}\int_0^\infty\int_y^\infty e^{\alpha y}dG(x)dy = \frac{\lambda}{c}\int_0^\infty\int_0^x e^{\alpha y}dydG(x) \\ &= \frac{\lambda}{c}\int_0^\infty\left(\frac{1}{\alpha}e^{\alpha x}-\frac{1}{\alpha}\right)dG(x) = \frac{\lambda}{c\alpha}\int_0^\infty e^{\alpha x}-1dG(x) \\ &= \frac{\lambda}{c\alpha}(\kappa_Y(\alpha)-1)\end{aligned}$$

and so we need to determine α such that

$$\frac{\lambda}{c\alpha}(\kappa_Y(\alpha)-1)=1.$$

We recognize this as the equation for the Lundberg coefficient. Hence $\alpha = R$ is the solution. We now have the (proper) renewal equation

$$Z_R(u) = z_R(u) + \int_0^u Z_R(u-y)dF_R(y).$$

Assuming that G (and hence F_R) is non-arithmetic and that z_R is directly Riemann-integrable, the second form of the renewal theorem yields

$$Z_R(u) \to C := \frac{1}{\mu_R}\int_0^\infty z_R(x)dx < \infty \quad \text{for} \quad u\to\infty.$$

where μ_R denotes the mean of the distribution given by F_R i.e.

$$\mu_R = \int_0^\infty xdF_R(x) = \int_0^\infty xe^{Rx}\frac{\lambda}{c}(1-G(x))dx.$$

Recalling that $\psi(u) = e^{-Ru}Z_R(u)$, the result can be stated as follows:

Theorem 11.1 (**The Cramér-Lundberg estimate**). *Assume the setup of the Cramér-Lundberg model with a non-arithmetic distribution function G for the claim sizes and that the adjustment coefficient R exists. Then*

$$\psi(u) \sim Ce^{-Ru} \quad \textit{for} \quad u\to\infty.$$

where C is given by

$$C = \frac{c-\lambda\mu_Y}{\lambda\kappa_Y'(R)-c}.$$

Two things need to be checked, namely

(i) z_R is directly Riemann integrable.

(ii) $C = (c-\lambda\mu_Y)/(\lambda\kappa_Y'(R)-c)$.

Both of these assertions are left as exercises. We have now seen a classical approach to estimate the ruin probability. Many alternative methods exist, but most of them are obsolete since we can achieve good estimates simply by simulating. One method which still carries modern relevance is the method of Laplace transforms.

12 Laplace transforms

Definition 12.1. Let $f : [0, \infty) \to \mathbb{R}$ be a function. The function

$$\hat{f}(\alpha) = \int_0^\infty e^{-\alpha x} f(x) dx$$

is called the (classical) *Laplace transform* of f.

There is a connection between Laplace transforms and random variables. Let Y be a non-negative random variable with density f. Then

$$\hat{f}(\alpha) = \int_0^\infty e^{-\alpha x} f(x) dx = \mathbf{E}[e^{-\alpha Y}] = \kappa_Y(-\alpha)$$

where κ_Y as usual denotes the moment generating function of Y. The Laplace transform has many nice properties. We collect these in the following proposition.

Proposition 12.2. *Let $f : [0, \infty) \to \mathbb{R}$ be a function. The following properties hold:*

(i) If $f(x) \geq 0$ a.e. then

$$\hat{f}(\alpha) \leq \hat{f}(\beta) \quad \Leftrightarrow \quad \alpha \geq \beta.$$

(ii) $\widehat{|f|}(\alpha) < \infty$ implies $\widehat{|f|}(\beta)$ for all $\beta \geq \alpha$.

(iii) If f' exists a.e. and $\widehat{|f|}(\alpha) < \infty$ then

$$\widehat{f'}(\alpha) = \alpha \hat{f}(\alpha) - f(0).$$

(iv) $\hat{f}$ is a convex function when f is positive a.e.

(v) $\hat{f}$ is differentiable on the interior of its domain.

(vi) $\hat{f}$ determines f uniquely (if $\hat{f} = \hat{g}$ then $f = g$).

Proof. Assertions (i) and (ii) are clear. To prove (iii), we use integration by parts as follows:

$$\widehat{f'}(\alpha) = \int_0^\infty e^{-\alpha x} f'(x) dx = \left[e^{-\alpha x} f(x)\right]_0^\infty + \int_0^\infty f(x) \alpha e^{-\alpha x} dx = \alpha \hat{f}(\alpha) - f(0).$$

In the last equality, we used that the Laplace transform of $|f|$ is finite in α which implies that $f(x)e^{-\alpha x} \to 0$ for $x \to \infty$. To prove (iv) let $g(\alpha) = \log \hat{f}(\alpha)$ and let $\lambda \in [0, 1]$. We then have

$$\begin{aligned}
g(\lambda\alpha + (1-\lambda)\beta) &= \log \int_0^\infty e^{-(\lambda\alpha + (1-\lambda)\beta)x} f(x) dx \\
&= \log \int_0^\infty (e^{-\alpha x} f(x))^\lambda (e^{-\beta x} f(x))^{1-\lambda} dx \\
&\leq \log \left(\left(\int_0^\infty e^{-\alpha x} f(x) dx \right)^\lambda \left(\int_0^\infty e^{-\beta x} f(x) dx \right)^{1-\lambda} \right) \\
&= \lambda \log \int_0^\infty e^{-\alpha x} f(x) dx + (1-\lambda) \log \int_0^\infty e^{-\beta x} f(x) dx \\
&= \lambda g(\alpha) + (1-\lambda) g(\beta)
\end{aligned}$$

where we apply Hölder's inequality in the third line. It follows that g is convex. Taking the exponential function on both sides gives the result. (v) follows by interchanging differentiation and integration. The details are left to the reader. A proof of (vi) can be found in section 3.1 of [8].

■

We will now compute the Laplace transform of the survival probability. This will allow us to determine $\delta(u)$ (and hence $\psi(u)$) explicitly in the case of exponential claims. Recall that we have the differential integral equation for the survival probability given by

$$\frac{c}{\lambda}\delta'(u) = \delta(u) - \int_0^u \delta(u-y)dG(y).$$

Taking the Laplace transform of both sides gives the equation

$$\frac{c}{\lambda}\left(\alpha\hat{\delta}(\alpha) - \delta(0)\right) = \hat{\delta}(\alpha) - \int_0^\infty \int_0^u \delta(u-y)dG(y)e^{-\alpha u}du$$

where we have used the properties of the Laplace transform. It remains to compute the integral on the right hand side:

$$\begin{aligned}\int_0^\infty \int_0^u \delta(u-y)dG(y)e^{-\alpha u}du &= \int_0^\infty \int_y^\infty \delta(u-y)e^{-\alpha u}dudG(y)\\ &= \int_0^\infty \int_0^\infty \delta(v)e^{-\alpha v}dve^{-\alpha y}dG(y)\\ &= \int_0^\infty \hat{\delta}(\alpha)e^{-\alpha y}dG(y) = \hat{\delta}(\alpha)\kappa_Y(-\alpha).\end{aligned}$$

Hence our equation becomes

$$\frac{c}{\lambda}\left(\alpha\hat{\delta}(\alpha) - \delta(0)\right) = \hat{\delta}(\alpha)(1-\kappa_Y(-\alpha)).$$

Earlier we computed

$$\psi(0) = \frac{\lambda}{c}\int_0^\infty (1-G(y))dy = \frac{\lambda\mu_Y}{c} \quad \text{and thus} \quad \delta(0) = 1 - \frac{\lambda\mu_Y}{c}.$$

Plugging this expression for $\delta(0)$ into the previous equation yields

$$-c + \lambda\mu_Y + c\alpha\hat{\delta}(\alpha) = \lambda\hat{\delta}(\alpha)(1+\kappa_Y(-\alpha))$$

and a bit of high school algebra gives us the final result

$$\hat{\delta}(\alpha) = \frac{c-\lambda\mu_Y}{c\alpha - \lambda(1-\kappa_Y(-\alpha))} \quad \text{for } \alpha > 0.$$

Note that if $\alpha \leq 0$ then $\hat{\delta}(\alpha) = \infty$ and nothing interesting happens. We are interested in the survival probability itself and not its Laplace transform. Due to the uniqueness property of the Laplace transform, we should be able to invert $\hat{\delta}(\alpha)$ and obtain $\delta(u)$. There exist numerical methods to do this but we are interested in an analytical expression if possible. This turns out to be possible in some rare cases.

Example 12.3. Assume that we have exponential claim sizes $Y_i \sim \text{Exp}(\theta)$. Then $\mu_Y = 1/\theta$ and

$$\kappa_Y(-\alpha) = \frac{\theta}{\theta + \alpha}.$$

In this case $\hat{\delta}(\alpha)$ becomes

$$\hat{\delta}(\alpha) = \frac{c - \lambda\frac{1}{\theta}}{c\alpha - \lambda\left(1 - \frac{\theta}{\theta+\alpha}\right)} = \frac{1}{\alpha} - \frac{\lambda}{c\theta}\frac{1}{\alpha + \left(\theta - \frac{\lambda}{c}\right)}.$$

Recall that $\psi(u)$ decays like a constant times an exponential function so there are not many choices for a candidate of $\delta(u)$. Consider a function of the form

$$f(u) = 1 - Ae^{-\beta u}$$

and let $\alpha > 0$. Then

$$\hat{f}(\alpha) = \int_0^\infty f(u)e^{-\alpha u}du = \int_0^\infty e^{-\alpha u} - Ae^{-(\alpha+\beta)u}du = \frac{1}{\alpha} - A\frac{1}{\alpha+\beta}.$$

By the uniqueness of the Laplace transform, $\delta(u)$ must be a function of the same form as f. We can identify $A = \lambda/c\theta$ and $\beta = \theta - \frac{\lambda}{c}$. Hence we obtain the exact survival and ruin probabilities

$$\delta(u) = 1 - \frac{\lambda}{c\theta}e^{-\left(\theta - \frac{\lambda}{c}\right)u}, \quad \text{and} \quad \psi(u) = \frac{\lambda}{c\theta}e^{-\left(\theta - \frac{\lambda}{c}\right)u}.$$

13 Subexponential distributions

Recall the setup for the important results so far. In the Cramér-Lundberg process

$$C_t = u + ct - \sum_{i=1}^{N_t} Y_i, \quad X_t = C_t - u, \quad Y_i \sim G$$

it only makes sense to talk about the Lundberg inequality and the Cramér-Lundberg estimate when the adjustment coefficient R exists. Recall that R is defined as the strictly positive number that satisfies $\Lambda(-R) = 0$ where

$$\Lambda(\alpha) = \log \mathbf{E}[e^{\alpha X_1}].$$

What are necessary conditions for R to exist? We definitely need that $\Lambda(-\alpha) < \infty$ for some $\alpha > 0$ where

$$\Lambda(-\alpha) = -\alpha c + \lambda(\kappa_Y(\alpha) - 1)$$

hence it is necessary that $\kappa_Y(\alpha) < \infty$ for some $\alpha > 0$. What is a sufficient condition for R to exist? If $\kappa_Y(\alpha) < \infty$ for some $\alpha > 0$ and κ_Y is continuous for all α where the function exists. The reason continuity matters is that a typical issue is that $\kappa_Y(\alpha) < \infty$ for $\alpha > 0$. Some classical distributions in non-life insurance include

- $Y_i \sim \text{Exp}(\theta)$ with density $f(x) = \theta e^{-\theta x}$ for $x > 0$.
- $Y_i \sim \text{Gamma}(\alpha, \beta)$ with density $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$ for $x > 0$.

- $Y_i \sim |W_i|$ with $W_i \sim \mathcal{N}(0,1)$ (the folded/truncated normal) .

The exponential and gamma distributions are used a lot because R can be computed explicitly. Usually R needs to be approximated numerically. But what if R does not exist at all? This happens for heavy-tailed distributions. We have

$$\kappa_Y(\alpha) = \int_0^\infty e^{\alpha y} dG(y)$$

so the question is whether G decays slower than exponential functions. If this is the case, $\kappa_Y(\alpha) = \infty$ for all $\alpha > 0$. Some examples of this behaviour follow.

Example 13.1. Consider the lognormal distribution with parameters (μ, σ^2). This distribution is obtained by taking the exponential function of a $\mathcal{N}(\mu, \sigma^2)$ distributed variable. In particular the distribution is supported on $(0, \infty)$. The density of this distribution is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}x} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}} \quad \text{for } x > 0.$$

It can be shown that $\kappa_Y(\alpha) = \infty$ for all $\alpha > 0$ for Y with this distribution.

Example 13.2. Consider the Pareto distribution with parameters $\alpha, \beta > 0$. The survival function of a variable Y with this distribution is given by

$$\overline{G}(x) = P(Y > x) = \frac{\beta^\alpha}{(\beta + x)^\alpha} \quad \text{for } x \geq 0.$$

This function decays slower than any exponential function and thus $\kappa_Y(\alpha) = \infty$ for any $\alpha > 0$. This distribution belongs to the class of *regularly varying* distributions . This class has survival functions of the form

$$\overline{G}(x) = L(x)x^{-\alpha}$$

where L is a *slowly varying function* . This is a function $L : (0, \infty) \to \mathbb{R}$ which satisfies

$$\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1$$

for all $t > 0$. Examples include logarithms and constants.

The rest of this lecture is dedicated to introducing a useful notion of being "heavy-tailed", namely the concept of subexponential distributions

Definition 13.3. A distribution function G on $[0, \infty)$ is *subexponential* if

$$\lim_{x\to\infty} \frac{1 - G^{*2}(x)}{1 - G(x)} = 2.$$

We denote the set of subexponential distribution functions by $\mathcal{S}$.

Remark 13.4. If $G \in \mathcal{S}$ and $Y_1, Y_2 \sim G$ are independent, the above definition says that

$$\lim_{x\to\infty} \frac{P(Y_1 + Y_2 > x)}{P(Y_1 > x)} = 2.$$

An interpretation of the definition is as follows. For general Y_1 and Y_2, we have

$$P(Y_1 + Y_2 > x) \geq P(Y_1 > x) + P(Y_2 > x) - P(Y_1 > x, Y_2 > x).$$

since one possible way for $Y_1 + Y_2 > x$ to occur is if exactly one of Y_1 or Y_2 surpasses x, that is, one large jump occurs. If Y_1 and Y_2 are independent,

$$P(Y_1 > x)+P(Y_2 > x)-P(Y_1 > x, Y_2 > x) = P(Y_1 > x)+P(Y_2 > x)-P(Y_1 > x)P(Y_2 > x)$$

and it follows that

$$\lim_{x\to\infty} \frac{P(Y_1 + Y_2 > x)}{P(Y_1 > x)} \geq 2.$$

Hence a distribution is subexponential if for independent $Y_1, Y_2 \sim G$, $Y_1 + Y_2 > x$ happens because of one large jump as x gets large. In particular, ruin will occur because of one large claim. This is very different from classical distributions where one can prove that ruin happens gradually.

Example 13.5. Let $Y_1, Y_2 \sim \text{Exp}(\theta)$ be independent. Then $Y_1 + Y_2 \sim \text{Gamma}(2, \theta)$. By L'Hospital's rule,

$$\lim_{x\to\infty} \frac{P(Y_1 + Y_2 > x)}{P(Y_1 > x)} = \lim_{x\to\infty} \frac{\int_x^\infty \theta^2 y e^{-\theta y} dy}{e^{-\theta x}} = \lim_{x\to\infty} \frac{-\theta^2 x e^{-\theta x}}{-\theta e^{-\theta x}} = \infty$$

so the exponential distribution is not subexponential.

We now consider some properties of subexponential distributions.

Proposition 13.6. *If $G \in \mathcal{S}$, then*

$$\lim_{x\to\infty} \frac{1 - G(x - a)}{1 - G(x)} = 1 \quad \textit{for all } a \in \mathbb{R}.$$

Proof. Let $a \geq 0$. We have

$$\begin{aligned}
\frac{1 - G^{*2}(x)}{1 - G(x)} - 1 &= \frac{G(x) - G^{*2}(x)}{1 - G(x)} \\
&= \int_0^a \frac{1 - G(x - y)}{1 - G(x)} dG(y) + \int_a^x \frac{1 - G(x - y)}{1 - G(x)} dG(y) \\
&\geq G(a) + \frac{1 - G(x - a)}{1 - G(x)} (G(x) - G(a)).
\end{aligned}$$

Rearranging gives

$$1 \leq \frac{1 - G(x - a)}{1 - G(x)} \leq (G(x) - G(a))^{-1} \left(\frac{1 - G^{*2}(x)}{1 - G(x)} - 1 - G(a) \right)$$

and letting $x \to \infty$ yields

$$1 \leq \lim_{x\to\infty} \frac{1 - G(x - a)}{1 - G(x)} \leq (1 - G(a))^{-1}(1 - G(a)) = 1$$

which proves the result for the case $a \geq 0$. If $a < 0$, then

$$\lim_{x\to\infty} \frac{1 - G(x - a)}{1 - G(x)} = \lim_{x\to\infty} \frac{1}{\frac{1-(G(x-a)-(-a))}{1-G(x-a)}} = 1$$

by the previous case. This completes the proof. ∎

This result has a natural interpretation. Let $Y \sim G$ where $G \in \mathcal{S}$. A special case of the proposition says that

$$\lim_{x\to\infty} \frac{P(Y > x + a)}{P(Y > x)} = \lim_{x\to\infty} P(Y > x + a \mid Y > x) = 1$$

for any $a > 0$. This means that if we know Y has attained some large value x, it is very likely to attain an even larger value as well. As an exercise the reader can verify that this is not the case for the exponential distribution for example.

Proposition 13.7. *Assume $G \in \mathcal{S}$. Then for all $n \in \mathbb{N}$, we have*

$$\lim_{x\to\infty} \frac{1 - G^{*n}(x)}{1 - G(x)} = n.$$

Proof. We use induction on n. By assumption, the assertion holds for $n = 2$. Assume that the claim holds for some n. Let $\varepsilon > 0$ and choose a such that

$$\left| \frac{1 - G^{*n}(x)}{1 - G(x)} - n \right| < \varepsilon$$

for $x \geq a$. Write

$$\begin{aligned}
\frac{1 - G^{*(n+1)}(x)}{1 - G(x)} &= \frac{1 - G(x) + G(x) - G^{*(n+1)}(x)}{1 - G(x)} = 1 + \frac{G(x) - G^{*(n+1)}(x)}{1 - G(x)} \\
&= 1 + \int_0^x \frac{1 - G^{*n}(x - y)}{1 - G(x)} dG(y) \\
&= 1 + \int_0^{x-a} \frac{1 - G^{*n}(x - y)}{1 - G(x - y)} \frac{1 - G(x - y)}{1 - G(x)} dG(y) + \int_{x-a}^x \frac{1 - G^{*n}(x - y)}{1 - G(x)} dG(y).
\end{aligned}$$

We can bound the second integral by

$$\int_{x-a}^x \frac{1 - G^{*n}(x - y)}{1 - G(x)} dG(y) \leq \frac{G(x) - G(x - a)}{1 - G(x)} = \frac{1 - G(x - a)}{1 - G(x)} - 1$$

and by the previous proposition this tends to zero for $x \to \infty$. Before turning to the other integral, consider

$$\int_0^{x-a} n \frac{1 - G(x - y)}{1 - G(x)} dG(y) = n \left(\frac{G(x) - G^{*2}(x)}{1 - G(x)} - \int_{x-a}^x \frac{1 - G(x - y)}{1 - G(x)} dG(y) \right).$$

The first term in the parantheses on the right converges to one and the other term converges to zero by the same argument as above. Hence the left hand side converges to n as $x \to \infty$. We now get

$$\begin{aligned}
&\left| \int_0^{x-a} \left(\frac{1 - G^{*n}(x - y)}{1 - G(x - y)} - n \right) \frac{1 - G(x - y)}{1 - G(x)} dG(y) \right| \\
&\leq \varepsilon \left(\frac{G(x) - G^{*2}(x)}{1 - G(x)} - \int_{x-a}^x \frac{1 - G(x - y)}{1 - G(x)} dG(y) \right) \to \varepsilon
\end{aligned}$$

as $x \to \infty$ using the same arguments as before. Combining all our arguments yields

$$\lim_{x\to\infty} \left| \frac{1 - G^{*(n+1)(x)}}{1 - G(x)} - (n + 1) \right| \leq \varepsilon.$$

As $\varepsilon > 0$ was chosen arbitrarily, the claim follows. ∎

The following property is useful for domination arguments.

Proposition 13.8. *Let $G \in \mathcal{S}$. For any $\varepsilon > 0$, there exists a finite constant D (dependent on ε) such that*

$$\frac{1 - G^{*n}(x)}{1 - G(x)} \leq D(1+\varepsilon)^n \quad \text{for all } x > 0 \text{ and } n \in \mathbb{N}.$$

Proof. We use induction on n. The claim is trivial for $n = 1$ so assume the result holds for some n. Then $(1 - G^{*n}(x)/(1 - G(x)$ is bounded, so we may define

$$\alpha_n = \sup_{t \geq 0} \frac{1 - G^{*n}(t)}{1 - G(t)}.$$

We note that

$$\frac{1 - G^{*2}(x)}{1 - G(x)} = 1 + \frac{G(x) - G^{*2}(x)}{1 - G(x)}$$

and since G is subexponential, we may find a $T \geq 0$ such that

$$\sup_{t \geq T} \frac{G(t) - G^{*2}(t)}{1 - G(t)} < 1 + \frac{\varepsilon}{2}.$$

We can now estimate

$$\begin{aligned}
\alpha_{n+1} &\leq 1 + \sup_{0 \leq t \leq T} \int_0^t \frac{1 - G^{*n}(t-y)}{1 - G(t)} dG(y) + \sup_{t \geq T} \int_0^t \frac{1 - G^{*n}(t-y)}{1 - G(t)} dG(y) \\
&\leq 1 + \frac{1}{1 - G(T)} + \sup_{t \geq T} \int_0^t \frac{1 - G^{*n}(t-y)}{1 - G(t-y)} \frac{1 - G(t-y)}{1 - G(t)} dG(y) \\
&\leq 1 + \frac{1}{1 - G(T)} + \alpha_n \sup_{t \geq T} \frac{G(t) - G^{*2}(t)}{1 - G(t)} \leq 1 + \frac{1}{1 - G(T)} + \alpha_n \left(1 + \frac{\varepsilon}{2}\right).
\end{aligned}$$

Now choose

$$D = \max\left\{ \frac{2(1 + 1/(1 - G(T)))}{\varepsilon}, 1 \right\}.$$

The reason for the max is to ensure that $\alpha_1 = 1 < D(1+\varepsilon)$.

■

In the start of the next lecture we will see some methods to determine whether a distribution is subexponential.

Week 4 - Ladder heights and subexponential ruin

14 Checking subexponentiality

A possible way to check that a distribution is subexponential is the following "closure property". If F is subexponential and G behaves in the sammer manner as F in the tail, G is also subexponential.

Proposition 14.1 (**Closure property of subexponential distributions**). *Suppose F and G are distribution functions on $[0,\infty)$ with $F \in \mathcal{S}$ and such that*

$$\lim_{x\to\infty} \frac{\overline{G}(x)}{\overline{F}(x)} = c$$

for some constant $c > 0$. Then $G \in \mathcal{S}$ also.

Proof. See Lemma 1.35 in [6]. ■

Remark 14.2. This proposition can be extended to more distribution functions in the following way. If G_i, $i = 1, 2$, satisfy

$$\overline{G_i}(x) \sim c_i \overline{F}(x) \quad \text{for } i = 1, 2, \quad x \to \infty$$

and $F \in \mathcal{S}$, then $G_1, G_2 \in \mathcal{S}$ also.

Before providing a more concrete method to establish subexponentiality, we need a definition.

Definition 14.3. Let G be a distribution function with density g. The *failure rate* of G is given by

$$\lambda(x) = \frac{g(x)}{\overline{G}(x)}.$$

Proposition 14.4. *Let G have density g and failure rate λ. Assume that $\lambda(x) \to 0$ for $x \to \infty$ and that there exists a fixed $x_0 \in \mathbb{R}$ such that for $x \geq x_0$, $\lambda(x)$ is decreasing. If*

$$\int_0^\infty e^{x\lambda(x)} g(x)dx < \infty,$$

then $G \in \mathcal{S}$.

Proof. This proof is from [2]. Splitting the area of integration into the two parts $[0, x_0]$ and $[x_0, \infty)$, we note that the integral over $[0, x_0]$ is always finite. Hence it suffices to consider the case $x_0 = 0$ so that $\lambda(x)$ is everywhere decreasing. Define

$$\Lambda(x) = \int_0^x \lambda(y)dy,$$

so that $\overline{G} = e^{-\Lambda(x)}$ (exercise). We get

$$\begin{aligned}
\frac{1 - G^{*2}(x)}{1 - G(x)} - 1 = \frac{1 - G^{*2}(x) - (1 - G(x))}{1 - G(x)} &= \frac{G(x) - G^{*2}(x)}{1 - G(x)} \\
&= \int_0^x \frac{1 - G(x - y)}{1 - G(x)} dG(y) = \int_0^x \frac{\overline{G}(x - y)}{\overline{G}(x)} g(y)dy \\
&= \int_0^x e^{\Lambda(x) - \Lambda(x-y) - \Lambda(y)} \lambda(y)dy \\
&= \int_0^{x/2} e^{\Lambda(x) - \Lambda(x-y) - \Lambda(y)} \lambda(y)dy + \int_0^{x/2} e^{\Lambda(x) - \Lambda(x-y) - \Lambda(y)} \lambda(x - y)dy
\end{aligned}$$

where the last equality follows from applying the substitution $y \mapsto x - y$. For $y < x/2$,

$$\Lambda(x) - \Lambda(x - y) = \int_{x-y}^x \lambda(v)dv \leq (x - (x - y))\lambda(x - y) = y\lambda(x - y) \leq y\lambda(y)$$

where we have used that λ is decreasing. Hence

$$\int_0^{x/2} e^{\Lambda(x) - \Lambda(x-y) - \Lambda(y)} \lambda(y)dy \leq \int_0^{x/2} e^{y\lambda(y) - \Lambda(y)} \lambda(y)dy = \int_0^{x/2} e^{y\lambda(y)} g(y)dy$$

which is finite by assumption. The bound $\Lambda(x) - \Lambda(x - y) \leq y\lambda(x - y)$ shows that $\Lambda(x) - \Lambda(x - y) \to 0$ for $x \to \infty$. We can now apply dominated convergence as follows:

$$\begin{aligned}
\lim_{x\to\infty} \int_0^{x/2} e^{\Lambda(x) - \Lambda(x-y) - \Lambda(y)} \lambda(y)dy &= \int_0^\infty \lim_{x\to\infty} 1_{[0,x/2]}(y) e^{\Lambda(x) - \Lambda(x-y) - \Lambda(y)} \lambda(y)dy \\
&= \int_0^\infty e^{-\Lambda(y)} \lambda(y)dy = \int_0^\infty g(y)dy = 1.
\end{aligned}$$

Using the inequality $\lambda(x-y) \leq \lambda(y)$ for $y < x/2$, we can again apply dominated convergence to the integral

$$\int_0^{x/2} e^{\Lambda(x) - \Lambda(x-y) - \Lambda(y)} \lambda(x - y)dy,$$

except now the limit is zero. Hence

$$\lim_{x\to\infty} \left(\frac{1 - G^{*2}(x)}{1 - G(x)} - 1 \right) = 1$$

proving that G is subexponential as desired.

■

Example 14.5. Consider the Pareto distribution with parameters α and β. Then

$$\overline{G}(x) = \frac{\beta^\alpha}{(\beta + x)^\alpha} = \frac{1}{(1 + x/\beta)^\alpha} \quad \text{for } x \geq 0.$$

The density is

$$g(x) = -\overline{G}'(x) = \frac{\alpha}{\beta} \frac{1}{(1 + x/\beta)^{\alpha+1}}$$

and the failure rate is

$$\lambda(x) = \frac{\alpha}{\beta} \frac{1}{(1 + x/\beta)}.$$

We see that λ is a decreasing function on $(0, \infty)$. Note also that

$$\lim_{x\to\infty} x \frac{\alpha}{\beta} \frac{1}{1 + x/\beta} = \alpha \quad \text{and} \quad \lim_{x\to 0} x \frac{\alpha}{\beta} \frac{1}{1 + x/\beta} = 0.$$

Thus $x\lambda(x)$ is a continuous bounded function and so $e^{x\lambda(x)} \leq C$ for some constant $C > 0$. This implies

$$\int_0^\infty e^{x\lambda(x)} g(x)dx \leq C \int_0^\infty g(x)dx = C < \infty$$

and using the above proposition, we can conclude that G is subexponential.

15 The method of ladder heights

Before we can study the ruin problem for the subexponential case, we need new techniques. We can no longer establish the Lundberg inequality or the Cramér-Lundberg estimate since the moment-generating function does not exist in a neighbourhood around zero for a subexponential distribution. In this section we will investigate the idea of *ladder heights* .

Consider again the Cramér-Lundberg process

$$C_t = u + ct - \sum_{i=1}^{N_t} Y_i$$

where as usual, $X_t = C_t - u$. The first time X_t goes below zero is called the *first ladder height time* and is denoted by T_1^-. Formally,

$$T_1^- = \inf\{t \geq 0 : X_t < 0\}.$$

Note that T_1^- does not need to be finite. If T_1^- is finite, it is of interest to study the size of the first negative value i.e. the size of the first "dip" below zero. The size of this absolute value L_1 is called the *first ladder height*. Formally,

$$L_1 = |X_{T_1^-}|.$$

We emphasise that L_1 is only defined conditional on the event $\{T_1^- < \infty\}$ (we will return to this issue). We can compute $P(T_1^- < \infty)$. Indeed, if we have a Cramér-Lundberg process with $u = 0$, then $X_t = C_t$ and thus $P(T_1^- < \infty) = \psi(0) = \lambda\mu_Y/c$. Note that this number is

in $(0,1)$ due to the NPC. We can now consider the second ladder height. The second ladder height time T_2^- is the first time X_t falls below $-L_1$ i.e.

$$T_2^- = \inf\{t \geq T_1^- : X_t < X_{T_1^-}\}$$

and the second ladder height is

$$L_2 = |X_{T_2^-}| - L_1$$

which is defined conditional on the event $\{T_2^- < \infty\}$. We can of course continue this construction. We summarise this discussion in the following definition.

Definition 15.1. For the Cramér-Lundberg process $\{C_t\}$ with $X_t = C_t - u$, the ladder height times are defined inductively by

$$T_0^- = 0 \quad \text{and} \quad T_i^- = \inf\{t \geq T_{i-1}^- : X_t < X_{T_{i-1}^-}\} \quad \text{for } i > 0$$

conditional on the event $\{T_{i-1}^- < \infty, \tau_i^- < \infty\}$. The τ_i^- are called the *inter-ladder height times* or *ladder epochs* and are defined by $\tau_1^- = T_1^-$ and $\tau_i^- = T_i^- - T_{i-1}^-$ for $i > 0$. The ladder heights are defined inductively by

$$L_0 = 0, \quad \text{and} \quad L_i = |X_{T_i^-} - X_{T_{i-1}^-}| = |X_{T_i^-}| - (L_1 + \cdots L_{i-1}) \quad \text{for } i > 0$$

conditional on the event $\{T_{i-1}^- < \infty, \tau_i^- < \infty\}$.

Proposition 15.2. *$\{\tau_i^-\}$ and $\{L_i\}$ are iid sequences.*

Proof. This follows from $\{X_t\}$ being a Lévy process. Indeed, the independent and stationary increments imply that at each ladder height time, we can think of the process as starting from scratch. ■

Remark 15.3. To get around the problem with the L_i being defined conditional on the event $\{T_{i-1}^- < \infty, \tau_i^- < \infty\}$ only, we implicitly extend L_i to the whole space by defining L_i to have the same distribution on $\{T_{i-1}^- = \infty\} \cup \{\tau_i^- = \infty\}$. Then L_i is independent of $\{\tau_i^- < \infty\}$.

The following result describes the distribution of the L_i.

Lemma 15.4. *The survival function $\overline{G}_0$ of L_1 is given by*

$$\overline{G}_0(x) = P(L_1 > x) = \frac{1}{\mu_Y} \int_x^\infty (1 - G(y))dy \quad \textit{for } x \geq 0.$$

We will provide a proof of this lemma later. A distribution function of this form is important enough to get its own name.

Definition 15.5. Let Y be a non-negative random variable with mean $\mu_Y < \infty$. The distribution given by the distribution function

$$\frac{1}{\mu_Y} \int_x^\infty (1 - G(y))dy \quad \text{for } x \geq 0$$

is called the *stationary excess distribution* or the *integrated tail distribution* of Y.

16 Ladder heights and the ruin probability for subexponential claims

We can now study the ruin probability using ladder heights. Each ladder height describes a new minimum of $\{X_t\}$. Let $K = \max\{i : T_i^- < \infty\}$ denote the number of observed ladder heights. Then $X_{T_K^-}$ is the smallest value of $\{X_t\}$ i.e.

$$\inf_{t\geq 0} X_t = -(L_1 + \cdots + L_K).$$

The probability of ruin may then be stated as

$$\psi(u) = P(\inf_{t\geq 0} X_t < -u) = P(L_1 + \cdots + L_K > u).$$

If we live in a heavy-tailed world with subexponential L_i we would expect the approximation $\psi(u) \approx \mathbf{E}[K]P(L_1 > u)$ to hold but we need to make this precise. The first step is to study the so-called *severity of ruin*. In the event that the company goes bankrupt a relevant question to ask is by how much we are ruined. This leads to the following definition.

Definition 16.1. Let $T = \inf\{t \geq 0 : X_t < -u\}$ denote the time of ruin. For $x \geq 0$, the quantity

$$\psi_x(u) = P(T < \infty, X_T < -(u + x))$$

is called the *severity of ruin*. x is called the *severity*.

We note that $\psi_x(0)$ describes the first ladder height. Indeed,

$$\psi_x(0) = P(T < \infty, X_T < -x) = P(T_1^- < \infty, L_1 > x).$$

The strategy now is to compute $\psi_x(u)$ generally and then specialize to the case $u = 0$. The first step is to find an integral-differential equation for $\psi_x(u)$ via a perturbation argument. This will be very similar to the argument that was given earlier in the course.

Consider a small interval $[0, h]$. We have three cases, namely

(i) $N_h = 0$ which occurs with probability $P(N_h = 0) = 1 - \lambda h + o(h)$,

(ii) $N_h = 1$ with probability $P(N_h = 1) = \lambda h + o(h)$ and

(iii) $N_h \geq 2$ which is negligible due to $P(N_h \geq 2) = o(h)$.

We consider each case.

(i) In this case the initial capital increases to $u + ch$. Hence $\psi_x(u)$ changes to $\psi_x(u + ch)$.

(ii) This is the most complicated case. Assume that we have one claim Y_1 of size $Y_1 = y$. We have

$$P(T < \infty, X_T < -(u + x) \mid N_h = 1, Y_1 = y) = \begin{cases} \psi_x(u - y) + o(1), & 0 < y \leq u \\ o(1), & u < y \leq u + x \\ 1 + o(1), & y > u + x \end{cases}.$$

The case $0 < y \leq u$ is the same as for the ordinary ruin probability. Since ruin does not occur, a new process starts at the lower initial capital $u - y$ and $o(1)$ takes care of the negligible premiums that the company receives. The case $u < y \leq u + x$ yields the probability $0 + o(1)$ because ruin has occured but the process has not surpassed the severity x. Formally, one can think of this case as intersecting $\{T < \infty, X_T < -(u+x)\}$ and $\{X_T \geq -(u + x)\}$ which is the empty set. The last case is self-explanatory.

(iii) This case is negligible.

Combining our considerations, we arrive at

$$\begin{aligned}\psi_x(u) &= (1 - \lambda h + o(h))\psi_x(u + ch) \\ &\quad + (\lambda h + o(h)) \left(\int_0^u \psi_x(u - y) + o(1) dG(y) + \int_{u+x}^{\infty} 1 dG(y) + o(1) \right) + o(h) \\ &= (1 - \lambda h + o(h))\psi_x(u + ch) + (\lambda h + o(h)) \left(\int_0^u \psi_x(u - y) + o(1) dG(y) + 1 - G(u + x) \right) + o(h).\end{aligned}$$

We rearrange this equation and obtain

$$c\frac{\psi_x(u + ch) - \psi_x(u)}{ch} = \lambda \left(\psi_x(u + ch) - \int_0^u \psi_x(u - y) + o(1) dG(y) - (1 - G(u + x)) \right) + o(1).$$

Letting $h \to 0^+$ gives

$$\frac{c}{\lambda}\psi_x'(u) = \psi_x(u) - \int_0^u \psi_x(u - y) dG(y) - (1 - G(u + x))$$

which is very reminiscent of the equation we got for the ordinary ruin probability. Just like we did then, we can integrate from 0 to u and obtain the integral equation

$$\frac{c}{\lambda}(\psi_x(u) - \psi_x(0)) = \int_0^u (1 - G(y))\psi_x(u - y) dy - \int_0^u (1 - G(y + x)) dy.$$

The detailed calculation is at the end of this week's material and is purely supplementary.

17 More on ladder heights

Recall that we derived the integral equation

$$\frac{c}{\lambda}(\psi_x(u) - \psi_x(0)) = \int_0^u (1 - G(y))\psi_x(u - y) dy - \int_0^u (1 - G(y + x)) dy.$$

for $\psi_x(u)$. The goal of this short section is to prove Lemma 15.4 which gave us the distribution of the first ladder height. We first need the following result.

Lemma 17.1. *$\psi_x(u) \to 0$ for $u \to \infty$.*

Proof. Note that $\psi_x(u) \leq \psi(u)$ for any $x > 0$. Define $Y = -\inf_{t \geq 0} X_t$. Then

$$\psi(u) = P\left(\inf_{t \geq 0} X_t < -u \right) = P(Y > u).$$

Noting that $Y < \infty$ a.s. since X_t has positive drift (because of NPC), we have $P(Y > u) \to 0$ for $u \to \infty$ and thus $\psi(u) \to 0$ for $u \to \infty$. This finishes the proof. ∎

Proof of Lemma 15.4. Letting $u \to \infty$ in the integral equation for $\psi_x(u)$ implies

$$-\frac{c}{\lambda}\psi_x(0) = \int_0^\infty (1 - G(x+y))dy = \int_x^\infty (1 - G(y))dy$$

since the first integral vanishes due to dominated convergence and the previous lemma. Now recall that

$$\psi_x(0) = P(T_1^- < \infty, L_1 > x)$$

so

$$P(T_1^- < \infty, L_1 > x) = \frac{\lambda}{c}\int_x^\infty (1 - G(y))dy.$$

Recall that we implicitly extend L_1 to the whole space such that $\{L_1 > x\}$ and $\{T_1^- < \infty\}$ are independent. The calculation can now be completed as follows

$$\begin{aligned} P(L_1 > x) &= \frac{P(L_1 > x, T_1^- < \infty)}{P(T_1^- < \infty)} = \frac{1}{P(T_1^- < \infty)}\frac{\lambda}{c}\int_x^\infty (1 - G(y))dy \\ &= \frac{1}{\frac{\lambda\mu_Y}{c}}\frac{\lambda}{c}\int_x^\infty (1 - G(y))dy = \frac{1}{\mu_Y}\int_x^\infty (1 - G(y))dy \end{aligned}$$

and the proof is complete.

■

18 Subexponential ruin: tail asymptotics

Recall that if $L_1, ..., L_K$ are the observed ladder heights, we can write the probability of ruin as

$$\psi(u) = P(L_1 + \cdots + L_K > u)$$

where $L_1 \sim G_0$ with

$$\overline{G}_0(x) = \frac{1}{\mu_Y}\int_x^\infty (1 - G(y))dy$$

the integrated tail distribution of the claims. In this section we study the case of subexponential ladder heights i.e. $G_0 \in \mathcal{S}$. From the previous week we have the following facts:

(i) $P(L_1 + \cdots + L_n > u) \sim nP(L_1 > u)$ as $u \to \infty$ and

(ii) for any $\varepsilon > 0$ there exists a constant D such that $P(L_1 + \cdots + L_n > u)/P(L_1 > u) \leq D(1+\varepsilon)^n$ for any x, n.

We will use these facts to prove the following theorem.

Theorem 18.1 (**Ruin asymptotics for subexponential ladder heights**). *Assume the Cramér-Lundberg model where the ladder heights are subexponential. Then*

$$\psi(u) \sim \frac{\lambda\mu_Y}{c - \lambda\mu_Y}\overline{G}_0(u) = \frac{\lambda\mu_Y}{c - \lambda\mu_Y}\frac{1}{\mu_Y}\int_u^\infty (1 - G(y))dy.$$

Proof. We start by noting that K is independent of $\{L_i\}$. We get

$$\begin{aligned}\lim_{u\to\infty}\frac{\psi(u)}{\overline{G}_0(u)} &= \lim_{u\to\infty}\frac{P(L_1+\cdots+L_K>u)}{\overline{G}_0(u)}\\ &= \lim_{u\to\infty}\sum_{n=0}^{\infty}P(K=n)\frac{P(L_1+\cdots L_n>u\mid K=n)}{\overline{G}_0(u)}\\ &= \lim_{u\to\infty}\sum_{n=0}^{\infty}P(K=n)\frac{P(L_1+\cdots L_n>u)}{\overline{G}_0(u)}.\end{aligned}$$

We wish to interchange the sum and the limit. In order to do so, we apply the dominated convergence theorem. Hence we need to show that the sum is bounded uniformly in u. Let $\varepsilon>0$, then by fact (ii) above, there is some constant D such that

$$\sum_{n=0}^{\infty}P(K=n)\frac{P(L_1+\cdots+L_n>u)}{\overline{G}_0(u)}\le\sum_{n=0}^{\infty}P(K=n)D(1+\varepsilon)^n.$$

To make this sum finite, we have to choose a proper ε. Recall from the exercises that K follows a geometric distribution with succes parameter $p=\lambda\mu_Y/c<1$. Hence $P(K=n)=p^n(1-p)$ and the above sum equals

$$(1-p)D\sum_{n=0}^{\infty}(p(1+\varepsilon))^n.$$

We have

$$p(1+\varepsilon)<1\quad\Leftrightarrow\quad\varepsilon<\frac{1}{p}-1.$$

Choosing such an ε, we see that the conditions of the dominated convergence theorem are satisfied. We can now continue the first calculation using fact (i)

$$\begin{aligned}\lim_{u\to\infty}\frac{\psi(u)}{\overline{G}_0(u)} &= \sum_{n=0}^{\infty}P(K=n)\lim_{u\to\infty}\frac{P(L_1+\cdots+L_K>u)}{\overline{G}_0(u)}\\ &= \sum_{n=0}^{\infty}P(K=n)n=\mathbf{E}[K]=\frac{\lambda\mu_Y}{c-\lambda\mu_Y}.\end{aligned}$$

The proof is now complete. ■

Remark 18.2. It is important to note that the theorem does not assume the claim sizes are subexponential. It assumes that the integrated tail distribution of the claims are subexponential.

We will see an application of this theorem in the exercises. Note the difference compared to the asymptotics for the ruin probability in the classical case. Here the ruin probability decays a lot slower. Moreover, the theorem tells us that ruin in this case is highly unpredictable. Ruin occurs because one large claim bankrupts the company where in the classical case, ruin is gradual.

19 Introduction to the renewal risk model

So far we have studied the Cramér-Lundberg model

$$C_t = u + ct - \sum_{i=1}^{N_t} Y_i, \quad \{Y_i\} \text{ iid}, Y_i \sim G, \quad \{N_t\} \text{ a Poisson process with intensity } \lambda.$$

This model can be generalized in many ways. We will extend the model so that $\{N_t\}$ is no longer necessarily a Poisson process but any renewal process. All other assumptions are kept including the independence of the claims and arrival times. This extended model is called the *renewal risk model* or the *Sparre Andersen model*.

A very essential consequence of $\{N_t\}$ being a Poisson process is that $X_t = C_t - u$ is a Lévy process. This no longer holds for general renewal processes. Hence we need a new way of studying $\{X_t\}$. To study this process we will use the *random walk representation*. We observe that ruin can only occur at an arrival time T_n. Indeed, between arrival times, the process increases since the company collects premiums. Hence we can simplify the study of $\{X_t\}$ into a problem concerning a discrete time process.

Definition 19.1. If $\{\tau_i\}$ denotes the iid sequence of interarrival times for the renewal process $\{N_t\}$ (so that $T_n = \tau_1 + \cdots + \tau_n$), define the random variables

$$Z_i = X_{T_i} = c\tau_i - Y_i.$$

We call the discrete time stochastic process $\{S_n\}$ given by

$$S_n = X_{T_n} = \sum_{i=1}^{n} Z_i$$

the *random walk representation* of $\{X_t\}$.

Note that Z_i is the increase in $\{X_t\}$ between claim $i-1$ and claim i. We also remark that $\{Z_i\}$ is an iid sequence. We can now write the probability of ruin as

$$\psi(u) = P(X_{T_i} < -u \text{ for some } i) = P(S_n < -u \text{ for some } n).$$

Supplementary: Deriving the integral equation for $\psi_x(u)$

We integrate the equation

$$\frac{c}{\lambda}\psi_x'(u) = \psi_x(u) - \int_0^u \psi_x(u-y)dG(y) - (1 - G(u+x))$$

from 0 to u (after replacing u with v) and obtain

$$\frac{c}{\lambda}(\psi_x(u) - \psi_x(0)) = \int_0^u \psi_x(v)dv - \int_0^u \int_0^v \psi_x(v-y)dG(y)dv - \int_0^u 1 - G(v+x)dv.$$

We give the first integral on the right hand side a name:

$$I(u) = \int_0^u \psi_x(v)dv,$$

then we note that

$$I(u-y) = \int_0^{u-y} \psi_x(v)dv = \int_y^u \psi_x(w-y)dw.$$

Now we consider the second integral. Interchanging the order of integration and applying integration by parts yields

$$\begin{aligned}
\int_0^u \int_0^v \psi_x(v-y)dG(y)dv &= \int_0^u \int_y^u \psi_x(v-y)dvdG(y) = \int_0^u I(u-y)dG(y) \\
&= -\int_0^u I(u-y)d(1-G)(y) \\
&= [-I(u-y)(1-G(y))]_0^u - \int_0^u (1-G(y))\psi_x(u-y)dy \\
&= I(u) - \int_0^u (1-G(y))\psi_x(u-y)dy.
\end{aligned}$$

Plugging back into our expression from before gives

$$\begin{aligned}
\frac{c}{\lambda}(\psi_x(u) - \psi_x(0)) &= I(u) - I(u) + \int_0^u (1-G(y))\psi_x(u-y)dy - \int_0^u 1 - G(v+x)dv \\
&= \int_0^u (1-G(y))\psi_x(u-y)dy - \int_0^u 1 - G(y+x)dy
\end{aligned}$$

which is the desired equation.

Week 5 - Cramér-Lundberg theory for the renewal risk model

20 The Lundberg inequality in the renewal risk model

To prove the Lundberg inequality for the renewal risk model, we will extend the ideas from the theory of martingales to this setting. Recall that the discrete sequence $\{Z_i\}$ given by $Z_i = c\tau_i - Y_i$ is iid and that $S_n = Z_1 + \cdots + Z_n$. As we have already seen, the processes $\{\tilde{M}_n\}$ and $\{M_n\}$ given by

$$\tilde{M}_n = S_n - n\mu, \quad \mu = \mathbf{E}[Z_1] < \infty$$

and

$$M_n = e^{\alpha S_n - n\Lambda(\alpha)}, \quad \Lambda(\alpha) = \log \mathbf{E}[e^{\alpha Z_1}] < \infty$$

are martingales. Now consider the random walk $\{S_n\}$. From the Strong Law of Large Numbers it follows that

$$\frac{1}{n} S_n \to \mu \ a.s.$$

and so if $\mu < 0$, $S_n \to -\infty$ a.s. Hence if $\mu < 0$ then ruin occurs with probability one. If $\mu > 0$ it follows from the previous discussion on ladder heights that $\psi(u) < 1$. Indeed, recall that the probability that the first ladder height time is finite is strictly less than one in this case (we shall see later in the lecture that this result still holds in this more general setting). Once again $\mu = 0$ is a difficult edge case. A theorem on random walks tells us that S_n will cross zero infinitely often in this case and thus $\psi(u) = 1$ in this case also. We conclude that we need $\mu > 0$ to avoid ruin with probability one. Letting $\mu_\tau = \mathbf{E}[\tau_1]$ and $\mu_Y = \mathbf{E}[Y_1]$, we have

$$\mu = \mathbf{E}[Z_1] = c\mu_\tau - \mu_Y > 0 \quad \Leftrightarrow \quad c > \frac{\mu_Y}{\mu_\tau}.$$

This leads to the following definition.

Definition 20.1 (**Net profit condition in the renewal risk model**)**.** In the setup of the renewal risk model, let $\mu_\tau = \mathbf{E}[\tau_1]$ and $\mu_Y = \mathbf{E}[Y_1]$. The assumption

$$c > \frac{\mu_Y}{\mu_\tau}$$

is called the *net profit condition*. .

Example 20.2. If the renewal process $\{N_t\}$ is a Poisson process with intensity $\lambda > 0$, we have $\mu_\tau = 1/\lambda$ since the interarrival times are Exp(λ) distributed. Hence the NPC reads

$$c > \lambda \mu_Y$$

which coincides with the previous definition.

We are now ready to consider the Lundberg inequality. To do so, we need the adjustment coefficient in the renewal risk model. Since we always assume the NPC, $\mathbf{E}[Z_1] > 0$ so if $\Lambda(\alpha)$ exists, $\Lambda'(0) > 0$ and Λ is convex. We now compute $\Lambda(\alpha)$. Letting $\kappa(\alpha) = \mathbf{E}[e^{\alpha Z_1}]$, we have

$$\kappa(\alpha) = \mathbf{E}[e^{\alpha(c\tau_1 - Y_1)}] = \mathbf{E}[e^{\alpha c\tau_1}]\mathbf{E}[e^{-\alpha Y_1}] = \kappa_\tau(\alpha c)\kappa_Y(-\alpha).$$

We have $\Lambda(-R) = 0$ if and only if $\kappa(-R) = 1$ so

$$1 = \kappa_\tau(-cR)\kappa_Y(R)$$

is the equation for the adjustment coefficient in the renewal risk model. We let this be a definition.

Definition 20.3. In the setup of the renewal risk model, the *adjustment coefficient* (if it exists) is the number $R > 0$ which satisfies the equation

$$\kappa_\tau(-cR)\kappa_Y(R) = 1.$$

Example 20.4. Let the renewal process $\{N_t\}$ be a Poisson process with intensity $\lambda > 0$. Then the interarrival times are $\text{Exp}(\lambda)$ distributed and so

$$\kappa_\tau(\alpha) = \frac{\lambda}{\lambda - \alpha}.$$

Plugging this expression into the equation for the adjustment coefficient yields

$$1 = \frac{\lambda}{\lambda + cR}\kappa_Y(R) \quad \Leftrightarrow \quad \lambda + cR = \lambda\kappa_Y(R) \quad \Leftrightarrow \quad cR - (\kappa_Y(R) - 1)\lambda = 0$$

which is the same equation for the adjustment coefficient as seen earlier.

We have now presented all the notions and tools needed to derive the Lundberg inequality.

Theorem 20.5 (**The Lundberg inequality in the renewal risk model**). *In the setup of the renewal risk model, assume that the adjustment coefficient $R > 0$ exists. Then*

$$\psi(u) \leq e^{-Ru}.$$

Proof. The proof very much resembles the one given earlier. Letting $\alpha = -R$ in the exponential martingale, we get that the process

$$M_t = e^{-RS_n}$$

is a martingale. Now let $T = \inf\{n \in \mathbb{N} : S_n < -u\}$ denote the time of ruin. Let $T_k = \min\{T, k\}$. Then T_k are bounded stopping times. Hence we can apply the optional sampling theorem and obtain

$$\mathbf{E}[M_{T_k}] = M_0 = 1.$$

We split the above mean into two cases:

$$\mathbf{E}[M_{T_k}] = \mathbf{E}[M_T 1_{\{T \leq k\}}] + \mathbf{E}[M_k 1_{\{T > k\}}].$$

The monotone convergence theorem gives

$$\lim_{k\to\infty} \mathbf{E}[M_T 1_{\{T\le k\}}] = \mathbf{E}[M_T \lim_{k\to\infty} 1_{\{T\le k\}}] = \mathbf{E}[M_T 1_{\{T<\infty\}}].$$

We know that $M_k \ge 0$ so taking limits on both sides of $\mathbf{E}[M_{T_k}] = 1$ gives

$$\begin{aligned} 1 &= \mathbf{E}[M_T 1_{T<\infty}] + \mathbf{E}[M_k 1_{\{T>k\}}] \ge \mathbf{E}[M_T 1_{T<\infty}] = \mathbf{E}[e^{-RS_T} 1_{\{T<\infty\}}] \\ &\ge \mathbf{E}[e^{-R(-u)} 1_{\{T<\infty\}}] = e^{Ru} P(T<\infty) = e^{Ru}\psi(u) \end{aligned}$$

where in the second line we used that at the time T, the process S_n is less than $-u$. Rearranging yields

$$\psi(u) \le e^{-Ru}$$

as desired. ■

21 The Cramér-Lundberg estimate in the renewal risk model

We now wish to establish the Cramér-Lundberg estimate in the renewal risk model. The perturbation argument we did for the classical Cramér-Lundberg model no longer works since it relied on the properties of a Poisson process. Instead we approach via ladder heights of $\{S_n\}$.

Definition 21.1. Set $T_0^- = 0$ and let

$$T_1^- = \inf\{n \ge 0 : S_n < 0\} \quad \text{and} \quad T_i^- = \inf\{n \ge T_{i-1}^- : S_n < S_{T_{i-1}^-}\} \quad \text{for } i > 1.$$

These are called the *negative ladder height times* . We define the *negative ladder heights* to be

$$L_i = |S_{T_i^-} - S_{T_{i-1}^-}|$$

and like before, these variables are defined conditional on the event that $T_{i-1}^- < \infty$ and $\tau_i^- < \infty$ where as before, $\tau_i^- = T_i^- - T_{i-1}^-$.

Like before we need to define the ladder heights conditional on $\{T_{i-1}^- < \infty, \tau_i^- < \infty\}$. Indeed, as discussed earlier, S_n drifts to infinity due to the NPC, so there is a positive probability that we never observe the first negative ladder height. Note also that since $\{N_t\}$ is a renewal process and S_n is a sum of iid variables, at the time T_1^-, the process essentially restarts at $-L_1$, i.e. the behaviour of the process after L_1 is independent of the process up to time T_1^-. The same argument applies to later negative ladder height times.

The method needed to establish the Cramér-Lundberg estimate is essentially the same as in the classical case. Recall that we constructed a renewal equation that $\psi(u)$ satisfies by making an exponential shift. Set

$$H(x) := P(L_1 \le x, T_1^- < \infty).$$

$H(x)$ is increasing and right-continuous but

$$H(\infty) := \lim_{x\to\infty} H(x) = P(T_1^- < \infty) < 1$$

due to the NPC. Hence $H(x)$ is not a proper distribution function (H is an example of a *defective* distribution function).

We can now make a perturbation argument. Observe that $\{S_n - S_{T_1^-}\}$ for $n > T_1^-$ is independent of $\{S_1, ..., S_{T_1^-}\}$. The idea is to condition on (T_1^-, L_1). We have three cases:

(i) $T_1^- = \infty$ i.e. we never observe any ladder heights. Thus $S_n > 0$ for all n and ruin never occurs. Also, $P(T_1^- = \infty) = 1 - P(T_1^- < \infty) = 1 - H(\infty)$.

(ii) $T_1^- < \infty$ and $L_1 = x \leq u$. Ruin does not occur and the ruin probability corresponds to the ruin probability with initial capital $u - x$. Also (informally),

$$P(L_1 \in [x, x + dx), T_1^- < \infty) = dH(x).$$

(iii) $T_1^- < \infty$ and $L_1 = x > u$. Ruin occurs and $P(L_1 > u, T_1^- < \infty) = H(\infty) - H(u)$.

We now collect these observations and condition to obtain

$$\begin{aligned}\psi(u) &= 0 \cdot (1 - H(\infty)) + \int_0^u \psi(u - x) dH(x) + 1 \cdot (H(\infty) - H(u)) \\ &= \int_0^u \psi(u - x) dH(x) + H(\infty) - H(u).\end{aligned}$$

Note that this is already an integral equation. In the classical case we needed to do some integral tricks to go from an integral differential equation to an integral equation. The equation looks like a renewal equation but since H is not a proper distribution function, we need a small modification. Multiply both sides by $e^{\alpha u}$ to obtain

$$e^{\alpha u}\psi(u) = \int_0^u e^{\alpha(u-x)}\psi(u - x)e^{\alpha x} dH(x) + e^{\alpha u}(H(\infty) - H(u)).$$

We want $e^{\alpha x} dH(x)$ to be a distribution function i.e. it should integrate to one. Such a distribution is called *exponentially shifted*. Based on earlier calculations, we expect $\alpha = R$ to be the solution to this problem. This turns out to be true but the calculations are more subtle since the known function is a lot less explicit than in the classical case. Let H_R denote the function with $dH_R = e^{Rx} dH(x)$.

Lemma 21.2. *H_R is a distribution function.*

Proof. Trivially, $H_R(x) \to 0$ for $x \to -\infty$. H_R is right-continuous and non-decreasing since H is. It only remains to show that $H_R(x) \to 1$ for $x \to \infty$ i.e.

$$\int_0^\infty dH_R(x) = 1.$$

We compute

$$\begin{aligned}\int_0^\infty dH_R(x) &= \int_0^\infty e^{Rx} dH(x) = \int_0^\infty e^{Rx} P(T_1^- < \infty) dP(L_1 \leq x \mid T_1^- < \infty) \\ &= P(T_1^- < \infty) \int_0^\infty e^{Rx} dP(L_1 \leq x \mid T_1^- < \infty) = P(T_1^- < \infty)\mathbf{E}[e^{RL_1} \mid T_1^- < \infty] \\ &= \mathbf{E}\left[e^{-RS_{T_1^-}} 1_{\{T_1^- < \infty\}}\right].\end{aligned}$$

To continue the computation, we apply martingale methods. Recall that the process

$$M_n = e^{\alpha S_n - n\Lambda(\alpha)}$$

is a martingale for any α. Choosing $\alpha = -R$, we obtain the martingale

$$M_n = e^{-RS_n}.$$

Define the sequence of bounded stopping times $J_k = T_1^- \wedge k$. We apply the optional sampling theorem and get

$$1 = M_0 = \mathbf{E}[M_{J_k}] = \mathbf{E}[M_{T_1^-} 1_{\{T_1^- \leq k\}}] + \mathbf{E}[M_k 1_{\{T_1^- > k\}}].$$

We consider each term on the right hand side. The integrand of the first term is increasing so by the monotone convergence theorem,

$$\lim_{k\to\infty} \mathbf{E}[M_{T_1^-} 1_{\{T_1^- \leq k\}}] = \mathbf{E}[M_{T_1^-} \lim_{k\to\infty} 1_{\{T_1^- \leq k\}}] = \mathbf{E}[M_{T_1^-} 1_{\{T_1^- < \infty\}}] = \mathbf{E}\left[e^{-RS_{T_1^-}} 1_{\{T_1^- < \infty\}}\right].$$

For the other term, note that if $T_1^- > k$, $S_k \geq 0$ so the second term is bounded by 1. Using the dominated convergence theorem,

$$\lim_{k\to\infty} \mathbf{E}\left[M_k 1_{\{T_1^- > k\}}\right] = \mathbf{E}\left[\lim_{k\to\infty} M_k 1_{\{T_1^- > k\}}\right] = \mathbf{E}[0] = 0.$$

We conclude that

$$\mathbf{E}\left[e^{-RS_{T_1^-}} 1_{\{T_1^- < \infty\}}\right] = 1$$

and this completes the proof.

■

We have thus established a proper renewal equation

$$Z_R(u) = z_R(u) + \int_0^u Z_R(u-x) dH_R(x)$$

with $Z_R(u) = e^{\alpha u}\psi(u)$ and $z_R(u) = e^{Ru}(H(\infty) - H(u))$. To establish the Cramér-Lundberg estimate, we need the following lemma.

Lemma 21.3. *z_R is directly Riemann integrable (given the proper technical conditions).*

We now have all the tools we need to prove the Cramér-Lundberg estimate.

Theorem 21.4 (**The Cramér-Lundberg estimate for the renewal risk model**). *In the context of the renewal risk model, assume $\Lambda(-R) = 0$ for some $R > 0$ and $\Lambda'(-R) < \infty$ and that H is non-arithmetic. Then*

$$\psi(u) \sim Ce^{-Ru} \quad \textit{for} \quad u \to \infty$$

where

$$C = \frac{1}{\mu_R} \int_0^\infty z_R(x) dx, \quad \mu_R = \int_0^\infty x dH_R(x).$$

Proof. One can show that $\mu_R < \infty$ if $\Lambda'(-R) < \infty$. As z_R is directly Riemann integrable, the second form of the renewal theorem (Theorem 9.8) yields

$$Z_R(u) \to C = \frac{1}{\mu_R} \int_0^\infty z_R(x)dx \quad \text{for} \quad u \to \infty, \quad \mu_R = \int_0^\infty x dH_R(x),$$

and this implies

$$\frac{\psi(u)}{Ce^{-Ru}} \to 1 \quad \text{for} \quad u \to \infty$$

which establishes the theorem. ■

In the exercises, we gave a very explicit expression for the constant C in the classical setting. In this more general setting we cannot provide an explicit expression, but C can be simplified. The following proposition provides one such simpler expression.

Proposition 21.5. *With the assumptions of the Cramér-Lundberg estimate above, we have*

$$C = \frac{1 - P(T_1^- < \infty)}{\mu_R R}.$$

Proof. The proof is a straightforward computation:

$$\begin{aligned} C &= \frac{1}{\mu_R} \int_0^\infty z_R(x)dx = \frac{1}{\mu_R} \int_0^\infty e^{Rx}(H(\infty) - H(x))dx = \frac{1}{\mu_R} \int_0^\infty e^{Rx} \int_x^\infty dH(y)dx \\ &= \frac{1}{\mu_R} \int_0^\infty \int_0^y e^{Rx} dx dH(y) = \frac{1}{\mu_R} \int_0^\infty \left[\frac{1}{R} e^{Rx} \right]_0^y dH(y) = \frac{1}{\mu_R R} \int_0^\infty (e^{Ry} - 1) dH(y) \\ &= \frac{1}{\mu_R R} \left(\int_0^\infty e^{Ry} dH(y) - \int_0^\infty dH(y) \right) = \frac{1}{\mu_R R} (1 - (H(\infty) - H(0))) \\ &= \frac{1 - H(\infty)}{\mu_R R} = \frac{1 - P(T_1^- < \infty)}{\mu_R R}. \end{aligned}$$

■

22 Simulating the Cramér-Lundberg process

In this section we discuss some aspects of estimating the probability of ruin via simulation. We wish to simulate

$$C_t = u + ct - \sum_{i=1}^{N_t} Y_i$$

where $\{N_t\}$ is a renewal process. It is a difficult task to simulate a continuous time process in general, but a Cramér-Lundberg process can be simulated simply by generating $\{(\tau_i, Y_i) : i = 1, 2, ...\}$ and this is easily done in many computer packages such as R. From the random walk representation we have $\psi(u) = P(S_n < u \text{ for some } n)$. A natural way to estimate $\psi(u)$ is thus to generate a large number of processes $\{S_n\}$ and compute the fraction of times that ruin happens. There are however two issues with this approach:

(i) Since $\psi(u) < 1$ there is a positive probability that a simulation of a process $\{S_n\}$ never terminates.

(ii) Since $\psi(u)$ is usually small, the error is often larger than the estimate of $\psi(u)$ itself.

Let us consider these issues in more detail. Let $T_u = \inf\{n \geq 1 : S_n < -u\}$ denote the time of ruin. We want to estimate the ruin probability

$$p_u := P(T_u < \infty) = \mathbf{E}[W]$$

for $W := 1_{\{T_u<\infty\}}$. We can estimate p_u by simulating an iid sequence $\{W_i : i = 1, ..., N\}$ with $W_i \sim W$ and compute the empirical mean $\widehat{W}_N$. By the Strong Law of Large Numbers,

$$p_u \approx \widehat{W}_N := \frac{1}{N}\sum_{i=1}^{N} W_i.$$

Note that $\widehat{W}_N$ is the proportion of simulations of $\{S_n\}$ where ruin occurs. We are interested in providing an error estimate for $\widehat{W}_N$. By the Central Limit Theorem, asymptotically we have

$$\widehat{W}_N \sim \mathcal{N}\left(p_u, \frac{\sigma_u^2}{N}\right), \quad \sigma_u^2 = \mathrm{Var}(W)$$

so

$$\widehat{W}_N - p_u \approx \frac{\sigma_u}{\sqrt{N}} Z \quad \text{where} \quad Z \sim \mathcal{N}(0,1).$$

We can now compute (asymptotic) confidence intervals. Let $z_{\alpha/2}$ denote the $\alpha/2$-quantile for the $\mathcal{N}(0,1)$ distribution i.e. $z_{\alpha/2}$ is the real number such that $P(Z > z_{\alpha/2}) = \alpha/2$. Then

$$\left(\widehat{W}_N - \frac{\sigma_u}{\sqrt{N}} z_{\alpha/2}, \widehat{W}_N + \frac{\sigma_u}{\sqrt{N}} z_{\alpha/2}\right)$$

is an asymptotic confidence interval for p_u with a coverage of $1 - \alpha$. We now consider the variance σ_u^2 of W. We have

$$\sigma_u^2 = \mathrm{Var}(W) = \mathbf{E}[1^2_{\{T_u<\infty\}}] - \mathbf{E}[1_{\{T_u<\infty\}}]^2 = p_u - p_u^2 = p_u(1 - p_u).$$

We can now define the *relative error RE* to be

$$RE := \frac{z_{\alpha/2}}{\sqrt{N}} \frac{\sigma_u}{p_u}.$$

We can now describe issue (ii) more formally. Indeed,

$$\frac{\sigma_u}{p_u} = \frac{\sqrt{p_u - p_u^2}}{p_u} = \frac{1}{\sqrt{p_u}}\sqrt{1 - p_u}$$

and $p_u \to 0$ for $u \to \infty$ so $\sigma_u/p_u \to \infty$ for $u \to \infty$. This means that the error of the estimate will get larger and larger as u gets larger. The error will thus dominate the estimate, effectively rendering the estimate useless. Thus, when estimating the ruin probability, we want the relative error to be bounded. We end this lecture with an example of a method which has this property.

Estimating the ruin probability

It turns out that replacing the distribution of Z_i with an exponentially shifted version solves both of the above discussed issues. We here present the general ideas and leave the proofs as exercises. Suppose Z_i has distribution function F and consider the exponential shift

$$dF_\alpha(x) = \frac{e^{\alpha x}}{\kappa(\alpha)} dF(x)$$

with $\kappa(\alpha) = \mathbf{E}[e^{\alpha Z_1}]$. This is a distribution function (exercise) and choosing $\alpha = -R$ turns out to be a good choice. If $\mathbf{E}_{-R}$ denotes expectation under the shifted measure with $\alpha = -R$, one can show that $\mathbf{E}_{-R}[Z_1] < 0$ so that $\{S_n\}$ goes below $-u$ with probability one under this measure. Moreover, one can show that

$$\psi(u) = \mathbf{E}_{-R}[e^{RS_{T_u}} 1_{\{T_u < \infty\}}]$$

where $T_u = \inf\{n \geq 1 : S_n < -u\}$ denotes the time of ruin. To summarise: The method works by first determining the shifted distribution F_{-R}. Second, simulate many $\{S_n\}$ with $S_n = Z_1 + \cdots + Z_n$, $Z_i \sim F_{-R}$ (not F!) and compute the empirical mean of $e^{RS_{T_u}}$. This yields an unbiased estimator of the ruin probability. Furthermore, it can be shown that the relative error is bounded for this method. In the exercises, we will apply this method to the case of exponential claims and compare to the theoretical expression for $\psi(u)$.

Week 6 - Ruin with stochastic investments and stochastic fixed point equations

See Jeffrey's note.

Week 7 - Claims reserving

23 Claims reserving

The material for this week is based on the notes by Jostein Paulsen, [5].

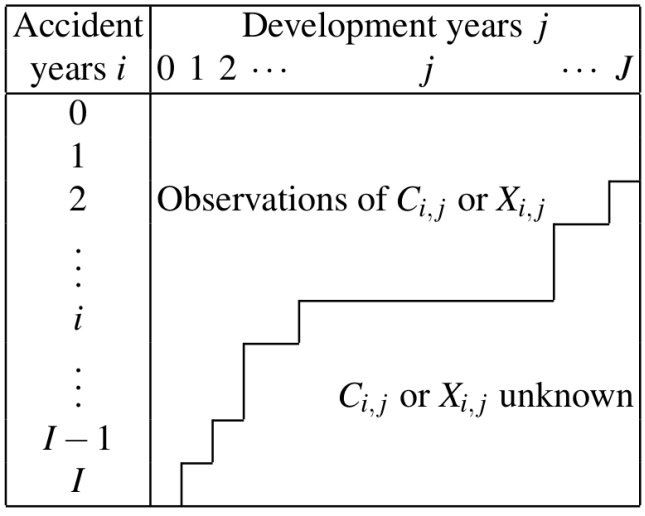
Up to now, all claims have been paid out immediately. In this final week, we will take a different (and more realistic) approach. In reality, claims are often reported and then settled at a later date. In some cases it can take years before a claim is paid to the policy-holder. Claims reserving can handle these cases. The goal is to model future payments for existing claims. We assume that over the time interval $[0, J]$, all claims will be settled. We take a simple approach and assume that time is discrete and counted in whole years. Let us fix the following notation.

Definition 23.1. We let X_{ij} denote the claims losses from year i paid j years after year i. We let

$$C_{ij} = \sum_{i=0}^{j} X_{ij}$$

denote the accumulated losses in year i and paid in years $\{0, 1, ..., j\}$.

Note also that we, in contrast to earlier, only work with sums of claims. Previously we worked with one claim at a time, while we now consider the total sum of claims in a given year. It is useful to present the data in a *runoff triangle* (taken from [5]):



As the figure indicates, the years where the accident takes place (indexed by i) are called *accident years* while the years where the claims are settled are called the *development years* or *runoff years* . We consider I total accident years, i.e. I is the current year. The longer we go back in time, the more information we have available. In year I, we only know the settlements for this year corresponding to X_{I0}. If we go one year back in time, we know both $X_{I-1,0}$ and $X_{I-1,1}$ and so on. Hence we have an upper triangle of information which is known and a lower triangle of information which is unknown. The goal is to predict the values in this lower triangle. Let us turn these observations into mathematics.

Definition 23.2. Define

$$\mathscr{D} = \sigma(\{X_{ij} : i + j \leq J, j \leq J\}), \quad \mathscr{E}_k = \sigma(\{X_{ij} : i \leq I, j \leq k\}).$$

We set $\mathscr{D} \vee \mathscr{E}_k = \sigma(\mathscr{D} \cup \mathscr{E}_k)$. We also let

$$C_T = \sum_{i=0}^{I} \sum_{j=0}^{J} X_{ij}$$

denote the total claims in the runoff triangle.

One should think of $\mathscr{D}$ as all the information available at time I i.e. the present time. This follows because all the information currently available is the information generated by the observations $X_{i0}, ...X_{i,I-i}$ i.e. the upper half of the triangle. One should think of $\mathscr{E}_k$ as all information generated by the observations up to runoff year k. Note that $\mathscr{D}$ and $\mathscr{E}_k$ contain different information in general. $\mathscr{D} \vee \mathscr{E}_k$ is the combined information in $\mathscr{D}$ and $\mathscr{E}_k$.

24 The chain ladder method

In order to make predictions, we need some assumptions on our data. The following two assumptions are standard.

(CL1): C_{ij} and $C_{i'j'}$ are independent for $i \neq i'$ i.e. observations from different accident years are independent.

(CL2): There exist factors (constants) $f_0, ..., f_{J-1}$ called *development factors* such that

$$\mathbf{E}[C_{i,j+1} \mid \mathscr{E}_j] = f_j C_{ij} \quad \text{for } j = 0, ..., J-1, i = 0, ..., I.$$

Under these two assumptions, it is possible to compute the expected value of future accumulated claims given the information available. Intuitively, we should start with $C_{i,I-i}$ and multiply with all the development factors $f_{I-i}, ..., f_{J-1}$. The following lemma confirms this intuition.

Lemma 24.1. *Under assumptions (CL1) and (CL2), the expected accumulated claims in year i is given by*

$$\mathbf{E}[C_{iJ} \mid \mathscr{D}] = C_{i,I-i} \prod_{j=I-i}^{J-1} f_j.$$

Proof. In the i'th accident year, $C_{i,I-i}$ is known and so

$$\begin{aligned}\mathbf{E}[C_{i,I-i+1} \mid \mathscr{D}] &= \mathbf{E}[\mathbf{E}[C_{i,I-i+1} \mid \mathscr{D} \vee \mathscr{E}_k] \mid \mathscr{D}] = \mathbf{E}[\mathbf{E}[C_{i,I-i+1} \mid \mathscr{E}_k] \mid \mathscr{D}] \\ &= \mathbf{E}[f_{I-i}C_{i,I-i} \mid \mathscr{D}] = f_{I-i}C_{i,I-i}.\end{aligned}$$

The first equality follows from the tower property. The second equality is a consequence of (CL1) since we can remove all information from observations not in year i. The third equality follows from (CL2). More generally, using the same calculations,

$$\mathbf{E}[C_{ij} \mid \mathscr{D}] = f_{j-1}\mathbf{E}[C_{i,j-1} \mid \mathscr{D}].$$

Iterating all the way from J to $I - i$, we obtain the desired formula. ■

The lemma leads to the following definition.

Definition 24.2. We define the *chain ladder estimate* to be

$$C_{iJ}^{\mathrm{CL}} = C_{i,I-i} \prod_{j=I-i}^{J-1} f_j.$$

The chain ladder estimate is exactly the expected losses for accident year i. To say something about the variance, we need another assumption.

(CL3): There exist constants $\sigma_0^2, ..., \sigma_{J-1}^2$ such that

$$\mathrm{Var}(C_{i,j+1} \mid \mathscr{E}_j) = \sigma_j^2 C_{ij}.$$

In order to do computations with the variance, the following lemma will be useful.

Lemma 24.3 (**Conditional variance formula**). *Let Y be a random variable and $\mathscr{F} \subseteq \mathscr{G}$ σ-algebras. Then*

$$\mathrm{Var}(Y \mid \mathscr{F}) = \mathbf{E}[\mathrm{Var}(Y \mid \mathscr{G}) \mid \mathscr{F}] + \mathrm{Var}(\mathbf{E}[Y \mid \mathscr{G}] \mid \mathscr{F}).$$

Proof. We have

$$\mathrm{Var}(Y \mid \mathscr{G}) = \mathbf{E}[Y^2 \mid \mathscr{G}] - \mathbf{E}[Y \mid \mathscr{G}]^2$$

so by the tower property,

$$\mathbf{E}[\mathrm{Var}(Y \mid \mathscr{G}) \mid \mathscr{F}] = \mathbf{E}[Y^2 \mid \mathscr{F}] - \mathbf{E}[\mathbf{E}[Y \mid \mathscr{G}]^2 \mid \mathscr{F}].$$

Note that

$$\mathbf{E}[\mathbf{E}[Y \mid \mathscr{G}] \mid \mathscr{F}] = \mathbf{E}[Y \mid \mathscr{F}],$$

and we may conclude

$$\begin{aligned}\mathrm{Var}(Y \mid \mathscr{G}) &= \mathbf{E}[Y^2 \mid \mathscr{G}] - \mathbf{E}[Y \mid \mathscr{F}]^2 + (\mathbf{E}[\mathbf{E}[Y \mid \mathscr{G}] \mid \mathscr{F}])^2 - \mathbf{E}[Y \mid \mathscr{G}]^2 \\ &= \mathrm{Var}(Y \mid \mathscr{F}) - \mathrm{Var}(\mathbf{E}[Y \mid \mathscr{G}] \mid \mathscr{F}).\end{aligned}$$

■

Remark 24.4. Note the similarity to the following well-known result from probability theory

$$\mathrm{Var}(Y) = \mathbf{E}[\mathrm{Var}(Y \mid \mathscr{F})] + \mathrm{Var}(\mathbf{E}[Y \mid \mathscr{F}])$$

sometimes called the *law of iterated variance*. This result is an easy consequence of the result above. Indeed, simply let $\mathscr{F}$ be trivial in the lemma.

Definition 24.5. We let

$$g_{ij} = \mathrm{Var}(C_{ij} \mid \mathscr{D}).$$

The following proposition allows us to compute g_{ij}.

Proposition 24.6. *Under the assumptions (CL1), (CL2) and (CL3), we have an iterative formula for g_{ij}, namely*

$$g_{ij} = \sigma_{j-1}^2 C_{i,I-i} \prod_{k=I-i}^{j-2} f_k + f_{j-1}^2 g_{i,j-1} = \sigma_{j-1}^2 C_{i,j-1}^{\mathrm{CL}} + f_{j-1}^2 g_{i,j-1},$$

and we have

$$g_{iJ} = (C_{iJ}^{\mathrm{CL}})^2 \sum_{j=I-i}^{J-1} \frac{\sigma_j^2}{f_j^2 C_{ij}^{\mathrm{CL}}}.$$

Proof. We apply the conditional variance formula and obtain

$$g_{ij} = \mathbf{E}[\mathrm{Var}(C_{ij} \mid \mathscr{E}_{j-1} \vee \mathscr{D}) \mid \mathscr{D}] + \mathrm{Var}(\mathbf{E}[C_{ij} \mid \mathscr{E}_{j-1} \vee \mathscr{D}] \mid \mathscr{D}).$$

Using the three assumptions, the first term becomes

$$\begin{aligned}\mathbf{E}[\mathrm{Var}(C_{ij} \mid \mathscr{E}_{j-1} \vee \mathscr{D}) \mid \mathscr{D}] &= \mathbf{E}[\mathrm{Var}(C_{ij} \mid \mathscr{E}_{j-1}) \mid \mathscr{D}] = \mathbf{E}[\sigma_{j-1}^2 C_{i,j-1} \mid \mathscr{D}] \\ &= \sigma_{j-1}^2 \mathbf{E}[C_{i,j-1} \mid \mathscr{D}] = \sigma_{j-1}^2 C_{i,I-i} \prod_{k=I-i}^{j-2} f_k.\end{aligned}$$

The second term becomes

$$\mathrm{Var}(\mathbf{E}[C_{ij} \mid \mathscr{E}_{j-1} \vee \mathscr{D}] \mid \mathscr{D}) = \mathrm{Var}(\mathbf{E}[C_{ij} \mid \mathscr{E}_{j-1}] \mid \mathscr{D}) = \mathrm{Var}(f_{j-1} C_{i,j-1}) = f_{j-1}^2 g_{i,j-1}.$$

Combining these results we get the first formula. To show the second formula, note that $g_{i,I-i} = \mathrm{Var}(C_{i,I-i} \mid \mathscr{D}) = 0$ since $C_{i,I-i}$ is already known. We can now compute

$$\begin{aligned}g_{iJ} &= \sigma_{J-1}^2 C_{i,J-1}^{\mathrm{CL}} + f_{J-1}^2 g_{i,J-1} = \sigma_{J-1}^2 C_{i,J-1}^{\mathrm{CL}} + f_{J-1}^2 (\sigma_{J-2}^2 C_{i,J-2}^{\mathrm{CL}} + f_{J-2}^2 g_{i,J-2}) \\ &= \sigma_{J-1}^2 C_{i,J-1}^{\mathrm{CL}} + f_{J-1}^2 \sigma_{J-2}^2 C_{i,J-2}^{\mathrm{CL}} + f_{J-2}^2 f_{J-1}^2 g_{i,J-2} \\ &= \ldots \\ &= \sum_{k=I-i}^{J-1} \left(\prod_{j=k+1}^{J-1} f_j^2 \right) \sigma_k^2 C_{ik}^{\mathrm{CL}} = \sum_{k=I-i}^{J-1} \left(\prod_{j=k}^{J-1} f_j \right) \frac{\sigma_k^2}{f_k^2} C_{ik}^{\mathrm{CL}} \\ &= \sum_{k=I-i}^{J-1} \left(\frac{C_{iJ}^{\mathrm{CL}}}{C_{ik}^{\mathrm{CL}}} \right)^2 \frac{\sigma_k^2}{f_k^2} C_{ik}^{\mathrm{Cl}} = (C_{iJ}^{\mathrm{CL}})^2 \sum_{k=I-i}^{J-1} \frac{\sigma_k^2}{f_k^2 C_{ik}^{\mathrm{CL}}}.\end{aligned}$$

■

We are now ready to estimate the f_j. To do so, first consider the ratio of claims in year i from development year j to $j+1$,

$$F_{ij} = \frac{C_{i,j+1}}{C_{ij}} \quad \text{for } i \leq I-j-1.$$

The condition on i follows from the fact that the accumulated claims C_{ij} are known for $i+j \leq I$. Note that

$$\mathbf{E}[F_{ij} \mid \mathscr{C}_j] = \frac{1}{C_{ij}}\mathbf{E}[C_{i,j+1} \mid \mathscr{C}_j] = \frac{1}{C_{ij}} f_j C_{ij} = f_j.$$

This observation leads to considering an estimator of the form

$$\widehat{f}_j = \sum_{i=0}^{I-j-1} a_i F_{ij}, \quad \text{where} \quad \sum_{i=0}^{I-j-1} a_i = 1.$$

The restriction on the a_i's makes $\widehat{f}_j$ an unbiased estimator i.e. $\mathbf{E}[\widehat{f}_j] = f_j$. We wish to choose the a_i such that $\mathrm{Var}(\widehat{f}_j \mid \mathscr{C}_j)$ is minimized. The following result provides the optimal weights a_i.

Proposition 24.7. *In the setup above and under assumptions (CL1), (CL2) and (CL3), the choice of a_i that minimizes $Var(\widehat{f}_j \mid \mathscr{C}_j)$ is given by*

$$a_i^* = \frac{C_{ij}}{C_{\bullet,j}}, \quad C_{\bullet,j} = \sum_{i=0}^{I-j-1} C_{ij}.$$

Proof. Let $\mathbf{a} = (a_0, ..., a_{I-j-1})$ and define the function

$$h(\mathbf{a}) = \mathrm{Var}(\widehat{f}_j \mid \mathscr{C}_j).$$

We want to minimize $h(\mathbf{a})$ given the constraint

$$q(\mathbf{a}) = \sum_{i=0}^{I-j-1} a_i = 1.$$

To solve this problem we use Lagrange multipliers which tells us that a minimizer $\mathbf{a}^*$ satisfies

$$\nabla h(\mathbf{a}^*) = \lambda \nabla q(\mathbf{a}^*)$$

for a real number λ. We compute using assumptions (CL1) and (CL3)

$$\begin{aligned} h(\mathbf{a}) &= \mathrm{Var}\left(\sum_{i=0}^{I-j-1} a_i F_{ij}\right) = \sum_{i=0}^{I-j-1} a_i^2 \mathrm{Var}(F_{ij} \mid \mathscr{C}_j) = \sum_{i=0}^{I-j-1} \frac{a_i^2}{C_{ij}^2}\mathrm{Var}(C_{i,j+1} \mid \mathscr{C}_j) \\ &= \sum_{i=0}^{I-j-1} \frac{a_i^2}{C_{ij}^2}\sigma_j^2 C_{ij} = \sum_{i=0}^{I-j-1} a_i^2 \frac{\sigma_j^2}{C_{ij}}. \end{aligned}$$

Therefore,

$$\frac{\partial}{\partial a_i} h(\mathbf{a}) = 2a_i \frac{\sigma_j^2}{C_{ij}} \quad \text{and} \quad \frac{\partial}{\partial a_i} q(\mathbf{a}) = 1.$$

The Lagrange multiplier equation thus becomes

$$2a_i^* \frac{\sigma_j^2}{C_{ij}} = \lambda, \quad i = 0, ..., I-j-1$$

i.e.

$$a_i^* = \frac{\lambda C_{ij}}{2\sigma_j^2}.$$

We can use our constraint to solve for λ as follows

$$1 = \sum_{i=0}^{I-j-1} a_i^* = \sum_{i=0}^{I-j-1} \frac{\lambda C_{ij}}{2\sigma_j^2} = \frac{\lambda}{2\sigma_j^2} \sum_{i=0}^{I-j-1} C_{ij} = \frac{\lambda}{2\sigma_j^2} C_{\bullet,j}.$$

This yields

$$\lambda = \frac{2\sigma_j^2}{C_{\bullet,j}}$$

and therefore

$$2a_i^* \frac{\sigma_j^2}{C_{ij}} = \frac{2\sigma_j^2}{C_{\bullet,j}}.$$

Rearranging completes the proof. ■

Remark 24.8. The notes by Jostein Paulsen [5] has an alternative proof that does not rely on Lagrange multipliers, see page 9.

Remark 24.9. Using the law of total variance, we have

$$\mathrm{Var}(\widehat{f}_j) = \mathbf{E}[\mathrm{Var}(\widehat{f}_j \mid \mathscr{E}_j)] + \mathrm{Var}(\mathbf{E}[\widehat{f}_j \mid \mathscr{E}_j]) = \mathbf{E}[\mathrm{Var}(\widehat{f}_j \mid \mathscr{E}_j)] + \mathrm{Var}(f_j) = \mathbf{E}[\mathrm{Var}(\widehat{f}_j \mid \mathscr{E}_j)],$$

so minimizing $\mathrm{Var}(\widehat{f}_j \mid \mathscr{E}_j)$ is equivalent to minimizing the unconditional variance $\mathrm{Var}(\widehat{f}_j)$. Plugging the optimal weights into the estimator $\widehat{f}_j$ gives the chain ladder estimators of f_j.

Definition 24.10. The chain ladder estimators of the development factors are given by

$$\widehat{f}_j^{\mathrm{CL}} = \sum_{i=0}^{I-j-1} \frac{C_{i,j+1}}{C_{\bullet,j}} = \frac{\sum_{i=0}^{I-j-1} C_{i,j+1}}{\sum_{i=0}^{I-j-1} C_{ij}},$$

and we define the chain ladder estimate of the accumulated losses in accident year i as

$$\widehat{C}_{iJ}^{\mathrm{CL}} = C_{i,I-i} \prod_{j=I-i}^{J-1} \widehat{f}_j.$$

25 Final comments

There are two aspects that we will briefly discuss, namely the question of estimating the mean squared error for the total losses estimate and parametric methods.

The mean squared error

We have an estimate for the total claims losses, namely

$$\widehat{C}_T = \sum_{i=0}^{I} \widehat{C}_{iJ}^{\mathrm{CL}}.$$

The mean squared error given $\mathscr{D}$ is given by

$$\begin{aligned}\mathrm{MSE}(\widehat{C}_T \mid \mathscr{D}) &= \mathbf{E}[(C_T - \widehat{C}_T)^2 \mid \mathscr{D}] = \mathbf{E}[C_T^2 \mid \mathscr{D}] - 2\widehat{C}_T \mathbf{E}[C_T \mid \mathscr{D}] + \widehat{C}_T^2 \\ &= (\mathbf{E}[C_T^2 \mid \mathscr{D}] - \mathbf{E}[C_T \mid \mathscr{D}]^2) + \widehat{C}_T^2 - 2\widehat{C}_T \mathbf{E}[C_T \mid \mathscr{D}] + \mathbf{E}[C_T \mid \mathscr{D}]^2 \\ &= \mathrm{Var}(C_T \mid \mathscr{D}) + (\mathbf{E}[C_T \mid \mathscr{D}] - \widehat{C}_T)^2\end{aligned}$$

where we used that $\widehat{C}_T$ is a function of the observations in $\mathscr{D}$. This equation is a variation of the famous *bias-variance tradeoff*. We already have the tools to compute the first term, and one can compute the empirical estimates of the σ_j^2 to compute $\mathrm{Var}(C_T \mid \mathscr{D})$ in practice. The other term is a lot more complicated. One can show

$$(\mathbf{E}[C_T \mid \mathscr{D}] - \widehat{C}_T)^2 = \left(\sum_{i=I-J+1}^{I} C_{i,I-i} \left(\prod_{j=I-i}^{J-1} f_j - \prod_{j=I-i}^{J-1} \widehat{f}_j \right) \right)^2,$$

and the problem with this expression is that it depends on both f_j and $\widehat{f}_j$. Futhermore, the natural idea of replacing f_j with $\widehat{f}_j$ just makes the term vanish. One way to remedy this issue was proposed by Mack. The idea is to condition on $\mathscr{E}_j$ in a clever way. See page 11 of [5] for the details.

Parametric methods

So far we have taken a nonparametric approach to chain ladder estimation i.e. to estimate f_j, we have not assumed a distribution on the claims. A parametric approach is to assume

$$C_{i,j+1} \mid \mathscr{E}_j \sim H_j$$

for some distribution H_j. An example could be

$$C_{i,j+1} \sim \mathcal{N}(C_{ij} f_j, C_{ij} \sigma_j^2).$$

The parameters in the model are then determined using maximum likelihood methods. For an introduction to this subject, see section 1.2.2 in [5].

Bibliography

[1] Gerold Alsmeyer. *Erneuerungstheorie*. Springer, 1 edition. ISBN 978-3-663-09977-2.

[2] Søren Asmussen and Mogens Steffensen. *Risk and Insurance - A Graduate Text*. Springer, 2019. ISBN 978-3-030-35175-5.

[3] William Feller. *An Introduction to Probability Theory and Its Applications - Volume II*. Wiley, 2 edition, 1971. ISBN 978-0471257097.

[4] George Lowther. *Optional Sampling - Almost Sure*. 2009. URL `https://almostsuremath.com/2009/12/20/optional-sampling/`.

[5] Jostein Paulsen. *Claim reservation in non-life insurance*. University of Copenhagen, 2016.

[6] Hanspeter Schmidli. *Lecture notes*.

[7] Alexander Sokol. *An introduction to stochastic integration with respect to general semimartingales*. 2013. URL `https://alexandersokol.dk/miscellaneous/2012_GeneralStochasticIntegration.pdf`.

[8] E. J. Watson. *Laplace Transforms and Applications*. Van Nostrand Reinhold Company Ltd., 1981. ISBN 0-442-30176-6.

Index