Lecture notes on

Risk Theory

by Hanspeter Schmidli

Institute of Mathematics
University of Cologne

Contents

1. Risk Models

1.1. Introduction

Let us consider a (collective) insurance contract in some fixed time period $(0, T]$, for instance $T = 1$ year. Let N denote the number of claims in $(0, T]$ and $Y_1, Y_2, \ldots, Y_N$ the corresponding claims. Then

$$S = \sum_{i=1}^{N} Y_i$$

is the accumulated sum of claims. We assume

i) The claim sizes are not affected by the number of claims.

ii) The amount of an individual claim is not affected by the others.

iii) All individual claims are exchangeable.

In probabilistic terms:

i) N and $(Y_1, Y_2, \ldots)$ are independent.

ii) $Y_1, Y_2, \ldots$ are independent.

iii) $Y_1, Y_2, \ldots$ have the same distribution function, G say.

We assume that $G(0) = 0$, i.e. the claim amounts are positive. Let $M_Y(r) = \mathbb{E}[\mathrm{e}^{rY_i}]$, $\mu_n = \mathbb{E}[Y_1^n]$ if the expressions exist and $\mu = \mu_1$. The distribution of S can be written as

$$\begin{aligned}\mathbb{P}[S \le x] = \mathbb{E}[\mathbb{P}[S \le x \mid N]] &= \sum_{n=0}^{\infty} \mathbb{P}[S \le x \mid N = n]\mathbb{P}[N = n] \\ &= \sum_{n=0}^{\infty} \mathbb{P}[N = n] G^{*n}(x)\,.\end{aligned}$$

This is in general not easy to compute. It is often enough to know some characteristics of a distribution.

$$\mathbb{E}[S] = \mathbb{E}\Bigl[\sum_{i=1}^{N} Y_i\Bigr] = \mathbb{E}\Bigl[\mathbb{E}\Bigl[\sum_{i=1}^{N} Y_i \Bigm| N\Bigr]\Bigr] = \mathbb{E}\Bigl[\sum_{i=1}^{N} \mu\Bigr] = \mathbb{E}[N\mu] = \mathbb{E}[N]\mu$$

and

$$\begin{aligned}\mathbb{E}[S^2] &= \mathbb{E}\Bigl[\mathbb{E}\Bigl[\Bigl(\sum_{i=1}^{N} Y_i\Bigr)^2 \Bigm| N\Bigr]\Bigr] = \mathbb{E}\Bigl[\mathbb{E}\Bigl[\sum_{i=1}^{N}\sum_{j=1}^{N} Y_iY_j \Bigm| N\Bigr]\Bigr] \\ &= \mathbb{E}[N\mu_2 + N(N-1)\mu^2] = \mathbb{E}[N^2]\mu^2 + \mathbb{E}[N](\mu_2 - \mu^2)\,,\end{aligned}$$

and thus

$$\operatorname{Var}[S] = \operatorname{Var}[N]\mu^2 + \mathbb{E}[N]\operatorname{Var}[Y_1]\,.$$

The moment generating function of S becomes

$$\begin{aligned}M_S(r) &= \mathbb{E}[\mathrm{e}^{rS}] = \mathbb{E}\Bigl[\exp\Bigl\{r\sum_{i=1}^{N} Y_i\Bigr\}\Bigr] = \mathbb{E}\Bigl[\prod_{i=1}^{N} \mathrm{e}^{rY_i}\Bigr] = \mathbb{E}\Bigl[\mathbb{E}\Bigl[\prod_{i=1}^{N} \mathrm{e}^{rY_i} \Bigm| N\Bigr]\Bigr] \\ &= \mathbb{E}\Bigl[\prod_{i=1}^{N} M_Y(r)\Bigr] = \mathbb{E}\left[(M_Y(r))^N\right] = \mathbb{E}\left[\mathrm{e}^{N\log(M_Y(r))}\right] = M_N(\log(M_Y(r)))\end{aligned}$$

where $M_N(r)$ is the moment generating function of N. The coefficient of skewness $\mathbb{E}[(S - \mathbb{E}[S])^3]/(\operatorname{Var}[S])^{3/2}$ can be computed from the moment generating function by using the formula

$$\mathbb{E}[(S - \mathbb{E}[S])^3] = \frac{\mathrm{d}^3}{\mathrm{d}r^3}\log(M_S(r))\Bigr|_{r=0}\,.$$

1.2. The Compound Binomial Model

We make in addition the following assumptions:

- The whole risk consists of several independent exchangeable single risks.
- The time interval under consideration can be split into several independent and exchangeable smaller intervals.
- There is at most one claim per single risk and interval.

Let the probability of a claim in a single risk in an interval be p. Then it follows from the above assumptions that

$$N \sim \mathrm{B}(n, p)$$

for some $n \in \mathbb{N}$. We obtain

$$\mathbb{E}[S] = np\mu\,,$$

$$\mathrm{Var}[S] = np(1-p)\mu^2 + np(\mu_2 - \mu^2) = np(\mu_2 - p\mu^2)$$

and

$$M_S(r) = (pM_Y(r) + 1 - p)^n \,.$$

Let us consider another characteristics of the distribution of S, the skewness. We need to compute $\mathbb{E}[(S - \mathbb{E}[S])^3]$.

$$\begin{aligned}\frac{\mathrm{d}^3}{\mathrm{d}r^3} n \log(pM_Y(r) + 1 - p) &= n\frac{\mathrm{d}^2}{\mathrm{d}r^2}\left(\frac{pM_Y'(r)}{pM_Y(r)+1-p}\right)\\ &= n\frac{\mathrm{d}}{\mathrm{d}r}\left(\frac{pM_Y''(r)}{pM_Y(r)+1-p} - \frac{p^2M_Y'(r)^2}{(pM_Y(r)+1-p)^2}\right)\\ &= n\left(\frac{pM_Y'''(r)}{pM_Y(r)+1-p} - \frac{3p^2M_Y''(r)M_Y'(r)}{(pM_Y(r)+1-p)^2} + \frac{2p^3(M_Y'(r))^3}{(pM_Y(r)+1-p)^3}\right) .\end{aligned}$$

For $r = 0$ we get

$$\mathbb{E}[(S - \mathbb{E}[S])^3] = n(p\mu_3 - 3p^2\mu_2\mu + 2p^3\mu^3)$$

from which the coefficient of skewness can be computed.

Example 1.1. Assume that the claim amounts are deterministic, y_0 say. Then

$$\mathbb{E}[(S - \mathbb{E}[S])^3] = ny_0^3(p - 3p^2 + 2p^3) = 2ny_0^3p(\tfrac{1}{2} - p)(1-p)\,.$$

Thus

$$\mathbb{E}[(S - \mathbb{E}[S])^3] \gtreqless 0 \iff p \lesseqgtr \tfrac{1}{2}\,.$$

■

1.3. The Compound Poisson Model

In addition to the compound binomial model we assume

- n is large and p is small.

Let $\lambda = np$. Because

$$B(n, \lambda/n) \longrightarrow \mathrm{Pois}(\lambda) \qquad\qquad \text{as } n \to \infty$$

it is natural to model

$$N \sim \mathrm{Pois}(\lambda)\,.$$

We get

$$\mathbb{E}[S] = \lambda\mu\,,$$

$$\operatorname{Var}[S] = \lambda\mu^2 + \lambda(\mu_2 - \mu^2) = \lambda\mu_2$$

and

$$M_S(r) = \exp\{\lambda(M_Y(r) - 1)\}\,.$$

Let us also compute the coefficient of skewness.

$$\frac{\mathrm{d}^3}{\mathrm{d}r^3}\log(M_S(r)) = \frac{\mathrm{d}^3}{\mathrm{d}r^3}(\lambda(M_Y(r) - 1)) = \lambda M_Y'''(r)$$

and thus

$$\mathbb{E}[(S - \mathbb{E}[S])^3] = \lambda\mu_3\,.$$

The coefficient of skewness is

$$\frac{\mathbb{E}[(S - \mathbb{E}[S])^3]}{(\operatorname{Var}[S])^{3/2}} = \frac{\mu_3}{\sqrt{\lambda\mu_2^3}} > 0\,.$$

Problem: The distribution of S is always positively skewed.

Example 1.2. (Fire insurance) An insurance company models claims caused by fire as $\mathrm{LN}(m,\, \sigma^2)$. Let us first compute the moments

$$\mu_n = \mathbb{E}[Y_1^n] = \mathbb{E}[\mathrm{e}^{n\log Y_1}] = M_{\log Y_1}(n) = \exp\{\sigma^2 n^2/2 + nm\}\,.$$

Thus

$$\mathbb{E}[S] = \lambda\exp\{\sigma^2/2 + m\}\,,$$

$$\operatorname{Var}[S] = \lambda\exp\{2\sigma^2 + 2m\}$$

and

$$\frac{\mathbb{E}[(S - \mathbb{E}[S])^3]}{(\operatorname{Var}[S])^{3/2}} = \frac{\exp\{9\sigma^2/2 + 3m\}}{\sqrt{\lambda\exp\{6\sigma^2 + 6m\}}} = \frac{\exp\{3\sigma^2/2\}}{\sqrt{\lambda}}\,.$$

■

The computation of the characteristics of a risk is much easier for the compound Poisson model than for the compound binomial model. But using a compound Poisson model has another big advantage. Assume that a portfolio consists of several independent single risks $S^{(1)}, S^{(2)}, \ldots, S^{(j)}$, each modelled as compound Poisson. For simplicity we use $j = 2$ in the following calculation. We want to find the moment generating function of $S^{(1)} + S^{(2)}$.

$$\begin{aligned} M_{S^{(1)}+S^{(2)}}(r) &= M_{S^{(1)}}(r) M_{S^{(2)}}(r) \\ &= \exp\{\lambda^{(1)}(M_{Y^{(1)}}(r) - 1)\} \exp\{\lambda^{(2)}(M_{Y^{(2)}}(r) - 1)\} \\ &= \exp\Bigl\{\lambda\Bigl(\frac{\lambda^{(1)}}{\lambda} M_{Y^{(1)}}(r) + \frac{\lambda^{(2)}}{\lambda} M_{Y^{(2)}}(r) - 1\Bigr)\Bigr\} \end{aligned}$$

where $\lambda = \lambda^{(1)} + \lambda^{(2)}$. It follows that $S^{(1)} + S^{(2)}$ is compound Poisson distributed with Poisson parameter λ and claim size distribution function

$$G(x) = \frac{\lambda^{(1)}}{\lambda} G^{(1)}(x) + \frac{\lambda^{(2)}}{\lambda} G^{(2)}(x)\,.$$

The claim size can be obtained by choosing it from the first risk with probability $\lambda^{(1)}/\lambda$ and from the second risk with probability $\lambda^{(2)}/\lambda$.

Let us now split the claim amounts into different classes. Let $A_1, A_2, \ldots, A_m$ be some disjoint sets with $\mathbb{P}[Y_1 \in \cup_{k=1}^m A_k] = 1$. Let $p_k = \mathbb{P}[Y_1 \in A_k]$ be the probability that a claim is in claim size class k. We can assume that $p_k > 0$ for all k. We denote by N_k the number of claims in claim size class k. Because the claim amounts are independent it follows that, given $N = n$, the vector $(N_1, N_2, \ldots, N_m)$ is conditionally multinomial distributed with parameters $p_1, p_2, \ldots, p_m, n$. We now want to find the unconditioned distribution of $(N_1, N_2, \ldots, N_m)$. Let $n_1, n_2, \ldots, n_m$ be natural numbers and $n = n_1 + n_2 + \cdots + n_m$.

$$\begin{aligned} \mathbb{P}[N_1 = n_1, N_2 = n_2, \ldots, N_m = n_m] &= \mathbb{P}[N_1 = n_1, N_2 = n_2, \ldots, N_m = n_m, N = n] \\ &= \mathbb{P}[N_1 = n_1, N_2 = n_2, \ldots, N_m = n_m \mid N = n]\mathbb{P}[N = n] \\ &= \frac{n!}{n_1! n_2! \cdots n_m!} p_1^{n_1} p_2^{n_2} \cdots p_m^{n_m} \frac{\lambda^n}{n!} \mathrm{e}^{-\lambda} = \prod_{k=1}^m \frac{(\lambda p_k)^{n_k}}{n_k!} \mathrm{e}^{-\lambda p_k} \end{aligned}$$

It follows that $N_1, N_2, \ldots, N_m$ are independent and that N_k is $\mathrm{Pois}(\lambda p_k)$ distributed. Because the claim sizes are independent of N the risks

$$S_k = \sum_{i=1}^N Y_i \mathbb{1}_{\{Y_i \in A_k\}}$$

are compound Poisson distributed with Poisson parameter λp_k and claim size distribution

$$G_k(x) = \mathbb{P}[Y_1 \le x \mid Y_1 \in A_k] = \frac{\mathbb{P}[Y_1 \le x, Y_1 \in A_k]}{\mathbb{P}[Y_1 \in A_k]}\,.$$

Moreover the sums $\{S_k\}$ are independent. In the special case where $A_k = (t_{k-1}, t_k]$ we get

$$G_k(x) = \frac{G(x) - G(t_{k-1})}{G(t_k) - G(t_{k-1})} \qquad (t_{k-1} \le x \le t_k).$$

1.4. The Compound Mixed Poisson Model

As mentioned before it is a disadvantage of the compound Poisson model that the distribution is always positively skewed. In practice it also often turns out that the model does not allow enough fluctuations. For instance $\mathbb{E}[N] = \mathrm{Var}[N]$. A simple way to allow for more fluctuation is to let the parameter λ be stochastic. Let H denote the distribution function of λ.

$$\mathbb{P}[N = n] = \mathbb{E}[\mathbb{P}[N = n \mid \lambda]] = \mathbb{E}\Big[\frac{\lambda^n}{n!}\mathrm{e}^{-\lambda}\Big] = \int_0^\infty \frac{\ell^n}{n!}\mathrm{e}^{-\ell}\,\mathrm{d}H(\ell)\;.$$

The moments are

$$\mathbb{E}[S] = \mathbb{E}[\mathbb{E}[S \mid \lambda]] = \mathbb{E}[\lambda\mu] = \mathbb{E}[\lambda]\mu\,, \tag{1.1a}$$

$$\mathbb{E}[S^2] = \mathbb{E}[\mathbb{E}[S^2 \mid \lambda]] = \mathbb{E}[\lambda\mu_2 + \lambda^2\mu^2] = \mathbb{E}[\lambda^2]\mu^2 + \mathbb{E}[\lambda]\mu_2 \tag{1.1b}$$

and

$$\mathbb{E}[S^3] = \mathbb{E}[\lambda\mu_3 + 3\lambda^2\mu_2\mu + \lambda^3\mu^3] = \mathbb{E}[\lambda^3]\mu^3 + 3\mathbb{E}[\lambda^2]\mu_2\mu + \mathbb{E}[\lambda]\mu_3\,. \tag{1.1c}$$

Thus the variance is

$$\mathrm{Var}[S] = \mathrm{Var}[\lambda]\mu^2 + \mathbb{E}[\lambda]\mu_2$$

and the third centralized moment becomes

$$\begin{aligned}
&\mathbb{E}[(S - \mathbb{E}[\lambda]\mu)^3] \\
&\qquad = \mathbb{E}[\lambda^3]\mu^3 + 3\mathbb{E}[\lambda^2]\mu_2\mu + \mathbb{E}[\lambda]\mu_3 - 3(\mathbb{E}[\lambda^2]\mu^2 + \mathbb{E}[\lambda]\mu_2)\mathbb{E}[\lambda]\mu + 2\mathbb{E}[\lambda]^3\mu^3 \\
&\qquad = \mathbb{E}[(\lambda - \mathbb{E}[\lambda])^3]\mu^3 + 3\,\mathrm{Var}[\lambda]\mu_2\mu + \mathbb{E}[\lambda]\mu_3\,.
\end{aligned}$$

We can see that the coefficient of skewness can also be negative.

It remains to compute the moment generating function

$$M_S(r) = \mathbb{E}\left[\mathbb{E}\left[\mathrm{e}^{rS} \mid \lambda\right]\right] = \mathbb{E}[\exp\{\lambda(M_Y(r) - 1)\}] = M_\lambda(M_Y(r) - 1)\,. \tag{1.2}$$

1.5. The Compound Negative Binomial Model

Let us first consider an example of the compound mixed Poisson model. Assume that $\lambda \sim \Gamma(\gamma, \beta)$. Then

$$M_N(r) = M_\lambda(\mathrm{e}^r - 1) = \Bigl(\frac{\beta}{\beta - (\mathrm{e}^r - 1)}\Bigr)^\gamma = \Bigl(\frac{\frac{\beta}{\beta+1}}{1 - (1 - \frac{\beta}{\beta+1})\mathrm{e}^r}\Bigr)^\gamma .$$

In this case N has a negative binomial distribution. Actuaries started using the negative binomial distribution for the number of claims a long time ago. They recognized that a Poisson distribution seldomly fits the real data. Estimating the parameters in a negative binomial distribution yielded satisfactory results.

Let now $N \sim \mathrm{NB}(\alpha, p)$. Then

$$\mathbb{E}[S] = \frac{\alpha(1-p)}{p}\mu$$

and

$$\mathrm{Var}[S] = \frac{\alpha(1-p)}{p^2}\mu^2 + \frac{\alpha(1-p)}{p}(\mu_2 - \mu^2) .$$

The moment generating function is

$$M_S(r) = \Bigl(\frac{p}{1 - (1-p)M_Y(r)}\Bigr)^\alpha .$$

The third centralized moment can be computed as

$$\mathbb{E}[(S - \mathbb{E}[S])^3] = \alpha\Bigl(\frac{1}{p} - 1\Bigr)\mu_3 + 3\alpha\Bigl(\frac{1}{p} - 1\Bigr)^2\mu_2\mu + 2\alpha\Bigl(\frac{1}{p} - 1\Bigr)^3\mu^3 .$$

Note that the compound negative binomial distribution is always positively skewed. Thus the compound negative binomial distribution does not fulfil all desired properties. Nevertheless in practice almost all risks are positively skewed.

1.6. A Note on the Individual Model

Assume that a portfolio consist of m independent individual contracts $(S^{(i)})_{i \le m}$. There can be at most one claim for each of the contracts. Such a claim occurs with probability $p^{(i)}$. Its size has distribution function $F^{(i)}$ and moment generating function $M^{(i)}(r)$. Let $\lambda = \sum_{i=1}^m p^{(i)}$. The moment generating function of the aggregate claims from the portfolio is

$$M_S(r) = \prod_{i=1}^m (1 + p^{(i)}(M^{(i)}(r) - 1)) .$$

The term $p^{(i)}(M^{(i)}(r)-1)$ is small for r not too large. Consider the logarithm

$$\begin{aligned}\log(M_S(r)) &= \sum_{i=1}^{m} \log(1+p^{(i)}(M^{(i)}(r)-1)) \\ &\approx \sum_{i=1}^{m} p^{(i)}(M^{(i)}(r)-1) = \lambda\Bigl(\sum_{i=1}^{m} \frac{p^{(i)}}{\lambda}(M^{(i)}(r)-1)\Bigr)\,.\end{aligned}$$

The last expression turns out to be the logarithm of the moment generating function of a compound Poisson distribution. In fact this derivation is the reason that the compound Poisson model is very popular amongst actuaries.

1.7. A Note on Reinsurance

For most of the insurance companies the premium volume is not big enough to carry the risk. This is the case especially in the case of large claim sizes, like insurance against damages caused by hurricanes or earthquakes. Therefore the insurers try to share part of the risk with other companies. Sharing the risk is done via reinsurance. Let S^I denote the part of the risk taken by the insurer and S^R the part taken by the reinsurer. Reinsurance can act on the individual claims or it can act on the whole risk S. Let f be an increasing function with $f(0)=0$ and $f(x) \le x$ for all $x \ge 0$. A reinsurance form acting on the individual claims is

$$S^I = \sum_{i=1}^{N} f(Y_i)\,, \qquad\qquad S^R = S - S^I\,.$$

The most common reinsurance forms are

- proportional reinsurance $\quad f(x) = \alpha x, \quad (0 < \alpha < 1)$,
- excess of loss reinsurance $\quad f(x) = \min\{x, M\}, \quad (M > 0)$.

We will consider these two reinsurance forms in the sequel. A reinsurance form acting on the whole risk is

$$S^I = f(S)\,, \qquad\qquad S^R = S - S^I\,.$$

The most common example of this reinsurance form is the

- stop loss reinsurance $\quad f(x) = \min\{x, M\}, \quad (M > 0)$.

1.7.1. Proportional reinsurance

For proportional reinsurance we have $S^I = \alpha S$, and thus

$$\mathbb{E}[S^I] = \alpha \mathbb{E}[S]\,,$$

$$\operatorname{Var}[S^I] = \alpha^2 \operatorname{Var}[S]\,,$$

$$\frac{\mathbb{E}[(S^I - \mathbb{E}[S^I])^3]}{(\operatorname{Var}[S^I])^{3/2}} = \frac{\mathbb{E}[(S - \mathbb{E}[S])^3]}{(\operatorname{Var}[S])^{3/2}}$$

and

$$M_{S^I}(r) = \mathbb{E}\left[\mathrm{e}^{r\alpha S}\right] = M_S(\alpha r)\,.$$

We can see that the coefficient of skewness does not change. But the variance is much smaller. The following consideration also suggest that the risk has decreased. Let the premium charged for this contract be p and assume that the insurer gets αp, the reinsurer $(1-\alpha)p$. Let the initial capital of the company be u. The probability that the company gets ruined after one year is

$$\mathbb{P}[\alpha S > \alpha p + u] = \mathbb{P}[S > p + u/\alpha]\,.$$

The effect for the insurer is thus like having a larger initial capital.

Remark. The computations for the reinsurer can be obtained by replacing α with $(1-\alpha)$. ■

1.7.2. Excess of loss reinsurance

Under excess of loss reinsurance we cannot give formulae of the above type for the cumulants and the moment generating function. They all have to be computed from the new distribution function of the claim sizes $Y_i^I = \min\{Y_i, M\}$. An indication that the risk has decreased for the insurer is that the claim sizes are bounded. This is wanted especially in the case of large claims.

Example 1.3. Let S be the compound Poisson model with parameter λ and $\operatorname{Pa}(\alpha, \beta)$ distributed claim sizes. Assume that $\alpha > 1$, i.e. that $\mathbb{E}[Y_i] < \infty$. Let us compute the expected value of the outgo paid by the insurer.

$$\begin{aligned}
\mathbb{E}[Y_i^I] &= \int_0^M x\frac{\alpha\beta^\alpha}{(\beta+x)^{\alpha+1}}\,\mathrm{d}x + \int_M^\infty M\frac{\alpha\beta^\alpha}{(\beta+x)^{\alpha+1}}\,\mathrm{d}x \\
&= \int_0^M (x+\beta)\frac{\alpha\beta^\alpha}{(\beta+x)^{\alpha+1}}\,\mathrm{d}x - \int_0^M \beta\frac{\alpha\beta^\alpha}{(\beta+x)^{\alpha+1}}\,\mathrm{d}x + M\frac{\beta^\alpha}{(\beta+M)^\alpha} \\
&= \frac{\alpha\beta^\alpha}{\alpha-1}\Bigl(\frac{1}{\beta^{\alpha-1}} - \frac{1}{(\beta+M)^{\alpha-1}}\Bigr) - \beta^{\alpha+1}\Bigl(\frac{1}{\beta^\alpha} - \frac{1}{(\beta+M)^\alpha}\Bigr) + M\frac{\beta^\alpha}{(\beta+M)^\alpha} \\
&= \frac{\beta}{\alpha-1} - \frac{\beta^\alpha}{(\alpha-1)(\beta+M)^{\alpha-1}} = \Bigl(1 - \Bigl(\frac{\beta}{\beta+M}\Bigr)^{\alpha-1}\Bigr)\frac{\beta}{\alpha-1} \\
&= \Bigl(1 - \Bigl(\frac{\beta}{\beta+M}\Bigr)^{\alpha-1}\Bigr)\mathbb{E}[Y_i]\,.
\end{aligned}$$

For $\mathbb{E}[S^I]$ follows that

$$\mathbb{E}[S^I] = \Bigl(1 - \Bigl(\frac{\beta}{\beta+M}\Bigr)^{\alpha-1}\Bigr)\mathbb{E}[S]\,.$$

■

Let

$$N^R = \sum_{i=1}^N \mathbb{1}_{\{Y_i>M\}}$$

denote the number of claims the reinsurer has to pay for. We denote by $q = \mathbb{P}[Y_i > M]$ the probability that a claim amount exceeds the level M. What is the distribution of N^R? We first note that the moment generating function of $\mathbb{1}_{\{Y_i>M\}}$ is $q\mathrm{e}^r + 1 - q$.

i) Let $N \sim \mathrm{B}(n,p)$. The moment generating function of N^R is

$$M_{N^R}(r) = (p(q\mathrm{e}^r + 1 - q) + 1 - p)^n = (pq\mathrm{e}^r + 1 - pq)^n\,.$$

Thus $N^R \sim \mathrm{B}(n,pq)$.

ii) Let $N \sim \mathrm{Pois}(\lambda)$. The moment generating function of N^R is

$$M_{N^R}(r) = \exp\{\lambda((q\mathrm{e}^r + 1 - q) - 1)\} = \exp\{\lambda q(\mathrm{e}^r - 1)\}\,.$$

Thus $N^R \sim \mathrm{Pois}(\lambda q)$.

iii) Let $N \sim \mathrm{NB}(\alpha,p)$. The moment generating function of N^R is

$$M_{N^R}(r) = \Bigl(\frac{p}{1-(1-p)(q\mathrm{e}^r + 1 - q)}\Bigr)^\alpha = \Bigl(\frac{\frac{p}{p+q-pq}}{1-\Bigl(1-\frac{p}{p+q-pq}\Bigr)\mathrm{e}^r}\Bigr)^\alpha\,.$$

Thus $N^R \sim \mathrm{NB}(\alpha, \frac{p}{p+q-pq})$.

1.8. Computation of the Distribution of S in the Discrete Case

Let us consider the case where the claim sizes (Y_i) have an arithmetic distribution. We can assume that $\mathbb{P}[Y_i \in \mathbb{N}] = 1$ by choosing the monetary unit appropriate. We denote by $p_k = \mathbb{P}[N = k]$, $f_k = \mathbb{P}[Y_i = k]$ and $g_k = \mathbb{P}[S = k]$. For simplicity let us assume that $f_0 = 0$. Let $f_k^{*n} = \mathbb{P}[Y_1 + Y_2 + \cdots Y_n = k]$ denote the convolutions of the claim size distribution. Note that

$$f_k^{*(n+1)} = \sum_{i=1}^{k-1} f_i^{*n} f_{k-i}\,.$$

We get the following identities:

$$g_0 = \mathbb{P}[S = 0] = \mathbb{P}[N = 0] = p_0\,,$$

$$g_n = \mathbb{P}[S = n] = \mathbb{E}[\mathbb{P}[S = n \mid N]] = \sum_{k=1}^{n} p_k f_n^{*k}\,.$$

We have explicit formulae for the distribution of S. But the computation of the f_n^{*k}'s is messy. An easier procedure is called for. Let us now make an assumption on the distribution of N.

Assumption 1.4. *Assume that there exist real numbers a and b such that*

$$p_r = \Bigl(a + \frac{b}{r}\Bigr) p_{r-1}$$

for $r \in \mathbb{N} \setminus \{0\}$.

Let us check the assumption for the distribution of the models we had considered.

i) **Binomial** $\mathrm{B}(n,p)$

$$\begin{aligned}\frac{p_r}{p_{r-1}} &= \frac{\frac{n!}{r!(n-r)!} p^r (1-p)^{n-r}}{\frac{n!}{(r-1)!(n-r+1)!} p^{r-1} (1-p)^{n-r+1}} = \frac{(n-r+1)p}{r(1-p)} \\ &= -\frac{p}{1-p} + \frac{(n+1)p}{r(1-p)}\,.\end{aligned}$$

Thus

$$a = -\frac{p}{1-p}\,, \qquad b = \frac{(n+1)p}{1-p}\,.$$

ii) **Poisson** Pois(λ)

$$\frac{p_r}{p_{r-1}} = \frac{\frac{\lambda^r}{r!}\mathrm{e}^{-\lambda}}{\frac{\lambda^{r-1}}{(r-1)!}\mathrm{e}^{-\lambda}} = \frac{\lambda}{r}\,.$$

Thus

$$a = 0\,, \qquad\qquad b = \lambda\,.$$

iii) **Negative binomial** NB(α, p)

$$\begin{aligned}\frac{p_r}{p_{r-1}} &= \frac{\frac{\Gamma(\alpha+r)}{r!\Gamma(\alpha)}p^\alpha(1-p)^r}{\frac{\Gamma(\alpha+r-1)}{(r-1)!\Gamma(\alpha)}p^\alpha(1-p)^{r-1}} = \frac{(\alpha+r-1)(1-p)}{r} \\ &= 1-p+\frac{(\alpha-1)(1-p)}{r}\,.\end{aligned}$$

Thus

$$a = 1-p\,, \qquad\qquad b = (\alpha-1)(1-p)\,.$$

These are in fact the only distributions satisfying the assumption. If we choose $a = 0$ we get by induction $p_r = p_0 b^r/r!$, which is the Poisson distribution.

If $a < 0$, then because $a + b/r$ is negative for r large enough there must be an $n_0 \in \mathbb{N}$ such that $b = -an_0$ in order that $p_r \ge 0$ for all r. Letting $n = n_0 - 1$ and $p = -a(1-a)$ we get the binomial distribution.

If $a > 0$ we need $a + b \ge 0$ in order that $p_1 \ge 0$. The case $a + b = 0$ can be considered as the degenerate case of a Poisson distribution with $\lambda = 0$. Suppose therefore that $a + b > 0$. In particular, $p_r > 0$ for all r. Let $k \in \mathbb{N} \setminus \{0\}$. Then

$$\begin{aligned}\sum_{r=1}^{k} rp_r = \sum_{r=1}^{k}(ar+b)p_{r-1} &= a\sum_{r=1}^{k}(r-1)p_{r-1} + (a+b)\sum_{r=1}^{k}p_{r-1} \\ &= a\sum_{r=1}^{k-1} rp_r + (a+b)\sum_{r=0}^{k-1}p_r\,.\end{aligned}$$

This can be written as

$$kp_k = (a-1)\sum_{r=1}^{k-1} rp_r + (a+b)\sum_{r=0}^{k-1}p_r\,.$$

Suppose $a \ge 1$. Then $kp_k \ge (a+b)p_0$ and therefore $p_k \ge (a+b)p_0/k$. In particular,

$$1 - p_0 = \sum_{r=1}^{\infty} p_r \ge (a+b)p_0\sum_{r=1}^{\infty}\frac{1}{r}\,.$$

This is a contradiction. Thus $a < 1$. If we now let $p = 1 - a$ and $\alpha = a^{-1}(a+b)$ we get the negative binomial distribution.

We will use the following lemma.

Lemma 1.5. *Let $n \geq 2$. Then*

i)
$$\mathbb{E}\Big[Y_1 \;\Big|\; \sum_{i=1}^{n} Y_i = r\Big] = \frac{r}{n}\,,$$

ii)
$$p_n f_r^{*n} = \sum_{k=1}^{r-1}\Big(a + \frac{bk}{r}\Big) f_k p_{n-1} f_{r-k}^{*(n-1)}\,.$$

Proof. i) Noting that the (Y_i) are iid. we get

$$\mathbb{E}\Big[Y_1 \;\Big|\; \sum_{i=1}^{n} Y_i = r\Big] = \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}\Big[Y_j \;\Big|\; \sum_{i=1}^{n} Y_i = r\Big] = \frac{1}{n}\mathbb{E}\Big[\sum_{j=1}^{n} Y_j \;\Big|\; \sum_{i=1}^{n} Y_i = r\Big] = \frac{r}{n}\,.$$

ii) Note that $f_0^{*(n-1)} = 0$. Thus

$$\begin{aligned}
p_{n-1}\sum_{k=1}^{r-1}\Big(a + \frac{bk}{r}\Big) f_k f_{r-k}^{*(n-1)} &= p_{n-1}\sum_{k=1}^{r}\Big(a + \frac{bk}{r}\Big) f_k f_{r-k}^{*(n-1)} \\
&= p_{n-1}\sum_{k=1}^{r}\Big(a + \frac{bk}{r}\Big)\mathbb{P}\Big[Y_1 = k, \sum_{j=2}^{n} Y_j = r - k\Big] \\
&= p_{n-1}\sum_{k=1}^{r}\Big(a + \frac{bk}{r}\Big)\mathbb{P}\Big[Y_1 = k, \sum_{j=1}^{n} Y_j = r\Big] \\
&= p_{n-1}\sum_{k=1}^{r}\Big(a + \frac{bk}{r}\Big)\mathbb{P}\Big[Y_1 = k \;\Big|\; \sum_{j=1}^{n} Y_j = r\Big] f_r^{*n} \\
&= p_{n-1}\mathbb{E}\Big[a + \frac{bY_1}{r} \;\Big|\; \sum_{j=1}^{n} Y_j = r\Big] f_r^{*n} = p_{n-1}\Big(a + \frac{b}{n}\Big) f_r^{*n} \\
&= p_n f_r^{*n}\,.
\end{aligned}$$

□

We use now the second formula to find a recursive expression for g_r. We know already that $g_0 = p_0$.

$$\begin{aligned}
g_r = \sum_{n=1}^{\infty} p_n f_r^{*n} &= p_1 f_r + \sum_{n=2}^{\infty} p_n f_r^{*n} \\
&= (a+b) p_0 f_r + \sum_{n=2}^{\infty}\sum_{i=1}^{r-1}\Big(a + \frac{bi}{r}\Big) f_i p_{n-1} f_{r-i}^{*(n-1)} \\
&= (a+b) g_0 f_r + \sum_{i=1}^{r-1}\Big(a + \frac{bi}{r}\Big) f_i \sum_{n=2}^{\infty} p_{n-1} f_{r-i}^{*(n-1)} \\
&= (a+b) f_r g_0 + \sum_{i=1}^{r-1}\Big(a + \frac{bi}{r}\Big) f_i g_{r-i} = \sum_{i=1}^{r}\Big(a + \frac{bi}{r}\Big) f_i g_{r-i}\,.
\end{aligned}$$

These formulae are called Panjer recursion. Note that the convolutions have vanished. Thus a simple recursion can be used to compute g_r.

Consider now the case where the claim size distribution is not arithmetic. Choose a bandwidth $\rho > 0$. Let

$$\bar{Y}_i = \inf\{n \in \mathbb{N} : Y_i \le \rho n\}\,, \quad \bar{S} = \sum_{i=1}^{N} \bar{Y}_i$$

and

$$\underline{Y_i} = \sup\{n \in \mathbb{N} : Y_i \ge \rho n\}\,, \quad \underline{S} = \sum_{i=1}^{N} \underline{Y_i}\,.$$

It is clear that

$$\rho\underline{S} \le S \le \rho\bar{S}\,.$$

The Panjer recursion can be used to compute the distribution of $\bar{S}$ and $\underline{S}$. Note that for $\underline{S}$ there is a positive probability of a claim of size 0. Thus the original distribution of N has to be changed slightly. The corresponding new claim number distribution can be found in Section 1.7.2 for the most common cases.

1.9. Approximations to S

We have seen that the distribution of S is rather hard to determine, especially if no computers are available. Therefore approximations are called for.

1.9.1. The Normal Approximation

It is quite often the case that the number of claims is large. For a deterministic number of claims we would use a normal approximation in such a case. So we should try a normal approximation to S.

We fit a normal distribution Z such that the first two moments coincide. Thus

$$\mathbb{P}[S \le x] \approx \mathbb{P}[Z \le x] = \Phi\Bigl(\frac{x - \mathbb{E}[N]\mu}{\sqrt{\operatorname{Var}[N]\mu^2 + \mathbb{E}[N](\mu_2 - \mu^2)}}\Bigr)\,.$$

In particular for a compound Poisson distribution the formula is

$$\mathbb{P}[S \le x] \approx \Phi\Bigl(\frac{x - \lambda\mu}{\sqrt{\lambda\mu_2}}\Bigr)\,.$$

Example 1.6. Let S have a compound Poisson distribution with parameter $\lambda = 20$ and Pa(4,3) distributed claim sizes. Note that $\mu = 1$ and $\mu_2 = 3$. Figure 1.1 shows

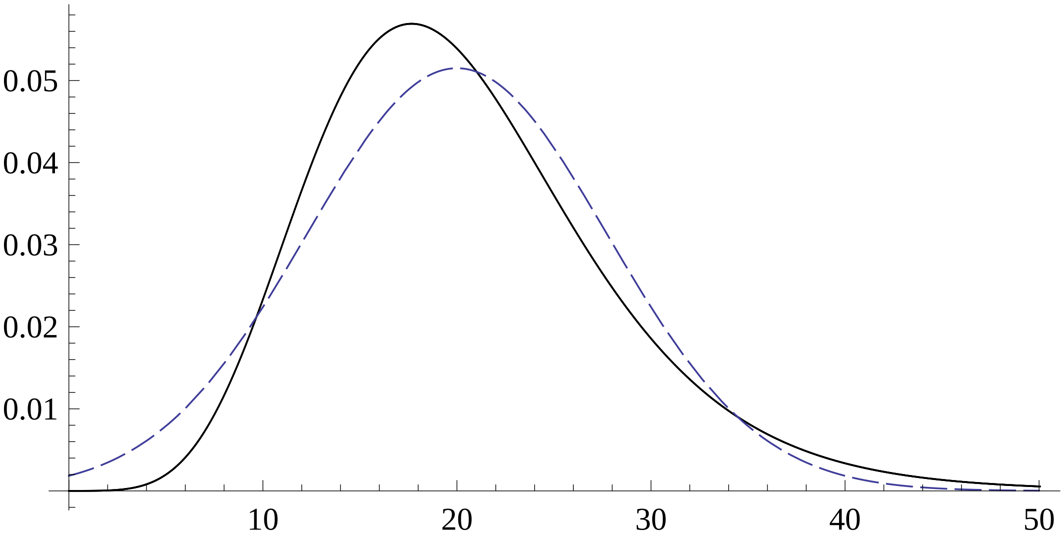


Figure 1.1: *Densities of S and its normal approximation*

the densities of S (solid line) and of its normal approximation (dashed line).

Assume that we want to find a premium p such that $\mathbb{P}[S > p] \le 0.05$. Using the normal approximation we find that

$$\mathbb{P}[S > p] \approx 1 - \Phi\Bigl(\frac{p-20}{\sqrt{60}}\Bigr) = 0.05\,.$$

Thus

$$\frac{p-20}{\sqrt{60}} = 1.6449$$

or $p = 32.7413$. If the accumulated claims only should exceed the premium in 1% of the cases we get

$$\frac{p-20}{\sqrt{60}} = 2.3263$$

or $p = 38.0194$. The exact values are 33.94 (5% quantile) and 42.99 (1% quantile). The approximation for the 5% quantile is quite good, but the approximation for the 1% quantile is not satisfactory. ■

Problem: It often turns out that loss distributions are skewed. The normal distribution is not skewed. Hence we cannot hope that the approximation fits well. We need an additional parameter to fit to the distribution.

1.9.2. The Translated Gamma Approximation

The idea of the translated gamma approximation is to approximate S by $k + Z$ where k is a real constant and Z has a gamma distribution $\Gamma(\gamma, \alpha)$. The first three moments should coincide. Note that $\mathbb{E}[((k+Z)-\mathbb{E}[k+Z])^3] = 2\gamma\alpha^{-3}$. Denote by $m = \mathbb{E}[S]$ the expected value of the risk, by σ^2 its variance and by β its coefficient of skewness. Then

$$\begin{aligned} k + \frac{\gamma}{\alpha} &= m\,, \\ \frac{\gamma}{\alpha^2} &= \sigma^2 \quad \text{and} \\ \frac{2\gamma\alpha^3}{\alpha^3\gamma^{3/2}} = \frac{2}{\sqrt{\gamma}} &= \beta\,. \end{aligned}$$

Solving this system yields

$$\gamma = \frac{4}{\beta^2}, \qquad \alpha = \frac{\sqrt{\gamma}}{\sigma} = \frac{2}{\beta\sigma}, \qquad k = m - \frac{\gamma}{\alpha} = m - \frac{2\sigma}{\beta}\,.$$

Remarks.

i) Note that $\beta > 0$ must be fulfilled.

ii) It may happen that k becomes negative. In that case the approximation has (as the normal approximation) some strictly positive mass on the negative half axis.

iii) The gamma distribution is not easy to calculate. However if $2\gamma \in \mathbb{N}$ then $2\alpha Z \sim \chi^2_{2\gamma}$. The latter distribution can be found in statistical tables. ■

Let us consider the compound Poisson case. Then

$$\gamma = \frac{4\lambda\mu_2^3}{\mu_3^2}, \qquad \alpha = \frac{2\sqrt{\lambda\mu_2^3}}{\mu_3\sqrt{\lambda\mu_2}} = \frac{2\mu_2}{\mu_3}, \qquad k = \lambda\mu - \frac{2\sqrt{\lambda\mu_2}\sqrt{\lambda\mu_2^3}}{\mu_3} = \lambda\Bigl(\mu - \frac{2\mu_2^2}{\mu_3}\Bigr)\,.$$

Example 1.6 (continued). A simple calculation shows that $\mu_3 = 27$. Thus

$$\gamma = \frac{80 \cdot 27}{729} = 2.96296, \qquad \alpha = \frac{2 \cdot 3}{27} = 0.222222, \qquad k = 20\Bigl(1 - \frac{2 \cdot 9}{27}\Bigr) = 6.66667.$$

Figure 1.2 shows the densities of S (solid line) and of its translated gamma approximation (dashed line).

Now $2\gamma = 5.92592 \approx 6$. Let $\bar{Z} \sim \chi^2_6$.

$$\mathbb{P}[S > p] \approx \mathbb{P}[Z > p - 6.66667] \approx \mathbb{P}[\bar{Z} > 0.444444(p - 6.66667)] = 0.05$$

and thus $0.444444(p - 6.66667) = 12.59$ or $p = 34.9942$. For the 1% case we obtain $0.444444(p - 6.66667) = 16.81$ or $p = 44.4892$. Because the coefficient of skewness

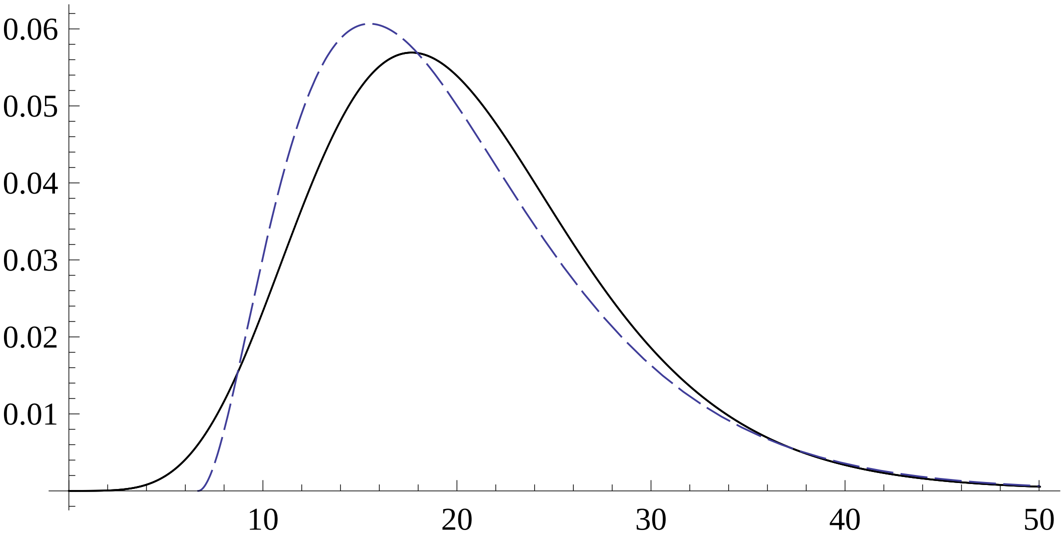


Figure 1.2: *Densities of S and its translated gamma approximation*

$\beta = 1.1619$ is not small it is not surprising that the premium becomes larger than the normal approximation premium especially if we are in the tail of the distribution. In any case the approximations of the quantiles are better than for the normal approximation. Recall that the exact values are 33.94 and 42.99 respectively. The approximation puts more weight in the tail of the distribution. In the domain where we did our computations the true value was overestimated. Note that for smaller quantiles the true value eventually will be underestimated. ■

1.9.3. The Edgeworth Approximation

Consider the random variable

$$Z = \frac{S - \mathbb{E}[S]}{\sqrt{\operatorname{Var}[S]}}\;.$$

The Taylor expansion of $\log M_Z(r)$ around $r = 0$ has the form

$$\log M_Z(r) = a_0 + a_1 r + a_2 \frac{r^2}{2} + a_3 \frac{r^3}{6} + a_4 \frac{r^4}{24} + \cdots$$

where

$$a_k = \frac{\mathrm{d}^k \log M_Z(r)}{\mathrm{d}r^k}\Big|_{r=0}\;.$$

We already know that $a_0 = 0$, $a_1 = \mathbb{E}[Z] = 0$, $a_2 = \operatorname{Var}[Z] = 1$ and $a_3 = \mathbb{E}[(Z - \mathbb{E}[Z])^3] = \beta$. For a_4 a simple calculation shows that

$$\begin{aligned} a_4 &= \mathbb{E}[Z^4] - 4\mathbb{E}[Z^3]\mathbb{E}[Z] - 3\mathbb{E}[Z^2]^2 + 12\mathbb{E}[Z^2]\mathbb{E}[Z]^2 - 6\mathbb{E}[Z]^4 = \mathbb{E}[Z^4] - 3 \\ &= \frac{\mathbb{E}[(S - \mathbb{E}[S])^4]}{\operatorname{Var}[S]^2} - 3\,. \end{aligned}$$

We truncate the Taylor series after the term involving r^4. The moment generating function of Z can be written as

$$M_Z(r) \approx \mathrm{e}^{r^2/2}\mathrm{e}^{a_3 r^3/6 + a_4 r^4/24} \approx \mathrm{e}^{r^2/2}\Bigl(1 + a_3\frac{r^3}{6} + a_4\frac{r^4}{24} + a_3^2\frac{r^6}{72}\Bigr)\,.$$

The inverse of $\exp\{r^2/2\}$ is easily found to be the normal distribution function $\Phi(x)$. For the other terms we derive

$$r^n \mathrm{e}^{r^2/2} = \int_{-\infty}^{\infty} (\mathrm{e}^{rx})^{(n)}\, \Phi'(x)\;\mathrm{d}x = (-1)^n \int_{-\infty}^{\infty} \mathrm{e}^{rx}\Phi^{(n+1)}(x)\;\mathrm{d}x\,.$$

Thus the inverse of $r^n\mathrm{e}^{r^2/2}$ is $(-1)^n$ times the n-th derivative of Φ. The approximation yields

$$\mathbb{P}[Z \le z] \approx \Phi(z) - \frac{a_3}{6}\Phi^{(3)}(z) + \frac{a_4}{24}\Phi^{(4)}(z) + \frac{a_3^2}{72}\Phi^{(6)}(z)\,.$$

It can be computed that

$$\begin{aligned} \Phi^{(3)}(z) &= \frac{1}{\sqrt{2\pi}}(z^2 - 1)\mathrm{e}^{-z^2/2}\,, \\ \Phi^{(4)}(z) &= \frac{1}{\sqrt{2\pi}}(-z^3 + 3z)\mathrm{e}^{-z^2/2}\,, \\ \Phi^{(6)}(z) &= \frac{1}{\sqrt{2\pi}}(-z^5 + 10z^3 - 15z)\mathrm{e}^{-z^2/2}\,. \end{aligned}$$

This approximation is called the **Edgeworth approximation**.

Consider now the compound Poisson case. One can see directly that $a_k = \lambda\mu_k(\lambda\mu_2)^{-k/2}$. Hence

$$\begin{aligned} \mathbb{P}[S \le x] &= \mathbb{P}\Bigl[Z \le \frac{x - \lambda\mu}{\sqrt{\lambda\mu_2}}\Bigr] \\ &\approx \Phi\Bigl(\frac{x - \lambda\mu}{\sqrt{\lambda\mu_2}}\Bigr) - \frac{\mu_3}{6\sqrt{\lambda\mu_2^3}}\Phi^{(3)}\Bigl(\frac{x - \lambda\mu}{\sqrt{\lambda\mu_2}}\Bigr) + \frac{\mu_4}{24\lambda\mu_2^2}\Phi^{(4)}\Bigl(\frac{x - \lambda\mu}{\sqrt{\lambda\mu_2}}\Bigr) \\ &\quad + \frac{\mu_3^2}{72\lambda\mu_2^3}\Phi^{(6)}\Bigl(\frac{x - \lambda\mu}{\sqrt{\lambda\mu_2}}\Bigr)\,. \end{aligned}$$

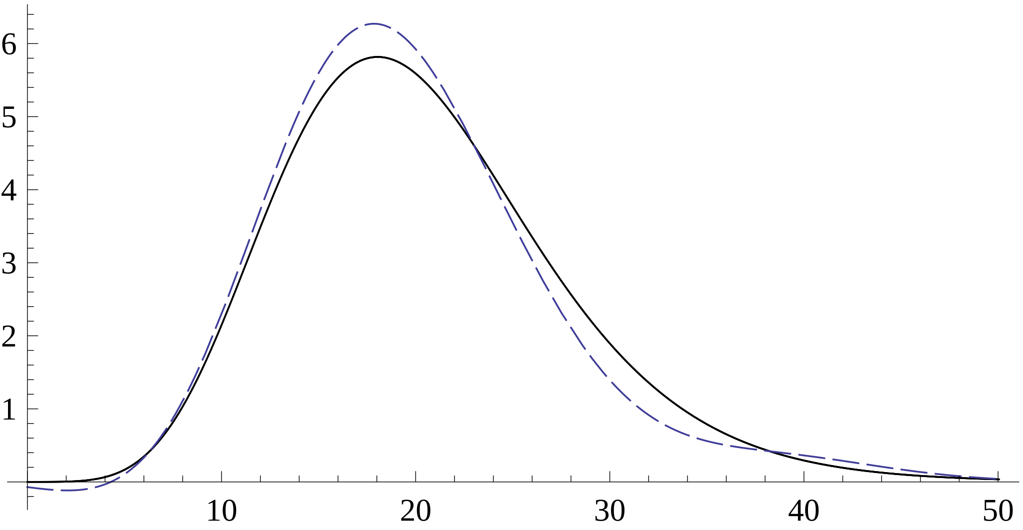


Figure 1.3: *Densities of S and its Edgeworth approximation*

Example 1.7. We cannot consider Example 1.6 because $\mu_4 = \infty$. Let us therefore consider the compound Poisson distribution with $\lambda = 20$ and Pa(5,4) distributed claim sizes. The we find $\mu = 1$, $\mu_2 = 8/3$, $\mu_3 = 16$ and $\mu_4 = 256$. Figure 1.3 gives us the densities of the exact distribution (solid line) and of its Edgeworth approximation (dashed line). The shape is approximated quite well close to the mean value 20. The tail is overestimated a little bit. Let us compare the 5% and 1% quantiles. Numerically we find 33.64 and 42.99, respectively. The exact values are 33.77 and 41.63, respectively. We see that the quantiles are approximated quite well. It is not surprising that the 5% quantile is approximated better. The tail of the approximation decays like $x^5 \mathrm{e}^{-cx^2}$ for some constant c while the exact tail decays like x^{-5} which is much slower. ■

1.9.4. The Normal Power Approximation

The idea of the normal power approximation is to approximate

$$Z = \frac{S - \mathbb{E}[S]}{\sqrt{\mathrm{Var}[S]}} \approx \tilde{Z} + \frac{\beta_S}{6}(\tilde{Z}^2 - 1) ,$$

where $\tilde{Z}$ is a standard normal variable and β_S is the coefficient of skewness of S. The mean value of the approximation is 0. But the variance becomes $1 + \beta_S^2/18$ and the skewness is $\beta_S + \beta_S^3/27$. If β_S is not too large the approximation meets the first three moments of Z quite well. Let us now consider the inequality

$$\tilde{Z} + \frac{\beta_S}{6}(\tilde{Z}^2 - 1) \leq z .$$

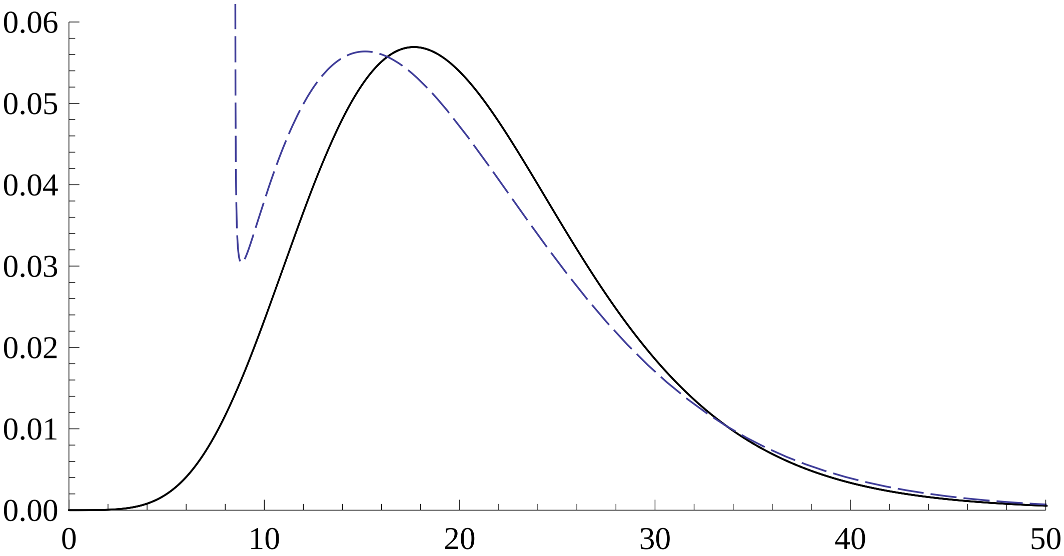


Figure 1.4: *Densities of S and its normal power approximation*

This is equivalent to

$$\Bigl(\tilde{Z} + \frac{3}{\beta_S}\Bigr)^2 \le 1 + \frac{6z}{\beta_S} + \frac{9}{\beta_S^2}\;.$$

We next consider positive $\tilde{Z} + 3/\beta_S$ only, because β_S should not be too large. Thus

$$\tilde{Z} + \frac{3}{\beta_S} \le \sqrt{1 + \frac{6z}{\beta_S} + \frac{9}{\beta_S^2}}\;.$$

We therefore get the approximation

$$\mathbb{P}[S \le x] \approx \Phi\biggl(\sqrt{1 + \frac{6(x - \mathbb{E}[S])}{\beta_S\sqrt{\operatorname{Var}[S]}} + \frac{9}{\beta_S^2}} - \frac{3}{\beta_S}\biggr)\;.$$

In the compound Poisson case the approximation reads

$$\mathbb{P}[S \le x] \approx \Phi\biggl(\sqrt{1 + \frac{6(x - \lambda\mu)\mu_2}{\mu_3} + \frac{9\lambda\mu_2^3}{\mu_3^2}} - \frac{3\sqrt{\lambda\mu_2^3}}{\mu_3}\biggr)\;.$$

Example 1.6 (continued). Plugging in the parameters we find the approximation

$$\begin{aligned}\mathbb{P}[S \le x] &\approx \Phi\biggl(\sqrt{1 + \frac{6(x-20)3}{27} + \frac{9 \cdot 20 \cdot 3^3}{27^2}} - \frac{3\sqrt{20 \cdot 3^3}}{27}\biggr)\\ &= \Phi\biggl(\sqrt{\frac{2x-17}{3}} - \sqrt{\frac{20}{3}}\biggr)\;.\end{aligned}$$

The approximation is only valid if $x \geq 8.5$. The density of S (solid line) and of the normal power approximation (dashed line) are given in Figure 1.4. The density of the approximation has a pole at $x = 8.5$. One can see that the approximation does not fit the tail as well as the translated gamma approximation. We get a clearer picture if we calculate the 5% quantile which is 35.2999. The 1% quantile is 44.6369. The approximation overestimates the true values by about 4%. As we also can see by comparision with Figure 1.2 the translated gamma approximation fits the tail much better than the normal power approximation. ■

1.10. Premium Calculation Principles

Let the annual premium for a certain risk be π and denote the losses in year i by S_i and assume that they are iid.. The company has a certain initial capital w. Then the capital of the company after year i is

$$X_i = w + \pi i - \sum_{j=1}^{i} S_j \,.$$

The stochastic process $\{X_i\}$ is called a random walk. From Lemma E.1 we conclude that

i) if $\pi < \mathbb{E}[S_1]$ then X_i converges to $-\infty$ a.s..

ii) if $\pi = \mathbb{E}[S_1]$ then a.s.

$$\varlimsup_{i\to\infty} X_i = -\varliminf_{i\to\infty} X_i = \infty \,.$$

iii) if $\pi > \mathbb{E}[S_1]$ then X_i converges to ∞ a.s. and there is a strictly positive probability that $X_i \geq 0$ for all $i \in \mathbb{N}$.

As an insurer we only have a chance to survive if $\pi > \mathbb{E}[S]$. Therefore the latter must hold for any reasonable premium calculation principle. Another argument will be given in Chapter 2.

1.10.1. The Expected Value Principle

The most popular premium calculation principle is

$$\pi = (1 + \theta)\mathbb{E}[S]$$

where θ is a strictly positive parameter called **safety loading**. The premium is very easy to calculate. One only has to estimate the quantity $\mathbb{E}[S]$. A disadvantage is that the principle is not sensible to heavy tailed distributions.

1.10.2. The Variance Principle

A way of making the premium principle more sensible to higher risks is to use the variance for determining the risk loading, i.e.

$$\pi = \mathbb{E}[S] + \alpha \operatorname{Var}[S]$$

for a strictly positive parameter α.

1.10.3. The Standard Deviation Principle

A principle similar to the variance principle is

$$\pi = \mathbb{E}[S] + \beta\sqrt{\operatorname{Var}[S]}$$

for a strictly positive parameter β. Observe that the loss can be written as

$$\pi - S = \sqrt{\operatorname{Var}[S]}\Bigl(\beta - \frac{S - \mathbb{E}[S]}{\sqrt{\operatorname{Var}[S]}}\Bigr)\,.$$

The standardized loss is therefore the risk loading parameter minus a random variable with mean 0 and variance 1.

1.10.4. The Modified Variance Principle

A problem with the variance principle is that changing the monetary unit changes the security loading. A possibility to change this is to use

$$\pi = \mathbb{E}[S] + \alpha\frac{\operatorname{Var}[S]}{\mathbb{E}[S]}$$

for some $\alpha > 0$.

1.10.5. The Principle of Zero Utility

This premium principle will be discussed in detail in Chapter 2. Denote by w the initial capital of the insurance company. The worst thing that may happen for the company is a very high accumulated sum of claims. Therefore high losses should be weighted stronger than small losses. Hence the company chooses a utility function v, which should have the following properties:

- $v(0) = 0$,

- $v(x)$ is strictly increasing,
- $v(x)$ is strictly concave.

The first property is for convenience only, the second one means that less losses are preferred and the last condition gives stronger weights for higher losses.

The premium for the next year is computed as the value π such that

$$v(w) = \mathbb{E}[v(w + p - S)] \,. \tag{1.3}$$

This means, the expected utility is the same whether the insurance contract is taken or not.

Lemma 1.8.

i) *If a solution to* (1.3) *exists, then it is unique.*

ii) *If* $\mathbb{P}[S = \mathbb{E}[S]] < 1$ *then* $\pi > \mathbb{E}[S]$.

iii) *The premium is independent of* w *for all loss distributions if and only if* $v(x) = A(1 - \mathrm{e}^{-\alpha x})$ *for some* $A > 0$ *and* $\alpha > 0$.

Proof. i) Let $\pi_1 > \pi_2$ be two solutions to (1.3). Then because $v(x)$ is strictly increasing

$$v(w) = \mathbb{E}[v(w + \pi_1 - S)] > \mathbb{E}[v(w + \pi_2 - S)] = v(w)$$

which is a contradiction.
ii) By Jensen's inequality

$$v(w) = \mathbb{E}[v(w + \pi - S)] < v(w + \pi - \mathbb{E}[S])$$

and thus because $v(x)$ is strictly increasing we get $w + \pi - \mathbb{E}[S] > w$.
iii) A simple calculation shows that the premium is independent of w if $v(x) = A(1 - \mathrm{e}^{-\alpha x})$. Assume now that the premium is independent of w. Let $\mathbb{P}[S = 1] = 1 - \mathbb{P}[S = 0] = q$ and denote by $\pi(q)$ be the corresponding premium. Then

$$qv(w + \pi(q) - 1) + (1 - q)v(w + \pi(q)) = v(w) \,. \tag{1.4}$$

Note that as a concave function $v(x)$ is differentiable almost everywhere. Taking the derivative of (1.4) (at points where it is allowed) with respect to w yields

$$qv'(w + \pi(q) - 1) + (1 - q)v'(w + \pi(q)) = v'(w) \,. \tag{1.5}$$

Varying q the function $\pi(q)$ takes all values in $(0, 1)$. Changing w and q in such a way that $w+\pi(q)$ remains constant shows that $v'(w)$ is continuous. Thus $v(w)$ must be continuously differentiable everywhere.

The derivative of (1.4) with respect to q is

$$v(w+\pi(q)-1)-v(w+\pi(q))+\pi'(q)[qv'(w+\pi(q)-1)+(1-q)v'(w+\pi(q))] = 0\,. \quad (1.6)$$

The implicit function theorem gives that $\pi(q)$ is indeed differentiable. Plugging in (1.5) into (1.6) yields

$$v(w + \pi(q) - 1) - v(w + \pi(q)) + \pi'(q)v'(w) = 0\,. \quad (1.7)$$

Note that $\pi'(q) > 0$ follows immediately. The derivative of (1.7) with respect to w is

$$v'(w + \pi(q) - 1) - v'(w + \pi(q)) + \pi'(q)v''(w) = 0$$

showing that $v(w)$ is twice continuously differentiable. From the derivative of (1.7) with respect to q it follows that

$$\pi'(q)[v'(w+\pi(q)-1)-v'(w+\pi(q))]+\pi''(q)v'(w) = -\pi'(q)^2v''(w)+\pi''(q)v'(w) = 0\,.$$

Thus $\pi''(q) \le 0$ and

$$\frac{v''(w)}{v'(w)} = \frac{\pi''(q)}{\pi'(q)^2}\,.$$

The left-hand side is independent of q and the right-hand side is independent of w. Thus it must be constant. Hence

$$\frac{v''(w)}{v'(w)} = -\alpha$$

for some $\alpha \ge 0$. $\alpha = 0$ would not lead to a strictly concave function. Solving the differential equation shows the assertion. $\square$

1.10.6. The Mean Value Principle

In contrast to the principle of zero utility it would be desirable to have a premium principle which is independent of the initial capital. The insurance company values its losses and compares it to the loss of the customer who has to pay its premium. Because high losses are less desirable they need to have a higher value. Similarly to the utility function we choose a value function v with properties

- $v(0) = 0$,
- $v(x)$ is strictly increasing and
- $v(x)$ is strictly convex.

The premium is determined by the equation

$$v(\pi) = \mathbb{E}[v(S)]$$

or equivalently

$$\pi = v^{-1}(\mathbb{E}[v(S)])\,.$$

Again, it follows from Jensen's inequality that $\pi > \mathbb{E}[S]$ provided $\mathbb{P}[S = \mathbb{E}[S]] < 1$.

1.10.7. The Exponential Principle

Let $\alpha > 0$ be a parameter. Choose the utility function $v(x) = 1 - \mathrm{e}^{-\alpha x}$. The premium can be obtained

$$1 - \mathrm{e}^{-\alpha w} = \mathbb{E}\left[1 - \mathrm{e}^{-\alpha(w+p-S)}\right] = 1 - \mathrm{e}^{-\alpha w}\mathrm{e}^{-\alpha p}\mathbb{E}\left[\mathrm{e}^{\alpha S}\right]$$

from which it follows that

$$\pi = \frac{\log(M_S(\alpha))}{\alpha}\,. \tag{1.8}$$

The same premium formula can be obtained by using the mean value principle with value function $v(x) = \mathrm{e}^{\alpha x} - 1$. Note that the principle only can be applied if the moment generating function of S exists at the point α.

How does the premium depend on the parameter α? In order to answer this question the following lemma is proved.

Lemma 1.9. *For a random variable X with moment generating function $M_X(r)$ the function $\log M_X(r)$ is convex. If X is not deterministic then $\log M_X(r)$ is strictly convex.*

Proof. Let $F(x)$ denote the distribution function of X. The second derivative of $\log M_X(r)$ is

$$(\log M_X(r))'' = \frac{M_X''(r)}{M_X(r)} - \Bigl(\frac{M_X'(r)}{M_X(r)}\Bigr)^2\,.$$

Consider

$$\frac{M_X''(r)}{M_X(r)} = \frac{\frac{\mathrm{d}^2}{\mathrm{d}r^2}\int_{-\infty}^{\infty} \mathrm{e}^{rx}\,\mathrm{d}F(x)}{M_X(r)} = \frac{\int_{-\infty}^{\infty} x^2\mathrm{e}^{rx}\,\mathrm{d}F(x)}{M_X(r)}$$

and

$$\frac{M_X'(r)}{M_X(r)} = \frac{\int_{-\infty}^{\infty} x\mathrm{e}^{rx}\,\mathrm{d}F(x)}{M_X(r)}\,.$$

Because

$$\frac{\int_{-\infty}^{\infty} \mathrm{e}^{rx}\,\mathrm{d}F(x)}{M_X(r)} = \frac{M_X(r)}{M_X(r)} = 1$$

it follows that $F_r(x) = (M_X(r))^{-1}\int_{-\infty}^{x} \mathrm{e}^{ry}\,\mathrm{d}F(y)$ is a distribution function. Let Z be a random variable with distribution function F_r. Then

$$(\log M_X(r))'' = \mathrm{Var}[Z] \ge 0\,.$$

The variance is strictly larger than 0 if Z is not deterministic. But the latter is equivalent to X is not deterministic. □

The question how the premium depends on α can now be answered.

Lemma 1.10. *Let $\pi(\alpha)$ be the premium determined by the exponential principle. Assume that $M_S''(\alpha) < \infty$. The function $\pi(\alpha)$ is strictly increasing provided S is not deterministic.*

Proof. Take the derivative of (1.8)

$$\pi'(\alpha) = \frac{M_S'(\alpha)}{\alpha M_S(\alpha)} - \frac{\log(M_S(\alpha))}{\alpha^2} = \frac{1}{\alpha}\Bigl(\frac{M_S'(\alpha)}{M_S(\alpha)} - \pi(\alpha)\Bigr)\,.$$

Furthermore

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\alpha}(\alpha^2\pi'(\alpha)) &= \frac{M_S'(\alpha)}{M_S(\alpha)} - \pi(\alpha) + \alpha\Bigl(\frac{M_S''(\alpha)}{M_S(\alpha)} - \Bigl(\frac{M_S'(\alpha)}{M_S(\alpha)}\Bigr)^2\Bigr) - \alpha\pi'(\alpha)\\ &= \alpha\Bigl(\frac{M_S''(\alpha)}{M_S(\alpha)} - \Bigl(\frac{M_S'(\alpha)}{M_S(\alpha)}\Bigr)^2\Bigr)\,.\end{aligned}$$

From Lemma 1.9 it follows that

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}(\alpha^2\pi'(\alpha)) \gtreqless 0 \text{ iff } \alpha \gtreqless 0\,.$$

Thus the function $\alpha^2\pi'(\alpha)$ has a minimum in 0. Because its value at 0 vanishes it follows that $\pi'(\alpha) > 0$. $\pi'(0) = \frac{1}{2}\mathrm{Var}[S] > 0$ can be verified directly with L'Hospital's rule. □

1.10.8. The Esscher Principle

A simple idea for pricing is to use a distribution $\tilde{F}(x)$ which is related to the distribution function of the risk $F(x)$, but does give more weight to larger losses. If exponential moments exist, a natural choice would be to consider the exponential class

$$F_\alpha(x) = (M_S(\alpha))^{-1} \int_{-\infty}^{x} \mathrm{e}^{ry} \, \mathrm{d}F_S(y) \, .$$

This yields the Esscher premium principle

$$\pi = \frac{\mathbb{E}[S\mathrm{e}^{\alpha S}]}{M_S(\alpha)} \, ,$$

where $\alpha > 0$. In order that the principle can be used we have to show that $\pi > \mathbb{E}[S]$ if $\operatorname{Var}[S] > 0$. Indeed, $\pi = (\log M_S(r))'|_{r=\alpha}$ and $\log M_S(r)$ is a strictly convex function by Lemma 1.9. Thus $(\log M_S(r))'$ is strictly increasing, in particular $\pi > (\log M_S(r))'|_{r=0} = \mathbb{E}[S]$.

1.10.9. The Proportional Hazard Transform Principle

A disadvantage of the Esscher principle is, that $M_S(\alpha)$ has to exist. We therefore look for a possibility to increase the weight for large claims, such that no exponential moments need to exist. A possibility is to change the distribution function $F_S(x)$ to $F_r(x) = 1 - (1 - F_S(x))^r$, where $r \in (0, 1)$. Then $F_r(x) < F_S(x)$ for all x such that $F_S(x) \in (0, 1)$. The proportional hazard transform principle is

$$\pi = \int_0^\infty (1 - F_S(x))^r \, \mathrm{d}x - \int_{-\infty}^0 1 - (1 - F_S(x))^r \, \mathrm{d}x \, .$$

Note that clearly $\pi > \mathbb{E}[S]$. Because $\mathrm{d}F_r(x) = r(1 - F_S(x))^{r-1} \, \mathrm{d}F_S(x)$ it follows that the distribution $F_r(x)$ and $F_S(x)$ are equivalent. A sufficient condition for π to be finite would be that $\mathbb{E}[\max\{S, 0\}^\alpha] < \infty$ for some $\alpha > 1/r$.

1.10.10. The Percentage Principle

The company wants to keep the risk that the accumulated sum of claims exceeds the premium income small. They choose a parameter ε and the premium

$$\pi = \inf\{x > 0 : \mathbb{P}[S > x] \le \varepsilon\} \, .$$

If the claims are bounded it would be possible to choose $\varepsilon = 0$. This principle is called *maximal loss principle*. It is clear that the latter cannot be of interest even though there had been villains using this principle.

1.10.11. Desirable Properties

- **Simplicity:** The premium calculation principle should be easy to calculate. This is the case for the principles 1.10.1, 1.10.2, 1.10.3 and 1.10.4.
- **Consistency:** If a deterministic constant is added to the risk then the premium should increase by the same constant. This is the case for the principles 1.10.2, 1.10.3, 1.10.5, 1.10.7, 1.10.8, 1.10.9 and 1.10.10. The property only holds for the mean value principle if it has an exponential value function (exponential principle) or a linear value function. This can be seen by a proof similar to the proof of Lemma 1.8 iii).
- **Additivity:** If a portfolio consists of two independent risks then its premium should be the sum of the individual premia. This is the case for the principles 1.10.1, 1.10.2, 1.10.7 and 1.10.8. For 1.10.5 and 1.10.6 it is only the case if they coincide with the exponential principle or if $\pi = \mathbb{E}[S]$. A proof can be found in [43, p.75].

Bibliographical Remarks

For further literature on the individual model (section 1.6) see also [43, p.51]. A reference to why reinsurance is necessary is [73]. The Panjer recursion is originally described in [64], see also [65] where also a more general algorithm can be found. Section 1.10 follows section 5 of [43].

2. Utility Theory

2.1. The Expected Utility Hypothesis

We consider an agent carrying a risk, for example a person who is looking for insurance, an insurance company looking for reinsurance, a financial engineer with a certain portfolio. Alternatively, we could consider an agent who has to decide whether to take over a risk, as for example an insurance or a reinsurance company. We consider a simple one-period model. At the beginning of the period the agent has wealth w, at the end of the period he has wealth $w - X$, where X is the loss occurred from the risk. For simplicity we assume $X \geq 0$.

The agent is now offered an **insurance contract**. An insurance contract is specified by a **compensation function** $r : \mathbb{R}_+ \to \mathbb{R}_+$. The insurer covers the amount $r(X)$, the insured covers $X - r(X)$. For the compensation of the risk taken over by the insurer, the insured has to pay the **premium** π. Let us first consider the properties $r(x)$ should have. Because the insured looks for security, the function $r(x)$ should be non-negative. Otherwise, the insured may take over a larger risk (for a possibly negative premium). On the other side, there should be no over-compensation, i.e. $0 \leq r(x) \leq x$. Otherwise, the insured could make a gain from a loss. Let $s(x) = x - r(x)$ be the **self-insurance function**, the amount covered by the insured. Sometimes $s(x)$ is also called **franchise**.

The insured has now the problem to find the right insurance for him. If he uses the policy $(\pi, r(x))$ he will have the wealth $Y = w - \pi - s(X)$ after the period. A first idea would be to consider the expected wealth $\mathbb{E}[Y]$ and to choose the policy with the largest expected wealth. Because some administration costs will be included in the premium, the optimal decision will then always be not to insure the risk. The agent wants to reduce his risk, and therefore is interested in buying insurance. He therefore would be willing to pay a premium π such that $\mathbb{E}[Y] < \mathbb{E}[w - X]$, provided the risk is reduced. A second idea would be to compare the variances. Let Y_1 and Y_2 be two wealths corresponding to different insurance treaties. We prefer the first insurance form if $\operatorname{Var}[Y_1] < \operatorname{Var}[Y_2]$. This will only make sense if $\mathbb{E}[Y_1] = \mathbb{E}[Y_2]$. The disadvantage of the above approach is, that a loss $X_1 = X - x$ will be considered as bad as a loss $X_2 = X + x$. Of course, for the insured the loss X_1 would be preferable to the loss X_2.

The above discussion inspires the following concept. The agent gives a value to each possible wealth, i.e. he has a function $u : I \to \mathbb{R}$, called **utility function**, giving a value to each possible wealth. Here I is the interval of possible wealths that could be attained by any insurance under consideration. I is assumed not to be a singleton. His criterion will now be **expected utility**, i.e. he will prefer insurance form one if $\mathbb{E}[u(Y_1)] > \mathbb{E}[u(Y_2)]$.

We now try to find properties a utility function should have. The agent will of course prefer larger wealth. Thus $u(x)$ must be *strictly increasing.* Moreover, the agent is risk averse, i.e. he will weight a loss higher for small wealth than for large wealth. Mathematically we can express this in the following way. For any $h > 0$ the function $y \mapsto u(y+h) - u(y)$ is *strictly decreasing.* Note that if $u(y)$ is a utility function fulfilling the above conditions, then also $\tilde{u}(y) = a + bu(y)$ for $a \in \mathbb{R}$ and $b > 0$ fulfils the conditions. Moreover, $\tilde{u}(y)$ and $u(y)$ will lead to the same decisions.

Let us now consider some possible utility functions.

- Quadratic utility
$$u(y) = y - \frac{y^2}{2c}, \qquad y \le c,\ c > 0. \tag{2.1a}$$
- Exponential utility
$$u(y) = -\mathrm{e}^{-cy}, \qquad y \in \mathbb{R},\ c > 0. \tag{2.1b}$$
- Logarithmic utility
$$u(y) = \log(c + y), \qquad y > -c. \tag{2.1c}$$
- HARA utility
$$u(y) = \frac{(y+c)^\alpha}{\alpha} \qquad y > -c,\ 0 < \alpha < 1, \tag{2.1d}$$
or
$$u(y) = -\frac{(c-y)^\alpha}{\alpha} \qquad y < c,\ \alpha > 1. \tag{2.1e}$$

The second condition for a utility function looks a little bit complicated. We therefore want to find a nicer equivalent condition.

Theorem 2.1. *A strictly increasing function $u(y)$ is a utility function if and only if it is strictly concave.*

Proof. Suppose $u(y)$ is a utility function. Let $x < z$. Then

$$u(x + \tfrac{1}{2}(z-x)) - u(x) > u(z) - u(x + \tfrac{1}{2}(z-x))$$

implying

$$u(\tfrac{1}{2}(x+z)) > \tfrac{1}{2}(u(x) + u(z))\,. \tag{2.2}$$

We fix now x and $z > x$. Define the function $v : [0,1] \to [0,1]$,

$$\alpha \mapsto v(\alpha) = \frac{u((1-\alpha)x + \alpha z) - u(x)}{u(z) - u(x)}\;.$$

Note that $u(y)$ concave is equivalent to $v(\alpha) \ge \alpha$ for $\alpha \in (0,1)$. By (2.2) $v(\frac{1}{2}) > \frac{1}{2}$. Let $n \in \mathbb{N}$ and let $\alpha_{n,j} = j2^{-n}$. Assume $v(\alpha_{n,j}) > \alpha_{n,j}$ for $1 \le j \le 2^n - 1$. Then clearly $v(\alpha_{n+1,2j}) = v(\alpha_{n,j}) > \alpha_{n,j} = \alpha_{n+1,2j}$. By (2.2) and $\frac{1}{2}(\alpha_{n,j} + \alpha_{n,j+1}) = \alpha_{n+1,2j+1}$,

$$v(\alpha_{n+1,2j+1}) > \tfrac{1}{2}(v(\alpha_{j,n}) + v(\alpha_{n,j+1})) > \tfrac{1}{2}(\alpha_{n,j} + \alpha_{n,j+1}) = \alpha_{n+1,2j+1}\;.$$

Let now α be arbitrary and let j_n such that $\alpha_{n,j_n} \le \alpha < \alpha_{n,j_n+1}$, $\alpha_n = \alpha_{n,j_n}$. Then

$$v(\alpha) \ge v(\alpha_n) > \alpha_n\;.$$

Because n is arbitrary we have $v(\alpha) \ge \alpha$. This implies that $u(y)$ is concave. Let now $\alpha < \frac{1}{2}$. Then

$$\begin{aligned} u((1-\alpha)x + \alpha z) &= u((1-2\alpha)x + 2\alpha\tfrac{1}{2}(x+z)) \ge (1-2\alpha)u(x) + 2\alpha u(\tfrac{1}{2}(x+z)) \\ &> (1-2\alpha)u(x) + 2\alpha\tfrac{1}{2}(u(x) + u(z)) = (1-\alpha)u(x) + \alpha u(z)\;. \end{aligned}$$

A similar argument applies if $\alpha > \frac{1}{2}$. The converse statement is left as an exercise.

□

Next we would like to have a measure, how much an agent likes the risk. Assume that $u(y)$ is twice differentiable. Intuitively, the more concave a utility function is, the more it is sensible to risk. So a risk averse agent will have a utility function with $-u''(y)$ large. But $-u''(y)$ is not a good measure for this purpose. Let $\tilde{u}(y) = a + bu(y)$. Then $-\tilde{u}''(y) = -bu''(y)$ would give a different risk aversion, even though it leads to the same decisions. In order to get rid of b define the **risk aversion function**

$$a(y) = -\frac{u''(y)}{u'(y)} = -\frac{\mathrm{d}}{\mathrm{d}y} \ln u'(y)\;.$$

Intuitively, we would assume that a wealthy person can take more risk than a poor person. This would mean that $a(y)$ should be decreasing. For our examples we have that (2.1a) and (2.1e) give increasing risk aversion, (2.1c) and (2.1d) give decreasing risk aversion, and (2.1b) gives constant risk aversion.

2.2. The Zero Utility Premium

Let us consider the problem that an agent has to decide whether to insure his risk or not. Let $(\pi, r(x))$ be an insurance treaty and assume that it is a non-trivial insurance, i.e. $\text{Var}[r(X)] > 0$. If the agent takes the insurance he will have the final wealth $w - \pi - s(X)$, if he does not take the insurance his final wealth will be $w - X$. Considering his expected utility, the agent will take the insurance if and only if

$$\mathbb{E}[u(w - \pi - s(X))] \ge \mathbb{E}[u(w - X)]\,.$$

Because insurance basically reduces the risk we assume here that insurance is taken if the expected utility is the same. The left hand side of the above inequality is a continuous and strictly decreasing function of π. The largest value that can be attained is for $\pi = 0$, $\mathbb{E}[u(w - s(X))] > \mathbb{E}[u(w - X)]$, the lowest possible value is difficult to obtain, because it depends on the choice of I. Let us assume that the lower bound is smaller or equal to $\mathbb{E}[u(w - X)]$. Then there exists a unique value $\pi_r(w)$ such that

$$\mathbb{E}[u(w - \pi_r - s(X))] = \mathbb{E}[u(w - X)]\,.$$

The premium π_r is the largest premium the agent is willing to pay for the insurance treaty $r(x)$. We call π_r the **zero utility premium**. In many situations the zero utility premium π_r will be larger then the "fair premium" $\mathbb{E}[r(X)]$.

Theorem 2.2. *If both* $r(x)$ *and* $s(x)$ *are increasing and* $\text{Var}[r(X)] > 0$ *then* $\pi_r(w) > \mathbb{E}[r(X)]$.

Proof. From the definition of a strictly concave function it follows that the function $v(y)$, $x \mapsto v(x) = u(w - x)$ is a strictly concave function. Define the functions $g_1(x) = s(x) + \mathbb{E}[r(X)]$ and $g_2(x) = x$. These two functions are increasing. The function $g_2(x) - g_1(x) = r(x) - \mathbb{E}[r(X)]$ is increasing. Thus there exists x_0 such that $g_1(x) \le g_2(x)$ for $x < x_0$ and $g_1(x) \ge g_2(x)$ for $x > x_0$. Note that $\mathbb{E}[g_2(x) - g_1(x)] = 0$. Thus the conditions of Corollary G.7 are fulfilled and

$$\begin{aligned}\mathbb{E}[u(w - \mathbb{E}[r(X)] - s(X))] &= \mathbb{E}[v(g_1(X))] > \mathbb{E}[v(g_2(X))] \\ &= \mathbb{E}[u(w - X)] = \mathbb{E}[u(w - \pi_r - s(X))]\,.\end{aligned}$$

Because $\pi \mapsto \mathbb{E}[u(w - \pi - s(X))]$ is a strictly decreasing function it follows that $\pi_r > \mathbb{E}[r(X)]$. □

Assume now that u is twice continuously differentiable. Then we find the following result.

Theorem 2.3. *Let $r(x) = x$. If the risk aversion function is decreasing (increasing), then the zero utility premium $\pi(w)$ is decreasing (increasing).*

Proof. Consider the function $v(y) = u'(u^{-1}(y))$. Taking the derivative yields

$$v'(y) = \frac{u''(u^{-1}(y))}{u'(u^{-1}(y))} = -a(u^{-1}(y))\,.$$

Because $y \mapsto u^{-1}(y)$ is increasing, this gives that $v'(y)$ is increasing (decreasing). Thus $v(y)$ is convex (concave).

We have

$$\pi(w) = w - u^{-1}(\mathbb{E}[u(w-X)]) \quad \text{and} \quad \pi'(w) = 1 - \frac{\mathbb{E}[u'(w-X)]}{u'(u^{-1}(\mathbb{E}[u(w-X)]))}\,.$$

Let now $a(y)$ be decreasing (increasing). By Jensen's inequality we have

$$u'(u^{-1}(\mathbb{E}[u(w-X)])) \le (\ge)\, \mathbb{E}[u'(u^{-1}(u(w-X)))] = \mathbb{E}[u'(w-X)]$$

and therefore $\pi'(w) \le (\ge)\, 0$. □

It is not always desirable that the decision depends on the initial wealth. We have the following result.

Lemma 2.4. *Let $u(x)$ be a utility function. Then the following are equivalent:*

i) *$u(x)$ is exponential, i.e. $u(x) = -A\mathrm{e}^{-cx} + B$ for some $A, c > 0$, $B \in \mathbb{R}$.*

ii) *For all losses X and all compensation functions $r(x)$, the zero utility premium π_r does not depend on the initial wealth w.*

iii) *For all losses X and for full reinsurance $r(x) = x$, the zero utility premium π_r does not depend on the initial wealth w.*

Proof. This follows directly from Lemma 1.8. □

2.3. Optimal Insurance

Let us now consider an agent carrying several risks $X_1, X_2, \ldots, X_n$. The total risk is then $X = X_1 + \cdots + X_n$. We consider here a compensation function $r(X_1, \ldots, X_n)$ and a self-insurance function $s(X_1, \ldots, X_n) = X - r(X_1, \ldots, X_n)$. We say an insurance treaty is **global** if $r(X_1, \ldots, X_n)$ depends on X only, or equivalently, $s(X_1, \ldots, X_n)$ depends on X only. Otherwise, we call an insurance treaty **local**. We next prove that, in some sense, the global treaties are optimal for the agent.

Theorem 2.5. (Pesonen-Ohlin) *If the premium depends on the pure net premium only, then to every local insurance treaty there exists a global treaty with the same premium but a higher expected utility.*

Proof. Let $(\pi, r(x_1, \ldots, x_n))$ be a local treaty. Consider the compensation function $R(x) = \mathbb{E}[r(X_1, \ldots, X_n) \mid X = x]$. If the premium depends on the pure net premium only, then clearly $R(X)$ is sold for the same premium, because

$$\mathbb{E}[R(X)] = \mathbb{E}[\mathbb{E}[r(X_1, \ldots, X_n) \mid X]] = \mathbb{E}[r(X_1, \ldots, X_n)]\,.$$

By Jensen's inequality

$$\begin{aligned}\mathbb{E}[u(w - \pi - s(X_1, \ldots, X_n))] &= \mathbb{E}[\mathbb{E}[u(w - \pi - s(X_1, \ldots, X_n)) \mid X]] \\ &< \mathbb{E}[u(\mathbb{E}[w - \pi - s(X_1, \ldots, X_n) \mid X])] = \mathbb{E}[u(w - \pi - (X - R(X)))]\,.\end{aligned}$$

The strict inequality follows from the fact that $s(X_1, \ldots, X_n) \neq S(X)$, because otherwise the insurance treaty would be global, and from the strict concavity. □

Remark. Note that the result does not say that the constructed global treaty is a good treaty. It only says, that for the investigation of good treaties we can restrict to global treaties. ■

Let us now find the optimal insurance treaty for the insured.

Theorem 2.6. (Arrow-Ohlin) *Assume the premium depends on the pure net premium only and let $(\pi, r(x))$ be an insurance treaty with $\mathbb{E}[r(X)] > 0$. Then there exists a unique deductible b such that the* **excess of loss insurance** *$r_b(x) = (x - b)^+$ has the same premium. Moreover, if $\mathbb{P}[r(X) \neq r_b(X)] > 0$ then the treaty $(\pi, r_b(x))$ yields a higher utility.*

Proof. The function $(x - b)^+$ is continuous and decreasing in b. We have $\mathbb{E}[(X - 0)^+] \geq \mathbb{E}[r(X)] > 0 = \lim_{b \to \infty} \mathbb{E}[(X - b)^+]$. Thus there exists a unique b such that $\mathbb{E}[(X - b)^+] = \mathbb{E}[r(X)]$. Thus $r_b(x)$ yields the same premium. Assume now $\mathbb{P}[r(X) \neq r_b(X)] > 0$. Let $s_b(x) = x - r_b(x)$. For $y < b$ we have

$$\mathbb{P}[s_b(X) \leq y] = \mathbb{P}[X \leq y] \leq \mathbb{P}[s(X) \leq y]\,,$$

and for $y > b$

$$\mathbb{P}[s_b(X) \leq y] = 1 \geq \mathbb{P}[s(X) \leq y]\,.$$

Note that $v(x) = u(w - \pi - x)$ is a strictly concave function. Thus Ohlin's lemma yields

$$\mathbb{E}[u(w - \pi - s_b(X))] = \mathbb{E}[v(s_b(X))] > \mathbb{E}[v(s(X))] = \mathbb{E}[u(w - \pi - s(X))]\,.$$

□

2.4. The Position of the Insurer

We now consider the situation from the point of view of an insurer. The discussion how to take decisions does also apply to the insurer. Assume therefore that the insurer has utility function $\bar{u}(x)$. Denote the initial wealth of the insurer by $\overline{w}$. For an insurance treaty $(\pi, r(x))$ the expected utility of the insurer becomes

$$\mathbb{E}[\bar{u}(\overline{w} + \pi - r(X))]\,.$$

The **zero utility premium** $\bar{\pi}_r(\overline{w})$ of the insurer is the the solution to

$$\mathbb{E}[\bar{u}(\overline{w} + \bar{\pi}_r(\overline{w}) - r(X))] = \bar{u}(\overline{w})\,.$$

The insurer will take over the risk if and only if $\pi \ge \bar{\pi}_r(\overline{w})$. Note that the insurer is interested in the contract if he gets the same expected utility. This is because without taking over the risk he will not have a chance to make some profit. By Jensen's inequality we have

$$\bar{u}(\overline{w}) = \mathbb{E}[\bar{u}(\overline{w} + \bar{\pi}_r(\overline{w}) - r(X))] < \bar{u}(\overline{w} + \bar{\pi}_r(\overline{w}) - \mathbb{E}[r(X)])$$

which yields $\bar{\pi}_r(\overline{w}) - \mathbb{E}[r(X)] > 0$. A insurance contract can thus be signed by both parties if and only if $\bar{\pi}_r(\overline{w}) \le \pi \le \pi_r(w)$. It follows as in Theorem 2.5 (replacing $s(x_1, \ldots, x_n)$ by $r(x_1, \ldots, x_n)$) that the insurer prefers global treaties to local ones. From Theorem 2.6 (replacing $s(x)$ by $r(x)$) it follows that a first risk deductible is optimal for the insurer. Thus the interests of insurer and insured contradict.

Let us now consider insurance forms that take care of the interests of the insured. The insured would like that the larger the loss the larger should be the fraction covered by the insurance company. We call a compensation function $r(x)$ a **Vajda compensation function** if $r(x)/x$ is an increasing function of x. Because $r_b(x) = (x - b)^+$ is a Vajda compensation function it is clear how the insured would choose $r(x)$. Let us now consider what is optimal for the insurer.

Theorem 2.7. *Suppose the premium depends on the pure net premium only. For any Vajda insurance treaty $(\pi, r(x))$ with $\mathbb{E}[r(X)] > 0$ there exists a* **proportional insurance treaty** $r_k(x) = (1-k)x$ *with the same premium. Moreover, if* $\mathbb{P}[r(X) \neq r_k(X)] > 0$ *then* $(\pi, r_k(x))$ *yields a higher utility for the insurer.*

Proof. The function $[0,1] \to \mathbb{R} : k \mapsto \mathbb{E}[(1-k)X]$ is decreasing and has its image in the interval $[0, \mathbb{E}[X]]$. Because $0 < \mathbb{E}[r(X)] \le \mathbb{E}[X]$ there is a unique k such that $\mathbb{E}[r_k(X)] = \mathbb{E}[r(X)]$. These two treaties have then the same premium. Suppose now that $\mathbb{P}[r(X) \neq r_k(X)] > 0$. Note that $r(x)$ and $r_k(x)$ are increasing. The difference

$$\frac{r(x) - r_k(x)}{x} = \Bigl(\frac{r(x)}{x} - (1-k)\Bigr)$$

is an increasing function in x. Thus there is x_0 such that $r(x) \le r_k(x)$ for $x < x_0$ and $r(x) \ge r_k(x)$ for $x > x_0$. The function $v(x) = \bar{u}(\overline{w} + \pi - x)$ is strictly concave. By Corollary G.7 we have

$$\mathbb{E}[\bar{u}(\overline{w} + \pi - r_k(X))] = \mathbb{E}[v(r_k(X))] > \mathbb{E}[v(r(X))] = \mathbb{E}[\bar{u}(\overline{w} + \pi - r_k(X))]\,.$$

This proves the result. □

As from the point of view of the insured it turns out that the zero utility premium is only independent of the initial wealth if the utility is exponential.

Proposition 2.8. *Let $\bar{u}(x)$ be a utility function. Then the following are equivalent:*

i) *$\bar{u}(x)$ is exponential, i.e. $\bar{u}(x) = -A\mathrm{e}^{-cx} + B$ for some $A, c > 0$, $B \in \mathbb{R}$.*

ii) *For all losses X and all compensation functions $r(x)$, the zero utility premium $\overline{\pi}_r$ does not depend on the initial wealth $\overline{w}$.*

Proof. The result follows analogously to the proof of Proposition 2.4. □

2.5. Pareto-Optimal Risk Exchanges

Consider now a market with n agents, i.e. insured, insurance companies, reinsurers. In the market there are risks $\boldsymbol{X} = (X_1, \dots, X_n)^\top$. Some of the risks may be zero, if the agent does not initially carry a risk. We allow here also for financial risks (investment). If X_i is positive, we consider it as a loss, if it is negative, we consider it as a gain. Making contracts, the agents redistribute the risk $\boldsymbol{X}$. Formally, a **risk exchange** is a function

$$\boldsymbol{f} = (f_1, \ldots, f_n)^\top : \mathbb{R}^n \to \mathbb{R}^n\,,$$

such that

$$\sum_{i=1}^n f_i(\boldsymbol{X}) = \sum_{i=1}^n X_i\,. \tag{2.3}$$

The latter condition assures that the whole risk is covered. A risk exchange means that agent i covers $f_i(\boldsymbol{X})$. This is a generalisation of the insurance market we had considered before.

Now each of the agents has an initial wealth w_i and a utility function $u_i(x)$. Under the risk exchange $\boldsymbol{f}$ the expected terminal utility of agent i becomes

$$V_i(\boldsymbol{f}) = \mathbb{E}[u_i(w_i - f_i(\boldsymbol{X}))]\,.$$

As we saw before, the interests of the agents will contradict. Thus it will in general not be possible to find $\boldsymbol{f}$ such that each agent gets an optimal utility. However, an agent may agree to a risk exchange $\boldsymbol{f}$ if $V_i(\boldsymbol{f}) \ge \mathbb{E}[u_i(w_i - X_i)]$. Hence, which risk exchange will be chosen depends upon negotiations. However, the agents will agree not to discuss certain risk exchanges.

Definition 2.9. *A risk exchange* $\boldsymbol{f}$ *is called* **Pareto-optimal** *if for any risk exchange* $\tilde{\boldsymbol{f}}$ *with*

$$V_i(\boldsymbol{f}) \le V_i(\tilde{\boldsymbol{f}})\ \forall i \quad \textit{it follows that} \quad V_i(\boldsymbol{f}) = V_i(\tilde{\boldsymbol{f}})\ \forall i\,.$$

For a risk exchange $\boldsymbol{f}$ that is not Pareto optimal we would have a risk exchange $\tilde{\boldsymbol{f}}$ that is at least as good for all agents, but better for at least one of them. Of course, it is not excluded that $V_i(\boldsymbol{f}) < V_i(\tilde{\boldsymbol{f}})$ for some i. If the latter is the case then there must be j such that $V_j(\boldsymbol{f}) > V_j(\tilde{\boldsymbol{f}})$.

For the negotiations, the agents can in principle restrict to Pareto-optimal risk exchanges. Of course, in reality, the agents do not know the utility functions of the other agents. So the concept of Pareto optimality is more a theoretical way to consider the market.

It seems quite difficult to decide whether a certain risk exchange is Pareto-optimal or not. It turns out that a simple criteria does the job.

Theorem 2.10. (Borch/du Mouchel) *Assume that the utility functions* $u_i(x)$ *are differentiable. A risk exchange* $\boldsymbol{f}$ *is Pareto-optimal if and only if there exist numbers* θ_i *such that (almost surely)*

$$u_i'(w_i - f_i(\boldsymbol{X})) = \theta_i u_1'(w_1 - f_1(\boldsymbol{X}))\,. \tag{2.4}$$

Remark. Note that necessarily $\theta_1 = 1$ and $\theta_i > 0$ because $u_i'(x) > 0$. Note also that we could make the comparisons with agent j instead of agent 1. Then we just need to divide the θ_i by θ_j, for example θ_1 would be replaced by θ_j^{-1}. ■

Proof. Assume that θ_i exists such that (2.4) is fulfilled. Let $\tilde{\boldsymbol{f}}$ be a risk exchange such that $V_i(\tilde{\boldsymbol{f}}) \ge V_i(\boldsymbol{f})$ for all i. By concavity

$$u_i(w_i - \tilde{f}_i(\boldsymbol{X})) \le u_i(w_i - f_i(\boldsymbol{X})) + u_i'(w_i - f_i(\boldsymbol{X}))(f_i(\boldsymbol{X}) - \tilde{f}_i(\boldsymbol{X}))\,.$$

This yields

$$\frac{u_i(w_i - \tilde{f}_i(\boldsymbol{X})) - u_i(w_i - f_i(\boldsymbol{X}))}{\theta_i} \le u_1'(w_1 - f_1(\boldsymbol{X}))(f_i(\boldsymbol{X}) - \tilde{f}_i(\boldsymbol{X}))\,.$$

Using $\sum_{i=1}^n f_i(\boldsymbol{X}) = \sum_{i=1}^n X_i = \sum_{i=1}^n \tilde{f}_i(\boldsymbol{X})$ we find by summing over all i,

$$\sum_{i=1}^n \frac{u_i(w_i - \tilde{f}_i(\boldsymbol{X})) - u_i(w_i - f_i(\boldsymbol{X}))}{\theta_i} \le 0\,.$$

Taking expectations yields

$$\sum_{i=1}^n \frac{V_i(\tilde{\boldsymbol{f}}) - V_i(\boldsymbol{f})}{\theta_i} \le 0\,.$$

From the assumption $V_i(\tilde{\boldsymbol{f}}) \ge V_i(\boldsymbol{f})$ it follows that $V_i(\tilde{\boldsymbol{f}}) = V_i(\boldsymbol{f})$. Thus $\boldsymbol{f}$ is Pareto-optimal.

Assume now that (2.4) does not hold. By renumbering, we can assume that there is no constant θ_2 such that (2.4) holds for $i = 2$. Let

$$\theta = \frac{\mathbb{E}[u_1'(w_1 - f_1(\boldsymbol{X}))u_2'(w_2 - f_2(\boldsymbol{X}))]}{\mathbb{E}[(u_1'(w_1 - f_1(\boldsymbol{X})))^2]}$$

and $W = u_2'(w_2 - f_2(\boldsymbol{X})) - \theta u_1'(w_1 - f_1(\boldsymbol{X}))$. We have then $\mathbb{E}[W u_1'(w_1 - f_1(\boldsymbol{X}))] = 0$ and by the assumption $\mathbb{E}[W^2] > 0$. Let now $\varepsilon > 0$, $\delta = \frac{1}{2}\mathbb{E}[W^2]/\mathbb{E}[u_2'(w_2 -$

$f_2(\boldsymbol{X}))] > 0$ and $\tilde{f}_1(\boldsymbol{X}) = f_1(\boldsymbol{X}) - (\delta - W)\varepsilon$, $\tilde{f}_2(\boldsymbol{X}) = f_2(\boldsymbol{X}) + (\delta - W)\varepsilon$, $\tilde{f}_i(\boldsymbol{X}) = f_i(\boldsymbol{X})$ for all $i > 2$. Then $\tilde{\boldsymbol{f}}$ is a risk exchange. Consider now

$$\begin{aligned}\frac{V_1(\tilde{\boldsymbol{f}}) - V_1(\boldsymbol{f})}{\varepsilon} &- \mathbb{E}[u_1'(w_1 - f_1(\boldsymbol{X}))]\delta \\ &= \mathbb{E}\Bigl[(\delta - W)\Bigl(\frac{u_1(w_1 - \tilde{f}_1(\boldsymbol{X})) - u_1(w_1 - f_1(\boldsymbol{X}))}{(\delta - W)\varepsilon} - u_1'(w_1 - f_1(\boldsymbol{X}))\Bigr)\Bigr]\,,\end{aligned}$$

where we used $\mathbb{E}[Wu_1'(w_1 - f_1(\boldsymbol{X}))] = 0$. The right-hand side tends to zero as ε tends to zero. Thus $V_1(\tilde{\boldsymbol{f}}) > V_1(\boldsymbol{f})$ for ε small enough. We also have

$$\begin{aligned}\frac{V_2(\tilde{\boldsymbol{f}}) - V_2(\boldsymbol{f})}{\varepsilon} &- \mathbb{E}[u_2'(w_2 - f_2(\boldsymbol{X}))(W - \delta)] \\ &= \mathbb{E}\Bigl[(W - \delta)\Bigl(\frac{u_2(w_2 - \tilde{f}_2(\boldsymbol{X})) - u_2(w_2 - f_2(\boldsymbol{X}))}{(W - \delta)\varepsilon} - u_2'(w_2 - f_2(\boldsymbol{X}))\Bigr)\Bigr]\,,\end{aligned}$$

Also here the right-hand side tends to zero. We have

$$\begin{aligned}\mathbb{E}[u_2'(w_2 - f_2(\boldsymbol{X}))(W - \delta)] &= \mathbb{E}[(W + \theta u_1'(w_1 - f_1(\boldsymbol{X})))W] - \delta\mathbb{E}[u_2'(w_2 - f_2(\boldsymbol{X}))] \\ &= \mathbb{E}[W^2] - \delta\mathbb{E}[u_2'(w_2 - f_2(\boldsymbol{X}))] = \tfrac{1}{2}\mathbb{E}[W^2] > 0\,.\end{aligned}$$

Thus also $V_2(\tilde{\boldsymbol{f}}) > V_2(\boldsymbol{f})$ for ε small enough. This shows that $\boldsymbol{f}$ is not Pareto-optimal. ☐

In order to find the Pareto-optimal risk exchanges one has to solve equation (2.4) subject to the constraint (2.3).

If a risk exchange is chosen the quantity $f_i(0, \dots, 0)$ is the amount agent i has to pay (obtains if negative) if no losses occur. Thus $f_i(0, \dots, 0)$ must be interpreted as a premium.

Note that for Pareto-optimality only the support of the distribution of $\boldsymbol{X}$, not the distribution itself, does has an influence. Hence the Pareto-optimal solution can be found without investigating the risk distribution, or the dependencies. This also shows that any Pareto optimal solution will depend on $\sum_{i=1}^n X_i$ only.

We now solve the problem explicitly. Because $u_i'(x)$ is strictly decreasing, it is invertible and its inverse $(u_i')^{-1}(y)$ is strictly decreasing. Moreover, if $u_i'(x)$ is continuous (as it is under the conditions of Theorem 2.10), then $(u_i')^{-1}(y)$ is continuous too. Choose $\theta_i > 0$, $\theta_1 = 1$. Then (2.4) yields

$$w_i - f_i(\boldsymbol{X}) = (u_i')^{-1}(\theta_i u_1'(w_1 - f_1(\boldsymbol{X})))\,.$$

Summing over i gives by use of (2.3)

$$\sum_{i=1}^n w_i - \sum_{i=1}^n X_i = \sum_{i=1}^n (u_i')^{-1}(\theta_i u_1'(w_1 - f_1(\boldsymbol{X})))\,.$$

The function

$$g(y) = \sum_{i=1}^n (u_i')^{-1}(\theta_i u_1'(y))$$

is then strictly increasing and therefore invertible. If all the $u_i'(x)$ are continuous, then also $g(y)$ is continuous and therefore $g^{-1}(x)$ is continuous too. This yields

$$f_1(\boldsymbol{X}) = w_1 - g^{-1}\Bigl(\sum_{i=1}^n w_i - \sum_{i=1}^n X_i\Bigr)\,.$$

For i arbitrary we then obtain

$$f_i(\boldsymbol{X}) = w_i - (u_i')^{-1}\Bigl(\theta_i u_1'\Bigl(g^{-1}\Bigl(\sum_{i=1}^n w_i - \sum_{i=1}^n X_i\Bigr)\Bigr)\Bigr)\,.$$

Thus we have found the following result.

Theorem 2.11. *A Pareto-optimal risk exchange is a pool, that is each individual agent contributes a share that depends on the total loss of the group only. Moreover, each agent must cover a genuine part of any increase of the total loss of the group. If all utility functions are differentiable, then the individual shares are continuous functions of the total loss.* ☐

Remark. Consider the situation insured-insurer, i.e. $n = 2$ and a risk of the form $(X, 0)^\top$, where $X \ge 0$. For certain choices of θ_2 it is possible that $f_1(0, 0) < 0$, i.e. the insurer has to pay a premium to take over the risk. This shows that not all possible choices of θ_i will be realistic. In fact, the risk exchange has to be chosen such that the expected utility of each agent is increased, i.e. $\mathbb{E}[u_i(w_i - f_i(\boldsymbol{X}))] \ge \mathbb{E}[u_i(w_i - X_i)]$ in order that agent i will be willing to participate. ■

Bibliographical Remarks

The results presented here can be found in [78].

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1			1							1							1	1		
2							1		1	1							1			
3			1				1		1				1							
4																	1			
5									1	1										
6						1			1		1									
7									1					1			1			
8											1									
9						1				1		1								
10									1		1						1		1	
Σ	0	0	2	0	0	2	2	0	6	4	3	1	1	1	0	0	5	1	1	0

Table 3.1: *Years with accidents*

3. Credibility Theory

3.1. Introduction

We start with an example. A company has insured twenty drivers in a small village during the past ten years. Table 3.1 shows, indicated by a 1, which driver had at least one accident in the corresponding year. The last row shows the number of years with at least one claim. Some of the drivers had no accidents. They will complain about their premia because they have to pay for the drivers with 4 and more years with accidents. They think that this is unfair. But could it not be possible that the probability of at least one accident in a year is the same for all the drivers? Had some of the drivers just been unlucky? We can test this by using a χ^2-test. Let $\hat{\mu}_i$ be the mean number of years with an accident for driver i and

$$\bar{\mu} = \frac{1}{20} \sum_{i=1}^{20} \hat{\mu}_i$$

be the mean number of years with an accident for a typical driver. Then

$$Z = \frac{10 \sum_{i=1}^{20} (\hat{\mu}_i - \bar{\mu})^2}{\bar{\mu}(1 - \bar{\mu})} = 49.1631\,.$$

Under the hypothesis that $\mathbb{E}[\mu_i]$ is the same for all drivers the statistic Z is approximately χ^2_{19}-distributed. But

$$\mathbb{P}[\chi^2_{19} > 49] \approx 0.0002\,.$$

Thus it is very unlikely that each driver has the same probability of having at least one accident in a year.

For the insurance company it is preferable to attract 'good' risks and to get rid of the 'poor' risks. Thus they would like to charge premia according to their experience with a particular customer. This procedure is called **experience rating** or **credibility**. Let us denote by Y_{ij} the losses of the i-th risk in year j, where $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$. Denote the mean losses of the i-th risk by

$$\bar{Y}_i := \frac{1}{m} \sum_{j=1}^{m} Y_{ij} \,.$$

We make the following assumptions:

i) There exists a parameter Θ_i belonging to the risk i. The parameters $(\Theta_i : 1 \le i \le n)$ are iid..

ii) The vectors $((Y_{i1}, Y_{i2}, \ldots, Y_{im}, \Theta_i) : 1 \le i \le n)$ are iid..

iii) For fixed i, given Θ_i the random variables $Y_{i1}, Y_{i2}, \ldots Y_{im}$ are conditionally iid..

For instance Θ_i can be the parameters of a family of distributions, or more generally, the distribution function of a claim of the i-th risk.

Denote by $m(\vartheta) = \mathbb{E}[Y_{ij} \mid \Theta_i = \vartheta]$ the conditional expected mean of an aggregate loss given $\Theta_i = \vartheta$, by $\mu = \mathbb{E}[m(\Theta_i)]$ the overall mean of the aggregate losses and by $s^2(\vartheta) = \operatorname{Var}[Y_{ij} \mid \Theta_i = \vartheta]$ the conditional variance of Y_{ij} given $\Theta_i = \vartheta$. Now instead of using the expected mean μ of a claim of a typical risk for the premium calculations one should use $m(\Theta_i)$ instead. Unfortunately we do not know Θ_i. But $\bar{Y}_i$ is an estimator for $m(\Theta_i)$. But it is not a good idea either to use $\bar{Y}_i$ for the premium calculation. Assume that somebody who is insured for the first year has a big claim. His premium for the next year would become larger than his claim. He would not be able to take insurance furthermore. Thus we have to find other methods to estimate $m(\Theta_i)$.

3.2. Bayesian Credibility

Suppose we know the distribution of Θ_i and the conditional distribution of Y_{ij} given Θ_i. We now want to find the best estimate M_0 of $m(\Theta_i)$ in the sense that

$$\mathbb{E}[(M - m(\Theta_i))^2] \ge \mathbb{E}[(M_0 - m(\Theta_i))^2]$$

for all measurable functions $M = M(Y_{i1}, Y_{i2}, \ldots, Y_{im})$. Let $M_0' = \mathbb{E}[m(\Theta_i) \mid Y_{i1}, Y_{i2}, \ldots, Y_{im}]$. Then

$$\begin{aligned}\mathbb{E}[(M - m(\Theta_i))^2] &= \mathbb{E}[\{(M - M_0') + (M_0' - m(\Theta_i))\}^2] \\ &= \mathbb{E}[(M - M_0')^2] + \mathbb{E}[(M_0' - m(\Theta_i))^2] + 2\mathbb{E}[(M - M_0')(M_0' - m(\Theta_i))] \;.\end{aligned}$$

We have then

$$\begin{aligned}\mathbb{E}[(M - M_0')(M_0' - m(\Theta_i))] &= \mathbb{E}[\mathbb{E}[(M - M_0')(M_0' - m(\Theta_i)) \mid Y_{i1}, Y_{i2}, \ldots, Y_{im}]] \\ &= \mathbb{E}[(M - M_0')(M_0' - \mathbb{E}[m(\Theta_i) \mid Y_{i1}, Y_{i2}, \ldots, Y_{im}])] \\ &= 0 \;.\end{aligned}$$

Thus

$$\mathbb{E}[(M - m(\Theta_i))^2] = \mathbb{E}[(M - M_0')^2] + \mathbb{E}[(M_0' - m(\Theta_i))^2] \;.$$

The second term is independent of the choice of M. That means that the best estimate is $M_0 = M_0'$. We call M_0 the Bayesian credibility estimator. In the sequel we will consider the two most often used Bayesian credibility models.

3.2.1. The Poisson-Gamma Model

We assume that Y_{ij} has a compound Poisson distribution with parameter λ_i and an individual claim size distribution which is the same for all the risks. It is therefore sufficient for the company to estimate λ_i. In the sequel we assume that all claim sizes are equal to 1, i.e. $Y_{ij} = N_{ij}$. Motivated by the compound negative binomial model we assume that $\lambda_i \sim \Gamma(\gamma, \alpha)$. Let

$$\bar{N}_i = \frac{1}{m} \sum_{j=1}^{m} N_{ij}$$

denote the mean number of claims of the i-th risk. We know the overall mean $\mathbb{E}[\lambda_i] = \gamma\alpha^{-1}$. For the best estimator for λ_i we get

$$\mathbb{E}[\lambda_i \mid N_{i1}, N_{i2}, \dots, N_{im}] = \frac{\displaystyle\int_0^\infty \ell \prod_{j=1}^m \Bigl(\frac{\ell^{N_{ij}}}{N_{ij}!}\mathrm{e}^{-\ell}\Bigr)\frac{\alpha^\gamma}{\Gamma(\gamma)}\ell^{\gamma-1}\mathrm{e}^{-\alpha\ell}\,\mathrm{d}\ell}{\displaystyle\int_0^\infty \prod_{j=1}^m \Bigl(\frac{\ell^{N_{ij}}}{N_{ij}!}\mathrm{e}^{-\ell}\Bigr)\frac{\alpha^\gamma}{\Gamma(\gamma)}\ell^{\gamma-1}\mathrm{e}^{-\alpha\ell}\,\mathrm{d}\ell}$$

$$= \frac{\displaystyle\int_0^\infty \ell^{m\bar{N}_i+\gamma}\mathrm{e}^{-(m+\alpha)\ell}\,\mathrm{d}\ell}{\displaystyle\int_0^\infty \ell^{m\bar{N}_i+\gamma-1}\mathrm{e}^{-(m+\alpha)\ell}\,\mathrm{d}\ell} = \frac{\Gamma(m\bar{N}_i+\gamma+1)(m+\alpha)^{m\bar{N}_i+\gamma}}{(m+\alpha)^{m\bar{N}_i+\gamma+1}\Gamma(m\bar{N}_i+\gamma)}$$

$$= \frac{m\bar{N}_i+\gamma}{m+\alpha} = \frac{m}{m+\alpha}\bar{N}_i + \frac{\alpha}{m+\alpha}\Bigl(\frac{\gamma}{\alpha}\Bigr) = \frac{m}{m+\alpha}\bar{N}_i + \Bigl(1-\frac{m}{m+\alpha}\Bigr)\mathbb{E}[\lambda_i]\,.$$

The best estimator is therefore of the form

$$Z\bar{N}_i + (1-Z)\mathbb{E}[\lambda_i]$$

with the **credibility factor**

$$Z = \frac{1}{1+\dfrac{\alpha}{m}}\,.$$

The credibility factor Z is increasing with the number of years m and converges to 1 as $m \to \infty$. Thus the more observations we have the higher is the weight put to the empirical mean $\bar{N}_i$.

The only problem that remains is to estimate the parameters γ and α. For an insurance company this quite often is no problem because they have big data sets.

3.2.2. The Normal-Normal Model

For the next model we do not assume that the individual claims have the same distribution for all risks. First we assume that each risk consists of a lot of sub-risks such that the random variables Y_{ij} are approximately normally distributed. Moreover, the conditional variance is assumed to be the same for all risks, i.e. $Y_{ij} \sim \mathrm{N}(\Theta_i, \sigma^2)$. Next we assume that the parameter $\Theta_i \sim \mathrm{N}(\mu, \eta^2)$ is also normally distributed. Note that μ and η^2 have to be chosen in such a way that $\mathbb{P}[Y_{ij} \le 0]$ is small.

Let us first consider the joint density of $Y_{i1}, Y_{i2}, \dots Y_{im}, \Theta_i$

$$(2\pi)^{-(m+1)/2}\eta^{-1}\sigma^{-m}\exp\Bigl\{-\Bigl(\frac{(\vartheta-\mu)^2}{2\eta^2} + \sum_{j=1}^m \frac{(y_{ij}-\vartheta)^2}{2\sigma^2}\Bigr)\Bigr\}\,.$$

Because we will mainly be interested in the posterior distribution of Θ_i we should write the exponent in the form $(\vartheta - \cdot)^2$. For abbreviation we write $m\bar{y}_i = \sum_{j=1}^m y_{ij}$.

$$\frac{(\vartheta-\mu)^2}{2\eta^2} + \sum_{j=1}^m \frac{(y_{ij}-\vartheta)^2}{2\sigma^2} = \vartheta^2\Bigl(\frac{1}{2\eta^2} + \frac{m}{2\sigma^2}\Bigr) - 2\vartheta\Bigl(\frac{\mu}{2\eta^2} + \frac{m\bar{y}_i}{2\sigma^2}\Bigr) + \frac{\mu^2}{2\eta^2} + \sum_{j=1}^m \frac{y_{ij}^2}{2\sigma^2}\,.$$

The joint density can therefore be written as

$$C(y_{i1}, y_{i2}\cdots, y_{im}) \exp\left\{ -\frac{\Bigl(\vartheta - \Bigl(\frac{\mu}{\eta^2} + \frac{m\bar{y}_i}{\sigma^2}\Bigr)\Bigl(\frac{1}{\eta^2} + \frac{m}{\sigma^2}\Bigr)^{-1}\Bigr)^2}{2\Bigl(\frac{1}{\eta^2} + \frac{m}{\sigma^2}\Bigr)^{-1}} \right\}$$

where C is a function of the data. Hence the posterior distribution of Θ_i is normally distributed with mean

$$\Bigl(\frac{\mu}{\eta^2} + \frac{m\bar{Y}_i}{\sigma^2}\Bigr)\Bigl(\frac{1}{\eta^2} + \frac{m}{\sigma^2}\Bigr)^{-1} = \frac{\mu\sigma^2 + m\eta^2\bar{Y}_i}{\sigma^2 + m\eta^2} = \frac{m\eta^2}{\sigma^2 + m\eta^2}\bar{Y}_i + \Bigl(1 - \frac{m\eta^2}{\sigma^2 + m\eta^2}\Bigr)\mu\,.$$

Again the credibility premium is a linear combination $Z\bar{Y}_i + (1-Z)\mathbb{E}[Y_{ij}]$ of the mean losses of the i-th risk and the overall mean μ. The credibility factor has a similar form as in the Poisson-gamma model

$$Z = \frac{1}{1 + \frac{\sigma^2}{\eta^2 m}}\,.$$

3.2.3. Is the Credibility Premium Formula Always Linear?

From the previous considerations one could conjecture that the credibility premium is always of the form $Z\bar{Y}_i + (1-Z)\mathbb{E}[Y_{ij}]$. We try to find a counterexample. Assume that Θ_i takes the values 1 and 2 with equal probabilities $\frac{1}{2}$. Given Θ_i the random variables $Y_{ij} \sim \operatorname{Pois}(\Theta_i)$ are Poisson distributed. Then

$$\begin{aligned}
&\mathbb{E}[\Theta_i \mid Y_{i1}, Y_{i2}, \dots, Y_{im}] \\
&\qquad = \mathbb{P}[\Theta_i = 1 \mid Y_{i1}, Y_{i2}, \dots, Y_{im}] + 2\mathbb{P}[\Theta_i = 2 \mid Y_{i1}, Y_{i2}, \dots, Y_{im}] \\
&\qquad = 2 - \mathbb{P}[\Theta_i = 1 \mid Y_{i1}, Y_{i2}, \dots, Y_{im}] = 2 - \frac{\frac{1}{2}\prod_{j=1}^m \frac{1}{Y_{ij}!}\mathrm{e}^{-1}}{\frac{1}{2}\Bigl(\prod_{j=1}^m \frac{1}{Y_{ij}!}\mathrm{e}^{-1} + \prod_{j=1}^m \frac{2^{Y_{ij}}}{Y_{ij}!}\mathrm{e}^{-2}\Bigr)} \\
&\qquad = 2 - \frac{1}{1 + \mathrm{e}^{-m}2^{m\bar{Y}_i}}\,.
\end{aligned}$$

It turns out that there is no possibility to get a linear formula.

3.3. Empirical Bayes Credibility

In order to calculate the Bayes credibility estimator one needs to know the joint distribution of $m(\Theta_i)$ and $(Y_{i1}, Y_{i,2}, \ldots, Y_{im})$. In practice one does not have distributions but only data. It seems therefore not feasible to estimate joint distributions if only one of the variables is observable. We therefore want to restrict now to linear estimators, i.e. estimators of the form

$$M = a_{i0} + a_{i1}Y_{i1} + a_{i2}Y_{i2} + \cdots + a_{im}Y_{im} \;.$$

The best estimator is called **linear Bayes estimator**. It turns out that it is enough to know the first two moments of $(m(\Theta_i), Y_i 1)$. One then has to estimate these quantities. We further do not know the mean values and the variances. In practice we will have to estimate these quantities. We will therefore now proceed in two steps: First we estimate the linear Bayes premium, and then we estimate the first two moments. The corresponding estimator is called **empirical Bayes estimator**.

3.3.1. The Bühlmann Model

In addition to the notation introduced before denote by $\sigma^2 = \mathbb{E}[s^2(\Theta_i)]$ and by $v^2 = \operatorname{Var}[m(\Theta_i)]$. Note that σ^2 is not the unconditional variance of Y_{ij}. In fact

$$\mathbb{E}[Y_{ij}^2] = \mathbb{E}[\mathbb{E}[Y_{ij}^2 \mid \Theta_i]] = \mathbb{E}[s^2(\Theta_i) + (m(\Theta_i))^2] = \sigma^2 + \mathbb{E}[(m(\Theta_i))^2]\,.$$

Thus the variance of Y_{ij} is

$$\operatorname{Var}[Y_{ij}] = \sigma^2 + \mathbb{E}[(m(\Theta_i))^2] - \mu^2 = \sigma^2 + v^2\,. \tag{3.1}$$

We consider credibility premia of the form

$$a_{i0} + \sum_{j=1}^m a_{ij}Y_{ij}\;.$$

Which parameters a_{ij} minimize the expected quadratic error

$$\mathbb{E}\Bigl[\Bigl(a_{i0} + \sum_{j=1}^m a_{ij}Y_{ij} - m(\Theta_i)\Bigr)^2\Bigr]\ ?$$

We first differentiate with respect to a_{i0}.

$$\begin{aligned}\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}a_{i0}}\mathbb{E}\Bigl[\Bigl(a_{i0} + \sum_{j=1}^m a_{ij}Y_{ij} - m(\Theta_i)\Bigr)^2\Bigr] &= \mathbb{E}\Bigl[a_{i0} + \sum_{j=1}^m a_{ij}Y_{ij} - m(\Theta_i)\Bigr] \\ &= a_{i0} + \Bigl(\sum_{j=1}^m a_{ij} - 1\Bigr)\mu = 0\,.\end{aligned}$$

It follows that

$$a_{i0} = \Bigl(1 - \sum_{j=1}^m a_{ij}\Bigr)\mu$$

and thus our estimator has the property that

$$\mathbb{E}\Bigl[a_{i0} + \sum_{j=1}^m a_{ij} Y_{ij}\Bigr] = \mu\,.$$

Let $1 \le k \le m$ and differentiate with respect to a_{ik}.

$$\begin{aligned}\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}a_{ik}}\mathbb{E}\Bigl[\Bigl(a_{i0} + \sum_{j=1}^m a_{ij}Y_{ij} - m(\Theta_i)\Bigr)^2\Bigr] &= \mathbb{E}\Bigl[Y_{ik}\Bigl(a_{i0} + \sum_{j=1}^m a_{ij}Y_{ij} - m(\Theta_i)\Bigr)\Bigr] \\ &= a_{i0}\mu + \sum_{\substack{j=1\\ j\neq k}}^m a_{ij}\mathbb{E}[Y_{ik}Y_{ik}] + a_{ik}\mathbb{E}[Y_{ik}^2] - \mathbb{E}[Y_{ik}m(\Theta_i)] = 0\,.\end{aligned}$$

We already know from (3.1) that $\mathbb{E}[Y_{ik}^2] = \sigma^2 + v^2 + \mu^2$. For $j \neq k$ we get

$$\mathbb{E}[Y_{ik}Y_{ij}] = \mathbb{E}[\mathbb{E}[Y_{ik}Y_{ij} \mid \Theta_i]] = \mathbb{E}[(m(\Theta_i))^2] = v^2 + \mu^2\,.$$

And finally

$$\mathbb{E}[Y_{ik}m(\Theta_i)] = \mathbb{E}[\mathbb{E}[Y_{ik}m(\Theta_i) \mid \Theta_i]] = \mathbb{E}[(m(\Theta_i))^2] = v^2 + \mu^2\,.$$

The equations to solve are

$$\begin{aligned}&\Bigl(1 - \sum_{j=1}^m a_{ij}\Bigr)\mu^2 + \sum_{\substack{j=1\\ j\neq k}}^m a_{ij}(v^2 + \mu^2) + a_{ik}(\sigma^2 + v^2 + \mu^2) - (v^2 + \mu^2) \\ &\qquad = a_{ik}\sigma^2 - \Bigl(1 - \sum_{j=1}^m a_{ij}\Bigr)v^2 = 0\,.\end{aligned}$$

The right hand side of

$$\sigma^2 a_{ik} = v^2\Bigl(1 - \sum_{j=1}^m a_{ij}\Bigr)$$

is independent of k. Thus

$$a_{i1} = a_{i2} = \cdots = a_{im} = \frac{v^2}{\sigma^2 + mv^2}\,. \tag{3.2}$$

The credibility premium is

$$\frac{mv^2}{\sigma^2 + mv^2}\bar{Y}_i + \Bigl(1 - \frac{mv^2}{\sigma^2 + mv^2}\Bigr)\mu = Z\bar{Y}_i + (1 - Z)\mu \tag{3.3}$$

where

$$Z = \frac{1}{1 + \dfrac{\sigma^2}{v^2 m}}\,.$$

The formula for the credibility premium does only depend on μ, σ^2, and v^2. This also were the only quantities we had assumed in our model. Hence the approach is quite general. We do not need any other assumptions on the distribution of Θ_i.

In order to apply the result we need to estimate the parameters μ, σ^2, and v^2. In order to make the right decision we look for unbiased estimators.

i) μ: The natural estimator of μ is

$$\hat{\mu} = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} Y_{ij} = \frac{1}{n}\sum_{i=1}^{n}\bar{Y}_i\,. \tag{3.4}$$

It is easy to see that $\hat{\mu}$ is unbiased. It can be shown that $\hat{\mu}$ is the best linear unbiased estimator.

ii) $\boldsymbol{\sigma^2}$: For the estimation of $s^2(\Theta_i)$ we would use the unbiased estimator

$$\frac{1}{m-1}\sum_{j=1}^{m}(Y_{ij} - \bar{Y}_i)^2\,.$$

Thus

$$\hat{\sigma}^2 = \frac{1}{n(m-1)}\sum_{i=1}^{n}\sum_{j=1}^{m}(Y_{ij} - \bar{Y}_i)^2 \tag{3.5}$$

is an unbiased estimator for σ^2.

iii) $\mathbf{v^2}$: $\bar{Y}_i$ is an unbiased estimator of $m(\Theta_i)$. Therefore a natural estimate of v^2 would be

$$\frac{1}{n-1}\sum_{i=1}^{n}(\bar{Y}_i - \hat{\mu})^2\,.$$

But is this estimator really unbiased?

$$\begin{aligned}
\frac{1}{n-1}\sum_{i=1}^{n}\mathbb{E}[(\bar{Y}_i-\hat{\mu})^2] &= \frac{1}{n-1}\sum_{i=1}^{n}\mathbb{E}\Bigl[\Bigl(\frac{n-1}{n}\bar{Y}_i-\frac{1}{n}\sum_{j\neq i}\bar{Y}_j\Bigr)^2\Bigr] \\
&= \frac{n}{n-1}\mathbb{E}\Bigl[\Bigl(\frac{n-1}{n}\bar{Y}_1-\frac{1}{n}\sum_{j=2}^{n}\bar{Y}_j\Bigr)^2\Bigr] \\
&= \frac{n}{n-1}\mathbb{E}\Bigl[\Bigl(\frac{n-1}{n}(\bar{Y}_1-\mu)-\frac{1}{n}\sum_{j=2}^{n}(\bar{Y}_j-\mu)\Bigr)^2\Bigr] \\
&= \frac{n}{n-1}\Bigl(\Bigl(\frac{n-1}{n}\Bigr)^2\mathbb{E}[(\bar{Y}_1-\mu)^2]+\frac{n-1}{n^2}\mathbb{E}[(\bar{Y}_1-\mu)^2]\Bigr) \\
&= \mathbb{E}\left[(\bar{Y}_1-\mu)^2\right] = \mathbb{E}\Bigl[\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}Y_{1i}Y_{1j}-\frac{2}{m}\mu\sum_{j=1}^{m}Y_{1j}+\mu^2\Bigr] \\
&= \frac{1}{m}(\sigma^2+v^2+\mu^2)+\frac{m-1}{m}(v^2+\mu^2)-\mu^2 = v^2+\frac{\sigma^2}{m}\,.
\end{aligned}$$

Our proposal for an estimator turns out to be biased. We have to correct the estimator and get

$$\hat{v}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\bar{Y}_i-\hat{\mu})^2-\frac{1}{m}\hat{\sigma}^2\,. \tag{3.6}$$

Example 3.1. A company has insured twenty similar collective risks over the last ten years. Table 3.2 shows the annual losses of each the risks. Thus the estimators are

$$\hat{\mu} = \frac{1}{20}\sum_{i=1}^{20}\bar{Y}_i = 102.02\,,$$

$$\hat{\sigma}^2 = \frac{1}{180}\sum_{i=1}^{20}\sum_{j=1}^{10}(Y_{ij}-\bar{Y}_i)^2 = 473.196$$

and

$$\hat{v}^2 = \frac{1}{19}\sum_{i=1}^{20}(\bar{Y}_i-\hat{\mu})^2-\frac{1}{10}\hat{\sigma}^2 = 806.6\,.$$

The credibility factor can now be computed

$$Z = \frac{1}{1+\dfrac{\hat{\sigma}^2}{\hat{v}^2 m}} = 0.944585\,.$$

Table 3.3 gives the credibility premia for next year for each of the risks. The data had been simulated using $\mu = 100$, $\sigma^2 = 400$ and $v^2 = 900$. The credibility model

i	1	2	3	4	5	6	7	8	9	10	$\bar{Y}_i$
1	67	71	50	56	64	69	77	94	56	48	65.2
2	80	82	109	61	89	113	149	91	127	108	100.9
3	125	118	101	135	120	117	101	101	100	122	114.0
4	96	144	152	124	94	132	155	143	94	113	124.7
5	176	161	153	191	139	157	192	151	175	147	164.2
6	89	129	95	88	131	72	91	122	79	56	95.2
7	70	116	102	129	81	69	75	120	95	93	95.0
8	22	33	48	20	56	25	43	53	48	64	41.2
9	121	106	55	103	87	81	130	113	98	97	99.1
10	126	101	158	129	110	112	107	114	117	95	116.9
11	125	106	104	135	83	177	157	95	101	128	121.1
12	116	134	133	167	142	150	134	153	134	109	137.2
13	53	62	89	76	63	66	69	70	57	49	65.4
14	110	136	75	52	49	85	110	89	68	34	80.8
15	125	112	145	114	61	74	168	98	114	97	110.8
16	87	95	85	111	89	101	85	88	90	164	99.5
17	92	46	74	82	68	94	116	109	81	89	85.1
18	44	74	67	93	89	57	83	53	82	38	68.0
19	95	107	147	154	121	125	146	83	117	139	123.4
20	103	105	139	138	145	145	140	148	137	127	132.7

Table 3.2: *Annual losses of the risks*

interprets the data as having more fluctuations in a single risk than in the mean values of the risks. This is not surprising because we only have 10 data for each risk but 20 risks. Thus the fluctuations of $\bar{Y}_i$ could also be due to larger fluctuations within the risk. ■

3.3.2. The Bühlmann Straub model

The main field of application for credibility theory are collective insurance contracts, e.g. employees insurance, fire insurance for companies, third party liability insurance for employees, travel insurance for the customers of a travel agent etc.. In reality the volume of the risk is not the same for each of the contracts. Looking at the credibility formulae calculated until now we can recognize that the credibility factors should be larger for higher risk volumes, because we have more data available. Thus we want to take the volume of the risk into consideration. The risk volume can be the number of employees, the sum insured, etc..

In order to see how to model different risk volumes we first consider an example.

Risk	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Premium	67.240	100.962	113.336	123.443	160.754	95.578	95.389	44.570	99.262	116.075	120.043	135.251	67.429	81.976	110.313	99.640	86.038	69.885	122.215	131.000

Table 3.3: *Credibility premia for the twenty risks*

Example 3.2. Assume that we insure taxi drivers. The losses of each driver are dependent on the driver itself, but also on the policies of the employer and the region where the company works. The i-th company employs P_i taxi drivers. There is a random variable Θ_i which determines the environment and the policies of the company. The random variable Θ_{ik} determines the risk class of each single driver. We assume that the vectors $(\Theta_{i1}, \Theta_{i2}, \dots, \Theta_{iP_i}, \Theta_i)$ are independent and that, given Θ_i, the parameters $\Theta_{i1}, \Theta_{i2}, \dots, \Theta_{iP_i}$ are conditionally iid.. The aggregate claims of different companies shall be independent. The aggregate claims of different drivers of company i are conditionally independent given Θ_i, and the aggregate claims Y_{ikj} of a driver are conditionally independent given Θ_{ik} and Y_{ikj} depends on Θ_{ik} only. Then the expected value of the annual aggregate claims of company i given Θ_i is

$$\mathbb{E}\Bigl[\sum_{k=1}^{P_i} Y_{ikj} \Bigm| \Theta_i\Bigr] = P_i \mathbb{E}[Y_{i1j} \mid \Theta_i]\,.$$

The conditional variance of company i is then

$$\mathbb{E}\Bigl[\Bigl(\sum_{k=1}^{P_i} Y_{ikj} - P_i\mathbb{E}[Y_{i1j} \mid \Theta_i]\Bigr)^2 \Bigm| \Theta_i\Bigr] = P_i\mathbb{E}\left[(Y_{i1j} - \mathbb{E}[Y_{i1j} \mid \Theta_i])^2 \middle| \Theta_i\right]\,.$$

Thus both the conditional mean and the conditional variance are proportional to characteristics of the company. ■

The example will motivate the model assumptions. Let P_{ij} denote the volume of the i-th risk in year j. We assume that there exists a parameter Θ_i for each risk. The parameters $\Theta_1, \Theta_2, \dots, \Theta_n$ are assumed to be iid.. Denote by Y_{ij} the aggregate claims of risk i in year j. We assume that the vectors $(Y_{i1}, Y_{i2}, \dots, Y_{im}, \Theta_i)$ are independent. Within each risk the random variables $Y_{i1}, Y_{i2}, \dots, Y_{im}$ are conditionally independent given Θ_i. Moreover, there exist functions $m(\vartheta)$ and $s^2(\vartheta)$ such that

$$\mathbb{E}[Y_{ij} \mid \Theta_i = \vartheta] = P_{ij} m(\vartheta)$$

and

$$\operatorname{Var}[Y_{ij} \mid \Theta_i = \vartheta] = P_{ij} s^2(\vartheta)\,.$$

As in the Bühlmann model we let $\mu = \mathbb{E}[m(\Theta_i)]$, $\sigma^2 = \mathbb{E}[s^2(\Theta_i)]$ and $v^2 = \operatorname{Var}[m(\Theta_i)]$. Denote by

$$P_{i\cdot} = \sum_{j=1}^{m} P_{ij}, \qquad\qquad P_{\cdot\cdot} = \sum_{i=1}^{n} P_{i\cdot}$$

the risk volume of the i-th risk over all m years and the risk volume over all years and companies.

First we normalize the annual losses. Let $X_{ij} = Y_{ij}/P_{ij}$. Note that $\mathbb{E}[X_{ij} \mid \Theta_i] = m(\Theta_i)$ and $\operatorname{Var}[X_{ij} \mid \Theta_i] = s^2(\Theta_i)/P_{ij}$. We next try to find the best linear estimator for $m(\Theta_i)$. Hence we have to minimize

$$\mathbb{E}\Bigl[\Bigl(a_{i0} + \sum_{j=1}^{m} a_{ij} X_{ij} - m(\Theta_i)\Bigr)^2\Bigr]\,.$$

It is clear that X_{kj} does not carry information on $m(\Theta_i)$ if $k \neq i$. We first differentiate with respect to a_{i0}.

$$\begin{aligned}\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}a_{i0}}\mathbb{E}\Bigl[\Bigl(a_{i0} + \sum_{j=1}^{m} a_{ij} X_{ij} - m(\Theta_i)\Bigr)^2\Bigr] &= \mathbb{E}\Bigl[a_{i0} + \sum_{j=1}^{m} a_{ij} X_{ij} - m(\Theta_i)\Bigr] \\ &= a_{i0} - \Bigl(1 - \sum_{j=1}^{m} a_{ij}\Bigr)\mu = 0\,.\end{aligned}$$

As in the Bühlmann model we get

$$a_{i0} = \Bigl(1 - \sum_{j=1}^{m} a_{ij}\Bigr)\mu$$

and therefore

$$\mathbb{E}\Bigl[a_{i0} + \sum_{j=1}^{m} a_{ij} X_{ij}\Bigr] = \mu\,.$$

Let us differentiate with respect to a_{ik},

$$\begin{aligned}\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}a_{ik}}\mathbb{E}\Bigl[\Bigl(a_{i0} + \sum_{j=1}^{m} a_{ij} X_{ij} - m(\Theta_i)\Bigr)^2\Bigr] &= \mathbb{E}\Bigl[X_{ik}\Bigl(a_{i0} + \sum_{j=1}^{m} a_{ij} X_{ij} - m(\Theta_i)\Bigr)\Bigr] \\ &= a_{i0}\mu + \sum_{\substack{j=1\\ j\neq k}}^{m} a_{ij}\mathbb{E}[X_{ik}X_{ij}] + a_{ik}\mathbb{E}\left[X_{ik}^2\right] - \mathbb{E}[X_{ik} m(\Theta_i)] = 0\,.\end{aligned}$$

Let us compute the terms. For $j \neq k$

$$\mathbb{E}[X_{ik}X_{ij}] = \mathbb{E}[\mathbb{E}[X_{ik}X_{ij} \mid \Theta_i]] = \mathbb{E}\left[(m(\Theta_i))^2\right] = v^2 + \mu^2 \,.$$

$$\mathbb{E}\left[X_{ik}^2\right] = \mathbb{E}\left[\mathbb{E}\left[X_{ik}^2 \middle| \Theta_i\right]\right] = \mathbb{E}\left[(m(\Theta_i))^2 + P_{ik}^{-1}s^2(\Theta_i)\right] = v^2 + \mu^2 + P_{ik}^{-1}\sigma^2 \,.$$

$$\mathbb{E}[X_{ik}m(\Theta_i)] = \mathbb{E}[\mathbb{E}[X_{ik}m(\Theta_i) \mid \Theta_i]] = \mathbb{E}\left[(m(\Theta_i))^2\right] = v^2 + \mu^2 \,.$$

Thus

$$\Bigl(1 - \sum_{j=1}^m a_{ij}\Bigr)\mu^2 + \sum_{\substack{j=1\\ j\neq k}}^m a_{ij}(v^2 + \mu^2) + a_{ik}(v^2 + \mu^2 + P_{ik}^{-1}\sigma^2) - (v^2 + \mu^2)$$

$$= a_{ik}P_{ik}^{-1}\sigma^2 - \Bigl(1 - \sum_{j=1}^m a_{ij}\Bigr)v^2 = 0 \,.$$

The right hand side of

$$\sigma^2 P_{ik}^{-1} a_{ik} = \Bigl(1 - \sum_{j=1}^m a_{ij}\Bigr)v^2$$

is independent of k and thus there exists a constant a such that $a_{ik} = P_{ik}a$. Then it follows readily

$$a_{ik} = P_{ik}\frac{v^2}{\sigma^2 + P_{i\cdot}v^2} \,.$$

Note that the formula is consistent with (3.2). Define

$$\bar{X}_i = \frac{1}{P_{i\cdot}}\sum_{j=1}^m P_{ij}X_{ij}$$

the weighted mean of the annual losses of the i-th risk. Then the credibility premium is

$$\Bigl(1 - \frac{P_{i\cdot}v^2}{\sigma^2 + P_{i\cdot}v^2}\Bigr)\mu + \frac{P_{i\cdot}v^2}{\sigma^2 + P_{i\cdot}v^2}\bar{X}_i = Z_i\bar{X}_i + (1 - Z_i)\mu$$

with the credibility factor

$$Z_i = \frac{1}{1 + \dfrac{\sigma^2}{P_{i\cdot}v^2}} \,.$$

Note that the premium is consistent with (3.3). Denote by $\hat{P}_{i(m+1)}$ the (estimated) risk volume of the next year. Then the credibility premium for the next year will be

$$\hat{P}_{i(m+1)}(Z_i\bar{X}_i + (1 - Z_i)\mu) \,.$$

Note that the credibility factor Z_i depends on $P_{i\cdot}$. This means that in general different risks will get different credibility factors.

It remains to find estimators for the parameters.

i) μ: We want to modify the estimator (3.4) for the Bühlmann Straub model. The loss X_{ij} should be weighted with P_{ij} because it originates from a volume of this size. Thus we get

$$\hat{\mu} = \frac{1}{P_{\cdot\cdot}} \sum_{i=1}^{n} \sum_{j=1}^{m} P_{ij} X_{ij} = \frac{1}{P_{\cdot\cdot}} \sum_{i=1}^{n} P_{i\cdot} \bar{X}_i \,.$$

It turns out that $\hat{\mu}$ is unbiased.

ii) $\boldsymbol{\sigma^2}$: We want to modify the estimator (3.5) for the Bühlmann Straub model. The conditional variance of X_{ij} is $P_{ij}^{-1} s^2(\Theta_i)$. This suggests that we should weight the term $(X_{ij} - \bar{X}_i)^2$ with P_{ij}. Therefore we try an estimator of the form

$$c \sum_{i=1}^{n} \sum_{j=1}^{m} P_{ij} (X_{ij} - \bar{X}_i)^2 \,.$$

In order to compute the expected value of this estimator we keep i fixed.

$$\begin{aligned}
\mathbb{E}\Bigl[\sum_{j=1}^{m} P_{ij}(X_{ij} - \bar{X}_i)^2\Bigr] &= \mathbb{E}\Bigl[\sum_{j=1}^{m} P_{ij} X_{ij}^2 - 2\sum_{j=1}^{m} P_{ij} X_{ij} \bar{X}_i + \sum_{j=1}^{m} P_{ij} \bar{X}_i^2\Bigr] \\
&= \mathbb{E}\Bigl[\sum_{j=1}^{m} P_{ij} X_{ij}^2\Bigr] - \mathbb{E}[P_{i\cdot} \bar{X}_i^2] \\
&= \sum_{j=1}^{m} P_{ij} \Bigl(\frac{\sigma^2}{P_{ij}} + v^2 + \mu^2\Bigr) - \frac{1}{P_{i\cdot}} \sum_{j=1}^{m} \sum_{k=1}^{m} P_{ij} P_{ik} \mathbb{E}[X_{ij} X_{ik}] \\
&= m\sigma^2 + P_{i\cdot}(v^2 + \mu^2) - P_{i\cdot}(v^2 + \mu^2) - \sigma^2 = (m-1)\sigma^2 \,. \qquad (3.7)
\end{aligned}$$

Hence the proposed estimator has mean value

$$\mathbb{E}\Bigl[c \sum_{i=1}^{n} \sum_{j=1}^{m} P_{ij} (X_{ij} - \bar{X}_i)^2\Bigr] = cn(m-1)\sigma^2 \,.$$

Thus our unbiased estimator is

$$\hat{\sigma}^2 = \frac{1}{n(m-1)} \sum_{i=1}^{n} \sum_{j=1}^{m} P_{ij} (X_{ij} - \bar{X}_i)^2 \,.$$

iii) $\mathbf{v^2}$: A closer look at the estimator (3.6) shows hat it contains the terms $(Y_{ij} - \hat{\mu})^2$, all with the same weight. For the estimator in the Bühlmann Straub model we propose an estimator that contains a term of the form

$$\sum_{i=1}^{n} \sum_{j=1}^{m} P_{ij} (X_{ij} - \hat{\mu})^2 \,.$$

P_{ij}		Year									
		1	2	3	4	5	6	7	8	9	10
	1	10	10	12	12	15	15	15	13	9	9
	2	5	5	5	5	5	5	5	5	5	5
Company	3	2	2	3	3	3	3	3	3	3	3
number	4	15	20	20	20	21	22	22	22	25	25
	5	5	5	5	3	3	3	3	3	3	5
	6	10	10	10	10	10	10	10	12	12	12

Table 3.4: *Premium volumes*

Let us compute its expected value.

$$\begin{aligned}
\mathbb{E}\Bigl[\sum_{i=1}^n\sum_{j=1}^m P_{ij}(X_{ij}-\hat\mu)^2\Bigr] &= \sum_{i=1}^n\sum_{j=1}^m P_{ij}\mathbb{E}[X_{ij}^2] - P_{\cdot\cdot}\mathbb{E}[\hat\mu^2] \\
&= \sum_{i=1}^n\sum_{j=1}^m P_{ij}\Bigl(\frac{\sigma^2}{P_{ij}}+\upsilon^2+\mu^2\Bigr) - \frac{1}{P_{\cdot\cdot}}\sum_{i=1}^n\sum_{j=1}^m\sum_{k=1}^n\sum_{l=1}^m P_{ij}P_{kl}\mathbb{E}[X_{ij}X_{kl}] \\
&= nm\sigma^2 + P_{\cdot\cdot}(\upsilon^2+\mu^2) - \frac{1}{P_{\cdot\cdot}}\Bigl(\sum_{i=1}^n\sum_{j=1}^m P_{ij}^2\Bigl(\frac{\sigma^2}{P_{ij}}+\upsilon^2+\mu^2\Bigr) \\
&\qquad + \sum_{i=1}^n\sum_{j=1}^m\sum_{\substack{l=1\\ l\neq j}}^m P_{ij}P_{il}(\upsilon^2+\mu^2) + \sum_{i=1}^n\sum_{j=1}^m\sum_{\substack{k=1\\ k\neq i}}^n\sum_{l=1}^m P_{ij}P_{kl}\mu^2\Bigr) \\
&= (nm-1)\sigma^2 + \Bigl(P_{\cdot\cdot} - \sum_{i=1}^n \frac{P_{i\cdot}^2}{P_{\cdot\cdot}}\Bigr)\upsilon^2\,. \qquad (3.8)
\end{aligned}$$

Let

$$P^* = \frac{1}{nm-1}\sum_{i=1}^n P_{i\cdot}\Bigl(1-\frac{P_{i\cdot}}{P_{\cdot\cdot}}\Bigr)\,.$$

Then

$$\hat\upsilon^2 = \frac{1}{P^*}\Bigl(\Bigl(\frac{1}{nm-1}\sum_{i=1}^n\sum_{j=1}^m P_{ij}(X_{ij}-\hat\mu)^2\Bigr) - \hat\sigma^2\Bigr)$$

is an unbiased estimator of υ^2.

Example 3.3. An insurance company has issued motor insurance policies for business cars to six companies. Table 3.4 shows the number of cars for each of the six companies in each of the past ten years. Table 3.5 shows the aggregate claims, measured in thousands of €, for each of the companies in each of the past ten years. The first step is to calculate the mean aggregate claims per car $X_{ij} = Y_{ij}/P_{ij}$ for each of the companies in each year, see Table 3.6. For the computation of the estimators

Y_{ij} Year

Company number	1	2	3	4	5	6	7	8	9	10
1	74	50	180	43	179	140	95	149	20	81
2	52	111	83	74	87	85	40	71	93	83
3	40	28	60	59	43	74	61	46	72	81
4	171	85	100	116	153	44	13	31	110	252
5	49	132	74	37	21	54	27	43	53	102
6	126	148	128	151	165	128	100	65	233	246

Table 3.5: *Aggregate annual losses*

Year

	1	2	3	4	5	6	7	8	9	10
1	7.4	5	15	3.583	11.933	9.333	6.333	11.462	2.222	9
2	10.4	22.2	16.6	14.8	17.4	17	8	14.2	18.6	16.6
3	20	14	20	19.667	14.333	24.667	20.333	15.333	24	27
4	11.4	4.25	5	5.8	7.286	2	0.591	1.409	4.4	10.08
5	9.8	26.4	14.8	12.333	7	18	9	14.333	17.667	20.4
6	12.6	14.8	12.8	15.1	16.5	12.8	10	5.417	19.417	20.5

Table 3.6: *Normalised losses*

the figures in Table 3.7 are needed. We further get

$$P_{\cdot\cdot} = 554, \qquad P^* = 7.08585\,.$$

We can now compute the following estimates.

$$\hat{\mu} = \frac{1}{P_{\cdot\cdot}} \sum_{i=1}^{6} P_{i\cdot} \bar{X}_i = 9.94765\,,$$

$$\hat{\sigma}^2 = \frac{1}{6} \sum_{i=1}^{6} \frac{1}{9} \sum_{j=1}^{10} P_{ij}(X_{ij} - \bar{X}_i)^2 = 157.808\,,$$

$$\hat{v}^2 = \frac{1}{P^*}\left[\frac{1}{59} \sum_{i=1}^{6} \sum_{j=1}^{10} P_{ij}(X_{ij} - \hat{\mu})^2 - \hat{\sigma}^2\right] = 28.7578\,.$$

The number of business cars each company plans to use next year and the credibility premium is given in Table 3.8. One can clearly see how the credibility factor increases with $P_{i\cdot}$. ■

3.3.3. The Bühlmann Straub model with missing data

In practice it is not realistic that all policy holders of a certain collective insurance type started to insure their risks in the same year. Thus some of the P_{ij}'s may be

Company	$P_{i\cdot}$	$\bar{X}_i$	$\sum_j P_{ij}(X_{ij} - \bar{X}_i)^2$	$\sum_j P_{ij}(X_{ij} - \hat{\mu})^2$
1	120	8.425	1659.62	1937.84
2	50	15.580	735.78	2321.95
3	28	20.143	494.10	3404.48
4	212	5.071	2310.63	7352.86
5	38	15.579	1289.26	2494.30
6	106	14.057	2032.23	3821.88

Table 3.7: *Quantities used in the estimators*

Company	Cars next year	Credibility factor	Credibility premium per unit volume	Actual premium
1	9	0.95627	8.4916	76.424
2	6	0.90110	15.0230	90.138
3	3	0.83613	18.4722	55.417
4	24	0.97477	5.1938	124.651
5	3	0.87382	14.8684	44.605
6	12	0.95078	13.8544	166.252

Table 3.8: *Premium next year*

0. The question arises whether this has an influence on the model. The answer is that its influence is only small. Denote by m_i the number of years with $P_{ij} \neq 0$.

For the credibility estimator it is clear that $a_{ij} = 0$ if $P_{ij} = 0$ because then also $Y_{ij} = 0$. For convenience we define $X_{ij} = 0$ if $P_{ij} = 0$. The computation of a_{ij} does not change if some of the P_{ij}'s are 0. Moreover, it is easy to see that $\hat{\mu}$ remains unbiased.

Next let us consider the estimator of σ^2. A closer look at (3.7) shows that

$$\mathbb{E}\Bigl[\sum_{j=1}^{m} P_{ij}(X_{ij} - \bar{X}_i)^2\Bigr] = (m_i - 1)\sigma^2 \,.$$

Hence the estimator of σ^2 must be changed to

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{m_i - 1} \sum_{j=1}^{m} P_{ij}(X_{ij} - \bar{X}_i)^2 \,.$$

Consider the estimator of v. The computation (3.8) changes to

$$\mathbb{E}\Bigl[\sum_{i=1}^{n}\sum_{j=1}^{m} P_{ij}(X_{ij} - \hat{\mu})^2\Bigr] = (m_{\cdot} - 1)\sigma^2 + \Bigl(P_{\cdot\cdot} - \sum_{i=1}^{n} \frac{P_{i\cdot}^2}{P_{\cdot\cdot}}\Bigr)v^2$$

where

$$m. = \sum_{i=1}^n m_i\,.$$

Thus one has to change

$$P^* = \frac{1}{m. - 1}\sum_{i=1}^n P_{i\cdot}\Bigl(1 - \frac{P_{i\cdot}}{P_{\cdot\cdot}}\Bigr)$$

and

$$\hat v^2 = \frac{1}{P^*}\Bigl(\Bigl(\frac{1}{m. - 1}\sum_{i=1}^n\sum_{j=1}^m P_{ij}(X_{ij} - \hat\mu)^2\Bigr) - \hat\sigma^2\Bigr)\,.$$

3.4. General Bayes Methods

We can generalise the approach above by summarise the observations in a vector $\boldsymbol{X}_i$. In the models considered before we would have $\boldsymbol{X}_i = (X_{i,1}, X_{i,2}, \ldots, X_{i,m_i})^\top$. We then want to estimate the vector $\boldsymbol{m}_i \in \mathbb{R}^s$. In the models above we had $\boldsymbol{m}_i = m(\Theta_i)$. We assume that the vectors $(\boldsymbol{X}_i^\top, \boldsymbol{m}_i^\top)^\top$ are independent and that second moments exist. In contrast to the models considered above we do not assume that the X_{ik} are conditionally independent nor identically distributed. For example, we could have $\boldsymbol{m}_i = (\mathbb{E}[X_{ik} \mid \Theta_i], \mathbb{E}[X_{ik}^2 \mid \Theta_i])^\top$ and $\boldsymbol{X}_i = (X_{i1}, X_{i1}^2, X_{i2}, X_{i2}^2, \ldots, X_{i,m_i}, X_{i,m_i}^2)^\top$. The goal is now to estimate $\boldsymbol{m}_i$.

The estimator $\boldsymbol{M}$ minimising $\mathbb{E}[\|\boldsymbol{M} - \boldsymbol{m}_i\|^2]$ is then $\boldsymbol{M} = \mathbb{E}[\boldsymbol{m}_i \mid \boldsymbol{X}_i]$. The problem with the estimator is again, that we need the joint distribution of the quantities. We therefore restrict again to linear estimators, i.e. estimators of the form $\boldsymbol{M} = \boldsymbol{g} + \boldsymbol{G}\boldsymbol{X}$, where we for the moment omit the index. We denote the entries of $\boldsymbol{g}$ by (g_{i0}). The quantity to minimise is then

$$\rho(\boldsymbol{M}) := \sum_i \mathbb{E}\Bigl[\Bigl(g_{i0} + \sum_{j=1}^m g_{ij}X_j - m_i\Bigr)^2\Bigr]\,.$$

In order to minimise $\rho(\boldsymbol{M})$ we first take the derivative with respect to g_{i0} and equate it to zero

$$\frac{\mathrm{d}\rho(\boldsymbol{M})}{\mathrm{d}g_{i0}} = 2\mathbb{E}\Bigl[g_{i0} + \sum_{\ell=1}^m g_{i\ell}X_\ell - m_i\Bigr] = 0\,.$$

The derivative with respect to g_{ij} yields

$$\frac{\mathrm{d}\rho(\boldsymbol{M})}{\mathrm{d}g_{ij}} = 2\mathbb{E}\Bigl[X_j\Bigl(g_{i0} + \sum_{\ell=1}^m g_{i\ell}X_\ell - m_i\Bigr)\Bigr] = 0\,.$$

In matrix form we can express the equations as

$$\begin{aligned}\mathbb{E}[\boldsymbol{g} + \boldsymbol{G}\boldsymbol{X} - \boldsymbol{m}] &= \boldsymbol{0}\,, \\ \mathbb{E}[(\boldsymbol{g} + \boldsymbol{G}\boldsymbol{X} - \boldsymbol{m})\boldsymbol{X}^\top] &= \boldsymbol{0}\,.\end{aligned}$$

The solution we look for is therefore

$$\boldsymbol{g} = \mathbb{E}[\boldsymbol{m}] - \boldsymbol{G}\mathbb{E}[\boldsymbol{X}]\,, \tag{3.9a}$$

$$\boldsymbol{G} = \operatorname{Cov}[\boldsymbol{m}, \boldsymbol{X}^\top]\operatorname{Var}[\boldsymbol{X}]^{-1}\,. \tag{3.9b}$$

Here we assume that $\operatorname{Var}[\boldsymbol{X}]$ is invertible. Note that $\operatorname{Var}[\boldsymbol{X}]$ is symmetric. Because for any vector $\boldsymbol{a}$ we have $\boldsymbol{a}^\top \operatorname{Var}[\boldsymbol{X}]\boldsymbol{a} = \operatorname{Var}[\boldsymbol{a}^\top \boldsymbol{X}]$ it follows that $\operatorname{Var}[\boldsymbol{X}]$ is positive semi-definite. If $\operatorname{Var}[\boldsymbol{X}]$ is not invertible, then there is $\boldsymbol{a}$ such that $\boldsymbol{a}^\top \operatorname{Var}[\boldsymbol{X}]\boldsymbol{a} = 0$, or equivalently, $\boldsymbol{a}^\top \boldsymbol{X} = 0$. Hence some of the entries of $\boldsymbol{X}$ can be expressed via the others. In this case, it is no loss of generality to reduce $\boldsymbol{X}$ to $\boldsymbol{X}^*$ such that $\operatorname{Var}[\boldsymbol{X}^*]$ becomes invertible.

If we now again consider n collective contracts $(\boldsymbol{X}_i, \boldsymbol{m}_i)$ the **linear Bayes estimator $\boldsymbol{M}_i$** for $\boldsymbol{m}_i$ can be written as

$$\boldsymbol{M}_i = \mathbb{E}[\boldsymbol{m}_i] + \operatorname{Cov}[\boldsymbol{m}_i, \boldsymbol{X}_i^\top]\operatorname{Var}[\boldsymbol{X}_i]^{-1}(\boldsymbol{X}_i - \mathbb{E}[\boldsymbol{X}_i])\,.$$

Thus we start with expected value $\mathbb{E}[\boldsymbol{m}_i]$ and correct it according to the deviation of $\boldsymbol{X}_i$ from the expected value $\mathbb{E}[\boldsymbol{X}_i]$. The correction is 'larger' if the covariance between $\boldsymbol{m}_i$ and $\boldsymbol{X}_i$ is 'larger', 'smaller' if the variance of $\boldsymbol{X}_i$ is 'larger'.

Example 3.4. Suppose there is an unobserved variable Θ_i that determines the risk. Suppose $\mathbb{E}[\boldsymbol{X}_i \mid \Theta_i] = \boldsymbol{Y}_i\boldsymbol{b}(\Theta_i)$ and $\operatorname{Var}[\boldsymbol{X}_i \mid \Theta_i] = \boldsymbol{P}_i\boldsymbol{V}_i(\Theta_i)$ for some known $\boldsymbol{Y}_i \in \mathbb{R}^{m\times q}$, $\boldsymbol{P}_i \in \mathbb{R}^{m\times m}$ and some functions $\boldsymbol{b}(\theta) \in \mathbb{R}^q$, $\boldsymbol{V}_i(\theta) \in \mathbb{R}^{m\times m}$ of the unobservable random variable Θ_i. We assume that $\boldsymbol{P}_i$ is invertible. The Bühlmann and the Bühlmann-Straub models are special cases.

We here are interested to estimate $\boldsymbol{b}(\Theta_i)$. Introduce the following quantities $\boldsymbol{\beta} = \mathbb{E}[\boldsymbol{b}(\Theta_i)]$, $\boldsymbol{\Lambda} = \operatorname{Var}[\boldsymbol{b}(\Theta_i)]$ and $\boldsymbol{\Phi}_i = \boldsymbol{P}_i\mathbb{E}[\boldsymbol{V}_i(\Theta_i)]$. The moments required are $\boldsymbol{\beta}$, $\mathbb{E}[\boldsymbol{X}_i] = \boldsymbol{Y}_i\boldsymbol{\beta}$,

$$\begin{aligned}\operatorname{Cov}[\boldsymbol{b}(\Theta_i), \boldsymbol{X}_i^\top] &= \operatorname{Cov}[\mathbb{E}[\boldsymbol{b}(\Theta_i) \mid \Theta_i], \mathbb{E}[\boldsymbol{X}_i^\top \mid \Theta_i]] + \mathbb{E}[\operatorname{Cov}[\boldsymbol{b}(\Theta_i), \boldsymbol{X}_i^\top \mid \Theta_i]] \\ &= \operatorname{Cov}[\boldsymbol{b}(\Theta_i), \boldsymbol{b}(\Theta_i)^\top \boldsymbol{Y}_i^\top] = \operatorname{Cov}[\boldsymbol{b}(\Theta_i), \boldsymbol{b}(\Theta_i)^\top]\boldsymbol{Y}_i^\top = \boldsymbol{\Lambda}\boldsymbol{Y}_i^\top\,,\end{aligned}$$

and

$$\operatorname{Var}[\boldsymbol{X}_i] = \operatorname{Var}[\mathbb{E}[\boldsymbol{X}_i \mid \Theta_i]] + \mathbb{E}[\operatorname{Var}[\boldsymbol{X}_i \mid \Theta_i]] = \boldsymbol{Y}_i\boldsymbol{\Lambda}\boldsymbol{Y}_i^\top + \boldsymbol{\Phi}_i\,.$$

We assume that $\mathbb{E}[\boldsymbol{V}(\Theta_i)]$ and therefore $\boldsymbol{\Phi}_i$ is invertible. This yields the solution

$$\bar{\boldsymbol{b}}_i = \boldsymbol{\Lambda}\boldsymbol{Y}_i^\top(\boldsymbol{Y}_i\boldsymbol{\Lambda}\boldsymbol{Y}_i^\top + \boldsymbol{\Phi}_i)^{-1}\boldsymbol{X}_i + (\boldsymbol{I} - \boldsymbol{\Lambda}\boldsymbol{Y}_i^\top(\boldsymbol{Y}_i\boldsymbol{\Lambda}\boldsymbol{Y}_i^\top + \boldsymbol{\Phi}_i)^{-1}\boldsymbol{Y}_i)\boldsymbol{\beta}\,. \tag{3.10}$$

Note that $\boldsymbol{\Lambda}$ is symmetric. If $\boldsymbol{\Lambda}$ would not be invertible could we write some coordinates of $\boldsymbol{b}(\Theta_i)$ as a linear function of the others. Thus we assume that $\boldsymbol{\Lambda}$ is invertible. A well know formula gives

$$(\boldsymbol{Y}_i\boldsymbol{\Lambda}\boldsymbol{Y}_i^\top + \boldsymbol{\Phi}_i)^{-1} = \boldsymbol{\Phi}_i^{-1} - \boldsymbol{\Phi}_i^{-1}\boldsymbol{Y}_i(\boldsymbol{\Lambda}^{-1} + \boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1}\boldsymbol{Y}_i)^{-1}\boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1}\,.$$

Define the matrix

$$\boldsymbol{Z}_i = \boldsymbol{\Lambda}\boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1}\boldsymbol{Y}_i(\boldsymbol{\Lambda}^{-1} + \boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1}\boldsymbol{Y}_i)^{-1}\boldsymbol{\Lambda}^{-1} = \boldsymbol{\Lambda}\boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1}\boldsymbol{Y}_i(\boldsymbol{\Lambda}\boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1}\boldsymbol{Y}_i + \boldsymbol{I})^{-1}\,.$$

Then we find

$$\boldsymbol{Z}_i = (\boldsymbol{\Lambda}\boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1}\boldsymbol{Y}_i + \boldsymbol{I} - \boldsymbol{I})(\boldsymbol{\Lambda}\boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1}\boldsymbol{Y}_i + \boldsymbol{I})^{-1} = \boldsymbol{I} - (\boldsymbol{\Lambda}\boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1}\boldsymbol{Y}_i + \boldsymbol{I})^{-1}\,.$$

We also find

$$\begin{aligned}
&\boldsymbol{\Lambda}\boldsymbol{Y}_i^\top(\boldsymbol{Y}_i\boldsymbol{\Lambda}\boldsymbol{Y}_i^\top + \boldsymbol{\Phi}_i)^{-1}\boldsymbol{Y}_i \\
&\qquad = \boldsymbol{\Lambda}\boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1}\boldsymbol{Y}_i - \boldsymbol{\Lambda}\boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1}\boldsymbol{Y}_i(\boldsymbol{\Lambda}^{-1} + \boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1}\boldsymbol{Y}_i)^{-1}\boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1}\boldsymbol{Y}_i \\
&\qquad = (\boldsymbol{I} - \boldsymbol{Z}_i)\boldsymbol{\Lambda}\boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1}\boldsymbol{Y}_i = (\boldsymbol{\Lambda}\boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1}\boldsymbol{Y}_i + \boldsymbol{I})^{-1}\boldsymbol{\Lambda}\boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1}\boldsymbol{Y}_i = \boldsymbol{Z}_i\,.
\end{aligned}$$

It follows that we can write the LB estimator as

$$\bar{\boldsymbol{b}}_i = (\boldsymbol{I} - \boldsymbol{Z}_i)(\boldsymbol{\Lambda}\boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1}\boldsymbol{X}_i + \boldsymbol{\beta})\,. \tag{3.11}$$

The matrix $\boldsymbol{Z}_i$ is called the **credibility matrix**.

Suppose that in addition $\boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1}\boldsymbol{Y}_i$ is invertible. This is equivalent to that $\boldsymbol{Y}_i$ has full rank q and $m \ge q$. Then

$$(\boldsymbol{\Lambda}\boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1}\boldsymbol{Y}_i + \boldsymbol{I})(\boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1}\boldsymbol{Y}_i)^{-1}\boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1} = (\boldsymbol{\Lambda} + (\boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1}\boldsymbol{Y}_i)^{-1})\boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1}\,.$$

Multiplying by $(\boldsymbol{\Lambda}\boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1}\boldsymbol{Y}_i + \boldsymbol{I})^{-1}$ from the left gives

$$(\boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1}\boldsymbol{Y}_i)^{-1}\boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1} = (\boldsymbol{\Lambda}\boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1}\boldsymbol{Y}_i + \boldsymbol{I})^{-1}(\boldsymbol{\Lambda} + (\boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1}\boldsymbol{Y}_i)^{-1})\boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1}\,.$$

Rearranging the term yields

$$\begin{aligned}
(\boldsymbol{I} - \boldsymbol{Z}_i)\boldsymbol{\Lambda}\boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1} &= (\boldsymbol{\Lambda}\boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1}\boldsymbol{Y}_i + \boldsymbol{I})^{-1}\boldsymbol{\Lambda}\boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1} \\
&= (\boldsymbol{I} - (\boldsymbol{\Lambda}\boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1}\boldsymbol{Y}_i + \boldsymbol{I})^{-1})(\boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1}\boldsymbol{Y}_i)^{-1}\boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1} \\
&= \boldsymbol{Z}_i(\boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1}\boldsymbol{Y}_i)^{-1}\boldsymbol{Y}_i^\top\boldsymbol{\Phi}_i^{-1}\,.
\end{aligned}$$

If we define

$$\hat{\boldsymbol{b}}_i = (\boldsymbol{Y}_i^\top \boldsymbol{\Phi}_i^{-1} \boldsymbol{Y}_i)^{-1} \boldsymbol{Y}_i^\top \boldsymbol{\Phi}_i^{-1} \boldsymbol{X}_i$$

we can express the estimator in credibility weighted form

$$\bar{\boldsymbol{b}}_i = \boldsymbol{Z}_i \hat{\boldsymbol{b}}_i + (\boldsymbol{I} - \boldsymbol{Z}_i)\boldsymbol{\beta}\,.$$

■

3.5. Hilbert Space Methods

In the section before we considered the space of quadratic integrable random variables $\mathcal{L}^2$. Choose the inner product $\langle \boldsymbol{X}, \boldsymbol{Y} \rangle = \mathbb{E}[\boldsymbol{X}^\top \boldsymbol{Y}]$. The problem was to minimise $\|\boldsymbol{M} - \boldsymbol{m}\|$, where $\boldsymbol{M}$ was an estimator from some linear subspace of estimators. We now want to consider the problem in the Hilbert space $(\mathcal{L}, \langle \cdot, \cdot \rangle)$.

Let $\boldsymbol{m} \in \mathcal{L}^2$ be an unknown quantity. Let $\mathcal{L} \subset \mathcal{L}^2$ be some subset. We now want to minimise $\rho(\boldsymbol{M}) = \|\boldsymbol{M} - \boldsymbol{m}\|^2$ over all estimators $\boldsymbol{M} \in \mathcal{L}$. If $\mathcal{L}$ is a closed subspace of $\mathcal{L}^2$ then the optimal estimator, called the **$\mathcal{L}$-Bayes estimator**, is the projection $\boldsymbol{m}_\mathcal{L} = \operatorname{pro}(\boldsymbol{m} \mid \mathcal{L})$. Because $\boldsymbol{m} - \boldsymbol{m}_\mathcal{L} \in \mathcal{L}^\perp$ we have $\rho(\boldsymbol{M}) = \|\boldsymbol{m} - \boldsymbol{m}_\mathcal{L}\|^2 + \|\boldsymbol{m}_\mathcal{L} - \boldsymbol{M}\|^2$ which clearly is minimised by choosing $\boldsymbol{M} = \boldsymbol{m}_\mathcal{L}$.

If $\mathcal{L}' \subset \mathcal{L}$ is a closed linear subspace then iterated projections give $\boldsymbol{m}_{\mathcal{L}'} = \operatorname{pro}(\boldsymbol{m}_\mathcal{L} \mid \mathcal{L}')$. More generally, if $\{\boldsymbol{0}\} = \mathcal{L}_0 \subset \mathcal{L}_1 \subset \cdots \subset \mathcal{L}_n = \mathcal{L}^2$ is a nested family of closed linear subspaces of $\mathcal{L}^2$. Then we have $\boldsymbol{m} = \boldsymbol{m}_{\mathcal{L}_n}$ and

$$\boldsymbol{m}_{\mathcal{L}_k} = \sum_{j=1}^{k} (\boldsymbol{m}_{\mathcal{L}_j} - \boldsymbol{m}_{\mathcal{L}_{j-1}})\,.$$

For exampe if $\mathcal{L}_1$ is the space of constants and $\mathcal{L}_2$ is the space of linear estimators we find $\boldsymbol{m}_{\mathcal{L}_1} = \mathbb{E}[\boldsymbol{m}_i]$ and

$$\boldsymbol{m}_{\mathcal{L}_2} - \boldsymbol{m}_{\mathcal{L}_1} = \operatorname{Cov}[\boldsymbol{m}_i, \boldsymbol{X}_I^\top] \operatorname{Var}[\boldsymbol{X}_i]^{-1} (\boldsymbol{X}_i - \mathbb{E}[\boldsymbol{X}_i])\,.$$

Bibliographical Remarks

Credibility theory originated in North America in the early part of the twentieth century. These early ideas are now known as *American credibility.* References can be found in the survey paper by Norberg [63]. The Bayesian approach to credibility is generally attributed to Bailey [14] and [15]. The empirical Bayes approach is due to Bühlmann [20] and Bühlmann and Straub [21].

4. The Cramér-Lundberg Model

4.1. Definition of the Cramér-Lundberg Process

We have seen that the compound Poisson model has nice properties. For instance it can be derived as a limit of individual models. This was the reason for Filip Lundberg to postulate a continuous time risk model where the aggregate claims in any interval have a compound Poisson distribution. Moreover, the premium income should be modelled. In a portfolio of insurance contracts the premium payments will be spread all over the year. Thus he assumed that that the premium income is continuous over time and that the premium income in any time interval is proportional to the interval length. This leads to the following model for the surplus of an insurance portfolio

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

u is the initial capital, c is the premium rate. The number of claims in $(0,t]$ is a Poisson process $\{N_t\}$ with rate λ. The claim sizes $\{Y_i\}$ are a sequence of iid. positive random variables independent of $\{N_t\}$. This model is called **Cramér-Lundberg process** or **classical risk process**.

We denote the distribution function of the claims by G, its moments by $\mu_n = \mathbb{E}[Y_1^n]$ and its moment generating function by $M_Y(r) = \mathbb{E}[\exp\{rY_1\}]$. Let $\mu = \mu_1$. We assume that $\mu < \infty$. Otherwise no insurance company would insure such a risk. Note that $G(x) = 0$ for $x < 0$. We will see later that it is no restriction to assume that $G(0) = 0$.

For an insurance company it is important that $\{C_t\}$ stays above a certain level. This level is given by legal restrictions. By adjusting the initial capital it is no loss of generality to assume this level to be 0. We define the ruin time

$$\tau = \inf\{t > 0 : C_t < 0\}\,, \qquad\qquad (\inf\emptyset = \infty)\,.$$

We will mostly be interested in the probability of ruin in a time interval $(0,t]$

$$\psi(u,t) = \mathbb{P}[\tau \le t \mid C_0 = u] = \mathbb{P}[\inf_{0<s\le t} C_s < 0 \mid C_0 = u]$$

and the probability of ultimate ruin

$$\psi(u) = \lim_{t\to\infty} \psi(u,t) = \mathbb{P}[\inf_{t>0} C_t < 0 \mid C_0 = u]\,.$$

It is easy to see that $\psi(u,t)$ is decreasing in u and increasing in t.

We denote the claim times by $T_1, T_2, \dots$ and by convention $T_0 = 0$. Let $X_i = c(T_i - T_{i-1}) - Y_i$. If we only consider the process at the claim times we can see that

$$C_{T_n} = u + \sum_{i=1}^{n} X_i$$

is a random walk. Note that $\psi(u) = \mathbb{P}[\inf_{n\in\mathbb{N}} C_{T_n} < 0]$. From the theory of random walks we can see that ruin occurs a.s. iff $\mathbb{E}[X_i] \le 0$, compare with Lemma E.1 or [40, p.396]. Hence we will assume in the sequel that

$$\mathbb{E}[X_i] > 0 \iff c\frac{1}{\lambda} - \mu > 0 \iff c > \lambda\mu \iff \mathbb{E}[C_t - u] > 0\,.$$

Recall that

$$\mathbb{E}\Bigl[\sum_{i=1}^{N_t} Y_i\Bigr] = \lambda t \mu\,.$$

The condition can be interpreted that the mean income is strictly larger than the mean outflow. Therefore the condition is also called the **net profit condition**.

If the net profit condition is fulfilled then C_{T_n} tends to infinity as $n \to \infty$. Hence

$$\inf\{C_t - u : t > 0\} = \inf\{C_{T_n} - u : n \ge 1\}$$

is a.s. finite. So we can conclude that

$$\lim_{u\to\infty} \psi(u) = 0\,.$$

4.2. A Note on the Model and Reality

In reality the mean number of claims in an interval will not be the same all the time. There will be a claim rate $\lambda(t)$ which may be periodic in time. Moreover, the number of individual contracts in the portfolio may vary with time. Let $a(t)$ be the volume of the portfolio at time t. Then the claim number process $\{N_t\}$ is an inhomogeneous Poisson process with rate $a(t)\lambda(t)$. Let

$$\Lambda(t) = \int_0^t a(s)\lambda(s)\,\mathrm{d}s$$

and $\Lambda^{-1}(t)$ be its inverse function. Then $\tilde{N}_t = N_{\Lambda^{-1}(t)}$ is a Poisson process with rate 1. Let now the premium rate vary with t such that $c_t = ca(t)\lambda(t)$ for some constant c. This assumption is natural for changes in the risk volume. It is artificial for the changes in the intensity. For instance we assume that the company gets more new customers at times with a higher intensity. This effect may arise because customers withdraw from their old insurance contracts and write new contracts with another company after claims occurred because they were not satisfied by the handling of claims by their old companies.

The premium income in the interval $(0,t]$ is $c\Lambda(t)$. Let $\tilde{C}_t = C_{\Lambda^{-1}(t)}$. Then

$$\tilde{C}_t = u + c\Lambda(\Lambda^{-1}(t)) - \sum_{i=1}^{N_{\Lambda^{-1}(t)}} Y_i = u + ct - \sum_{i=1}^{\tilde{N}_t} Y_i$$

is a Cramér-Lundberg process. Thus we should not consider time to be the real time but **operational time**.

The event of ruin does almost never occur in practice. If an insurance company observes that their surplus is decreasing they will immediately increase their premia. On the other hand an insurance company is built up on different portfolios. Ruin in one portfolio does not mean bankruptcy. Therefore ruin is only a technical term. The probability of ruin is used for decision taking, for instance the premium calculation or the computation of reinsurance retention levels. For an actuary it is important to be able to take a good decision in reasonable time. Therefore it is fine to simplify the model in order to be able to check whether a decision has the desired effect or not.

The surplus will also be a technical term in practice. If the business is going well then the share holders will decide to get a higher dividend. To model this we would have to assume a premium rate dependent on the surplus. But then it would be hard to obtain any useful results. For some references see for instance [42] and [11].

In Section 1 about risk models we saw that a negative binomial distribution for the number of claims in a certain time interval would be preferable. We will later consider such models. But actuaries use the Cramér-Lundberg model for their calculations, even though there is a certain lack of reality. The reason is that this is more or less the only model which they can handle.

4.3. A Differential Equation for the Ruin Probability

We first prove that $\{C_t\}$ is a strong Markov process.

Lemma 4.1. *Let $\{C_t\}$ be a Cramér-Lundberg process and T be a finite stopping time. Then the stochastic process $\{C_{T+t} - C_T : t \geq 0\}$ is a Cramér-Lundberg process with initial capital 0 and independent of $\mathcal{F}_T$.*

Proof. We can write $C_{T+t} - C_T$ as

$$ct - \sum_{i=N_T+1}^{N_{T+t}} Y_i \, .$$

Because the claim amounts are iid. and independent of $\{N_t\}$ we only have to prove that $\{N_{T+t} - N_T\}$ is a Poisson process independent of $\mathcal{F}_T$. Because $\{N_t\}$ is a renewal process it is enough to show that $T_{N_T+1} - T$ is $\mathrm{Exp}(\lambda)$ distributed and independent of $\mathcal{F}_T$. Condition on T_{N_T} and T. Then

$$\begin{aligned}&\mathbb{P}[T_{N_T+1} - T > x \mid T_{N_T}, T] \\ &\qquad = \mathbb{P}[T_{N_T+1} - T_{N_T} > x + T - T_{N_T} \mid T_{N_T}, T, T_{N_T+1} - T_{N_T} > T - T_{N_T}] = \mathrm{e}^{-\lambda x}\end{aligned}$$

by the lack of memory property of the exponential distribution. The assertion follows because T_{N_T+1} depends on $\mathcal{F}_T$ via T_{N_T} and T only. The latter, even though intuitively clear, follows readily noting that $\mathcal{F}_T$ is generated by sets of the form $\{N_T = n, A_n\}$ where $n \in \mathbb{N}$ and $A_n \in \sigma(Y_1, \ldots, Y_n, T_1, \ldots, T_n)$. □

Let h be small. If ruin does not occur in the interval $(0, T_1 \wedge h]$ then a new Cramér-Lundberg process starts at time $T_1 \wedge h$ with new initial capital $C_{T_1 \wedge h}$. Let $\delta(u) = 1 - \psi(u)$ denote the probability that ruin does not occur. Using that the interarrival times are exponentially distributed we have the density $\lambda \mathrm{e}^{-\lambda t}$ of the distribution of T_1 and $\mathbb{P}[T_1 > h] = \mathrm{e}^{-\lambda h}$. We obtain noting that $\delta(x) = 0$ for $x < 0$

$$\delta(u) = \mathrm{e}^{-\lambda h} \delta(u + ch) + \int_0^h \int_0^{u+ct} \delta(u + ct - y) \; \mathrm{d}G(y) \lambda \mathrm{e}^{-\lambda t} \; \mathrm{d}t \, .$$

Letting h tending to 0 shows that $\delta(u)$ is right continuous. Rearranging the terms and dividing by h yields

$$c\,\frac{\delta(u + ch) - \delta(u)}{ch} = \frac{1 - \mathrm{e}^{-\lambda h}}{h} \delta(u + ch) - \frac{1}{h} \int_0^h \int_0^{u+ct} \delta(u + ct - y) \; \mathrm{d}G(y) \lambda \mathrm{e}^{-\lambda t} \; \mathrm{d}t \, .$$

Letting h tend to 0 shows that $\delta(u)$ is differentiable from the right and

$$c\delta'(u) = \lambda \Bigl[\delta(u) - \int_0^u \delta(u - y) \; \mathrm{d}G(y) \Bigr] \, . \tag{4.1}$$

Replacing u by $u - ch$ gives

$$\delta(u - ch) = \mathrm{e}^{-\lambda h} \delta(u) + \int_0^h \int_0^{u - c(h-t)} \delta(u - (c - h)t - y) \; \mathrm{d}G(y) \lambda \mathrm{e}^{-\lambda t} \; \mathrm{d}t \, .$$

We conclude that $\delta(u)$ is continuous. Rearranging the terms, dividing by h and letting $h \downarrow 0$ yields the equation

$$c\delta'(u) = \lambda\Bigl[\delta(u) - \int_0^{u-} \delta(u-y)\,\mathrm{d}G(y)\Bigr]\,.$$

We see that $\delta(u)$ is differentiable at all points u where $G(u)$ is continuous. If $G(u)$ has a jump then the derivatives from the left and from the right do not coincide. But we have shown that $\delta(u)$ is absolutely continuous. If we now let $\delta'(u)$ be the derivative from the right we have that $\delta'(u)$ is a density of $\delta(u)$.

The difficulty with equation (4.1) is that it contains both the derivative of $\delta(u)$ and an integral. Let us try to get rid of the derivative. We find

$$\begin{aligned}
\frac{c}{\lambda}(\delta(u)-\delta(0)) = \frac{1}{\lambda}\int_0^u c\delta'(x)\,\mathrm{d}x &= \int_0^u \delta(x)\,\mathrm{d}x - \int_0^u\!\int_0^x \delta(x-y)\,\mathrm{d}G(y)\,\mathrm{d}x \\
&= \int_0^u \delta(x)\,\mathrm{d}x - \int_0^u\!\int_y^u \delta(x-y)\,\mathrm{d}x\,\mathrm{d}G(y) \\
&= \int_0^u \delta(x)\,\mathrm{d}x - \int_0^u\!\int_0^{u-y} \delta(x)\,\mathrm{d}x\,\mathrm{d}G(y) \\
&= \int_0^u \delta(x)\,\mathrm{d}x - \int_0^u\!\int_0^{u-x} \mathrm{d}G(y)\,\delta(x)\,\mathrm{d}x \\
&= \int_0^u \delta(x)(1-G(u-x))\,\mathrm{d}x = \int_0^u \delta(u-x)(1-G(x))\,\mathrm{d}x
\end{aligned}$$

Note that $\int_0^\infty (1-G(x))\,\mathrm{d}x = \mu$. Letting $u \to \infty$ we can by the bounded convergence theorem interchange limit and integral and get

$$c(1-\delta(0)) = \lambda\int_0^\infty (1-G(x))\,\mathrm{d}x = \lambda\mu$$

where we used that $\delta(u) \to 1$. It follows that

$$\delta(0) = 1 - \frac{\lambda\mu}{c}, \qquad \psi(0) = \frac{\lambda\mu}{c}\,.$$

Replacing $\delta(u)$ by $1-\psi(u)$ we obtain

$$\begin{aligned}
c\psi(u) &= \lambda\mu - \lambda\int_0^u (1-\psi(u-x))(1-G(x))\,\mathrm{d}x \\
&= \lambda\Bigl(\int_u^\infty (1-G(x))\,\mathrm{d}x + \int_0^u \psi(u-x)(1-G(x))\,\mathrm{d}x\Bigr)\,. \qquad (4.2)
\end{aligned}$$

In Section 4.8 we shall obtain a natural interpretation of (4.2).

Example 4.2. Let the claims be $\text{Exp}(\alpha)$ distributed. Then the equation (4.1) can be written as

$$c\delta'(u) = \lambda\Big[\delta(u) - \mathrm{e}^{-\alpha u}\int_0^u \delta(y)\alpha\mathrm{e}^{\alpha y}\,\mathrm{d}y\Big]\,.$$

Differentiating yields

$$c\delta''(u) = \lambda\Big[\delta'(u) + \alpha\mathrm{e}^{-\alpha u}\int_0^u \delta(y)\alpha\mathrm{e}^{\alpha y}\,\mathrm{d}y - \alpha\delta(u)\Big] = \lambda\delta'(u) - \alpha c\delta'(u)\,.$$

The solution to this differential equation is

$$\delta(u) = A + B\mathrm{e}^{-(\alpha-\lambda/c)u}\,.$$

Because $\delta(u) \to 1$ as $u \to \infty$ we get $A = 1$. Because $\delta(0) = 1 - \lambda/(\alpha c)$ the solution is

$$\delta(u) = 1 - \frac{\lambda}{\alpha c}\mathrm{e}^{-(\alpha-\lambda/c)u}$$

or

$$\psi(u) = \frac{\lambda}{\alpha c}\mathrm{e}^{-(\alpha-\lambda/c)u}\,.$$

■

4.4. The Adjustment Coefficient

Let

$$\theta(r) = \lambda(M_Y(r) - 1) - cr \tag{4.3}$$

provided $M_Y(r)$ exists. Then we find the following martingale.

Lemma 4.3. *Let $r \in \mathbb{R}$ such that $M_Y(r) < \infty$. Then the stochastic process*

$$\{\exp\{-rC_t - \theta(r)t\}\} \tag{4.4}$$

is a martingale.

Proof. By the Markov property we have for $s < t$

$$\begin{aligned}\mathbb{E}[\mathrm{e}^{-rC_t-\theta(r)t} \mid \mathcal{F}_s] &= \mathbb{E}[\mathrm{e}^{-r(C_t-C_s)}]\mathrm{e}^{-rC_s-(\lambda(M_Y(r)-1)-cr)t}\\ &= \mathbb{E}[\mathrm{e}^{r\sum_{N_s+1}^{N_t} Y_i}]\mathrm{e}^{-rC_s-\lambda(M_Y(r)-1)t+crs}\\ &= \mathrm{e}^{-rC_s-\lambda(M_Y(r)-1)s+crs} = \mathrm{e}^{-rC_s-\theta(r)s}\,.\end{aligned}$$

□

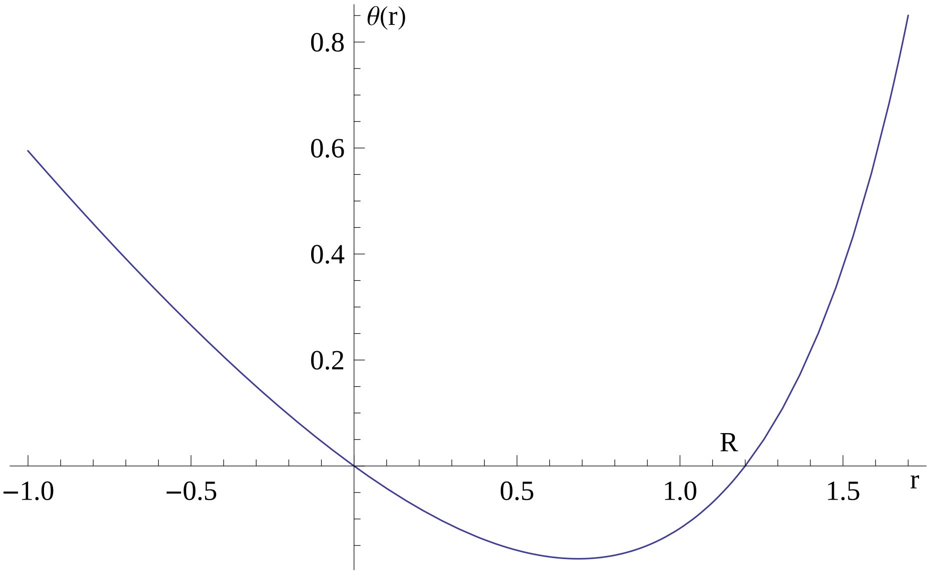


Figure 4.1: *The function $\theta(r)$ and the adjustment coefficient R*

It would be nice to have a martingale that only depends on the surplus and not explicitly on time. Let us therefore consider the equation $\theta(r) = 0$. This equation has obviously the solution $r = 0$. We differentiate $\theta(r)$.

$$\theta'(r) = \lambda M_Y'(r) - c\,,$$

$$\theta''(r) = \lambda M_Y''(r) = \lambda \mathbb{E}[Y^2 e^{rY}] > 0\,.$$

The function $\theta(r)$ is strictly convex. For $r = 0$

$$\theta'(0) = \lambda M_Y'(0) - c = \lambda\mu - c < 0$$

by the net profit condition. There might be at most one additional solution R to the equation $\theta(R) = 0$ and $R > 0$. If this solution exists we call it the **adjustment coefficient** or the **Lundberg exponent**. The Lundberg exponent will play an important rôle in the estimation of the ruin probabilities.

Example 4.2 (continued). For Exp(α) distributed claims we have to solve

$$\lambda\Bigl(\frac{\alpha}{\alpha - r} - 1\Bigr) - cr = \frac{\lambda r}{\alpha - r} - cr = 0\,.$$

For $r \neq 0$ we find $R = \alpha - \lambda/c$. Note that

$$\psi(u) = \frac{\lambda}{\alpha c} e^{-Ru}\,.$$

∎

In general the adjustment coefficient is difficult to calculate. So we try to find some bounds for the adjustment coefficient. Note that the second moment μ_2 exists if the Lundberg exponent exists. Consider the function $\theta(r)$ for $r > 0$.

$$\begin{aligned}\theta''(r) &= \lambda \mathbb{E}[Y^2 \mathrm{e}^{rY}] > \lambda \mathbb{E}[Y^2] = \lambda\mu_2\,,\\ \theta'(r) &= \theta'(0) + \int_0^r \theta''(s)\,\mathrm{d}s > -(c-\lambda\mu) + \lambda\mu_2 r\,,\\ \theta(r) &= \theta(0) + \int_0^r \theta'(s)\,\mathrm{d}s > \lambda\mu_2\frac{r^2}{2} - (c-\lambda\mu)r\,.\end{aligned}$$

The last inequality yields for $r = R$

$$0 = \theta(R) > R\Bigl(\lambda\mu_2\frac{R}{2} - (c-\lambda\mu)\Bigr)$$

from which an upper bound of R

$$R < \frac{2(c-\lambda\mu)}{\lambda\mu_2}$$

follows.

We are not able to find a lower bound in general. But in the case of bounded claims we get a lower bound for R. Assume that $Y_1 \le M$ a.s..

Let us first consider the function

$$f(x) = \frac{x}{M}\left(\mathrm{e}^{RM} - 1\right) - \left(\mathrm{e}^{Rx} - 1\right)\,.$$

Its second derivative is

$$f''(x) = -R^2\mathrm{e}^{Rx} < 0\,.$$

$f(x)$ is concave with $f(0) = f(M) = 0$. Thus $f(x) > 0$ for $0 < x < M$. The function $h(x) = x\mathrm{e}^x - \mathrm{e}^x + 1$ has a minimum in 0. Thus $h(x) > h(0) = 0$ for $x \neq 0$, in particular

$$\frac{1}{RM}\left(\mathrm{e}^{RM} - 1\right) < \mathrm{e}^{RM}\,.$$

We calculate

$$M_Y(R) - 1 = \int_0^M \left(\mathrm{e}^{Rx} - 1\right)\,\mathrm{d}G(x) < \int_0^M \frac{x}{M}\left(\mathrm{e}^{RM} - 1\right)\,\mathrm{d}G(x) = \frac{\mu}{M}\left(\mathrm{e}^{RM} - 1\right)\,.$$

From the equation determining R we get

$$0 = \lambda(M_Y(R) - 1) - cR < \frac{\lambda\mu}{M}\left(\mathrm{e}^{RM} - 1\right) - cR < \lambda\mu R\mathrm{e}^{RM} - cR\,.$$

It follows readily

$$R > \frac{1}{M}\log\frac{c}{\lambda\mu}\,.$$

4.5. Lundberg's Inequality

We will now connect the adjustment coefficient and ruin probabilities.

Theorem 4.4. *Assume that the adjustment coefficient R exists. Then*

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

Proof. Assume that the theorem does not hold. Let

$$u_0 = \inf\{u \ge 0 : \psi(u) \ge \mathrm{e}^{-Ru}\}\,.$$

Because $\psi(u)$ is continuous we get

$$\psi(u_0) = \mathrm{e}^{-Ru_0}\,.$$

Because $\psi(0) < 1$ we conclude that $u_0 > 0$. Consider the equation (4.2) for $u = u_0$.

$$\begin{aligned}
c\psi(u_0) &= \lambda\Bigl[\int_{u_0}^\infty (1-G(x))\,\mathrm{d}x + \int_0^{u_0} \psi(u_0-x)(1-G(x))\,\mathrm{d}x\Bigr]\\
&< \lambda\Bigl[\int_{u_0}^\infty (1-G(x))\,\mathrm{d}x + \int_0^{u_0} \mathrm{e}^{-R(u_0-x)}(1-G(x))\,\mathrm{d}x\Bigr]\\
&\le \lambda\int_0^\infty \mathrm{e}^{-R(u_0-x)}(1-G(x))\,\mathrm{d}x = \lambda\mathrm{e}^{-Ru_0}\int_0^\infty\int_x^\infty \mathrm{e}^{Rx}\,\mathrm{d}G(y)\,\mathrm{d}x\\
&= \lambda\mathrm{e}^{-Ru_0}\int_0^\infty\int_0^y \mathrm{e}^{Rx}\,\mathrm{d}x\,\mathrm{d}G(y) = \lambda\mathrm{e}^{-Ru_0}\int_0^\infty \frac{1}{R}(\mathrm{e}^{Ry}-1)\,\mathrm{d}G(y)\\
&= \frac{\lambda}{R}\mathrm{e}^{-Ru_0}(M_Y(R)-1) = c\mathrm{e}^{-Ru_0}
\end{aligned}$$

which is a contradiction. This proves the theorem. □

We now give an alternative proof of the theorem. We will use the positive martingale (4.4) for $r = R$. By the stopping theorem

$$\mathrm{e}^{-Ru} = \mathrm{e}^{-RC_0} = \mathbb{E}[\mathrm{e}^{-RC_{\tau\wedge t}}] \ge \mathbb{E}[\mathrm{e}^{-RC_{\tau\wedge t}};\tau \le t] = \mathbb{E}[\mathrm{e}^{-RC_\tau};\tau \le t]\,.$$

Letting t tend to infinity yields by monotone convergence

$$\mathrm{e}^{-Ru} \ge \mathbb{E}[\mathrm{e}^{-RC_\tau};\tau < \infty] > \mathbb{P}[\tau < \infty] = \psi(u) \tag{4.5}$$

because $C_\tau < 0$. We have got an upper bound for the ruin probability. We would like to know whether R is the best possible exponent in an exponential upper bound. This question will be answered in the next section.

Example 4.5. Let $G(x) = 1 - p\mathrm{e}^{-\alpha x} - (1-p)\mathrm{e}^{-\beta x}$ where $0 < \alpha < \beta$ and $0 < p \le \beta(\beta-\alpha)^{-1}$. Note that the mean value is $\frac{p}{\alpha} + \frac{1-p}{\beta}$ and thus

$$c > \frac{\lambda p}{\alpha} + \frac{\lambda(1-p)}{\beta}\,.$$

For $r < \alpha$

$$M_Y(r) = \int_0^\infty \mathrm{e}^{rx}\Bigl(\alpha p\mathrm{e}^{-\alpha x} + \beta(1-p)\mathrm{e}^{-\beta x}\Bigr)\,\mathrm{d}x = \frac{\alpha p}{\alpha - r} + \frac{\beta(1-p)}{\beta - r}\,.$$

Thus we have to solve

$$\lambda\Bigl(\frac{\alpha p}{\alpha - r} + \frac{\beta(1-p)}{\beta - r} - 1\Bigr) - cr = \lambda\Bigl(\frac{pr}{\alpha - r} + \frac{(1-p)r}{\beta - r}\Bigr) - cr = 0\,.$$

We find the obvious solution $r = 0$. If $r \neq 0$ then

$$\lambda p(\beta - r) + \lambda(1-p)(\alpha - r) = c(\alpha - r)(\beta - r)$$

or equivalently

$$cr^2 - ((\alpha+\beta)c - \lambda)r + \alpha\beta c - \lambda((1-p)\alpha + p\beta) = 0\,. \tag{4.6}$$

The solution is

$$\begin{aligned} r &= \frac{1}{2}\left(\alpha + \beta - \frac{\lambda}{c} \pm \sqrt{\Bigl(\alpha + \beta - \frac{\lambda}{c}\Bigr)^2 - 4\Bigl(\alpha\beta - \frac{\lambda}{c}p\beta - \frac{\lambda}{c}(1-p)\alpha\Bigr)}\right) \\ &= \frac{1}{2}\left(\alpha + \beta - \frac{\lambda}{c} \pm \sqrt{\Bigl(\beta - \alpha - \frac{\lambda}{c}\Bigr)^2 + 4p\frac{\lambda}{c}(\beta - \alpha)}\right)\,. \end{aligned}$$

Now we got three solutions. But there should only be two. Note that

$$\alpha\beta - \frac{\lambda}{c}p\beta - \frac{\lambda}{c}(1-p)\alpha = \frac{\alpha\beta}{c}\Bigl(c - \lambda\Bigl(\frac{p}{\alpha} + \frac{1-p}{\beta}\Bigr)\Bigr) > 0$$

and thus both solutions are positive. The larger of the solutions can be written as

$$\alpha + \frac{1}{2}\left(\beta - \alpha - \frac{\lambda}{c} + \sqrt{\Bigl(\beta - \alpha - \frac{\lambda}{c}\Bigr)^2 + 4p\frac{\lambda}{c}(\beta - \alpha)}\right)$$

and is thus larger than α. But the moment generating function does not exist for $r \ge \alpha$. Thus

$$R = \frac{1}{2}\left(\alpha + \beta - \frac{\lambda}{c} - \sqrt{\Bigl(\beta - \alpha - \frac{\lambda}{c}\Bigr)^2 + 4p\frac{\lambda}{c}(\beta - \alpha)}\right)\,.$$

From Lundberg's inequality it follows that

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

■

4.6. The Cramér-Lundberg Approximation

Consider the equation (4.2). This equation looks almost like a renewal equation, but

$$\int_0^\infty \frac{\lambda}{c}(1-G(x))\,\mathrm{d}x = \frac{\lambda\mu}{c} < 1$$

is not a probability distribution. Can we manipulate (4.2) to get a renewal equation? We try for some measurable function h and $\bar{h}$

$$\psi(u)h(u) = h(u)\int_u^\infty \frac{\lambda}{c}(1-G(x))\,\mathrm{d}x + \int_0^u \psi(u-x)h(u-x)\bar{h}(x)\frac{\lambda}{c}(1-G(x))\,\mathrm{d}x \quad (4.7)$$

where $h(u) = h(u-x)\bar{h}(x)$. We can assume that $h(0) = 1$. Setting $x = u$ shows that $\bar{h}(u) = h(u)$. From a general theorem it follows that $h(u) = \mathrm{e}^{ru}$ where $r = \log h(1)$. In order that (4.7) is a renewal equation we need

$$1 = \int_0^\infty \mathrm{e}^{rx}\frac{\lambda}{c}(1-G(x))\,\mathrm{d}x = \frac{\lambda}{c}\int_0^\infty\int_x^\infty \mathrm{e}^{rx}\,\mathrm{d}G(y)\,\mathrm{d}x = \frac{\lambda(M_Y(r)-1)}{cr}\,.$$

The only solution to the latter equation is $r = R$. Let us assume that R exists. Moreover, we need that

$$\int_0^\infty x\mathrm{e}^{Rx}\frac{\lambda}{c}(1-G(x))\,\mathrm{d}x < \infty \quad (4.8)$$

which is equivalent to $M_Y'(R) < \infty$, see below. The equation is now

$$\psi(u)\mathrm{e}^{Ru} = \mathrm{e}^{Ru}\int_u^\infty \frac{\lambda}{c}(1-G(x))\,\mathrm{d}x + \int_0^u \psi(u-x)\mathrm{e}^{R(u-x)}\mathrm{e}^{Rx}\frac{\lambda}{c}(1-G(x))\,\mathrm{d}x\,.$$

It can be shown that

$$\mathrm{e}^{Ru}\int_u^\infty \frac{\lambda}{c}(1-G(x))\,\mathrm{d}x$$

is directly Riemann integrable. Thus we get from the renewal theorem

Theorem 4.6. *Assume that the Lundberg exponent exists and that* (4.8) *is fulfilled. Then*

$$\lim_{u\to\infty}\psi(u)\mathrm{e}^{Ru} = \frac{c-\lambda\mu}{\lambda M_Y'(R)-c}\,.$$

Proof. It only remains to compute the limit.

$$\begin{aligned}\int_0^\infty \mathrm{e}^{Ru}\int_u^\infty \frac{\lambda}{c}(1-G(x))\,\mathrm{d}x\,\mathrm{d}u &= \frac{\lambda}{c}\int_0^\infty\int_0^x \mathrm{e}^{Ru}\,\mathrm{d}u\,(1-G(x))\,\mathrm{d}x \\ &= \frac{\lambda}{cR}\int_0^\infty(\mathrm{e}^{Rx}-1)(1-G(x))\,\mathrm{d}x = \frac{1}{R}-\frac{\lambda\mu}{cR} = \frac{1}{cR}(c-\lambda\mu)\,.\end{aligned}$$

The mean value of the distribution is

$$\begin{aligned}\int_0^\infty x\mathrm{e}^{Rx}\frac{\lambda}{c}(1-G(x))\,\mathrm{d}x &= \frac{\lambda}{c}\int_0^\infty\int_x^\infty x\mathrm{e}^{Rx}\,\mathrm{d}G(y)\,\mathrm{d}x\\ &= \frac{\lambda}{c}\int_0^\infty\int_0^y x\mathrm{e}^{Rx}\,\mathrm{d}x\,\mathrm{d}G(y) = \frac{\lambda}{cR^2}\int_0^\infty(Ry\mathrm{e}^{Ry}-\mathrm{e}^{Ry}+1)\,\mathrm{d}G(y)\\ &= \frac{\lambda(RM_Y'(R)-M_Y(R)+1)}{cR^2} = \frac{\lambda RM_Y'(R)-cR}{cR^2} = \frac{\lambda M_Y'(R)-c}{cR}\,.\end{aligned}$$

The limit value follows readily. □

Remark. The theorem shows that it is not possible to obtain an exponential upper bound for the ruin probability with an exponent strictly larger than R. ■

The theorem can be written in the following form

$$\psi(u) \sim \frac{c-\lambda\mu}{\lambda M_Y'(R)-c}\mathrm{e}^{-Ru}\,.$$

Thus for large u we get an approximation to $\psi(u)$. This approximation is called the **Cramér-Lundberg approximation**.

Example 4.2 (continued). From $M_Y(r)$ we get that

$$M_Y'(r) = \frac{\alpha}{(\alpha-r)^2}$$

and thus

$$\lim_{u\to\infty}\psi(u)\mathrm{e}^{Ru} = \frac{c-\frac{\lambda}{\alpha}}{\frac{\lambda\alpha}{(\alpha-R)^2}-c} = \frac{c-\frac{\lambda}{\alpha}}{\frac{\lambda\alpha}{\left(\frac{\lambda}{c}\right)^2}-c} = \frac{\lambda}{\alpha c}\frac{\alpha c-\lambda}{\alpha c-\lambda} = \frac{\lambda}{\alpha c}\,.$$

Hence the Cramér-Lundberg approximation

$$\psi(u) \sim \frac{\lambda}{\alpha c}\mathrm{e}^{-(\alpha-\lambda/c)u}$$

becomes exact in this case. ■

Example 4.7. Let $c = \lambda = 1$ and $G(x) = 1-\frac{1}{3}(\mathrm{e}^{-x}+\mathrm{e}^{-2x}+\mathrm{e}^{-3x})$. The mean value of claim sizes is $\mu = 0.611111$, i.e. the net profit condition is fulfilled. One can show that

$$\psi(u) = 0.550790\,\mathrm{e}^{-0.485131u}+0.0436979\,\mathrm{e}^{-1.72235u}+0.0166231\,\mathrm{e}^{-2.79252u}\,.$$

From Theorem 4.6 it follows that $app(u) = 0.550790\,\mathrm{e}^{-0.485131u}$ is the Cramér-Lundberg approximation to $\psi(u)$. Table 4.1 below shows the ruin function $\psi(u)$, its Cramér-Lundberg approximation $app(u)$ and the relative error $(app(u)-\psi(u))/\psi(u)$ multiplied by 100 (Er). Note that the the relative error is below 1% for $u \ge 1.71358 = 2.8\mu$. ■

u	0	0.25	0.5	0.75	1	1.25	1.5	1.75	2	2.25
$\psi(u)$	0.6111	0.5246	0.4547	0.3969	0.3479	0.3059	0.2696	0.2379	0.2102	0.1858
$app(u)$	0.5508	0.4879	0.4322	0.3828	0.3391	0.3003	0.2660	0.2357	0.2087	0.1849
Er	-9.87	-6.99	-4.97	-3.54	-2.54	-1.82	-1.32	-0.95	-0.69	-0.50

Table 4.1: *Cramér-Lundberg approximation to ruin probabilities*

4.7. Reinsurance and Ruin

4.7.1. Proportional Reinsurance

Recall that for proportional insurance the insurer covers $Y_i^I = \alpha Y_i$ of each claim, the reinsurer covers $Y_i^R = (1-\alpha)Y_i$. Denote by c^I the insurer's premium rate. The insurer's adjustment coefficient is obtained from the equation

$$\lambda(M_{\alpha Y}(r) - 1) - c^I r = 0\,.$$

The new moment generating function is

$$M_{\alpha Y}(r) = \mathbb{E}[\mathrm{e}^{r\alpha Y_i}] = M_Y(\alpha r)\,.$$

Assume that both insurer and reinsurer use an expected value premium principle with the same safety loading. Then $c^I = \alpha c$ and we have to solve

$$\lambda(M_Y(\alpha r) - 1) - c\alpha r = 0\,.$$

This is almost the original equation, hence $R = \alpha R^I$ where R^I is the adjustment coefficient under reinsurance. The new adjustment coefficient is larger, hence the risk has become smaller.

Example 4.8. Let $Y_i \sim \mathrm{Exp}(\beta)$ and let the gross premium rate be fixed $(1+\xi)\lambda\mathbb{E}[Y] = (1+\xi)\frac{\lambda}{\beta}$. We want to reinsure the risk via proportional reinsurance with retention level α. The reinsurer charges the premium $(1+\vartheta)(1-\alpha)\frac{\lambda}{\beta}$. We assume that $\vartheta \ge \xi$, otherwise the insurer could choose $\alpha = 0$ and he would make a profit without any risk. How shall we choose α in order to maximize the insurer's adjustment coefficient? The case $\vartheta = \xi$ had been considered before. We only have to consider the case $\vartheta > \xi$. Because the net profit condition must be fulfilled there is a lower bound for α.

$$c^I = (1+\xi)\frac{\lambda}{\beta} - (1+\vartheta)(1-\alpha)\frac{\lambda}{\beta} = (\alpha(1+\vartheta) - (\vartheta-\xi))\frac{\lambda}{\beta} > \alpha\frac{\lambda}{\beta}\,.$$

Thus

$$\alpha > \frac{\vartheta-\xi}{\vartheta} = 1 - \frac{\xi}{\vartheta}\,. \tag{4.9}$$

The moment generating function of αY is

$$M_{\alpha Y}(r) = M_Y(\alpha r) = \frac{\beta}{\beta - \alpha r} = \frac{\frac{\beta}{\alpha}}{\frac{\beta}{\alpha} - r}\,.$$

The claims after reinsurance are $\mathrm{Exp}(\beta/\alpha)$ distributed and the adjustment coefficient is

$$R^I(\alpha) = \frac{\beta}{\alpha} - \frac{\lambda}{(\alpha(1+\vartheta) - (\vartheta-\xi))\frac{\lambda}{\beta}} = \beta\Bigl(\frac{1}{\alpha} - \frac{1}{\alpha(1+\vartheta) - (\vartheta-\xi)}\Bigr)\,.$$

Setting the derivative to 0 yields

$$\alpha^2 = \frac{(\alpha(1+\vartheta) - (\vartheta-\xi))^2}{1+\vartheta} = (1+\vartheta)\alpha^2 - 2\alpha(\vartheta-\xi) + \frac{(\vartheta-\xi)^2}{1+\vartheta}\,.$$

The solution to this equation is

$$\begin{aligned}\alpha &= \frac{1}{\vartheta}\Bigl(\vartheta - \xi \pm \sqrt{(\vartheta-\xi)^2 - \frac{\vartheta}{1+\vartheta}(\vartheta-\xi)^2}\Bigr)\\ &= \Bigl(1-\frac{\xi}{\vartheta}\Bigr)\Bigl(1 \pm \sqrt{1 - \frac{\vartheta}{1+\vartheta}}\Bigr) = \Bigl(1-\frac{\xi}{\vartheta}\Bigr)\Bigl(1 + \sqrt{\frac{1}{1+\vartheta}}\Bigr)\end{aligned}$$

where the last equality follows from (4.9). This solution is strictly larger than 1 if $\vartheta > \xi^2 + 2\xi$. Thus the solution is

$$\alpha_{\max} = \min\Bigl\{\Bigl(1-\frac{\xi}{\vartheta}\Bigr)\Bigl(1 + \sqrt{\frac{1}{1+\vartheta}}\Bigr), 1\Bigr\}\,.$$

By plugging in the values $1 - \frac{\xi}{\vartheta}$ and 1 into the derivative $R^{I\prime}(\alpha)$ shows that indeed $R^I(\alpha)$ has a maximum at $\alpha_{\max}$. ■

4.7.2. Excess of Loss Reinsurance

Under excess of loss reinsurance with retention level M the insurer has to pay $Y_i^I = \min\{Y_i, M\}$ of each claim. The adjustment coefficient is the strictly positive solution to the equation

$$\lambda\Bigl(\int_0^M \mathrm{e}^{rx}\,\mathrm{d}G(x) + \mathrm{e}^{rM}(1 - G(M)) - 1\Bigr) - c^I r = 0\,.$$

There is no possibility to find the solution from the problem without reinsurance. We have to solve the equation for every M separately. But note that R^I exists in any case. Especially for heavy tailed distributions this shows that the risk has become much smaller. By the Cramér-Lundberg approximation the ruin probability decreases exponentially as the initial capital increases.

We will now show that for the insurer the excess of loss reinsurance is optimal. We assume that both insurer and reinsurer use an expected value principle.

Proposition 4.9. *Let all premia be computed via the expected value principle. Under all reinsurance forms acting on individual claims with premium rates c^I and c^R fixed the excess of loss reinsurance maximizes the insurer's adjustment coefficient.*

Proof. Let $h(x)$ be an increasing function with $0 \le h(x) \le x$ for $x \ge 0$. We assume that the insurer pays $Y_i^I = h(Y_i)$. Let $h^*(x) = \min\{x, U\}$ be the excess of loss reinsurance. Because c^I is fixed U can be determined from

$$\int_0^U y \,\mathrm{d}G(y) + U(1 - G(U)) = \mathbb{E}[h^*(Y_i)] = \mathbb{E}[h(Y_i)] = \int_0^\infty h(y)\,\mathrm{d}G(y)\,.$$

Because $\mathrm{e}^z \ge 1 + z$ we obtain

$$\mathrm{e}^{r(h(y)-h^*(y))} \ge 1 + r(h(y) - h^*(y))$$

and

$$\mathrm{e}^{rh(y)} \ge \mathrm{e}^{rh^*(y)}(1 + r(h(y) - h^*(y)))\,.$$

Thus

$$\begin{aligned} M_{h(Y)}(r) &= \int_0^\infty \mathrm{e}^{rh(y)}\,\mathrm{d}G(y) \ge \int_0^\infty \mathrm{e}^{rh^*(y)}(1 + r(h(y) - h^*(y)))\,\mathrm{d}G(y) \\ &= M_{h^*(Y)}(r) + r\int_0^\infty (h(y) - h^*(y))\,\mathrm{e}^{rh^*(y)}\,\mathrm{d}G(y)\,. \end{aligned}$$

For $y \le U$ we have $h(y) \le y = h^*(y)$. We obtain for $r > 0$

$$\begin{aligned} &\int_0^\infty (h(y) - h^*(y))\mathrm{e}^{rh^*(y)}\,\mathrm{d}G(y) \\ &\quad= \int_0^U (h(y) - h^*(y))\mathrm{e}^{rh^*(y)}\,\mathrm{d}G(y) + \int_U^\infty (h(y) - h^*(y))\mathrm{e}^{rh^*(y)}\,\mathrm{d}G(y) \\ &\quad\ge \int_0^U (h(y) - h^*(y))\mathrm{e}^{rU}\,\mathrm{d}G(y) + \int_U^\infty (h(y) - h^*(y))\mathrm{e}^{rU}\,\mathrm{d}G(y) \\ &\quad= \mathrm{e}^{rU}\int_0^\infty (h(y) - h^*(y))\,\mathrm{d}G(y) \\ &\quad= \mathrm{e}^{rU}(\mathbb{E}[h(Y)] - \mathbb{E}[h^*(Y)]) = 0\,. \end{aligned}$$

It follows that $M_{h(Y)}(r) \ge M_{h^*(Y)}(r)$ for $r > 0$. Since

$$0 = \theta(R^I) = \lambda(M_{h(Y)}(R^I) - 1) - c^I R^I \ge \lambda(M_{h^*(Y)}(R^I) - 1) - c^I R^I = \theta^*(R^I)$$

and the assertion follows from the convexity of $\theta^*(r)$. ☐

We consider now the portfolio of the reinsurer. What is the claim number process of the claims the reinsurer is involved? Because the claim amounts are independent of the claim arrival process we delete independently points from the Poisson process with probability $G(M)$ and do not delete them with probability $1 - G(M)$. By Proposition C.3 this process is a Poisson process with rate $\lambda(1 - G(M))$. Because the claim sizes are iid. and independent of the claim arrival process the surplus of the reinsurer is a Cramér-Lundberg process with intensity $\lambda(1 - G(M))$ and claim size distribution

$$\tilde{G}(x) = \mathbb{P}[Y_i - M \le x \mid Y_i > M] = \frac{G(M+x) - G(M)}{1 - G(M)}\,.$$

4.8. The Severity of Ruin and the Distribution of $\inf\{C_t : t \ge 0\}$

For an insurance company ruin is not so dramatic if $-C_\tau$ is small, but it could ruin the whole company if $-C_\tau$ is very large. So we are interested in the distribution of $-C_\tau$ if ruin occurs. Let

$$\psi_x(u) = \mathbb{P}[\tau < \infty, C_\tau < -x]\,.$$

We proceed as in Section 4.3. For h small

$$\psi_x(u) = \mathrm{e}^{-\lambda h}\psi_x(u+ch) + \int_0^h \Bigl[\int_0^{u+ct} \psi_x(u+ct-y)\,\mathrm{d}G(y) + (1 - G(u+ct+x))\Bigr]\lambda \mathrm{e}^{-\lambda t}\,\mathrm{d}t\,.$$

It follows that $\psi_x(u)$ is right continuous. We obtain the integro-differential equation

$$c\psi_x'(u) = \lambda\Bigl[\psi_x(u) - \int_0^u \psi_x(u-y)\,\mathrm{d}G(y) - (1 - G(u+x))\Bigr]\,.$$

Replacing u by $u - ch$ gives the derivative from the left. Integration of the integro-differential equation yields

$$\begin{aligned}\frac{c}{\lambda}(\psi_x(u) - \psi_x(0)) &= \int_0^u \psi_x(u-y)(1 - G(y))\,\mathrm{d}y - \int_0^u (1 - G(y+x))\,\mathrm{d}y \\ &= \int_0^u \psi_x(u-y)(1 - G(y))\,\mathrm{d}y - \int_x^{x+u} (1 - G(y))\,\mathrm{d}y\,.\end{aligned}$$

Because $\psi_x(u) \le \psi(u)$ we can see that $\psi_x(u) \to 0$ as $u \to \infty$. By the bounded convergence theorem we obtain

$$-\frac{c}{\lambda}\psi_x(0) = -\int_x^\infty (1 - G(y))\,\mathrm{d}y$$

and

$$\mathbb{P}[C_\tau < -x \mid \tau < \infty, C_0 = 0] = \frac{\dfrac{\lambda}{c}\displaystyle\int_x^\infty (1 - G(y))\,\mathrm{d}y}{\dfrac{\lambda\mu}{c}} = \frac{1}{\mu}\int_x^\infty (1 - G(y))\,\mathrm{d}y\,. \tag{4.10}$$

The random variable $-C_\tau$ is called the **severity of ruin**.

Remark. In order to get an equation for the ruin probability we can consider the first time point τ_1 where the surplus is below the initial capital. At this point, by the strong Markov property, a new Cramér-Lundberg process starts. We get three possibilities:

- The process gets never below the initial capital.
- $\tau_1 < \infty$ but $C_{\tau_1} \ge 0$.
- Ruin occurs at τ_1.

Thus we get

$$\psi(u) = \Bigl(1 - \frac{\lambda\mu}{c}\Bigr)0 + \frac{\lambda}{c}\int_0^u \psi(u-y)(1 - G(y))\,\mathrm{d}y + \frac{\lambda}{c}\int_u^\infty 1\,(1 - G(y))\,\mathrm{d}y\,.$$

This is equation (4.2). Thus we have now a natural interpretation of (4.2). ■

Let $\tau_0 = 0$ and $\tau_i = \inf\{t > \tau_{i-1} : C_t < C_{\tau_{i-1}}\}$, called the **ladder times**, and define $L_i = C_{\tau_{i-1}} - C_{\tau_i}$, called the **ladder heights**. Note that L_i only is defined if $\tau_i < \infty$. By Lemma 4.1 we find $\mathbb{P}[\tau_i < \infty \mid \tau_{i-1} < \infty] = \lambda\mu/c$. Let $K = \sup\{i \in \mathbb{N} : \tau_i < \infty\}$ be the number of ladder epochs. We have just seen that $K \sim \mathrm{NB}(1, 1 - \lambda\mu/c)$ and that, given K, the random variables $(L_i : i \le K)$ are iid. and absolutely continuous with density $(1 - G(x))/\mu$. We only have to condition on K because L_i is not defined for $i > K$. If we assume that all $(L_i : i \ge 1)$ have the same distribution, then we can drop the conditioning on K and (L_i) is independent of K. Then

$$\inf\{C_t : t \ge 0\} = u - \sum_{i=1}^K L_i$$

and

$$\mathbb{P}[\tau < \infty] = \mathbb{P}[\inf\{C_t : t \geq 0\} < 0] = \mathbb{P}\Bigl[\sum_{i=1}^{K} L_i > u\Bigr]\,.$$

Denote by

$$B(x) = \frac{1}{\mu}\int_0^x (1 - G(y))\,\mathrm{d}y$$

the distribution function of L_i. We can use Panjer recursion to approximate $\psi(u)$ by using an appropriate discretization.

A formula that is useful for theoretical considerations, but too messy to use for the computation of $\psi(u)$ is the Pollaczek-Khintchine formula.

$$\begin{aligned}\psi(u) &= \mathbb{P}\Bigl[\sum_{i=1}^{K} L_i > u\Bigr] = \sum_{n=1}^{\infty}\mathbb{P}\Bigl[\sum_{i=1}^{n} L_i > u\Bigr]\mathbb{P}[K = n] \\ &= \Bigl(1 - \frac{\lambda\mu}{c}\Bigr)\sum_{n=1}^{\infty}\Bigl(\frac{\lambda\mu}{c}\Bigr)^n (1 - B^{*n}(u))\,. \qquad (4.11)\end{aligned}$$

4.9. The Laplace Transform of ψ

Definition 4.10. *Let f be a real function on $[0, \infty)$. The transform*

$$\hat{f}(s) := \int_0^\infty \mathrm{e}^{-sx} f(x)\,\mathrm{d}x \qquad (s \in \mathbb{R})$$

is called the **Laplace transform** *of f.*

If X is an absolutely continuous positive random variable with density f then $\hat{f}(s) = M_X(-s)$. The Laplace transform has the following properties.

Lemma 4.11.

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

$$\hat{f}(s_1) \leq \hat{f}(s_2) \quad \iff \quad s_1 \geq s_2.$$

ii)

$$\widehat{|f|}(s_1) < \infty \quad \Longrightarrow \quad \widehat{|f|}(s) < \infty \text{ for all } s \geq s_1.$$

iii)

$$\widehat{f'}(s) = \int_0^\infty \mathrm{e}^{-sx} f'(x)\,\mathrm{d}x = s\hat{f}(s) - f(0)$$

provided $f'(x)$ exists a.e. and $\widehat{|f|}(s) < \infty$.

iv)

$$\lim_{s\to\infty} s\hat{f}(s) = \lim_{x\to 0} f(x)$$

provided $f'(x)$ exists a.e., $\lim_{x\to 0} f(x)$ exists and $\widehat{|f|}(s) < \infty$ for an s large enough.

v) $$\lim_{s\downarrow 0} s\hat{f}(s) = \lim_{x\to\infty} f(x)$$

provided $f'(x)$ *exists a.e.,* $\lim_{x\to\infty} f(x)$ *exists and* $\widehat{|f|}(s) < \infty$ *for all* $s > 0$.

vi) $$\hat{f}(s) = \hat{g}(s) \text{ on } (s_0, s_1) \quad \Longrightarrow \quad f(x) = g(x) \text{ Lebesgue a.e.} \quad \forall x \in [0,\infty).$$

□

We want to find the Laplace transform of $\delta(u)$. We multiply (4.1) with e^{-su} and then integrate over u. Let $s > 0$.

$$c\int_0^\infty \delta'(u)\mathrm{e}^{-su}\,\mathrm{d}u = \lambda\int_0^\infty \delta(u)\mathrm{e}^{-su}\,\mathrm{d}u - \lambda\int_0^\infty\int_0^u \delta(u-y)\,\mathrm{d}G(y)\mathrm{e}^{-su}\,\mathrm{d}u.$$

We have to determine the last integral.

$$\begin{aligned}\int_0^\infty\int_0^u \delta(u-y)\,\mathrm{d}G(y)\mathrm{e}^{-su}\,\mathrm{d}u &= \int_0^\infty\int_y^\infty \delta(u-y)\mathrm{e}^{-su}\,\mathrm{d}u\,\mathrm{d}G(y)\\ &= \int_0^\infty\int_0^\infty \delta(u)\mathrm{e}^{-s(u+y)}\,\mathrm{d}u\,\mathrm{d}G(y) = \hat{\delta}(s)M_Y(-s)\,.\end{aligned}$$

Thus we get the equation

$$c(s\hat{\delta}(s) - \delta(0)) = \lambda\hat{\delta}(s)(1 - M_Y(-s))$$

which has the solution

$$\hat{\delta}(s) = \frac{c\delta(0)}{cs - \lambda(1 - M_Y(-s))} = \frac{c - \lambda\mu}{cs - \lambda(1 - M_Y(-s))}\,.$$

The Laplace transform of ψ can easily be found as

$$\hat{\psi}(s) = \int_0^\infty (1 - \delta(u))\mathrm{e}^{-su}\,\mathrm{d}u = \frac{1}{s} - \hat{\delta}(s)\,.$$

Example 4.2 (continued). For exponentially distributed claims we obtain

$$\hat{\delta}(s) = \frac{c - \lambda/\alpha}{cs - \lambda(1 - \alpha/(\alpha+s))} = \frac{c - \lambda/\alpha}{s(c - \lambda/(\alpha+s))} = \frac{(c - \lambda/\alpha)(\alpha+s)}{s(c(\alpha+s) - \lambda)}\,.$$

and

$$\begin{aligned}\hat{\psi}(s) &= \frac{1}{s} - \frac{(c - \lambda/\alpha)(\alpha+s)}{s(c(\alpha+s) - \lambda)} = \frac{c(\alpha+s) - \lambda - (c - \lambda/\alpha)(\alpha+s)}{s(c(\alpha+s) - \lambda)}\\ &= \frac{\lambda}{\alpha}\,\frac{1}{c(\alpha+s) - \lambda} = \frac{\lambda}{\alpha c}\,\frac{1}{\alpha - \lambda/c + s}\,.\end{aligned}$$

By comparison with the moment generating function of the exponential distribution we recognize that

$$\psi(u) = \frac{\lambda}{\alpha c}\mathrm{e}^{-(\alpha-\lambda/c)u}\,.$$

■

Example 4.5 (continued). For the Laplace transform of δ we get

$$\begin{aligned}\hat{\delta}(s) &= \frac{c-\lambda p/\alpha-\lambda(1-p)/\beta}{cs-\lambda(1-\alpha p/(\alpha+s)-\beta(1-p)/(\beta+s))}\\ &= \frac{c-\lambda p/\alpha-\lambda(1-p)/\beta}{s(c-\lambda p/(\alpha+s)-\lambda(1-p)/(\beta+s))}\end{aligned}$$

and for the Laplace transform of ψ

$$\begin{aligned}\hat{\psi}(s) &= \frac{c-\lambda p/(\alpha+s)-\lambda(1-p)/(\beta+s)-c+\lambda p/\alpha+\lambda(1-p)/\beta}{s(c-\lambda p/(\alpha+s)-\lambda(1-p)/(\beta+s))}\\ &= \lambda\frac{sp/(\alpha(\alpha+s))+s(1-p)/(\beta(\beta+s))}{s(c-\lambda p/(\alpha+s)-\lambda(1-p)/(\beta+s))}\\ &= \lambda\frac{p(\beta+s)/\alpha+(1-p)(\alpha+s)/\beta}{c(\alpha+s)(\beta+s)-\lambda p(\beta+s)-\lambda(1-p)(\alpha+s)}\\ &= \lambda\frac{p(\beta+s)/\alpha+(1-p)(\alpha+s)/\beta}{cs^2+((\alpha+\beta)c-\lambda)s+\alpha\beta c-\lambda((1-p)\alpha+p\beta)}\,.\end{aligned}$$

The denominator can be written as $c(s+R)(s+\bar{R})$ where R and $\bar{R}$ are the two solutions to (4.6) with $R<\bar{R}$. The Laplace transform of ψ can be written in the form

$$\hat{\psi}(s)=\frac{A}{R+s}+\frac{B}{\bar{R}+s}$$

for some constants A and B. Hence

$$\psi(u)=A\mathrm{e}^{-Ru}+B\mathrm{e}^{-\bar{R}u}\,.$$

A must be the constant appearing in the Cramér-Lundberg approximation. The constant B can be found from $\psi(0)$. We can see that in this case the Cramér-Lundberg approximation is not exact. ■

Recall that τ_1 is the first ladder epoch. On the set $\{\tau_1<\infty\}$ the random variable $\sup\{u-C_t : t\ge 0\}$ is absolutely continuous with distribution function

$$\mathbb{P}[\sup\{u-C_t : t\ge 0\}\le x \mid \tau_1<\infty]=1-\frac{c}{\lambda\mu}\psi(x)\,.$$

Let Z be a random variable with the above distribution. Its moment generating function is

$$\begin{aligned}M_Z(r) &= \int_0^\infty \mathrm{e}^{ru}\frac{c}{\lambda\mu}\delta'(u)\,\mathrm{d}u = \frac{c}{\lambda\mu}\Bigl(-r\frac{c-\lambda\mu}{-cr-\lambda(1-M_Y(r))}-\Bigl(1-\frac{\lambda\mu}{c}\Bigr)\Bigr)\\ &= 1-\frac{c}{\lambda\mu}+\frac{c(c-\lambda\mu)}{\lambda\mu}\,\frac{r}{cr-\lambda(M_Y(r)-1)}\,.\end{aligned}$$

We will later need the first two moments of the above distribution function. Assume that $\mu_2 < \infty$. The first derivative of the moment generating function is

$$\begin{aligned} M_Z'(r) &= \frac{c(c-\lambda\mu)}{\lambda\mu} \frac{cr - \lambda(M_Y(r)-1) - r(c-\lambda M_Y'(r))}{(cr-\lambda(M_Y(r)-1))^2} \\ &= \frac{c(c-\lambda\mu)}{\mu} \frac{rM_Y'(r) - (M_Y(r)-1)}{(cr-\lambda(M_Y(r)-1))^2} \,. \end{aligned}$$

Note that

$$\lim_{r\to 0} \frac{1}{r}(cr - \lambda(M_Y(r)-1)) = c - \lambda\mu \,.$$

We find

$$\lim_{r\to 0} \frac{rM_Y'(r) - (M_Y(r)-1)}{r^2} = \lim_{r\to 0} \frac{M_Y'(r) + rM_Y''(r) - M_Y'(r)}{2r} = \frac{\mu_2}{2}$$

and thus

$$\mathbb{E}[Z] = \frac{c(c-\lambda\mu)}{\mu} \frac{\mu_2}{2(c-\lambda\mu)^2} = \frac{c\mu_2}{2\mu(c-\lambda\mu)} \,. \tag{4.12}$$

Assume now that $\mu_3 < \infty$. The second derivative of $M_Z(r)$ is

$$\begin{aligned} M_Z''(r) = \frac{c(c-\lambda\mu)}{\mu} \frac{1}{(cr-\lambda(M_Y(r)-1))^3} \Bigl(& rM_Y''(r)(cr-\lambda(M_Y(r)-1)) \\ & - 2(rM_Y'(r) - (M_Y(r)-1))(c-\lambda M_Y'(r)) \Bigr) \,. \end{aligned}$$

For the limit to 0 we find

$$\begin{aligned} \lim_{r\to 0} \frac{1}{r^3} \Bigl(& rM_Y''(r)(cr-\lambda(M_Y(r)-1)) \\ & - 2(rM_Y'(r) - (M_Y(r)-1))(c-\lambda M_Y'(r)) \Bigr) \\ = \lim_{r\to 0} \frac{1}{3r^2} \Bigl(& (M_Y''(r) + rM_Y'''(r))(cr-\lambda(M_Y(r)-1)) \\ & + rM_Y''(r)(c-\lambda M_Y'(r)) - 2rM_Y''(r)(c-\lambda M_Y'(r)) \\ & + 2(rM_Y'(r) - (M_Y(r)-1))\lambda M_Y''(r) \Bigr) \\ = \frac{\mu_3}{3}&(c-\lambda\mu) + \lambda\mu_2 \lim_{r\to 0} \frac{rM_Y'(r) - (M_Y(r)-1)}{r^2} \\ = \frac{\mu_3}{3}&(c-\lambda\mu) + \frac{\lambda}{2}\mu_2^2 \,. \end{aligned}$$

Thus the second moment of Z becomes

$$\mathbb{E}[Z^2] = \frac{c(c-\lambda\mu)}{\mu} \frac{\mu_3(c-\lambda\mu)/3 + \lambda\mu_2^2/2}{(c-\lambda\mu)^3} = \frac{c}{\mu}\left(\frac{\mu_3}{3(c-\lambda\mu)} + \frac{\lambda\mu_2^2}{2(c-\lambda\mu)^2} \right) \,. \tag{4.13}$$

4.10. Approximations to ψ

4.10.1. Diffusion Approximations

Diffusion approximations are based on the following

Proposition 4.12. *Let $\{C_t^{(n)}\}$ be a sequence of Cramér-Lundberg processes with initial capital $u^{(n)} = u$, claim arrival intensities $\lambda^{(n)} = \lambda n$, claim size distributions $G^{(n)}(x) = G(x\sqrt{n})$ and premium rates*

$$c^{(n)} = \Bigl(1 + \frac{c - \lambda\mu}{\lambda\mu\sqrt{n}}\Bigr)\lambda^{(n)}\mu^{(n)} = c + (\sqrt{n} - 1)\lambda\mu\,.$$

Let $\mu = \int_0^\infty y \, \mathrm{d}G(y)$ and assume that $\mu_2 = \int_0^\infty y^2 \, \mathrm{d}G(y) < \infty$. Then

$$\{C_t^{(n)}\} \xrightarrow{\mathrm{d}} \{u + W_t\}$$

in distribution in the topology of uniform convergence on finite intervals where $\{W_t\}$ is a $(c - \lambda\mu, \lambda\mu_2)$-Brownian motion.

Proof. See [55] or [45]. □

Intuitively we let the number of claims in a unit time interval go to infinity and make the claim sizes smaller in such a way that the distribution of $C_1^{(n)} - u$ tends to a normal distribution and $\mathbb{E}[C_1^{(n)} - u] = c - \lambda\mu$. Let $\tau^{(n)}$ denote the ruin time of $\{C_t^{(n)}\}$ and $\tau = \inf\{t \ge 0 : u + W_t < 0\}$ the ruin probability of the Brownian motion. Then

Proposition 4.13. *Let $(C_t^{(n)})$ and (W_t) be as above. Then*

$$\lim_{n\to\infty} \mathbb{P}[\tau^{(n)} \le t] = \mathbb{P}[\tau \le t]$$

and

$$\lim_{n\to\infty} \mathbb{P}[\tau^{(n)} < \infty] = \mathbb{P}[\tau < \infty]\,.$$

Proof. The result for a finite time horizon is a special case of [87, Thm.9], see also [55] or [45]. The result for the infinite time horizon can be found in [70]. □

The idea of the diffusion approximation is to approximate $\mathbb{P}[\tau^{(1)} \le t]$ by $\mathbb{P}[\tau \le t]$ and $\mathbb{P}[\tau^{(1)} < \infty]$ by $\mathbb{P}[\tau < \infty]$. Thus we need the ruin probabilities of the Brownian motion.

Lemma 4.14. *Let* $\{W_t\}$ *be a* (m, η^2)*-Brownian motion with* $m > 0$ *and* $\tau = \inf\{t \ge 0 : u + W_t < 0\}$. *Then*

$$\mathbb{P}[\tau < \infty] = \mathrm{e}^{-2um/\eta^2}$$

and

$$\mathbb{P}[\tau \le t] = 1 - \Phi\Bigl(\frac{mt + u}{\eta\sqrt{t}}\Bigr) + \mathrm{e}^{-2um/\eta^2}\Phi\Bigl(\frac{mt - u}{\eta\sqrt{t}}\Bigr)\,.$$

Proof. By Lemma D.3 the process

$$\exp\Bigl\{-\frac{2m(u + W_t)}{\eta^2}\Bigr\}$$

is a martingale. By the stopping theorem

$$\exp\Bigl\{-\frac{2m(u + W_{\tau\wedge t})}{\eta^2}\Bigr\}$$

is a positive bounded martingale. Thus, because $\lim_{t\to\infty} W_t = \infty$,

$$\exp\Bigl\{-\frac{2um}{\eta^2}\Bigr\} = \mathbb{E}\Bigl[\exp\Bigl\{-\frac{2m(u + W_\tau)}{\eta^2}\Bigr\}\Bigr] = \mathbb{P}[\tau < \infty]\,.$$

It is easy to see that $\{sW_{1/s} - m\}$ is a $(0, \eta^2)$-Brownian motion (see [58, p.351]) and thus $\{s(u + W_{1/s}) - m\}$ is (u, η^2)-Brownian motion. Denote the latter process by $\{\tilde{W}_s\}$. Then

$$\begin{aligned}
\mathbb{P}[\tau \le t] &= \mathbb{P}[\inf\{u + W_s : 0 < s \le t\} < 0] \\
&= \mathbb{P}[\inf\{s(u + W_{1/s}) : s \ge 1/t\} < 0] \\
&= \mathbb{E}[\mathbb{P}[\inf\{m + \tilde{W}_s : s \ge 1/t\} < 0 \mid \tilde{W}_{1/t}]] \\
&= \int_{-\infty}^{-m} \frac{1}{\sqrt{2\pi\eta^2/t}}\mathrm{e}^{-\frac{(y-u/t)^2}{2\eta^2/t}}\,\mathrm{d}y + \int_{-m}^{\infty} \mathrm{e}^{-\frac{2u(y+m)}{\eta^2}}\frac{1}{\sqrt{2\pi\eta^2/t}}\mathrm{e}^{-\frac{(y-u/t)^2}{2\eta^2/t}}\,\mathrm{d}y \\
&= \Phi\Bigl(\frac{-m - u/t}{\eta/\sqrt{t}}\Bigr) + \mathrm{e}^{-\frac{2um}{\eta^2}}\int_{-m}^{\infty} \frac{1}{\sqrt{2\pi\eta^2/t}}\mathrm{e}^{-\frac{(y+u/t)^2}{2\eta^2/t}}\,\mathrm{d}y \\
&= 1 - \Phi\Bigl(\frac{mt + u}{\eta\sqrt{t}}\Bigr) + \mathrm{e}^{-\frac{2um}{\eta^2}}\Phi\Bigl(\frac{mt - u}{\eta\sqrt{t}}\Bigr)\,.
\end{aligned}$$

□

Diffusion approximations only work well if $c/(\lambda\mu)$ is close to 1. There also exist corrected diffusion approximations which work much better, see [77] or [7].

u	0	0.25	0.5	0.75	1	1.25	1.5	1.75	2	2.25
$\psi(u)$	0.6111	0.5246	0.4547	0.3969	0.3479	0.3059	0.2696	0.2379	0.2102	0.1858
DA	1.0000	0.8071	0.6514	0.5258	0.4244	0.3425	0.2765	0.2231	0.1801	0.1454
Er	63.64	53.87	43.26	32.49	21.98	11.96	2.54	-6.22	-14.32	-21.78

Table 4.2: *Diffusion approximation to ruin probabilities*

Example 4.7 (continued). Let $c = \lambda = 1$ and $G(x) = 1 - \frac{1}{3}(\mathrm{e}^{-x} + \mathrm{e}^{-2x} + \mathrm{e}^{-3x})$. We find $c - \lambda\mu = 7/18$ and $\lambda\mu_2 = 49/54$. This leads to the diffusion approximation $\psi(u) \approx \exp\{-6u/7\}$. Table 4.2 shows exact values ($\psi(u)$), the diffusion approximation (DA) and the relative error multiplied by 100 (Er). Here we have $c/(\lambda\mu) = 18/11 = 1.63636$ is not close to one. This is also indicated by the figures. ■

4.10.2. The deVylder Approximation

In the case of exponentially distributed claim amounts we know the ruin probabilities explicitly. The idea of the deVylder approximation is to replace $\{C_t\}$ by $\{\tilde{C}_t\}$ where $\{\tilde{C}_t\}$ has exponentially distributed claim amounts and

$$\mathbb{E}[(C_t - u)^k] = \mathbb{E}[(\tilde{C}_t - u)^k] \qquad \text{for } k = 1, 2, 3.$$

The first three (centralized) moments are

$$\mathbb{E}[C_t - u] = (c - \lambda\mu)t = \Bigl(\tilde{c} - \frac{\tilde{\lambda}}{\tilde{\alpha}}\Bigr)t\,,$$

$$\operatorname{Var}[C_t] = \operatorname{Var}[u + ct - C_t] = \lambda\mu_2 t = \frac{2\tilde{\lambda}}{\tilde{\alpha}^2}t$$

and

$$\mathbb{E}[(C_t - \mathbb{E}[C_t])^3] = -\mathbb{E}[(u + ct - C_t - \mathbb{E}[u + ct - C_t])^3] = -\lambda\mu_3 t = -\frac{6\tilde{\lambda}}{\tilde{\alpha}^3}t\,.$$

The parameters of the approximation are

$$\tilde{\alpha} = \frac{3\mu_2}{\mu_3}\,,$$

u	0	0.25	0.5	0.75	1	1.25	1.5	1.75	2	2.25
$\psi(u)$	0.6111	0.5246	0.4547	0.3969	0.3479	0.3059	0.2696	0.2379	0.2102	0.1858
DV	0.5774	0.5102	0.4509	0.3984	0.3520	0.3110	0.2748	0.2429	0.2146	0.1896
Er	-5.51	-2.73	-0.86	0.38	1.18	1.67	1.95	2.07	2.09	2.03

Table 4.3: *DeVylder approximation to ruin probabilities*

$$\tilde{\lambda} = \frac{\lambda \mu_2 \tilde{\alpha}^2}{2} = \frac{9\mu_2^3}{2\mu_3^2}\lambda$$

and

$$\tilde{c} = c - \lambda\mu + \frac{\tilde{\lambda}}{\tilde{\alpha}} = c - \lambda\mu + \frac{3\mu_2^2}{2\mu_3}\lambda\,.$$

Thus the approximation to the probability of ultimate ruin is

$$\psi(u) \approx \frac{\tilde{\lambda}}{\tilde{\alpha}\tilde{c}} \mathrm{e}^{-\left(\tilde{\alpha}-\frac{\tilde{\lambda}}{\tilde{c}}\right)u}\,.$$

There is also a formula for the probability of ruin within finite time. Let $\eta = \sqrt{\tilde{\lambda}/(\tilde{\alpha}\tilde{c})}$. Then

$$\psi(u,t) \approx \frac{\tilde{\lambda}}{\tilde{\alpha}\tilde{c}} \mathrm{e}^{-(\tilde{\alpha}-\tilde{\lambda}/\tilde{c})u} - \frac{1}{\pi}\int_0^\pi f(x)\,\mathrm{d}x \tag{4.14}$$

where

$$f(x) = \eta \frac{\exp\{2\eta\tilde{\alpha}\tilde{c}t\,\cos x - (\tilde{\alpha}\tilde{c}+\tilde{\lambda})t + \tilde{\alpha}u(\eta\cos x - 1)\}}{1+\eta^2 - 2\eta\cos x} \times (\cos(\tilde{\alpha}u\eta\sin x) - \cos(\tilde{\alpha}u\eta\sin x + 2x))\,.$$

A numerical investigation shows that the approximation is quite accurate.

Example 4.7 (continued). In addition to the previously calculated values we also need $\mu_3 = 251/108$. The approximation parameters are $\tilde{\alpha} = 1.17131$, $\tilde{\lambda} = 0.622472$ and $\tilde{c} = 0.920319$. This leads to the approximation $\psi(u) \approx 0.577441\mathrm{e}^{-0.494949u}$. It turns out that the approximation works well. ■

4.10.3. The Beekman-Bowers Approximation

Recall from (4.12) and (4.13) that

$$F(u) = 1 - \frac{c}{\lambda\mu}\psi(u)$$

is a distribution function and that

$$\int_0^\infty z \,\mathrm{d}F(z) = \frac{c\mu_2}{2\mu(c-\lambda\mu)}$$

and that

$$\int_0^\infty z^2 \,\mathrm{d}F(z) = \frac{c}{\mu}\Bigl(\frac{\mu_3}{3(c-\lambda\mu)} + \frac{\lambda\mu_2^2}{2(c-\lambda\mu)^2}\Bigr)\,.$$

The idea is to approximate the distribution function F by the distribution function $\tilde{F}(u)$ of a $\Gamma(\gamma,\alpha)$ distributed random variable such that the first two moments coincide. Thus the parameters γ and α have to fulfil

$$\frac{\gamma}{\alpha} = \frac{c\mu_2}{2\mu(c-\lambda\mu)}\,,$$
$$\frac{\gamma(\gamma+1)}{\alpha^2} = \frac{c}{\mu}\Bigl(\frac{\mu_3}{3(c-\lambda\mu)} + \frac{\lambda\mu_2^2}{2(c-\lambda\mu)^2}\Bigr)\,.$$

The Beekman-Bowers approximation to the ruin probability is

$$\psi(u) = \frac{\lambda\mu}{c}(1-F(u)) \approx \frac{\lambda\mu}{c}(1-\tilde{F}(u))\,.$$

Remark. If $2\gamma \in \mathbb{N}$ then $2\alpha Z \sim \chi^2_{2\gamma}$ is χ^2 distributed. ■

Example 4.7 (continued). For the Beekman-Bowers approximation we have to solve the equations

$$\frac{\gamma}{\alpha} = 1.90909\,, \qquad \frac{\gamma(\gamma+1)}{\alpha^2} = 7.71429\,,$$

which yields the parameters $\gamma = 0.895561$ and $\alpha = 0.469104$. From this the Beekman Bowers approximation can be obtained. Here we have $2\gamma = 1.79112$ which is not close to an integer. Anyway, one can interpolate between the χ^2_1 and the χ^2_2 distribution function to get the approximation

$$0.20888\chi^2_1(2\alpha u) + 0.79112\chi^2_2(2\alpha u)$$

to $1 - c/(\lambda\mu)\,\psi(u)$. Table 4.4 shows the exact values ($\psi(u)$), the Beekman-Bowers approximation (BB1) and the approximation obtained by interpolating the χ^2 distributions (BB2). The relative errors (Er) are given in percent. One can clearly see that all the approximations work well. ■

u	0	0.25	0.5	0.75	1	1.25	1.5	1.75	2	2.25
$\psi(u)$	0.6111	0.5246	0.4547	0.3969	0.3479	0.3059	0.2696	0.2379	0.2102	0.1858
BB1	0.6111	0.5227	0.4553	0.3985	0.3498	0.3076	0.2709	0.2387	0.2106	0.1859
Er	0.00	-0.35	0.12	0.42	0.54	0.54	0.47	0.34	0.19	0.04
BB2	0.6111	0.5105	0.4456	0.3914	0.3450	0.3046	0.2693	0.2383	0.2110	0.1869
Er	0.00	-2.68	-2.02	-1.38	-0.83	-0.42	-0.11	0.18	0.40	0.59

Table 4.4: *Beekman-Bowers approximation to ruin probabilities*

4.11. Subexponential Claim Size Distributions

Let us now consider subexponential claim size distributions. In this case the Lundberg exponent does not exist (Lemma F.3).

Theorem 4.15. *Assume that the ladder height distribution*

$$\frac{1}{\mu}\int_0^x (1 - G(y))\,\mathrm{d}y$$

is subexponential. Then

$$\lim_{u\to\infty} \frac{\psi(u)}{\int_u^\infty (1 - G(y))\,\mathrm{d}y} = \frac{\lambda}{c - \lambda\mu}\,.$$

Remark. Recall that the probability that ruin occurs at the first ladder time given there is a first ladder epoch is

$$\frac{1}{\mu}\int_u^\infty (1 - G(y))\,\mathrm{d}y\,.$$

Hence the ruin probability is asymptotically $(c - \lambda\mu)^{-1}\lambda\mu$ times the probability of ruin at the first ladder time given there is a first ladder epoch. But $(c - \lambda\mu)^{-1}\lambda\mu$ is the expected number of ladder times. Intuitively for u large ruin will occur if one of the ladder heights is larger than u. ■

Proof. Let $B(x)$ denote the distribution function of the first ladder height L_1, i.e.

$$B(x) = \frac{1}{\mu}\int_0^x (1 - G(y))\,\mathrm{d}y\,.$$

Choose $\varepsilon > 0$ such that $\lambda\mu(1+\varepsilon) < c$. By Lemma F.6 there exists D such that

$$\frac{1 - B^{*n}(x)}{1 - B(x)} \le D(1+\varepsilon)^n\,.$$

From the Pollaczek-Khintchine formula (4.11) we obtain

$$\begin{aligned}\frac{\psi(u)}{1 - B(u)} &= \Bigl(1 - \frac{\lambda\mu}{c}\Bigr)\sum_{n=1}^\infty \Bigl(\frac{\lambda\mu}{c}\Bigr)^n \frac{1 - B^{*n}(u)}{1 - B(u)} \\ &\le D\Bigl(1 - \frac{\lambda\mu}{c}\Bigr)\sum_{n=1}^\infty \Bigl(\frac{\lambda\mu}{c}\Bigr)^n (1+\varepsilon)^n < \infty\,.\end{aligned}$$

Thus we can interchange sum and limit. Recall from Lemma F.7 that

$$\lim_{u\to\infty}\frac{1 - B^{*n}(u)}{1 - B(u)} = n\,.$$

Thus

$$\begin{aligned}\lim_{u\to\infty}\frac{\psi(u)}{1 - B(u)} &= \Bigl(1 - \frac{\lambda\mu}{c}\Bigr)\sum_{n=1}^\infty n\Bigl(\frac{\lambda\mu}{c}\Bigr)^n = \Bigl(1 - \frac{\lambda\mu}{c}\Bigr)\sum_{n=1}^\infty\sum_{m=1}^n \Bigl(\frac{\lambda\mu}{c}\Bigr)^n \\ &= \Bigl(1 - \frac{\lambda\mu}{c}\Bigr)\sum_{m=1}^\infty\sum_{n=m}^\infty \Bigl(\frac{\lambda\mu}{c}\Bigr)^n = \sum_{m=1}^\infty \Bigl(\frac{\lambda\mu}{c}\Bigr)^m = \frac{\lambda\mu/c}{1 - \lambda\mu/c} = \frac{\lambda\mu}{c - \lambda\mu}\,.\end{aligned}$$

□

Example 4.16. Let $G \sim \mathrm{Pa}(\alpha,\beta)$. Then (see Example F.5)

$$\lim_{x\to\infty}\frac{1 - B(zx)}{1 - B(x)} = z\lim_{x\to\infty}\frac{1 - G(zx)}{1 - G(x)} = z^{-(\alpha-1)}\,.$$

By Lemma F.4 B is a subexponential distribution. Note that we assume that $\mu < \infty$ and thus $\alpha > 1$. Because

$$\int_x^\infty \Bigl(\frac{\beta}{\beta + y}\Bigr)^\alpha \mathrm{d}y = \frac{\beta}{\alpha - 1}\Bigl(\frac{\beta}{\beta + x}\Bigr)^{\alpha-1}$$

we obtain

$$\psi(u) \approx \frac{\lambda\beta/(\alpha-1)}{c - \lambda\beta/(\alpha-1)}\Bigl(\frac{\beta}{\beta+u}\Bigr)^{\alpha-1} = \frac{\lambda\beta}{c(\alpha-1) - \lambda\beta}\Bigl(\frac{\beta}{\beta+u}\Bigr)^{\alpha-1}.$$

u	$\psi(u)$	App	Er
1	0.364	$8.79 \cdot 10^{-3}$	-97.588
2	0.150	$1.52 \cdot 10^{-4}$	-99.898
3	$6.18 \cdot 10^{-2}$	$8.58 \cdot 10^{-6}$	-99.986
4	$2.55 \cdot 10^{-2}$	$9.22 \cdot 10^{-7}$	-99.996
5	$1.05 \cdot 10^{-2}$	$1.49 \cdot 10^{-7}$	-99.999
10	$1.24 \cdot 10^{-4}$	$3.47 \cdot 10^{-10}$	-100
20	$1.75 \cdot 10^{-8}$	$5.40 \cdot 10^{-13}$	-99.997
30	$2.50 \cdot 10^{-12}$	$1.10 \cdot 10^{-14}$	-99.56
40	$1.60 \cdot 10^{-15}$	$6.71 \cdot 10^{-16}$	-58.17
50	$1.21 \cdot 10^{-16}$	$7.56 \cdot 10^{-17}$	-37.69

Table 4.5: *Approximations for subexponential claim sizes*

Choose now $c = 1$, $\lambda = 9$, $\alpha = 11$ and $\beta = 1$. Table 4.5 gives the exact value ($\psi(u)$), the approximation (App) and the relative error in percent (Er). Consider for instance $u = 20$. The ruin probability is so small, that it is not interesting for practical purposes anymore. But the approximation still underestimates the true value by almost 100%. That means we are still not far out enough in the tail. This is the problem by using the approximation. It should, however, be remarked, that for small values of α the approximation works much better. Values $\alpha \in (1, 2)$ are also more interesting from a practical point of view. ■

Remark. The conditions of the theorem are also fulfilled for $\mathrm{LN}(\mu, \sigma^2)$, for $\mathrm{LG}(\gamma, \alpha)$ and for $\mathrm{Wei}(\alpha, c)$ ($\alpha < 1$) distributed claims, see [34] and [60]. ■

4.12. The Time of Ruin

Consider the function

$$f_\alpha(u) = \mathbb{E}[\mathrm{e}^{-\alpha\tau} 1\!\!1_{\{\tau<\infty\}} \mid C_0 = u]\,.$$

The function is defined at least for $\alpha \ge 0$. We will first find a differential equation for $f_\alpha(u)$.

Lemma 4.17. *The function $f_\alpha(u)$ is absolutely continuous and fulfils the equation*

$$c f_\alpha'(u) + \lambda\Bigl[\int_0^u f_\alpha(u-y)\,\mathrm{d}G(y) + 1 - G(u) - f_\alpha(u)\Bigr] - \alpha f_\alpha(u) = 0\,. \tag{4.15}$$

Proof. For h small we get

$$f_\alpha(u) = \mathrm{e}^{-\lambda h}\mathrm{e}^{-\alpha h} f_\alpha(u+ch) + \int_0^h \Bigl[\int_0^{u+ct} \mathrm{e}^{-\alpha t} f_\alpha(u+ct-y)\,\mathrm{d}G(y) + \mathrm{e}^{-\alpha t}(1-G(u+ct))\Bigr]\lambda \mathrm{e}^{-\lambda t}\,\mathrm{d}t\;.$$

We see that $f_\alpha(u)$ is right continuous. Reordering of the terms and dividing by h yields

$$c\,\frac{f_\alpha(u+ch)-f_\alpha(u)}{ch} - \frac{1-\mathrm{e}^{-(\lambda+\alpha)h}}{h} f_\alpha(u+ch) + \frac{1}{h}\int_0^h \Bigl[\int_0^{u+ct} f_\alpha(u+ct-y)\,\mathrm{d}G(y) + (1-G(u+ct))\Bigr]\lambda \mathrm{e}^{-(\lambda+\alpha)t}\,\mathrm{d}t = 0\,.$$

Letting $h \to 0$ shows that $f_\alpha(u)$ is differentiable and that (4.15) for the derivative from the right holds. Replacing u by $u-ch$ shows the derivative from the left. □

This differential equation is hard to solve. Let us take the Laplace transform with respect to the initial capital. Let $\hat{f}_\alpha(s) = \int_0^\infty \mathrm{e}^{-su} f_\alpha(u)\,\mathrm{d}u$. For the moment we assume $s > 0$. Note that (see Section 4.9)

$$\int_0^\infty f'_\alpha(u)\mathrm{e}^{-su}\,\mathrm{d}u = s\hat{f}_\alpha(s) - f_\alpha(0)\;,$$

$$\int_0^\infty \int_0^u f_\alpha(u-y)\,\mathrm{d}G(y)\mathrm{e}^{-su}\,\mathrm{d}u = \hat{f}_\alpha(s) M_Y(-s)$$

and

$$\int_0^\infty \int_u^\infty \mathrm{d}G(y)\mathrm{e}^{-su}\,\mathrm{d}u = \int_0^\infty \int_0^y \mathrm{e}^{-su}\,\mathrm{d}u\,\mathrm{d}G(y) = \frac{1-M_Y(-s)}{s}\;.$$

Multiplying (4.15) by e^{-su} and integrating yields

$$c(s\hat{f}_\alpha(s) - f_\alpha(0)) + \lambda\Bigl[\hat{f}_\alpha(s)M_Y(-s) + \frac{1-M_Y(-s)}{s} - \hat{f}_\alpha(s)\Bigr] - \alpha\hat{f}_\alpha(s) = 0\,.$$

Solving for $\hat{f}_\alpha(s)$ yields

$$\hat{f}_\alpha(s) = \frac{cf_\alpha(0) - \lambda s^{-1}(1-M_Y(-s))}{cs - \lambda(1-M_Y(-s)) - \alpha}\;. \tag{4.16}$$

We know that $\hat{f}_\alpha(s)$ exists if $\alpha > 0$ and $s > 0$ and is positive. The denominator

$$cs - \lambda(1-M_Y(-s)) - \alpha$$

is convex, has value $-\alpha < 0$ at 0 and converges to ∞ as $s \to \infty$. Thus there exists a strictly positive root $s(\alpha)$ of the denominator. Because $\hat{f}_\alpha(s)$ exists also for $s = s(\alpha)$ the numerator must have a root at $s(\alpha)$ too. Thus

$$cf_\alpha(0) = \lambda s(\alpha)^{-1}(1 - M_Y(-s(\alpha)))\,.$$

The function $s(\alpha)$ is differentiable by the implicit function theorem

$$s'(\alpha)(c - \lambda M_Y'(-s(\alpha))) - 1 = 0\,. \tag{4.17}$$

Because $s(0) = 0$ we obtain that $\lim_{\alpha \to 0} s(\alpha) = 0$.

Example 4.18. Let the claims be $\mathrm{Exp}(\beta)$ distributed. We have to solve

$$cs - \frac{\lambda s}{\beta + s} - \alpha = 0$$

which admits the two solutions

$$s_\pm(\alpha) = \frac{-(\beta c - \lambda - \alpha) \pm \sqrt{(\beta c - \lambda - \alpha)^2 + 4\alpha\beta c}}{2c}$$

where $s_-(\alpha) < 0 \le s_+(\alpha)$. Thus

$$\hat{f}_\alpha(s) = \frac{\lambda/(\beta + s_+(\alpha)) - \lambda/(\beta + s)}{cs - \lambda s/(\beta + s) - \alpha} = \frac{\lambda}{c(\beta + s_+(\alpha))(s - s_-(\alpha))}\,.$$

It follows that

$$f_\alpha(u) = \frac{\lambda}{c(\beta + s_+(\alpha))}\mathrm{e}^{s_-(\alpha)\,u}\,.$$

■

Noting that

$$\mathbb{E}[\tau 1\!\!1_{\{\tau<\infty\}}] = \lim_{\alpha \downarrow 0} \mathbb{E}[\tau \mathrm{e}^{-\alpha\tau} 1\!\!1_{\{\tau<\infty\}}] = \lim_{\alpha \downarrow 0} -\frac{\mathrm{d}}{\mathrm{d}\alpha}\mathbb{E}[\mathrm{e}^{-\alpha\tau} 1\!\!1_{\{\tau<\infty\}}]$$

we see

$$\int_0^\infty \mathbb{E}[\tau 1\!\!1_{\{\tau<\infty\}} \mid C_0 = u]\mathrm{e}^{-su}\,\mathrm{d}u = \lim_{\alpha\downarrow 0} -\frac{\mathrm{d}}{\mathrm{d}\alpha}\int_0^\infty f_\alpha(u)\mathrm{e}^{-su}\,\mathrm{d}u = \lim_{\alpha \downarrow 0} -\frac{\mathrm{d}}{\mathrm{d}\alpha}\hat{f}_\alpha(s)\,.$$

We can find the following ‘explicit’ formula.

Lemma 4.19. *Assume $\mu_2 < \infty$. Then*

$$\mathbb{E}[\tau 1\!\!1_{\{\tau<\infty\}}] = \frac{1}{c-\lambda\mu}\Bigl[\frac{\lambda\mu_2}{2(c-\lambda\mu)}\delta(u) - \int_0^u \psi(u-y)\delta(y)\,\mathrm{d}y\Bigr]\,. \tag{4.18}$$

Proof. We get

$$\begin{aligned}-\frac{\mathrm{d}}{\mathrm{d}\alpha}\hat{f}_\alpha(s) = {}& \frac{\lambda s'(\alpha)}{cs-\lambda(1-M_Y(-s))-\alpha}\,\frac{1-M_Y(-s(\alpha))-M_Y'(-s(\alpha))s(\alpha)}{s(\alpha)^2} \\ &- \frac{\lambda((1-M_Y(-s(\alpha)))s(\alpha)^{-1}-(1-M_Y(-s))s^{-1})}{(cs-\lambda(1-M_Y(-s))-\alpha)^2}\,. \end{aligned}\tag{4.19}$$

It follows from (4.17) that

$$s'(0) = \frac{1}{c-\lambda\mu}$$

and we have already seen that

$$\lim_{\alpha\downarrow 0}\frac{1-M_Y(-s(\alpha))}{s(\alpha)} = \lim_{s\downarrow 0}\frac{1-M_Y(-s)}{s} = \lim_{s\downarrow 0}M_Y'(-s) = \mu\,.$$

Moreover,

$$\begin{aligned}\lim_{\alpha\downarrow 0}\frac{(1-M_Y(-s(\alpha)))-M_Y'(-s(\alpha))s(\alpha)}{s(\alpha)^2} &= \lim_{s\downarrow 0}\frac{(1-M_Y(-s))-M_Y'(-s)s}{s^2} \\ &= \lim_{s\downarrow 0}\frac{sM_Y''(-s)}{2s} = \frac{\mu_2}{2}\,.\end{aligned}$$

Letting α tend to 0 in (4.19) yields

$$\begin{aligned}&\int_0^\infty \mathbb{E}[\tau 1\!\!1_{\{\tau<\infty\}} \mid C_0 = u]\,\mathrm{e}^{-su}\,\mathrm{d}u \\ &\quad= \frac{\lambda\mu_2}{2(c-\lambda\mu)(cs-\lambda(1-M_Y(-s)))} - \frac{\lambda\mu-\lambda s^{-1}(1-M_Y(-s))}{(cs-\lambda(1-M_Y(-s)))^2} \\ &\quad= \frac{\lambda\mu_2}{2(c-\lambda\mu)^2}\,\frac{c-\lambda\mu}{cs-\lambda(1-M_Y(-s))} \\ &\qquad\qquad - \frac{1}{c-\lambda\mu}\,\frac{c-\lambda\mu}{cs-\lambda(1-M_Y(-s))}\Bigl(\frac{1}{s}-\frac{c-\lambda\mu}{cs-\lambda(1-M_Y(-s))}\Bigr) \\ &\quad= \frac{1}{c-\lambda\mu}\Bigl[\frac{\lambda\mu_2}{2(c-\lambda\mu)}\hat{\delta}(s) - \hat{\delta}(s)\hat{\psi}(s)\Bigr]\,.\end{aligned}$$

But this is the Laplace transform of the assertion. □

Corollary 4.20. *Let $t > 0$. Then*

$$\mathbb{P}[t < \tau < \infty] < \frac{\lambda\mu_2}{2(c-\lambda\mu)^2 t}\,.$$

Proof. Because $\delta(u)$ and $\psi(u)$ take values in $[0,1]$ it is clear from (4.18) that

$$\mathbb{E}[\tau 1\!\!1_{\{\tau<\infty\}}] < \frac{\lambda\mu_2}{2(c-\lambda\mu)^2}\,.$$

By Markov's inequality

$$\mathbb{P}[t < \tau < \infty] = \mathbb{P}[\tau 1\!\!1_{\{\tau<\infty\}} > t] < \frac{1}{t}\mathbb{E}[\tau 1\!\!1_{\{\tau<\infty\}}] < \frac{\lambda\mu_2}{2(c-\lambda\mu)^2 t}\,.$$

□

From (4.18) it is possible to get an explicit expression for $u = 0$

$$\mathbb{E}[\tau 1\!\!1_{\{\tau<\infty\}} \mid C_0 = 0] = \frac{\lambda\mu_2}{2(c-\lambda\mu)^2}\Bigl(1 - \frac{\lambda\mu}{c}\Bigr) = \frac{\lambda\mu_2}{2c(c-\lambda\mu)}$$

and

$$\mathbb{E}[\tau \mid \tau < \infty, C_0 = 0] = \frac{\mu_2}{2\mu(c-\lambda\mu)}\,.$$

Example 4.18 (continued). For Exp(β) distributed claims we get for $R = \beta - \lambda/c$

$$\begin{aligned}
&\mathbb{E}[\tau 1\!\!1_{\{\tau<\infty\}}]\\
&\quad= \frac{\beta}{c\beta-\lambda}\Bigl[\frac{2\lambda}{2\beta(c\beta-\lambda)}\Bigl(1-\frac{\lambda}{\beta c}e^{-Ru}\Bigr) - \int_0^u \frac{\lambda}{\beta c}e^{-R(u-y)}\Bigl(1-\frac{\lambda}{\beta c}e^{-Ry}\Bigr)\,dy\Bigr]\\
&\quad= \frac{\beta}{c\beta-\lambda}\Bigl[\frac{\lambda}{\beta(c\beta-\lambda)}\Bigl(1-\frac{\lambda}{\beta c}e^{-Ru}\Bigr) - \frac{\lambda}{\beta c}e^{-Ru}\Bigl(\frac{c}{\beta c-\lambda}(e^{Ru}-1) - \frac{\lambda}{\beta c}u\Bigr)\Bigr]\\
&\quad= \frac{\lambda}{\beta c^2(\beta c-\lambda)}e^{-Ru}(\lambda u + c) = \frac{1}{c(\beta c-\lambda)}\psi(u)(\lambda u + c)\,.
\end{aligned}$$

The conditional expectation of the time of ruin is linear in u

$$\mathbb{E}[\tau \mid \tau < \infty] = \frac{\lambda u + c}{c(\beta c-\lambda)}\,.$$

■

4.13. Seal's Formulae

We consider now the probability of ruin within finite time $\psi(u,t)$. But first we want to find the conditional finite ruin probability given C_t for some t fixed.

Lemma 4.21. *Let t be fixed, $u = 0$ and $0 < y \le ct$. Then*

$$\mathbb{P}[C_s \ge 0, 0 \le s \le t \mid C_t = y] = \frac{y}{ct}\,.$$

Proof. Consider

$$C_t - C_{(t-s)-} = cs - \sum_{i=N_{(t-s)-}+1}^{N_t} Y_i\,.$$

This is also a Cramér-Lundberg model. Thus

$$\begin{aligned}\mathbb{P}[C_s \ge 0, 0 \le s \le t \mid C_t = y] &= \mathbb{P}[C_t - C_{t-s} \ge 0, 0 \le s \le t \mid C_t - C_0 = y]\\ &= \mathbb{P}[C_s \le C_t, 0 \le s \le t \mid C_t = y]\,.\end{aligned}$$

Let $S_s = cs - C_s$ denote the aggregate claims up to time s. Denote by σ the permutations of $\{1, 2, \dots, n\}$. Then

$$\begin{aligned}\mathbb{E}[S_s \mid S_t = y, N_t = n, N_s = k] &= \frac{1}{n!}\mathbb{E}\Bigl[\sum_\sigma \sum_{i=1}^k Y_{\sigma(i)} \Bigm| S_t = y, N_t = n, N_s = k\Bigr]\\ &= \frac{k(n-1)!}{n!}\mathbb{E}\Bigl[\sum_{i=1}^n Y_i \Bigm| S_t = y, N_t = n, N_s = k\Bigr] = \frac{ky}{n}\,.\end{aligned}$$

Because, given $N_t = n$, the claim times are uniformly distributed in $[0,t]$ (Proposition C.2) we obtain

$$\begin{aligned}\mathbb{E}[S_s \mid S_t = y, N_t = n] &= \sum_{k=0}^n \binom{n}{k}\Bigl(\frac{s}{t}\Bigr)^k \Bigl(\frac{t-s}{t}\Bigr)^{n-k}\frac{ky}{n}\\ &= \sum_{k=1}^n \binom{n-1}{k-1}\Bigl(\frac{s}{t}\Bigr)^k \Bigl(\frac{t-s}{t}\Bigr)^{n-k} y = \frac{sy}{t}\,.\end{aligned}$$

This is independent of n. Thus

$$\mathbb{E}[S_s \mid S_t = y] = \frac{sy}{t}$$

and

$$\mathbb{E}[C_s \mid C_t = y] = \mathbb{E}[cs - S_s \mid S_t = ct - y] = cs - \frac{s(ct-y)}{t} = \frac{sy}{t}\,.$$

Then for $v \le s < t$

$$\mathbb{E}\Big[\frac{y - C_s}{t - s} \;\Big|\; C_v, C_t = y\Big] = \frac{y - C_v - \mathbb{E}[(C_s - C_v) \mid C_v, C_t = y]}{t - s}$$
$$= \frac{y - C_v - (s - v)(y - C_v)/(t - v)}{t - s} = \frac{y - C_v}{t - v}$$

and the process

$$M_s = \frac{y - C_s}{t - s} \qquad (0 \le s < t)$$

is a conditional martingale. Note that $\lim_{s\uparrow t} M_s = c$. Let $T = \inf\{s \ge 0 : C_s = y\}$. Because

$$0 \le M_s 1\!\!1_{\{T>s\}} \le c$$

on $\{C_t = y\}$ we have

$$\lim_{s\uparrow t} \mathbb{E}[M_s 1\!\!1_{\{T>s\}} \mid C_t = y] = \mathbb{E}\Big[\lim_{s\uparrow t} M_s 1\!\!1_{\{T>s\}} \;\Big|\; C_t = y\Big] = c\mathbb{P}[T = t \mid C_t = y]\,.$$

Note that by the stopping theorem $\{M_{T\wedge t}\}$ is a bounded martingale. Thus because $M_T = 0$ on $\{T < t\}$

$$\frac{y}{t} = M_0 = \lim_{s\uparrow t} \mathbb{E}[M_{T\wedge s} \mid C_t = y] = \lim_{s\uparrow t} \mathbb{E}[M_s 1\!\!1_{\{T>s\}} \mid C_t = y] = c\mathbb{P}[T = t \mid C_t = y]\,.$$

Note that $\mathbb{P}[T = t \mid C_t = y] = \mathbb{P}[C_s \le C_t, 0 \le s \le t \mid C_t = y]$. □

Conditioning on C_t is in fact a conditioning on the aggregate claim size S_t. Denote by $F(x;t)$ the distribution function of S_t. For the integration we use $\mathrm{d}F(\cdot;t)$ for an integration with respect to the measure $F(\cdot;t)$. Moreover, let $\delta(u,t) = 1 - \psi(u,t) = \mathbb{P}[\tau > t]$.

Theorem 4.22. *For initial capital 0 we have*

$$\delta(0,t) = \frac{1}{ct}\mathbb{E}[C_t \vee 0] = \frac{1}{ct}\int_0^{ct} F(y;t)\,\mathrm{d}y\,.$$

Proof. By Lemma 4.21 it follows readily that

$$\delta(0,t) = \mathbb{E}[\mathbb{P}[\tau > t \mid C_t]] = \mathbb{E}\Big[\frac{C_t \vee 0}{ct}\Big]\,.$$

Moreover,

$$\mathbb{E}[C_t \vee 0] = \mathbb{E}[(ct - S_t) \vee 0] = \int_0^{ct} (ct - z)\,\mathrm{d}F(z;t)$$
$$= \int_0^{ct}\int_z^{ct} \mathrm{d}y\,\mathrm{d}F(z;t) = \int_0^{ct}\int_0^{y} \mathrm{d}F(z;t)\,\mathrm{d}y = \int_0^{ct} F(y;t)\,\mathrm{d}y\,.$$

□

Let now the initial capital be arbitrary.

Theorem 4.23. *With the notation used above we have for $u > 0$*

$$\delta(u,t) = F(u+ct;t) - \int_u^{u+ct} \delta\Bigl(0, t - \frac{v-u}{c}\Bigr) F\Bigl(\mathrm{d}v; \frac{v-u}{c}\Bigr)\,.$$

Proof. Note that

$$\delta(u,t) = \mathbb{P}[C_t > 0] - \mathbb{P}[\exists 0 \le s < t : C_s = 0, C_v > 0 \text{ for } s < v \le t]$$

and $\mathbb{P}[C_t > 0] = F(u+ct;t)$. Let $T = \sup\{0 \le s \le t : C_s = 0\}$ and set $T = \infty$ if $C_s > 0$ for all $s \in [0,t]$. Then, noting that $\mathbb{P}[C_s > 0 : 0 < s \le t \mid C_0 = 0] = \mathbb{P}[C_s \ge 0 : 0 < s \le t \mid C_0 = 0]$,

$$\begin{aligned}\mathbb{P}[T \in [s, s+\mathrm{d}s)] &= \mathbb{P}[C_s \in (-c\,\mathrm{d}s, 0], C_v > 0 \text{ for } v \in [s+\mathrm{d}s, t]] \\ &= (F(u+c(s+\mathrm{d}s);s) - F(u+cs;s))\delta(0, t-s-\mathrm{d}s) \\ &= \delta(0,t-s)c\,\mathrm{d}F(u+cs;s)\,.\end{aligned}$$

Thus

$$\begin{aligned}\delta(u,t) &= F(u+ct;t) - \int_0^t \delta(0,t-s)c\,\mathrm{d}F(u+cs;s) \\ &= F(u+ct;t) - \int_u^{u+ct} \delta\Bigl(0, t-\frac{v-u}{c}\Bigr)\,\mathrm{d}F\Bigl(v, \frac{v-u}{c}\Bigr)\,.\end{aligned}$$

□

4.14. Finite Time Lundberg Inequalities

In this section we derive an upper bound for probabilities of the form

$$\mathbb{P}[\underline{y}u < \tau \le \bar{y}u]$$

for $0 \le \underline{y} < \bar{y} < \infty$. Assume that the Lundberg exponent R exists. We will use the martingale (4.4) and the stopping theorem. For $r \ge 0$ such that $M_Y(r) < \infty$,

$$\begin{aligned}\mathrm{e}^{-ru} &= \mathbb{E}\left[\mathrm{e}^{-rC_{\tau\wedge\bar{y}u} - \theta(r)(\tau\wedge\bar{y}u)}\right] > \mathbb{E}\left[\mathrm{e}^{-rC_{\tau\wedge\bar{y}u} - \theta(r)(\tau\wedge\bar{y}u)}; \underline{y}u < \tau \le \bar{y}u\right] \\ &= \mathbb{E}\left[\mathrm{e}^{-rC_\tau - \theta(r)\tau} \,\middle|\, \underline{y}u < \tau \le \bar{y}u\right] \mathbb{P}[\underline{y}u < \tau \le \bar{y}u] \\ &> \mathbb{E}\left[\mathrm{e}^{-\theta(r)\tau} \,\middle|\, \underline{y}u < \tau \le \bar{y}u\right] \mathbb{P}[\underline{y}u < \tau \le \bar{y}u] \\ &> \mathrm{e}^{-\max\{\theta(r)\underline{y}u, \theta(r)\bar{y}u\}} \mathbb{P}[\underline{y}u < \tau \le \bar{y}u]\,.\end{aligned}$$

Thus

$$\mathbb{P}[\underline{y}u < \tau \leq \bar{y}u \mid C_0 = u] < \mathrm{e}^{-(r-\max\{\theta(r)\underline{y},\theta(r)\bar{y}\})u}\,.$$

Choosing the exponent as small as possible we obtain

$$\mathbb{P}[\underline{y}u < \tau \leq \bar{y}u \mid C_0 = u] < \mathrm{e}^{-R(\underline{y},\bar{y})u} \tag{4.20}$$

where

$$R(\underline{y},\bar{y}) = \sup_{r\in\mathbb{R}} \min\{r - \theta(r)\underline{y}, r - \theta(r)\bar{y}\}\,.$$

The supremum over $r \in \mathbb{R}$ makes sense because $\theta(r) > 0$ for $r < 0$ and $\theta(r) = \infty$ if $M_Y(r) = \infty$. Since $\theta(R) = 0$ it follows that $R(\underline{y},\bar{y}) \geq R$. Hence (4.20) could be useful.

For further investigation of $R(\underline{y},\bar{y})$ we consider the function $f_y(r) = r - y\theta(r)$. Let $r_\infty = \sup\{r \geq 0 : M_Y(r) < \infty\}$. Then $f_y(r)$ exists in the interval $(-\infty, r_\infty)$ and $f_y(r) = -\infty$ for $r \in (r_\infty, \infty)$. If $M_Y(r_\infty) < \infty$ then $|f_y(r_\infty)| < \infty$. Since $f_y''(r) = -yM_Y''(r) < 0$ the function $f_y(r)$ is concave. Thus there exists a unique r_y such that $f_y(r_y) = \sup\{f_y(r) : r \in \mathbb{R}\}$. Because $f_y(0) = 0$ and $f_y'(0) = 1 - y\theta'(0) = 1 + y(c - \lambda\mu) > 0$ we conclude that $r_y > 0$. r_y is either the solution to the equation

$$1 + y(c - \lambda M_Y'(r_y)) = 0$$

or $r_y = r_\infty$. We assume now that $R < r_\infty$ and let

$$y_0 = (\lambda M_Y'(R) - c)^{-1}\,.$$

It follows $r_{y_0} = R$. We call y_0 the **critical value**.

Because $\frac{\mathrm{d}}{\mathrm{d}y} f_y(r) = -\theta(r)$ it follows readily that

$$\frac{\mathrm{d}}{\mathrm{d}y} f_y(r) \gtreqless 0 \iff r \lesseqgtr R\,.$$

We get

$$f_{\underline{y}}(r) \lesseqgtr f_{\bar{y}}(r) \text{ as } r \lesseqgtr R\,.$$

Because $M_Y'(r)$ is an increasing function it follows that

$$y \gtreqless y_0 \iff r_y \lesseqgtr R\,.$$

Thus

$$R(\underline{y},\bar{y}) = \begin{cases} R & \text{if } \underline{y} \leq y_0 \leq \bar{y}\,, \\ f_{\bar{y}}(r_{\bar{y}}) & \text{if } \bar{y} < y_0, \\ f_{\underline{y}}(r_{\underline{y}}) & \text{if } \underline{y} > y_0. \end{cases}$$

If $y_0 \in [\underline{y}, \bar{y}]$ then we got again Lundberg's inequality. If $\bar{y} < y_0$ then $R(\underline{y}, \bar{y})$ does not depend on $\underline{y}$. By choosing $\underline{y}$ as small as possible we obtain

$$\mathbb{P}[0 < \tau \le yu \mid C_0 = u] < \mathrm{e}^{-R(0,y)u} \tag{4.21}$$

for $y < y_0$. Note that $R(0, y) > R$. Analogously

$$\mathbb{P}[yu < \tau < \infty \mid C_0 = u] < \mathrm{e}^{-R(y,\infty)u} \tag{4.22}$$

for $y > y_0$. The strict inequality follows from

$$\mathrm{e}^{-ru} \ge \mathbb{E}[\mathrm{e}^{-rC_\tau - \theta(r)\tau}; yu < \tau < \infty] > \mathrm{e}^{-\theta(r)yu}\mathbb{P}[yu < \tau < \infty]$$

if $0 < r < R$, compare also (4.5). Again $R(y, \infty) > R$.

We see that $\mathbb{P}[\tau \notin ((y_0 - \varepsilon)u, (y_0 + \varepsilon)u)]$ goes faster to 0 than $\mathbb{P}[\tau \in ((y_0 - \varepsilon)u, (y_0 + \varepsilon)u)]$. Intuitively ruin should therefore occur close to $y_0 u$.

Theorem 4.24. *Assume that $R < r_\infty$. Then*

$$\frac{\tau}{u} \longrightarrow y_0$$

in probability on the set $\{\tau < \infty\}$ as $u \to \infty$.

Proof. Choose $\varepsilon > 0$. By the Cramér-Lundberg approximation $\mathbb{P}[\tau < \infty \mid C_0 = u]\exp\{Ru\} \to C$ for some $C > 0$. Then

$$\begin{aligned}
&\mathbb{P}\Big[\Big|\frac{\tau}{u} - y_0\Big| > \varepsilon \;\Big|\; \tau < \infty, C_0 = u\Big] \\
&\quad = \frac{\mathbb{P}[\tau < (y_0 - \varepsilon)u \mid C_0 = u] + \mathbb{P}[(y_0 + \varepsilon)u < \tau < \infty \mid C_0 = u]}{\mathbb{P}[\tau < \infty \mid C_0 = u]} \\
&\quad \le \frac{\exp\{-R(0, y_0 - \varepsilon)u\} + \exp\{-R(y_0 + \varepsilon, \infty)u\}}{\mathbb{P}[\tau < \infty \mid C_0 = u]} \\
&\quad = \frac{\exp\{-(R(0, y_0 - \varepsilon) - R)u\} + \exp\{-(R(y_0 + \varepsilon, \infty) - R)u\}}{\mathbb{P}[\tau < \infty \mid C_0 = u]\mathrm{e}^{Ru}} \longrightarrow 0
\end{aligned}$$

as $u \to \infty$. □

Bibliographical Remarks

The Cramér-Lundberg model was introduced by Filip Lundberg [61]. The basic results, like Lundberg's inequality and the Cramér-Lundberg approximation go back to Lundberg [62] and Cramér [24], see also [25], [46] or [66]. The differential equation for the ruin probability (4.2) and the Laplace transform of the ruin probability can for instance be found in [25]. The renewal theory proof of the Cramér-Lundberg approximation is due to Feller [40]. The martingale (4.4) was introduced by Gerber [41]. The latter paper contains also the martingale proof of Lundberg's inequality and the proof of (4.21). The proof of (4.22) goes back to Grandell [47]. Section 4.14 can also be found in [35]. Some similar results are obtained in [4] and [5]. Segerdahl [76] proved the asymptotic normality of the ruin time conditioned on ruin occurs.

The results on reinsurance and ruin can be found in [86] and [22]. Proposition 4.9 can be found in [43, p.130].

The term **severity of ruin** was introduced in [44]. More results on the topic are obtained in [33], [32], [72] and [66]. A diffusion approximation, even though not obtained mathematically correct, was already used by Hadwiger [51]. The modern approach goes back to Iglehart [55], see also [45], [7] and [70]. The deVylder approximation was introduced in [31]. Formula (4.14) was found by Asmussen [7]. It is however stated incorrectly in the paper. For the correct formula see [16] or [66]. The Beekman-Bowers approximation can be found in [17].

Theorem 4.15 was obtain by von Bahr [85] in the special case of Pareto distributed claim sizes and by Thorin and Wikstad [84] in the case of lognormally distributed claims. The theorem in the present form is basically found in [34], see also [38].

Section 4.12 is taken from [69]. Parts can also be found in [70] and [16]. (4.16) was first obtained in [25]. The expected value of the ruin time for exponentially distributed claims can be found in [43, p.138]. Seal's formulae were first obtained by Takács [80, p.56]. Seal [74] and [75] introduced them to risk theory. The present presentation follows [43].

5. The Renewal Risk Model

5.1. Definition of the Renewal Risk Model

The easiest generalisation of the classical risk model is to assume that the claim number process is a renewal process. Then the risk process is not Markovian anymore, because the distribution of the time of the next claim is dependent on the past via the time of the last claim. We therefore cannot find a differential equation for $\psi(u)$ as in the classical case.

Let

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

where

- $\{N_t\}$ is a renewal process with event times $0 = T_0, T_1, T_2, \ldots$ with interarrival distribution function F, mean λ^{-1} and moment generating function $M_T(r)$. We denote by T a generic random variable with distribution function F. The first claim time T_1 has distribution function F^1. If nothing else is said we consider the ordinary case where $F^1 = F$.
- The claim amounts $\{Y_i\}$ build an iid. sequence with distribution function G, moments $\mu_k = \mathbb{E}[Y_i^k]$ and moment generating function $M_Y(r)$. As in the classical model $\mu = \mu_1$.
- $\{N_t\}$ and $\{Y_i\}$ are independent.

This model is called the **renewal risk model** or, after the person who introduced the model, the **Sparre Andersen model**.

As in the classical model we define the ruin time $\tau = \inf\{t > 0 : C_t < 0\}$ and the ruin probabilities $\psi(u,t) = \mathbb{P}[\tau \le t]$ and $\psi(u) = \mathbb{P}[\tau < \infty]$. Considering the random walk $\{C_{T_i}\}$ we can conclude that

$$\psi(u) = 1 \qquad \Longleftrightarrow \qquad \mathbb{E}[C_{T_2} - C_{T_1}] = c/\lambda - \mu \le 0\,.$$

In order to avoid ruin a.s. we assume the net profit condition $c > \lambda\mu$. Note that in contrary to the classical case

$$\mathbb{E}[C_t - u] \neq (c - \lambda\mu)t \text{ for most } t \in \mathbb{R}$$

except if T has an exponential distribution. But

$$\lim_{t\to\infty} \frac{1}{t}(C_t - u) = c - \lambda\mu$$

by the law of large numbers. It follows again that $\psi(u) \to 0$ as $u \to \infty$ because $\inf\{C_t - u, t \ge 0\}$ is finite.

5.2. The Adjustment Coefficient

We were able to construct some martingales of the form $\{\exp\{-rC_t - \theta(r)t\}\}$ in the classical case. Because the risk process $\{C_t\}$ is not Markov anymore we cannot hope to get such a martingale again because Lemma 4.1 does not hold any more. But the claim times are renewal times because $\{C_{T_1+t} - C_{T_1}\}$ is a renewal risk model independent of T_1 and Y_1.

Lemma 5.1. *Let $\{C_t\}$ be an (ordinary) renewal risk model. For any $r \in \mathbb{R}$ such that $\inf\{1/M_T(s) : s \ge 0, M_T(s) < \infty\} \le M_Y(r) < \infty$ let $\theta(r)$ be the unique solution to*

$$M_Y(r)M_T(-\theta(r) - cr) = 1\,. \tag{5.1}$$

Then the discrete time process

$$\{\exp\{-rC_{T_i} - \theta(r)T_i\}\}$$

is a martingale.

Remark. In the classical case with $F(t) = 1 - \mathrm{e}^{-\lambda t}$ equation (5.1) is

$$M_Y(r)\frac{\lambda}{\lambda + \theta(r) + cr} = 1$$

which is equivalent to the equation (4.3) in the classical case. ■

Proof. Note that $1/M_Y(r) \le \sup\{M_T(s) : M_T(s) < \infty\}$. Because $M_T(r)$ is an increasing and continuous function defined for all $r \le 0$ and because $M_T(r) \to 0$ as $r \to -\infty$ there exists a unique solution $\theta(r)$ to (5.1). Then

$$\begin{aligned}\mathbb{E}[\mathrm{e}^{-rC_{T_{i+1}}-\theta(r)T_{i+1}} \mid \mathcal{F}_{T_i}] &= \mathbb{E}[\mathrm{e}^{-r(c(T_{i+1}-T_i)-Y_{i+1})-\theta(r)(T_{i+1}-T_i)} \mid \mathcal{F}_{T_i}]\mathrm{e}^{-rC_{T_i}-\theta(r)T_i}\\ &= \mathbb{E}[\mathrm{e}^{rY_{i+1}}\mathrm{e}^{-(cr+\theta(r))(T_{i+1}-T_i)}]\mathrm{e}^{-rC_{T_i}-\theta(r)T_i}\\ &= M_Y(r)M_T(-cr-\theta(r))\mathrm{e}^{-rC_{T_i}-\theta(r)T_i}\\ &= \mathrm{e}^{-rC_{T_i}-\theta(r)T_i}\,.\end{aligned}$$

□

As in the classical case we are interested in the case $\theta(r) = 0$. In the classical case there existed, besides the trivial solution $r = 0$, at most a second solution to the equation $\theta(r) = 0$. This was the case because $\theta(r)$ was a convex function. Let us therefore compute the second derivative of $\theta(r)$. Define $m_Y(r) = \log M_Y(r)$ and $m_T(r) = \log M_T(r)$. Then $\theta(r)$ is the solution to the equation

$$m_Y(r) + m_T(-\theta(r) - cr) = 0\,.$$

By the implicit function theorem $\theta(r)$ is differentiable and

$$m_Y'(r) - (\theta'(r) + c)m_T'(-\theta(r) - cr) = 0\,. \tag{5.2}$$

Hence $\theta(r)$ is infinitely often differentiable and

$$m_Y''(r) - \theta''(r)m_T'(-\theta(r) - cr) + (\theta'(r) + c)^2 m_T''(-\theta(r) - cr) = 0\,.$$

Note that $M_T(r)$ is a strictly increasing function and so is $m_T(r)$. Thus $m_T'(-\theta(r) - cr) > 0$. By Lemma 1.9 both $m_Y''(r)$ and $m_T''(r)$ are positive. We can assume that not both Y and T are deterministic, otherwise $\psi(u) = 0$. Then at least one of the two functions $m_Y''(r)$ and $m_T''(r)$ is strictly positive. From (5.2) it follows that $\theta'(r) + c > 0$. Thus $\theta''(r) > 0$ and $\theta(r)$ is a strictly convex function.

Let $r = 0$, i.e. $\theta(0) = 0$. Then

$$\theta'(0) = \frac{M_Y'(0)M_T(0)}{M_Y(0)M_T'(0)} - c = \lambda\mu - c < 0 \tag{5.3}$$

by the net profit condition. Thus there exists at most a second solution to $\theta(R) = 0$ with $R > 0$. Again we call this solution **adjustment coefficient** or **Lundberg exponent**. Note that R is the strictly positive solution to the equation

$$M_Y(r)M_T(-cr) = 1\,.$$

Example 5.2. Let $\{C_t\}$ be a renewal risk model with Exp(1) distributed claim amounts, premium rate $c = 5$ and interarrival time distribution

$$F(t) = 1 - \tfrac{1}{2}(\mathrm{e}^{-3t} + \mathrm{e}^{-7t})\,.$$

It follows that $M_Y(r)$ exists for $r < 1$, $M_T(r)$ exists for $r < 3$ and $\lambda = 4.2$. The net profit condition $5 > 4.2$ is fulfilled. The equation to solve is

$$\frac{1}{1-r}\,\frac{1}{2}\Bigl(\frac{3}{3+5r} + \frac{7}{7+5r}\Bigr) = 1\,.$$

Thus

$$3(7+5r)+7(3+5r)=2(1-r)(3+5r)(7+5r)$$

or equivalently

$$25r^3+25r^2-4r=0\,.$$

We find the obvious solution $r=0$ and the two other solutions

$$r_{1/2}=\frac{-25\pm\sqrt{1025}}{50}=\frac{-5\pm\sqrt{41}}{10}=\begin{cases}0.140312,\\-1.14031.\end{cases}$$

From the theory we conclude that $R=0.140312$. But we proved that there is no negative solution. Why do we get $r_2=-1.14031$? Obviously $M_Y(r_2)<1<\infty$. But $-cr_2=5.70156>3$ and thus $M_T(-cr_2)=\infty$. Thus r_2 is not a solution to the equation $M_Y(r)M_T(-cr)=1$. ■

Example 5.3. Let $T\sim\Gamma(\gamma,\alpha)$. Assume that there exists an r_∞ such that $M_Y(r)<\infty\iff r<r_\infty$. Then

$$\lim_{r\uparrow r_\infty}M_Y(r)=\infty\,.$$

Let $R(\gamma,\alpha)$ denote the Lundberg exponent which exists in this case and $\theta(r;\gamma,\alpha)$ denote the corresponding $\theta(r)$. Then for $h>0$

$$\Bigl(\frac{\alpha}{\alpha+cR(\gamma,\alpha)}\Bigr)^{\gamma+h}M_Y(R(\gamma,\alpha))=\Bigl(\frac{\alpha}{\alpha+cR(\gamma,\alpha)}\Bigr)^{h}<1\,.$$

That means that $\theta(R(\gamma,\alpha);\gamma+h,\alpha)<0$ and therefore $R(\gamma+h,\alpha)>R(\gamma,\alpha)$. Similarly

$$\Bigl(\frac{\alpha+h}{\alpha+h+cR(\gamma,\alpha)}\Bigr)^{\gamma}M_Y(R(\gamma,\alpha))>\Bigl(\frac{\alpha}{\alpha+cR(\gamma,\alpha)}\Bigr)^{\gamma}M_Y(R(\gamma,\alpha))=1$$

and thus $R(\gamma,\alpha+h)<R(\gamma,\alpha)$. It would be interesting to see what happens if we let $\alpha\to0$ and $\gamma\to0$ in such a way that the mean value remains constant. Let therefore $\alpha=\lambda\gamma$. Recall that $m_Y(r)=\log M_Y(r)$.

Observe that

$$\lim_{\gamma\to0}\gamma\log\Bigl(\frac{\lambda\gamma}{\lambda\gamma+cr}\Bigr)+m_Y(r)=m_Y(r)$$

and therefore $R(\gamma)=R(\gamma,\lambda\gamma)\to0$ as $\gamma\to0$. The latter follows because convergence to a continuous function is uniform on compact intervals. Let now $r(\gamma)=\gamma^{-1}R(\gamma)$. Then $r(\gamma)$ has to solve the equation

$$\log\Bigl(\frac{\lambda}{\lambda+cr}\Bigr)+\frac{m_Y(\gamma r)}{\gamma}=0\,. \tag{5.4}$$

Letting $\gamma \to 0$ yields

$$\log\Bigl(\frac{\lambda}{\lambda + cr}\Bigr) + r m_Y'(0) = \log\Bigl(\frac{\lambda}{\lambda + cr}\Bigr) + r\mu = 0\,.$$

The latter function is convex, has the root $r = 0$, the derivative $-c/\lambda + \mu < 0$ at 0 and converges to infinity as $r \to \infty$. Thus there exists exactly one additional solution r_0 and $r(\gamma) \to r_0$ as $\gamma \to 0$.

Replace r by $r(\gamma)$ in (5.4) and take the derivative, which exists by the implicit function theorem,

$$-\frac{cr'(\gamma)}{\lambda + cr(\gamma)} + \frac{(\gamma r'(\gamma) + r(\gamma)) m_Y'(\gamma r(\gamma))}{\gamma} - \frac{m_Y(\gamma r(\gamma))}{\gamma^2} = 0\,.$$

One obtains

$$r'(\gamma) = \frac{\frac{\gamma r(\gamma) m_Y'(\gamma r(\gamma)) - m_Y(\gamma r(\gamma))}{\gamma^2}}{\frac{c}{\lambda + cr(\gamma)} - m_Y'(\gamma r(\gamma))}\,.$$

Note that by Taylor expansion

$$\gamma r(\gamma) m_Y'(\gamma r(\gamma)) - m_Y(\gamma r(\gamma)) = \frac{\sigma^2}{2}\gamma^2 r(\gamma)^2 + O(\gamma^3 r(\gamma)^3) = \frac{\sigma^2}{2}\gamma^2 r_0^2 + O(\gamma^3)$$

uniformly on compact intervals for which $R(\gamma)$ exists. The second equality holds because $\lim_{\gamma\downarrow 0} r'(\gamma) = r_0 > 0$ exists. We denoted by σ^2 the variance of the claims. We found

$$r'(\gamma) = \frac{\sigma^2 r_0^2(\lambda + cr_0)}{2(c - \lambda\mu - c\mu r_0)} + O(\gamma)\,.$$

Because the convergence is uniform it follows that

$$r(\gamma) = r_0 + \frac{\sigma^2 r_0^2(\lambda + cr_0)}{2(c - \lambda\mu - c\mu r_0)}\gamma + O(\gamma^2)$$

or

$$R(\gamma) = \gamma r_0 + \frac{\sigma^2 r_0^2(\lambda + cr_0)}{2(c - \lambda\mu - c\mu r_0)}\gamma^2 + O(\gamma^3)\,.$$

Note that the moment generating function of the interarrival times

$$\Bigl(\frac{\lambda\gamma}{\lambda\gamma - r}\Bigr)^{\gamma} = \Bigl(1 - \frac{r}{\lambda\gamma}\Bigr)^{-\gamma} \to \mathrm{e}^{r/\lambda}$$

converges to the moment generating function of deterministic interarrival times $T = 1/\lambda$. That means that $T(\gamma) \to 1/\lambda$ in distribution as $\gamma \to 0$. But the Lundberg exponents do not converge to the Lundberg exponent of the model with deterministic interarrival times. ■

5.3. Lundberg's Inequality

5.3.1. The Ordinary Case

In order to find an upper bound for the ruin probability we try to proceed as in the classical case. Unfortunately we cannot find an analogue to (4.1). But recall the interpretation of (4.2) as conditioning on the ladder heights. Let $H(x) = \mathbb{P}[\tau < \infty, C_\tau \geq -x \mid C_0 = 0]$. Then

$$\psi(u) = \int_0^u \psi(u-x)\,\mathrm{d}H(x) + (H(\infty) - H(u)) \tag{5.5}$$

where $H(\infty) = \mathbb{P}[\tau < \infty \mid C_0 = 0]$. Unfortunately we cannot find an explicit expression for $H(u)$. In order to find a result we have to link the Lundberg exponent R to H.

Lemma 5.4. *Let $\{C_t\}$ be an ordinary renewal risk model with Lundberg exponent R. Then*

$$\int_0^\infty \mathrm{e}^{Rx}\,\mathrm{d}H(x) = \mathbb{E}\left[\mathrm{e}^{-RC_\tau}\mathbb{1}_{\{\tau<\infty\}}\middle| C_0 = 0\right] = 1\,.$$

Proof. Let $B(x)$ be the weekly descending ladder height of the process $\{-C_{T_n}\}$. Multiplying the Wiener-Hopf-factorization (E.1) by e^{rx} and integrating yields

$$M_Y(r)M_T(-cr) = M_H(r) + M_B(r) - M_H(r)M_B(r)\,.$$

We see that $M_H(r) < \infty$ iff $M_Y(r) < \infty$ and $M_B(r) < \infty$ iff $M_T(-cr) < \infty$. For $r = R$ the left hand side is one and we can rewrite the equation as $(M_H(R) - 1)(1 - M_B(R)) = 0$. Since $B(x)$ is distribution on the negative half line we have $M_B(R) < 1$. Thus $M_H(R) = 1$, which is the assertion. □

Also here we can give a martingale proof. The discrete time process $\{\mathrm{e}^{-RC_{T_i}}\}$ is a martingale with initial value 1 if $C_0 = 0$. We will drop the conditioning on $C_0 = 0$ for the rest of the proof. By the stopping theorem

$$1 = \mathbb{E}\left[\mathrm{e}^{-RC_{\tau\wedge T_n}}\right] = \mathbb{E}\left[\mathrm{e}^{-RC_\tau}\mathbb{1}_{\{\tau\leq T_n\}}\right] + \mathbb{E}\left[\mathrm{e}^{-RC_{T_n}}\mathbb{1}_{\{\tau>T_n\}}\right]\,.$$

The function $\mathrm{e}^{-RC_\tau}\mathbb{1}_{\{\tau\leq T_n\}}$ is increasing with n. By the monotone limit theorem it follows that

$$\lim_{n\to\infty}\mathbb{E}[\mathrm{e}^{-RC_\tau}\mathbb{1}_{\{\tau\leq T_n\}}] = \mathbb{E}\Big[\lim_{n\to\infty}\mathrm{e}^{-RC_\tau}\mathbb{1}_{\{\tau\leq T_n\}}\Big] = \mathbb{E}[\mathrm{e}^{-RC_\tau}\mathbb{1}_{\{\tau<\infty\}}]\,.$$

For the second term we have $\mathrm{e}^{-RC_{T_n}}\mathbb{1}_{\{\tau>T_n\}} \le 1$ and we know that $C_{T_n} \to \infty$ as $n \to \infty$. Thus

$$\lim_{n\to\infty} \mathbb{E}[\mathrm{e}^{-RC_{T_n}}\mathbb{1}_{\{\tau>T_n\}}] = \mathbb{E}\Bigl[\lim_{n\to\infty} \mathrm{e}^{-RC_{T_n}}\mathbb{1}_{\{\tau>T_n\}}\Bigr] = 0\,.$$

We are now ready to prove Lundberg's inequality.

Theorem 5.5. *Let $\{C_t\}$ be an ordinary renewal risk model and assume that the Lundberg exponent R exists. Then*

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

Proof. Assume that the assertion were wrong and let

$$u_0 = \inf\{u \ge 0 : \psi(u) \ge \mathrm{e}^{-Ru}\}\,.$$

Let $u_0 \le u < u_0 + \varepsilon$ such that $\psi(u) \ge \mathrm{e}^{-Ru}$. Then

$$\psi(u_0) \ge \psi(u) \ge \mathrm{e}^{-Ru} > \mathrm{e}^{-R(u_0+\varepsilon)}\,.$$

Thus $\psi(u_0) \ge \mathrm{e}^{-Ru_0}$. Because $\mathrm{e}^{-RC_\tau} > 1$ it follows from Lemma 5.4 that $\psi(0) = H(\infty) < 1$ and therefore $u_0 \ge -R^{-1}\log H(\infty) > 0$. Assume first that $H(u_0-) = 0$. Then it follows from (5.5) that

$$\begin{aligned}\mathrm{e}^{-Ru_0} \le \psi(u_0) &= \psi(0)H(u_0) + \int_{u_0}^\infty \mathrm{d}H(x) < H(u_0) + \int_{u_0}^\infty \mathrm{e}^{R(x-u_0)}\,\mathrm{d}H(x)\\ &= \int_{u_0-}^\infty \mathrm{e}^{R(x-u_0)}\,\mathrm{d}H(x) = \mathrm{e}^{-Ru_0}\int_0^\infty \mathrm{e}^{Rx}\,\mathrm{d}H(x) = \mathrm{e}^{-Ru_0}\end{aligned}$$

which is a contradiction. Thus $H(u_0-) > 0$.

As in the classical case we obtain from (5.5) and Lemma 5.4

$$\begin{aligned}\mathrm{e}^{-Ru_0} \le \psi(u_0) &= \int_0^{u_0} \psi(u_0 - x)\,\mathrm{d}H(x) + \int_{u_0}^\infty \mathrm{d}H(x)\\ &< \int_0^{u_0} \mathrm{e}^{-R(u_0-x)}\,\mathrm{d}H(x) + \int_{u_0}^\infty \mathrm{d}H(x)\\ &\le \int_0^\infty \mathrm{e}^{-R(u_0-x)}\,\mathrm{d}H(x) = \mathrm{e}^{-Ru_0}\,.\end{aligned}$$

This is a contradiction and the theorem is proved. □

As in the classical case we prove the result using martingale techniques. By the stopping theorem

$$\begin{aligned}\mathrm{e}^{-Ru} = \mathrm{e}^{-RC_{T_0}} &= \mathbb{E}\left[\mathrm{e}^{-RC_{\tau\wedge T_n}}\right] \ge \mathbb{E}\left[\mathrm{e}^{-RC_{\tau\wedge T_n}}; \tau \le T_n\right] \\ &= \mathbb{E}\left[\mathrm{e}^{-RC_\tau}; \tau \le T_n\right] .\end{aligned}$$

Letting $n \to \infty$ yields by monotone convergence

$$\mathrm{e}^{-Ru} \ge \mathbb{E}\left[\mathrm{e}^{-RC_\tau}; \tau < \infty\right] > \mathbb{P}[\tau < \infty]$$

because $C_\tau < 0$.

Example 5.6. Let Y_i be $\mathrm{Exp}(\alpha)$ distributed and

$$F(t) = 1 - p\mathrm{e}^{-\beta t} - (1-p)\mathrm{e}^{-\gamma t}$$

where $0 < \beta < \gamma$ and $0 < p \le \gamma(\gamma-\beta)^{-1}$. The net profit condition can then be written as

$$\alpha c(\beta + p(\gamma-\beta)) > \beta\gamma .$$

The equation for the adjustment coefficient is

$$\frac{\alpha}{\alpha - r}\Bigl(\frac{p\beta}{\beta + cr} + \frac{(1-p)\gamma}{\gamma + cr}\Bigr) = 1 .$$

This is equivalent to

$$c^2r^3 - c(\alpha c - \beta - \gamma)r^2 - (\alpha c(\beta + p(\gamma-\beta)) - \beta\gamma)r = 0$$

from which the solutions $r_0 = 0$ and

$$r_{1/2} = \frac{\alpha c - \beta - \gamma \pm \sqrt{(\alpha c - \gamma + \beta)^2 + 4\alpha c p(\gamma - \beta)}}{2c}$$

follow. Note that by the net profit condition $r_1 > 0 > r_2$. By the condition on p

$$\begin{aligned} r_2 < r_1 &\le \frac{\alpha c - \beta - \gamma + \sqrt{(\alpha c - \gamma + \beta)^2 + 4\alpha c\gamma}}{2c} \\ &< \frac{\alpha c - \beta - \gamma + \sqrt{(\alpha c + \gamma + \beta)^2}}{2c} = \alpha .\end{aligned}$$

Thus $M_Y(r_{1/2}) < \infty$. And

$$\begin{aligned} -cr_2 &= \frac{\beta + \gamma - \alpha c + \sqrt{(\alpha c - \gamma + \beta)^2 + 4\alpha c p(\gamma - \beta)}}{2} \\ &> \frac{\beta + \gamma - \alpha c + \sqrt{(\alpha c - \gamma + \beta)^2}}{2} \ge \beta \end{aligned}$$

and thus $M_T(-cr_2) = \infty$, i.e. r_2 is not a solution. Therefore

$$R = \frac{\alpha c - \beta - \gamma + \sqrt{(\alpha c - \gamma + \beta)^2 + 4\alpha cp(\gamma - \beta)}}{2c}\,.$$

It follows from Lundberg's inequality that

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

■

5.3.2. The General Case

Let now F^1 be arbitrary. Let us denote the ruin probability in the ordinary model by $\psi^{\mathrm{o}}(u)$. We know that $\{C_{T_1+t}\}$ is an ordinary model with initial capital C_{T_1}. There are two possibilities: Ruin occurs at T_1 or $C_{T_1} \ge 0$. Thus

$$\begin{aligned}
\psi(u) &= \int_0^\infty \Bigl(\int_0^{u+ct} \psi^{\mathrm{o}}(u + ct - y)\,\mathrm{d}G(y) + \int_{u+ct}^\infty \mathrm{d}G(y)\Bigr)\,\mathrm{d}F^1(t)\\
&< \int_0^\infty \Bigl(\int_0^{u+ct} \mathrm{e}^{-R(u+ct-y)}\,\mathrm{d}G(y) + \int_{u+ct}^\infty \mathrm{e}^{-R(u+ct-y)}\,\mathrm{d}G(y)\Bigr)\,\mathrm{d}F^1(t)\\
&= \int_0^\infty \int_0^\infty \mathrm{e}^{-R(u+ct-y)}\,\mathrm{d}G(y)\,\mathrm{d}F^1(t)\\
&= \mathrm{e}^{-Ru}\mathbb{E}\left[\mathrm{e}^{R(Y_1-cT_1)}\right] = M_Y(R)M_{T_1}(-cR)\mathrm{e}^{-Ru}\,.
\end{aligned}$$

Thus $\psi(u) < C\mathrm{e}^{-Ru}$ for $C = M_Y(R)M_{T_1}(-cR)$. In the cases considered so far we always had $C = 1$. But in the general case $C \in (0, M_Y(R))$ can be any value of this interval (let T_1 be deterministic). If $F^1(t) = F(t)$ then $C = 1$ by the equation determining the adjustment coefficient.

5.4. The Cramér-Lundberg Approximation

5.4.1. The Ordinary Case

In order to obtain the Cramér-Lundberg approximation we proceed as in the classical case and multiply (5.5) by e^{Ru}.

$$\psi(u)\mathrm{e}^{Ru} = \int_0^u \psi(u-x)\mathrm{e}^{R(u-x)}\mathrm{e}^{Rx}\,\mathrm{d}H(x) + \mathrm{e}^{Ru}(H(\infty) - H(u))\,.$$

We obtain the following

Theorem 5.7. *Let $\{C_t\}$ be an ordinary renewal risk model. Assume that R exists and that there exists an $r > R$ such that $M_Y(r) < \infty$. If the distribution of $Y_1 - cT_1$ given $Y_1 - cT_1 > 0$ is not arithmetic then*

$$\lim_{u\to\infty} \psi(u)\mathrm{e}^{Ru} = \frac{1 - H(\infty)}{R\int_0^\infty x\mathrm{e}^{Rx}\,\mathrm{d}H(x)} =: C\,.$$

If the distribution of $Y_1 - cT_1$ is arithmetic with span γ then for $x \in [0,\gamma)$

$$\lim_{n\to\infty} \psi(x + n\gamma)\mathrm{e}^{R(x+n\gamma)} = C\mathrm{e}^{Rx}\frac{1 - \mathrm{e}^{-R\gamma}}{R}\,.$$

Proof. We prove only the non-arithmetic case. The arithmetic case can be proved similarly. Recall from Lemma 5.4 that $\int_0^x \mathrm{e}^{Ry}\,\mathrm{d}H(y)$ is a proper distribution function. It follows as in the classical case that $\mathrm{e}^{Ru}(H(\infty) - H(u))$ is directly Riemann integrable. From $M_Y(r) < \infty$ for an $r > R$ it follows that $\int_0^\infty x\mathrm{e}^{Rx}\,\mathrm{d}H(x) < \infty$. Thus by the renewal theorem

$$\lim_{u\to\infty} \psi(u)\mathrm{e}^{Ru} = \frac{\int_0^\infty \mathrm{e}^{Ru}(H(\infty) - H(u))\,\mathrm{d}u}{\int_0^\infty x\mathrm{e}^{Rx}\,\mathrm{d}H(x)}\,.$$

Simplifying the numerator we find

$$\int_0^\infty \mathrm{e}^{Ru}\int_u^\infty \mathrm{d}H(x)\,\mathrm{d}u = \int_0^\infty\int_0^x \mathrm{e}^{Ru}\,\mathrm{d}u\,\mathrm{d}H(x) = \frac{1}{R}(1 - H(\infty))$$

where we used Lemma 5.4 in the last equality. □

It should be remarked that in general there is no explicit expression for the ladder height distribution $H(x)$. It is therefore difficult to find an explicit expression for C. Nevertheless we know that $0 < C \le 1$.

Assume that we know C. Then

$$\psi(u) \approx C\mathrm{e}^{-Ru}$$

for u large.

Example 5.8. Assume that Y_i is Exp(α) distributed. Because $M_Y(r) \to \infty$ as $r \to \alpha$ it follows that R exists and that the conditions of Theorem 5.7 are fulfilled. Consider the martingale $\{\mathrm{e}^{-RC_{T_i}}\}$. By the stopping theorem

$$\mathrm{e}^{-Ru} = \mathbb{E}\left[\mathrm{e}^{-RC_{\tau\wedge T_i}}\right] = \mathbb{E}\left[\mathrm{e}^{-RC_\tau}\mathbb{1}_{\{\tau\le T_i\}}\right] + \mathbb{E}\left[\mathrm{e}^{-RC_{T_i}}\mathbb{1}_{\{\tau > T_i\}}\right]\,.$$

As in the proof of Lemma 5.4 it follows that

$$\mathrm{e}^{-Ru} = \mathbb{E}\left[\mathrm{e}^{-RC_\tau}\mathbb{1}_{\{\tau<\infty\}}\right] = \mathbb{E}\left[\mathrm{e}^{-RC_\tau}\,\middle|\,\tau<\infty\right]\psi(u)\,.$$

Assume for the moment that we know $C_{\tau-}$. Let $Z = C_{\tau-} - C_\tau$ denote the size of the claim leading to ruin. The only information on Z we have is that $Z > C_{\tau-}$ because ruin occurs at time τ. Then

$$\begin{aligned}\mathbb{P}[-C_\tau > x \mid C_{\tau-} = y, \tau < \infty] &= \mathbb{P}[Z > y + x \mid C_{\tau-} = y, \tau < \infty] \\ &= \mathbb{P}[Y_1 > y + x \mid Y_1 > y] = \mathrm{e}^{-\alpha x}\,.\end{aligned}$$

It follows that

$$\mathbb{E}\left[\mathrm{e}^{-RC_\tau} \mid \tau < \infty\right] = \int_0^\infty \mathrm{e}^{Rx}\alpha\mathrm{e}^{-\alpha x}\,\mathrm{d}x = \frac{\alpha}{\alpha - R}\,.$$

Thus we obtain the explicit solution

$$\psi(u) = \frac{\alpha - R}{\alpha}\mathrm{e}^{-Ru}$$

to the ruin problem. Therefore, as in the classical case, the Cramér-Lundberg approximation is exact and we have an explicit expression for C. ■

5.4.2. The general case

Consider the equation

$$\psi(u) = \int_0^\infty \Bigl(\int_0^{u+ct} \psi^{\mathrm{o}}(u + ct - y)\,\mathrm{d}G(y) + (1 - G(u + ct))\Bigr)\,\mathrm{d}F^1(t)\,.$$

We have to multiply the above equation by e^{Ru}. We obtain by Lundberg's inequality

$$\begin{aligned}&\int_0^\infty\int_0^{u+ct} \psi^{\mathrm{o}}(u + ct - y)\mathrm{e}^{Ru}\,dG(y)\,\mathrm{d}F^1(t) \\ &\qquad = \int_0^\infty\int_0^{u+ct} \psi^{\mathrm{o}}(u + ct - y)\mathrm{e}^{R(u+ct-y)}\mathrm{e}^{Ry}\,\mathrm{d}G(y)\mathrm{e}^{-cRt}\,\mathrm{d}F^1(t) \\ &\qquad < \int_0^\infty\int_0^{u+ct} \mathrm{e}^{Ry}\,\mathrm{d}G(y)\mathrm{e}^{-cRt}\,\mathrm{d}F^1(t) \qquad (5.6) \\ &\qquad \le \int_0^\infty\int_0^\infty \mathrm{e}^{Ry}\,\mathrm{d}G(y)\mathrm{e}^{-cRt}\,\mathrm{d}F^1(t) = M_Y(R)M_{T_1}(-cR) < \infty\,.\end{aligned}$$

We can therefore interchange limit and integration

$$\begin{aligned}&\lim_{u\to\infty}\int_0^\infty\int_0^{u+ct} \psi^{\mathrm{o}}(u + ct - y)\mathrm{e}^{R(u+ct-y)}\mathrm{e}^{Ry}\,\mathrm{d}G(y)\mathrm{e}^{-cRt}\,\mathrm{d}F^1(t) \\ &\qquad = \int_0^\infty\int_0^\infty C\mathrm{e}^{Ry}\,\mathrm{d}G(y)\mathrm{e}^{-cRt}\,\mathrm{d}F^1(t) = M_Y(R)M_{T_1}(-cR)C\,.\end{aligned}$$

The remaining term is converging to 0 because

$$\begin{aligned}\int_0^\infty\int_{u+ct}^\infty \mathrm{e}^{Ru}\,\mathrm{d}G(y)\,\mathrm{d}F^1(t) &\le \int_0^\infty\int_{u+ct}^\infty \mathrm{e}^{R(y-ct)}\,\mathrm{d}G(y)\,\mathrm{d}F^1(t)\\ &\le \int_0^\infty\int_{u}^\infty \mathrm{e}^{Ry}\,\mathrm{d}G(y)\,\mathrm{e}^{-cRt}\,\mathrm{d}F^1(t)\\ &= M_{T_1}(-cR)\int_u^\infty \mathrm{e}^{Ry}\,\mathrm{d}G(y)\,.\end{aligned}$$

Hence

$$\lim_{u\to\infty}\psi(u)\mathrm{e}^{Ru} = M_Y(R)M_{T_1}(-cR)C \tag{5.7}$$

where C was the constant obtained in the ordinary case.

Example 5.8 (continued). In the general case we have to change the martingale because in general $\mathbb{E}[\mathrm{e}^{-RC_{T_1}}] \ne \mathrm{e}^{-Ru}$. Note that

$$\mathbb{E}[\mathrm{e}^{-RC_{T_1}}] = \mathrm{e}^{-Ru}M_Y(R)M_{T_1}(-cR) = \frac{\alpha}{\alpha-R}\mathrm{e}^{-Ru}M_{T_1}(-cR)\,.$$

Let

$$M_n = \begin{cases}\mathrm{e}^{-RC_{T_n}} & \text{if } n\ge 1,\\ \frac{\alpha}{\alpha-R}\mathrm{e}^{-Ru}M_{T_1}(-cR) & \text{if } n=0.\end{cases}$$

As in the ordinary case we obtain

$$\frac{\alpha}{\alpha-R}\mathrm{e}^{-Ru}M_{T_1}(-cR) = \frac{\alpha}{\alpha-R}\psi(u)$$

or equivalently

$$\psi(u) = M_{T_1}(-cR)\mathrm{e}^{-Ru}\,.$$

■

5.5. Diffusion Approximations

As in the classical case it is possible to show the following

Proposition 5.9. *Let $\{C_t^{(n)}\}$ be a sequence of renewal risk models with initial capital u, interarrival time distribution $F^{(n)}(t) = F(nt)$, claim size distribution $G^{(n)}(x) = G(x\sqrt{n})$ and premium rate*

$$c^{(n)} = \Bigl(1+\frac{\frac{c-\lambda\mu}{\lambda\mu}}{\sqrt{n}}\Bigr)\lambda^{(n)}\mu^{(n)} = c+(\sqrt{n}-1)\lambda\mu$$

where $\lambda = \lambda^{(1)}$ *and* $\mu = \mu^{(1)}$. *Then*

$$\{C_t^{(n)}\} \to \{u + W_t\}$$

uniformly on finite intervals where (W_t) *is a* $(c - \lambda\mu, \lambda\mu_2)$*-Brownian motion.*

Proof. See for instance [50, p.172]. □

Denote the ruin time of the renewal model by $\tau^{(n)}$ and the ruin time of the Brownian motion $\tau = \inf\{t \geq 0 : u + W_t < 0\}$. The diffusion approximation is based on the following

Proposition 5.10. *Let* $\{C_t^{(n)}\}$ *and* $\{W_t\}$ *as above. Then*

$$\lim_{n\to\infty} \mathbb{P}[\tau^{(n)} \leq t] = \mathbb{P}[\tau \leq t]\,.$$

Proof. See [87, Thm.9]. □

As in the classical case numerical investigations show that the approximation only works well if $c/(\lambda\mu)$ is close to 1.

5.6. Subexponential Claim Size Distributions

In the classical case we had assumed that the ladder height distribution is subexponential. If we assume that the ladder height distribution H is subexponential then the asymptotic behaviour would follow as in the classical case. Unfortunately we do hot have an explicit expression for the ladder height distribution in the renewal case. But the following proposition helps us out of this problem. We assume the ordinary case.

Proposition 5.11. *Let* U *be the distribution of* $u - C_{T_1} = Y_1 - cT_1$ *and let* $m = \int_0^\infty (1 - U(x))\,\mathrm{d}x$. *Let* $G_1(x) = \mu^{-1}\int_0^x (1 - G(y))\,\mathrm{d}y$ *and* $U_1(x) = m^{-1}\int_0^x (1 - U(y))\,\mathrm{d}y$. *Then*

i) *If* G_1 *is subexponential, then* U_1 *is subexponential and*

$$\lim_{x\to\infty} \frac{1 - U_1(x)}{1 - G_1(x)} = \frac{\mu}{m}\,.$$

ii) *If* U_1 *is subexponential then* $H/H(\infty)$ *is subexponential and*

$$\lim_{x\to\infty}\frac{1-U_1(x)}{H(\infty)-H(x)}=\frac{c-\lambda\mu}{\lambda m(1-H(\infty))}\,.$$

Proof.

i) Using Fubini's theorem

$$\begin{aligned}m(1-U_1(x))&=\int_x^\infty\int_0^\infty(1-G(y+ct))\,\mathrm{d}F(t)\,\mathrm{d}y\\&=\int_0^\infty\int_x^\infty(1-G(y+ct))\,\mathrm{d}y\,\mathrm{d}F(t)\\&=\mu\int_0^\infty(1-G_1(x+ct))\,\mathrm{d}F(t)\,.\end{aligned}$$

It follows by the bounded convergence theorem and Lemma F.2 that

$$\lim_{x\to\infty}\frac{1-U_1(x)}{1-G_1(x)}=\lim_{x\to\infty}\frac{\mu}{m}\int_0^\infty\frac{1-G_1(x+ct)}{1-G_1(x)}\,\mathrm{d}F(t)=\frac{\mu}{m}\,.$$

By Lemma F.8 U_1 is subexponential.

ii) Consider the random walk $\sum_{i=1}^n(Y_i-cT_i)$. We use the notation of Lemma E.2. By the Wiener-Hopf factorisation for $y>0$

$$1-U(y)=\int_{-\infty}^0(H(y-z)-H(y))\,\mathrm{d}\rho(z)\,.$$

Let $0<x<b$. Then using Fubini's theorem

$$\begin{aligned}\int_x^b(1-U(y))\,\mathrm{d}y&=\int_{-\infty}^0\int_x^b(H(y-z)-H(y))\,\mathrm{d}y\,\mathrm{d}\rho(z)\\&=\int_{-\infty}^0\Bigl(\int_x^{x-z}(H(\infty)-H(y))\,\mathrm{d}y-\int_b^{b-z}(H(\infty)-H(y))\,\mathrm{d}y\Bigr)\,\mathrm{d}\rho(z)\,.\end{aligned}$$

Because $\int_{-\infty}^0|z|\,\mathrm{d}\rho(z)$ is a finite upper bound (Lemma E.2) we can interchange the integral and the limit $b\to\infty$ and obtain

$$\int_x^\infty(1-U(y))\,\mathrm{d}y=\int_{-\infty}^0\int_x^{x-z}(H(\infty)-H(y))\,\mathrm{d}y\,\mathrm{d}\rho(z)\,.$$

We find

$$\int_{-\infty}^0|z|(H(\infty)-H(x-z))\,\mathrm{d}\rho(z)\le m(1-U_1(x))\le(H(\infty)-H(x))\int_{-\infty}^0|z|\,\mathrm{d}\rho(z)\,.$$

For $s > 0$ we obtain

$$(H(\infty)-H(x+s))\int_{-s}^{0}|z|\,\mathrm{d}\rho(z) \le \int_{-s}^{0}|z|(H(\infty)-H(x-z))\,\mathrm{d}\rho(z) \le m(1-U_1(x))$$

and therefore

$$1 \le \int_{-\infty}^{0}|z|\,\mathrm{d}\rho(z)\frac{H(\infty)-H(x+s)}{m(1-U_1(x+s))} \le \frac{\int_{-\infty}^{0}|z|\,\mathrm{d}\rho(z)}{\int_{-s}^{0}|z|\,\mathrm{d}\rho(z)}\,\frac{1-U_1(x)}{1-U_1(x+s)}\,.$$

Thus by Lemma F.2 $(H(\infty)-H(x))^{-1}(1-U_1(x)) \to m^{-1}\int_{-\infty}^{0}|z|\ \mathrm{d}\rho(z)$ and $H/H(\infty)$ is subexponential by Lemma F.8. The explicit value of the limit follows from Lemma E.2 noting that $\int_{-\infty}^{0}|z|\ \mathrm{d}\rho(z)$ is the expected value of the first descending ladder height of the random walk $\sum_{i=1}^{n}(Y_i - cT_i)$.

□

We can now proof the asymptotic behaviour of the ruin probability.

Theorem 5.12. *Let $\{C_t\}$ be a renewal risk process. Assume that the integrated tail of the claim size distribution*

$$G_1(x) = \mu^{-1}\int_0^x (1-G(y))\,\mathrm{d}y$$

is subexponential. Then

$$\lim_{u\to\infty}\frac{\psi(u)}{\int_u^\infty(1-G(y))\,\mathrm{d}y} = \frac{\lambda}{c-\lambda\mu}\,.$$

Remark. Note that this is exactly Theorem 4.15. The difference is only that $G_1(x)$ is not the ladder height distribution anymore. ■

Proof. Let us first consider an ordinary renewal model. By repeating the proof of Theorem 4.15 we find that

$$\lim_{u\to\infty}\frac{\psi(u)}{1-H(u)/H(\infty)} = \frac{H(\infty)}{1-H(\infty)}$$

because $H(x)/H(\infty)$ is subexponential by Proposition 5.11. Moreover, by Proposition 5.11

$$\begin{aligned}\lim_{u\to\infty}\frac{\psi(u)}{\int_u^\infty(1-G(y))\,\mathrm{d}y} &= \lim_{u\to\infty}\frac{\psi(u)}{1-H(u)/H(\infty)}\,\frac{1}{H(\infty)}\,\frac{H(\infty)-H(u)}{1-U_1(u)}\,\frac{1-U_1(u)}{\mu(1-G_1(u))}\\ &= \frac{H(\infty)}{1-H(\infty)}\,\frac{1}{H(\infty)}\,\frac{\lambda m(1-H(\infty))}{c-\lambda\mu}\,\frac{1}{m} = \frac{\lambda}{c-\lambda\mu}\,.\end{aligned}$$

For an arbitrary renewal risk model the assertion follows similarly as in Section 5.4.2. But two properties of subexponential distributions that are not proved here will be needed. □

5.7. Finite Time Lundberg Inequalities

Let $0 \le \underline{y} < \bar{y} < \infty$ and define $T = \inf\{T_n : n \in \mathbb{N}, T_n \ge \bar{y}u\}$. Using the stopping theorem

$$\begin{aligned}\mathrm{e}^{-ru} &= \mathbb{E}\left[\mathrm{e}^{-rC_{\tau\wedge T\wedge T_n}-\theta(r)(\tau\wedge T\wedge T_n)}\right] \\ &= \mathbb{E}\left[\mathrm{e}^{-rC_{\tau\wedge T}-\theta(r)(\tau\wedge T)}; T_n \ge \tau\wedge T\right] + \mathbb{E}\left[\mathrm{e}^{-rC_{T_n}-\theta(r)T_n}; T_n < \tau\wedge T\right] .\end{aligned}$$

By monotone convergence

$$\lim_{n\to\infty} \mathbb{E}\left[\mathrm{e}^{-rC_{\tau\wedge T}-\theta(r)(\tau\wedge T)}; T_n \ge \tau\wedge T\right] = \mathbb{E}\left[\mathrm{e}^{-rC_{\tau\wedge T}-\theta(r)(\tau\wedge T)}\right] .$$

For the second term we obtain

$$\lim_{n\to\infty} \mathbb{E}\left[\mathrm{e}^{-rC_{T_n}-\theta(r)T_n}; T_n < \tau\wedge T\right] \le \lim_{n\to\infty} \max\{\mathrm{e}^{-\theta(r)\bar{y}u}, 1\}\mathbb{P}[T_n < \tau\wedge T] = 0 .$$

Therefore

$$\begin{aligned}\mathrm{e}^{-ru} &= \mathbb{E}\left[\mathrm{e}^{-rC_{\tau\wedge T}-\theta(r)(\tau\wedge T)}\right] > \mathbb{E}\left[\mathrm{e}^{-rC_{\tau\wedge T}-\theta(r)(\tau\wedge T)}; \underline{y}u < \tau \le \bar{y}u\right] \\ &= \mathbb{E}\left[\mathrm{e}^{-rC_{\tau}-\theta(r)\tau}\,\middle|\, \underline{y}u < \tau \le \bar{y}u\right] \mathbb{P}[\underline{y}u < \tau \le \bar{y}u] \\ &> \mathbb{E}\left[\mathrm{e}^{-\theta(r)\tau}\,\middle|\, \underline{y}u < \tau \le \bar{y}u\right] \mathbb{P}[\underline{y}u < \tau \le \bar{y}u] \\ &> \mathrm{e}^{-\max\{\theta(r)\underline{y}u, \theta(r)\bar{y}u\}}\mathbb{P}[\underline{y}u < \tau \le \bar{y}u] .\end{aligned}$$

We obtain

$$\mathbb{P}[\underline{y}u < \tau \le \bar{y}u] < \mathrm{e}^{-(r-\max\{\theta(r)\underline{y}, \theta(r)\bar{y}\})u} = \mathrm{e}^{-\min\{r-\theta(r)\underline{y}, r-\theta(r)\bar{y}\}u}$$

which leads to the finite time Lundberg inequality

$$\mathbb{P}[\underline{y}u < \tau \le \bar{y}u] < \mathrm{e}^{-R(\underline{y},\bar{y})u} .$$

Here, as in the classical case,

$$R(\underline{y}, \bar{y}) := \sup\{\min\{r - \theta(r)\underline{y}, r - \theta(r)\bar{y}\} : r \in \mathbb{R}\} .$$

We have already discussed $R(\underline{y}, \bar{y})$ in the classical case (section 4.14). Assume that $r_\infty = \sup\{r \ge 0 : M_Y(r) < \infty\} > R$. Let

$$y_0 = (\theta'(R))^{-1} = \Bigl(\frac{M_Y'(R)M_T(-cR)}{M_Y(R)M_T'(-cR)} - c\Bigr)^{-1}$$

denote the critical value. Then we obtain as in the classical case

$$\mathbb{P}[0 < \tau \le yu \mid C_0 = u] < \mathrm{e}^{-R(0,y)u}$$

and $R(0, y) > R$ if $y < y_0$, and

$$\mathbb{P}[yu < \tau < \infty \mid C_0 = u] < \mathrm{e}^{-R(y,\infty)u}$$

and $R(y, \infty) > R$ if $y > y_0$. We can prove in the same way as Theorem 4.24 was proved.

Theorem 5.13. *Assume that* $R < r_\infty$. *Then*

$$\frac{\tau}{u} \longrightarrow y_0$$

in probability on the set $\{\tau < \infty\}$. □

Bibliographical Remarks

The renewal risk model was introduced by E. Sparre Andersen [3]. A good reference are also the papers by Thorin [81], [82] and [83]. Example 5.3 is due to Grandell [46, p.71]. Lundberg's inequality in the ordinary case was first proved by Andersen [3, p.224]. Example 5.6 can be found in [46, p.60]. The Cramér-Lundberg approximation was first proved by Thorin, [81, p.94] in the ordinary case and [82, p.97] in the stationary case. The results on subexponential claim sizes go back to Embrechts and Veraverbeke [38]. The results on finite time Lundberg inequalities can be found in [35]. More results on the renewal risk model is to be found in Chapter 6 of [66].

6. Some Aspects of Heavy Tails

6.1. Why Heavy Tailed Distributions Are Dangerous

In most of the considerations so far only the first or the first two moments of the claim size distributions were needed. And mean value and variance are easy to estimate via the estimators

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

and

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \hat{\mu})^2 \,.$$

These estimators are unbiased under any underlying claim size distribution. Nevertheless, we will see that using the estimator $\hat{\mu}$ is dangerous in the heavy claim case.

For the modelling of large claims the Pareto distribution is very popular. Let $\{Y_i : 1 \le i \le n\}$ be a sample of iid. Pa(α,1) distributed random variables where $\alpha > 1$. Note that the second parameter is only a scale parameter. Thus it is no loss of generality to choose it equal to 1. Assume that we got the knowledge that

$$Y_n^{\max} := \max_{i \le n} Y_i \le M$$

where M is a constant. What is the expected value of $\hat{\mu}$ conditioned on $\{Y_n^{\max} \le M\}$ and what is the relative error compared to the true mean $(\alpha-1)^{-1}$?

For the conditional mean value we get

$$\begin{aligned}
\mathbb{E}[\hat{\mu} \mid Y_n^{\max} \le M] &= \mathbb{E}[\hat{\mu} \mid Y_1 \le M, \dots, Y_n \le M] = \mathbb{E}[Y_1 \mid Y_1 \le M, \dots, Y_n \le M] \\
&= \mathbb{E}[Y_1 \mid Y_1 \le M] = \frac{\int_0^M x \frac{\alpha}{(1+x)^{\alpha+1}} \,\mathrm{d}x}{\int_0^M \frac{\alpha}{(1+x)^{\alpha+1}} \,\mathrm{d}x} \\
&= \frac{\int_0^M (x+1) \frac{\alpha}{(1+x)^{\alpha+1}} \,\mathrm{d}x}{\int_0^M \frac{\alpha}{(1+x)^{\alpha+1}} \,\mathrm{d}x} - 1 \\
&= \frac{\alpha}{\alpha-1}\,\frac{1-(1+M)^{-(\alpha-1)}}{1-(1+M)^{-\alpha}} - 1 \,.
\end{aligned}$$

The relative error is

$$\begin{aligned}
\frac{\mathbb{E}[\hat{\mu} \mid Y_n^{\max} \le M] - (\alpha-1)^{-1}}{(\alpha-1)^{-1}} &= \frac{\alpha(1-(1+M)^{-(\alpha-1)})}{1-(1+M)^{-\alpha}} - (\alpha-1) - 1 \\
&= -\alpha\frac{(1+M)^{-(\alpha-1)} - (1+M)^{-\alpha}}{1-(1+M)^{-\alpha}} \,.
\end{aligned}$$

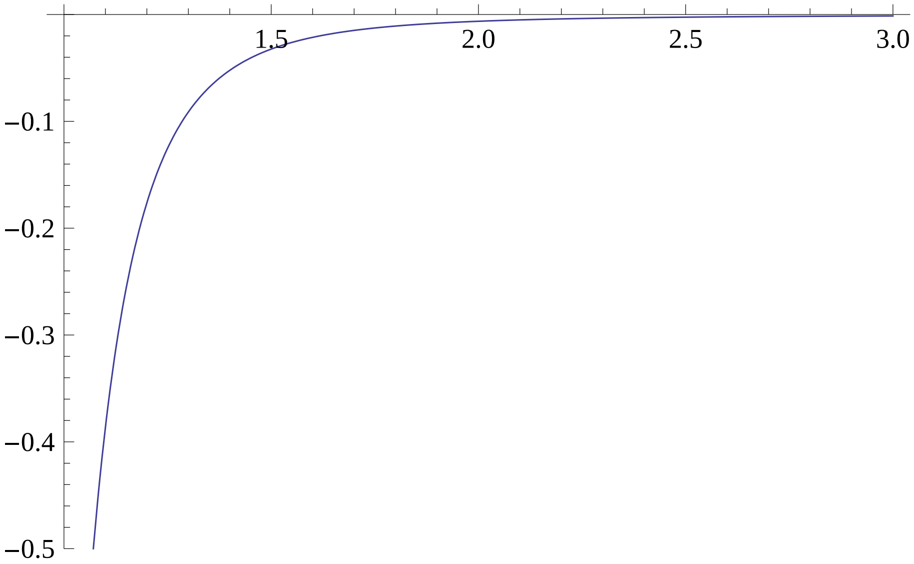


Figure 6.1: *Relative error for* $p = 0.99$ *and* $n = 1000$.

Let us fix $p \in (0, 1)$ and let us choose M such that $\mathbb{P}[Y_n^{\max} \le M] = p$, i.e. we fix the probability that it is no mistake to assume that $Y_n^{\max} \le M$.

$$\begin{aligned}\mathbb{P}[Y_n^{\max} \le M] &= \mathbb{P}[Y_1 \le M, \ldots, Y_n \le M] = \mathbb{P}[Y_1 \le M]^n \\ &= \bigl(1 - (1 + M)^{-\alpha}\bigr)^n = p\end{aligned}$$

or equivalently

$$(1 + M)^{-1} = \bigl(1 - p^{1/n}\bigr)^{1/\alpha} .$$

The relative error is therefore

$$-\alpha \frac{(1 - p^{1/n})^{(\alpha-1)/\alpha} - (1 - p^{1/n})}{p^{1/n}} . \tag{6.1}$$

Figure 6.1 shows the relative error made by choosing $p = 0.99$ and $n = 1000$. $n = 1000$ is a quite large sample size (a typical example is fire insurance). With probability $p = 0.99$ the maximum value is below M, i.e. on a set with probability 0.99 the expected relative error is larger than (6.1).

Quite often estimation yields values for α between 1 and 1.5. Some relative errors may be found in the table below:

α	1.05	1.1	1.15	1.2	1.25	1.3	1.4	1.5
error	-60.7%	-38.6%	-25.6%	-17.6%	-12.5%	-9.1%	-5.2%	-3.2%

If for instance the true parameter is 1.25 then the premium computed would be "at least 12.5% too low in 99% of all cases". This is quite a crucial error for any insurance company. We conclude that the usual estimation procedure for the mean value should not be used in the Pareto case. The same is true for other heavy tailed distribution functions.

Let us next consider a reinsurance problem. A company wants to reinsure Pa(α,1) distributed claims via an excess of loss reinsurance with retention level M. In order to fix the reinsurance premium the mean value of the part of a claim involving the reinsurer is needed.

$$\begin{aligned}
\mathbb{E}[Y_i - M \mid Y_i > M] &= \frac{\int_M^\infty x \frac{\alpha}{(1+x)^{\alpha+1}}\,\mathrm{d}x}{\int_M^\infty \frac{\alpha}{(1+x)^{\alpha+1}}\,\mathrm{d}x} - M = \frac{\int_M^\infty \frac{\alpha}{(1+x)^{\alpha}}\,\mathrm{d}x}{\int_M^\infty \frac{\alpha}{(1+x)^{\alpha+1}}\,\mathrm{d}x} - (M+1) \\
&= \frac{\frac{\alpha}{\alpha-1}(1+M)^{-(\alpha-1)}}{(1+M)^{-\alpha}} - (M+1) = \frac{\alpha(M+1)}{\alpha-1} - (M+1) \\
&= \frac{M+1}{\alpha-1}\,.
\end{aligned}$$

It follows that $\mathbb{E}[Y_i - M \mid Y_i > M]$ increases linearly with M. Thus in order to estimate $\mathbb{E}[Y_i - M \mid Y_i > M]$ only the large values $Y_i > M$ should be used for the estimation procedure. But this event might be rare. In contrast consider the Exp(1) case. Here $\mathbb{E}[Y_i - M \mid Y_i > M] = 1$ for all M and thus the mean value needed can be inferred from the full sample.

An argument often heard is the following: “The aggregate value of all goods on earth is very large, but finite. Thus a claim must be bounded which means light tailed. Therefore there is no need to consider the heavy tailed case.” But this contradicts the actuarial experience. Quite often claim amounts show a behaviour typical to heavy tails. For instance considering damages due to hurricanes shows that the aggregate sum of claims is determined by a few of the largest claim. Thus an insurance company has to be prepared that the next very large claim is much larger. Moreover the aggregate value of all goods is so large that it can be considered to be infinite.

We conclude that one has to be careful as soon as heavy tailed claims are involved. Actuaries use the phrase **heavy tailed claims are dangerous**. Thus one has to treat the heavy tailed claim case specially.

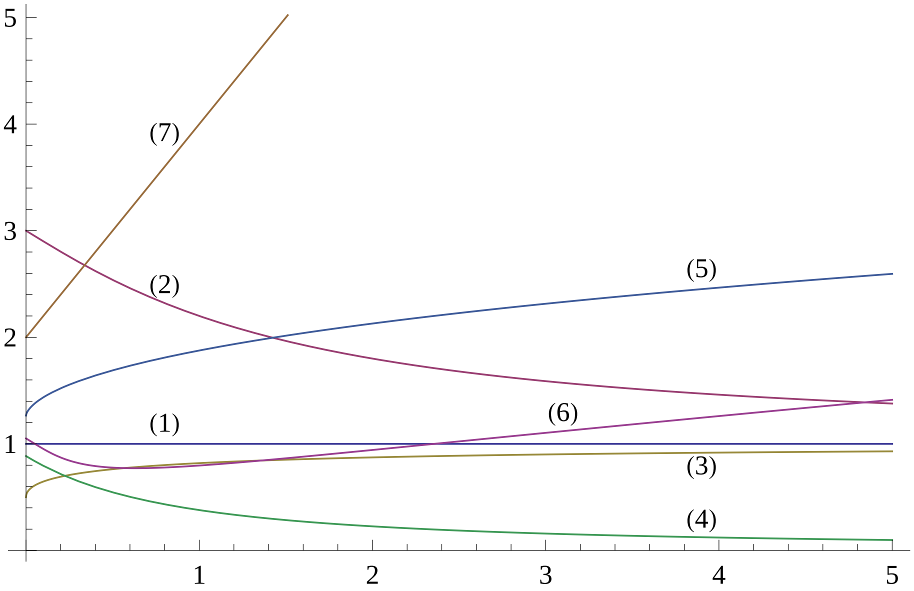


Figure 6.2: *Mean residual life functions:* (1) Exp*(1)*, (2) Γ*(3,1)*, (3) Γ*(0.5,1)*, (4) Wei*(2,1)*, (5) Wei*(0.7,1)*, (6) LN*(-0.2,1)*, (7) Pa*(1.5,1)*.

6.2. How to Find Out Whether a Distribution Is Heavy Tailed or Not

6.2.1. The Mean Residual Life

We considered before the quantity $\mathbb{E}[Y_i - M \mid Y_i > M]$, which is called **mean residual life** in survival analysis. Let us consider some other distributions and the limits for $M \to \infty$. We first note that

$$\mathbb{E}[Y_i - M \mid Y_i > M] = \frac{\int_M^\infty \int_M^y \, \mathrm{d}z \, \mathrm{d}F(y)}{1 - F(M)} = \frac{\int_M^\infty (1 - F(z)) \, \mathrm{d}z}{1 - F(M)} \,.$$

i) Exp(α)

$$\mathbb{E}[Y_i - M \mid Y_i > M] = \alpha^{-1} \to \alpha^{-1} \,.$$

ii) $\Gamma(\gamma, \alpha)$

$$\mathbb{E}[Y_i - M \mid Y_i > M] = \frac{\int_M^\infty \int_z^\infty y^{\gamma-1} \mathrm{e}^{-\alpha y} \, \mathrm{d}y \, \mathrm{d}z}{\int_M^\infty y^{\gamma-1} \mathrm{e}^{-\alpha y} \, \mathrm{d}y} \to \alpha^{-1} \,.$$

iii) Wei(α, c)

$$\mathbb{E}[Y_i - M \mid Y_i > M] = \frac{\int_M^\infty \mathrm{e}^{-cz^\alpha} \, \mathrm{d}z}{\mathrm{e}^{-cM^\alpha}} \to \begin{cases} 0, & \text{if } \alpha > 1, \\ c^{-1}, & \text{if } \alpha = 1, \\ \infty, & \text{if } \alpha < 1. \end{cases}$$

iv) $\mathrm{LN}(\mu, \sigma^2)$

$$\mathbb{E}[Y_i - M \mid Y_i > M] = \frac{\int_M^\infty \Phi\left(\frac{\mu - \log z}{\sigma}\right)\,\mathrm{d}z}{\Phi\left(\frac{\mu - \log M}{\sigma}\right)} \to \infty\,.$$

v) $\mathrm{Pa}(\alpha, \beta)$

$$\mathbb{E}[Y_i - M \mid Y_i > M] = \frac{M + \beta}{\alpha - 1} \to \infty\,.$$

Figure 6.2 shows the mean residual life functions of several distribution functions. From the computations above we can see that the mean residual life tends to infinity as $M \to \infty$ in the heavy tailed cases, and to a finite value in the small claim cases. One can also show that in the case of a subexponential distribution the mean residual life function tends to infinity as $M \to \infty$.

The idea of the approach is to estimate the mean residual life function. If the function is 'increasing to infinity', then the distribution function should be considered to be heavy tailed. If it is 'converging' to a finite constant then we can model it to be light tailed.

6.2.2. How to Estimate the Mean Residual Life Function

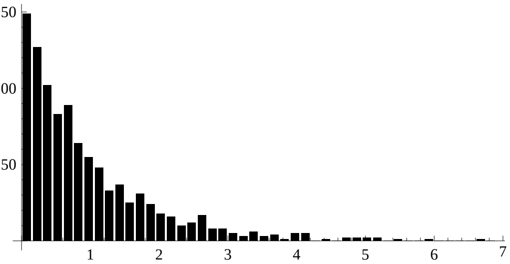
Let $\{Y_i : 1 \leq i \leq n\}$ be a sample of iid. random variables. The natural estimator of $\mathbb{E}[Y_i - M \mid Y_i > M]$ is

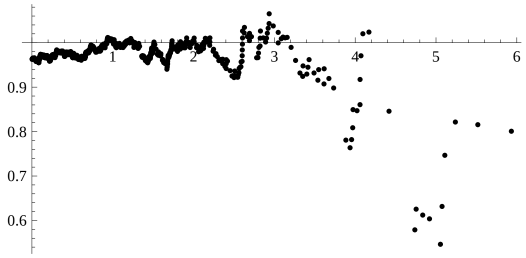
$$\hat{e}(M) = \frac{1}{\sum_{i=1}^n \mathbb{1}_{\{Y_i > M\}}} \sum_{i=1}^n \mathbb{1}_{\{Y_i > M\}}(Y_i - M)\,.$$

$\hat{e}(M)$ is called **empirical mean residual life function**. We have seen before that it is dangerous to estimate the mean value in this way, especially if the claim sizes are heavy tailed. But fortunately, we are not really interested in the real mean residual life function. We are only interested in the shape of its graph. And for this purpose the estimator $\hat{e}(M)$ is good enough.

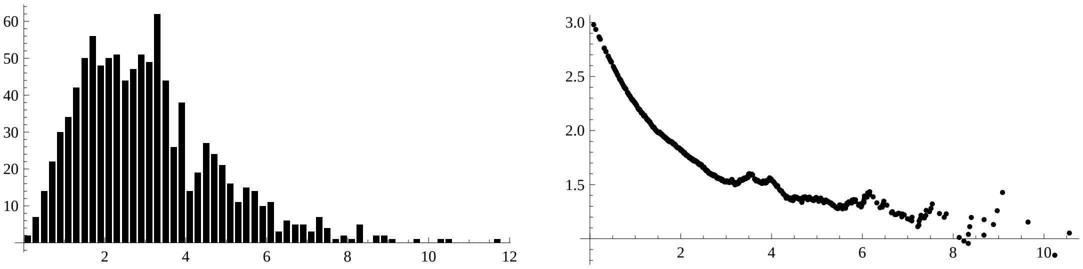
The following pictures show histograms and empirical residual life functions generated by a simulated sample of size 1000 each.

- Exp(1)

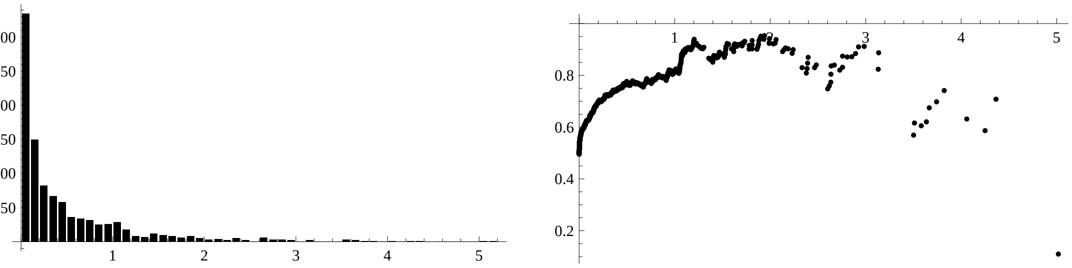




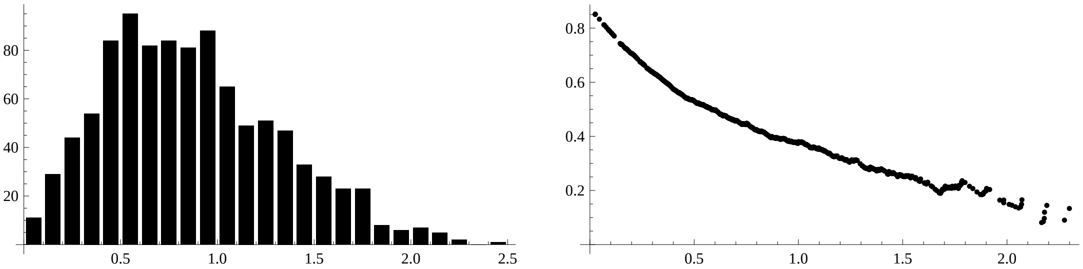
- $\Gamma(3,1)$



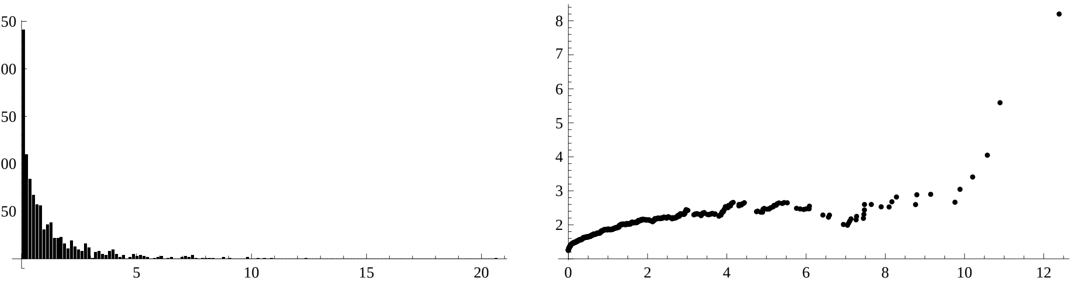
- $\Gamma(0.5,1)$



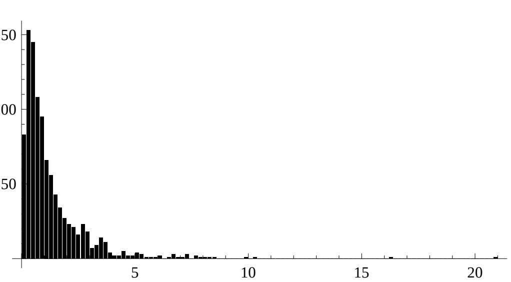
- Wei(2,1)

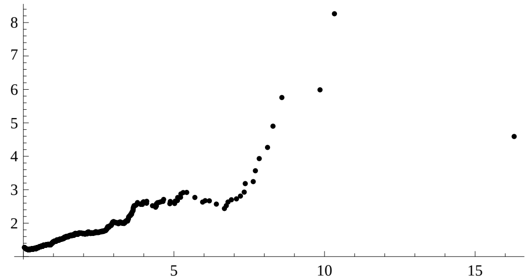


- Wei(0.7,1)

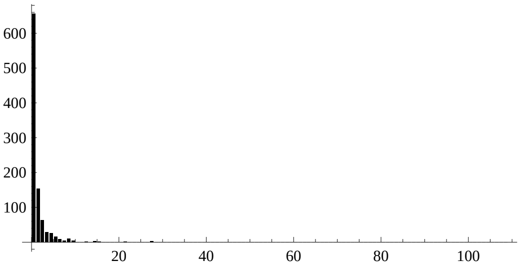


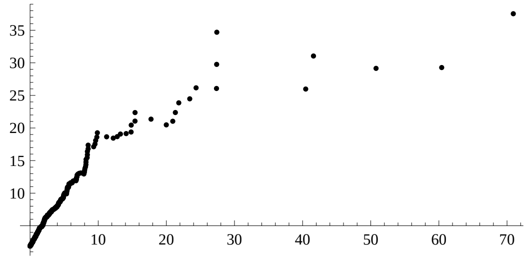
- LN(-0.2,1)





- Pa(1.5,1)

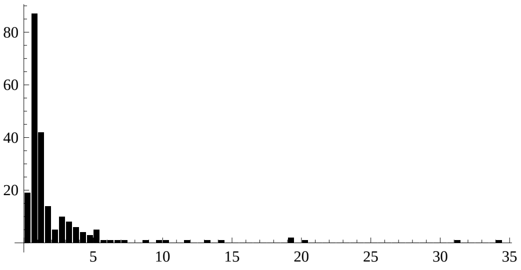


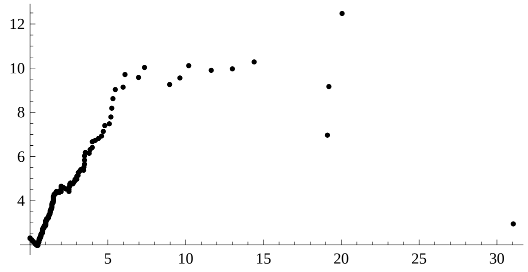


One can see clearly that the exponential, the two gamma and the first Weibull distributions are light tailed. The lognormal and the Pareto distributions are clearly heavy tailed (note that the last few points are not significant because only few data were used for its calculation). The second Weibull distribution yields problems. Though the empirical mean residual life function is increasing, it looks like it would converge. On the other hand there are points far out. The largest value is 20.6, the largest but one is 12.4. This behaviour is typical for heavy tailed data. This example shows that it is important also to consider a plot of the data in order to find the right model.

To end this section we consider real data from Swedish and Danish fire insurance.

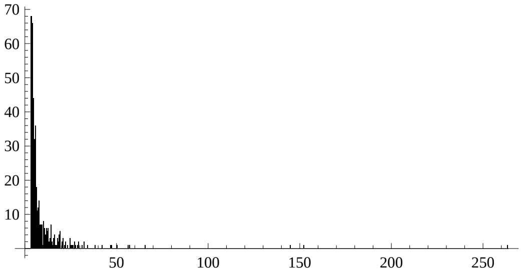
Example 6.1. (Swedish fire insurance) The sample consists of 218 claims for 1982 in units of Millions of Swedish kroner.

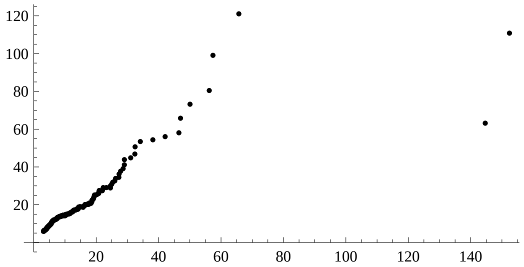




The histogram (some claims are far out) and the mean residual life function (which is clearly increasing) indicate that the distribution function is heavy tailed. The empirical mean residual life function is first decreasing and after that linearly increasing. This indicates that the larger claims follow a Pareto or a lognormal distribution. ■

Example 6.2. (Danish fire insurance) The sample consists of 500 large claims from 1st January 1980 till 31st December 1990. The unit is millions of Danish Kroner (1985 prices). The data are also analysed in [68] and [36].





As in Example 6.1 the histogram and the mean residual life function indicate that the distribution function is heavy tailed. Most likely the larger claims follow a Pareto or a lognormal distribution. ■

6.3. Parameter Estimation

If a model for the data is chosen then it remains to estimate the parameters.

6.3.1. The Exponential Distribution

For the exponential distribution both the maximum likelihood estimator and the method of moments yield the estimator

$$\hat{\alpha} = \Bigl(\frac{1}{n}\sum_{i=1}^n Y_i\Bigr)^{-1} .$$

6.3.2. The Gamma Distribution

The log likelihood function is

$$\sum_{i=1}^n (\gamma \log \alpha - \log \Gamma(\gamma) + (\gamma - 1)\log Y_i - \alpha Y_i) .$$

This yields the equations

$$\gamma = \frac{\alpha}{n} \sum_{i=1}^{n} Y_i$$

and

$$\log \alpha - \frac{\Gamma'(\gamma)}{\Gamma(\gamma)} + \frac{1}{n} \sum_{i=1}^{n} \log Y_i = 0\,.$$

These equations can only be solved numerically.

Because Gamma distributions are light tailed the method of moments can be used. This yields the estimators

$$\hat{\alpha} = \frac{\bar{Y}}{\frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2}$$

and

$$\hat{\gamma} = \frac{\bar{Y}^2}{\frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2}$$

where $\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$. Because of the simplicity the second estimator should be prefered.

6.3.3. The Weibull Distribution

The log likelihood function is

$$\sum_{i=1}^{n} (\log \alpha + \log c + (\alpha - 1) \log Y_i - c Y_i^{\alpha})\,.$$

We can assume that $\alpha \neq 1$. The we obtain the equations

$$c \frac{1}{n} \sum_{i=1}^{n} Y_i^{\alpha} = 1$$

and

$$\frac{\alpha}{n} \sum_{i=1}^{n} (c Y_i^{\alpha} - 1) \log Y_i = 1\,.$$

Also these equations have to be solved numerically.

The equations obtained from the method of moments are even harder to solve. Anyway, taking into account that the distribution is heavy tailed for $\alpha < 1$, the method of moments should not be used, except if it is clear that $\alpha > 1$.

6.3.4. The Lognormal Distribution

Here obviously the best unbiased estimator is (compare with the normal distribution)

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} \log Y_i$$

and

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\log Y_i - \hat{\mu})^2 \,.$$

6.3.5. The Pareto Distribution

Solving the maximum likelihood equations yields the two equations

$$\frac{\alpha}{n}\sum_{i=1}^{n} \log(1 + Y_i/\beta) = 1$$

and

$$1 = \frac{\alpha+1}{n}\sum_{i=1}^{n}(1+\beta/Y_i)^{-1} \,.$$

Also these equations have to be solved numerically.

We shall try to find another approach. Assume for the moment that β is known. Then the maximum likelihood estimator for α

$$\hat{\alpha} = \Bigl(\frac{1}{n}\sum_{i=1}^{n} \log(1 + Y_i/\beta)\Bigr)^{-1}$$

looks similar to the estimator in the exponential case. In fact,

$$\mathbb{P}[\log(1 + Y_i/\beta) > x] = \mathbb{P}[Y_i > \beta(\mathrm{e}^x - 1)] = \mathrm{e}^{-\alpha x}$$

is exponentially distributed. Because, for Y_i large, $\log(1 + Y_i/\beta) \approx \log(Y_i/\beta) = \log Y_i - \log \beta$, we find that intuitively $\log Y_i$ is approximately $\mathrm{Exp}(\alpha)$ distributed for large values of Y_i. Explicitly

$$\mathbb{P}[\log Y_i > x] = (1 + \mathrm{e}^x/\beta)^{-\alpha} = (\beta\mathrm{e}^{-x} + 1)^{-\alpha}\beta^\alpha\mathrm{e}^{-\alpha x} \sim \beta^\alpha\mathrm{e}^{-\alpha x}$$

and

$$\mathbb{P}[\log Y_i > M + x \mid \log Y_i > M] = \Bigl(\frac{\beta\mathrm{e}^{-M-x} + 1}{\beta\mathrm{e}^{-M} + 1}\Bigr)^{-\alpha}\mathrm{e}^{-\alpha x} \,.$$

If we choose M large enough then

$$\hat{\alpha} = \Bigl(\frac{1}{\sum_{i=1}^n \mathbb{1}_{\{\log Y_i > M\}}} \sum_{i=1}^n \mathbb{1}_{\{\log Y_i > M\}} (\log Y_i - M)\Bigr)^{-1}$$

is an estimator for α.

There remains one problem: How shall we choose M?

- The larger M the closer is the distribution of $\log Y_i - M$ to an exponential distribution.
- The larger M the less data are used in the estimator.

It is clear that M can be chosen to be large if n is larger. Thus M must depend on n. Moreover the optimal M will depend on the parameters α and β, which are not known. Thus M must depend on the sample as well. This can be achieved if we estimate α only using the largest $k(n) + 1$ data. Let $\{Y_{i:n} : i \le n\}$ denote the order statistics $Y_{1:n} \le Y_{2:n} \le \cdots \le Y_{n:n}$ of $\{Y_i : i \le n\}$. We define the estimator

$$\hat{\alpha} = \Bigl(\frac{1}{k(n)} \sum_{i=1}^{k(n)} (\log Y_{n+1-i:n} - \log Y_{n-k(n):n})\Bigr)^{-1} . \tag{6.2}$$

Intuitively, this estimator will converge to α as $n \to \infty$ provided $k(n) \to \infty$ and $Y_{k(n):n} \to \infty$ a.s.. It can be shown that whenever $k(n)/n \to 0$ and $k(n)/\log\log n \to \infty$ then $\hat{\alpha} \to \alpha$ a.s..

It remains to estimate β. Note that for $n \to \infty$ and $\gamma \in (0, 1)$

$$1 - (1 + Y_{\lfloor \gamma n \rfloor :n}/\beta)^{-\alpha} \to \gamma \text{ a.s.},$$

which we will see later (Section 6.4.1). Thus for n large we find the estimator

$$\hat{\beta} = \frac{Y_{\lfloor (1-\gamma) n \rfloor :n}}{\gamma^{-1/\hat{\alpha}} - 1} .$$

Because quite often the distribution will not be explicitly Pareto, only the tail will behave like the tail of a Pareto distribution, we wish to let $\gamma \to 0$ as $n \to \infty$. Choose therefore $\gamma n = k(n)$, i.e. our estimator is

$$\hat{\beta} = \frac{Y_{n-k(n):n}}{(n/k(n))^{1/\hat{\alpha}} - 1} .$$

In order to simplify the estimator note that asymptotically this is the same as

$$\hat{\beta} = \Bigl(\frac{k(n)}{n}\Bigr)^{1/\hat{\alpha}} Y_{n-k(n):n} . \tag{6.3}$$

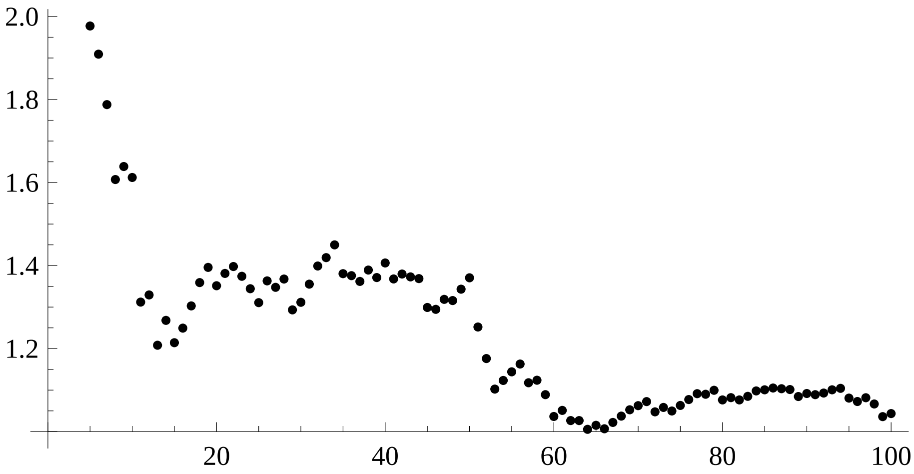


Figure 6.3: *Hill's estimator as function of k for the Swedish fire insurance data.*

One can show that $\hat{\beta} \to \beta$ a.s. as $n \to \infty$.

The last question of interest is how to choose $k(n)$ optimally. It can be shown that in many situations the following holds. If, for any $\varepsilon > 0$, $k(n)n^{-2/3+2\varepsilon}$ converges to 0 then the rate of convergence of $\hat{\alpha}$ to α is $n^{-1/3+\varepsilon}$. Thus, in practice, one would take ε as small as possible and thus use $k(n) = \lfloor n^{2/3} \rfloor$. Another possibility is to plot the estimator for $k = 1, 2, 3 \ldots$. Then choose the first point, where the function $\hat{\alpha}(k)$ stabilises (for a short time). More about this and also other estimators can be found in [36].

6.3.6. Parameter Estimation for the Fire Insurance Data

We now fit a Pareto and a lognormal model to the Swedish and Danish fire insurance data considered before. We already have seen that a heavy tailed distribution function should be used in order to model the fire insurance data. The explicit figures for the claims have not been given here, so the reader cannot do the calculations himself.

Example 6.1 (continued). In the Pareto case we choose $k(218) = \lfloor 218^{2/3} \rfloor = 36$. Figure 6.3 indicates that k indeed should be chosen close to 36. We find the estimates

$$\hat{\alpha} = 1.37567, \qquad \hat{\beta} = 0.879807.$$

Note that we find the mean value $\hat{\beta}/(\hat{\alpha} - 1) = 2.34199$, whereas the empirical mean value is 2.28172, only slightly smaller. The big difference lies in the estimate for

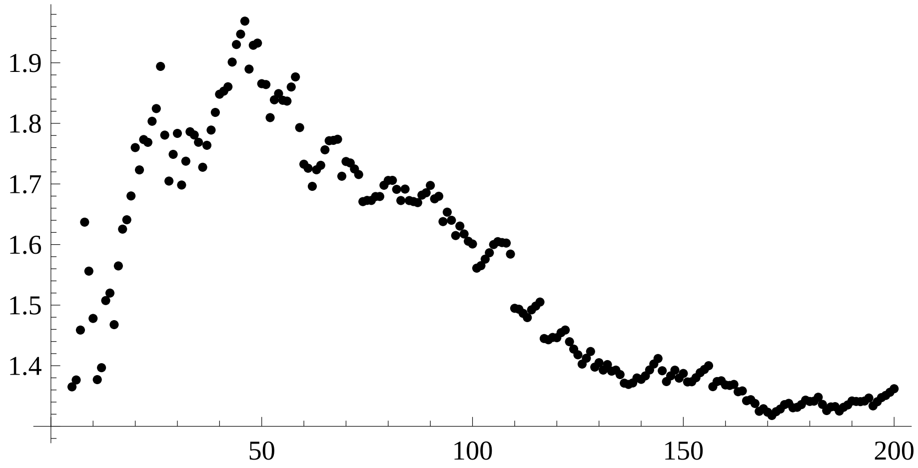


Figure 6.4: *Hill's estimator as function of k for the Danish fire insurance data.*

the variance. The empirical variance is 14.3158, whereas in the Pareto model the variance is infinite.

Fitting a lognormal model we find the parameters

$$\hat{\mu} = 0.218913, \qquad \hat{\sigma}^2 = 0.945877.$$

In the lognormal model we obtain for the mean value 1.99741 which is even smaller than the empirical mean value. The variance of the lognormal model is 6.28397, which again is smaller than the empirical variance. This shows that choosing a Pareto model the insurance company is on the safer side. Therefore the Pareto model is very popular among practitioners. The comparison with the empirical mean value and the empirical variance shows that we should not choose this model.

The problem we have with the lognormal model is that there are too many small claims, i.e. a real lognormal distribution would have more claims close to $e^{\hat{\mu}}$. These small claims are responsible that $\hat{\mu}$ is relative small. The Pareto distribution has a lot of very small claims, but for the estimation they are not considered. That is another advantage of the Pareto model. For the estimation only the tail is considered. And the tail is the important part of the distribution function in practice. ■

Example 6.2 (continued). In the Pareto case we choose $k(500) = \lfloor 500^{2/3} \rfloor = 62$. Figure 6.4 shows that the Hill estimator as a function of k is very unstable. For a short range it stabilizes around 70. Thus 62 seems not to be a too bad choice. It starts really to stabilize at about $k = 170$. This seems to be too large for k. Anyway, we calculate the parameters for both $k = 62$ and $k = 180$.

For $k = 62$ the estimates are

$$\hat{\alpha} = 1.69605, \qquad \hat{\beta} = 4.20413.$$

The mean value is 6.03997, whereas the empirical mean was 9.08176. This indicates that $k = 62$ is not a good choice. However, the Pareto distribution will have a lot of very small claims, which here is not the case. We should here only model the claims over a certain threshold to be Pareto distributed. For example, if we only take claims over 10 into account, i.e. the largest 109 claims the empirical mean will be 24.0818, whereas the chosen parametrization will yields a conditional mean of 30.4067.

For $k = 180$ the estimates are

$$\hat{\alpha} = 1.34103, \qquad \hat{\beta} = 2.87879.$$

The mean value is 8.4415, which still is smaller than the empirical mean value. Taking only values over 10 into account the conditional mean will be 47.7646, which is much larger than the empirical conditional mean 24.0818.

Fitting a lognormal distribution the parameters are

$$\hat{\mu} = 1.84616, \qquad \hat{\sigma}^2 = 0.461025.$$

In the lognormal model we obtain for the mean value 7.97787 which is also much smaller than the empirical mean value. The variance of the lognormal model is 100.924, which again is much smaller than the empirical variance 270.933. This shows that choosing a Pareto model the insurance company is on the safer side. ■

6.4. Verification of the Chosen Model

6.4.1. Q-Q-Plots

Let $\{Y_i : 1 \le i \le n\}$ be iid. with distribution function G, and for simplicity we assume that G is continuous. Consider the random variables $\{G(Y_i)\}$. It is clear that the $\{G(Y_i)\}$'s are iid.. We want to find the distribution function of $G(Y_i)$.

Let $G^{-1} : (0,1) \to \mathbb{R}, x \mapsto \inf\{y \in \mathbb{R} : G(y) > x\}$ denote the **generalised inverse function** of G. Let $x \in (0,1)$. Because G is increasing and continuous we find that $G(G^{-1}(x)) = x$ and $G(z) \le x \iff z \le G^{-1}(x)$. Therefore

$$\mathbb{P}[G(Y_i) \le x] = \mathbb{P}[Y_i \le G^{-1}(x)] = G(G^{-1}(x)) = x\,.$$

It follows that $G(Y_i)$ is uniformly distributed on $(0,1)$.

Let $X_i = G(Y_i)$. Considering the order statistics it is easy to see that $X_{i:n} = G(Y_{i:n})$. The distribution function of $X_{k:n}$ is

$$\mathbb{P}[X_{k:n} > x] = \sum_{i=0}^{k-1} \mathbb{P}[X_{i:n} \le x, X_{i+1:n} > x] = \sum_{i=0}^{k-1} \binom{n}{i} x^i (1-x)^{n-i} \,.$$

From this we can compute the first two moments of $X_{k:n}$.

$$\begin{aligned}\mathbb{E}[X_{k:n}] &= \int_0^1 \sum_{i=0}^{k-1} \binom{n}{i} x^i (1-x)^{n-i} \,\mathrm{d}x = \sum_{i=0}^{k-1} \binom{n}{i} \int_0^1 x^i (1-x)^{n-i} \,\mathrm{d}x \\ &= \sum_{i=0}^{k-1} \binom{n}{i} \frac{\Gamma(i+1)\Gamma(n-i+1)}{\Gamma((i+1)+(n-i+1))} = \sum_{i=0}^{k-1} \frac{1}{n+1} = \frac{k}{n+1} \,.\end{aligned}$$

$$\begin{aligned}\mathbb{E}[X_{k:n}^2] &= \int_0^1 2x \sum_{i=0}^{k-1} \binom{n}{i} x^i (1-x)^{n-i} \,\mathrm{d}x = 2\sum_{i=0}^{k-1} \binom{n}{i} \int_0^1 x^{i+1} (1-x)^{n-i} \,\mathrm{d}x \,, \\ &= 2\sum_{i=0}^{k-1} \binom{n}{i} \frac{\Gamma(i+2)\Gamma(n-i+1)}{\Gamma(n+3)} = \frac{2}{(n+1)(n+2)} \sum_{i=0}^{k-1} (i+1) \\ &= \frac{k(k+1)}{(n+1)(n+2)} \,.\end{aligned}$$

Thus the variance of $X_{k:n}$ is

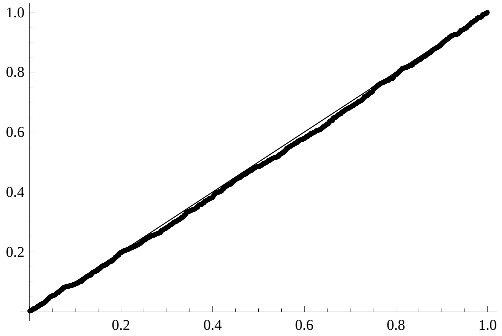
$$\operatorname{Var}[X_{k:n}] = \frac{k(n+1-k)}{(n+1)^2(n+2)} \le \frac{1}{4(n+2)} \,.$$

The variance is therefore converging to 0 as $n \to \infty$, i.e. $X_{k:n}$ lies close to $k/(n+1)$. The result can be made more precise. Because $[0,1]$ is compact and the pointwise limit distribution is continuous, we have that $\sup\{|X_{k:n} - k/(n+1)| : k \le n\}$ converges to zero.

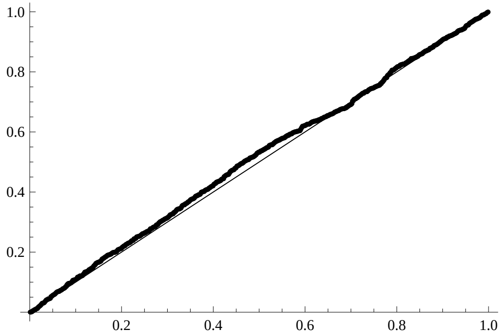
Consider now the graph generated by the points $(k/(n+1), G(Y_{k:n}))$. The points are expected to lie close to the line $y = x$. This graph is called $Q - Q - plot$ (quantile-quantile-plot).

We now consider the Q-Q-plots of the simulated data verse the correct distributions.

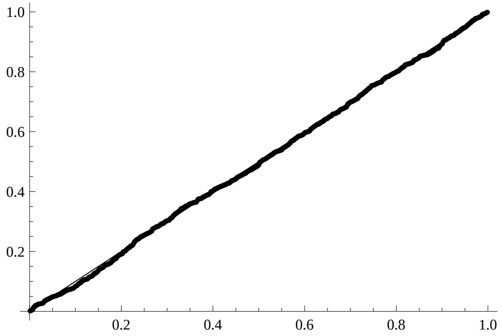
- Exp(1)



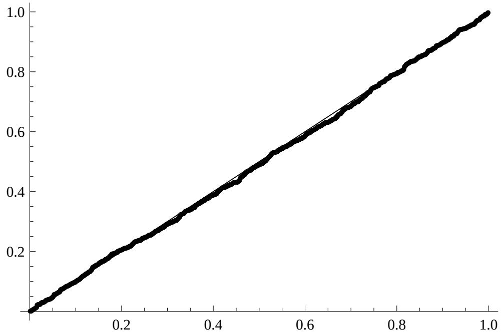
- $\Gamma(3,1)$



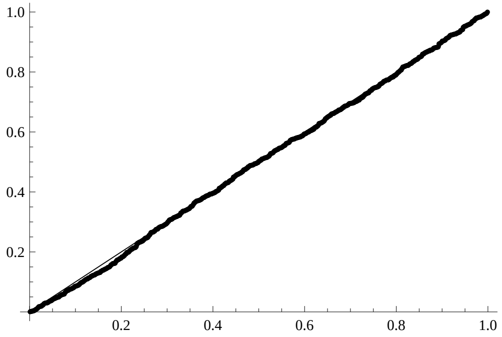
- $\Gamma(0.5,1)$



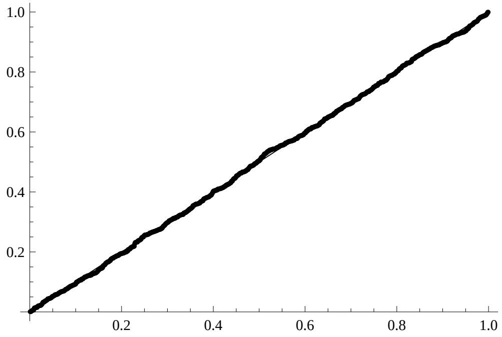
- Wei(2,1)



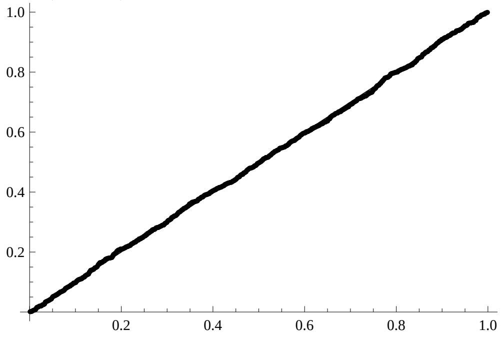
- Wei(0.7,1)



- LN(-0.2,1)

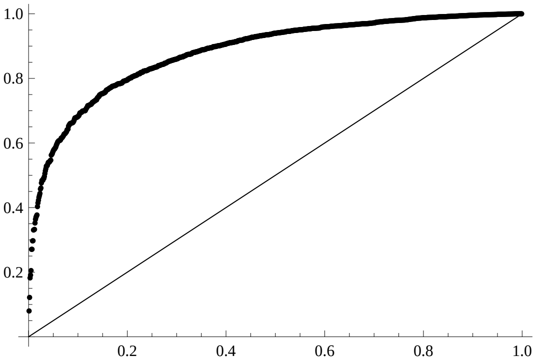


- Pa(1.5,1)

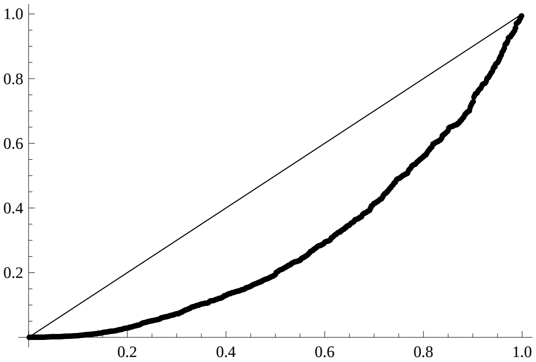


We see that the graphs are very close to the line $y = x$. Thus the random number generator yields satisfactory results.

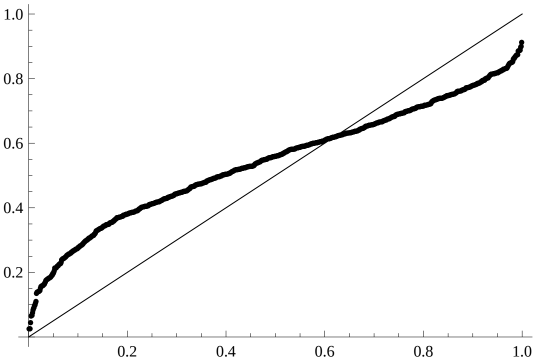
If we did not choose the right distribution then the following may happen.



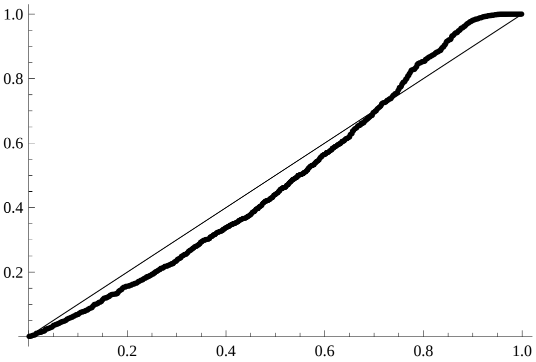
$G(Y_{k:n})$ is too large. The correct function $G(x)$ should be smaller than the chosen one, i.e. the chosen distribution has too much weight close to 0. The correct distribution is therefore 'less skewed' than the chosen one.



$G(Y_{k:n})$ is too small. The correct function $G(x)$ should be larger than the chosen one, i.e. the chosen distribution has not enough weight close to 0. The correct distribution is therefore 'more skewed' than the chosen one.



$G(Y_{k:n})$ is too large for small k and too small for large k. The correct distribution should have less small values and less very large values, i.e. the correct distribution should have a lighter tail than the chosen distribution.



$G(Y_{k:n})$ is too small for small k and too large for large k. The correct distribution should have more small values and more very large values, i.e. the correct distribution should have a thicker tail than the chosen distribution.

Remarks.

i) Note that the Q-Q-plot is sensitive to the parameter estimation. A bad Q-Q-plot must not mean that the model is wrong, it can also mean that the parameters have been estimated badly.

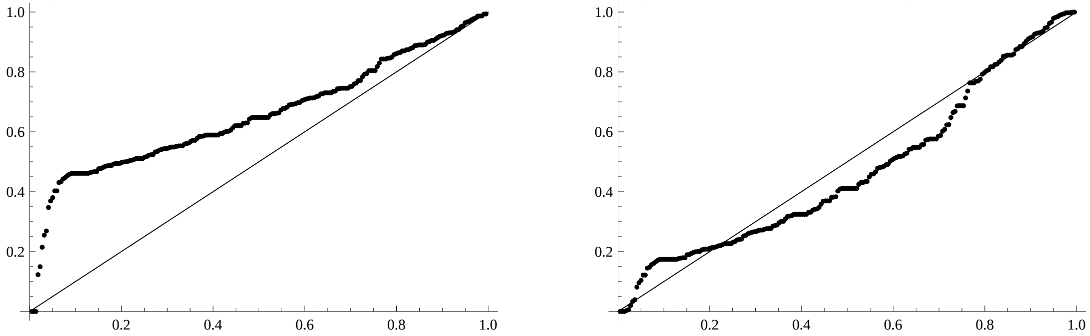
ii) In order to make the right decision the sample size has to be taken into account.

iii) Note that the picture only depends on the sample and not on the whole distribution. If for instance a class of distribution functions is chosen with a too light tail, then the large values of the sample will force the estimated parameters to be from a skewed model. The Q-Q-plot will then suggest that the tail should be lighter than the chosen one even though it is vice versa. ■

6.4.2. The Fire Insurance Example

Let us now consider the fire insurance example again. We had fitted a Pareto and a lognormal model to the data.

Example 6.1 (continued). The Q-Q-plots in this example are



- Pa(1.37567, 0.879807)
- LN(0.218913, 0.945877)

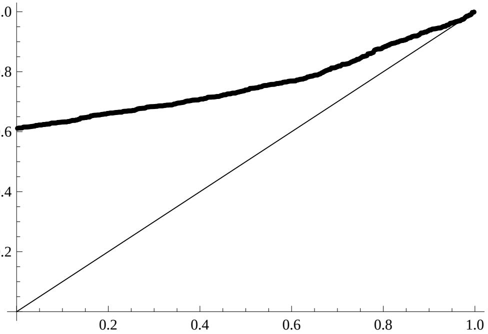
All investigations so far had indicated that the underlying distribution is Pareto. But now the Q-Q-plot suggests that the lognormal distribution is closer to the true distribution. The Pareto distribution is too much skewed. The reason might be that the α-parameter is too far away from the true value. On the other hand the lognormal distribution seems to fit quite well in the tail. It is not too far away from the right distribution for smaller values. We can of course not expect that one of the considered distributions fits perfectly. But we try to find a distribution with a tail similar to the correct distribution function. The Q-Q-plot suggest that the lognormal distribution has the right tail.

We conclude that it is important to consider a larger set of distribution functions, even if some evidence shows that another distribution function should be prefered. As in our case it might turn out that not the preferred distribution function is the correct one. One of the problems here would be that 218 is not a very large sample size. In practice one would be able to collect data for several years, which might help to get a clearer picture of the distribution. ■

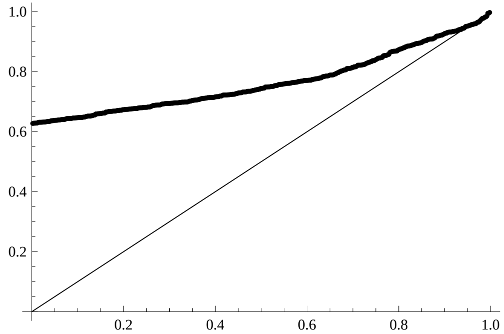
Example 6.2 (continued).

The Q-Q-plots in this example are

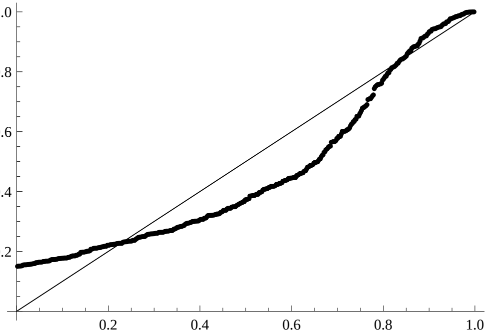
- Pa(1.69605, 4.20413)



- Pa(1.34103, 2.87879)



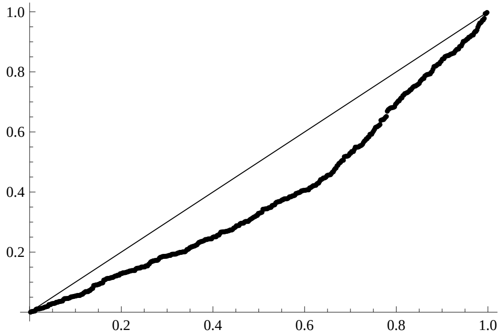
- LN(1.84616, 0.461025)



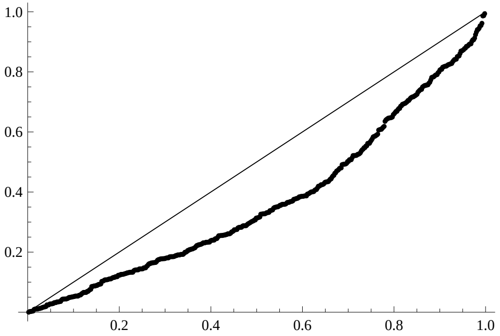
We clearly see that there are not enough very small data, but far out in the tails of the Pareto distributions seem to be right. The lognormal distribution clearly is not the right distribution.

Because the absence of the very small claims cause problems we should have considered a conditional Pareto distribution instead. Condition on $Y_i > Y_{1:n}$. Then the Q-Q-plots are

- Pa(1.69605, 4.20413)



- Pa(1.34103, 2.87879)



The chosen parametrization both do not have enough weight close to 0 and near infinity. This indicates that the correct parameter α should be chosen smaller than the estimated ones or the correct parameter β is larger than the estimated one. This was also indicated by the fact that the mean value of the estimated distributions were smaller than the empirical mean value. ■

Bibliographical Remarks

The graphical tool for recognizing heavy tailed distributions, the mean residual life function, can be found for instance in [54] or [37]. The Swedish fire insurance data are taken from [57]. The data can also be found in [37]. The estimator (6.2) was introduced by Hill [53] and is therefore often called the **Hill estimator**. That the estimators (6.2) and (6.3) are strongly consistent was proved in [29]. The optimal choice of $k(n)$ can be found in [52]. A comprehensive introduction into the field can be found in the book [36]. This book mainly discusses the approach via extreme value theory.

7. The Ammeter Risk Model

7.1. Mixed Poisson Risk Processes

We have seen that a continuous time risk model is "easy" to handle if the claim number process is a Poisson process. Considering a risk in a fixed time interval we learned that a compound Poisson model does not fit real data very well. We had to use a compound mixed Poisson model, as for instance the compound negative binomial model. The reason that these models are closer to reality is that they allow more fluctuation.

It is natural to ask whether we can allow for more fluctuation in the Poisson process. We already saw one way to allow for more fluctuation, that is the renewal risk model. But this is not an analogue of the compound mixed Poisson model in a fixed time interval. The easiest way in order to obtain a mixed Poisson distribution for the number of claims in any finite time interval is to let the claim arrival intensity to be stochastic. Such a model is called **mixed Poisson risk process**.

Let $H(x)$ be the distribution function of λ. Then, using the usual notation,

$$\psi(u) = \int_0^\infty \psi^{\mathrm{c}}(u,\ell)\,\mathrm{d}H(\ell) = \int_0^{c/\mu} \psi^{\mathrm{c}}(u,\ell)\,\mathrm{d}H(\ell) + (1 - H(c/\mu))$$

where $\psi^{\mathrm{c}}(u,\ell)$ denotes the ruin probability in the Cramér-Lundberg model with claim arrival intensity ℓ. It follows immediately that there can only exist an exponential upper bound if there exists an ℓ_0 such that $H(\ell_0) = 1$ and $c > \mu\ell_0$. Otherwise, as for instance if λ is gamma distributed, Lundberg's inequality or the Cramér-Lundberg approximation cannot hold. In particular, there is no possibility to formulate a net profit condition without knowing an upper bound for λ.

Assume we have observed the process up to time t. We want to show that $\mathbb{P}[\lambda \le \ell \mid \mathcal{F}_t] = \mathbb{P}[\lambda \le \ell \mid N_t]$. Because $\{N_t\}$ and $\{Y_i\}$ are independent it follows that $\mathbb{P}[\lambda \le \ell \mid \mathcal{F}_t] = \mathbb{P}[\lambda \le \ell \mid N_t, T_1, T_2, \ldots, T_{N_t}]$. Recall that, given $N_t = n$ the claim times $T_1, T_2, \ldots, T_n$ have the same distribution as the order statistics of n independent uniformly in $[0,t]$ distributed random variables (Proposition C.2). Let

$n \in \mathbb{N} \setminus \{0\}$ and let $0 \le t_i \le t$ for $1 \le i \le n$. Then

$$\begin{aligned}
&\mathbb{P}[\lambda \le \ell, N_t = n, T_1 \le t_1, \ldots, T_n \le t_n] \\
&\quad = \mathbb{P}[\lambda \le \ell, T_1 \le t_1, \ldots, T_n \le t_n, T_{n+1} > t] \\
&\quad = \int_0^\ell \int_0^{t_1} x\mathrm{e}^{-xs_1} \int_{s_1}^{t_2 \vee s_1} x\mathrm{e}^{-x(s_2-s_1)} \cdots \int_{s_{n-1}}^{t_n \vee s_{n-1}} x\mathrm{e}^{-x(s_n-s_{n-1})} \\
&\qquad\qquad \times \mathrm{e}^{-x(t-s_n)} \,\mathrm{d}s_n \,\cdots\, \mathrm{d}s_1 \,\mathrm{d}H(x) \\
&\quad = \int_0^\ell \frac{(xt)^n}{n!} \mathrm{e}^{-xt} \int_0^{t_1} \int_{s_1}^{t_2 \vee s_1} \cdots \int_{s_{n-1}}^{t_n \vee s_{n-1}} n! t^{-n} \,\mathrm{d}s_n \,\ldots\, \mathrm{d}s_1 \,\mathrm{d}H(x) \\
&\quad = \mathbb{P}[\lambda \le \ell, N_t = n] \mathbb{P}[T_1 \le t_1, \ldots, T_n \le t_n \mid N_t = n] \\
&\quad = \mathbb{P}[\lambda \le \ell \mid N_t = n] \mathbb{P}[T_1 \le t_1, \ldots, T_n \le t_n, N_t = n]
\end{aligned}$$

and it follows that

$$\mathbb{P}[\lambda \le \ell \mid N_t = n, T_1 \le t_1, \ldots, T_n \le t_n] = \mathbb{P}[\lambda \le \ell \mid N_t = n]\,.$$

Thus $\mathbb{P}[\lambda \le \ell \mid \mathcal{F}_t] = \mathbb{P}[\lambda \le \ell \mid N_t]$.

We know from the renewal theory (using the strong law of large numbers) that, given λ, $t^{-1}N_t$ converges to λ a.s.. Thus

$$\mathbb{P}[\lim_{t\to\infty} t^{-1}N_t = \lambda] = 1\,.$$

Moreover, $\operatorname{Var}[\lambda \mid N_t] \to 0$ as $t \to \infty$. Thus as time increases the desired variability disappears. Hence this model is not the desired generalisation of the mixed Poisson model in a finite time interval.

7.2. Definition of the Ammeter Risk Model

In 1948 the Swiss actuary Hans Ammeter had the idea to combine a sequence of mixed Poisson models. He wanted to start a new independent model at the beginning of each year. That means a new intensity level is chosen at the beginning of each year independently of the past intensity levels. This has the advantage that the variability does not disappear with time.

For a formal definition let Δ denote the length of the time interval in which the intensity level remains constant, for instance one year. Let $\{L_i\}$ be a sequence of positive random variables with distribution function H. We allow for the possibility that $H(0) \ne 0$. The intensity at time t is

$$\lambda_t := L_i, \qquad\qquad \text{if } (i-1)\Delta \le t < i\Delta.$$

The stochastic process $\{\lambda_t\}$ is called **claim arrival intensity process** of the risk model. Define $\Lambda_t = \int_0^t \lambda_s \, \mathrm{d}s$ and let $\{\tilde{N}_t\}$ be a homogeneous Poisson process with rate 1. The claim number process is then

$$N_t := \tilde{N}_{\Lambda_t} \,.$$

Intuitively the point process $\{N_t\}$ can be obtained in the following way. First generate a realization $\{\lambda_t\}$. Given $\{\lambda_t : 0 \le t < \infty\}$ let $\{N_t\}$ be an inhomogeneous Poisson process with rate $\{\lambda_t\}$ (compare with Proposition C.5). A process obtained in this way is called **doubly stochastic point process** or **Cox process**. The process $\{N_t\}$ constructed above and the mixed Poisson process are special cases of Cox processes.

As for the models considered before let the claims $\{Y_i\}$ be a sequence of iid. positive random variables with distribution function G. The Ammeter risk model is the the process

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

where u is the initial capital and c is the premium rate. As before let $\tau := \inf\{t > 0 : C_t < 0\}$ denote the time of ruin and $\psi(u) = \mathbb{P}[\tau < \infty \mid C_0 = u]$ denote the ruin probability.

In order to investigate the model we define also the probability of ruin obtained by considering only the times $\{\Delta, 2\Delta, \dots\}$ where the intensity changes, i.e. let $\tau_\Delta = \inf\{i\Delta : i \in \mathbb{N}, C_{i\Delta} < 0\}$ and $\psi_\Delta(u) := \mathbb{P}[\tau_\Delta < \infty]$. Only considering the points $i\Delta$ is as though all the claims of the interval $((i-1)\Delta, i\Delta]$ are occurring at the time point $i\Delta$. But this is a renewal risk model

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

where $\{N_t^\Delta\}$ is a renewal process with interarrival time distribution $F_\Delta(t) = \mathbb{1}_{\{t \ge \Delta\}}$ and claim sizes $\{Y_i^\Delta\}$ with a compound mixed Poisson distribution. More explicitly,

$$G_\Delta(x) = \mathbb{P}[Y_i^\Delta \le x] = \mathbb{P}\Bigl[\sum_{i=1}^{N_\Delta} Y_i \le x\Bigr] \,.$$

Given L_1 the conditional expected number of claims and thus the Poisson parameter is ΔL_1. From (1.2) we already know the moment generating function

$$M_{Y^\Delta}(r) = M_L(\Delta(M_Y(r) - 1))$$

of Y_i^Δ where $M_L(r)$ is the moment generating function of the intensity levels L_i.

In order to avoid $\psi(u) = 1$ we need that the random walk $\{C_{k\Delta}\}$ tends to infinity, or equivalently that $\mathbb{E}[C_\Delta - u] > 0$. The mean value of the claims up to time Δ is $\mathbb{E}[L_1]\mu\Delta$ (see (1.1)) and the income is $c\Delta$. Thus the net profit condition must be

$$c > \mathbb{E}[L_1]\mu\,.$$

We assume in the future that the net profit condition is fulfilled.

The ruin probabilities of $\{C_t\}$ and of $\{C_t^\Delta\}$ can be compared in the following way.

Lemma 7.1. *For $\psi(u)$ and $\psi_\Delta(u)$ defined above we have*

$$\psi(u + c\Delta) \le \psi_\Delta(u) \le \psi(u) \le \psi_\Delta(u - c\Delta)\,.$$

Proof. Assume that there exists $t_0 > 0$ such that $C_{t_0} + c\Delta < 0$, i.e. ruin occurs for the Ammeter risk model with initial capital $u + c\Delta$. Let k_0 such that $t_0 \in ((k_0 - 1)\Delta, k_0\Delta]$. Then

$$0 > c\Delta + C_{t_0} = c\Delta + C_{k_0\Delta} - c(k_0\Delta - t_0) + \sum_{i=N_{t_0}+1}^{N_{k_0\Delta}} Y_i > C_{k_0\Delta}\,.$$

Thus $\psi(u + c\Delta) \le \psi_\Delta(u)$ by noting that $C_{k_0\Delta} = C_{k_0\Delta}^\Delta$. The last inequality follows from the first one and the second inequality is trivial. □

7.3. Lundberg's Inequality and the Cramér-Lundberg Approximation

7.3.1. The Ordinary Case

The first problem is to find the analogue of the Lundberg exponent for the Ammeter risk model. In the classical case and in the renewal case we constructed an exponential martingale. Here, as in the renewal case, the process $\{C_t\}$ is not a Markov process anymore. But for each $k \in \mathbb{N}$ the process $\{C_{t+k\Delta} - C_{k\Delta}\}$ is independent of $\mathcal{F}_{k\Delta}$. Let us therefore only consider the time points $0, \Delta, 2\Delta, \ldots$.

Lemma 7.2. *Let $\{C_t\}$ be an Ammeter risk model. For any $r \in \mathbb{R}$ such that $M_L(\Delta(M_Y(r) - 1)) < \infty$ let $\theta(r)$ be the unique solution to*

$$\mathrm{e}^{-(\theta(r)+cr)\Delta} M_L(\Delta(M_Y(r) - 1)) = 1\,.$$

Then the discrete time process

$$\{\exp\{-rC_{k\Delta} - \theta(r)k\Delta\}\}$$

is a martingale. Moreover, $\theta(r)$ is strictly convex, $\theta(0) = 0$ and $\theta'(0) = \mathbb{E}[L_1]\mu - c < 0$.

Proof. Considering the process at the time points $0, \Delta, 2\Delta, \ldots$ we find $C_{k\Delta} = C^{\Delta}_{k\Delta}$. These points are exactly the claim times of the renewal risk process $\{C^{\Delta}_{k\Delta}\}$. The assertion follows from Lemma 5.1 and from (5.3). □

Because $\theta(r)$ is a convex function and $\theta(0) = 0$ there may exist a second solution R to the equation $\theta(r) = 0$. R is, if it exists, unique and strictly positive. R is then called the **Lundberg exponent** or **adjustment coefficient** of the Ammeter risk model.

Next we proof Lundberg's inequality.

Theorem 7.3. *Let $\{C_t\}$ be an Ammeter risk model and assume that the Lundberg exponent R exists. Then*

$$\psi(u) < \mathrm{e}^{cR\Delta}\mathrm{e}^{-Ru}\,.$$

If there exists an $r > R$ such that $M_L(\Delta(M_Y(r) - 1)) < \infty$ then R is the right exponent in the sense that there exists a constant $\underline{C} > 0$ such that

$$\psi(u) \ge \underline{C}\mathrm{e}^{-Ru}\,.$$

Remark. The constant $\mathrm{e}^{cR\Delta}$ might be too large, especially if Δ is large. If possible, one should therefore try to find an alternative for the constant. An upper bound, based on the martingale (7.2), is found in [35] and [66]. However, one has to accept that the constant might be larger than 1. Recall that this was also the case for Lundberg's inequality in the general case of the renewal risk model. ■

Proof. Using Lemma 7.1 the proof becomes almost trivial. From Theorem 5.5 it follows that $\psi_\Delta(u) < \mathrm{e}^{-Ru}$. Thus

$$\psi(u) \le \psi_\Delta(u - c\Delta) < \mathrm{e}^{-R(u-c\Delta)} = \mathrm{e}^{cR\Delta}\mathrm{e}^{-Ru}\,.$$

If there exist an $r > R$ such that $M_L(\Delta(M_Y(r) - 1)) < \infty$ then by Theorem 5.7 there exists a constant $\underline{C} > 0$ such that $\psi_\Delta(u) \ge \underline{C}\mathrm{e}^{-Ru}$. Thus also

$$\psi(u) \ge \psi_\Delta(u) \ge \underline{C}\mathrm{e}^{-Ru}\,.$$

□

The constant $\mathrm{e}^{cR\Delta}$ seems to increase to infinity as Δ tends to infinity. But R also dependends on Δ. At least in some cases $\mathrm{e}^{cR\Delta}$ remains bounded as $\Delta \to \infty$.

Proposition 7.4. *Assume that R exists and that there exists a solution $r_0 \neq 0$ to*

$$M_L(\mu r)\mathrm{e}^{-cr} = 1\,.$$

Then

$$\mathrm{e}^{cR\Delta} \le \mathrm{e}^{cr_0}\,. \tag{7.1}$$

In particular, the condition will be fulfilled if R exists and if there exists $\ell_\infty \le \infty$ such that $M_L(r) < \infty \iff r < \ell_\infty$.

Remark. Note that r_0 depends on the claim size distribution via μ only. ■

Proof. The function $\log(M_L(\mu r)) - cr$ has second derivative

$$\mu^2\left(\frac{M_L''(\mu r)}{M_L(\mu r)} - \Bigl(\frac{M_L'(\mu r)}{M_L(\mu r)}\Bigr)^2\right) > 0$$

(see Lemma 1.9) and is therefore strictly convex. Because

$$\mu\Bigl(\frac{M_L'(0)}{M_L(0)}\Bigr) - c = \mu\mathbb{E}[L_i] - c < 0$$

r_0 has to be strictly positive. In particular, if

$$M_L(\mu r)\mathrm{e}^{-cr} = \exp\{\log(M_L(\mu r)) - cr\} \le 1$$

then $r \le r_0$. If $M_L(r) < \infty \iff r < \ell_\infty$ then by monotone convergence

$$\lim_{r\uparrow\ell_\infty} M_L(r) = M_L(\ell_\infty) = \infty\,.$$

Thus by the convexity a strictly positive solution to $\log(M_L(\mu r)) - cr = 0$ has to exist.

Recall that $M_L(r)$ is increasing and that $M_Y(r) - 1 \ge rM_Y'(0) = r\mu$. Since

$$1 = \mathrm{e}^{-cR\Delta}M_L(\Delta(M_Y(R) - 1)) \ge \mathrm{e}^{-cR\Delta}M_L(\Delta R\mu)$$

it follows from the consideration above that $R\Delta \le r_0$. □

Example 7.5. Assume that $L_i \sim \Gamma(\gamma, \alpha)$. The case $\Delta = 1$ was the case considered in Ammeter's original paper [2]. In order to obtain the Lundberg exponent we have to solve the equation

$$\mathrm{e}^{-cr\Delta}\Bigl(\frac{\alpha}{\alpha - \Delta(M_Y(r) - 1)}\Bigr)^{\gamma} = 1$$

or equivalently

$$\mathrm{e}^{-cr\Delta/\gamma}\frac{\alpha}{\alpha - \Delta(M_Y(r) - 1)} = 1\,.$$

This can be written as

$$[\alpha - \Delta(M_Y(r) - 1)]\,\mathrm{e}^{cr\Delta/\gamma} = \alpha\,.$$

It follows immediately that the above equation admits only a solution in closed form for special cases. For example, if $Y_i = 1$ is deterministic then one has to solve

$$\Delta\mathrm{e}^{(1+c\Delta/\gamma)r} - (\alpha + \Delta)\mathrm{e}^{(c\Delta/\gamma)r} + \alpha = 0\,.$$

One can only hope for a solution in closed form if $c\Delta/\gamma \in \mathbb{Q}$. For instance if $c\Delta/\gamma = 1$ then $R = \log(\alpha/\Delta) = \log(\alpha c/\gamma)$.

Let us now consider (7.1). r_0 is the strictly positive solution to

$$\Bigl(\frac{\alpha}{\alpha - r\mu}\Bigr)^{\gamma}\mathrm{e}^{-cr} = 1$$

or

$$(\alpha - r\mu)\mathrm{e}^{cr/\gamma} = \alpha\,.$$

The latter equation also has to be solved numerically. ■

From Theorem 7.3 we find that

$$\underline{C} \le \varliminf_{u\to\infty} \psi(u)\mathrm{e}^{Ru} \le \varlimsup_{u\to\infty} \psi(u)\mathrm{e}^{Ru} \le \mathrm{e}^{cR\Delta}\,.$$

The question is: exists there a constant C such that

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

i.e. does the Cramér-Lundberg approximation hold? This question is answered in the following theorem.

Theorem 7.6. *Let $\{C_t\}$ be an Ammeter risk model. Assume that R exists and that there exists an $r > R$ such that $M_L(\Delta(M_Y(r) - 1)) < \infty$. Then there exists a constant $0 < C \le \mathrm{e}^{cR\Delta}$ such that*

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

Proof. This is a special case of Theorem 2 of [71]. □

We next want to derive an alternative proof of Lundberg's inequality. For this reason we have to find an exponential martingale in continuous time. In applications it is usually rather easy to find a martingale if one deals with Markov processes. But we have remarked before that $\{C_t\}$ is not a Markov process. The future of the process is dependent on the present level λ_t of the intensity. Let us therefore consider the process $\{(C_t, \lambda_t)\}$. But this is not a homogeneous Markov process neither. The future of the process is also dependent on time because the future intensity level is known until the next change of the intensity. Instead of considering $\{(C_t, \lambda_t, t)\}$ we consider $\{(C_t, \lambda_t, V_t)\}$ where V_t is the time remaining till the next change of the intensity, i.e.

$$V_t = k\Delta - t, \qquad\qquad \text{if } (k-1)\Delta \le t < k\Delta.$$

Remarks.

i) The process $\{\lambda_t\}$ cannot be observed directly. The only information on λ_t are the number of claims occured in the interval $[t + V_t - \Delta, t]$. Thus, provided L_i is not deterministic, we can, in the best case, find the distribution of λ_t given $\mathcal{F}_t^C$, the information generated by $\{C_s : s \le t\}$ up to time t. But for our purpose it is no problem to work with the filtration $\mathcal{F}_t$ generated by $\{(C_t, \lambda_t, V_t)\}$. The results will not depend on the chosen filtration.

ii) As an alternative we could consider the time elapsed since the last change of the intensity $W_t = \Delta - V_t$. For the present choice of the Markovization see also Section 8.2. ■

We find the following martingale.

Lemma 7.7. *Let r such that $M_L(\Delta(M_Y(r) - 1)) < \infty$. Then the process*

$$M_t^r = \mathrm{e}^{-rC_t}\mathrm{e}^{(\lambda_t(M_Y(r)-1)-cr-\theta(r))V_t}\mathrm{e}^{-\theta(r)t} \tag{7.2}$$

is a martingale with respect to the natural filtration of $\{(C_t, \lambda_t, V_t)\}$.

Proof. Let first $(k-1)\Delta \le s < t < k\Delta$. Then $\lambda_t = \lambda_s$ and

$$\begin{aligned}\mathbb{E}[M_t^r \mid \mathcal{F}_s] &= \mathbb{E}\left[\mathrm{e}^{-rC_t}\mathrm{e}^{(\lambda_t(M_Y(r)-1)-cr-\theta(r))V_t}\mathrm{e}^{-\theta(r)t}\middle|\,\mathcal{F}_s\right] \\ &= M_s^r\mathbb{E}\left[\mathrm{e}^{-r(C_t-C_s)}\middle|\,\mathcal{F}_s\right]\mathrm{e}^{-(\lambda_t(M_Y(r)-1)-cr)(t-s)} \\ &= M_s^r\mathrm{e}^{(\lambda_t(M_Y(r)-1)-cr)(t-s)}\mathrm{e}^{-(\lambda_t(M_Y(r)-1)-cr)(t-s)} = M_s^r\,.\end{aligned}$$

Let now $(k-1)\Delta \le s < t = k\Delta$. Then, using the definition of $\theta(r)$,

$$\begin{aligned}\mathbb{E}[M_t^r \mid \mathcal{F}_s] &= M_s^r\mathbb{E}\left[\mathrm{e}^{-r(C_t-C_s)}\middle|\,\mathcal{F}_s\right]\mathbb{E}\left[\mathrm{e}^{(L_{k+1}(M_Y(r)-1)-cr-\theta(r))\Delta}\right] \\ &\qquad\times\mathrm{e}^{-(\lambda_s(M_Y(r)-1)-cr-\theta(r))(t-s)}\mathrm{e}^{-\theta(r)(t-s)} \\ &= M_s^r\mathrm{e}^{(\lambda_s(M_Y(r)-1)-cr)(t-s)}M_L(\Delta(M_Y(r)-1))\mathrm{e}^{-(cr+\theta(r))\Delta}\mathrm{e}^{-(\lambda_s(M_Y(r)-1)-cr)(t-s)} \\ &= M_s^r\,.\end{aligned}$$

The assertion follows by iteration. ☐

From the martingale stopping theorem we find

$$\mathrm{e}^{-Ru} = \mathbb{E}[M_0^R] = \mathbb{E}[M_{\tau\wedge t}^R] \ge \mathbb{E}[M_\tau^R; \tau \le t]\,.$$

By monotone convergence it follows that

$$\mathrm{e}^{-Ru} \ge \mathbb{E}[M_\tau^R; \tau < \infty] = \mathbb{E}[M_\tau^R \mid \tau < \infty]\mathbb{P}[\tau < \infty]\,,$$

from which

$$\psi(u) \le \frac{\mathrm{e}^{-Ru}}{\mathbb{E}[M_\tau^R \mid \tau < \infty]}$$

follows. Note that $C_\tau < 0$ on $\{\tau < \infty\}$. Thus

$$\mathbb{E}[M_\tau^R \mid \tau < \infty] > \mathbb{E}[\mathrm{e}^{(\lambda_\tau(M_Y(R)-1)-cR)V_\tau} \mid \tau < \infty] > \mathbb{E}[\mathrm{e}^{-cRV_\tau} \mid \tau < \infty] > \mathrm{e}^{-cR\Delta}\,.$$

We have recovered Lundberg's inequality

$$\psi(u) < \mathrm{e}^{cR\Delta}\mathrm{e}^{-Ru}\,.$$

7.3.2. The General Case

The second proof of Lundberg's inequality is in particular useful if the process does not start at a point where the intensity changes. For instance, the premium for the next year has to be determined before the end of the year. In this case the actuary has some information about the present intensity via the number of claims occured in the first part of the year. Let us therefore assume that (λ_0, V_0) has the joint distribution function $F^1(\ell, v)$. We assume that $\mathbb{P}[V_0 \le \Delta] = 1$. By allowing V_0 to be stochastic we can also treat cases where no information on the intensity changing times is available. This might be a little bit strange from a practical point of view, but it is nevertheless interesting from a theoretical point of view. One special case is the case of a stationary intensity process where

$$F^1(\ell, v) = \Bigl(\frac{v}{\Delta}\mathbb{1}_{0\le v\le\Delta} + \mathbb{1}_{v>\Delta}\Bigr)H(\ell)\,.$$

We get the following version of Lundberg's inequality.

Theorem 7.8. *Let $\{C_t\}$ be an Ammeter risk model. Assume that the Lundberg exponent R exists and that*

$$\mathbb{E}[\mathrm{e}^{\lambda_0(M_Y(R)-1)V_0}] < \infty\,.$$

Then

$$\psi(u) < \mathrm{e}^{cR\Delta}\mathbb{E}[\mathrm{e}^{(\lambda_0(M_Y(R)-1)-cR)V_0}]\mathrm{e}^{-Ru}\,.$$

Remark. The condition

$$\mathbb{E}\left[\mathrm{e}^{\lambda_0(M_Y(R)-1)V_0}\right] < \infty$$

will be fulfilled in all reasonable situations. For instance, if (λ_0, V_0) is distributed such that $\{\lambda_t\}$ becomes stationary, then

$$\mathbb{E}[\mathrm{e}^{\lambda_0(M_Y(R)-1)V_0}] = \frac{1}{\Delta}\int_0^\Delta M_L(v(M_Y(R)-1))\,\mathrm{d}v < M_L(\Delta(M_Y(R)-1)) < \infty\,.$$

■

Proof. From the stopping theorem it follows that

$$\mathrm{e}^{-Ru}\mathbb{E}\left[\mathrm{e}^{(\lambda_0(M_Y(R)-1)-cR)V_0}\right] = \mathbb{E}[M_0^R] = \mathbb{E}[M_{\tau\wedge t}^R] \ge \mathbb{E}[M_\tau^R; \tau \le t]\,.$$

The assertion follows now as before. ☐

For the Cramér-Lundberg approximation we need the following

Lemma 7.9. *Assume that an $r > R$ exists such that $M_L(\Delta(M_Y(r)-1)) < \infty$ and that*

$$\mathbb{E}[\mathrm{e}^{\lambda_0(M_Y(r)-1)V_0}] < \infty\,.$$

Then

$$\lim_{u\to\infty} \mathbb{P}[\tau \le V_0 \mid C_0 = u]\mathrm{e}^{Ru} = 0\,.$$

Proof. Again using the stopping theorem we get

$$\mathrm{e}^{-ru}\mathbb{E}[\mathrm{e}^{(\lambda_0(M_Y(r)-1)-cr-\theta(r))V_0}] = \mathbb{E}[M_0^r] = \mathbb{E}[M_{\tau\wedge V_0}^r] \ge \mathbb{E}[M_\tau^r; \tau \le V_0]\,.$$

As before it follows that

$$\begin{aligned}\mathbb{P}[\tau \le V_0] &\le \mathrm{e}^{(rc+\theta(r))\Delta}\mathbb{E}[\mathrm{e}^{(\lambda_0(M_Y(r)-1)-cr-\theta(r))V_0}]\mathrm{e}^{-ru}\\ &< \mathrm{e}^{(rc+\theta(r))\Delta}\mathbb{E}[\mathrm{e}^{\lambda_0(M_Y(r)-1)V_0}]\mathrm{e}^{-ru}\end{aligned}$$

where we used that $\theta(r) > 0$. Multiplying with e^{Ru} and then letting $u \to \infty$ yields the assertion. □

The Cramér-Lundberg approximation in the general case can now be proved.

Theorem 7.10. *Let $\{C_t\}$ be an Ammeter risk model. Assume that R exists and that there exists an $r > R$ such that both $M_L(\Delta(M_Y(r)-1)) < \infty$ and*

$$\mathbb{E}[\mathrm{e}^{\lambda_0(M_Y(r)-1)V_0}] < \infty\,.$$

Then

$$\lim_{u\to\infty} \psi(u)\mathrm{e}^{Ru} = \mathbb{E}[\mathrm{e}^{(\lambda_0(M_Y(R)-1)-cR)V_0}]\,C$$

where C is the constant obtained in Theorem 7.6.

Proof. Let $\psi^{\mathrm{o}}(u)$ denote the ruin probability in the ordinary case and define

$$B(x;\ell,v) := \mathbb{P}[C_v - u \le x \mid \lambda_0 = \ell, V_0 = v]\,.$$

Then

$$\begin{aligned}\psi(u) = \int_{\ell=0}^{\infty}\int_{v=0}^{\Delta}\int_{y=-u}^{cv} &\psi^{\mathrm{o}}(u+y)\mathbb{P}[\tau > v \mid \lambda_0 = \ell, V_0 = v, C_v = u+y]\\ &\times B(\mathrm{d}y;\ell,v)F^1(\mathrm{d}\ell,\mathrm{d}v) + \mathbb{P}[\tau \le V_0]\,.\end{aligned}$$

We multiply the above equation by e^{Ru}. The last term tends to 0 by Lemma 7.9. The first term on the right hand side can be written as

$$\begin{aligned}\int_{\ell=0}^{\infty}\int_{v=0}^{\Delta}\int_{y=-u}^{cv} \psi^{\mathrm{o}}(u+y)\mathrm{e}^{R(u+y)}\mathbb{P}[\tau > v \mid \lambda_0=\ell, V_0=v, C_v=u+y] \\ \times\, \mathrm{e}^{-Ry}B(\mathrm{d}y;\ell,v)F^1(\mathrm{d}\ell,\mathrm{d}v)\,.\end{aligned}$$

Noting that by Theorem 7.6 $\psi^{\mathrm{o}}(u+y)\mathrm{e}^{R(u+y)}$ is uniformly bounded by a constant $\bar{C}$ we see that

$$\begin{aligned}\int_{\ell=0}^{\infty}\int_{v=0}^{\Delta}\int_{y=-u}^{cv} \psi^{\mathrm{o}}(u+y)\mathrm{e}^{R(u+y)}\mathbb{P}[\tau > v \mid \lambda_0=\ell, V_0=v, C_v=u+y] \\ \times\, \mathrm{e}^{-Ry}B(\mathrm{d}y;\ell,v)F^1(\mathrm{d}\ell,\mathrm{d}v) \\ \le \bar{C}\mathbb{E}[\mathrm{e}^{-R(C_{V_0}-u)}] = \bar{C}\mathbb{E}[\mathrm{e}^{(\lambda_0(M_Y(R)-1)-cR)V_0}] < \infty\,.\end{aligned}$$

We can therefore interchange integration and limit. Because $\psi^{\mathrm{o}}(u+y)\mathrm{e}^{R(u+y)}$ tends to C and $\mathbb{P}[\tau > v \mid \lambda_0=\ell, V_0=v, C_v=u+y]$ tends to 1 as $u \to \infty$ the desired limit is

$$\begin{aligned}\lim_{u\to\infty}\psi(u)\mathrm{e}^{Ru} &= C\int_{\ell=0}^{\infty}\int_{v=0}^{\Delta}\int_{y=-\infty}^{cv} \mathrm{e}^{-Ry}B(\mathrm{d}y;\ell,v)F^1(\mathrm{d}\ell,\mathrm{d}v) \\ &= \mathbb{E}[\mathrm{e}^{(\lambda_0(M_Y(R)-1)-cR)V_0}]\,C\,.\end{aligned}$$

□

7.4. The Subexponential Case

A subexponential behaviour of the ruin probability in the Ammeter risk model can be caused by two sources. First the claim sizes can be subexponential. More explicitly, if there exists an $\varepsilon > 0$ such that $M_L(\varepsilon) < \infty$ and Y_i has a subexponential distribution, then Y_i^Δ has a subexponential distribution. But it is also possible that L_i has a distribution which makes Y_i^Δ subexponential. In any case we have the following

Proposition 7.11. *Assume that*

$$\frac{1}{\Delta\mathbb{E}[L_i]\mu}\int_0^x (1-G_\Delta(y))\,\mathrm{d}y$$

is subexponential. Then

$$\lim_{u\to\infty}\frac{\psi(u)}{\int_u^\infty (1-G_\Delta(y))\,\mathrm{d}y} = \frac{1}{c\Delta - \Delta\mathbb{E}[L_i]\mu}\,.$$

Proof. From Theorem 5.12 the asymptotic behaviour in the assertion is true for $\psi(u)$ replaced by $\psi_\Delta(u)$. It is therefore sufficient to show that $\psi(u)/\psi_\Delta(u)$ tends to 1 as $u \to \infty$. We have from Lemma 7.1

$$1 \le \frac{\psi(u)}{\psi_\Delta(u)} \le \frac{\psi_\Delta(u - c\Delta)}{\psi_\Delta(u)}\,.$$

Using Theorem 5.12 and Lemma F.2 we obtain

$$\lim_{u\to\infty} \frac{\psi_\Delta(u-c\Delta)}{\psi_\Delta(u)} = \lim_{u\to\infty} \frac{\int_{u-c\Delta}^\infty (1 - G_\Delta(y))\,\mathrm{d}y}{\int_u^\infty (1 - G_\Delta(y))\,\mathrm{d}y} = 1\,.$$

This proves the assertion. □

The next result deals with the tail of compound distributions with a regularly varying tail.

Lemma 7.12. *Let N be a random variable with values in $\mathbb{N}$ such that $\mathbb{E}[N] < \infty$, $\{Y_i\}$ be a sequence of positive iid. random variables independent of N with distribution function $G(y)$ and mean μ, and let $\tilde{G}(y)$ be the distribution function of $\sum_{i=1}^N Y_i$. Let $\delta > 1$ and $L(x)$ be a slowly varying function. Assume that*

$$\lim_{x\to\infty} \frac{x^\delta(1-G(x))}{L(x)} = \phi, \qquad \lim_{n\to\infty} \frac{n^\delta \mathbb{P}[N > n]}{L(n)} = \gamma$$

where ϕ and γ are positive constants. Then

$$\lim_{x\to\infty} \frac{x^\delta(1-\tilde{G}(x))}{L(x)} = \mu^\delta \gamma + \mathbb{E}[N]\phi\,.$$

Proof. See [79]. □

Next we link the distribution function $H(\ell)$ of L_i to the distribution function of N_Δ.

Lemma 7.13. *Assume that there exists a $\delta > 0$ and a slowly varying function $L(x)$ such that*

$$\lim_{\ell\to\infty} \frac{(1-H(\ell))\ell^\delta}{L(\ell)} = 0\,.$$

Then

$$\lim_{n\to\infty} \frac{n^\delta \mathbb{P}[N_\Delta > n]}{L(n)} = 0\,.$$

Proof. First note that for $n \geq 1$

$$\begin{aligned}
\mathbb{P}[N_\Delta > n] &= \int_0^\infty \sum_{m=n+1}^\infty \frac{(\Delta\ell)^m}{m!} \mathrm{e}^{-\Delta\ell} \,\mathrm{d}H(\ell) \\
&= \Delta \int_0^\infty \sum_{m=n+1}^\infty \int_0^\ell \Bigl(\frac{(\Delta y)^{m-1}}{(m-1)!} - \frac{(\Delta y)^m}{m!} \Bigr) \mathrm{e}^{-\Delta y} \,\mathrm{d}y \,\mathrm{d}H(\ell) \\
&= \Delta \int_0^\infty \int_y^\infty \sum_{m=n+1}^\infty \Bigl(\frac{(\Delta y)^{m-1}}{(m-1)!} - \frac{(\Delta y)^m}{m!} \Bigr) \mathrm{e}^{-\Delta y} \,\mathrm{d}H(\ell) \,\mathrm{d}y \\
&= \Delta \int_0^\infty \frac{(\Delta y)^n}{n!} \mathrm{e}^{-\Delta y} (1 - H(y)) \,\mathrm{d}y = \int_0^\infty \frac{y^n}{n!} \mathrm{e}^{-y} (1 - H(y/\Delta)) \,\mathrm{d}y \,.
\end{aligned}$$

Choose now $a \leq \mathrm{e}^{-1}$. Then

$$\begin{aligned}
\int_0^{an} \frac{y^n}{n!} \mathrm{e}^{-y} (1 - H(y/\Delta)) \,\mathrm{d}y \leq \int_0^{an} \frac{y^n}{n!} \mathrm{e}^{-y} \,\mathrm{d}y &\leq \frac{(an)^{n+1}}{n!} \mathrm{e}^{-an} \\
&\leq \frac{n^{n+1}}{n!} \mathrm{e}^{-an-n}
\end{aligned}$$

because $y^n \mathrm{e}^{-y}$ is increasing in the interval $(0, n)$. By Stirling's formula there exists $K > 0$ such that $n! \geq K n^{n+\frac{1}{2}} \mathrm{e}^{-n}$ from which it follows that

$$n^\delta L(n)^{-1} \int_0^{an} \frac{y^n}{n!} \mathrm{e}^{-y} (1 - H(y/\Delta)) \,\mathrm{d}y$$

converges to 0 as $n \to \infty$. The latter follows because $x^{-\varepsilon}/L(x)$ converges to 0 for any $\varepsilon > 0$.

For the remaining part we obtain

$$\begin{aligned}
\int_{an}^\infty \frac{y^n}{n!} \mathrm{e}^{-y} (1 - H(y/\Delta)) \,\mathrm{d}y &\leq \int_{an}^\infty \frac{1}{\Gamma(n+1)} y^n \mathrm{e}^{-y} \,\mathrm{d}y \,(1 - H(an/\Delta)) \\
&\leq 1 - H(an/\Delta) \,.
\end{aligned}$$

Thus

$$n^\delta L(n)^{-1} \int_{an}^\infty \frac{y^n}{n!} \mathrm{e}^{-y} (1 - H(y/\Delta)) \,\mathrm{d}y$$

converges to 0 because

$$n^\delta L(n)^{-1} (1 - H(an/\Delta)) = \Bigl(\frac{\Delta}{a} \Bigr)^\delta \frac{(an/\Delta)^\delta (1 - H(an/\Delta))}{L(an/\Delta)} \frac{L(an/\Delta)}{L(n)}$$

converges to 0. □

Lemma 7.14. *Let $L(x)$ be a locally bounded slowly varying function and let $\delta > 0$. Then*

$$\lim_{y\to\infty} \frac{\int_0^\infty x^{y-\delta}\mathrm{e}^{-x}L(x)\,\mathrm{d}x}{\Gamma(y-\delta+1)L(y)} = 1\,.$$

Proof. There exists a constant K such that $L(x) \le Kx^\delta$. Assume that y is so large that $L(y) > y^{-\delta}$. Then

$$\int_0^{\mathrm{e}^{-1}y} \frac{x^{y-\delta}\mathrm{e}^{-x}L(x)}{\Gamma(y-\delta+1)}\,\mathrm{d}x \Big/ L(y) \le \frac{Ky^\delta}{\Gamma(y-\delta+1)} \int_0^{\mathrm{e}^{-1}y} x^y\mathrm{e}^{-x}\,\mathrm{d}x < \frac{Ky^{y+\delta+1}\mathrm{e}^{-(1+\mathrm{e}^{-1})y}}{\Gamma(y-\delta+1)}$$

which converges to 0 as $y \to \infty$ by Stirling's formula. The last inequality follows because $x^y\mathrm{e}^{-x}$ is increasing on $[0, y]$.

Analogously

$$\begin{aligned}\int_{2y}^\infty \frac{x^{y-\delta}\mathrm{e}^{-x}L(x)}{\Gamma(y-\delta+1)}\,\mathrm{d}x \Big/ L(y) &\le \frac{Ky^\delta}{\Gamma(y-\delta+1)} \int_{2y}^\infty x^y\mathrm{e}^{-x/2}\mathrm{e}^{-x/2}\,\mathrm{d}x \\ &\le \frac{Ky^\delta(2y)^y}{\Gamma(y-\delta+1)}\mathrm{e}^{-y}2\mathrm{e}^{-y}\end{aligned}$$

which again converges to 0 as $y \to \infty$ by Stirling's formula.

Because $L(xy)/L(y)$ converges uniformly as $y \to \infty$ for $\mathrm{e}^{-1} \le x \le 2$ it follows that

$$\lim_{y\to\infty} \frac{\int_{\mathrm{e}^{-1}y}^{2y} x^{y-\delta}\mathrm{e}^{-x}L(x)\,\mathrm{d}x}{\Gamma(y-\delta+1)L(y)} = \lim_{y\to\infty} \frac{\int_{\mathrm{e}^{-1}y}^{2y} x^{y-\delta}\mathrm{e}^{-x}\,\mathrm{d}x}{\Gamma(y-\delta+1)} = 1$$

where the last equality follows from the considerations above with $L(x) = 1$. □

Lemma 7.15. *Assume that there exists a $\delta > 0$ and a locally bounded slowly varying function $L(x)$ such that*

$$\lim_{\ell\to\infty} \frac{(1-H(\ell))\ell^\delta}{L(\ell)} = 1\,.$$

Then

$$\lim_{n\to\infty} \frac{n^\delta \mathbb{P}[N_\Delta > n]}{L(n)} = \Delta^\delta\,.$$

Proof. Let us consider

$$\begin{aligned}&\left| \frac{\mathbb{P}[N_\Delta > n] - \Delta^\delta \int_0^\infty \frac{\ell^{n-\delta}}{n!}\mathrm{e}^{-\ell}L(\ell/\Delta)\,\mathrm{d}\ell}{\Delta^\delta \int_0^\infty \frac{\ell^{n-\delta}}{n!}\mathrm{e}^{-\ell}L(\ell/\Delta)\,\mathrm{d}\ell} \right| \\ &\qquad = \left| \frac{\int_0^\infty \ell^{n-\delta}\mathrm{e}^{-\ell}L(\ell/\Delta)\left(\frac{(\ell/\Delta)^\delta(1-H(\ell/\Delta))}{L(\ell/\Delta)} - 1\right)\,\mathrm{d}\ell}{\int_0^\infty \ell^{n-\delta}\mathrm{e}^{-\ell}L(\ell/\Delta)\,\mathrm{d}\ell} \right|.\end{aligned}$$

Choose x_0 such that

$$1-\frac{\varepsilon}{2}<\frac{(\ell/\Delta)^\delta(1-H(\ell/\Delta))}{L(\ell/\Delta)}<1+\frac{\varepsilon}{2}$$

for $\ell \ge x_0$. Then

$$\left|\frac{\int_{x_0}^\infty \ell^{n-\delta}\mathrm{e}^{-\ell}L(\ell/\Delta)\left(\frac{(\ell/\Delta)^\delta(1-H(\ell/\Delta))}{L(\ell/\Delta)}-1\right)\,\mathrm{d}\ell}{\int_0^\infty \ell^{n-\delta}\mathrm{e}^{-\ell}L(\ell/\Delta)\,\mathrm{d}\ell}\right|<\frac{\varepsilon}{2}\,.$$

The remaining term can be estimated denoting $\bar{L}=\max\{L(x/\Delta):x\le x_0\}$

$$\begin{aligned}&\frac{\Delta^{-\delta}\int_0^{x_0}\ell^n\mathrm{e}^{-\ell}(1-H(\ell/\Delta))\,\mathrm{d}\ell+\int_0^{x_0}\ell^{n-\delta}\mathrm{e}^{-\ell}L(\ell/\Delta)\,\mathrm{d}\ell}{\int_0^\infty \ell^{n-\delta}\mathrm{e}^{-\ell}L(\ell/\Delta)\,\mathrm{d}\ell}\\ &\qquad\le\frac{\Delta^{-\delta}x_0^{n+1}+x_0^{n-\delta+1}\bar{L}}{\int_{2x_0}^\infty \ell^{n-\delta}\mathrm{e}^{-\ell}L(\ell/\Delta)\,\mathrm{d}\ell}\le\Bigl(\frac{1}{2}\Bigr)^{n-\delta}\frac{\Delta^{-\delta}x_0^{\delta+1}+x_0\bar{L}}{\int_{2x_0}^\infty \mathrm{e}^{-\ell}L(\ell/\Delta)\,\mathrm{d}\ell}\end{aligned}$$

Choosing n large enough the latter can be made smaller than $\varepsilon/2$. Thus

$$\lim_{n\to\infty}\frac{\mathbb{P}[N_\Delta>n]}{\int_0^\infty \frac{\ell^{n-\delta}}{n!}\mathrm{e}^{-\ell}L(\ell/\Delta)\,\mathrm{d}\ell}=\Delta^\delta\,.$$

It follows from Stirling's formula that

$$\lim_{n\to\infty}\frac{n^\delta\Gamma(n-\delta+1)}{n!}=1$$

and the assertion follows from Lemma 7.14 noting that $L(x/\Delta)$ is also a slowly varying function. ☐

We can now prove the asymptotic behaviour in the case where the heavier tail of G and H is regularly varying.

Theorem 7.16. *Assume that there exist a slowly varying function $L(x)$ and positive constants δ, ω and ϕ with $\delta>1$ and $\omega+\phi>0$ such that*

$$\lim_{\ell\to\infty}\frac{\ell^\delta(1-H(\ell))}{L(\ell)}=\omega,\qquad\lim_{x\to\infty}\frac{x^\delta(1-G(x))}{L(x)}=\phi\,.$$

Then

$$\lim_{u\to\infty}\frac{\psi(u)u^{\delta-1}}{L(u)}=\frac{\phi\mathbb{E}[L_i]\Delta+\omega(\mu\Delta)^\delta}{(c-\mathbb{E}[L_i]\mu)\Delta(\delta-1)}\,.$$

Proof. By Lemma 7.13 and Lemma 7.15

$$\lim_{n\to\infty}\frac{n^\delta\mathbb{P}[N_\Delta>n]}{L(n)}=\omega\Delta^\delta\,.$$

From Lemma 7.12 one obtains

$$\lim_{x\to\infty}\frac{x^\delta(1-G_\Delta(x))}{L(x)}=\mathbb{E}[L_i]\Delta\phi+\mu^\delta\omega\Delta^\delta\,.$$

The result follows now from Proposition 7.11 and Karamata's theorem. ☐

Example 7.5 (continued). Assume that the claims are $\mathrm{Pa}(\delta,\beta)$ distributed with $\delta > 1$. Then

$$\lim_{x\to\infty} x^\delta(1-G(x)) = \beta^\delta$$

and

$$\lim_{x\to\infty} x^\delta(1-H(x)) = \lim_{x\to\infty} \frac{\alpha^\gamma x^{\gamma-1}\mathrm{e}^{-\alpha x}}{\Gamma(\gamma)\delta x^{-(\delta+1)}} = 0\,.$$

Thus

$$\lim_{u\to\infty} u^{\delta-1}\psi(u) = \frac{\beta^\delta\gamma/\alpha\,\Delta}{(c-\gamma/\alpha\,\beta/(\delta-1))\Delta(\delta-1)} = \frac{\beta^\delta\gamma}{(c\alpha(\delta-1)-\gamma\beta)}\,.$$

■

Example 7.17. Let now L_i be $\mathrm{Pa}(\delta,\beta)$ distributed and assume that

$$\lim_{x\to\infty} x^\delta(1-G(x)) = 0\,.$$

Then it follows that

$$\lim_{u\to\infty} u^{\delta-1}\psi(u) = \frac{\beta^\delta(\mu\Delta)^\delta}{(c-\beta/(\delta-1)\,\mu)\Delta(\delta-1)} = \frac{(\beta\mu\Delta)^\delta}{(c(\delta-1)-\beta\mu)\Delta}\,.$$

■

7.5. Finite Time Lundberg Inequalities

Let us now return to the small claim case. As in the classical case it is possible to find exponential inequalities for ruin probabilities of the form

$$\mathbb{P}[\underline{y}u < \tau \le \bar{y}u]\,.$$

Recall that

$$M_t^r = \mathrm{e}^{-rC_t}\mathrm{e}^{(\lambda_t(M_Y(r)-1)-cr-\theta(r))V_t}\mathrm{e}^{-\theta(r)t}$$

is a martingale provided $M_L(\Delta(M_Y(r)-1)) < \infty$.

Let now $0 \le \underline{y} < \bar{y} < \infty$. Using the stopping theorem we get

$$\begin{aligned}
\mathrm{e}^{-ru}\mathbb{E}[\mathrm{e}^{(\lambda_0(M_Y(r)-1)-cr-\theta(r))V_0}] &= \mathbb{E}[M_0^r] = \mathbb{E}[M_{\tau\wedge\bar{y}u}] \\
&> \mathbb{E}[\mathrm{e}^{-rC_\tau}\mathrm{e}^{(\lambda_\tau(M_Y(r)-1)-cr-\theta(r))V_\tau}\mathrm{e}^{-\theta(r)\tau} \mid \underline{y}u < \tau \le \bar{y}u]\mathbb{P}[\underline{y}u < \tau \le \bar{y}u] \\
&> \mathrm{e}^{-(cr+\theta(r))\Delta}\mathrm{e}^{-\max\{\theta(r)\underline{y},\theta(r)\bar{y}\}u}\mathbb{P}[\underline{y}u < \tau \le \bar{y}u]\,.
\end{aligned}$$

Thus

$$\mathbb{P}[\underline{y}u < \tau \le \bar{y}u] < \mathrm{e}^{(cr+\theta(r))\Delta}\mathbb{E}[\mathrm{e}^{(\lambda_0(M_Y(r)-1)-cr-\theta(r))V_0}]\mathrm{e}^{-\min\{r-\theta(r)\underline{y},r-\theta(r)\bar{y}\}u}$$

and we can define the finite time Lundberg exponent.

$$R(\underline{y},\bar{y}) := \sup\{\min\{r-\theta(r)\underline{y}, r-\theta(r)\bar{y}\} : r \in \mathbb{R}\}\,.$$

Let $r_\infty = \sup\{r \ge 0 : M_L(\Delta(M_Y(r)-1)) < \infty\}$ and assume that $r_\infty > R$. We have discussed $R(\underline{y},\bar{y})$ in the classical case (Section 4.14). Let r_y be the argument maximizing $r - \theta(r)y$. We obtain the critical value

$$y_0 = (\theta'(R))^{-1} = \Bigl(\frac{M_L'(\Delta(M_Y(R)-1))M_Y'(R)}{M_L(\Delta(M_Y(R)-1))} - c\Bigr)^{-1}$$

and the two finite time Lundberg inequalities

$$\mathbb{P}[0 < \tau \le yu \mid C_0 = u] < \mathrm{e}^{(cr_y+\theta(r_y))\Delta}\mathbb{E}[\mathrm{e}^{(\lambda_0(M_Y(r_y)-1)-cr_y-\theta(r_y))V_0}]\mathrm{e}^{-R(0,y)u}$$

where $R(0,y) > R$ if $y < y_0$, and

$$\mathbb{P}[yu < \tau < \infty \mid C_0 = u] < \mathrm{e}^{(cr_y+\theta(r_y))\Delta}\mathbb{E}[\mathrm{e}^{(\lambda_0(M_Y(r_y)-1)-cr_y-\theta(r_y))V_0}]\mathrm{e}^{-R(y,\infty)u}$$

where $R(y,\infty) > R$ if $y > y_0$. The next result can be proved in the same way as Theorem 4.24 was proved.

Theorem 7.18. *Assume that $R < r_\infty$. Then*

$$\frac{\tau}{u} \longrightarrow y_0$$

in probability on the set $\{\tau < \infty\}$. □

Bibliographical Remarks

For a comprehensive introduction to mixed Poisson processes see [49]. The Ammeter risk process was introduced by Ammeter [2]. The process, as defined here, was first considered by Björk and Grandell [18], where also Lundberg's inequality was proved. The approach used here can be found in [48]. In particular, Lemma 7.1, Proposition 7.4, Proposition 7.11, Lemma 7.13 and Theorem 7.16 are taken from [48]. Lemma 7.7 was first obtained in [35]. Lemma 7.15 goes back to [49]. The finite time Lundberg inequalities were first proved in [35].

8. Change of Measure Techniques

In the risk models considered so far we constructed exponential martingales. Using the stopping theorem we were able to prove an upper bound for the ruin probabilities. But the technique did not allow to prove a lower bound or the Cramér-Lundberg approximation. In the Ammeter risk model it had been possible to find a lower bound by means of a comparison with a renewal risk model. The reason that this was possible was that the lengths of the intervals, in which the intensity was constant, was deterministic. This made the interarrival times and the claim amounts in the corresponding renewal risk model independent.

In order to be able to treat models where the intervals with constant intensity are stochastic we have to find another method. A possibility is always the following. Let $\{M_t\}$ be a strictly positive martingale. Suppose that $M_t \to 0$ on $\{\tau = \infty\}$ and that M_t is bounded on $\{\tau > t\}$. Then

$$M_0 = \mathbb{E}[M_{\tau \wedge t}] = \mathbb{E}[M_\tau; \tau \le t] + \mathbb{E}[M_t; \tau > t] \;.$$

Now monotone convergence yields $\mathbb{E}[M_\tau; \tau \le t] \to \mathbb{E}[M_\tau; \tau < \infty]$ and bounded convergence yields $\mathbb{E}[M_t; \tau > t] \to 0$ as $t \to \infty$. Thus

$$M_0 = \mathbb{E}[M_\tau; \tau < \infty] = \mathbb{E}[M_\tau \mid \tau < \infty] \mathbb{P}[\tau < \infty] \;.$$

This gives the formula

$$\mathbb{P}[\tau < \infty] = \frac{M_0}{\mathbb{E}[M_\tau \mid \tau < \infty]} \;.$$

The problem is, that in order to prove a Cramér-Lundberg approximation one needs that M_τ and $\{\tau < \infty\}$ are independent. This is ususally not an easy task to verify.

The method we will use here are ‘change of measure techniques’. We will see that the exponential martingales can be used to change the measure. Under the new measure the risk model will be of the same type, except that the net profit condition will not be fulfilled and ruin will always occur within finite time. Thus the ruin problem will become easier to handle.

We found the asymptotic behaviour of the ruin probability. But in the renewal as well as in the Ammeter case we were not able to give an explicit formula for the constant involved. And the results only are valid for ruin in infinite time. If we are interested in an exact figure for the (finite or infinite time) ruin probability or for the distribution of the severity of ruin then the only method available is simulation.

But the event of interest is rare. One will have to simulate a lot of paths of the process which will never lead to ruin. If dealing with the event of ruin in infinite time, then we will have to decide at which point we declare the path as "not leading to ruin" because we have to stop the simulation within finite time.

The change of measure technique will give us the opportunity to express the quantity of interest as a quantity under the new measure. But because under the new measure ruin occurs almost surely the problems discussed above vanish. Thus the change of measure technique provides a clever way to simulate.

8.1. The Cramér-Lundberg Case

The easiest model in order to illustrate the change of measure technique is the classical risk model. Let $\{C_t\}$ be a Cramér-Lundberg process and assume that the Lundberg exponent R exists and that there is an $r > R$ such that $M_Y(r) < \infty$. We have seen that for any r such that $M_Y(r) < \infty$ the process

$$L_t^r := \mathrm{e}^{-r(C_t-u)}\mathrm{e}^{-\theta(r)t}$$

is a martingale where $\theta(r) = \lambda(M_Y(r) - 1) - cr$. This martingale has the following properties

- $L_t^r > 0$ for all $t \geq 0$,
- $\mathbb{E}[L_t^r] = \mathbb{E}[L_0^r] = 1$ for all $t \geq 0$.

These properties are exactly the properties needed to define an equivalent measure on the σ-algebra $\mathcal{F}_t$. If $A \in \mathcal{F}_t$ then

$$Q_r[A] := \mathbb{E}_{\mathbb{P}}[L_t^r; A]$$

is a measure on $\mathcal{F}_t$ equivalent to $\mathbb{P}$. In fact it looks as though the measure also depends on t, but the next lemma shows that this is not the case. We suppose here that $\{\mathcal{F}_t\}$ is the smallest right continuous filtration such that $\{C_t\}$ is adapted. From iii) of the lemma below we can see that it is possible to extend the measures Q_r on $\mathcal{F}_t$ to the whole σ-algebra $\mathcal{F}$.

Lemma 8.1.

i) *Let* $t > s$ *and* $A \in \mathcal{F}_s \subset \mathcal{F}_t$. *Then* $\mathbb{E}_{\mathbb{P}}[L_t^r; A] = \mathbb{E}_{\mathbb{P}}[L_s^r; A]$.

ii) *Let T be a stopping time and $A \in \mathcal{F}_T$ such that $A \subset \{T < \infty\}$. Then*

$$Q_r[A] = \mathbb{E}_{\mathbb{P}}[L_T^r; A]\,.$$

iii) *$\{C_t\}$ is a Cramér-Lundberg process under the measure Q_r with claim size distribution $\tilde{G}(x) = \int_0^x \mathrm{e}^{ry}\, dG(y)/M_Y(r)$ and claim arrival intensity $\tilde{\lambda} = M_Y(r)\lambda$. In particular $\mathbb{E}_{Q_r}[C_1 - u] = -\theta'(r)$.*

iv) *If $r \neq 0$ then Q_r and P are singular on the whole σ-algebra $\mathcal{F}$.*

Proof. i) This is an easy consequence of the martingale property

$$\mathbb{E}_{\mathbb{P}}[L_t^r; A] = \mathbb{E}_{\mathbb{P}}[\mathbb{E}_{\mathbb{P}}[L_t^r; A \mid \mathcal{F}_s]] = \mathbb{E}_{\mathbb{P}}[\mathbb{E}_{\mathbb{P}}[L_t^r \mid \mathcal{F}_s]; A] = \mathbb{E}_{\mathbb{P}}[L_s^r; A]\,.$$

ii) By iii) and Kolmogorov's extension theorem Q_r can be extended uniquely to a measure on $\mathcal{F}$. We have by the stopping theorem and the monotone convergence theorem

$$\begin{aligned} Q_r[A] &= \lim_{t\to\infty} Q_r[A \cap \{T \le t\}] = \lim_{t\to\infty} \mathbb{E}_{\mathbb{P}}[L_t^r; A \cap \{T \le t\}] \\ &= \lim_{t\to\infty} \mathbb{E}_{\mathbb{P}}[\mathbb{E}_{\mathbb{P}}[L_t^r \mid \mathcal{F}_T]; A \cap \{T \le t\}] = \lim_{t\to\infty} \mathbb{E}_{\mathbb{P}}[L_T^r; A \cap \{T \le t\}] \\ &= \mathbb{E}_{\mathbb{P}}[L_T^r; A]\,. \end{aligned}$$

iii) For all $n \ge 1$ we have to find the joint distribution of $T_1, T_2 - T_1, \ldots, T_n - T_{n-1}$ and $Y_1, Y_2, \ldots, Y_n$. Note that these random variables are independent under $\mathbb{P}$. Let $t \ge 0$, A be a Borel set of $\mathbb{R}_+^n$ and $B_1, B_2, \ldots, B_n$ be Borel sets on $\mathbb{R}_+$. Then

$$\begin{aligned} &Q_r[N_t = n, (T_1, T_2, \ldots, T_n) \in A, Y_1 \in B_1, Y_2 \in B_2, \ldots, Y_n \in B_n] \\ &\quad = \mathbb{E}_{\mathbb{P}}[L_t^r; N_t = n, (T_1, T_2, \ldots, T_n) \in A, Y_1 \in B_1, Y_2 \in B_2, \ldots, Y_n \in B_n] \\ &\quad = \mathbb{E}_{\mathbb{P}}\Bigl[\exp\Bigl\{r \sum_{i=1}^n Y_i\Bigr\} \exp\{-(\theta(r) + cr)t\}; N_t = n, (T_1, \ldots, T_n) \in A, \\ &\qquad\qquad\qquad\qquad\qquad\qquad Y_1 \in B_1, Y_2 \in B_2, \ldots, Y_n \in B_n\Bigr] \\ &\quad = \frac{(\lambda t)^n}{n!} \mathrm{e}^{-\lambda t - (\theta(r) + cr)t} \mathbb{P}[(T_1, T_2, \ldots, T_n) \in A \mid N_t = n] \prod_{i=1}^n \mathbb{E}_{\mathbb{P}}[\mathrm{e}^{rY_i}; Y_i \in B_i] \\ &\quad = \frac{(\lambda M_Y(r) t)^n}{n!} \mathrm{e}^{-\lambda M_Y(r) t} \mathbb{P}[(T_1, T_2, \ldots, T_n) \in A \mid N_t = n] \prod_{i=1}^n \frac{\mathbb{E}_{\mathbb{P}}[\mathrm{e}^{rY_i}; Y_i \in B_i]}{M_Y(r)}\,. \end{aligned}$$

Thus the number of claims is Poisson distributed with parameter $\lambda M_Y(r)$. Given $\{N_t\}$, the claim occurrence times do not change the conditional law. By Proposition C.2 the process $\{N_t\}$ is a Poisson process with rate $\lambda M_Y(r)$. The claim sizes are independent and have the claimed distribution. In particular

$$\mathbb{E}_{Q_r}[Y_i^n] = \frac{\int_0^\infty y^n \mathrm{e}^{ry} \,\mathrm{d}G(y)}{M_Y(r)} = \frac{M_Y^{(n)}(r)}{M_Y(r)}$$

from which iii) follows.

iv) This follows because

$$Q_r\Bigl[\lim_{t\to\infty} \frac{C_t}{t} = -\theta'(r)\Bigr] = 1$$

and because $\theta'(r)$ is strictly increasing. □

We next express the ruin probability using the measure Q_r.

$$\psi(u) = \mathbb{P}[\tau < \infty] = \mathbb{E}_{Q_r}[(L_\tau^r)^{-1}; \tau < \infty] = \mathbb{E}_{Q_r}[\mathrm{e}^{rC_\tau}\mathrm{e}^{\theta(r)\tau}; \tau < \infty]\mathrm{e}^{-ru}\,.$$

The latter is much more complicated because the joint distribution of τ and C_τ has to be known unless $\theta(r) = 0$. Let us therefore choose $r = R$ and $Q = Q_R$. First observe that $\mathbb{E}_Q[C_1 - u] = -\theta'(R) < 0$. Thus the net profit condition is not fulfilled and $\tau < \infty$ a.s. under the measure Q. Then

$$\psi(u) = \mathbb{E}_Q[\mathrm{e}^{RC_\tau}; \tau < \infty]\mathrm{e}^{-Ru} = \mathbb{E}_Q[\mathrm{e}^{RC_\tau}]\mathrm{e}^{-Ru}\,.$$

Lundberg's inequality (Theorem 4.4) becomes now trivial. Because $C_\tau < 0$ one has $\psi(u) < \mathrm{e}^{-Ru}$.

In order to prove the Cramér-Lundberg approximation we have to show that $\mathbb{E}_Q[\mathrm{e}^{RC_\tau}]$ converges as $u \to \infty$. Denote by $B(x) = Q[C_\tau \ge -x \mid C_0 = 0]$ the descending ladder height distribution under Q. Note that (4.10) does not apply because the limit of $\psi_x(u)$ is not 0 anymore as $u \to \infty$. Let $Z(u) = \mathbb{E}_Q[\mathrm{e}^{RC_\tau} \mid C_0 = u]$. Then $Z(u)$ fulfils the renewal equation

$$Z(u) = \int_0^u Z(u-y)\,\mathrm{d}B(y) + \int_u^\infty \mathrm{e}^{-R(y-u)}\,\mathrm{d}B(y)\,.$$

It is not difficult to show that

$$\int_u^\infty \mathrm{e}^{-R(y-u)}\,\mathrm{d}B(y)$$

is directly Riemann integrable and its integral is

$$\frac{1}{R}\Bigl(1 - \int_0^\infty \mathrm{e}^{-Ry}\,\mathrm{d}B(y)\Bigr)\,.$$

Thus by the renewal theorem the limit of $Z(u)$ exists and is equal to

$$\frac{1}{R\int_0^\infty u\,\mathrm{d}B(u)}\Bigl(1-\int_0^\infty \mathrm{e}^{-Ry}\,\mathrm{d}B(y)\Bigr)\,.$$

Consider first $\int_0^\infty u\,\mathrm{d}B(u)$. We find

$$\begin{aligned}\int_0^\infty u\,\mathrm{d}B(u) &= -\mathbb{E}_Q[C_\tau \mid C_0=0] = -\mathbb{E}_{\mathbb{P}}[C_\tau \mathrm{e}^{-RC_\tau};\tau<\infty \mid C_0=0]\\ &= \frac{\lambda}{c}\int_0^\infty x\mathrm{e}^{Rx}(1-G(x))\,\mathrm{d}x = \frac{\lambda}{R^2c}(RM_Y'(R)-(M_Y(R)-1))\,.\end{aligned}$$

The integral in the numerator of the limit is

$$\int_0^\infty \mathrm{e}^{-Ry}\,\mathrm{d}B(y) = \mathbb{E}_Q[\mathrm{e}^{RC_\tau} \mid C_0=0] = \mathbb{P}[\tau<\infty \mid C_0=0] = \frac{\lambda\mu}{c}\,.$$

Thus

$$\begin{aligned}\lim_{u\to\infty} Z(u) &= \frac{\frac{1}{R}\bigl(1-\frac{\lambda\mu}{c}\bigr)}{\frac{\lambda}{R^2c}(RM_Y'(R)-(M_Y(R)-1))}\\ &= \frac{c-\lambda\mu}{\lambda(M_Y'(R)-(M_Y(R)-1)/R)} = \frac{c-\lambda\mu}{\lambda M_Y'(R)-c}\end{aligned} \tag{8.1}$$

where we used the definition of the Lundberg exponent R. This proves the Cramér-Lundberg approximation (Theorem 4.6).

8.2. The Renewal Case

In Section 5 we only considered the process at the claim times. But for the change of the measure we need a continuous time martingale. As mentioned in Section 5 the process $\{C_t\}$ is not a Markov process. For our approach it is important to have the Markov property. Thus we need to Markovize the process. The future of the process is dependent on the time of the last claim. Let $W_t = t - T_{N_t}$ denote the time elapsed since the last claim. Then $\{(C_t, W_t)\}$ is a Markov process. Alternatively we can consider $V_t = T_{N_t+1} - t$ the time left till the next claim. This seems a little bit strange because $\{V_t\}$ is not observable. But recall that in the Ammeter model we worked with a non-observable stochastic process which was useful in order to obtain Lundberg's inequality in the general case. The stochastic process $\{(C_t, V_t)\}$ is also a Markov process. But note that the natural filtrations of $\{C_t\}$ and of $\{(C_t, V_t)\}$ are different.

8.2.1. Markovization Via the Time Since the Last Claim

In order to construct a martingale we have to assume that the interarrival time distribution F is absolutely continuous with density $F'(t)$. The following martingale follows from the theory of piecewise deterministic Markov processes, see [66].

Lemma 8.2. *Assume that $M_Y(r) < \infty$ such that the unique solution $\theta(r)$ to*

$$M_Y(r)M_T(-\theta - cr) = 1$$

exists (such a solution exists if $r \geq 0$). Then the process

$$L_t^r = \mathrm{e}^{-r(C_t-u)}\, M_Y(r)\frac{\mathrm{e}^{(\theta(r)+cr)W_t}}{1-F(W_t)}\int_{W_t}^{\infty} \mathrm{e}^{-(\theta(r)+cr)s} F'(s)\;\mathrm{d}s\, \mathrm{e}^{-\theta(r)t}$$

is a martingale.

Remark. Considering the above martingale at the times $0, T_1, \ldots$ only yields the martingale obtained in Lemma 5.1. ■

Proof. See Dassios and Embrechts [28, Thm.10]. □

Define now the measures for $A \in \mathcal{F}_t$

$$Q_r[A] = \mathbb{E}_{\mathbb{P}}[L_t^r; A]\,.$$

Lemma 8.1 remains true also in the renewal case. Also here, we suppose that $\{\mathcal{F}_t\}$ is the smallest right continuous filtration such that $\{(C_t, W_t)\}$ is adapted. Also here it turns out that the measure Q_r can be extended to the whole σ-algebra $\mathcal{F}$.

Lemma 8.3.

i) *Let $t > s$ and $A \in \mathcal{F}_s \subset \mathcal{F}_t$. Then $\mathbb{E}_{\mathbb{P}}[L_t^r; A] = \mathbb{E}_{\mathbb{P}}[L_s^r; A]$.*

ii) *Let T be a stopping time and $A \in \mathcal{F}_T$ such that $A \subset \{T < \infty\}$. Then*

$$Q_r[A] = \mathbb{E}_{\mathbb{P}}[L_T^r; A]\,.$$

iii) *$\{(C_t, W_t)\}$ is a renewal risk model under the measure Q_r with claim size distribution $\tilde{G}(x) = M_T(-\theta(r) - cr)\int_0^x \mathrm{e}^{ry}\;\mathrm{d}G(y)$ and interarrival distribution $\tilde{F}(t) = M_Y(r)\int_0^t \mathrm{e}^{-(\theta(r)+cr)s}\;\mathrm{d}F(s)$. In particular*

$$\lim_{t\to\infty} \frac{1}{t}(C_t - u) = -\theta'(r)$$

Q_r-a.s..

iv) *If* $r \neq 0$ *then* Q_r *and* P *are singular on the whole* σ*-algebra* $\mathcal{F}$*.*

Proof. i), ii) and iv) are proved as in Lemma 8.1.

iii) Let $n \in \mathbb{N}$ and $A_1, A_2, \ldots, A_n, B_1, B_2, \ldots, B_n$ be Borel sets. As long as we did not prove that the measure Q_r can be extended to $\mathcal{F}$ we have to choose $A_1, \ldots, A_n$ such that T_n remains finite. Then because $W_{T_n} = 0$

$$\begin{aligned}
&Q_r[T_1 \in A_1, T_2 - T_1 \in A_2, \ldots, T_n - T_{n-1} \in A_n, Y_1 \in B_1, Y_2 \in B_2, \ldots, Y_n \in B_n] \\
&\quad = \mathbb{E}_{\mathbb{P}}[L^r_{T_n}; T_1 \in A_1, T_2 - T_1 \in A_2, \ldots, T_n - T_{n-1} \in A_n, \\
&\qquad\qquad Y_1 \in B_1, Y_2 \in B_2, \ldots, Y_n \in B_n] \\
&\quad = \mathbb{E}_{\mathbb{P}}\Bigl[\exp\Bigl\{r\sum_{i=1}^n Y_i\Bigr\}\exp\Bigl\{-(\theta(r)+cr)\sum_{i=1}^n (T_i - T_{i-1})\Bigr\}; T_1 \in A_1, \\
&\qquad\qquad T_2 - T_1 \in A_2, \ldots, T_n - T_{n-1} \in A_n, Y_1 \in B_1, Y_2 \in B_2, \ldots, Y_n \in B_n\Bigr] \\
&\quad = \prod_{i=1}^n \mathbb{E}_{\mathbb{P}}[\mathrm{e}^{rY_i}; Y_i \in B_i]\mathbb{E}_{\mathbb{P}}[\mathrm{e}^{-(\theta(r)+cr)(T_i - T_{i-1})}; (T_i - T_{i-1}) \in A_i]\,.
\end{aligned}$$

Thus $\{(C_t, W_t)\}$ is the renewal risk model with the parameters given by the assertion. A straightforward calculation yields

$$\lim_{t\to\infty} \frac{1}{t}(C_t - u) = c - \tilde{\lambda}\tilde{\mu} = -\theta'(r)$$

Q_r-a.s.. □

The ruin probability can be expressed as

$$\begin{aligned}
\psi(u) &= \mathbb{E}_{Q_r}\Bigl[\mathrm{e}^{rC_\tau}\frac{(1 - F(W_\tau))\mathrm{e}^{-(\theta(r)+cr)W_\tau}}{M_Y(r)\int_{W_\tau}^\infty \mathrm{e}^{-(\theta(r)+cr)s}F'(s)\,\mathrm{d}s}\,\mathrm{e}^{\theta(r)\tau}; \tau < \infty\Bigr]\mathrm{e}^{-ru} \\
&= \mathbb{E}_{Q_r}[\mathrm{e}^{rC_\tau}\mathrm{e}^{\theta(r)\tau}; \tau < \infty]\mathrm{e}^{-ru}
\end{aligned}$$

where we used that $W_\tau = 0$. As before we have to choose $r = R$ which we assume to exist. Then, because $\theta'(R) > 0$, the net profit condition is violated and

$$\psi(u) = \mathbb{E}_Q[\mathrm{e}^{RC_\tau}]\mathrm{e}^{-Ru}\,.$$

Lundberg's inequality (Theorem 5.5) follows immediately and the Cramér-Lundberg approximation (Theorem 5.7) can be proved in the same way as in the classical case.

8.2.2. Markovization Via the Time Till the Next Claim

Choosing the Markovization $\{(C_t, V_t)\}$ simplifies the situation. The martingale looks nicer and we do not have to assume that F is absolutely continuous. Let t_0 be fixed. Then the process is deterministic in the interval $(t_0, t_0 + V_{t_0})$, in fact $(C_t, V_t) = (C_{t_0} + c(t - t_0), V_{t_0} - (t - t_0))$. Because no event can happen the martingale must be constant in the interval $(t_0, t_0 + V_{t_0})$.

Lemma 8.4. *Assume that* $M_Y(r) < \infty$ *such that the unique solution* $\theta(r)$ *to*

$$M_Y(r)M_T(-\theta - cr) = 1$$

exists (such a solution exists if $r \ge 0$*). Then the process*

$$L_t^r = M_Y(r)\mathrm{e}^{-r(C_t-u)}\mathrm{e}^{-(\theta(r)+cr)V_t}\mathrm{e}^{-\theta(r)t}$$

is a martingale.

Proof. See [35, p.27]. □

Note that $\mathbb{E}[L_{T_i}^r \mid \mathcal{F}_{T_i}^C]$ with $\{\mathcal{F}_t^C\}$ denoting the natural filtration of the risk process $\{C_t\}$ is the martingale considered in Section 5. Define as before $Q_r[A] = \mathbb{E}_{\mathbb{P}}[L_t^r; A]$ for $A \in \mathcal{F}_t$. From the lemma below it again follows that the measure can be extended to a measure on the whole σ-algebra $\mathcal{F}$. We again have

Lemma 8.5.

i) *Let* $t > s$ *and* $A \in \mathcal{F}_s \subset \mathcal{F}_t$. *Then* $\mathbb{E}_{\mathbb{P}}[L_t^r; A] = \mathbb{E}_{\mathbb{P}}[L_s^r; A]$.

ii) *Let* T *be a stopping time and* $A \in \mathcal{F}_T$ *such that* $A \subset \{T < \infty\}$. *Then*

$$Q_r[A] = \mathbb{E}_{\mathbb{P}}[L_T^r; A]\,.$$

iii) $\{(C_t, V_t)\}$ *is a renewal risk model under the measure* Q_r *with claim size distribution* $\tilde{G}(x) = M_T(-\theta(r) - cr)\int_0^x \mathrm{e}^{ry}\,\mathrm{d}G(y)$ *and interarrival distribution* $\tilde{F}(t) = M_Y(r)\int_0^t \mathrm{e}^{-(\theta(r)+cr)s}\,\mathrm{d}F(s)$. *In particular*

$$\lim_{t\to\infty} \frac{1}{t}(C_t - u) = -\theta'(r)$$

Q_r*-a.s..*

iv) *If* $r \ne 0$ *then* Q_r *and* P *are singular on the whole* σ*-algebra* $\mathcal{F}$.

Proof. Left as an exercise. □

For the ruin probability we obtain

$$\begin{aligned}\psi(u) &= (M_Y(r))^{-1}\mathbb{E}_{Q_r}[\mathrm{e}^{rC_\tau}\mathrm{e}^{(\theta(r)+cr)V_\tau}\mathrm{e}^{\theta(r)\tau};\tau<\infty]\mathrm{e}^{-ru}\\ &= \mathbb{E}_{Q_r}[\mathrm{e}^{rC_\tau}\mathrm{e}^{\theta(r)\tau};\tau<\infty]\mathrm{e}^{-ru}\,.\end{aligned}$$

Note that V_τ is independent of $\mathcal{F}_{\tau-}$ and has the same distribution as $T_{n+1}-T_n$. Thus we again have to choose $r=R$. Lundberg's inequality (Theorem 5.5) and the Cramér-Lundberg approximation (Theorem 5.7) follow as before.

Another advantage of the second approach is that we can treat the general case, where T_1 has distribution F^1. This causes problems in the first approach, because it is difficult to find the distribution of W_0 such that T_1 has the right distribution. Such a distribution does not even exist in all cases.

Change the martingale to

$$L_t^r = \frac{\mathrm{e}^{-r(C_t-u)}\mathrm{e}^{-(\theta(r)+cr)V_t}\mathrm{e}^{-\theta(r)t}}{\mathbb{E}_{\mathbb{P}}[\mathrm{e}^{-(\theta(r)+cr)V_0}]}\,.$$

Note that this includes the ordinary case. Then

$$\psi(u) = M_Y(r)\mathbb{E}_{\mathbb{P}}[\mathrm{e}^{-(\theta(r)+cr)V_0}]\,\mathbb{E}_{Q_r}[\mathrm{e}^{rC_\tau}\mathrm{e}^{\theta(r)\tau};\tau<\infty]\mathrm{e}^{-ru}\,.$$

Lundberg's inequality (Theorem 5.6) is trivial again. For the Cramér-Lundberg approximation we have to show that $\mathbb{E}_Q[\mathrm{e}^{RC_\tau}]$ converges as $u\to\infty$. Letting $Z(u)=\mathbb{E}_Q[\mathrm{e}^{RC_\tau}\mid C_0=u]$ and $Z^{\mathrm{o}}(u)=\mathbb{E}_Q^{\mathrm{o}}[\mathrm{e}^{RC_\tau}\mid C_0=u]$ be the corresponding quantity in the ordinary case. Then

$$Z(u)=\int_0^\infty\Bigl(\int_0^{u+ct}Z^{\mathrm{o}}(u+ct-y)\,\mathrm{d}\tilde G(y)+\int_{u+ct}^\infty \mathrm{e}^{-R(y-u-ct)}\,\mathrm{d}\tilde G(y)\Bigr)\,\mathrm{d}\tilde F^1(t)\,.$$

Because $Z^{\mathrm{o}}(u)$ is bounded by 1 the Cramér-Lundberg approximation (5.7) follows from the bounded convergence theorem.

8.3. The Ammeter Risk Model

For the Ammeter risk model we already found the martingale (Lemma 7.7)

$$L_t^r = \mathrm{e}^{-r(C_t-u)}\mathrm{e}^{(\lambda_t(M_Y(r)-1)-cr-\theta(r))V_t}\mathrm{e}^{-\theta(r)t}$$

where V_t is the time remaining till the next change of the intensity. With this martingale we define the new measure $Q_r[A]=\mathbb{E}_{\mathbb{P}}[L_t^r;A]$ for $A\in\mathcal{F}_t$.

In the classical case we obtained the new intensity $\tilde{\lambda} = \lambda M_Y(r)$. Thus it might be that under the new measure the process $\{\lambda_t\}$ is not anymore the intensity process. Let us therefore denote the intensity under the measure Q_r by $\tilde{\lambda}_t$ and the intensity levels by $\tilde{L}_i$. From the lemma below it follows again that Q_r can be extended to the whole σ-algebra $\mathcal{F}_t$.

Lemma 8.6.

i) *Let* $t > s$ *and* $A \in \mathcal{F}_s \subset \mathcal{F}_t$. *Then* $\mathbb{E}_{\mathbb{P}}[L_t^r; A] = \mathbb{E}_{\mathbb{P}}[L_s^r; A]$.

ii) *Let* T *be a stopping time and* $A \in \mathcal{F}_T$ *such that* $A \subset \{T < \infty\}$. *Then*

$$Q_r[A] = \mathbb{E}_{\mathbb{P}}[L_T^r; A]\,.$$

iii) $\{(C_t, \tilde{\lambda}_t, V_t)\}$ *is an Ammeter risk model under the measure* Q_r *with claim size distribution* $\tilde{G}(x) = \int_0^x \mathrm{e}^{ry}\, \mathrm{d}G(y)/M_Y(r)$, *claim arrival intensity* $\tilde{\lambda}_t = M_Y(r)\lambda_t$, $\tilde{L}_i = M_Y(r)L_i$ *and* $\tilde{H}(\ell) = \mathrm{e}^{-(cr+\theta(r))\Delta} \int_0^\ell \mathrm{e}^{l\Delta(M_Y(r)-1)}\, \mathrm{d}H(l)$. *In particular,* $\mathbb{E}_{Q_r}[C_\Delta - u] = -\theta'(r)\Delta$.

iv) *If* $r \neq 0$ *then* Q_r *and* P *are singular on the whole* σ*-algebra* $\mathcal{F}$.

Proof. i), ii) and iv) are proved as in Lemma 8.1.

iii) We first want to show that the processes $\{(C_t - C_{(i-1)\Delta} : (i-1)\Delta \le t < i\Delta)\}$ are independent and follow the same law. Let A_i be a set measurable with respect to $\sigma(L_i, N_{i\Delta} - N_{(i-1)\Delta}, T_{N_{(i-1)\Delta}+1}, \ldots, T_{N_{i\Delta}}, Y_{N_{(i-1)\Delta}+1}, \ldots, Y_{N_{i\Delta}})$. Then for $n \in \mathbb{N}$

$$\begin{aligned}
Q_r[A_1, \ldots, A_n] &= \mathbb{E}_{\mathbb{P}}[\mathrm{e}^{-r(C_{n\Delta}-u)}\mathrm{e}^{(L_{n+1}(M_Y(r)-1)-cr-\theta(r))\Delta}\mathrm{e}^{-\theta(r)n\Delta}; A_1, \ldots, A_n] \\
&= \mathbb{E}_{\mathbb{P}}[\mathrm{e}^{-r(C_{n\Delta}-u)}\mathrm{e}^{-\theta(r)n\Delta}; A_1, \ldots, A_n] \\
&= \prod_{i=1}^n \mathrm{e}^{-\theta(r)\Delta}\mathbb{E}_{\mathbb{P}}[\mathrm{e}^{-r(C_{i\Delta}-C_{(i-1)\Delta})}; A_i]\,.
\end{aligned}$$

Next we show that the claim sizes are iid., independent of $\{N_t\}$ and have the required distribution. Let $n \in \mathbb{N}$ and A be a $\sigma(L_1, T_1, \ldots, T_n)$-measurable set. Let B_i be Borel sets. Then

$$\begin{aligned}
&Q_r[N_\Delta = n, A, Y_1 \in B_1, \ldots, Y_n \in B_n] \\
&\qquad = \mathrm{e}^{-\theta(r)\Delta}\mathbb{E}_{\mathbb{P}}[\mathrm{e}^{-r(C_\Delta-u)}; N_\Delta = n, A, Y_1 \in B_1, \ldots, Y_n \in B_n] \\
&\qquad = \mathrm{e}^{-(cr+\theta(r))\Delta}\mathbb{P}[N_\Delta = n, A]\prod_{i=1}^n \mathbb{E}_{\mathbb{P}}[\mathrm{e}^{rY_i}; Y_i \in B_i]\,.
\end{aligned}$$

Next we verify the distribution of L_1 and N_Δ. Let A be Borel measurable. Then

$$\begin{aligned}Q_r[L_1 \in A, N_\Delta = n] &= \mathrm{e}^{-(cr+\theta(r))\Delta} M_Y(r)^n \int_A \frac{(\ell\Delta)^n}{n!} \mathrm{e}^{-\ell\Delta}\,\mathrm{d}H(\ell)\\ &= \int_A \frac{(M_Y(r)\ell\Delta)^n}{n!} \mathrm{e}^{-M_Y(r)\ell\Delta} \mathrm{e}^{-(cr+\theta(r))\Delta} \mathrm{e}^{\ell\Delta(M_Y(r)-1)}\,\mathrm{d}H(\ell)\,.\end{aligned}$$

Thus L_1 and N_Δ have the desired distribution. As a last thing we have to show that given $N_\Delta = n$ then times $(T_1, \dots, T_n)$ have the same distribution as the order statistics of n iid. uniformly in $(0, \Delta)$ distributed random variables (Proposition C.2), i.e. the same conditional distribution as under $\mathbb{P}$. Let $n \in \mathbb{N}$, A be a $\sigma(T_1, \dots, T_n)$-measurable set and B be a Borel set such that $\mathbb{P}[L_1 \in B] \neq 0$. Then

$$\begin{aligned}Q_r[A \mid L_1 \in B, N_\Delta = n] &= \frac{\mathrm{e}^{-(cr+\theta(r))\Delta}\mathbb{P}[A, L_1 \in B, N_\Delta = n] M_Y(r)^n}{\mathrm{e}^{-(cr+\theta(r))\Delta}\mathbb{P}[L_1 \in B, N_\Delta = n] M_Y(r)^n}\\ &= \mathbb{P}[A \mid L_1 \in B, N_\Delta = n] = \mathbb{P}[A \mid N_\Delta = n]\,.\end{aligned}$$

The last assertion follows straightforwardly noting that

$$\mathbb{E}_{Q_r}[C_\Delta - u] = (c - \mathbb{E}_{Q_r}[M_Y(r)L_i]\mathbb{E}_{Q_r}[Y_i])\Delta\,.$$

□

We consider now the ordinary case $F^1(\ell, v) = H(\ell)\mathbb{1}_{v \ge \Delta}$. As before it turns out to be convenient to choose $r = R$ and that then the net profit condition is violated. For the ruin probability we obtain

$$\psi(u) = \mathbb{E}_Q[\mathrm{e}^{RC_\tau} \mathrm{e}^{-(\lambda_\tau(M_Y(R)-1)-cR)V_\tau}]\,\mathrm{e}^{-Ru}\,.$$

The Lundberg inequality follows immediately noting that $\lambda_\tau > 0$ and therefore $\mathrm{e}^{-(\lambda_\tau(M_Y(R)-1)-cR)V_\tau} < \mathrm{e}^{cR\Delta}$. For the Cramér-Lundberg approximation we have a problem. How can we find the asymptotic distribution of $(C_\tau, \lambda_\tau, V_\tau)$ as $u \to \infty$? In order to use the approach which was successful in the Cramér-Lundberg and in the renewal case this distribution has to be known. Anyway, with a slightly modified renewal approach it is possible to prove the Cramér-Lundberg approximation, see [71] for details.

In order to find a lower bound for the ruin probability define $\tau_1 := \inf\{k\Delta : k \in \mathbb{N}, C_{k\Delta} < 0\}$ the first time when the process becomes negative at a regeneration time. We assume now that there exists an $r > R$ such that $M_L(\Delta(M_Y(r) - 1)) < \infty$. Obviously $\mathbb{P}[\tau < \infty] \ge \mathbb{P}[\tau_1 < \infty]$ and

$$\begin{aligned}\mathbb{P}[\tau_1 < \infty] &= \mathbb{E}_Q[\mathrm{e}^{RC_{\tau_1}} \mathrm{e}^{-(\lambda_{\tau_1}(M_Y(R)-1)-cR)V_{\tau_1}}]\mathrm{e}^{-Ru}\\ &= \mathbb{E}_Q[\mathrm{e}^{RC_{\tau_1}}]\mathrm{e}^{-Ru} \ge \mathrm{e}^{R\mathbb{E}_Q[C_{\tau_1}]}\mathrm{e}^{-Ru}\,.\end{aligned}$$

Because $\mathbb{E}_Q[\mathrm{e}^{-(r-R)(C_t-u)}]$ exists all moments of the increments of the random walk $C_{k\Delta} - u$ exist. Using the renewal theorem one can show that $\mathbb{E}_Q[C_{\tau_1}]$ converges to a finite value as $u \to \infty$.

The case with a general $F^1(\ell, v)$ is left to the reader as an exercise.

Bibliographical Remarks

The change of measure method is frequently used in the literature without mentioning the connection to martingales, see for instance [6], [7] or [8]. The technique used in this section can also be found in [9] or [66]. Lemma 8.4 was used in an early version of [27], but does not appear in the final version.

9. The Markov Modulated Risk Model

In the Ammeter model the claim arrival intensity was constant during the interval $((k-1)\Delta, k\Delta)$. For instance Δ was one year. It would be preferable, if the intensity was allowed to change more frequently. This can be achieved by choosing Δ smaller. But it would also be preferable that different intensities levels could be holding for time intervals of different lengths. A model fulfilling the above requirements is the Markov modulated risk model. One first constructs a continuous time Markov chain on a finite state space. This Markov chain represents an environment process for the risk business. The environment determines the intensity level. Moreover, also the claim size distribution can differ for different states of the environment.

9.1. Definition of the Markov Modulated Risk Model

Let $\{J_t\}$ be a continuous time Markov chain with state space $(1, 2, \ldots, \mathcal{J})$ and **intensity matrix** $\boldsymbol{\eta} = (\eta_{ij})$, where $\mathcal{J} \in \mathbb{N} \setminus \{0\}$ and $\eta_{ij} \ge 0$ for $i \ne j$ and $\eta_{ii} = -\sum_{j \ne i} \eta_{ij}$. This means that, if $J_t = i$, then $J_{t+s} = i$ for a $\mathrm{Exp}(-\eta_{ii})$ distributed time and the next state will be $j \ne i$ with probability $\eta_{ij}/(-\eta_{ii})$. In order to obtain an ergodic model we assume that $\eta_{ii} < 0$ for all $i \le \mathcal{J}$. Let $\{L_i : i \le \mathcal{J}\}$ denote the intensity levels and let $\{G_i(y) : i \le \mathcal{J}\}$ be distribution functions of positive random variables. L_i will be the claim arrival intensity and $G_i(y)$ will be the claim size distribution during the time where the environment process is in state i. For a formal definition let $\Sigma_0 = 0$ and for $n \in \mathbb{N}$

$$\Sigma_{n+1} = \inf\{t > \Sigma_n : J_t \ne J_{t-}\}$$

be times at which the environment changes. Let $\{C_t^{(i)}\}$ be independent Cramér-Lundberg models with initial capital 0 and premium rate c where the i-th model has claim arrival intensity L_i and claim size distribution $G_i(y)$. The claim number process of the i-th model is denoted by $\{N_t^{(i)}\}$. Let $C_{0-}^{(i)} = 0$, $N_{0-}^{(i)} = 0$, $N_{0-} = 0$ and $C_{0-} = u$. Define recursively

$$N_t := N_{\Sigma_n -} + N_t^{(J_t)} - N_{\Sigma_n -}^{(J_t)}, \qquad \Sigma_n \le t < \Sigma_{n+1}$$

and

$$C_t := C_{\Sigma_n -} + C_t^{(J_t)} - C_{\Sigma_n -}^{(J_t)}, \qquad \Sigma_n \le t < \Sigma_{n+1}\,.$$

$\{N_t\}$ is the number of claims of $\{C_t\}$ in the interval $(0, t]$. Moreover, $\{N_t\}$ is as Cox point process with intensity $\lambda_t = L_{J_t}$, i.e. given $\{J_t : t \ge 0\}$, $\{N_t\}$ is an inhomogeneous Poisson process with rate $\{\lambda_t\}$ and $\mathbb{E}[N_t \mid J_s, 0 \le s \le t] = \int_0^t \lambda_s \,\mathrm{d}s$. A claim occurring at time t has distribution function $G_{J_t}(y)$.

For later use we will need some further notation. Let $\mu_i = \int_0^\infty (1 - G_i(y))\,\mathrm{d}y$ be the mean value and $M_Y^{(i)}(r) = \int_0^\infty \mathrm{e}^{ry}\,\mathrm{d}G_i(y)$ be the moment generating function of a claim in state i. By $M_Y(r) = \max\{M_Y^{(i)}(r) : i \le \mathcal{J}\}$ we denote the maximal value of the moment generating functions. Denote by $\boldsymbol{I}$ the identity matrix. For r such that $M_Y(r) < \infty$ let $\boldsymbol{S}(r)$ be the diagonal matrix with $S_{ii}(r) = L_i(M_Y^{(i)}(r) - 1)$ and $\boldsymbol{\Theta}(r) := \boldsymbol{\eta} + \boldsymbol{S}(r) - cr\boldsymbol{I}$.

Definition 9.1. *Let $\boldsymbol{A}$ be a square matrix. Then we denote by $\mathrm{e}^{t\boldsymbol{A}}$ the matrix*

$$\mathrm{e}^{t\boldsymbol{A}} := \sum_{n=0}^{\infty} \frac{t^n}{n!} \boldsymbol{A}^n \,.$$

Lemma 9.2. *For the Markov modulated risk model $\{C_t\}$ one has*

$$\mathbb{E}[\mathrm{e}^{-r(C_t-u)} 1\!\!1_{\{J_t=j\}} \mid J_0 = i] = (\mathrm{e}^{t\boldsymbol{\Theta}(r)})_{ij}$$

provided $M_Y(r) < \infty$. In particular $\mathbb{P}[J_t = j \mid J_0 = i] = (\mathrm{e}^{t\boldsymbol{\eta}})_{ij}$.

Proof. Let $f_{ij}(t) = \mathbb{E}[\mathrm{e}^{-r(C_t-u)} 1\!\!1_{\{J_t=j\}} \mid J_0 = i]$. For a small time interval we have

$$\mathbb{P}[N_{t+h} - N_t = 0 \mid J_t = k] = 1 - L_k h + o(h), \quad \mathbb{P}[N_{t+h} - N_t = 1 \mid J_t = k] = L_k h + o(h)$$

and

$$\mathbb{P}[J_{t+h} = l \mid J_t = k] = I_{kl} + \eta_{kl} h + o(h)\,.$$

Therefore, for h small,

$$\begin{aligned} f_{ij}(t+h) = f_{ij}(t)(1 + \eta_{jj}h + o(h))\Big((1 - L_j h + o(h))\mathrm{e}^{-rch} \\ + (L_j h + o(h))\mathrm{e}^{-rch} M_Y^{(j)}(r)\Big) + \sum_{\substack{k=1\\k\ne j}}^{\mathcal{J}} f_{ik}(t)(\eta_{kj}h + o(h))\mathrm{e}^{-rch} + o(h) \\ = f_{ij}(t)\mathrm{e}^{-rch} + f_{ij}(t)L_j h(M_Y^{(j)}(r) - 1) + \sum_{k=1}^{\mathcal{J}} f_{ik}(t)\eta_{kj}h + o(h)\,. \end{aligned}$$

Thus $f_{ij}(t)$ fulfils the differential equation

$$f'_{ij}(t) = -crf_{ij}(t) + f_{ij}(t)L_j(M_Y^{(j)}(r) - 1) + \sum_{k=1}^{\mathcal{J}} f_{ik}(t)\eta_{kj} = \sum_{k=1}^{\mathcal{J}} f_{ik}(t)\Theta_{kj}(r)$$

with initial condition $f_{ij}(0) = I_{ij}$. The only solution to the above system of differential equations is $f_{ij}(t) = (\mathrm{e}^{t\boldsymbol{\Theta}(r)})_{ij}$. □

Of special interest is the case where J_0 is distributed such that $\{J_t\}$ becomes stationary. The initial distribution $\boldsymbol{\pi}$ has to fulfil

$$\boldsymbol{\pi}\mathrm{e}^{t\boldsymbol{\eta}} = \boldsymbol{\pi}$$

where we write $\boldsymbol{\pi}$ as a row vector. The condition is equivalent to $\boldsymbol{\pi\eta} = \mathbf{0}$. We assume that the Markov chain is irreducible, which means that $\boldsymbol{\pi}$ is unique and $\pi_i > 0$ for all $i \le \mathcal{J}$.

Let for the moment $u = 0$ and let $V_i(t) = \int_0^t \mathbb{I}_{\{J_s=i\}}\,\mathrm{d}s$ denote the time, $\{J_s\}$ has spent in state i. Then

$$\frac{C_t}{t} = \sum_{i=1}^{\mathcal{J}} \frac{V_i(t)}{t} \frac{\int_0^t \mathbb{I}_{\{J_s=i\}}\,\mathrm{d}C_s}{V_i(t)}$$

where

$$\int_0^t \mathbb{I}_{\{J_s=i\}}\,\mathrm{d}C_s = \int_0^t \mathbb{I}_{\{J_s=i\}} c\,\mathrm{d}s - \sum_{j=1}^{N_t} \mathbb{I}_{\{J_{T_j}=i\}} Y_i\,.$$

It is well-known (and follows from the law of large numbers) that $V_i(t)/t \to \pi_i$. The random variable $\int_0^t \mathbb{I}_{\{J_s=i\}}\,\mathrm{d}C_s$ has the same distribution as $C^{(i)}_{V_i(t)}$ and it follows from the classical model that $C^{(i)}_{V_i(t)}/V_i(t)$ tends to $c - L_i\mu_i$ a.s.. Thus C_t/t converges a.s. to $c - \sum_{i=1}^{\mathcal{J}} \pi_i L_i \mu_i$. In order to avoid $\psi(u) = 1$ for all u the latter limit should be strictly positive. This means the net profit condition reads

$$c > \sum_{i=1}^{\mathcal{J}} \pi_i L_i \mu_i\,.$$

9.2. The Lundberg Exponent and Lundberg's Inequality

The matrix $\mathrm{e}^{\boldsymbol{\Theta}(r)}$ has only strictly positive elements by Lemma 9.2. Let $\mathrm{e}^{\theta(r)} = \operatorname{spr} \mathrm{e}^{\boldsymbol{\Theta}(r)}$ denote the largest absolute value of the eigenvalues. By the Frobenius theorem $\mathrm{e}^{\theta(r)}$ is an eigenvalue of $\mathrm{e}^{\boldsymbol{\Theta}(r)}$, it is the only algebraic eigenvalue with absolute value $\operatorname{spr} \mathrm{e}^{\boldsymbol{\Theta}(r)}$ and its eigenvector $\boldsymbol{g}(r)$ consists of strictly positive elements. Moreover $\boldsymbol{g}(r)$ is the only non-trivial eigenvector consisting of only positive elements. It follows in particular that $\boldsymbol{g}(0) = \mathbf{1}$ because $\boldsymbol{\eta}\mathbf{1} = \mathbf{0}$. We normalize $\boldsymbol{g}(r)$ such that $\boldsymbol{\pi}\boldsymbol{g}(r) = 1$. We find the following martingale.

Lemma 9.3. *The stochastic process*

$$L_t^r = g_{J_t}(r)/\mathbb{E}[g_{J_0}(r)]\,\mathrm{e}^{-r(C_t-u)-\theta(r)t}$$

is a martingale. Moreover the function $\theta(r)$ is convex and

$$\theta'(0) = -\Bigl(c - \sum_{i=1}^{\mathcal{J}} \pi_i L_i \mu_i\Bigr) < 0\,.$$

Proof. In order to simplify the notation we write g_i instead of $g_i(r)$. Note that

$$\mathbb{E}[g_{J_t}\mathrm{e}^{-r(C_t-u)} \mid J_0 = i] = (\mathrm{e}^{t\boldsymbol{\Theta}(r)}g)_i = \mathrm{e}^{\theta(r)t}g_i\,.$$

Because the increments of $\{C_{t+s} : s \ge 0\}$ depend on $\mathcal{F}_t$ via J_t only, the martingale property is proved.

It follows from Lemma 1.9 that $\log\sum_{i=1}^{\mathcal{J}} \mathbb{E}[\mathrm{e}^{-r(C_t-u)}; J_t = j \mid J_0 = i]$ is strictly convex. It is clear that

$$\varlimsup_{n\to\infty}((\mathrm{e}^{\boldsymbol{\Theta}(r)})^n\mathbf{1})_i^{1/n} \le \mathrm{e}^{\theta(r)}\,.$$

Denote by $\bar{g} = \max\{g_i\}$. Then

$$\varliminf_{n\to\infty}((\mathrm{e}^{\boldsymbol{\Theta}(r)})^n\mathbf{1})_i^{1/n} = \varliminf_{n\to\infty}((\mathrm{e}^{\boldsymbol{\Theta}(r)})^n\bar{g}\mathbf{1})_i^{1/n} \ge \varliminf_{n\to\infty}((\mathrm{e}^{\boldsymbol{\Theta}(r)})^n g)_i^{1/n} = \mathrm{e}^{\theta(r)}\,.$$

Thus there exists a series $f_n(r)$ of strictly convex functions converging to $\theta(r)$. Let $r_1 < r_2$ such that $M_Y(r_2) < \infty$. Let $\alpha \in (0,1)$. Then

$$\begin{aligned}\theta(\alpha r_1 + (1-\alpha)r_2) &= \lim_{n\to\infty} f_n(\alpha r_1 + (1-\alpha)r_2)\\ &\le \lim_{n\to\infty} \alpha f_n(r_1) + (1-\alpha)f_n(r_2) = \alpha\theta(r_1) + (1-\alpha)\theta(r_2)\end{aligned}$$

and $\theta(r)$ is convex.

Note that $\theta(r)$ is an eigenvalue of $\boldsymbol{\Theta}(r)$, i.e. a solution to the equation $\det(\boldsymbol{\Theta}(r) - \theta\boldsymbol{I}) = 0$. By the implicit function theorem $\theta(r)$ is differentiable and therefore also $\boldsymbol{g}(r)$ is differentiable. We have $\boldsymbol{\pi}\boldsymbol{\Theta}(r)\boldsymbol{g}(r) = \theta(r)\boldsymbol{\pi}\boldsymbol{g}(r) = \theta(r)$. In particular

$$\theta'(r) = \boldsymbol{\pi}\boldsymbol{\Theta}'(r)\boldsymbol{g}(r) + \boldsymbol{\pi}\boldsymbol{\Theta}(r)\boldsymbol{g}'(r)\,.$$

Letting $r = 0$ the result for $\theta'(0)$ follows. □

Because $\theta(r)$ is convex there might exist a second solution $R \neq 0$ to the equation $\theta(r) = 0$. If such a solution exists then $R > 0$ and the solution is unique. We call it again **Lundberg exponent** or **adjustment coefficient**. Let us now assume that R exists. As in Chapter 8 let us define the new measure $Q_r[A] = \mathbb{E}_{\mathbb{P}}[L_t^r; A]$. It will again follow from the lemma below that Q_r can be extended to $\mathcal{F}$.

Lemma 9.4.

i) *Let $t > s$ and $A \in \mathcal{F}_s \subset \mathcal{F}_t$. Then* $\mathbb{E}_{\mathbb{P}}[L^r_t; A] = \mathbb{E}_{\mathbb{P}}[L^r_s; A]$.

ii) *Let T be a stopping time and $A \in \mathcal{F}_T$ such that $A \subset \{T < \infty\}$. Then*

$$Q_r[A] = \mathbb{E}_{\mathbb{P}}[L^r_T; A]\,.$$

iii) *Under the measure Q_r, $\{(C_t, J_t)\}$ is a Markov modulated risk model with intensity matrix $\tilde{\eta}_{ij} = (g_i(r))^{-1} g_j(r) \eta_{ij}$ if $i \neq j$, claim size distributions $\tilde{G}_i(x) = \int_0^x \mathrm{e}^{ry} \,\mathrm{d}G_i(y)/M_Y^{(i)}(r)$ and claim arrival intensities $\tilde{L}_i = L_i M_Y^{(i)}(r)$. In particular*

$$\lim_{t \to \infty} \frac{1}{t}(C_t - u) = -\theta'(r)$$

Q_r-a.s..

Proof. i) and ii) are proved as in Lemma 8.1.

iii) In order to simplify the notation we write g_i instead of $g_i(r)$. Let B_k be $\sigma(\Sigma_k - \Sigma_{k-1}, N_{\Sigma_k} - N_{\Sigma_{k-1}}, T_j - T_{j-1}, Y_j; N_{\Sigma_{k-1}} < j \le N_{\Sigma_k})$-measurable sets, $n \in \mathbb{N}$ and $i_k \le \mathcal{J}$. Before one has proved that Q_r can be extended to $\mathcal{F}$ one has to choose sets B_k such that we only consider the process in finite time. Let $i_k \in (1, \dots, \mathcal{J})$. Then

$$\begin{aligned}
&Q_r[J_{\Sigma_{k-1}} = i_k, J_{\Sigma_n} = i_{n+1}, B_k, 1 \le k \le n] \\
&\qquad = \frac{g_{i_{n+1}}}{\mathbb{E}_{\mathbb{P}}[g_{J_0}]} \mathbb{E}_{\mathbb{P}}\Bigl[\prod_{k=1}^n \mathrm{e}^{-r(C_{\Sigma_k} - C_{\Sigma_{k-1}})} \mathrm{e}^{-\theta(r)(\Sigma_k - \Sigma_{k-1})}; \\
&\qquad\qquad\qquad\qquad J_{\Sigma_{k-1}} = i_k, J_{\Sigma_n} = i_{n+1}, B_k, 1 \le k \le n\Bigr] \\
&\qquad = \frac{g_{i_{n+1}}}{\mathbb{E}_{\mathbb{P}}[g_{J_0}]} \mathbb{E}_{\mathbb{P}}[\mathrm{e}^{-r(C_{\Sigma_1} - u)} \mathrm{e}^{-\theta(r)\Sigma_1}; J_0 = i_1, B_1] \\
&\qquad\quad \times \prod_{k=2}^n \mathbb{E}_{\mathbb{P}}[\mathrm{e}^{-r(C_{\Sigma_k} - C_{\Sigma_{k-1}})} \mathrm{e}^{-\theta(r)(\Sigma_k - \Sigma_{k-1})}; J_{\Sigma_{k-1}} = i_k, B_k \mid J_{\Sigma_{k-2}} = i_{k-1}] \\
&\qquad\quad \times \mathbb{P}[J_{\Sigma_n} = i_{n+1} \mid J_{\Sigma_{n-1}} = i_n]
\end{aligned}$$

and therefore the process in the interval $[\Sigma_k, \Sigma_{k+1})$ depends on $\mathcal{F}_{\Sigma_k-}$ via $J_{\Sigma_{k-1}}$ only. In particular

$$\begin{aligned}
&Q_r[J_{\Sigma_k} = i_{k+1}, B_k, 1 \le k \le n \mid J_0 = i_1] \\
&\qquad = \frac{g_{i_{n+1}}}{g_{i_1}} \mathbb{E}_{\mathbb{P}}[\mathrm{e}^{-r(C_{\Sigma_1} - u)} \mathrm{e}^{-\theta(r)\Sigma_1}; B_1 \mid J_0 = i_1] \\
&\qquad\quad \times \prod_{k=2}^n \mathbb{E}_{\mathbb{P}}[\mathrm{e}^{-r(C_{\Sigma_k} - C_{\Sigma_{k-1}})} \mathrm{e}^{-\theta(r)(\Sigma_k - \Sigma_{k-1})}; J_{\Sigma_{k-1}} = i_k, B_k \mid J_{\Sigma_{k-2}} = i_{k-1}] \\
&\qquad\quad \times \mathbb{P}[J_{\Sigma_n} = i_{n+1} \mid J_{\Sigma_{n-1}} = i_n]
\end{aligned}$$

and

$$\begin{aligned}Q[B_2 \mid J_0 = i_1, J_{\Sigma_1} = i_2] &= \mathbb{E}_{\mathbb{P}}[\mathrm{e}^{-r(C_{\Sigma_2}-C_{\Sigma_1})}\mathrm{e}^{-\theta(r)(\Sigma_2-\Sigma_1)}; B_2 \mid J_0 = i_1, J_{\Sigma_1} = i_2] \\ &= \mathbb{E}_{\mathbb{P}}[\mathrm{e}^{-r(C_{\Sigma_2}-C_{\Sigma_1})}\mathrm{e}^{-\theta(r)(\Sigma_2-\Sigma_1)}; B_2 \mid J_{\Sigma_1} = i_2] = Q[B_2 \mid J_{\Sigma_1} = i_2]\end{aligned}$$

and therefore the process in the interval $[\Sigma_n, \Sigma_{n+1})$ depends on $\mathcal{F}_{\Sigma_n}$ via J_{Σ_n} only. Next we compute the distribution of Σ_1.

$$\begin{aligned}Q[\Sigma_1 > v \mid J_0 = i] &= \mathbb{E}_{\mathbb{P}}[\mathrm{e}^{-r(C_v-u)-\theta(r)v}; \Sigma_1 > v \mid J_0 = i] \\ &= \mathrm{e}^{(L_i(M_Y^{(i)}(r)-1)-cr-\theta(r))v}\mathrm{e}^{\eta_{ii}v}\end{aligned}$$

and therefore Σ_1 is $\mathrm{Exp}(-(\boldsymbol{\Theta}(r) - \theta(r)\boldsymbol{I})_{ii})$ distributed. It follows that $\{J_t\}$ is a continuous time Markov chain under Q_r. In order to find the intensity we compute

$$\begin{aligned}Q[J_{\Sigma_1} = j \mid J_0 = i] &= \frac{g_j \mathbb{P}[J_{\Sigma_1} = j \mid J_0 = i]}{g_i}\mathbb{E}_{\mathbb{P}}[\mathrm{e}^{(L_i(M_Y^{(i)}(r)-1)-cr-\theta(r))\Sigma_1} \mid J_0 = i] \\ &= \frac{g_j\eta_{ij}}{-g_i(\boldsymbol{\Theta}(r) - \theta(r)\boldsymbol{I})_{ii}}\end{aligned}$$

and the desired intensity matrix under Q_r follows. Note that we should have considered $Q[J_{\Sigma_1} = j \mid J_0 = i, \Sigma_1 \le t]$ which would have given the same result.

Let $n \in \mathbb{N}$, A be a $\sigma(T_1, \dots, T_n)$-measurable set and B_k be Borel sets. Then

$$\begin{aligned}&Q[\Sigma_1 > v, N_v = n, A, Y_k \in B_k, 1 \le k \le n \mid J_0 = i] \\ &\qquad = \mathrm{e}^{-(cr+\theta(r))v}\mathbb{P}[\Sigma_1 > v, N_v = n, A \mid J_0 = i]\prod_{k=1}^{n}\int_{B_k}\mathrm{e}^{ry}\,\mathrm{d}G_i(y)\end{aligned}$$

and thus the claim size distribution is as desired. Moreover

$$Q[A \mid N_v = n, \Sigma_1 > v, J_0 = i] = \mathbb{P}[A \mid N_v = n, \Sigma_1 > v, J_0 = i]$$

from which it follows that, conditioned on the number of claims, the claim times are distributed as in a Poisson process. Finally

$$Q[N_v = n \mid \Sigma_1 > v, J_0 = i] = \frac{(L_i M_Y^{(i)}(r)v)^n}{n!}\mathrm{e}^{-L_i M_Y^{(i)}(r)v}$$

which shows that $\{(C_t, J_t)\}$ is a Markov modulated risk model.

Under the measure Q_r the process

$$\tilde{L}_t^s = \frac{g_{J_t}(r+s)}{g_{J_t}(r)}\mathrm{e}^{-s(C_t-u)}\mathrm{e}^{-(\theta(s+r)-\theta(r))t}$$

is a martingale because for $t \ge v$

$$\begin{aligned}
&\mathbb{E}_{Q_r}\Bigl[\frac{g_{J_t}(r+s)}{g_{J_t}(r)}\mathrm{e}^{-s(C_t-u)-(\theta(s+r)-\theta(r))t}\Bigm| \mathcal{F}_v\Bigr] \\
&\quad= \frac{\mathbb{E}_{\mathbb{P}}\Bigl[\frac{g_{J_t}(r+s)}{g_{J_t}(r)}\mathrm{e}^{-s(C_t-u)-(\theta(s+r)-\theta(r))t}\frac{g_{J_t}(r)}{\mathbb{E}_{\mathbb{P}}[g_{J_0}(r)]}\mathrm{e}^{-r(C_t-u)-\theta(r)t}\Bigm| \mathcal{F}_v\Bigr]}{\frac{g_{J_v}(r)}{\mathbb{E}_{\mathbb{P}}[g_{J_0}(r)]}\mathrm{e}^{-r(C_v-u)-\theta(r)v}} \\
&\quad= \mathbb{E}_{\mathbb{P}}\Bigl[\frac{g_{J_t}(r+s)}{g_{J_v}(r)}\mathrm{e}^{-(r+s)(C_t-u)-\theta(r+s)t}\Bigm| \mathcal{F}_v\Bigr]\mathrm{e}^{r(C_v-u)+\theta(r)v} \\
&\quad= \frac{g_{J_v}(r+s)}{g_{J_v}(r)}\mathrm{e}^{-(r+s)(C_v-u)-\theta(r+s)v}\mathrm{e}^{r(C_v-u)+\theta(r)v} = \tilde{L}_v^s\,.
\end{aligned}$$

As in Lemma 9.3 it follows that

$$\lim_{t\to\infty}\frac{1}{t}(C_t-u) = -\frac{\mathrm{d}}{\mathrm{d}s}(\theta(r+s)-\theta(r))|_{s=0} = -\theta'(r)\,.$$

□

Let again τ denote the time of ruin and $\psi(u)$ denote the ruin probability. If we write $\psi(u)$ in terms of the measure Q_r we obtain

$$\psi(u) = \mathbb{E}_{\mathbb{P}}[g_{J_0}(r)]\mathbb{E}_{Q_r}[(g_{J_\tau}(r))^{-1}\mathrm{e}^{rC_\tau}\mathrm{e}^{\theta(r)\tau}; \tau < \infty]\mathrm{e}^{-ru}\,. \tag{9.1}$$

The latter expression is only simpler than $\mathbb{P}[\tau < \infty]$ if $r = R$. In the sequel we write Q instead of Q_R and g_i instead of $g_i(R)$.

We now can prove Lundberg's inequality.

Theorem 9.5. *Let $\{(C_t, J_t)\}$ be a Markov modulated risk model and assume that R exists. Let $\{g_i : 1 \le i \le \mathcal{J}\}$ be as above and $g_{\min} = \min\{g_i : 1 \le i \le \mathcal{J}\}$. Then*

$$\psi(u) < \mathbb{E}_{\mathbb{P}}[g_{J_0}]/g_{\min}\,\mathrm{e}^{-Ru}\,.$$

Proof. Note that $g_{\min} > 0$. Thus the assertion follows readily from (9.1) noting that $\mathrm{e}^{RC_\tau} < 1$. □

9.3. The Cramér-Lundberg Approximation

For proving the Cramér-Lundberg approximation in the classical case (8.1) and in the renewal case it was no problem to use the renewal equation because each ladder time was a regeneration point. This is not the case anymore for the Markov modulated risk model. We therefore have to find a slightly different version of the ladder epochs.

Assume that $L_1 \neq 0$ and that $J_0 = 1$. Let

$$\tau_- := \inf\{t > 0 : J_t = 1 \text{ and } C_t = \inf_{0 \le s \le t} C_s\}\,.$$

Because of the lack of memory property of the exponential distribution τ_- is a regeneration point. Note that $Q[\tau_- < \infty] = 1$. If $C_{\tau_-} \ge 0$ then ruin has not yet occurred. Denote by $B(x) = Q[u - C_{\tau_-} \le x]$ the modified ladder height distribution under Q. Thus

$$Z(u) = \mathbb{E}_Q[(g_{J_\tau})^{-1} \mathrm{e}^{RC_\tau} \mid C_0 = u, J_0 = 1]$$

fulfils the renewal equation

$$Z(u) = \int_0^u Z(u-y)\,\mathrm{d}B(y) + z(u)$$

where

$$z(u) = \mathbb{E}_Q[(g_{J_\tau})^{-1} \mathrm{e}^{RC_\tau}; C_{\tau_-} < 0 \mid C_0 = u, J_0 = 1]\,.$$

In order to show that $Z(u)$ converges as $u \to \infty$ we have to show that $z(u)$ is directly Riemann integrable. To avoid $\lim_{u\to\infty} Z(u) = 0$ we assume that there exists an $r > R$ such that $M_Y(r) < \infty$.

Let us first show that $z(u)$ is a continuous function. Define

$$z_i(u) = \mathbb{E}_Q[(g_{J_\tau})^{-1} \mathrm{e}^{RC_\tau}; C_{\tau_-} < 0 \mid C_0 = u, J_0 = i]\,.$$

Note that $z_i(u)$ is bounded. Then for $h > 0$ small we condition on time $T \wedge h$ where $T = T_1 \wedge \Sigma_1$ and T_1 is the time of the first claim. Note that T is under the measure Q exponentially distributed with parameter $\tilde{L}_1 - \tilde{\eta}_{11}$. Then

$$\begin{aligned} z(u) = \mathrm{e}^{-(\tilde{L}_1-\tilde{\eta}_{11})h} z(u+ch) + \int_0^h \int_0^\infty z(u+ct-y)\,\mathrm{d}\tilde{G}_1(y) \tilde{L}_1 \mathrm{e}^{-(\tilde{L}_1-\tilde{\eta}_{11})t}\,\mathrm{d}t \\ + \sum_{j=2}^{\mathcal{J}} \int_0^h z_j(u+ct) \tilde{\eta}_{1j} \mathrm{e}^{-(\tilde{L}_1-\tilde{\eta}_{11})t}\,\mathrm{d}t\;. \end{aligned}$$

Letting $h \to 0$ shows that $z(u)$ is right continuous. Replacing u by $u - ch$ gives

$$z(u-ch) = \mathrm{e}^{-(\tilde{L}_1 - \tilde{\eta}_{11})h} z(u) + \int_0^h \int_0^\infty z(u - c(h-t) - y)\, \mathrm{d}\tilde{G}_1(y) \tilde{L}_1 \mathrm{e}^{-(\tilde{L}_1 - \tilde{\eta}_{11})t}\, \mathrm{d}t + \sum_{j=2}^{\mathcal{J}} \int_0^h z_j(u - c(h-t)) \tilde{\eta}_{1j} \mathrm{e}^{-(\tilde{L}_1 - \tilde{\eta}_{11})t}\, \mathrm{d}t\,.$$

and $z(u)$ is left continuous too.

It remains to show that $z(u)$ has a directly Riemann integrable upper bound. Then it will follow from Lemma C.11 that $z(u)$ is directly Riemann integrable. It is clear that

$$g_{\min} z(u) \le Q[C_{\tau_-} < 0 \mid C_0 = u, J_0 = 1] = Q[-C_{\tau_-} > u \mid J_0 = 1, C_0 = 0]\,.$$

The latter function is monotone. Thus it is enough to show that it is integrable. But

$$\int_0^\infty Q[-C_{\tau_-} > u \mid J_0 = 1, C_0 = 0]\, \mathrm{d}u = \mathbb{E}_Q[-C_{\tau_-} \mid J_0 = 1, C_0 = 0] < \infty\,.$$

The latter follows from the Wiener-Hopf factorization because

$$\begin{aligned} &\mathbb{E}_Q[\mathrm{e}^{(r-R)/2\,(-C_{\tau_-})} \mid J_0 = 1, C_0 = 0] \\ &\quad = \mathbb{E}_{\mathbb{P}}[\mathrm{e}^{(r+R)/2\,(-C_{\tau_-})}; \tau_- < \infty \mid J_0 = 1, C_0 = 0] < \infty \end{aligned}$$

since $\mathbb{E}_{\mathbb{P}}[\mathrm{e}^{-rC_t}] < \infty$ for all $t \ge 0$. It follows readily that for $\tau^* = \inf\{T_n : n \ge 0, J_{T_n} = 1\}$ we also have $\mathbb{E}_{\mathbb{P}}[\exp\{-rC_{\tau^*}\}] < \infty$. We have just shown that $z(u)$ is directly Riemann integrable and therefore that the limit of $Z(u)$ exists as $u \to \infty$.

Theorem 9.6. *Let $\{(C_t, J_t)\}$ be a Markov modulated risk model and let $\{g_i : 1 \le i \le \mathcal{J}\}$ as above. Assume that R exists and that there is an $r > R$ such that $M_Y(r) < \infty$. Then*

$$\lim_{u\to\infty} \psi(u)\mathrm{e}^{Ru} = \mathbb{E}_{\mathbb{P}}[g_{J_0}]\, C$$

for some constant $C > 0$ which is independent of the distribution of J_0.

Proof. We have already seen that the assertion is true if $J_0 = 1$ a.s.. Denote the limit of $Z(u)$ by C. Let $\tilde{\tau} = \inf\{t \ge 0 : J_t = 1\}$ and $B_1(x) = Q[u - C_{\tilde{\tau}} \le x]$. Denote by

$$\tilde{Z}(u) = \mathbb{E}_Q[(g_{J_\tau})^{-1} \mathrm{e}^{RC_\tau} \mid C_0 = u]$$

the analogue of $Z(u)$. Then

$$\tilde{Z}(u) = \int_{-\infty}^{u} Z(u-y)\,\mathrm{d}B_1(y) + \tilde{z}(u)$$

where

$$\tilde{z}(u) = \mathbb{E}_Q[(g_{J_\tau})^{-1}\mathrm{e}^{RC_\tau}; \tau \le \tilde{\tau} \mid C_0 = u]\,.$$

Because $Q[\tau \le \tilde{\tau} \mid C_0 = u]$ tends to 0 as $u \to \infty$ we find that $\tilde{z}(u)$ tends to 0 as $u \to \infty$. Because $Z(x)$ is a bounded function we can interchange the limit and the integration to find that

$$\lim_{u\to\infty} \tilde{Z}(u) = \int_{-\infty}^{\infty} \lim_{u\to\infty} Z(u-y)\,\mathrm{d}B_1(y) = C\,.$$

□

9.4. Subexponential Claim Sizes

Assume that one or several of the claim size distributions G_i are subexponential. Assume that $G_1(x)$ has the heaviest tail of all claim size distributions, more specifically, there exists constants $0 \le K_i < \infty$ such that

$$\lim_{y\to\infty} \frac{1-G_i(y)}{1-G_1(y)} = K_i$$

for all $i \le \mathcal{J}$. We assume without mentioning that $L_1 > 0$. The following result is much more complicated to prove than the corresponding result in the renewal risk model. We therefore state the result without proof.

Theorem 9.7. *Assume that*

$$\lim_{y\to\infty} \frac{1-G_i(y)}{1-G_1(y)} = K_i < \infty\,,$$

that $L_1 > 0$ *and that*

$$\frac{1}{\mu_1}\int_0^x (1-G_1(y))\,\mathrm{d}y$$

is subexponential. Let $K = \sum_i \pi_i L_i K_i$. *Then*

$$\lim_{u\to\infty} \frac{\psi(u)}{\int_u^\infty (1-G_1(y))\,\mathrm{d}y} = \frac{K}{c - \sum_i \pi_i L_i \mu_i}$$

independent of the distribution of the initial state. □

Let us consider the problem intuitively. Assume the process starts in state 1. Consider the random walk generated by the process by considering only the time points when the process J_t jumps to state 1. The increments of this random walk have a left tail that is asymptotically comparable with the tail of G_1 (only the heaviest tail contributes). Thus it follows as in the renewal case that the ruin probability of the random walk has the desired behaviour.

The time between two regeneration points has a light tailed distribution. The difference between the infimum of the random walk and the infimum of the risk process is smaller than c times the distance between the corresponding two regeneration points. The infimum of the random walk has a tail proportional to $\int_u^\infty (1-G_1(y))\,\mathrm{d}y$ which will asymptotically dominate the light tail of the distance between two regeneration points. Thus the tail of the distribution of the infimum of the random walk and tail of the distribution of the infimum of the risk process are asymptotically equivalent.

9.5. Finite Time Lundberg Inequalities

Let $0 \le \underline{y} < \bar{y} < \infty$. As for the other models we want to consider probabilities of the form $\mathbb{P}[\underline{y}u < \tau \le \bar{y}u]$. We can proceed exactly as in the classical case (Section 4.14)

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}}[g_{J_0}(r)]\mathrm{e}^{-ru} &= \mathbb{E}_{\mathbb{P}}[g_{J_{\tau\wedge\bar{y}u}}(r)\mathrm{e}^{-rC_{\tau\wedge\bar{y}u}}\mathrm{e}^{-\theta(r)(\tau\wedge\bar{y}u)}] \\
&> \mathbb{E}_{\mathbb{P}}[g_{J_\tau}(r)\mathrm{e}^{-rC_\tau}\mathrm{e}^{-\theta(r)\tau};\underline{y}u < \tau \le \bar{y}u] \\
&> g_{\min}(r)\mathbb{E}_{\mathbb{P}}[\mathrm{e}^{-\theta(r)\tau} \mid \underline{y}u < \tau \le \bar{y}u]\mathbb{P}[\underline{y}u < \tau \le \bar{y}u] \\
&\ge g_{\min}(r)\mathrm{e}^{-\max\{\underline{y}\theta(r),\bar{y}\theta(r)\}u}\mathbb{P}[\underline{y}u < \tau \le \bar{y}u]\,.
\end{aligned}$$

Hence we obtain the inequality

$$\mathbb{P}[\underline{y}u < \tau \le \bar{y}u] < \frac{\mathbb{E}_{\mathbb{P}}[g_{J_0}(r)]}{g_{\min}(r)}\mathrm{e}^{-\min\{r-\underline{y}\theta(r),r-\bar{y}\theta(r)\}u}\,.$$

Defining the finite time Lundberg exponent

$$R(\underline{y},\bar{y}) = \sup\{\min\{r-\underline{y}\theta(r), r-\bar{y}\theta(r)\} : r \in \mathbb{R}\}$$

we obtain the finite time Lundberg inequality

$$\mathbb{P}[\underline{y}u < \tau \le \bar{y}u] < \frac{\mathbb{E}_{\mathbb{P}}[g_{J_0}(\tilde{r})]}{g_{\min}(\tilde{r})}\mathrm{e}^{-R(\underline{y},\bar{y})u}$$

where $\tilde{r}$ is the argument maximizing $\min\{r-\underline{y}\theta(r), r-\bar{y}\theta(r)\}$. In the discussion of $R(\underline{y},\bar{y})$ in Section 4.14 we only used to property that $\theta(r)$ is convex. Thus the same discussion applies. Let r_y be the argument maximizing $r - y\theta(r)$ and let $y_0 = (\theta'(R))^{-1}$ denote the critical value. Then we obtain the two finite time Lundberg inequalities

$$\mathbb{P}[0 < \tau \le yu \mid C_0 = u] < \frac{\mathbb{E}_{\mathbb{P}}[g_{J_0}(r_y)]}{g_{\min}(r_y)} \mathrm{e}^{-R(0,y)u}$$

where $R(0, y) > R$ if $y < y_0$, and

$$\mathbb{P}[yu < \tau < \infty \mid C_0 = u] < \frac{\mathbb{E}_{\mathbb{P}}[g_{J_0}(r_y)]}{g_{\min}(r_y)} \mathrm{e}^{-R(y,\infty)u}$$

where $R(y, \infty) > R$ if $y > y_0$.

Let $r_\infty = \sup\{r \in \mathbb{R} : M_Y(r) < \infty\}$. The next result can be proved in the same way as Theorem 4.24 was proved.

Theorem 9.8. *Assume that* $R < r_\infty$. *Then*

$$\frac{\tau}{u} \longrightarrow y_0$$

in probability on the set $\{\tau < \infty\}$. □

Bibliographical Remarks

The Markov modulated risk model was introduced as a semi-Markov model by Janssen [56]. The useful formulation as a model in continuous time goes back to Asmussen [8], where also Lemma 9.2, Theorem 9.5 and Theorem 9.6 were proved. A version of Theorem 9.7, where the limiting constant was dependent on the initial distribution, can be found in [10]. The Theorem presented here is proved in [12]. The model is also discussed in [66].

A. Stochastic Processes

We start with some definitions.

Definition A.1. *Let* I *be either* $\mathbb{N}$ *or* $[0, \infty)$*. A* **stochastic process** *on* I *with* **state space** E *is a family of random variables* $\{X_t : t \in I\}$ *on* E*. Let* E *be a topological space. A stochastic process is called* **cadlag** *if it has a.s. right continuous paths and the limits from the left exist. It is called* **continuous** *if its paths are a.s. continuous. We will often identify a cadlag (continuous, respectively) stochastic process with a random variable on the space of cadlag (continuous) functions. For the rest of these notes we will always assume that all stochastic processes on* $[0, \infty)$ *are cadlag.*

Definition A.2. *A cadlag stochastic process* $\{N_t\}$ *on* $[0, \infty)$ *is called* **point process** *if a.s.*

i) $N_0 = 0$,

ii) $N_t \geq N_s$ *for all* $t \geq s$ *and*

iii) $N_t \in \mathbb{N}$ *for all* $t \in (0, \infty)$.

We denote the jump times by $T_1, T_2, \ldots$*, i.e.* $T_k = \inf\{t \geq 0 : N_t \geq k\}$ *for all* $k \in \mathbb{N}$*. In particular* $T_0 = 0$*. A point process is called* **simple** *if* $T_0 < T_1 < T_2 < \cdots$.

Definition A.3. *A stochastic process* $\{X_t\}$ *is said to have* **independent increments** *if for all* $n \in \mathbb{N}$ *and all real numbers* $0 = t_0 < t_1 < t_2 < \cdots < t_n$ *the random variables* $X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ *are independent. A stochastic process is said to have* **stationary increments** *if for all* $n \in \mathbb{N}$ *and all real numbers* $0 = t_0 < t_1 < t_2 < \cdots < t_n$ *and all* $h > 0$ *the random vectors* $(X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}})$ *and* $(X_{t_1+h} - X_{t_0+h}, X_{t_2+h} - X_{t_1+h}, \ldots, X_{t_n+h} - X_{t_{n-1}+h})$ *have the same distribution.*

Definition A.4. *An increasing family* $\{\mathcal{F}_t\}$ *of* σ*-algebras is called* **filtration** *if* $\mathcal{F}_t \subset \mathcal{F}$ *for all* t*. A stochastic process* $\{X_t\}$ *is called* $\{\mathcal{F}_t\}$**-adapted** *if* X_t *is* $\mathcal{F}_t$*-measurable for all* $t \geq 0$*. The* **natural filtration** $\{\mathcal{F}_t^X\}$ *of a stochastic process* $\{X_t\}$ *is the smallest filtration such that the process is adapted. If a stochastic process is considered then we use its natural filtration if nothing else is mentioned. A filtration is called* **right continuous** *if* $\mathcal{F}_{t+} := \bigcap_{s>t} \mathcal{F}_s = \mathcal{F}_t$ *for all* $t \geq 0$.

Definition A.5. *A* $\{\mathcal{F}_t\}$-**stopping time** *is a random variable* T *on* $[0, \infty]$ *or* $\mathbb{N}$ *such that* $\{T \leq t\} \in \mathcal{F}_t$ *for all* $t \geq 0$. *If we stop a stochastic process then we usually assume that* T *is a stopping time with respect to the natural filtration of the stochastic process. The* σ*-algebra*

$$\mathcal{F}_T := \{A \in \mathcal{F} : A \cap \{T \leq t\} \in \mathcal{F}_t \textit{ for all } t \geq 0\}$$

is called the information up to the stopping time T.

A.1. Bibliographical Remarks

For further literature on stochastic processes see for instance [30], [39], [58], [59], [67] or [88].

B. Martingales

Definition B.1. *A stochastic process* $\{M_t\}$ *with state space* $\mathbb{R}$ *is called a* $\{\mathcal{F}_t\}$-**martingale** *(***-submartingale, -supermartingale***, respectively) if*

i) $\{M_t\}$ *is adapted,*

ii) $\mathbb{E}[M_t]$ *exists for all* $t \geq 0$*,*

iii) $\mathbb{E}[M_t \mid \mathcal{F}_s] = (\geq, \leq)\, M_s$ *a.s. for all* $t \geq s \geq 0$*.*

We simply say $\{M_t\}$ *is a martingale (submartingale, supermartingale) if it is a martingale (submartingale, supermartingale) with respect to its natural filtration.*

The following two propositions are very important for dealing with martingales.

Proposition B.2. (Martingale stopping theorem) *Let* $\{M_t\}$ *be a* $\{\mathcal{F}_t\}$*-martingale (-submartingale, -supermartingale, respectively) and* T *be a* $\{\mathcal{F}_t\}$*-stopping time. Assume that* $\{\mathcal{F}_t\}$ *is right continuous. Then also the stochastic process* $\{M_{T\wedge t} : t \geq 0\}$ *is a* $\{\mathcal{F}_t\}$*-martingale (-submartingale, -supermartingale). Moreover,* $\mathbb{E}[M_t \mid \mathcal{F}_T] = (\geq, \leq) M_{T\wedge t}$*.*

Proof. See [30, Thm. 12], [39, Thm. II.2.13] or [66, Thm. 10.2.5]. □

Proposition B.3. (Martingale convergence theorem) *Let* $\{M_t\}$ *be a* $\{\mathcal{F}_t\}$*-martingale such that* $\lim_{t\to\infty} \mathbb{E}[M_t^-] < \infty$ *(or equivalently* $\sup_{t\geq 0} \mathbb{E}[|M_t|] < \infty$*). If* $\{\mathcal{F}_t\}$ *is right continuous then the random variable* $M_\infty := \lim_{t\to\infty} M_t$ *exists a.s. and is integrable.*

Proof. See for instance [30, Thm. 6] or [66, Thm. 10.2.2]. □

Note that in general

$$\mathbb{E}[M_\infty] = \mathbb{E}[\lim_{t\to\infty} M_t] \neq \lim_{t\to\infty} \mathbb{E}[M_t] = \mathbb{E}[M_0]\,.$$

B.1. Bibliographical Remarks

The theory of martingales can also be found in [30], [39] or [66].

C. Renewal Processes

C.1. Poisson Processes

Definition C.1. *A point process $\{N_t\}$ is called (homogeneous)* **Poisson process** *with rate λ if*

i) *$\{N_t\}$ has stationary and independent increments,*

ii) $$\mathbb{P}[N_h = 0] = 1 - \lambda h + o(h) \qquad \text{as } h \to 0,$$

iii) $$\mathbb{P}[N_h = 1] = \lambda h + o(h) \qquad \text{as } h \to 0.$$

Remark. It follows readily that $\mathbb{P}[N_h \geq 2] = o(h)$ and that the point process is simple. ■

We give now some alternative definitions of the Poisson process.

Proposition C.2. *Let $\{N_t\}$ be a point process. Then the following are equivalent:*

i) *$\{N_t\}$ is a Poisson process with rate λ.*

ii) *$\{N_t\}$ has independent increments and $N_t \sim \mathrm{Pois}(\lambda t)$ for each fixed $t \geq 0$.*

iii) *The interarrival times $\{T_k - T_{k-1} : k \geq 1\}$ are independent and $\mathrm{Exp}(\lambda)$ distributed.*

iv) *For each fixed $t \geq 0$, $N_t \sim \mathrm{Pois}(\lambda t)$ and given $\{N_t = n\}$ the occurrence points have the same distribution as the order statistics of n independent uniformly on $[0, t]$ distributed points.*

v) *$\{N_t\}$ has independent increments such that $\mathbb{E}[N_1] = \lambda$ and given $\{N_t = n\}$ the occurrence points have the same distribution as the order statistics of n independent uniformly on $[0, t]$ distributed points.*

vi) *$\{N_t\}$ has independent and stationary increments such that $\mathbb{P}[N_h \geq 2] = o(h)$ and $\mathbb{E}[N_1] = \lambda$.*

Proof. "*i*) $\Longrightarrow$ *ii*)" Let $p_n(t) = \mathbb{P}[N_t = n]$. Then show that $p_n(t)$ is continuous and differentiable. Finding the differential equations and solving them shows *ii*). The details are left as an exercise.

"*ii*) $\Longrightarrow$ *iii*)" We first show that $\{N_t\}$ has stationary increments. It is enough to show that $N_{t+h} - N_h$ is Pois(λt) distributed. For the moment generating functions we obtain

$$\mathrm{e}^{\lambda(t+h)(\mathrm{e}^r-1)} = \mathbb{E}[\mathrm{e}^{r(N_{t+h}-N_h)}\mathrm{e}^{rN_h}] = \mathbb{E}[\mathrm{e}^{r(N_{t+h}-N_h)}]\mathrm{e}^{\lambda h(\mathrm{e}^r-1)}\,.$$

It follows that

$$\mathbb{E}[\mathrm{e}^{r(N_{t+h}-N_h)}] = \mathrm{e}^{\lambda t(\mathrm{e}^r-1)}\,.$$

This proves that $N_{t+h} - N_h$ is Pois(λt) distributed.

Let $t_0 = 0 \le s_1 < t_1 \le s_2 < t_2 \le \cdots \le s_n < t_n$. Then

$$\begin{aligned}
&\mathbb{P}[s_k < T_k \le t_k, 1 \le k \le n] \\
&\qquad= \mathbb{P}[N_{s_k} - N_{t_{k-1}} = 0, N_{t_k} - N_{s_k} = 1,\ 1 \le k \le n-1, \\
&\qquad\qquad\qquad\qquad N_{s_n} - N_{t_{n-1}} = 0, N_{t_n} - N_{s_n} \ge 1] \\
&\qquad= \mathrm{e}^{-\lambda(s_n-t_{n-1})}(1-\mathrm{e}^{-\lambda(t_n-s_n)})\prod_{k=1}^{n-1}\mathrm{e}^{-\lambda(s_k-t_{k-1})}\lambda(t_k-s_k)\mathrm{e}^{-\lambda(t_k-s_k)} \\
&\qquad= (\mathrm{e}^{-\lambda s_n} - \mathrm{e}^{-\lambda t_n})\lambda^{n-1}\prod_{k=1}^{n-1}(t_k - s_k) \\
&\qquad= \int_{s_1}^{t_1}\cdots\int_{s_n}^{t_n}\lambda^n\mathrm{e}^{-\lambda y_n}\,\mathrm{d}y_n\cdots\mathrm{d}y_1 \\
&\qquad= \int_{s_1}^{t_1}\int_{s_2-z_1}^{t_2-z_1}\cdots\int_{s_n-z_1-\cdots-z_{n-1}}^{t_n-z_1-\cdots-z_{n-1}}\lambda^n\mathrm{e}^{-\lambda(z_1+\cdots+z_n)}\,\mathrm{d}z_n\cdots\mathrm{d}z_1\,.
\end{aligned}$$

It follows that the joint density of $T_1, T_2 - T_1, \ldots, T_n - T_{n-1}$ is

$$\lambda^n\mathrm{e}^{-\lambda(z_1+\cdots+z_n)}$$

and therefore they are independent and Exp(λ) distributed.

"*iii*) $\Longrightarrow$ *iv*)" Note that T_n is $\Gamma(n,\lambda)$ distributed. Thus $\mathbb{P}[N_t = 0] = \mathbb{P}[T_1 > t] = \mathrm{e}^{-\lambda t}$ and for $n \ge 1$

$$\begin{aligned}
\mathbb{P}[N_t = n] &= \mathbb{P}[N_t \ge n] - \mathbb{P}[N_t \ge n+1] = \mathbb{P}[T_n \le t] - \mathbb{P}[T_{n+1} \le t] \\
&= \int_0^t \frac{\lambda^n s^{n-1}}{(n-1)!}\mathrm{e}^{-\lambda s}\,\mathrm{d}s - \int_0^t \frac{\lambda^{n+1}s^n}{n!}\mathrm{e}^{-\lambda s}\,\mathrm{d}s \\
&= \int_0^t \frac{\mathrm{d}}{\mathrm{d}s}\frac{(\lambda s)^n}{n!}\mathrm{e}^{-\lambda s}\,\mathrm{d}s = \frac{(\lambda t)^n}{n!}\mathrm{e}^{-\lambda t}\,.
\end{aligned}$$

The joint density of $T_1, \ldots, T_{n+1}$ is for $t_0 = 0$

$$\prod_{k=1}^{n+1} \lambda \mathrm{e}^{-\lambda(t_k - t_{k-1})} = \lambda^{n+1} \mathrm{e}^{-\lambda t_{n+1}} \,.$$

Thus the joint conditional density of $T_1, \ldots, T_n$ given $N_t = n$ is

$$\frac{\int_t^\infty \lambda^{n+1} \mathrm{e}^{-\lambda t_{n+1}} \; \mathrm{d}t_{n+1}}{\int_0^t \int_{t_1}^t \cdots \int_{t_{n-1}}^t \int_t^\infty \lambda^{n+1} \mathrm{e}^{-\lambda t_{n+1}} \; \mathrm{d}t_{n+1} \cdots \mathrm{d}t_1} = \frac{n!}{t^n} \,.$$

This is the claimed conditional distribution.

"*iv*) $\Longrightarrow$ *v*)" It is clear that $\mathbb{E}[N_1] = \lambda$. Let $x_k \in \mathbb{N}$ and $t_0 = 0 < t_1 < \cdots < t_n$. Then for $x = x_1 + \cdots + x_n$

$$\begin{aligned} &\mathbb{P}[N_{t_k} - N_{t_{k-1}} = x_k, 1 \le k \le n] \\ &\qquad = \mathbb{P}[N_{t_k} - N_{t_{k-1}} = x_k, 1 \le k \le n \mid N_{t_n} = x] \mathbb{P}[N_{t_n} = x] \\ &\qquad = \frac{(\lambda t_n)^x}{x!} \mathrm{e}^{-\lambda t_n} \frac{x!}{x_1! \cdots x_n!} \prod_{k=1}^n \Bigl(\frac{t_k - t_{k-1}}{t_n} \Bigr)^{x_k} = \prod_{k=1}^n \frac{(\lambda(t_k - t_{k-1}))^{x_k}}{x_k!} \mathrm{e}^{-\lambda(t_k - t_{k-1})} \end{aligned}$$

and therefore $\{N_t\}$ has independent increments.

"*v*) $\Longrightarrow$ *i*)" Note that for $x_k \in \mathbb{N}$, $t_0 = 0 < t_1 < \cdots < t_n$ and $h > 0$

$$\begin{aligned} &\mathbb{P}[N_{t_k + h} - N_{t_{k-1}+h} = x_k, 1 \le k \le n \mid N_{t_n + h}] \\ &\qquad = \mathbb{P}[N_{t_k} - N_{t_{k-1}} = x_k, 1 \le k \le n \mid N_{t_n + h}] \,. \end{aligned}$$

Integrating with respect to N_{t_n+h} yields that $\{N_t\}$ has stationary increments. For $\mathbb{P}[N_t = 0]$ we obtain

$$\mathbb{P}[N_{t+s} = 0] = \mathbb{P}[N_{t+s} - N_s = 0, N_s = 0] = \mathbb{P}[N_t = 0] \mathbb{P}[N_s = 0] \,.$$

It is left as an exercise to show that $\mathbb{P}[N_t = 0] = (\mathbb{P}[N_1 = 0])^t$. Let now $\lambda_0 := -\log \mathbb{P}[N_1 = 0]$. Then $\mathbb{P}[N_t = 0] = \mathrm{e}^{-\lambda_0 t}$. Moreover, for $0 < h < t$,

$$\mathrm{e}^{-\lambda_0(t-h)} = \mathbb{E}[\mathbb{P}[N_{t-h} = 0 \mid N_t]] = \mathrm{e}^{-\lambda_0 t} + \frac{h}{t} \mathbb{P}[N_t = 1] + \sum_{k=2}^\infty \Bigl(\frac{h}{t} \Bigr)^k \mathbb{P}[N_t = k] \,.$$

Reordering the terms yields

$$\frac{\mathrm{e}^{-\lambda_0(t-h)} - \mathrm{e}^{-\lambda_0 t}}{h} t = \mathbb{P}[N_t = 1] + \frac{h}{t} \sum_{k=2}^\infty \Bigl(\frac{h}{t} \Bigr)^{k-2} \mathbb{P}[N_t = k] \,.$$

Because the sum is bounded by 1, letting $h \to 0$ gives $\mathbb{P}[N_t = 1] = \lambda_0 t \mathrm{e}^{-\lambda_0 t}$. In particular, $\{N_t\}$ is a Poisson process with rate λ_0. From "*i*) $\Longrightarrow$ *v*)" we conclude

that $\lambda_0 = \mathbb{E}[N_1] = \lambda$. This implies $i)$.
"$i) \Longrightarrow vi)$" Follows readily from $i)$ and $v)$.
"$vi) \Longrightarrow i)$" As in the proof of "$v) \Longrightarrow i)$" it follows that $\mathbb{P}[N_t = 0] = \mathrm{e}^{-\lambda_0 t}$. Then $\mathbb{P}[N_h = 1] = 1 - \mathbb{P}[N_h = 0] - \mathbb{P}[N_h \geq 2] = \lambda_0 h + o(h)$. Thus $\{N_t\}$ is a Poisson process with rate λ_0. From "$i) \Longleftrightarrow v)$" we conclude that $\lambda_0 = \mathbb{E}[N_1] = \lambda$. □

Remark. The condition $\mathbb{E}[N_1] = \lambda$ in $v)$ and $vi)$ is only used for identifying the parameter λ. We did not use in the proof that $\mathbb{E}[N_1] < \infty$. ■

The Poisson process has the following properties.

Proposition C.3. *Let $\{N_t\}$ and $\{\tilde{N}_t\}$ be two independent Poisson processes with rates λ and $\tilde{\lambda}$ respectively. Let $\{I_i : i \in \mathbb{N}\}$ be an iid. sequence of random variables independent of $\{N_t\}$ with $\mathbb{P}[I_i = 1] = 1 - \mathbb{P}[I_i = 0] = q$ for some $q \in (0, 1)$. Furthermore let $a > 0$ be a real number. Then*

i) $\{N_t + \tilde{N}_t\}$ *is a Poisson process with rate $\lambda + \tilde{\lambda}$.*

ii) $\left\{\sum_{i=1}^{N_t} I_i\right\}$ *is a Poisson process with rate λq.*

iii) $\{N_{at}\}$ *is a Poisson process with rate λa.*

Proof. Exercise. □

Definition C.4. *Let $\Lambda(t)$ be an increasing right continuous function on $[0, \infty)$ with $\Lambda(0) = 0$. A point process $\{N_t\}$ on $[0, \infty)$ is called* **inhomogeneous Poisson process** *with* **intensity measure** *$\Lambda(t)$ if*

i) $\{N_t\}$ *has independent increments,*

ii) $N_t - N_s \sim \mathrm{Pois}(\Lambda(t) - \Lambda(s))$.

If there exists a function $\lambda(t)$ such that $\Lambda(t) = \int_0^t \lambda(s)\,\mathrm{d}s$ then $\lambda(t)$ is called **intensity** *or* **rate** *of the inhomogeneous Poisson process.*

Note that a homogeneous Poisson process is a special case with $\Lambda(t) = \lambda t$. Define $\Lambda^{-1}(x) = \sup\{t \geq 0 : \Lambda(t) \leq x\}$ the inverse function of $\Lambda(t)$.

Proposition C.5. *Let $\{\tilde{N}_t\}$ be a homogeneous Poisson process with rate 1. Define $N_t = \tilde{N}_{\Lambda(t)}$. Then $\{N_t\}$ is an inhomogeneous Poisson process with intensity measure $\Lambda(t)$. Conversely, let $\{N_t\}$ be an inhomogeneous Poisson process with intensity measure $\Lambda(t)$. Let $\tilde{N}_t = N_{\Lambda^{-1}(t)}$ at all points where $\Lambda(\Lambda^{-1}(t)) = t$. On intervals $(\Lambda(\Lambda^{-1}(t)-), \Lambda(\Lambda^{-1}(t)))$ where $\Lambda(\Lambda^{-1}(t)) \neq t$ let there be $N_{\Lambda^{-1}(t)} - N_{(\Lambda^{-1}(t)-)}$ occurrence points uniformly distributed on the interval $(\Lambda(\Lambda^{-1}(t)-), \Lambda(\Lambda^{-1}(t)))$ independent of $\{N_t\}$. Then $\{\tilde{N}_t\}$ is a homogeneous Poisson process with rate* 1.

Proof. Exercise. $\square$

For an inhomogeneous Poisson process we can construct the following martingales.

Lemma C.6. *Let $r \in \mathbb{R}$. The following processes are martingales.*

i) $\{N_t - \Lambda(t)\}$,

ii) $\{(N_t - \Lambda(t))^2 - \Lambda(t)\}$,

iii) $\{\exp\{rN_t - \Lambda(t)(\mathrm{e}^r - 1)\}\}$.

Proof. i) Because $\{N_t\}$ has independent increments

$$\mathbb{E}[N_t - \Lambda(t) \mid \mathcal{F}_s] = \mathbb{E}[N_t - N_s] + N_s - \Lambda(t) = N_s - \Lambda(s)\,.$$

ii) Analogously,

$$\begin{aligned}
&\mathbb{E}[(N_t - \Lambda(t))^2 - \Lambda(t) \mid \mathcal{F}_s] \\
&\quad = \mathbb{E}[(N_t - N_s - \{\Lambda(t) - \Lambda(s)\} + N_s - \Lambda(s))^2 \mid \mathcal{F}_s] - \Lambda(t) \\
&\quad = \mathbb{E}[(N_t - N_s - \{\Lambda(t) - \Lambda(s)\})^2] \\
&\qquad\qquad + 2(N_s - \Lambda(s))\mathbb{E}[N_t - N_s - \{\Lambda(t) - \Lambda(s)\}] + (N_s - \Lambda(s))^2 - \Lambda(t) \\
&\quad = \Lambda(t) - \Lambda(s) + 2 \cdot 0 + (N_s - \Lambda(s))^2 - \Lambda(t) = (N_s - \Lambda(s))^2 - \Lambda(s)\,.
\end{aligned}$$

iii) Analogously

$$\mathbb{E}[\mathrm{e}^{rN_t - \Lambda(t)(\mathrm{e}^r - 1)} \mid \mathcal{F}_s] = \mathbb{E}[\mathrm{e}^{r(N_t - N_s)}]\mathrm{e}^{rN_s - \Lambda(t)(\mathrm{e}^r - 1)} = \mathrm{e}^{rN_s - \Lambda(s)(\mathrm{e}^r - 1)}\,.$$

$\square$

C.2. Renewal Processes

Definition C.7. *A simple point process* $\{N_t\}$ *is called an* **ordinary renewal process** *if the interarrival times* $\{T_k - T_{k-1} : k \geq 1\}$ *are iid.. If* T_1 *has a different distribution then* $\{N_t\}$ *is called a* **delayed renewal process**. *If* $\lambda^{-1} = \mathbb{E}[T_2 - T_1]$ *exists and*

$$\mathbb{P}[T_1 \leq x] = \lambda \int_0^x \mathbb{P}[T_2 - T_1 > y] \, \mathrm{d}y \tag{C.1}$$

then $\{N_t\}$ *is called a* **stationary renewal process**. *If* $\{N_t\}$ *is an ordinary renewal process then the function* $U(t) = 1\!\!I_{\{t \geq 0\}} + \mathbb{E}[N_t]$ *is called the* **renewal measure**.

In the rest of this section we denote by F the distribution function of $T_2 - T_1$. For simplicity we let T be a random variable with distribution F. Note that because the point process is simple we implicitly assume that $F(0) = 0$. If nothing else is said we consider in the sequel only ordinary renewal processes.

Recall that F^{*0} is the indicator function of the interval $[0, \infty)$.

Lemma C.8. *The renewal measure can be written as*

$$U(t) = \sum_{n=0}^{\infty} F^{*n}(t) \, .$$

Moreover, $U(t) < \infty$ *for all* $t \geq 0$ *and* $U(t) \to \infty$ *as* $t \to \infty$.

Proof. Exercise. □

In the renewal theory one often has to solve equations of the form

$$Z(x) = z(x) + \int_0^x Z(x-y) \, \mathrm{d}F(y) \qquad (x \geq 0) \tag{C.2}$$

where $z(x)$ is a known function and $Z(x)$ is unknown. This equation is called the **renewal equation**. The equation can be solved explicitly.

Proposition C.9. *If* $z(x)$ *is bounded on bounded intervals then*

$$Z(x) = \int_{0-}^x z(x-y) \, \mathrm{d}U(y) = z * U(x)$$

is the unique solution to (C.2) *that is bounded on bounded intervals.*

Proof. We first show that $Z(x)$ is a solution. This follows from

$$Z(x) = z * U(x) = z * \sum_{n=0}^{\infty} F^{*n}(x) = \sum_{n=0}^{\infty} z * F^{*n}(x)$$
$$= z(x) + \sum_{n=1}^{\infty} z * F^{*(n-1)} * F(x) = z(x) + z * U * F(x) = z(x) + Z * F(x)\,.$$

Let $Z_1(x)$ be a solution that is bounded on bounded intervals. Then

$$|Z_1(x)-Z(x)| = \Big|\int_0^x (Z_1(x-y)-Z(x-y))\,\mathrm{d}F(y)\Big| \le \int_0^x |Z_1(x-y)-Z(x-y)|\,\mathrm{d}F(y)$$

and by induction

$$|Z_1(x) - Z(x)| \le \int_0^x |Z_1(x-y) - Z(x-y)|\,\mathrm{d}F^{*n}(y) \le \sup_{0\le y\le x} |Z_1(y) - Z(y)| F^{*n}(x)\,.$$

The latter tends to 0 as $n \to \infty$. Thus $Z_1(x) = Z(x)$. ☐

Let $z(x)$ be a bounded function and $h > 0$ be a real number. Define

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

and the Riemann sums

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

Definition C.10. *A function $z(x)$ is called* **directly Riemann integrable** *if*

$$-\infty < \varliminf_{h\downarrow 0} \underline{\sigma}(h) = \varlimsup_{h\downarrow 0} \bar{\sigma}(h) < \infty\,.$$

The following lemma gives some criteria for a function to be directly Riemann integrable.

Lemma C.11.

- *The space of directly Riemann integrable functions is a linear space.*
- *If $z(t)$ is monotone and $\int_0^\infty z(t)\,\mathrm{d}t < \infty$ then $z(t)$ is directly Riemann integrable.*
- *If $a(t)$ and $b(t)$ are directly Riemann integrable and $z(t)$ is continuous Lebesgue a.e. such that $a(t) \le z(t) \le b(t)$ then $z(t)$ is directly Riemann integrable.*
- *If $z(t) \ge 0$, $z(t)$ is continuous Lebesgue a.e. and $\bar{\sigma}(h) < \infty$ for some $h > 0$ then is $z(t)$ directly Riemann integrable.*

Proof. See for instance [1, p.69]. ☐

Definition C.12. *A distribution function F of a positive random variable X is called* **arithmetic** *if for some γ one has $\mathbb{P}[X \in \{\gamma, 2\gamma, \dots\}] = 1$. The* **span** *$\gamma$ is the largest number such that the above relation is fulfilled.*

The most important result in renewal theory is a result on the asymptotic behaviour of the solution to (C.2).

Proposition C.13. (Renewal theorem) *If $z(x)$ is a directly Riemann integrable function then the solution $Z(x)$ to the renewal equation* (C.2) *satisfies*

$$\lim_{t\to\infty} Z(t) = \lambda \int_0^\infty z(y)\,\mathrm{d}y$$

if F is non-arithmetic and

$$\lim_{n\to\infty} Z(t+n\gamma) = \gamma\lambda \sum_{j=0}^{\infty} z(t+j\gamma)$$

for $0 \le t < \gamma$ if F is arithmetic with span γ.

Proof. See for instance [40, p.364] or [66, p.218]. □

Example C.14. Assume that F is non-arithmetic and that $\mathbb{E}[T_1^2] < \infty$. We consider the function $Z(t) = \mathbb{E}[T_{N_t+1} - t]$, the expected time till the next occurrence of an event. Consider first $\mathbb{E}[T_{N_t+1} - t \mid T_1 = s]$. If $s \le t$ then there is a renewal process starting at s and thus $\mathbb{E}[T_{N_t+1} - t \mid T_1 = s] = Z(t-s)$. If $s > t$ then $T_{N_t+1} = T_1 = s$ and thus $\mathbb{E}[T_{N_t+1} - t \mid T_1 = s] = s - t$. Hence we get the renewal equation

$$Z(t) = \mathbb{E}[\mathbb{E}[T_{N_t+1} - t \mid T_1]] = \int_t^\infty (s-t)\,\mathrm{d}F(s) + \int_0^t Z(t-s)\,\mathrm{d}F(s)\,.$$

Now

$$z(t) = \int_t^\infty (s-t)\,\mathrm{d}F(s) = \int_t^\infty \int_t^s \mathrm{d}y\,\mathrm{d}F(s) = \int_t^\infty (1 - F(y))\,\mathrm{d}y\,.$$

The function is monotone and

$$\int_0^\infty \int_t^\infty (s-t)\,\mathrm{d}F(s)\,\mathrm{d}t = \int_0^\infty \int_0^s (s-t)\,\mathrm{d}t\,\mathrm{d}F(s) = \frac{1}{2}\mathbb{E}[T_1^2]\,.$$

Thus $z(t)$ is directly Riemann integrable and by the renewal theorem

$$\lim_{t\to\infty} Z(t) = \frac{\lambda}{2}\mathbb{E}\left[T_1^2\right]\,.$$

■

Example C.15. Assume that $\lambda > 0$ and that F is non-arithmetic. What is the asymptotic distribution of the waiting time till the next event? Let $x > 0$ and set $Z(t) = \mathbb{P}[T_{N_t+1} - t > x]$. Conditioning on $T_1 = s$ yields

$$\mathbb{P}[T_{N_t+1} - t > x \mid T_1 = s] = \begin{cases} Z(t-s) & \text{if } s \le t, \\ 0 & \text{if } t < s \le t + x, \\ 1 & \text{if } t + x < s. \end{cases}$$

This gives the renewal equation

$$Z(t) = 1 - F(t+x) + \int_0^t Z(t-s) \, \mathrm{d}F(s) \, .$$

$z(t) = 1 - F(t+x)$ is decreasing in t and its integral is bounded by λ^{-1}. Thus it is directly Riemann integrable. It follows from the renewal theorem that

$$\lim_{t \to \infty} Z(t) = \lambda \int_0^\infty (1 - F(t+x)) \, \mathrm{d}t = \lambda \int_x^\infty (1 - F(t)) \, \mathrm{d}t \, .$$

Thus the stationary distribution of the next event must be the distribution given by (C.1). ■

C.3. Bibliographical Remarks

The theory of point processes can also be found in [19] or [26]. For further literature on renewal theory see also [1] and [40].

D. Brownian Motion

Definition D.1. *A stochastic process $\{W_t\}$ is called a (m, η^2)-***Brownian motion** *if a.s.*

- $W_0 = 0$,
- $\{W_t\}$ *has independent increments,*
- $W_t \sim \mathrm{N}(mt, \eta^2 t)$ *and*
- $\{W_t\}$ *has cadlag paths.*

A $(0,1)$-Brownian motion is called a **standard Brownian motion**.

It can be shown that Brownian motion exists. Moreover, one can proof that a Brownian motion has continuous paths.

From the definition it follows that $\{W_t\}$ has also stationary increments.

Lemma D.2. *A (m, η^2)-Brownian motion has stationary increments.*

Proof. Exercise. $\square$

We construct the following martingales.

Lemma D.3. *Let $\{W_t\}$ be a (m, η^2)-Brownian motion and $r \in \mathbb{R}$. The following processes are martingales.*

i) $\{W_t - mt\}$.

ii) $\{(W_t - mt)^2 - \eta^2 t\}$.

iii) $\{\exp\{r(W_t - mt) - \frac{\eta^2}{2} r^2 t\}\}$.

Proof. i) Using the stationary and independent increments property

$$\mathbb{E}[W_t - mt \mid \mathcal{F}_s] = \mathbb{E}[W_t - W_s] + W_s - mt = W_s - ms\,.$$

ii) It suffices to consider the case $m = 0$.

$$\mathbb{E}[W_t^2 - \eta^2 t \mid \mathcal{F}_s] = \mathbb{E}[(W_t - W_s)^2 + 2(W_t - W_s)W_s + W_s^2 \mid \mathcal{F}_s] - \eta^2 t = W_s^2 - \eta^2 s\,.$$

iii) It suffices to consider the case $m = 0$.

$$\mathbb{E}[\mathrm{e}^{rW_t - \frac{\eta^2}{2} r^2 t} \mid \mathcal{F}_s] = \mathbb{E}[\mathrm{e}^{r(W_t - W_s)}]\mathrm{e}^{rW_s - \frac{\eta^2}{2} r^2 t} = \mathrm{e}^{rW_s - \frac{\eta^2}{2} r^2 s}\,.$$

$\square$

E. Random Walks and the Wiener-Hopf Factorization

Let $\{X_i\}$ be a sequence of iid. random variables. Let $U(x)$ be the distribution function of X_i. Define

$$S_n = \sum_{i=1}^{n} X_i$$

and $S_0 = 0$.

Lemma E.1.

i) *If* $\mathbb{E}[X_1] < 0$ *then* S_n *converges to* $-\infty$ *a.s..*

ii) *If* $\mathbb{E}[X_1] = 0$ *then a.s.*

$$\overline{\lim_{n\to\infty}} S_n = -\underline{\lim_{n\to\infty}} S_n = \infty\,.$$

iii) *If* $\mathbb{E}[X_1] > 0$ *then* S_n *converges to* ∞ *a.s. and there is a strictly positive probability that* $S_n \geq 0$ *for all* $n \in \mathbb{N}$.

Proof. See for instance [40, p.396] or [66, p.233]. □

For the rest of this section we assume $-\infty < \mathbb{E}[X_i] < 0$. Let $\tau_+ = \inf\{n > 0 : S_n > 0\}$ and $\tau_- = \inf\{n > 0 : S_n \leq 0\}$, $H(x) = \mathbb{P}[\tau_+ < \infty, S_{\tau_+} \leq x]$ and $\rho(x) = \mathbb{P}[S_{\tau_-} \leq x]$. Note that τ_- is defined a.s. because $S_n \to -\infty$ as $n \to \infty$, see Lemma E.1. Define $\psi_0(x) = 1\!\!1_{\{x\geq 0\}}$ and

$$\psi_n(x) = \mathbb{P}[S_1 > 0, S_2 > 0, \ldots, S_n > 0, S_n \leq x]$$

for $n \geq 1$. Let

$$\psi(x) = \sum_{n=0}^{\infty} \psi_n(x)\,.$$

Lemma E.2. *We have for* $n \geq 1$

$$\psi_n(x) = \mathbb{P}[S_n > S_j,\, (0 \leq j \leq n-1), S_n \leq x]$$

and therefore

$$\psi(x) = \sum_{n=0}^{\infty} H^{*n}(x)\,.$$

Moreover, $\psi(\infty) = (1 - H(\infty))^{-1}$, $\mathbb{E}[\tau_-] = \psi(\infty)$ *and* $\mathbb{E}[S_{\tau_-}] = \mathbb{E}[\tau_-]\mathbb{E}[X_i]$.

Proof. Let for n fixed

$$S_k^* = S_n - S_{n-k} = \sum_{i=n-k+1}^{n} X_i \,.$$

Then $\{S_k^* : k \le n\}$ follows the same law as $\{S_k : k \le n\}$. Thus

$$\begin{aligned}\mathbb{P}[S_n > S_j,\, (0 \le j \le n-1), S_n \le x] &= \mathbb{P}[S_n^* > S_j^*,\, (0 \le j \le n-1), S_n^* \le x] \\ &= \mathbb{P}[S_n > S_n - S_{n-j},\, (0 \le j \le n-1), S_n \le x] \\ &= \mathbb{P}[S_j > 0,\, (1 \le j \le n), S_n \le x] = \psi_n(x)\,.\end{aligned}$$

Denote by $\tau_{n+1} := \inf\{k > 0 : S_k > S_{\tau_n}\}$ the $n+1$-st ascending ladder time. Note that S_{τ_n} has distribution $H^{*n}(x)$. We found that $\psi_n(x)$ is the probability that S_n is a maximum of the random walk and lies in the interval $(0, x]$. Thus $\psi_n(x)$ is the probability that there is a ladder height at n and $S_n \le x$. We obtain

$$\begin{aligned}\sum_{n=1}^{\infty} \psi_n(x) &= \sum_{n=1}^{\infty} \mathbb{P}[\exists k : \tau_k = n, 0 < S_n \le x] = \sum_{n=1}^{\infty}\sum_{k=1}^{\infty} \mathbb{P}[\tau_k = n, 0 < S_{\tau_k} \le x] \\ &= \sum_{k=1}^{\infty} \mathbb{P}[0 < S_{\tau_k} \le x, \tau_k < \infty] = \sum_{k=1}^{\infty} H^{*k}(x)\,.\end{aligned}$$

Because $\mathbb{E}[X_i] < 0$ we have $H(\infty) < 1$ (Lemma E.1) and $H^{*n}(\infty) = (H(\infty))^n$ from which $\psi(\infty) = (1 - H(\infty))^{-1}$ follows.

For the expected value of τ_-

$$\mathbb{E}[\tau_-] = \sum_{n=0}^{\infty} \mathbb{P}[\tau_- > n] = \sum_{n=0}^{\infty} \psi_n(\infty) = \psi(\infty)\,.$$

Finally let $\tau_-(n)$ denote the n-th descending ladder time and let L_n be then n-th descending ladder height $S_{\tau_-(n-1)} - S_{\tau_-(n)}$. Then

$$\frac{L_1 + \cdots + L_n}{n} = \frac{-S_{\tau_-(n)}}{\tau_-(n)} \frac{\tau_-(n)}{n}$$

from which the last assertion follows by the strong law of large numbers. □

Let now

$$\rho_n(x) = \mathbb{P}[S_1 > 0, S_2 > 0, \ldots, S_{n-1} > 0, S_n \le x]\,.$$

Note that for $x \le 0$

$$\rho(x) = \sum_{n=1}^{\infty} \rho_n(x)\,,$$

that $\psi_1(x) = U(x) - U(0)$ if $x > 0$ and $\rho_1(x) = U(x)$ if $x \le 0$. For $n \ge 1$ one obtains for $x > 0$

$$\psi_{n+1}(x) = \int_0^\infty (U(x-y) - U(-y))\psi_n(\mathrm{d}y)$$

and for $x \le 0$

$$\rho_{n+1}(x) = \int_0^\infty U(x-y)\psi_n(\mathrm{d}y)\,.$$

If we sum over all n we obtain for $x > 0$

$$\psi(x) - 1 = \int_{0-}^\infty (U(x-y) - U(-y))\psi(\mathrm{d}y)$$

and for $x \le 0$

$$\rho(x) = \int_{0-}^\infty U(x-y)\psi(\mathrm{d}y)\,.$$

Note that $\rho(0) = \psi(0) = 1$ and thus for $x \ge 0$

$$\psi(x) = \int_{0-}^\infty U(x-y)\psi(\mathrm{d}y)\,.$$

The last two equations can be written as a single equation

$$\psi + \rho = \psi_0 + \psi * U\,.$$

Then

$$\begin{aligned}\psi - \psi_0 + \rho * H = \psi * H + \rho * H &= H + \psi * H * U = H + \psi * U - U \\ &= H - U + \psi + \rho - \psi_0\end{aligned}$$

from which it follows that

$$U = H + \rho - H * \rho\,. \tag{E.1}$$

The latter is called the **Wiener-Hopf factorization**.

E.1. Bibliographical Remarks

The results of this section are taken from [40].

F. Subexponential Distributions

Definition F.1. *A distribution function F with $F(x) = 0$ for $x < 0$ is called* **subexponential** *if*

$$\lim_{t\to\infty} \frac{1 - F^{*2}(t)}{1 - F(t)} = 2\,.$$

The intuition behind the definition is the following. One often observes that the aggregate sum of claims is determined by a few of the largest claims. The definition can be rewritten as $\mathbb{P}[X_1 + X_2 > x] \sim \mathbb{P}[\max\{X_1, X_2\} > x]$. We will prove in Lemma F.7 below that subexponential is equivalent to $\mathbb{P}[X_1 + X_2 + \cdots + X_n > x] \sim \mathbb{P}[\max\{X_1, X_2, \ldots, X_n\} > x]$ for some $n \ge 2$.

We want to show that the moment generating function of such a distribution does not exist for strictly positive values. We first need the following

Lemma F.2. *If F is subexponential then for all $t \in \mathbb{R}$*

$$\lim_{x\to\infty} \frac{1 - F(x-t)}{1 - F(x)} = 1\,.$$

Proof. Let $t \ge 0$. We have

$$\begin{aligned} \frac{1 - F^{*2}(x)}{1 - F(x)} - 1 &= \frac{F(x) - F^{*2}(x)}{1 - F(x)} \\ &= \int_{0-}^{t} \frac{1 - F(x-y)}{1 - F(x)}\,\mathrm{d}F(y) + \int_{t}^{x} \frac{1 - F(x-y)}{1 - F(x)}\,\mathrm{d}F(y) \\ &\ge F(t) + \frac{1 - F(x-t)}{1 - F(x)}(F(x) - F(t))\,. \end{aligned}$$

Thus

$$1 \le \frac{1 - F(x-t)}{1 - F(x)} \le (F(x) - F(t))^{-1}\Bigl(\frac{1 - F^{*2}(x)}{1 - F(x)} - 1 - F(t)\Bigr)\,.$$

The assertion for $t \ge 0$ follows by letting $x \to \infty$. If $t < 0$ then

$$\lim_{x\to\infty} \frac{1 - F(x-t)}{1 - F(x)} = \lim_{x\to\infty} \frac{1}{\dfrac{1 - F((x-t) - (-t))}{1 - F(x-t)}} = \lim_{y\to\infty} \frac{1}{\dfrac{1 - F(y - (-t))}{1 - F(y)}} = 1$$

where $y = x - t$. □

Lemma F.3. *Let F be subexponential and $r > 0$. Then*

$$\lim_{t\to\infty} \mathrm{e}^{rt}(1 - F(t)) = \infty\,,$$

in particular

$$\int_{0-}^{\infty} \mathrm{e}^{rx}\,\mathrm{d}F(x) = \infty\,.$$

Proof. We first observe that for any $0 < x < t$

$$\mathrm{e}^{rt}(1 - F(t)) = \frac{1 - F(t)}{1 - F(t - x)}(1 - F(t - x))\mathrm{e}^{r(t-x)}\mathrm{e}^{rx} \tag{F.1}$$

and by Lemma F.2 the function $\mathrm{e}^{rn}(1 - F(n))$ ($n \in \mathbb{N}$) is increasing for n large enough. Thus there exist a limit in $(0, \infty]$. Letting $n \to \infty$ in (F.1) shows that this limit can only be infinite.

Let now $\{t_n\}$ be an arbitrary sequence such that $t_n \to \infty$. Then

$$\begin{aligned}\mathrm{e}^{rt_n}(1 - F(t_n)) &= \frac{1 - F(t_n)}{1 - F([t_n])}(1 - F([t_n]))\mathrm{e}^{r[t_n]}\mathrm{e}^{r(t_n-[t_n])}\\ &\ge \frac{1 - F(t_n)}{1 - F(t_n - 1)}(1 - F([t_n]))\end{aligned}$$

which tends to infinity as $n \to \infty$.

The moment generating function can be written as

$$\begin{aligned}\int_{0-}^{\infty} \mathrm{e}^{rx}\,\mathrm{d}F(x) &= 1 + \int_0^\infty\int_0^x r\mathrm{e}^{ry}\,\mathrm{d}y\,\mathrm{d}F(x)\\ &= 1 + r\int_0^\infty\int_y^\infty \mathrm{d}F(x)\mathrm{e}^{ry}\,\mathrm{d}y = 1 + r\int_0^\infty \mathrm{e}^{ry}(1 - F(y))\,\mathrm{d}y = \infty\end{aligned}$$

because the integrand diverges to infinity. □

The following lemma gives a condition for subexponentiality.

Lemma F.4. *Assume that for all $z \in (0, 1]$ the limit*

$$\gamma(z) = \lim_{x\to\infty} \frac{1 - F(zx)}{1 - F(x)}$$

exists and that $\gamma(z)$ is left-continuous at 1. Then F is a subexponential distribution function.

Proof. Note first that $F^{*2}(x) = \mathbb{P}[X_1 + X_2 \le x] \le \mathbb{P}[X_1 \le x, X_2 \le x] = F^2(x)$. For simplicity assume $F(0) = 0$. Hence

$$\varliminf_{x\to\infty} \frac{1 - F^{*2}(x)}{1 - F(x)} \ge \varliminf_{x\to\infty} \frac{1 - F^2(x)}{1 - F(x)} = \varliminf_{x\to\infty} 1 + F(x) = 2\,.$$

Let $n \ge 1$ be fixed. Then

$$\begin{aligned}\varlimsup_{x\to\infty} \frac{1 - F^{*2}(x)}{1 - F(x)} &= 1 + \varlimsup_{x\to\infty} \int_0^x \frac{1 - F(x-y)}{1 - F(x)}\,\mathrm{d}F(y) \\ &\le 1 + \varlimsup_{x\to\infty} \sum_{k=1}^n \frac{1 - F(x - kx/n)}{1 - F(x)}(F(kx/n) - F((k-1)x/n)) \\ &= 1 + \gamma(1 - 1/n)\,.\end{aligned}$$

Because n is arbitrary and $\gamma(z)$ is left continuous at 1 the assertion follows. □

Example F.5. Consider the $\mathrm{Pa}(\alpha, \beta)$ distribution.

$$\frac{1 - F(zx)}{1 - F(x)} = \frac{(\frac{\beta}{\beta + zx})^\alpha}{(\frac{\beta}{\beta + x})^\alpha} = \Bigl(\frac{\beta + x}{\beta + zx}\Bigr)^\alpha \to z^{-\alpha}$$

as $x \to \infty$. It follows that the Pareto distribution is subexponential. ■

Next we give an upper bound for the tails of the convolutions.

Lemma F.6. *Let F be subexponential. Then for any $\varepsilon > 0$ there exist a $D \in \mathbb{R}$ such that*

$$\frac{1 - F^{*n}(t)}{1 - F(t)} \le D(1 + \varepsilon)^n$$

for all $t > 0$ and $n \in \mathbb{N}$.

Proof. Let

$$\alpha_n = \sup_{t\ge 0} \frac{1 - F^{*n}(t)}{1 - F(t)}$$

and note that $1 - F^{*(n+1)}(t) = 1 - F(t) + F * (1 - F^{*n})(t)$. Choose $T \ge 0$ such that

$$\sup_{t\ge T} \frac{F(t) - F^{*2}(t)}{1 - F(t)} < 1 + \frac{\varepsilon}{2}\,.$$

Let $A_T = (1 - F(T))^{-1}$. For $n \ge 1$

$$\begin{aligned}\alpha_{n+1} &\le 1 + \sup_{0\le t\le T} \int_{0-}^t \frac{1 - F^{*n}(t-y)}{1 - F(t)}\,\mathrm{d}F(y) + \sup_{t\ge T} \int_{0-}^t \frac{1 - F^{*n}(t-y)}{1 - F(t)}\,\mathrm{d}F(y) \\ &\le 1 + A_T + \sup_{t\ge T} \int_{0-}^t \frac{1 - F^{*n}(t-y)}{1 - F(t-y)}\,\frac{1 - F(t-y)}{1 - F(t)}\,\mathrm{d}F(y) \\ &\le 1 + A_T + \alpha_n \sup_{t\ge T} \frac{F(t) - F^{*2}(t)}{1 - F(t)} \le 1 + A_T + \alpha_n\Bigl(1 + \frac{\varepsilon}{2}\Bigr)\,.\end{aligned}$$

Choose

$$D = \max\Bigl\{\frac{2(1+A_T)}{\varepsilon}, 1\Bigr\}$$

and note that $\alpha_1 = 1 < D(1+\varepsilon)$. The assertion follows by induction. ☐

Lemma F.7. *Let $F(x) = 0$ for $x < 0$. The following are equivalent:*

i) *F is subexponential.*

ii) *For all $n \in \mathbb{N}$*

$$\lim_{x\to\infty} \frac{1 - F^{*n}(x)}{1 - F(x)} = n\,. \tag{F.2}$$

iii) *There exists $n \ge 2$ such that* (F.2) *holds.*

Proof. "*i) $\Longrightarrow$ ii)*" We proof the assertion by induction. The assertion is trivial for $n = 2$. Assume that it is proved for some $n \in \mathbb{N}$ with $n \ge 2$. Let $\varepsilon > 0$. Choose t such that

$$\Bigl|\frac{1 - F^{*n}(x)}{1 - F(x)} - n\Bigr| < \varepsilon$$

for $x \ge t$.

$$\begin{aligned}\frac{1 - F^{*(n+1)}(x)}{1 - F(x)} = 1 &+ \int_{0-}^{x-t} \frac{1 - F^{*n}(x-y)}{1 - F(x-y)}\,\frac{1 - F(x-y)}{1 - F(x)}\,\mathrm{d}F(y) \\ &+ \int_{x-t}^{x} \frac{1 - F^{*n}(x-y)}{1 - F(x)}\,\mathrm{d}F(y)\,.\end{aligned}$$

The second integral is bounded by

$$\int_{x-t}^{x} \frac{1 - F^{*n}(x-y)}{1 - F(x)}\,\mathrm{d}F(y) \le \frac{F(x) - F(x-t)}{1 - F(x)} = \frac{1 - F(x-t)}{1 - F(x)} - 1$$

which tends to 0 as $x \to \infty$ by Lemma F.2. The expression

$$\int_{0-}^{x-t} n\,\frac{1 - F(x-y)}{1 - F(x)}\,\mathrm{d}F(y) = n\Bigl(\frac{F(x) - F^{*2}(x)}{1 - F(x)} - \int_{x-t}^{x} \frac{1 - F(x-y)}{1 - F(x)}\,\mathrm{d}F(y)\Bigr) \to n$$

as $x \to \infty$ by the same arguments as above. Then

$$\begin{aligned}&\Bigl|\int_{0-}^{x-t} \Bigl(\frac{1 - F^{*n}(x-y)}{1 - F(x-y)} - n\Bigr)\,\frac{1 - F(x-y)}{1 - F(x)}\,\mathrm{d}F(y)\Bigr| \\ &\qquad \le \varepsilon\Bigl(\frac{F(x) - F^{*2}(x)}{1 - F(x)} - \int_{x-t}^{x} \frac{1 - F(x-y)}{1 - F(x)}\,\mathrm{d}F(y)\Bigr) \to \varepsilon\end{aligned}$$

by the arguments used before. Thus

$$\varlimsup_{x\to\infty}\left|\frac{1-F^{*(n+1)}(x)}{1-F(x)}-(n+1)\right|\le\varepsilon\,.$$

This proves the assertion because ε was arbitrary.
"*ii*) $\Longrightarrow$ *iii*)" Trivial.
"*iii*) $\Longrightarrow$ *i*)" Assume now (F.2) for some $n>2$. We show that the assertion also holds for n replaced by $n-1$. Since

$$\frac{1-F^{*n}(x)}{1-F(x)}-1=\int_{0-}^{x}\frac{1-F^{*(n-1)}(x-y)}{1-F(x)}\,\mathrm{d}F(y)\ge F(x)\frac{1-F^{*(n-1)}(x)}{1-F(x)}$$

it follows that

$$\varlimsup_{x\to\infty}\frac{1-F^{*(n-1)}(x)}{1-F(x)}\le n-1\,.$$

Because $F^{*(n-1)}(x)\le F(x)^{n-1}$ it is trivial that

$$\varliminf_{x\to\infty}\frac{1-F^{*(n-1)}(x)}{1-F(x)}\ge n-1\,.$$

It follows that (F.2) holds for $n=2$ and F is a subexponential distribution. □

Lemma F.8. *Let U and V be two distribution functions with $U(x)=V(x)=0$ for all $x<0$. Assume that*

$$1-V(x)\sim a(1-U(x))\qquad\qquad \textit{as } x\to\infty$$

for some $a>0$. If U is subexponential then also V is subexponential.

Proof. We have to show that

$$\varlimsup_{x\to\infty}\int_{0-}^{x}\frac{1-V(x-y)}{1-V(x)}\,\mathrm{d}V(y)\le 1\,.$$

Let $0<\varepsilon<1$. There exists y_0 such that for all $y\ge y_0$

$$(1-\varepsilon)a\le\frac{1-V(y)}{1-U(y)}\le(1+\varepsilon)a\,.$$

Let $x>y_0$. Note that

$$\int_{x-y_0}^{x}\frac{1-V(x-y)}{1-V(x)}\,\mathrm{d}V(y)\le\frac{V(x)-V(x-y_0)}{1-V(x)}=\frac{\frac{1-V(x-y_0)}{1-U(x-y_0)}}{\frac{1-V(x)}{1-U(x)}}\,\frac{1-U(x-y_0)}{1-U(x)}-1\to 0$$

by Lemma F.2. Moreover,

$$\begin{aligned}\int_{0-}^{x-y_0} \frac{1-V(x-y)}{1-V(x)} \,\mathrm{d}V(y) &\le \frac{1+\varepsilon}{1-\varepsilon} \int_{0-}^{x-y_0} \frac{1-U(x-y)}{1-U(x)} \,\mathrm{d}V(y) \\ &\le \frac{1+\varepsilon}{1-\varepsilon} \int_{0-}^{x} \frac{1-U(x-y)}{1-U(x)} \,\mathrm{d}V(y)\,.\end{aligned}$$

The last integral is

$$\begin{aligned}\int_{0-}^{x} \frac{1-U(x-y)}{1-U(x)} \,\mathrm{d}V(y) &= \frac{V(x)-U*V(x)}{1-U(x)} \\ &= 1-\frac{1-V(x)}{1-U(x)} + \frac{U(x)-V*U(x)}{1-U(x)} \\ &= 1-\frac{1-V(x)}{1-U(x)} + \int_{0-}^{x} \frac{1-V(x-y)}{1-U(x)} \,\mathrm{d}U(y)\,.\end{aligned}$$

$1-(1-U(x))^{-1}(1-V(x))$ tends to $1-a$ as $x\to\infty$. And

$$\int_{x-y_0}^{x} \frac{1-V(x-y)}{1-U(x)} \,\mathrm{d}U(y) \le \frac{U(x)-U(x-y_0)}{1-U(x)}$$

tends to 0. Finally

$$\begin{aligned}\int_{0-}^{x-y_0} \frac{1-V(x-y)}{1-U(x)} \,\mathrm{d}U(y) &\le (1+\varepsilon)a \int_{0-}^{x-y_0} \frac{1-U(x-y)}{1-U(x)} \,\mathrm{d}U(y) \\ &\le (1+\varepsilon)a \int_{0-}^{x} \frac{1-U(x-y)}{1-U(x)} \,\mathrm{d}U(y)\end{aligned}$$

which tends to $(1+\varepsilon)a$. Putting all these limits together we obtain

$$\overline{\lim_{x\to\infty}} \int_{0-}^{x} \frac{1-V(x-y)}{1-V(x)} \,\mathrm{d}V(y) \le \frac{(1+\varepsilon)(1+a\varepsilon)}{1-\varepsilon}\,.$$

Because ε was arbitrary the assertion follows. □

F.1. Bibliographical Remarks

The class of subexponential distributions was introduced by Chistyakov [23]. The results presented here can be found in [23], [13] and [34].

G. Concave and Convex Functions

In this appendix we let I be an interval, finite or infinite, but not a singleton.

Definition G.1. *A function* $u : I \to \mathbb{R}$ *is called* **(strictly) concave** *if for all* $x, z \in I$, $x \neq z$, *and all* $\alpha \in (0,1)$ *one has*

$$u((1-\alpha)x + \alpha z) \geq (>) \, (1-\alpha)u(x) + \alpha u(z)\,.$$

u *is called* **(strictly) convex** *if* $-u$ *is (strictly) concave.*

Because results on concave functions can easily translated for convex functions we will only consider concave functions in the sequel.

Concave functions have nice properties.

Lemma G.2. *A concave function* $u(y)$ *is continuous, differentiable from the left and from the right. The derivative is decreasing, i.e. for* $x < y$ *we have* $u'(x-) \geq u'(x+) \geq u'(y-) \geq u'(y+)$. *If* $u(y)$ *is strictly concave then* $u'(x+) > u'(y-)$.

Remark. The theorem implies that $u(y)$ is differentiable almost everywhere. ■

Proof. Let $x < y < z$. Then

$$u(y) = u\Bigl(\frac{z-y}{z-x}x + \frac{y-x}{z-x}z\Bigr) \geq \frac{z-y}{z-x}u(x) + \frac{y-x}{z-x}u(z)$$

or equivalently

$$(z-x)u(y) \geq (z-y)u(x) + (y-x)u(z)\,. \tag{G.1}$$

This implies immediately

$$\frac{u(y)-u(x)}{y-x} \geq \frac{u(z)-u(x)}{z-x} \geq \frac{u(z)-u(y)}{z-y}\,. \tag{G.2}$$

Thus the function $h \mapsto h^{-1}(u(y) - u(y-h))$ is increasing in h and bounded from below by $(z-y)^{-1}(u(z) - u(y))$. Thus the derivative $u'(y-)$ from the left exists. Analogously, the derivative from the right $u'(y+)$ exists. The assertion in the concave case follows now from (G.2). The strict inequality in the strictly concave case follows analogously. □

Concave functions have also the following property.

Lemma G.3. *Let $u(y)$ be a concave function. There exists a function $k : I \to \mathbb{R}$ such that for any $y, x \in I$*

$$u(x) \le u(y) + k(y)(x - y)\,. \tag{G.3}$$

Moreover, the function $k(y)$ is decreasing. If $u(y)$ is strictly concave then the above inequality is strict for $x \ne y$ and $k(y)$ is strictly decreasing. Conversely, if a function $k(y)$ exists such that (G.3) is fulfilled, then $u(y)$ is concave, strictly concave if the strict inequality holds for $x \ne y$.

Proof. Left as an exercise. $\square$

Corollary G.4. *Let $u(y)$ be a twice differentiable function. Then $u(y)$ is concave if and only if its second derivative is negative. It is strictly concave if and only if its second derivative is strictly negative almost everywhere.*

Proof. This follows readily from Theorem G.2 and Lemma G.3. $\square$

The following result is very useful.

Theorem G.5. (Jensen's inequality) *The function $u(y)$ is (strictly) concave if and only if*

$$\mathbb{E}[u(Y)] \le (<)\, u(\mathbb{E}[Y]) \tag{G.4}$$

for all I-valued integrable random variables Y with $\mathbb{P}[Y \ne \mathbb{E}[Y]] > 0$.

Proof. Assume (G.4) for all random variables Y. Let $\alpha \in (0, 1)$. Let $\mathbb{P}[Y = z] = 1 - \mathbb{P}[Y = x] = \alpha$. Then the (strict) concavity follows. Assume $u(y)$ is strictly concave. Then it follows from Lemma G.3 that

$$u(Y) \le u(\mathbb{E}[Y]) + k(\mathbb{E}[Y])(Y - \mathbb{E}[Y])\,.$$

The strict inequality holds if $u(y)$ is strictly concave and $Y \ne \mathbb{E}[Y]$. Taking expected values gives (G.4). $\square$

Also the following result is often useful.

Theorem G.6. (Ohlin's lemma) *Let $F_i(y)$, $i = 1, 2$ be two distribution functions defined on I. Assume*

$$\int_I y \,\mathrm{d}F_1(y) = \int_I y \,\mathrm{d}F_2(y) < \infty$$

and that there exists $y_0 \in I$ such that

$$F_1(y) \le F_2(y), \quad y < y_0, \qquad\qquad F_1(y) \ge F_2(y), \quad y > y_0\,.$$

Then for any concave function $u(y)$

$$\int_I u(y) \,\mathrm{d}F_1(y) \ge \int_I u(y) \,\mathrm{d}F_2(y)$$

provided the integrals are well defined. If $u(y)$ is strictly concave and $F_1 \ne F_2$ then the inequality holds strictly.

Proof. Recall the formulae $\int_0^\infty y \,\mathrm{d}F_i(y) = \int_0^\infty (1 - F_i(y)) \,\mathrm{d}y$ and $\int_{-\infty}^0 y \,\mathrm{d}F_i(y) = -\int_{-\infty}^0 F_i(y) \,\mathrm{d}y$, which can be proved for example by Fubini's theorem. Thus it follows that $\int_I (F_2(y) - F_1(y)) \,\mathrm{d}y = 0$. We know that $u(y)$ is differentiable almost everywhere and continuous. Thus $u(y) = u(y_0) + \int_{y_0}^y u'(z) \,\mathrm{d}z$, where we can define $u'(y)$ as the right derivative. This yields

$$\begin{aligned}\int_{-\infty}^{y_0} u(y) \,\mathrm{d}F_i(y) &= u(y_0)F_i(y_0) - \int_{-\infty}^{y_0}\int_y^{y_0} u'(z) \,\mathrm{d}z \,\mathrm{d}F_i(y) \\ &= u(y_0)F_i(y_0) - \int_{-\infty}^{y_0} F_i(z)u'(z) \,\mathrm{d}z\,.\end{aligned}$$

Analogously

$$\int_{y_0}^\infty u(y) \,\mathrm{d}F_i(y) = u(y_0)(1 - F_i(y_0)) + \int_{y_0}^\infty (1 - F_i(z))u'(z) \,\mathrm{d}z\,.$$

Putting the results together we find

$$\int_{-\infty}^\infty u(y) \,\mathrm{d}F_1(y) - \int_{-\infty}^\infty u(y) \,\mathrm{d}F_2(y) = \int_{-\infty}^\infty (F_2(y) - F_1(y))u'(y) \,\mathrm{d}y\,.$$

If $y < y_0$ then $F_2(y) - F_1(y) \ge 0$ and $u'(y) \ge u'(y_0)$. If $y > y_0$ then $F_2(y) - F_1(y) \le 0$ and $u'(y) \le u'(y_0)$. Thus

$$\int_{-\infty}^\infty u(y) \,\mathrm{d}F_1(y) - \int_{-\infty}^\infty u(y) \,\mathrm{d}F_2(y) \ge \int_{-\infty}^\infty (F_2(y) - F_1(y))u'(y_0) \,\mathrm{d}y = 0\,.$$

The strictly concave case follows analogously. □

Corollary G.7. *Let X be a real random variable taking values in some interval I_1, and let $g_i : I_1 \to I_2$, $i = 1, 2$ be increasing functions with values on some interval I_2. Suppose*

$$\mathbb{E}[g_1(X)] = \mathbb{E}[g_2(X)] < \infty\,.$$

Let $u : I_2 \to \mathbb{R}$ be a concave function such that $\mathbb{E}[u(g_i(X))]$ is well-defined. If there exists x_0 such that

$$g_1(x) \geq g_2(x), \quad x < x_0, \qquad\qquad g_1(x) \leq g_2(x), \quad x > x_0,$$

then

$$\mathbb{E}[u(g_1(X))] \geq \mathbb{E}[u(g_2(X))]\,.$$

Moreover, if $u(y)$ is strictly concave and $\mathbb{P}[g_1(X) \neq g_2(X)] > 0$ then the inequality is strict.

Proof. Choose $F_i(y) = \mathbb{P}[g_i(X) \leq y]$. Let $y_0 = g_1(x_0)$. If $y < y_0$ then

$$F_1(y) = \mathbb{P}[g_1(X) \leq y] = \mathbb{P}[g_1(X) \leq y, X < x_0] \leq \mathbb{P}[g_2(X) \leq y, X < x_0] \leq F_2(y)\,.$$

If $y > y_0$ then

$$1 - F_1(y) = \mathbb{P}[g_1(X) > y, X > x_0] \leq \mathbb{P}[g_2(X) > y, X > x_0] \leq 1 - F_2(y)\,.$$

The result follows now from Theorem G.6. $\square$

Table of Distribution Functions

Exponential distribution: Exp(α)

Parameters:	$\alpha > 0$	
Distribution function:	$F(x)$	$= 1 - \mathrm{e}^{-\alpha x}$
Density :	$f(x)$	$= \alpha \mathrm{e}^{-\alpha x}$
Expected value:	$\mathbb{E}[X]$	$= \alpha^{-1}$
Variance:	$\mathrm{Var}(X)$	$= \alpha^{-2}$
Moment generating function:	$M_X(r)$	$= \alpha(\alpha - r)^{-1}$

Pareto distribution: Pa(α, β)

Parameters:	$\alpha > 0,\ \beta > 0$	
Distribution function:	$F(x)$	$= 1 - \beta^{\alpha}(\beta + x)^{-\alpha}$
Density :	$f(x)$	$= \alpha\beta^{\alpha}(\beta + x)^{-\alpha-1}$
Expected value:	$\mathbb{E}[X]$	$= \beta(\alpha - 1)^{-1}$ if $\alpha > 1$
Variance:	$\mathrm{Var}(X)$	$= \alpha\beta^2(\alpha - 1)^{-2}(\alpha - 2)^{-1}$ if $\alpha > 2$
Moment generating function:	$M_X(r)$	$=$ — —

Weibull distribution: Wei(α, c)

Parameters:	$\alpha > 0,\ c > 0$	
Distribution function:	$F(x)$	$= 1 - \mathrm{e}^{-cx^{\alpha}}$
Density :	$f(x)$	$= \alpha c x^{\alpha-1} \mathrm{e}^{-cx^{\alpha}}$
Expected value:	$\mathbb{E}[X]$	$= c^{-1/\alpha}\Gamma(1 + \frac{1}{\alpha})$
Variance:	$\mathrm{Var}(X)$	$= c^{-2/\alpha}(\Gamma(1 + \frac{2}{\alpha}) - \Gamma(1 + \frac{1}{\alpha})^2)$
Moment generating function:	$M_X(r)$	$=$ — —

Gamma distribution: $\Gamma(\gamma, \alpha)$

Parameters:	$\gamma > 0,\ \alpha > 0$	
Distribution function:	$F(x)$	$=$ — —
Density :	$f(x)$	$= \alpha^{\gamma}(\Gamma(\gamma))^{-1} x^{\gamma-1} \mathrm{e}^{-\alpha x}$
Expected value:	$\mathbb{E}[X]$	$= \gamma\alpha^{-1}$
Variance:	$\mathrm{Var}(X)$	$= \gamma\alpha^{-2}$
Moment generating function:	$M_X(r)$	$= (\alpha(\alpha - r)^{-1})^{\gamma}$

Loggamma distribution: LG(γ, α)

Parameters:	$\gamma > 0,\ \alpha > 0$	
Distribution function:	$F(x)$	$=$ — —
Density :	$f(x)$	$= \alpha^{\gamma}(\Gamma(\gamma))^{-1} x^{-\alpha-1}(\log x)^{\gamma-1} \quad (x > 1)$
Expected value:	$\mathbb{E}[X]$	$= (\alpha(\alpha - 1)^{-1})^{\gamma}$ if $\alpha > 1$
Variance:	$\mathrm{Var}(X)$	$= (\alpha/(\alpha - 2))^{\gamma} - (\alpha/(\alpha - 1))^{2\gamma}$, $\alpha > 2$
Moment generating function:	$M_X(r)$	$=$ — —

Normal distribution: $\mathbf{N}(\mu, \sigma^2)$

Parameters:	$\mu \in \mathbb{R},\ \sigma^2 > 0$	
Distribution function:	$F(x)$	$= \Phi(\sigma^{-1}(x - \mu))$
Density :	$f(x)$	$= (2\pi\sigma^2)^{-1/2}\mathrm{e}^{-(x-\mu)^2/(2\sigma^2)}$
Expected value:	$\mathbb{E}[X]$	$= \mu$
Variance:	$\mathrm{Var}(X)$	$= \sigma^2$
Moment generating function:	$M_X(r)$	$= \exp\{(\sigma^2/2)r^2 + r\mu\}$

Lognormal distribution: $\mathbf{LN}(\mu, \sigma^2)$

Parameters:	$\mu \in \mathbb{R},\ \sigma^2 > 0$	
Distribution function:	$F(x)$	$= \Phi(\sigma^{-1}(\log x - \mu))$
Density :	$f(x)$	$= (2\pi\sigma^2)^{-1/2}x^{-1}\mathrm{e}^{-(\log x-\mu)^2/(2\sigma^2)}$
Expected value:	$\mathbb{E}[X]$	$= \mathrm{e}^{(\sigma^2/2)+\mu}$
Variance:	$\mathrm{Var}(X)$	$= \mathrm{e}^{\sigma^2+2\mu}(\mathrm{e}^{\sigma^2} - 1)$
Moment generating function:	$M_X(r)$	$=$ — —

Binomial distribution: $\mathbf{B}(k, p)$

Parameters:	$k \in \mathbb{N},\ p \in (0, 1)$	
Probabilities:	$\mathbb{P}[X = n]$	$= \binom{k}{n} p^n (1-p)^{k-n}$
Expected value:	$\mathbb{E}[X]$	$= kp$
Variance:	$\mathrm{Var}(X)$	$= kp(1-p)$
Moment generating function:	$M_X(r)$	$= (p\mathrm{e}^r + 1 - p)^k$

Poisson distribution: $\mathbf{Pois}(\lambda)$

Parameters:	$\lambda > 0$	
Probabilities:	$\mathbb{P}[X = n]$	$= (n!)^{-1}\lambda^n \mathrm{e}^{-\lambda}$
Expected value:	$\mathbb{E}[X]$	$= \lambda$
Variance:	$\mathrm{Var}(X)$	$= \lambda$
Moment generating function:	$M_X(r)$	$= \exp\{\lambda(\mathrm{e}^r - 1)\}$

Negative binomial distribution: $\mathbf{NB}(\alpha, p)$

Parameters:	$\alpha > 0,\ p \in (0, 1)$	
Probabilities:	$\mathbb{P}[X = n]$	$= \binom{\alpha+n-1}{n} p^\alpha (1-p)^n$
Expected value:	$\mathbb{E}[X]$	$= \alpha(1-p)/p$
Variance:	$\mathrm{Var}(X)$	$= \alpha(1-p)/p^2$
Moment generating function:	$M_X(r)$	$= p^\alpha(1 - (1-p)\mathrm{e}^r)^{-\alpha}$

References

[1] **Alsmeyer, G.** (1991). *Erneuerungstheory*. Teubner, Stuttgart.

[2] **Ammeter, H.** (1948). A generalization of the collective theory of risk in regard to fluctuating basic probabilities. *Skand. Aktuar Tidskr.*, 171–198.

[3] **Andersen, E.Sparre** (1957). *On the collective theory of risk in the case of contagion between the claims*. Transactions XVth International Congress of Actuaries, New York, II, 219–229.

[4] **Arfwedson, G.** (1954). Research in collective risk theory. Part 1. *Skand. Aktuar Tidskr.*, 191–223.

[5] **Arfwedson, G.** (1955). Research in collective risk theory. Part 2. *Skand. Aktuar Tidskr.*, 53–100.

[6] **Asmussen, S.** (1982). Conditioned limit theorems relating a random walk to its associate, with applications to risk reserve processes and the GI/G/1 Queue. *Adv. in Appl. Probab.* **14**, 143–170.

[7] **Asmussen, S.** (1984). Approximations for the probability of ruin within finite time. *Scand. Actuarial J.*, 31–57.

[8] **Asmussen, S.** (1989). Risk theory in a Markovian environment. *Scand. Actuarial J.*, 66–100.

[9] **Asmussen, S.** (2000). *Ruin Probabilities*. World Scientific, Singapore.

[10] **Asmussen, S., Fløe Henriksen, F. and Klüppelberg, C.** (1994). Large claims approximations for risk processes in a Markovian environment. *Stochastic Process. Appl.* **54**, 29–43.

[11] **Asmussen, S. and Nielsen, H.M.** (1995). Ruin probabilities via local adjustment coefficients. *J. Appl. Probab.*, 736–755.

[12] **Asmussen, S., Schmidli, H. and Schmidt, V.** (1999). Tail probabilities for non-standard risk and queueing processes with subexponential jumps. *Adv. in Appl. Probab.* **31**, to appear.

[13] **Athreya, K.B. and Ney, P.** (1972). *Branching Processes*. Springer-Verlag, Berlin.

[14] **Bailey, A.L.** (1945). A generalised theory of credibility. *Proceedings of the Casualty Actuarial Society* **32**, 13–20.

[15] **Bailey, A.L.** (1950). Credibility procedures, La Place's generalisation of Bayes' rule, and the combination of collateral knowledge with observed data. *Proceedings of the Casualty Actuarial Society* **37**, 7–23 and 94–115.

[16] **Barndorff-Nielsen, O.E. and Schmidli, H.** (1994). Saddlepoint approximations for the probability of ruin in finite time. *Scand. Actuarial J.*, 169–186.

[17] **Beekman, J.** (1969). A ruin function approximation. *Trans. Soc. Actuaries* **21**, ,. 41–48 and 275–279

[18] **Björk, T. and Grandell, J.** (1988). Exponential inequalities for ruin probabilities in the Cox case. *Scand. Actuarial J.*, 77–111.

[19] **Brémaud, P.** (1981). *Point Processes and Queues*. Springer-Verlag, New York.

[20] **Bühlmann, H.** (1967). Experience rating and credibility. *ASTIN Bulletin* **4**, 199–207.

[21] **Bühlmann, H. and Straub, E.** (1970). Glaubwürdigkeit für Schadensätze. *Schweiz. Verein. Versicherungsmath. Mitt.* **70**, 111–133.

[22] **Canteno, L.** (1986). Measuring the effects of reinsurance by the adjustment coefficient. *Insurance Math. Econom.* **5**, 169–182.

[23] **Chistyakov, V.P.** (1964). A theorem on sums of independent, positive random variables and its applications to branching processes. *Theory Probab. Appl.* **9**, 640–648.

[24] **Cramér, H.** (1930). *On the Mathematical Theory of Risk*. Skandia Jubilee Volume, Stockholm.

[25] **Cramér, H.** (1955). *Collective Risk Theory*. Skandia Jubilee Volume, Stockholm.

[26] **Daley, D.J. and Vere-Jones, D.** (1988). *An Introduction to the Theory of Point Processes*. Springer-Verlag, New York.

[27] **Dassios, A.** (1987). Ph.D. Thesis. *Imperial College, London.*

[28] **Dassios, A. and Embrechts, P.** (1989). Martingales and insurance risk. *Commun. Statist. Stochastic Models* **5**, 181–217.

[29] **Deheuvels, P., Haeusler, E. and Mason, D.M.** (1988). Almost sure convergence of the Hill estimator. *Math. Proc. Camb. Phil. Soc.* **104**, 371–381.

[30] **Dellacherie, C. and Meyer, P.A.** (1980). *Probabilités et Potentiel*. Ch. VI, Hermann, Paris.

[31] **DeVylder, F.** (1978). A practical solution to the problem of ultimate ruin probability. *Scand. Actuarial J.*, 114–119.

[32] **Dickson, D.C.M.** (1992). On the distribution of the surplus prior to ruin. *Insurance Math. Econom.* **11**, 191–207.

[33] **Dufresne, F. and Gerber, H.U.** (1988). The surpluses immediately before and at ruin, and the amount of the claim causing ruin. *Insurance Math. Econom.* **7**, 193–199.

[34] **Embrechts, P., Goldie, C.M. and Veraverbeke, N.** (1979). Subexponentiality and infinite divisibility. *Z. Wahrsch. verw. Geb.* **49**, 335–347.

[35] **Embrechts, P., Grandell, J. and Schmidli, H.** (1993). Finite-time Lundberg inequalities in the Cox case. *Scand. Actuarial J.*, 17–41.

[36] **Embrechts, P., Klüppelberg, C. and Mikosch, T.** (1997). *Modelling Extremal Events.* Springer-Verlag, Berlin.

[37] **Embrechts, P. and Schmidli, H.** (1994). Modelling of extremal events in insurance and finance. *ZOR—Math. Methods Oper. Res.* **39**, 1–33.

[38] **Embrechts, P. and Veraverbeke, N.** (1982). Estimates for the probability of ruin with special emphasis on the possibility of large claims. *Insurance Math. Econom.* **1**, 55–72.

[39] **Ethier, S.N. and Kurtz, T.G.** (1986). *Markov Processes.* Wiley, New York.

[40] **Feller, W.** (1971). *An Introduction to Probability Theory and its Applications.* Volume II, Wiley, New York.

[41] **Gerber, H.U.** (1973). Martingales in risk theory. *Schweiz. Verein. Versicherungsmath. Mitt.* **73**, 205–216.

[42] **Gerber, H.U.** (1975). The surplus process as a fair game – utilitywise. *ASTIN Bulletin* **8**, 307–322.

[43] **Gerber, H.U.** (1979). *An Introduction to Mathematical Risk Theory.* Huebner Foundation Monographs, Philadelphia.

[44] **Gerber, H.U., Goovaerts, M.J. and Kaas, R.** (1987). On the probability and severity of ruin. *ASTIN Bulletin* **17**, 151–163.

[45] **Grandell, J.** (1977). A class of approximations of ruin probabilities. *Scand. Actuarial J.*, 37–52.

[46] **Grandell, J.** (1991). *Aspects of Risk Theory.* Springer-Verlag, New York.

[47] **Grandell, J.** (1991). Finite time ruin probabilities and martingales. *Informatica* **2**, 3–32.

[48] **Grandell, J.** (1995). Some remarks on the Ammeter risk process. *Schweiz. Verein. Versicherungsmath. Mitt.* **95**, 43–71.

[49] **Grandell, J.** (1997). *Mixed Poisson Processes.* Chapman & Hall, London.

[50] **Gut, A.** (1988). *Stopped Random Walks.* Springer-Verlag, New York.

[51] **Hadwiger, H.** (1940). Über die Wahrscheinlichkeit des Ruins bei einer grossen Zahl von Geschäften. *Archiv für mathematische Wirtschafts- und Sozialforschung* **6**, 131–135.

[52] **Hall, P.** (1982). On some simple estimates of an exponent of regular variation. *J. R. Statist. Soc. B* **44**, 37–42.

[53] **Hill, B.M.** (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3**, 1163–1174.

[54] **Hogg, R.V. and Klugman, S.A.** (1984). *Loss Distributions.* Wiley, New York.

[55] **Iglehart, D.L.** (1969). Diffusion approximations in collective risk theory. *J. Appl. Probab.* **6**, 285–292.

[56] **Janssen, J.** (1980). Some transient results on the M/SM/1 special semi-Markov model in risk and queueing theories. *ASTIN Bulletin* **11**, 41–51.

[57] **Jung, J.** (1982). Association of Swedish Insurance Companies. *Statistical Department, Stockholm.*

[58] **Karlin, S. and Taylor, H.M.** (1975). *A First Course in Stochastic Processes.* Academic Press, New York.

[59] **Karlin, S. and Taylor, H.M.** (1981). *A Second Course in Stochastic Processes.* Academic Press, New York.

[60] **Klüppelberg, C.** (1988). Subexponential distributions and integrated tails. *J. Appl. Probab.* **25**, 132–141.

[61] **Lundberg, F.** (1903). *I. Approximerad Framställning av Sannolikhetsfunktionen. II. Återförsäkring av Kollektivrisker.* Almqvist & Wiksell, Uppsala.

[62] **Lundberg, F.** (1926). *Försäkringsteknisk Riskutjämning.* F. Englunds boktryckeri A.B., Stockholm.

[63] **Norberg, R.** (1979). The credibility approach to experience rating. *Scand. Actuarial J.*. 181–221

[64] **Panjer, H.H.** (1981). Recursive evaluation of a family of compound distributions. *ASTIN Bulletin* **12**, 22–26.

[65] **Panjer, H.H. and Willmot, G.E.** (1992). *Insurance Risk Models.* Society of Actuaries.

[66] **Rolski, T., Schmidli, H., Schmidt, V. and Teugels, J.L.** (1999). *Stochastic Processes for Insurance and Finance.* Wiley, Chichester.

[67] **Ross, S.M.** (1983). *Stochastic Processes.* Wiley, New York.

[68] **Rytgaard, M.** (1996). Simulation experiments on the mean residual life function $m(x)$. *Proceedings of the XXVII ASTIN Colloquium, Copenhagen, Denmark,* vol. 1, 59–81.

[69] **Schmidli, H.** (1992). A general insurance risk model. *Ph.D Thesis, ETH Zürich.*

[70] **Schmidli, H.** (1994). Diffusion approximations for a risk process with the possibility of borrowing and investment. *Comm. Statist. Stochastic Models* **10**, 365–388.

[71] **Schmidli, H.** (1997). An extension to the renewal theorem and an application to risk theory. *Ann. Appl. Probab.* **7**, 121–133.

[72] **Schmidli, H.** (1998). On the distribution of the surplus prior and at ruin. *ASTIN Bull.*, to appear.

[73] **Schnieper, R.** (1990). Insurance premiums, the insurance market and the need for reinsurance. *Schweiz. Verein. Versicherungsmath. Mitt.* **90**, 129–147.

[74] **Seal, H.L.** (1972). Numerical calculation of the probability of ruin in the Poisson/exponential case. *Schweiz. Verein. Versicherungsmath. Mitt.* **72**, 77–100.

[75] **Seal, H.L.** (1974). The numerical calculation of $U(w, t)$, the probability of non-ruin in an interval $(0, t)$. *Scand. Actuarial J.*, 121-139.

[76] **Segerdahl, C.-O.** (1955). When does ruin occur in the collective theory of risk?. *Skand. Aktuar Tidskr.*, 22–36.

[77] **Siegmund, D.** (1979). Corrected diffusion approximation in certain random walk problems. *Adv. in Appl. Probab.* **11**, 701–719.

[78] **Sundt, B.** (1999). *An Introduction to Non-Life Insurance Mathematics*. Verlag Versicherungswirtschaft e.V., Karlsruhe.

[79] **Stam, A.J.** (1973). Regular variation of the tail of a subordinated probability distribution. *Adv. Appl. Probab.* **5**, 308–327.

[80] **Takács, L.** (1962). *Introduction to the Theory of Queues*. Oxford University Press, New York.

[81] **Thorin, O.** (1974). On the asymptotic behavior of the ruin probability for an infinite period when the epochs of claims form a renewal process. *Scand. Actuarial J.*, 81–99.

[82] **Thorin, O.** (1975). Stationarity aspects of the Sparre Andersen risk process and the corresponding ruin probabilities. *Scand. Actuarial J.*, 87–98.

[83] **Thorin, O.** (1982). Probabilities of ruin. *Scand. Actuarial J.*, 65–102.

[84] **Thorin, O. and Wikstad, N.** (1977). Calculation of ruin probabilities when the claim distribution is lognormal. *ASTIN Bulletin* **9**, 231–246.

[85] **von Bahr, B.** (1975). Asymptotic ruin probabilities when exponential moments do not exist. *Scand. Actuarial J.*, 6–10.

[86] **Waters, H.R.** (1983). Some mathematical aspects of reinsurance insurance. *Insurance Math. Econom.* **2**, 17–26.

[87] **Whitt, W.** (1970). Weak convergence of probability measures on the function space $C[0,\infty)$. *Ann. Statist.* **41**, 939–944.

[88] **Williams, D.** (1979). *Diffusions, Markov Processes and Martingales.* Volume I, Wiley, New York.

Index