Chapter I

Introduction

1 The risk process

In this chapter, we give a very brief summary of some of the models, results and topics to be studied in the rest of the book, and some terminology is introduced.

A *risk reserve process* $\{R_t\}_{t\geq 0}$, as defined in broad terms, is a model for the time evolution of the reserves of an insurance company. We denote throughout the initial reserve by $u = R_0$. The probability $\psi(u)$ of ultimate ruin is the probability that the reserve ever drops below zero,

$$\psi(u) \;=\; \mathbb{P}\Bigl(\inf_{t\geq 0} R_t < 0\Bigr) \;=\; \mathbb{P}\Bigl(\inf_{t\geq 0} R_t < 0 \Bigm| R_0 = u\Bigr).$$

The probability of ruin before time T is

$$\psi(u,T) \quad=\quad \mathbb{P}\Bigl(\inf_{0\leq t\leq T} R_t < 0\Bigr).$$

We also refer to $\psi(u)$ and $\psi(u,T)$ as ruin probabilities with infinite horizon and finite horizon, respectively. *They are the main topics of study of the present book.*

For mathematical purposes, it is frequently more convenient to work with the *claim surplus process* (also called *aggregate loss process*) $\{S_t\}_{t\geq 0}$ defined by $S_t = u - R_t$. Letting

$$\tau(u) \;=\; \inf\{t\geq 0 : R_t < 0\} \;=\; \inf\{t \geq 0 : S_t > u\}\,, \tag{1.1}$$

$$M \;=\; \sup_{0\leq t<\infty} S_t, \qquad M_T \;=\; \sup_{0\leq t\leq T} S_t, \tag{1.2}$$

be the time to ruin and the maxima with infinite and finite horizon, respectively, the ruin probabilities can then alternatively be written as

$$\psi(u) = \mathbb{P}(\tau(u) < \infty) = \mathbb{P}(M > u), \tag{1.3}$$

$$\psi(u,T) = \mathbb{P}(\tau(u) \le T) = \mathbb{P}(M_T > u). \tag{1.4}$$

So far we have not imposed any assumptions on the risk reserve process. However, the following set-up will cover a main part of the book:

- There are only finitely many claims in finite time intervals. That is, the number N_t of arrivals in $[0,t]$ is finite. We denote the interarrival times of claims by $T_2, T_3, \dots$ and T_1 is the time of the first claim. Thus, the time of arrival of the nth claim is $\sigma_n = T_1 + \cdots + T_n$, and $N_t = \min\{n \ge 0 : \sigma_{n+1} > t\} = \max\{n \ge 0 : \sigma_n \le t\}$.
- The size of the nth claim is denoted by U_n.
- Premiums flow in at rate p, say, per unit time.

Putting things together, we see that

$$R_t = u + pt - \sum_{k=1}^{N_t} U_k, \quad S_t = \sum_{k=1}^{N_t} U_k - pt. \tag{1.5}$$

The sample paths of $\{R_t\}$ and $\{S_t\}$ and the connection between the two processes are illustrated in Fig. I.1.

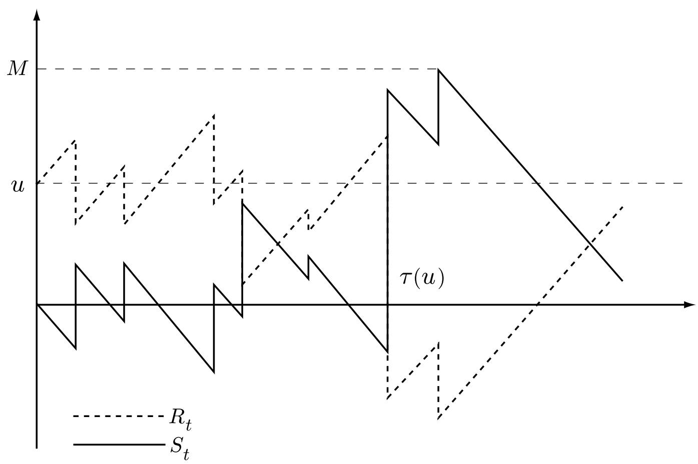

FIGURE I.1

Note that it is a matter of taste (or mathematical convenience) whether one allows $\{R_t\}$ and/or $\{S_t\}$ to continue its evolution after the time $\tau(u)$ of ruin. Thus, for example, one could well replace R_t by $R_{t\wedge\tau(u)}$ or $R_{t\wedge\tau(u)} \vee 0$. For the purpose of studying ruin probabilities this distinction is, of course, immaterial.

Some main examples of models not incorporated in the above set-up are:

- Models which are non-homogeneous in space, for example with a premium depending on the reserve (i.e. on Fig. I.1 the slope of $\{R_t\}$ should depend also on the level). We study this case in Chapter VIII.

- Brownian motion or more general diffusions. Traditionally, Brownian motion has mainly been used as an approximation to the risk process rather than as a model of intrinsic merit and we look at this in Chapter V. However, since any modeling involves some approximative assumptions, it has (partly inspired from the modeling in mathematical finance) become more and more common to use Brownian motion as an intrinsically reasonable model.

- General Lévy processes (defined as continuous time processes with stationary independent increments) where the jump component has infinite Lévy measure, allowing a countable infinity of jumps on Fig. I.1. We treat Lévy processes in Chapter XI.

The models we consider will often have the property that there exists a constant ρ such that

$$\frac{1}{t}\sum_{k=1}^{N_t} U_k \overset{\text{a.s.}}{\to} \rho, \quad t \to \infty. \tag{1.6}$$

The interpretation of ρ is as the average amount of claim per unit time. A further basic quantity is the *safety loading* (or the *security loading*) η defined as the relative amount by which the premium rate p exceeds ρ,

$$\eta = \frac{p-\rho}{\rho}.$$

It is sometimes stated in the theoretical literature that the typical values of the safety loading η are relatively small, say $10\% - 20\%$; we shall, however, not discuss whether this actually corresponds to practice. It would appear obvious, however, that the insurance company should try to ensure $\eta > 0$, and in fact:

Proposition 1.1 *Assume that* (1.6) *holds. If $\eta < 0$, then $M = \infty$ a.s. and hence $\psi(u) = 1$ for all u. If $\eta > 0$, then $M < \infty$ a.s. and hence $\psi(u) < 1$ for all sufficiently large u.*

Proof. It follows from (1.6) that

$$\frac{S_t}{t} \;=\; \frac{\sum_{k=1}^{N_t} U_k - pt}{t} \;\overset{\text{a.s.}}{\to}\; \rho - p, \quad t \to \infty.$$

If $\eta < 0$, then this limit is > 0 which implies $S_t \overset{\text{a.s.}}{\to} \infty$ and hence $M = \infty$ a.s. If $\eta > 0$, then similarly $\lim S_t/t < 0$, $S_t \overset{\text{a.s.}}{\to} -\infty$, $M < \infty$ a.s. □

In concrete models, we obtain typically a somewhat stronger conclusion, namely that $M = \infty$ a.s., $\psi(u) = 1$ for all u holds also when $\eta = 0$, and that $\psi(u) < 1$ for *all* $u > 0$ when $\eta > 0$. However, this needs to be verified in each separate case.

The simplest concrete example (to be studied in Chapter IV) is the *Cramér-Lundberg* or *compound Poisson* model, where $\{N_t\}$ is a Poisson process with rate β (say) and $U_1, U_2, \ldots$ are i.i.d. and independent of $\{N_t\}$. Here it is easy to see that $\rho = \beta \mathbb{E} U$ (on the average, β claims arrive per unit time and the mean of a single claim is $\mathbb{E}U$) and that also

$$\lim_{t\to\infty} \mathbb{E}\, \frac{1}{t} \sum_{k=1}^{N_t} U_k \;=\; \rho. \tag{1.7}$$

Again, (1.7) is a property which we will typically encounter. However, not all models considered in the literature have this feature:

Example 1.2 (Cox processes) Here $\{N_t\}$ is a Poisson process with *random rate* $\beta(t)$ (say) at time t. If $U_1, U_2, \ldots$ are i.i.d. and independent of $\bigl\{(\beta(t), N_t)\bigr\}$, it is not too difficult to show that ρ as defined by (1.6) is given by

$$\rho \;=\; \mathbb{E}U \cdot \lim_{t\to\infty} \frac{1}{t} \int_0^t \beta(s)\,\mathrm{d}s$$

(provided the limit exists). Thus ρ may well be random for such processes, namely, if $\bigl\{\beta(t)\bigr\}$ is non-ergodic. The simplest example is $\beta(t) \equiv V$ where V is a r.v. This case is referred to as the *mixed Poisson process*, with the most notable special case being V having a Gamma distribution, corresponding to the *Pólya process.* □

We shall only encounter a few instances of a Cox process, in connection with risk processes in a Markovian or periodic environment (Chapter VII), and here (1.6), (1.7) hold with ρ constant.

Proposition 1.3 *Assume $p \neq 1$ and define $\widetilde{R}_t = R_{t/p}$. Then the connection between the ruin probabilities for the given risk process $\{R_t\}$ and those $\widetilde{\psi}(u)$, $\widetilde{\psi}(u,T)$ for $\{\widetilde{R}_t\}$ is given by*

$$\psi(u) = \widetilde{\psi}(u), \quad \psi(u,T) = \widetilde{\psi}(u,Tp).$$

The proof is trivial. Since $\{\widetilde{R}_t\}$ has premium rate 1, the role of the result is to justify taking $p = 1$, which is feasible since in most cases the process $\{\widetilde{R}_t\}$ has a similar structure as $\{R_t\}$ (for example, the claim arrivals are Poisson or renewal at the same time). Note that when $p = 1$, the assumption $\eta > 0$ is equivalent to $\rho < 1$; in a number of models, we shall be able to identify ρ with the traffic intensity of an associated queue, and in fact $\rho < 1$ is the fundamental assumption of queueing theory ensuring steady-state behavior (existence of a limiting stationary distribution).

Notes and references The study of ruin probabilities, often referred to as *collective risk theory* or just *risk theory*, was largely initiated in Sweden in the first half of the century. Some of the main general ideas were laid down by Lundberg [614], while the first mathematically substantial results were given in Lundberg [615] and Cramér [265]; another important early Swedish work is Täcklind [826]. The Swedish school was pioneering not only in risk theory, but also in probability and applied probability as a whole; in particular, many results and methods in random walk theory originate from there and the area was ahead of related ones like queueing theory.

Some early surveys are given in Cramér [265], Segerdahl [792] and Philipson [699]. Some main later textbooks are (in alphabetical order) Bühlmann [208], Dickson [309], Daykin, Pentikäinen & Pesonen [279], De Vylder [300], Gerber [398], Grandell [429], Rolski, Schmidli, Schmidt & Teugels [746] and Seal [784, 788]. Besides in standard journals in probability and applied probability, the research literature is often published in journals like *Astin Bulletin*, *Insurance: Mathematics and Economics*, the *North American Actuarial Journal*, the *Scandinavian Actuarial Journal* and *Mitteilungen der Schweizerischen Aktuarvereinigung*. Note that the latter has recently been merged with *Blätter der Deutschen Gesellschaft für Versicherungs- und Finanzmathematik* and a number of further Actuarial Bulletins of European countries into *The European Actuarial Journal*.

The term *risk theory* is often interpreted in a broader sense than as just to comprise the study of ruin probabilities. An idea of the additional topics and problems one may incorporate under risk theory can be obtained from the survey paper [665] by Norberg; see also Chapter XVI. In the even more general area of non-life insurance mathematics, some main texts (typically incorporating some ruin theory but emphasizing the topic to a varying degree) are Bowers *et al.* [195], Bühlmann [208], Daykin *et al.* [279], Embrechts *et al.* [349], Heilmann [458], Hipp & Michel [468], Kaas *et al.* [515], Klugman, Panjer & Willmot [536], Mikosch [638], Schmidt [782], Straub [818], Sundt

[820] and Taylor [840]. Note that life insurance (e.g. Gerber [402]) has a rather different flavor, and we do not get near to the topic anywhere in this book.

Cox processes are treated extensively in Grandell [429]. For mixed Poisson processes and Pólya processes, see e.g. the recent survey by Grandell [431] and references therein.

2 Claim size distributions

This section contains a brief survey of some of the most popular classes of distributions B which have been used to model the claims $U_1, U_2, \ldots$ We roughly classify these into two groups, *light-tailed distributions* (sometimes the term 'Cramér-type conditions' is used), and *heavy-tailed distributions*. Here light-tailed means that the tail $\overline{B}(x) = 1 - B(x)$ satisfies $\overline{B}(x) = \mathrm{O}(\mathrm{e}^{-sx})$ for some $s > 0$. Equivalently, the m.g.f. $\widehat{B}[s]$ is finite for some $s > 0$. In contrast, B is *heavy-tailed* if $\widehat{B}[s] = \infty$ for all $s > 0$, but different more restrictive definitions are often used: subexponential, regularly varying (see below) or even regularly varying with infinite variance. On the more heuristical side, one could mention also the folklore in actuarial practice to consider B heavy-tailed if '20% of the claims account for more than 80% of the total claims', i.e. if

$$\frac{1}{\mu_B}\int_{b_{0.2}}^{\infty} x\, B(\mathrm{d}x) \;\geq\; 0.8,$$

where $\overline{B}(b_{0.2}) = 0.2$ and μ_B is the mean of B.

2a Light-tailed distributions

Example 2.1 (THE EXPONENTIAL DISTRIBUTION) Here the density is

$$b(x) = \delta \mathrm{e}^{-\delta x}. \tag{2.1}$$

The parameter δ is referred to as the *rate* or the *intensity*, and can also be interpreted as the (constant) failure rate $b(x)/\overline{B}(x)$.

As in a number of other applied probability areas, the exponential distribution is by far the simplest to deal with in risk theory as well. In particular, for the compound Poisson model with exponential claim sizes the ruin probability $\psi(u)$ can be found in closed form. The crucial feature is the *lack of memory*: if U is exponential with rate δ, then the conditional distribution of $U - x$ given $U > x$ is again exponential with rate δ (this is essentially equivalent to the failure rate being constant). For example in the compound Poisson model, a simple stopping time argument shows that this implies that the conditional distribution of the overshoot $S_{\tau(u)} - u$ at the time of ruin given $\tau(u)$ is again exponential with rate δ, a fact which turns out to contain considerable information. $\square$

Example 2.2 (THE GAMMA DISTRIBUTION) The gamma distribution with parameters p, δ has density

$$b(x) = \frac{\delta^p}{\Gamma(p)} x^{p-1} \mathrm{e}^{-\delta x} \tag{2.2}$$

and m.g.f.

$$\widehat{B}[s] = \left(\frac{\delta}{\delta - s}\right)^p, \quad s < \delta.$$

The mean $\mathbb{E}U$ is p/δ and the variance $\mathbb{V}ar\, U$ is p/δ^2. In particular, the squared coefficient of variation (s.c.v.)

$$\frac{\mathbb{V}ar\, U}{(\mathbb{E}U)^2} = \frac{1}{p}$$

is < 1 for $p > 1$, > 1 for $p < 1$ and $= 1$ for $p = 1$ (the exponential case).

The exact form of the tail $\overline{B}(x)$ is given by the incomplete Gamma function $\Gamma(x; p)$,

$$\overline{B}(x) = \frac{\Gamma(\delta x; p)}{\Gamma(p)} \quad \text{where} \quad \Gamma(x; p) = \int_x^\infty t^{p-1} \mathrm{e}^{-t}\, \mathrm{d}t.$$

Asymptotically, one has

$$\overline{B}(x) \sim \frac{\delta^{p-1}}{\Gamma(p)} x^{p-1} \mathrm{e}^{-\delta x}.$$

In the sense of the theory of infinitely divisible distributions, the Gamma density (2.2) can be considered as the pth power of the exponential density (2.1) (or the $1/p$th root if $p < 1$). In particular, if p is integer and U has the gamma distribution p, δ, then $U \stackrel{\mathscr{D}}{=} X_1 + \cdots + X_p$ where $X_1, X_2, \ldots$ are i.i.d. and exponential with rate δ. This special case is referred to as the *Erlang distribution with p stages*, or just the Erlang(p) distribution. An appealing feature is its simple connection to the Poisson process: $\overline{B}(x) = \mathbb{P}(U_1 + \cdots + U_p > x)$ is the probability of at most $p - 1$ Poisson events in $[0, x]$ so that

$$\overline{B}(x) = \sum_{i=0}^{p-1} \mathrm{e}^{-\delta x} \frac{(\delta x)^i}{i!}.$$

In the present text, we develop computationally tractable results mainly for the Erlang case (i.e. $p \in \mathbb{N}$). Ruin probabilities for the general case have been studied, among others, by Grandell & Segerdahl [433] and Thorin [847]. $\square$

Example 2.3 (THE HYPEREXPONENTIAL DISTRIBUTION) This is defined as a finite mixture of exponential distributions,

$$b(x) = \sum_{i=1}^{p} \alpha_i \delta_i \mathrm{e}^{-\delta_i x} \tag{2.3}$$

where $\sum_1^p \alpha_i = 1$, $0 \leq \alpha_i \leq 1$, $i = 1, \ldots, p$. An important property of the hyperexponential distribution is that its s.c.v. is > 1.

If $\alpha_i \in \mathbb{R}$, then one speaks of the distribution as a *combination of exponentials* and this class is dense in the set of all distributions on the positive halfline. □

Example 2.4 (PHASE-TYPE DISTRIBUTIONS) A phase-type distribution is the distribution of the absorption time in a Markov process with finitely many states, of which one is absorbing and the rest transient. Important special cases are the exponential, the Erlang and the hyperexponential distributions. *This class of distributions plays a major role in this book as one within computationally tractable exact forms of the ruin probability $\psi(u)$ can be obtained.*

The parameters of a phase-type distribution are the set E of transient states, the restriction $\boldsymbol{T}$ of the intensity matrix of the Markov process to E and the row vector $\boldsymbol{\alpha} = (\alpha_i)_{i \in E}$ of initial probabilities. The density and c.d.f. are

$$b(x) \;=\; \boldsymbol{\alpha} \mathrm{e}^{\boldsymbol{T}x} \boldsymbol{t}, \quad \text{resp. } B(x) \;=\; \boldsymbol{\alpha} \mathrm{e}^{\boldsymbol{T}x} \boldsymbol{e}, \qquad x \geq 0,$$

where $\boldsymbol{t} = \boldsymbol{T}\boldsymbol{e}$ and $\boldsymbol{e} = (1 \ldots 1)^{\mathsf{T}}$ is the column vector with 1 at all entries. The couple $(\boldsymbol{\alpha}, \boldsymbol{T})$ or sometimes the triple $(E, \boldsymbol{\alpha}, \boldsymbol{T})$ is called the *representation*. We give a more comprehensive treatment in IX.1 and defer further details to Chapter IX. □

Example 2.5 (DISTRIBUTIONS WITH RATIONAL TRANSFORMS) A distribution B has a rational m.g.f. (or, equivalently, a rational Laplace transform) if $\widehat{B}[r] = p(r)/q(r)$ with $p(r)$ and $q(r)$ being polynomials of finite degree. An equivalent characterization is that the density $b(x)$ is the solution of a homogeneous ordinary differential equation with constant coefficients

$$b^{(q)}(x) + d_{q-1} b^{(q-1)}(x) + \cdots + d_0 = 0; \quad d_j \in \mathbb{R}, d_0 \neq 0,$$

where one of the initial conditions is determined by $\int_0^\infty b(x)\,\mathrm{d}x = 1$. Consequently the density $b(x)$ has one of the forms

$$b(x) \;=\; \sum_{j=0}^{q} c_j x^j \mathrm{e}^{\eta_j x}, \tag{2.4}$$

$$b(x) \;=\; \sum_{j=0}^{q_1} c_j x^j \mathrm{e}^{\eta_j x} + \sum_{j=0}^{q_2} d_j x^j \cos(a_j x) \mathrm{e}^{\delta_j x} + \sum_{j=0}^{q_3} e_j x^j \sin(b_j x) \mathrm{e}^{\epsilon_j x}, \tag{2.5}$$

where the parameters in (2.4) are possibly complex-valued but the parameters in (2.5) are real-valued. This class of distributions is popular in the literature on both risk theory and queues, but often the attention is restricted to the class of phase-type distributions, which is slightly smaller but more amenable to probabilistic reasoning. We give some theory for matrix-exponential distributions in IX.6. □

Example 2.6 (DISTRIBUTIONS WITH BOUNDED SUPPORT) This example (i.e. there exists a $x_0 < \infty$ such that $\overline{B}(x) = 0$ for $x \geq x_0$, $\overline{B}(x) > 0$ for $x < x_0$) is of course a trivial instance of a light-tailed distribution. However, it is notable from a practical point of view because of reinsurance: if excess-of-loss reinsurance has been arranged with retention level x_0, then the claim size which is relevant from the point of view of the insurance company itself is $U \wedge x_0$ rather than U (the excess $(U - x_0)^+$ is covered by the reinsurer). See XVI.6. □

2b Heavy-tailed distributions

Example 2.7 (THE WEIBULL DISTRIBUTION) This distribution originates from reliability theory. Here failure rates $\delta(x) = b(x)/\overline{B}(x)$ play an important role, the exponential distribution representing the simplest example since here $\delta(x)$ is constant. However, in practice one may observe that $\delta(x)$ is either decreasing or increasing and may try to model smooth (increasing or decreasing) deviations from constancy by $\delta(x) = dx^{r-1}$ $(0 < r < \infty)$. Writing $c = d/r$, we obtain the Weibull distribution

$$\overline{B}(x) \;=\; \mathrm{e}^{-cx^r}, \quad b(x) \;=\; crx^{r-1}\mathrm{e}^{-cx^r}, \tag{2.6}$$

which is heavy-tailed when $0 < r < 1$. All moments are finite. Another interpretation is that it is the distribution of $X^{1/r}$, where X is exponential with parameter c. □

Example 2.8 (THE LOGNORMAL DISTRIBUTION) The *lognormal distribution* with parameters σ^2, μ is defined as the distribution of e^V where $V \sim N(\mu, \sigma^2)$, or equivalently as the distribution of $\mathrm{e}^{\sigma W + \mu}$ where $W \sim N(0,1)$. It follows that the density is

$$\begin{aligned} b(x) \;&=\; \frac{\mathrm{d}}{\mathrm{d}x}\Phi\Big(\frac{\log x - \mu}{\sigma}\Big) \;=\; \frac{1}{\sigma x}\varphi\Big(\frac{\log x - \mu}{\sigma}\Big) \\ &=\; \frac{1}{x\sigma\sqrt{2\pi}}\exp\Big\{-\frac{1}{2}\Big(\frac{\log x - \mu}{\sigma}\Big)^2\Big\}. \end{aligned} \tag{2.7}$$

Asymptotically, the tail is

$$\overline{B}(x) \;\sim\; \frac{\sigma}{\log x\sqrt{2\pi}}\exp\Big\{-\frac{1}{2}\Big(\frac{\log x - \mu}{\sigma}\Big)^2\Big\}, \tag{2.8}$$

which is heavier than the one of the Weibull distribution. The lognormal distribution has moments of all orders. In particular, the mean is $\mathrm{e}^{\mu+\sigma^2/2}$ and the second moment is $\mathrm{e}^{2\mu+2\sigma^2}$. □

Example 2.9 (THE PARETO DISTRIBUTION) Here the essence is that the tail $\overline{B}(x)$ decreases like a power of x. There are various variants of the definition around, the simplest one being

$$\overline{B}(x) = x^{-\alpha}, \quad x \geq 1, \tag{2.9}$$

which can be interpreted as the distribution of e^X for an exponential r.v. X with parameter α. Another variant is often referred to as US-Pareto and defined by

$$\overline{B}(x) = \frac{a^\alpha}{(a+x)^\alpha}, \quad b(x) = \frac{\alpha a^\alpha}{(a+x)^{\alpha+1}}, \quad x \geq 0, \tag{2.10}$$

for some $a > 0$. The pth moment is finite if and only if $p < \alpha - 1$.

The Laplace-Stieltjes transform of the Pareto distribution defined in (2.9) can be expressed through the incomplete Gamma function by

$$\widehat{B}[-s] = \int_1^\infty \mathrm{e}^{-sx} \frac{\alpha}{x^{\alpha+1}}\,\mathrm{d}x = \alpha\, s^\alpha \Gamma(-\alpha, s).$$

Similarly, the Laplace-Stieltjes transform of the US Pareto distribution is $\widehat{B}[-s] = \alpha\,(as)^\alpha \mathrm{e}^{as}\Gamma(-\alpha, as)$. These relatively simple expressions have not always been noted.

Abate, Choudhury & Whitt [1] introduced a somewhat related class of random variables called Pareto mixture of exponentials, which are products of Pareto and exponential r.v.'s and lead to quite explicit Laplace-Stieltjes transforms. □

Example 2.10 (THE LOGGAMMA DISTRIBUTION) The *loggamma distribution with parameters* p, δ is defined as the distribution of e^V where V has the gamma density (2.2). The density is

$$b(x) = \frac{\delta^p (\log x)^{p-1}}{x^{\delta+1}\Gamma(p)}. \tag{2.11}$$

The pth moment is finite if $p < \delta$ and infinite if $p > \delta$. For $p = 1$, the loggamma distribution is a Pareto distribution. □

Example 2.11 (DISTRIBUTIONS WITH REGULARLY VARYING TAILS) The tail $\overline{B}(x)$ of a distribution B is said to be *regularly varying with index* α if

$$\overline{B}(x) \sim \frac{L(x)}{x^\alpha}, \quad x \to \infty, \tag{2.12}$$

where $L(x)$ is *slowly varying*, i.e. satisfies $L(xt)/L(x) \to 1$ as $x \to \infty$ (any L having a limit in $(0, \infty)$ is slowly varying; another standard example is $(\log x)^\eta$). Thus, examples of distributions with regularly varying tails are the Pareto distribution (2.10) (here $L(x) \to 1$), the loggamma distribution (with index δ) and a Pareto mixture of exponentials. □

Example 2.12 (THE SUBEXPONENTIAL CLASS OF DISTRIBUTIONS) We say that a distribution B is subexponential if

$$\lim_{x\to\infty} \frac{\overline{B^{*2}}(x)}{\overline{B}(x)} \;=\; 2. \tag{2.13}$$

It can be proved (see X.1) that any distribution with a regularly varying tail is subexponential. Also, for example the lognormal distribution is subexponential (but not regularly varying), though the proof of this is non-trivial, and so is the Weibull distribution with $0 < r < 1$. Thus, the subexponential class of distributions provide a convenient framework for studying large classes of heavy-tailed distributions. We return to a closer study in X.1. □

When studying ruin probabilities, it will be seen that we obtain completely different results depending on whether the claim size distribution is exponentially bounded or heavy-tailed. From a practical point of view, this phenomenon represents one of the true controversies of the area. Namely, the knowledge of the claim size distribution will typically be based upon statistical data, and based upon such information it seems questionable to extrapolate to tail behavior. However, one may argue that this difficulty is not restricted to ruin probability theory alone. Similar discussion applies to the distribution of the accumulated claims (XVI.2) or even to completely different applied probability areas like extreme value theory: if we are using a Gaussian process to predict extreme value behavior, we may know that such a process (with a covariance function estimated from data) is a reasonable description of the behavior of the system under study in typical conditions, but can never be sure whether this is also so for atypical levels for which far less detailed statistical information is available. We give some discussion on standard methods to distinguish between light and heavy tails in Chapter X.

3 The arrival process

For the purpose of modeling a risk process, the claim size distribution represents of course only one aspect (though a major one). At least as important is the specification of the structure of the point process $\{N_t\}$ of claim arrivals and its possible dependence with the claims.

By far the most prominent case is the compound Poisson (Cramér-Lundberg) model where $\{N_t\}$ is Poisson and independent of the claim sizes $U_1, U_2, \ldots$ The reason is in part mathematical since this model is the easiest to analyze, but the model also admits a natural interpretation: a large portfolio of insurance holders, which each have a (time-homogeneous) small rate of experiencing a claim, gives rise to an arrival process which is very close to a Poisson process, in just the same way as the Poisson process arises in telephone traffic (a large number of subscribers each calling with a small rate), radioactive decay (a huge number of atoms each splitting with a tiny rate) and many other applications. The compound Poisson model is studied in detail in Chapters IV, V (and, with the extension to premiums depending on the reserve, in Chapter VIII).

To the authors' knowledge, not so many detailed studies of the goodness-of-fit of the Poisson model in insurance are available. Some of them have concentrated on the marginal distribution of N_T (say T = one year), found the Poisson distribution to be inadequate and suggested various other univariate distributions as alternatives, e.g. the negative binomial distribution. The difficulty in such an approach lies in that it may be difficult or even impossible to imbed such a distribution into the continuous set-up of $\{N_t\}$ evolving over time, and also that the ruin problem may be hard to analyze. Nevertheless, getting away from the simple Poisson process seems a crucial step in making the model more realistic, in particular to allow for certain inhomogeneities.

Historically, the first extension to be studied in detail was $\{N_t\}$ to be renewal (the interarrival times $T_1, T_2, \ldots$ are i.i.d. but with a general not necessarily exponential distribution). This model, to be studied in Chapter VI, has some mathematically appealing random walk features, which facilitate the analysis. However, it is more questionable whether it provides a model with a similar intuitive content as the Poisson model. One could possibly argue that renewal models are a compromise between choosing a tractable model and taking into account statistical information that may indicate that exponential interarrival time distributions do not calibrate given data well enough. Of course, one is then still left to believe in the independence assumption and – with the introduced memory between claims – one has to be aware that the resulting model is for most applications to be seen as an interpolation rather than a causal model.

A more appealing way to allow for inhomogeneity is by means of an intensity $\beta(t)$ fluctuating over time. An obvious example is $\beta(t)$ depending on the time of the year (the season), so that $\beta(t)$ is a periodic function of t; we study this case in VII.6. Another one is Cox processes, where $\{\beta(t)\}_{t\geq 0}$ is an arbitrary stochastic process. In order to prove reasonably substantial and interesting results, Cox processes are, however, too general and one needs to specialize to more concrete assumptions. The one we focus on (Chapter VII) is a Markovian environment: the environmental conditions are described by a finite Markov process $\{J_t\}_{t\geq 0}$, such that $\beta(t) = \beta_i$ when $J_t = i$. I.e. with a common term $\{N_t\}$ is a *Markov-modulated Poisson process*; its basic feature is to allow more variation (bursty arrivals) than inherent in the simple Poisson process. This model can be intuitively understood in some simple cases like $\{J_t\}$ describing weather conditions in car insurance, epidemics in life insurance etc. In others, it may be used in a purely descriptive way when it is empirically observed that the claim arrivals are more bursty than allowed for by the simple Poisson process.

Mathematically, the periodic and the Markov-modulated models also have attractive features. The point of view we take here is Markov-dependent random walks in continuous time (Markov additive processes), see III.4. This applies also to the case where the claim size distribution depends on the time of the year or the environment (VII.6), and which seems well motivated from a practical point of view as well.

4 A summary of main results and methods

4a Duality with other applied probability models

Risk theory may be viewed as one of many applied probability areas, others being branching processes, genetics models, queueing theory, dam/storage processes, reliability, interacting particle systems, stochastic differential equations, time series and Gaussian processes, extreme value theory, stochastic geometry, point processes and so on. Some of these have a certain resemblance in flavor and methodology, others are quite different.

The ones which appear most related to risk theory are queueing theory and dam/storage processes. In fact, it is a recurrent theme of this book to stress this connection which was often neglected in the early specialized literature on risk theory. Mathematically, the classical result is that the ruin probabilities for the compound Poisson model are related to the workload (virtual waiting time) process $\{V_t\}_{t\geq 0}$ of an initially empty M/G/1 queue by means of

$$\psi(u,T) = \mathbb{P}(V_T > u), \quad \psi(u) = \mathbb{P}(V > u), \tag{4.1}$$

where V is the limit in distribution of V_t as $t \to \infty$. The M/G/1 workload process $\{V_t\}$ may also be seen as one of the simplest storage models, with Poisson arrivals and constant release rule $p(x) \equiv 1$. A general release rule $p(x)$ means that $\{V_t\}$ decreases according to the differential equation $\dot{V} = -p(V)$ in between jumps, and here (4.1) holds as well provided the risk process has a premium rule depending on the reserve, $\dot{R} = p(R)$ in between jumps. Similarly, ruin probabilities for risk processes with an input process which is renewal, Markov-modulated or periodic can be related to queues with similar characteristics. Thus, it is desirable to have a set of formulas like (4.1) permitting to translate freely between risk theory and the queueing/storage setting. In Chapter VIII we will also see a direct and natural link between the maximum workload of an M/G/1 queue and the ruin probability in a compound Poisson risk model in terms of excursions. In general, methods or modeling ideas developed in one area often have relevance for the other one as well.

A stochastic process $\{V_t\}$ is said to be in the *steady state* if it is strictly stationary (in the Markov case, this amounts to V_0 having the stationary distribution of $\{V_t\}$), and the limit $t \to \infty$ is the *steady-state limit*. The study of the steady state is by far the most dominant topic of queueing and storage theory, and a lot of information on steady-state r.v.'s like V is available. It should be noted, however, that quite often the emphasis is on computing expected values like $\mathbb{E}V$. In the setting of (4.1), this gives only $\int_0^\infty \psi(u)\mathrm{d}u$ which is of limited intrinsic interest. Thus, the two areas, though overlapping, have to some extent a different flavor.

A prototype of the duality results in this book is Theorem III.2.1, which gives a sample path version of (4.1) in the setting of a general premium rule $p(x)$: the events $\{V_T > u\}$ and $\{\tau(u) \le T\}$ coincide when the risk process and the storage process are coupled in a suitable way (via time-reversion). The infinite horizon (steady state) case is covered by letting $T \to \infty$. The fact that Theorem III.2.1 is a sample path relation should be stressed: in this way the approach also applies to models having supplementary r.v.'s like the environmental process $\{J_t\}$ in a Markov-modulated setting.

4b Exact solutions

Of course, the ideal is to be able to come up with closed form solutions for the ruin probabilities $\psi(u)$, $\psi(u,T)$. The cases where this is possible are basically the following for the infinite horizon ruin probability $\psi(u)$:

- The compound Poisson model with constant premium rate $p = 1$ and exponential claim size distribution B, $\overline{B}(x) = \mathrm{e}^{-\delta x}$. Here $\psi(u) = \rho \mathrm{e}^{-\gamma u}$ where β is the arrival intensity, $\rho = \beta/\delta$ and $\gamma = \delta - \beta$.

- The compound Poisson model with constant premium rate $p = 1$ and B being phase-type with just a few phases. Here $\psi(u)$ is given in terms of a matrix-exponential function (Corollary IX.3.1), which can be expanded into a sum of exponential terms by diagonalization (see, e.g., Example IX.3.2). The qualifier 'with just a few phases' refers to the fact that the diagonalization has to be carried out numerically in higher dimensions.

- The compound Poisson model with a claim size distribution degenerate at one point, see Corollary IV.3.7.
- The compound Poisson model with some rather special heavy-tailed claim size distributions, see Boxma & Cohen [193] and Abate & Whitt [3].
- The compound Poisson model with premium rate $p(x)$ depending on the reserve and exponential claim size distribution B. Here $\psi(u)$ is explicit provided that, as is typically the case, the functions

$$\omega(x) \;=\; \int_0^x \frac{1}{p(y)}\,\mathrm{d}y, \qquad \int_0^\infty \frac{1}{p(x)}\mathrm{e}^{\beta\omega(x)-\delta x}\,\mathrm{d}x$$

  can be written in closed form, see Corollary VIII.1.9.
- The compound Poisson model with a piecewise constant premium rule $p(x)$ and B being phase-type with just a few phases, see IX.7.
- Renewal models with exponential claim sizes, see Theorem VI.2.2.
- Renewal model variants of the above cases for which the interclaim time is phase-type with just a few phases.
- Any Lévy model where the risk reserve process (not the claim surplus process!) is downward skipfree (Theorem XI.2.3). This includes Brownian motion.
- Any Lévy model for which the scale function is explicitly available, see XI.3 (for an early example cf. Furrer [381]).

A further notable fact (see again XVI.1) is the explicit form of the ruin probability when $\{R_t\}$ is a diffusion with infinitesimal drift and variance $\mu(x), \sigma^2(x)$:

$$\psi(u) \;=\; \frac{\int_u^\infty \exp\left\{-\int_0^x 2\mu(y)/\sigma^2(y)\,\mathrm{d}y\right\}\,\mathrm{d}x}{\int_0^\infty \exp\left\{-\int_0^x 2\mu(y)/\sigma^2(y)\,\mathrm{d}y\right\}\,\mathrm{d}x} \;=\; 1-\frac{S(u)}{S(\infty)} \tag{4.2}$$

where

$$S(u) \;=\; \int_0^u \exp\left\{-\int_0^x 2\mu(y)/\sigma^2(y)\,\mathrm{d}y\right\}\,\mathrm{d}x$$

is the natural scale.

The finite horizon ruin probability $\psi(u,T)$ is explicit for Brownian motion (III.1) and the compound Poisson model with exponential claim size distribution (V.1). Later in the book a number of further rather specific cases will be discussed for which explicit expressions exist.

4c Numerical methods

Next to a closed-form solution, the second best alternative is a numerical procedure which allows to calculate the exact values of the ruin probabilities. Here are some of the main approaches:

Laplace transform inversion Often, it is easier to find the Laplace transforms

$$\widehat{\psi}[-s] = \int_0^\infty \mathrm{e}^{-su}\psi(u)\,\mathrm{d}u\,, \quad \widehat{\psi}[-s,-\omega] = \int_0^\infty\int_0^\infty \mathrm{e}^{-su-\omega T}\psi(u,T)\,\mathrm{d}u\,\mathrm{d}T$$

in closed form rather than the ruin probabilities $\psi(u)$, $\psi(u,T)$ themselves. In that case $\psi(u)$, $\psi(u,T)$ can be calculated numerically by some method for transform inversion, say the fast Fourier transform (FFT) as implemented in Grübel [438] for infinite horizon ruin probabilities for the renewal model. We do not discuss Laplace transform inversion much; other relevant references are Abate & Whitt [2], Embrechts, Grübel & Pitts [346] and Grübel & Hermesmeier [439]; see also Albrecher, Avram & Kortschak [14] and the Bibliographical Notes in [746, p. 191].

Matrix-analytic methods This approach is relevant when the claim size distribution is of phase-type (or matrix-exponential), and in quite a few cases (Chapter IX), $\psi(u)$ is then given in terms of a matrix-exponential function $\mathrm{e}^{\boldsymbol{U}u}$ (here $\boldsymbol{U}$ is some suitable matrix) which can be computed by diagonalization, as the solution of linear differential equations or by some series expansion (not necessarily the straightforward $\sum_0^\infty \boldsymbol{U}^n u/n!$ one!). In the compound Poisson model with $p=1$, $\boldsymbol{U}$ is explicit in terms of the model parameters, whereas for the renewal arrival model and the Markovian environment model $\boldsymbol{U}$ has to be calculated numerically, either as the iterative solution of a fixed point problem or by finding the diagonal form in terms of the complex roots to certain transcendental equations.

Differential- and integral equations The idea here is to express $\psi(u)$ or $\psi(u,T)$ as the solution to a differential- or integral equation, and carry out the solution by some standard numerical method. One example where this is feasible is the renewal equation for $\psi(u)$ (Corollary IV.3.3) in the compound Poisson model which is an integral equation of Volterra type. The method is, however, restricted to models that have a certain degree of Markovian structure in which case conditioning (or applying the more formal tool of generators, see II.4a) leads to equations that often involve both differential and integral terms. We will discuss cases where this approach can even lead to explicit solutions (see e.g. IX.7 and XII.3c). In many more cases, numerical solution methods are applicable, although the initial or boundary conditions can be a challenge.

If an integral equation is available, it is often possible to define a contractive integral operator and to identify the ruin probability as its fixed point, in which case the ruin probability can be approximated by iterated application of the integral operator to some starting function. The resulting high-dimensional integral can then be calculated by standard Monte Carlo and Quasi-Monte Carlo techniques (see e.g. Albrecher *et al.* [31, 24]). In comparison to the alternative of direct simulation of the risk process (as discussed in Section 4g), this technique often has significant computational advantages over the latter.

4d Approximations

The Cramér-Lundberg approximation This is one of the most celebrated results of risk theory (and probability theory as a whole). For the compound Poisson model with $p = 1$ and claim size distribution B with moment generating function (m.g.f.) $\widehat{B}[s]$, it states that

$$\psi(u) \ \sim \ C\mathrm{e}^{-\gamma u}, \quad u \to \infty, \tag{4.3}$$

where $C = (1-\rho)/(\beta\widehat{B}'[\gamma]-1)$ and $\gamma > 0$ is the solution of the Lundberg equation

$$\beta\bigl(\widehat{B}[\gamma]-1\bigr) - \gamma \ = \ 0, \tag{4.4}$$

which can equivalently be written as

$$\widehat{B}[\gamma] \ = \ 1 + \frac{\gamma}{\beta}\,. \tag{4.5}$$

It is rather standard to call γ the *adjustment coefficient* but a variety of other terms are also frequently encountered (and often the notation R instead of γ is used in the literature). The Cramér-Lundberg approximation is renowned not only for its mathematical beauty but also for being very precise, often for *all* $u \geq 0$ and not just for large u. It has generalizations to other Lévy models, to the models with renewal arrivals, a Markovian environment or periodically varying parameters. However, in such cases the evaluation of C is more cumbersome. In fact, when the claim size distribution is of phase-type, the exact solution is as easy to compute as the Cramér-Lundberg approximation in some of these models.

The shape of the l.h.s. of equation (4.4) and its extensions to other models lie at the heart of ruin theory. Its level sets (not only the one at 0) reveal a lot of (in particular asymptotic) properties of ruin-related quantities and will play an important role in this book.

Diffusion approximations Here the idea is simply to approximate the risk process by a Brownian motion (or a more general diffusion) by fitting the first and second moments, and to use the fact that first passage probabilities are more readily calculated for diffusions than for the risk process itself. Diffusion approximations are easy to calculate, but typically not very precise in their first naive implementation. However, incorporating correction terms may change the picture dramatically. In particular, corrected diffusion approximations (see V.6) are by far the best one can do in terms of finite horizon ruin probabilities $\psi(u,T)$.

Large claims approximations In order for the Cramér-Lundberg approximation to be valid, the claim size distribution should have an exponentially decreasing tail $\overline{B}(x)$. In the case of heavy-tailed distributions, other approaches are thus required. Approximations for $\psi(u)$ as well as for $\psi(u,T)$ for large u are available in most of the models we discuss. For example, for the compound Poisson model under certain assumptions on B

$$\psi(u) \sim \frac{\rho}{\mu_B(1-\rho)} \int_u^\infty \overline{B}(x)\,\mathrm{d}x, \quad u \to \infty. \tag{4.6}$$

In fact, in some cases the results are even more complete than for light tails. See Chapter X.

This list of approximations does by no means exhaust the topic; some further possibilities are surveyed in IV.7 and V.2.

4e Bounds and inequalities

The outstanding result in the area is Lundberg's inequality

$$\psi(u) \le \mathrm{e}^{-\gamma u}. \tag{4.7}$$

Compared to the Cramér-Lundberg approximation (4.3), it has the advantage of not involving approximations and also, as a general rule, of being somewhat easier to generalize beyond the compound Poisson setting. We return to various extensions and sharpenings of Lundberg's inequality (finite horizon versions, lower bounds etc.) at various places and in various settings.

When comparing different risk models, it is a general principle that *adding random variation to a model increases the risk*. For example, one expects a model with a deterministic claim size distribution B, say degenerate at m, to have smaller ruin probabilities than when B is non-degenerate with the same mean m. This is proved for the compound Poisson model in IV.8 (see also further ordering results for dependent risks in Section XIII.8). Empirical evidence shows that the general principle holds in a broad variety of settings, though precise mathematical results are not always available.

4f Statistical methods

Any of the approaches and results above assume that the parameters of the model are completely known. In practice, they have however to be estimated from data, obtained say by observing the risk process in $[0, T]$. This procedure in itself is fairly straightforward; e.g., in the compound Poisson model, it splits up into the estimation of the Poisson intensity (the estimator is $\widehat{\beta} = N_T/T$) and of the parameter(s) of the claim size distribution, which is a standard statistical problem since the claim sizes $U_1, \ldots, U_{N_T}$ are i.i.d. given N_T. However, the difficulty comes in when drawing inference about the ruin probabilities. How do we produce a confidence interval? And, more importantly, can we trust the confidence intervals for the large values of u which are of interest? In the present authors' opinion, this is extrapolation from data due to the extreme sensitivity of the ruin probabilities to the tail of the claim size distribution in particular (in contrast, fitting a parametric model to $U_1, \ldots, U_{N_T}$ may be viewed as an interpolation or smoothing of the histogram). For example, one may question whether it is possible to distinguish between claim size distributions which are heavy-tailed and those that have an exponentially decaying tail. This issue will be further discussed in Section X.6.

4g Simulation

The development of modern computers has made simulation a popular experimental tool in all branches of applied probability and statistics, and of course the method is relevant in risk theory as well. Simulation may be used just to get some vague insight in the process under study: simulate one or several sample paths, and look at them to see whether they exhibit the expected behavior or some surprises come up. However, the more typical situation is to perform a Monte Carlo experiment to estimate probabilities (or expectations or distributions) which are not analytically available. For example, this is a straightforward way to estimate finite horizon ruin probabilities.

The infinite horizon case presents a difficulty, because it appears to require an infinitely long simulation. Truncation to a finite horizon (or above a certain surplus level) has been used, but is not very satisfying. Still, good methods exist in a number of models and are based upon representing the ruin probability $\psi(u)$

as the expected value of a r.v. (or a functional of the expectation of a set of r.v's) which can be generated by simulation. The problem is entirely analogous to estimating steady-state characteristics by simulation in queueing/storage theory, and in fact methods from that area can often be used in risk theory as well. We look at a variety of such methods in Chapter XV, and also discuss how to develop methods which are efficient in terms of producing a small variance for a fixed simulation budget. A main problem is that ruin is typically a rare event (i.e. having small probability) and that therefore it is expensive or even infeasible in terms of computer time to obtain reasonably precise estimates of the ruin probability through naive simulation. Variance reduction techniques to improve the situation are discussed in Chapter XV.