# Introduction to actuarial mathematics

Lecture notes for *Forsikring og jura*

Second edition

Christian Furrer

Department of Mathematical Sciences
University of Copenhagen

Department of Mathematical Sciences
University of Copenhagen
Universitetsparken 5
DK-2100 Copenhagen Ø


furrer@math.ku.dk

Second edition, 2024
ISBN: 978-87-7125-287-3

# Preface

**First edition.** These lecture notes have been prepared for the BSc Programme in Actuarial Mathematics under the Department of Mathematical Sciences at the University Copenhagen, more specifically for the first year course *Insurance and law* (in Danish: *Forsikring og jura*). This course introduces the students to conceps and methods from actuarial mathematics, ranging from basic notions of risk to actuarial and financial valuation principles, and is deeply formative in nature. While motivated by practical applications, the style of presentation emphasizes mathematical rigor and is, consequently, first and foremost targeted a mathematical audience. In particular, the reader is expected to be acquainted with mathematical analysis, including the notions of continuity, differentiability, and Riemann integrability as well as the fundamental theorems of calculus, and well-versed in elementary probability theory corresponding to the material covered in for example *Introduction to Probability, Second Edition* by Joseph K. Blitzstein & Jessica Hwang (2019).

Numerous exercises are provided at the end of each chapter; they are essential to the didactic purpose of the lecture notes by underpinning and putting into perspective the concepts and results of the primary text, but also by illustrating how the abstract methods might be applied in practice. In addition, links to actuarial practice are provided throughout the text.

In total, the material is well suited for an introductory course on actuarial mathematics of about 5 ECTS, but it is also possible to handpick sections in connection with more specialized courses. For example, Section 2.1, Chapter 3, and Section 4.1 could constitute the first part of a course on the mathematics of life insurance.

This presentation draws to a great extent on past lecture notes for *Insurance and law* by my colleagues Jostein Paulsen and Mogens Steffensen. I would like to offer my sincere thanks to Jostein and Mogens for sharing their material with me. I am also grateful to Mie Kano Glückstad, who as a teaching assistant for *Insurance and law* proofread an early version of the notes and contributed with many insightful comments, questions, and suggestions. On a final note, I should also like to thank my many former students who suffered through earlier versions of the notes, identified mistakes, and suggested improvements.

Please do not hesitate to contact me via e-mail at furrer@math.ku.dk in case you should happen to stumble upon any errors or typographical mistakes.

Christian Furrer
Copenhagen, February 2023

**Second edition.** The second edition does not differ substantially from the first edition. However, a few inconsistencies and inaccuracies have been corrected. I should like to thank Alexander ter Braak for useful feedback.

Christian Furrer
Copenhagen, April 2024

# Contents

# Chapter 1

# Introduction

> *"Certain forms of uncertainty make life interesting but less safe. I was always intrigued by the various forms of risk that are associated with human life and activity and how they can be mitigated for the individual by contractual risk exchange between two or more parties. Certain forms of certainty make life interesting and more safe. I was always attracted to mathematics because it allows for statements that are non-trivial and still indisputably true. These two areas of interest synthesize perfectly into actuarial/financial mathematics, which gives precise contents to notions of risk and develops methods for measuring and controlling it."*
>
> — Ragnar Norberg (1945–2017)

These lectures notes serve as an introduction to actuarial mathematics. The word *actuarial* refers to the title of *actuary* (in Danish: *aktuar*); actuaries are business professionals that apply mathematical-statistical methods to model, assess, and control risk – in particular in the context of insurance. Actuarial mathematics is exactly the branch of applied mathematics from which these mathematical-statistical methods are drawn. It includes a number of interrelated subjects, ranging from probability theory over statistics to mathematical finance, and is of course first and foremost quantitative. In Denmark, many if not almost all actuaries are employed in the insurance and pension sector; this is a sector which has always been a beneficiary of much mathematical talent.

The terms actuarial mathematics and insurance mathematics are often used interchangeably. After all, insurance is chiefly concerned with risk modeling and risk control. In some sense, insurance is really nothing but a wager: be it on whether your bicycle gets stolen, whether your house burns down, or whether you or your loved ones become injured or pass away. But betting in this context is quite different from say playing the lottery, since the benefits are paid conditional on an event that has already caused financial or personal distress to the insured. So at its heart, insurance involves the transfer of risk from an individual, the insured, to some institution, the insurer, who is better equipped to manage the risk. This institution could be a cooperative, a private or public company, or the state.

One would hardly say that society in general encourages its members to partake in gambling, yet this is the case for the betting game that is insurance. How come? It stems from the fact that insurance has a stabilizing effect on the financial position of individuals and families and thereby on society as a whole. Consequently, the running of an insurance business is deeply regulated, and politics are ripe with discussions concerning individual and collective bearing of burdens in relation to insurance.

In actuarial mathematics specifically but also more broadly, it is common to distinguish between *life insurance* (in Danish: *livsforsikring*) and *non-life insurance* (in Danish: *skadeforsikring*). We should like to stress that since there is no universally agreed-upon definition of the term life insurance, this distinction is not set in stone. From the point of view of actuarial mathematics, the term non-life insurance is first and foremost a testimony to the historical development of the field with mathematical-statistical methods for life insurance (in a narrow sense) preceding mathematical-statistical methods targeted for example fire insurance. Today, the distinction between life insurance and non-life insurance is derived not only from historical developments within the field, but also from differences in regulation and mathematical methodology.

Throughout this presentation, we adopt the point of view that life insurance encompasses all forms of insurance that in one way or another relates to the health status of the insured, including disability and health insurance. By presenting a wide range of methods from both the mathematics of life insurance and the mathematics of non-life insurance, we hope to enable the reader to make up their own mind regarding the scope of application of these methods to different types of insurance risk.

These lecture notes are not intended to offer a comprehensive overview of the basics of actuarial mathematics; rather, they serve as an introduction to the peculiar notions and particular ways of thinking that set actuarial mathematics apart from other branches of applied mathematics. In particular, their main purpose is first and foremost to prepare the reader for further in-depth studies concerning both the mathematics of life insurance as well as the mathematics of non-life insurance.

The presentation is structured as follows. In Chapter 2, we introduce and study three basic types of insurance risk: mortality, severity, and frequency. In Chapter 3, we discuss actuarial and financial valuation techniques. Chapter 4 on Thiele's differential equation (from the mathematics of life insurance) and the collective risk model (from the mathematics of non-life insurance) concludes.

# Chapter 2

# Basic insurance risks

At the heart of insurance is the transfer of risk from an insured to an insurer. According to the Oxford English Dictionary, *risk* is defined as "(Exposure to) the possibility of loss, injury, or other adverse or unwelcome circumstances; a chance or situation involving such a possibility". By talking about possibilities, it is clear that the concept of risk is closely related to uncertainty; as an example, gambling is *risky* since it involves games of chance.

In this chapter, we review certain aspects of some basic insurance risks from a probabilistic point of view. The main goal is to develop a range of methods to quantify risk and (qualitatively) compare different risks. These methods then serve as the basis for valuation of insurance liabilities, in particular pricing of insurance contracts, which we study in Chapter 3.

The historical and methodological distinction between life insurance and non-life insurance, which we established in the previous chapter, is also retained here. In the context of life insurance, we focus on unsystematic mortality risk and on the concept of mortality rates. This is the focal point of Section 2.1. In the context of non-life insurance, we discuss both severity risk (Section 2.2) and frequency risk (Section 2.3). In combination, mortality, severity, and frequency risk encompass the most basic insurance risks.

## 2.1 Mortality risk

In this section, we give an introduction to some key aspects of mortality risk. In combination, financial risk and mortality risk constitute the basic risks of life insurance. Our interpretation of life insurance, as discussed previously, shall be quite broad and thus encompass a range of insurance contracts tied to the health status of the insured, including life insurances, disability insurances, and pension schemes. These products have in common that the contributions and benefits, to a lesser or greater extent, depend on the survival of the insured. In other words, the contracts involve a transfer of mortality risk from the insured to the insurer.

It is important to distinguish between two types of mortality risk, namely unsystematic and systematic mortality risk; the latter is often also denoted longevity risk. Unsystematic mortality risk is composed of the uncertainty that stems from the fact that survival times are random. Systematic mortality risk is composed of the uncertainty that stems from the fact that future life expectancies are uncertain, both due to so-called short-term mortality shocks (think: pandemics, cure of cancer) and long-term lifetime improvements, which have been observed throughout the 20th century and the beginning of the 21th century. Unsystematic mortality risk is broadly speaking diversifiable, meaning that it vanishes as the size of the insurance portfolio (the number of insured) goes to infinity. This is a consequence of the Law of Large Numbers. Systematic mortality risk, on the other hand, is not diversifiable. In these lecture notes, we focus on unsystematic mortality risk.

In the following, we consider a non-negative random variable $T$, describing the survival time of some individual, say an insured. The methodology we develop also applies to other areas, for example in reliability engineering, where $T$ would describe the lifetime of a system or component. But in the following, we are going to stress the demographical point of view. Since we focus on unsystematic mortality risk, we shall think of the time axis as describing the age of the insured.

Let $F$ denote the (cumulative) distribution function of $T$. Since the random variable $T$ is non-negative, we may take the domain of $F$ to be $[0, \infty)$. For a fixed $t \geq 0$, we interpret $F(t) = \mathbb{P}(T \leq t)$ as the probability that the insured dies before or exactly at time $t$.

**Definition 2.1.1.** *The **survival function** of $T$ is*

$$\overline{F}(t) = \mathbb{P}(T > t), \qquad t \geq 0.$$

For some fixed $t \geq 0$, we interpret $\overline{F}(t)$ as the probability that the insured is yet to die (still alive) at time $t$. Since the complement of the event $(T \leq t)$ is exactly the event $(T > t)$, we find that $\overline{F} = 1 - F$. The survival function is sometimes denoted $S$ rather than $\overline{F}$.

The following assumptions are primarily made for convenience and mathematical tractability; they allow us to proceed via elementary means.

**Assumption 2.1.2.** *We assume that*

$$F(t) = \int_0^t f(s)\,\mathrm{d}s, \qquad t \geq 0,$$

*for some probability density function $f$ with domain $[0, \infty)$. Furthermore, we assume that $f$ is continuous on $[0, \infty)$ and strictly positive on $(0, \infty)$.*

By the first assumption, the survival time $T$ is a continuous random variable, so that $\mathbb{P}(T \leq t) = \mathbb{P}(T < t)$ and, equivalently, $\mathbb{P}(T > t) = \mathbb{P}(T \geq t)$. We may interpret this as the probability of dying at a specific time $t$ being zero. Since $f$ is continuous, it follows from the First Fundamental Theorem of Calculus that

$$\frac{\mathrm{d}}{\mathrm{d}t} F(t) = f(t), \qquad t > 0. \tag{2.1.1}$$

Similarly, we have

$$-\frac{\mathrm{d}}{\mathrm{d}t} \overline{F}(t) = f(t), \qquad t > 0, \tag{2.1.2}$$

since $\overline{F} = 1 - F$. The assumption that $f$ is strictly positive on $(0, \infty)$ implies that $F$ is strictly increasing and $\overline{F}$ is strictly decreasing. Recall that $\lim_{t \to \infty} F(t) = 1$, so that $\lim_{t \to \infty} \overline{F}(t) = 0$. In combination, we find that $\overline{F} > 0$.

*Remark* 2.1.3. It follows from the above considerations that in the context of survival times, the assumption that $f$ is strictly positive on $(0, \infty)$ rules out the existence of some maximal biological lifetime $\tau > 0$ for which $\overline{F}(\tau) = 0$. One could allow for such a maximal biological lifetime by requiring the random variable $T$ to instead take values only in $[0, \tau]$ and suitably adapting Assumption 2.1.2. The results we derive in this would then still by and large hold, although only on the time interval from zero up until time $\tau$ rather than $[0, \infty)$. ▽

The life expectancy (expected survival time) is the expected value or mean of $T$, which we denote by $\mathbb{E}[T]$; it may be computed according to the expression

$$\mathbb{E}[T] = \int_0^\infty t f(t) \, \mathrm{d}t. \tag{2.1.3}$$

The following result provides an alternative formula. Note that the proof only relies on the assumption that $T$ is a non-negative continuous random variable.

**Proposition 2.1.4.** *It holds that*

$$\mathbb{E}[T] = \int_0^\infty \overline{F}(t) \, \mathrm{d}t.$$

*Proof.* By (2.1.3) and since $t = \int_0^t 1 \, \mathrm{d}u$, we find that

$$\begin{aligned} \mathbb{E}[T] &= \int_0^\infty t f(t) \, \mathrm{d}t \\ &= \int_0^\infty \int_0^t 1 \, \mathrm{d}u \; f(t) \, \mathrm{d}t \\ &= \int_0^\infty \int_0^t f(t) \, \mathrm{d}u \, \mathrm{d}t. \end{aligned}$$

Interchanging the order of integration yields

$$\mathbb{E}[T] = \int_0^\infty \int_u^\infty f(t)\,\mathrm{d}t\,\mathrm{d}u.$$

Recall that

$$\int_u^\infty f(t)\,\mathrm{d}t = \int_0^\infty f(t)\,\mathrm{d}t - \int_0^u f(t)\,\mathrm{d}t = 1 - F(u).$$

The latter equals $\overline{F}(u)$ since $\overline{F} = 1 - F$. Consequently,

$$\mathbb{E}[T] = \int_0^\infty \overline{F}(u)\,\mathrm{d}u$$

as desired. □

*Remark* 2.1.5. Consider the transformed random variable $T^k$ for some $k \in \mathbb{N}$. The survival function of $T^k$ is given by $\overline{F}(t^{1/k})$, $t \geq 0$, since $\mathbb{P}(T^k > t) = \mathbb{P}(T > t^{1/k})$ for $t \geq 0$. From Proposition 2.1.4 and using the substitution $u = t^{1/k}$, we may then conclude that

$$\begin{aligned}\mathbb{E}[T^k] &= \int_0^\infty \overline{F}(t^{1/k})\,\mathrm{d}t \\ &= k\int_0^\infty u^{k-1}\overline{F}(u)\,\mathrm{d}u,\end{aligned}$$

which extends the formula of Proposition 2.1.4 to moments of all order. ▽

**Exercise 2.1.6.** Let $N$ be a non-negative discrete random variable. Denote by $p$ the probability mass function of $N$. We may take the domain of $p$ to be $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$, when $p(n) = \mathbb{P}(N = n)$ for all $n \in \mathbb{N}_0$. Recall that

$$\mathbb{E}[N] = \sum_{n=0}^{\infty} np(n).$$

Show that

$$\mathbb{E}[N] = \sum_{n=0}^{\infty} \mathbb{P}(N > n)$$

by following an approach similar to the one found in the proof of Proposition 2.1.4. ◇

The expected value $\mathbb{E}[T]$ of $T$ provides the life expectancy of a newborn. In the context of life insurance, we are rather interested in residual (remaining) lifetimes for insured of a certain age, say $x \geq 0$. This leads us to introduce the notion of conditional distribution functions and conditional survival functions.

**Definition 2.1.7.** *The* ***conditional distribution function*** *of $T-x$ given the event $(T>x)$ for an $x \geq 0$ is*

$$F_x(t) = \mathbb{P}(T \leq x+t \,|\, T>x), \qquad t \geq 0.$$

*The* ***conditional survival function*** *of $T-x$ given the event $(T>x)$ for an $x \geq 0$ is*

$$\overline{F}_x(t) = \mathbb{P}(T > x+t \,|\, T>x), \qquad t \geq 0.$$

*Remark* 2.1.8. In the classic actuarial literature, notations such as ${}_tq_x$ for $\mathbb{P}(T \leq x+t \,|\, T>x)$ and ${}_tp_x$ for $\mathbb{P}(T > x+t \,|\, T>x)$ are quite common and constitute examples of the so-called Hamza notation. In these lecture notes, Hamza notation is not used.

Note that $\overline{F}_x = 1 - F_x$. It is quite straightforward to show that conditional distribution functions are in fact distribution functions. Thus we may introduce random variables $T_x$, $x \geq 0$, which satisfy[1]

$$\mathbb{P}(T_x \leq t) = F_x(t), \qquad \mathbb{P}(T_x > t) = \overline{F}_x(t), \qquad t \geq 0. \tag{2.1.4}$$

We think of $T_x$ as the residual lifetime of an insured of age $x$. Note that since $\mathbb{P}(T=0)=0$ due to Assumption 2.1.2, we have that $F_0 = F$ and $\overline{F}_0 = \overline{F}$.

In the following, it will be useful to note that for $x \geq 0$ and $t \geq 0$,

$$F_x(t) = \frac{\mathbb{P}(T \leq x+t, T>x)}{\mathbb{P}(T>x)} = \frac{F(x+t)-F(x)}{\overline{F}(x)}, \tag{2.1.5}$$

$$\overline{F}_x(t) = \frac{\mathbb{P}(T > x+t)}{\mathbb{P}(T>x)} = \frac{\overline{F}(x+t)}{\overline{F}(x)}, \tag{2.1.6}$$

by definition and since $(T > x+t) \subset (T > x)$. In particular, we find that $\overline{F}_x(t) \geq \overline{F}(x+t)$, with equality if and only if $x=0$. We interpret this as the probability of surviving till age $x+t$ being higher for an insured of age $x$ than for a newborn. Note also the clear link between $\overline{F}_x$ and $\overline{F}$ given in (2.1.6); it reveals that the conditional survival functions, rather than the conditional distribution functions, might be the natural mathematical objects of interest.

The following result characterizes the conditional probability density functions, that is the probability density functions of the conditional distributions.

**Proposition 2.1.9.** *For any fixed $x \geq 0$ it holds that*

$$F_x(t) = \int_0^t f_x(s)\,\mathrm{d}s, \qquad t \geq 0,$$

[1] This statement is non-trivial. Further (measure-theoretic) details lie outside the scope of these lecture notes.

*where $f_x$ is given by*

$$f_x(t) = \frac{f(x+t)}{\overline{F}(x)}, \qquad t \geq 0.$$

*Proof.* Plugging in we find that

$$\begin{aligned}\int_0^t f_x(s)\,\mathrm{d}s &= \int_0^t \frac{f(x+s)}{\overline{F}(x)}\,\mathrm{d}s \\ &= \frac{1}{\overline{F}(x)} \int_0^t f(x+s)\,\mathrm{d}s \\ &= \frac{1}{\overline{F}(x)} \int_x^{x+t} f(s)\,\mathrm{d}s.\end{aligned}$$

Note that

$$\int_x^{x+t} f(s)\,\mathrm{d}s = \int_0^{x+t} f(s)\,\mathrm{d}s - \int_0^x f(s)\,\mathrm{d}s = F(x+t) - F(x),$$

so that by (2.1.5) we may conclude that

$$\int_0^t f_x(s)\,\mathrm{d}s = \frac{F(x+t) - F(x)}{\overline{F}(x)} = F_x(t)$$

as desired. □

Note that $f_x$ is continuous on $[0,\infty)$ and strictly positive on $(0,\infty)$. In particular, it follows from the First Fundamental Theorem of Calculus that

$$\frac{\mathrm{d}}{\mathrm{d}t} F_x(t) = f_x(t), \qquad t > 0. \tag{2.1.7}$$

Similarly, we have

$$-\frac{\mathrm{d}}{\mathrm{d}t} \overline{F}_x(t) = f_x(t), \qquad t > 0,$$

since $\overline{F}_x = 1 - F_x$.

**Example 2.1.10.** Suppose that $T$ follows an exponential distribution with mean $\lambda > 0$, so that the probability density function is given by

$$f(t) = \frac{1}{\lambda} \exp\left\{-\frac{1}{\lambda} t\right\}.$$

By integrating $f$, we find that the survival function takes the form

$$\overline{F}(t) = 1 - F(t) = \exp\left\{-\frac{1}{\lambda} t\right\}.$$

Consequently, the conditional survival functions read

$$\begin{aligned}\overline{F}_x(t) &= \frac{\overline{F}(x+t)}{\overline{F}(x)}\\ &= \frac{\exp\{-\frac{1}{\lambda}(x+t)\}}{\exp\{-\frac{1}{\lambda}x\}}\\ &= \exp\left\{-\frac{1}{\lambda}t\right\}\\ &= \overline{F}(t).\end{aligned}$$

This property is also called memorylessness, and among all continuous distributions it is only obeyed by the exponential distribution. In our context of mortality, it essentially states that the probability of surviving a time period of length $t$, say 10 years, is independent of the current age of the insured. This is highly improbable, so we should focus on other distributions than the exponential distribution. ◦

The residual life expectancy is the expected value of $T_x$; it may be cast as

$$\mathbb{E}[T_x] = \mathbb{E}[T - x \,|\, T > x]$$

since $T_x$ follows the same distribution as $T - x$ given the event $(T > x)$, and it may be computed according to the expression

$$\mathbb{E}[T_x] = \int_0^\infty t f_x(t)\,\mathrm{d}t = \frac{1}{\overline{F}(x)}\int_x^\infty t f(t)\,\mathrm{d}t - x. \tag{2.1.8}$$

The following result provides an alternative formula. Note that the proof only relies on the assumption that $T$ is a non-negative continuous random variable and that $\overline{F}_x$ is well-defined for the $x \geq 0$ of interest, that is $\overline{F}(x) > 0$. This is also the case of (2.1.8).

**Proposition 2.1.11.** *For any fixed $x \geq 0$ it holds that*

$$\mathbb{E}[T_x] = \int_0^\infty \overline{F}_x(t)\,\mathrm{d}t = \frac{1}{\overline{F}(x)}\int_0^\infty \overline{F}(x+t)\,\mathrm{d}t.$$

*Proof.* Apply Proposition 2.1.4 on $T_x$ in place of $T$ and use the fact that

$$\overline{F}_x(t) = \frac{\overline{F}(x+t)}{\overline{F}(x)}$$

according to (2.1.6). □

Until now we have described the distribution of the survival time $T$ via the distribution function $F$, the survival function $\overline{F}$, or the probability density function $f$. But as we already noticed in Example 2.1.10, it is hard to tell directly from these objects and without further calculations whether a distribution is viable or not. So we are looking for a concept that better captures our intuition concerning mortality; the following is an obvious candidate.

**Definition 2.1.12.** *The* ***mortality rate*** *of $T$ is*

$$\mu(x) = \lim_{h \downarrow 0} \frac{1}{h} F_x(h), \qquad x \geq 0. \tag{2.1.9}$$

The mortality rate is also known as the *force of mortality*. In other contexts, one may also use the nomenclature hazard rate or failure rate for $\mu$. While (2.1.9) might look complicated at first, the notion of mortality rates is actually quite intuitive. For $h > 0$ the quantity

$$F_x(h) = \mathbb{P}(T \leq x + h \,|\, T > x)$$

is the probability that the insured who has survived until and including age $x$ dies within the time interval $(x, x + h]$. Since the conditional distribution function is continuously differentiable, this probability is proportional to $h$ as $h$ goes to zero. The mortality rate is then exactly this infinitesimal proportionality factor.

The following result shows how the mortality rate is related to the probability density function and the survival function.

**Proposition 2.1.13.** *It holds that*

$$\mu(x) = f_x(0) = \frac{f(x)}{\overline{F}(x)} = -\frac{\mathrm{d}}{\mathrm{d}x} \log\{\overline{F}(x)\}, \qquad x \geq 0,$$

*so that*

$$\overline{F}(x) = \exp\Big\{ - \int_0^x \mu(s)\,\mathrm{d}s \Big\}, \qquad x \geq 0. \tag{2.1.10}$$

*Proof.* Since $F_x(0) = 0$, it follows from Proposition 2.1.9 that

$$\begin{aligned} \mu(x) &= \lim_{h \downarrow 0} \frac{1}{h} F_x(h) \\ &= \frac{\mathrm{d}}{\mathrm{d}t} F_x(t) \Big|_{t=0} \\ &= f_x(0) \\ &= \frac{f(x)}{\overline{F}(x)}. \end{aligned}$$

According to (2.1.2) the derivative of $-\overline{F}(x)$ with respect to $x$ is $f(x)$. Thus

$$\mu(x) = \frac{f(x)}{\overline{F}(x)} = -\frac{\mathrm{d}}{\mathrm{d}x} \log\{\overline{F}(x)\},$$

where we have also utilized that the derivative of $\log\{y\}$ with respect to $y$ is $\frac{1}{y}$. From this we conclude that

$$\int_0^x -\mu(s)\,\mathrm{d}s = \log\{\overline{F}(x)\},$$

so that

$$\overline{F}(x) = \exp\Big\{-\int_0^x \mu(s)\,\mathrm{d}s\Big\},$$

which completes the proof. □

Note that the mortality rate is continuous on $[0,\infty)$ and strictly positive on $(0,\infty)$ and that since $\overline{F}(t) \to 0$ as $t \to \infty$, it holds that $\int_0^t \mu(s)\,\mathrm{d}s \to \infty$ as $t \to \infty$. It is important to stress that the mortality rate actually characterizes the distribution of $T$. That is, given a function $\mu$ with domain $[0,\infty)$ which is continuous on $[0,\infty)$, strictly positive on $(0,\infty)$, and satisfies $\int_0^t \mu(s)\,\mathrm{d}s \to \infty$ as $t \to \infty$, then

$$F(t) = 1 - \exp\Big\{-\int_0^t \mu(s)\,\mathrm{d}s\Big\}, \qquad t \geq 0,$$

can be shown to define a distribution function satisfying Assumption 2.1.2 with corresponding mortality rate $\mu$.

**Exercise 2.1.14.** Let $\mu \equiv \frac{1}{\lambda}$ for some $\lambda > 0$. Show that $T$ follows an exponential distribution with mean $\lambda$. ◇

The mortality rate, rather than the distribution function or the probability density function, is both from a statistical as well as an actuarial point of view considered the fundamental quantity of interest. This is not only due to the fact that it provides an infinitesimal and hence somewhat intuitive starting point, but also because it conveniently allows one to cast conditional probabilities and expectations in a consistent manner. The latter is exactly the content of the following result.

**Corollary 2.1.15.** *For any fixed $x \geq 0$ it holds that*

$$\overline{F}_x(t) = \exp\Big\{-\int_x^{x+t} \mu(s)\,\mathrm{d}s\Big\} = \exp\Big\{-\int_0^t \mu(x+s)\,\mathrm{d}s\Big\}, \qquad t \geq 0.$$

*Furthermore,*

$$\mathbb{E}[T_x] = \int_0^\infty \exp\Big\{-\int_x^{x+t} \mu(s)\,\mathrm{d}s\Big\}\,\mathrm{d}t = \int_0^\infty \exp\Big\{-\int_0^t \mu(x+s)\,\mathrm{d}s\Big\}\,\mathrm{d}t.$$

*Proof.* According to (2.1.6) and (2.1.10), we have

$$\begin{aligned}
\overline{F}_x(t) &= \frac{\overline{F}(x+t)}{\overline{F}(x)} \\
&= \exp\Big\{-\int_0^{x+t} \mu(s)\,\mathrm{d}s\Big\} \Big/ \exp\Big\{-\int_0^{x} \mu(s)\,\mathrm{d}s\Big\} \\
&= \exp\Big\{-\Big(\int_0^{x+t} \mu(s)\,\mathrm{d}s - \int_0^{x} \mu(s)\,\mathrm{d}s\Big)\Big\} \\
&= \exp\Big\{-\int_x^{x+t} \mu(s)\,\mathrm{d}s\Big\} = \exp\Big\{-\int_0^{t} \mu(x+s)\,\mathrm{d}s\Big\},
\end{aligned}$$

which proves the first part of the corollary. The second part follows from the first part and Proposition 2.1.11. □

In Section 3.3 and in particular in Section 4.1, conditional distributions of residual lifetimes play a prominent role. To be specific, it turns out to be quite important to be able to characterize the conditional distribution of $T_x$ given the event $(T_x > t)$. To this end, we introduce the shorthand notation

$$p_x^{\mathrm{aa}}(t,s) := \mathbb{P}(T_x > s \,|\, T_x > t) \quad \text{and} \quad p_x^{\mathrm{ad}}(t,s) := \mathbb{P}(T_x \le s \,|\, T_x > t) \tag{2.1.11}$$

for $0 \le t \le s$. Note that $p_x^{\mathrm{ad}} = 1 - p_x^{\mathrm{aa}}$. The superscripts 'a' and 'd' are used to indicate whether the insured is 'alive' or 'dead', respectively. The following lemma establishes the fact that the conditional distribution of $T_x$ given the event $(T_x > t)$ corresponds to the distribution of $T_{x+t} + t$, while the subsequent lemma provides an expression for the probability density function of $T_{x+t} + t$ in terms of $\mu$ and $p_x^{\mathrm{aa}}$.

**Lemma 2.1.16.** *For any fixed $x \ge 0$ and $t \ge 0$ it holds that*

$$\begin{aligned}
p_x^{aa}(t,s) &= \mathbb{P}(T_{x+t} > s-t) = \overline{F}_{x+t}(s-t), \qquad & s \ge t,\\
p_x^{ad}(t,s) &= \mathbb{P}(T_{x+t} \le s-t) = F_{x+t}(s-t), \qquad & s \ge t.
\end{aligned}$$

*Furthermore,*

$$p_x^{aa}(t,s) = \exp\Big\{-\int_t^{s} \mu(x+u)\,\mathrm{d}u\Big\}, \qquad s \ge t. \tag{2.1.12}$$

*In particular,*

$$p_x^{aa}(0,t) = p_x^{aa}(0,u)p_x^{aa}(u,t) = p_x^{aa}(0,u)p_{x+u}^{aa}(0,t-u), \qquad 0 \le u \le t.$$

*Proof.* For the first part, confer with Exercise VI and use the fact that $p_x^{\mathrm{ad}} = 1 - p_x^{\mathrm{aa}}$. For the second part, use Corollary 2.1.15 to obtain

$$p_x^{\mathrm{aa}}(t,s) = \exp\Big\{-\int_{x+t}^{x+s} \mu(u)\,\mathrm{d}u\Big\} = \exp\Big\{-\int_t^{s} \mu(x+u)\,\mathrm{d}u\Big\} \tag{2.1.13}$$

as desired. The third part follows immediatly from the previous parts. □

**Lemma 2.1.17.** *For any fixed $x \geq 0$ and $t \geq 0$ it holds that*

$$f_{x+t}(s-t) = \mu(x+s)p_x^{aa}(t,s), \qquad s \geq t.$$

*Proof.* Let $x \geq 0$ and $t \geq 0$ be fixed. Recall from Proposition 2.1.9 that

$$f_{x+t}(s-t) = \frac{f(x+s)}{\overline{F}(x+t)}, \qquad s \geq t,$$

so that according to Proposition 2.1.13 it holds that

$$f_{x+t}(s-t) = \frac{f(x+s)}{\overline{F}(x+s)}\frac{\overline{F}(x+s)}{\overline{F}(x+t)} = \mu(x+s)\frac{\mathbb{P}(T > x+s)}{\mathbb{P}(T > x+t)}.$$

By invoking the definition of conditional probabilities and using the fact that the distribution of $T_{x+t}$ corresponds to the conditional distribution of $T-(x+t)$ given the event $(T > x+t)$, we find that

$$\frac{\mathbb{P}(T > x+s)}{\mathbb{P}(T > x+t)} = \frac{\mathbb{P}(T-(x+t) > s-t)}{\mathbb{P}(T > x+t)} = \mathbb{P}(T_{x+t} > s-t).$$

Thus according to Lemma 2.1.16, it holds that

$$\frac{\mathbb{P}(T > x+s)}{\mathbb{P}(T > x+t)} = p_x^{\mathrm{aa}}(t,s).$$

Collecting results completes the proof. □

We conclude this section by discussing some aspects of mortality modeling in Denmark. Historically, the Gompertz-Makeham distribution (or mortality law) has played a central role. For this model, the mortality rate takes the form

$$\mu(t) = \alpha + \beta\exp\{\gamma t\}, \qquad t \geq 0,$$

for parameters $\alpha > 0$, $\beta > 0$, and $\gamma \in \mathbb{R}$. It turns out that the Gompertz-Makeham mortality law strikes a decent balance between simplicity (few parameters) and mathematical tractability on one side and flexibility (many parameters) and descriptive prowess on the other side. In the Danish mortality table for males used by Danish insurance companies from 1982 (known also as G82M), the parameters are taken to be

$$\alpha = 5 \cdot 10^{-4}, \qquad \beta = 7.5858 \cdot 10^{-5}, \qquad \gamma = \log\{1.09144\},$$

so that

$$\mu(t) = 5 \cdot 10^{-4} + 7.5858 \cdot 10^{-5} \cdot 1.09144^t, \qquad t \geq 0. \tag{2.1.14}$$

Since 2011, Danish life insurance companies have had to use the so-called *Levetidsmodel* of the Danish Financial Supervisory Authority (FSA). This framework consists of

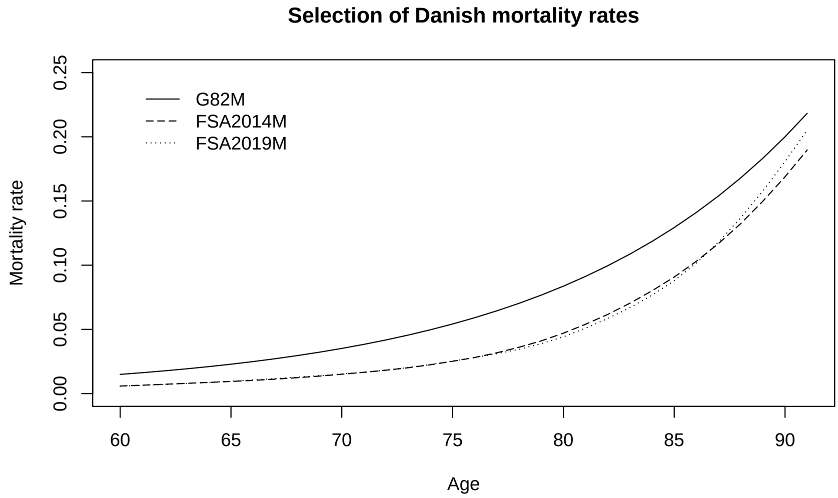


**Figure 2.1:** *Selection of male mortality rates for ages* 60 *to* 90*. We observe significant lifetime improvements since 1982.*

multiple components, including average mortality rates (in Danish: observerede nuværende dødeligheder) for insured of male and female sex, respectively, which are updated on a yearly basis. Figure 2.1 serves to compare the mortality rate of (2.1.14), that is the mortality rate for males from 1982, with the Danish FSA's average mortality rate for insured of male sex in 2014 as well as 2019. The mortality rates are significantly lower in both 2014 and 2019 than in 1982, which points to significant lifetime improvements and thus systematic mortality risk. This in turn stresses the need to model the mortality rate not only as a function of age, as has been our focus, but also as a function of calendar time. The Danish FSA's Levetidsmodel addresses such issues by including a component which quantifies expected future lifetime improvements (in Danish: forventede fremtidige levetidsforbedringer).

## 2.2 Severity risk

In non-life insurance, one typically distinguishes between severity risk and frequency risk. Severity risk concerns the uncertainty of the size of insurance claims, while frequency risk concerns the uncertainty regarding the number of claims within the contractual period, which is typically a year. In this section, we give an introduction to some key aspects of severity risk. Section 2.3 is then devoted to frequency risk.

Denote by $X$ a non-negative random variable describing the size of an insurance claim, and let $G$ and $\overline{G} = 1 - G$ denote the distribution function and survival function of $X$, respectively. To keep things somewhat simple and closely related to Section 2.1, we are going to assume that $X$ is a continuous random variable with probability density function $g$, whose domain we may take to be $[0, \infty)$ since $X$ is non-negative.

Popular claim size distributions include the Gamma, log-normal, and Pareto distribution. These distributions have significantly different characteristics, and one may be more suitable than the others depending on the type of insurance contract and the risks involved. When they are finite, we may compare distributions via central moments. Quite often, however, the total claim amount generated by an insurance portfolio stems from a few large claims, and thus higher-order central moments might not be finite. This is reflected by a popular rule of thumb known as the Pareto principle, which states that one fifth of all claims account for four fifths of the total claim amount in an insurance portfolio. Essentially, the claim size distribution is in some sense heavy-tailed, whatever that exactly means, and to appropriately capture this characteristic is crucial.

If $X$ follows an exponential distribution, say with mean $\lambda > 0$, then the survival function reads

$$\overline{G}(x) = \exp\left\{-\frac{1}{\lambda}x\right\}, \qquad x \geq 0.$$

This survival function exhibits exponential decay, that is, the rate with which the survival function decreases to zero as $x$ increases to infinity is proportional to its current value. Continuous distributions with survival functions that admit exponential decay are considered *light-tailed*.

We say that $X$ follows a Pareto distribution of shape $\alpha > 0$ and scale $\theta > 0$ if

$$\overline{G}(x) = \left(\frac{\theta}{\theta + x}\right)^{\alpha}, \qquad x \geq 0.$$

We may also write $X \sim \text{Pareto}(\alpha, \theta)$. The survival function of a Pareto distribution exhibits power decay (of order $\alpha$). Survival functions that exhibit slower decay, for example logarithmic decay, are rare. So to contrast the exponential distribution and other light-tailed distributions, the Pareto distribution and other continuous distribution that admit power decay or slower decay are considered *heavy-tailed*. The categorization of distributions with survival functions that admit decay at a rate between power decay and exponential decay depends on the exact mathematical notion of heavy-tailedness, of which there exists quite a few; examples include the concepts of regularly varying distributions and subexponential distributions. Later, in Example 2.2.4, we discuss the notion of regular variation, while subexponentiality is outside the scope of these lecture notes.

To further examine the tail behavior of claim size distributions, we now introduce the mean excess function. It describes the expected excess of a claim size over some threshold as a function of the threshold.

**Definition 2.2.1.** *The **mean excess function** of $X$ is*

$$e(d) = \mathbb{E}[X - d \,|\, X > d], \qquad d \geq 0.$$

Note that the mean excess function is only properly defined at $d \geq 0$ whenever $\mathbb{P}(X > d) > 0$. In the following, we typically implicitly assume that $\mathbb{P}(X > d) > 0$. If we would had interpreted $X$ as a survival time rather than the size of an insurance claim, then the mean excess function of $X$ would correspond to the residual life expectancy discussed in Section 2.1. Actually, one may take interest not only in the mean of $X - d$ given the event $(X > d)$ but also in the actual distribution function and survival function of $X - d$ given the event $(X > d)$, which we in the spirit of Section 2.1 shall denote $G_d$ and $\overline{G}_d$, respectively. Because if the insurance contract stipulates a *deductible* $d \geq 0$ (in Danish: *selvrisiko*) and the distribution $G$ of $X$ corresponds to the distribution of incurring damages, then $G_d$ gives exactly the distribution of actual claim sizes reported to the insurer. It should be stressed that one may by and large apply the results of Section 2.1 to calculate $G_d$. For example, we have that

$$\overline{G}_d(x) = \frac{\overline{G}(d + x)}{\overline{G}(d)}, \qquad x \geq 0,$$

analogously to (2.1.6). The following results also follow directly from Section 2.1:

**Proposition 2.2.2.** *For any fixed $d \geq 0$ which satisfies $\mathbb{P}(X > d) > 0$ it holds that*

$$e(d) = \frac{1}{\overline{G}(d)} \int_d^\infty x g(x) \,\mathrm{d}x - d.$$

*Alternatively, under the same conditions,*

$$e(d) = \int_0^\infty \overline{G}_d(x) \,\mathrm{d}x = \frac{1}{\overline{G}(d)} \int_0^\infty \overline{G}(d + x) \,\mathrm{d}x.$$

*Proof.* Combine suitably the results of Section 2.1. □

**Exercise 2.2.3.** Suppose that $X$ follows an exponential distribution with mean $\lambda > 0$. Referring to Example 2.1.10, show that $e \equiv \lambda$.

A plot of $e(d)$ against $d$ is called a mean excess plot and provides a convenient graphical tool to compare distributions and validate models against actual data. Figure 2.2 compares the mean excess function of certain Pareto, Gamma, and exponential distributions, confer also with Example 2.2.4, Exercise IX, and Exercise 2.2.3. The behavior of the mean excess function of the Gamma distribution as the threshold goes to infinity resembles that of the exponential distribution considerably more than that of the Pareto distribution. Therefore one might classify the Gamma distribution as a light-tailed distribution.

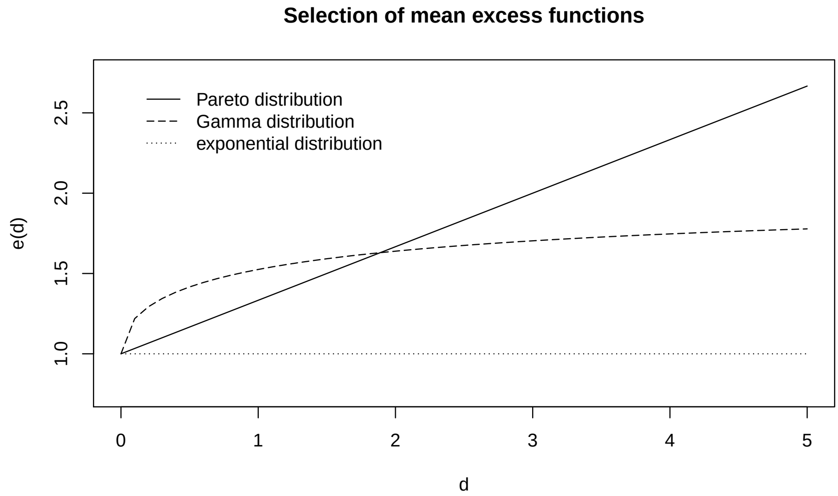


**Figure 2.2:** *Mean excess function of Pareto distribution with mean one and variance two, of Gamma distribution with mean one and variance two, and of exponential distribution with mean one, confer with Example 2.2.4, Exercise IX, and Exercise 2.2.3.*

We conclude this section by discussing in more detail two common claim size distributions, namely the Pareto and log-normal distributions. From this point onward, we write $h(x) \sim \ell(x)$ (as $x \to \infty$) for functions $h : [0, \infty) \to (0, \infty)$ and $\ell : [0, \infty) \to (0, \infty)$ if and only if

$$\lim_{x \to \infty} \frac{h(x)}{\ell(x)} = 1.$$

**Example 2.2.4.** This example is dedicated to the Pareto distribution and closely related distributions. If $X \sim \text{Pareto}(\alpha, \theta)$ for $\alpha > 0$ and $\theta > 0$, then

$$\overline{G}(x; \alpha, \theta) = \left( \frac{\theta}{\theta + x} \right)^{\alpha}, \qquad x \geq 0.$$

We first study the moments of $X$. To this end, let $k \in \mathbb{N}$ be given. The survival function of $X + \theta$ is given by $\overline{G}(x - \theta; \alpha, \theta)$, $x \geq 0$, since $\mathbb{P}(X + \theta > x) = \mathbb{P}(X > x - \theta)$, and where we have set $\overline{G}(x; \alpha, \theta) = 1$ for $x < 0$. According to Remark 2.1.5, it then holds that

$$\begin{aligned} \mathbb{E}[(X + \theta)^k] &= k \int_0^{\infty} x^{k-1} \overline{G}(x - \theta; \alpha, \theta) \, \mathrm{d}x \\ &= k\theta^{\alpha} \int_{\theta}^{\infty} x^{k-1-\alpha} \, \mathrm{d}x + k \int_0^{\theta} x^{k-1} \, \mathrm{d}x. \end{aligned}$$

Since $\frac{\mathrm{d}}{\mathrm{d}x} x^k = k x^{k-1}$, the second term equals $\theta^k$. Regarding the first term, note that if $\alpha < k$, then

$$\int_\theta^\infty x^{k-1-\alpha}\,\mathrm{d}x = \frac{1}{k-\alpha}\left[x^{k-\alpha}\right]_{x=\theta}^{x=\infty} = \infty,$$

so that $\mathbb{E}[(X+\theta)^k] = \infty$. If instead $\alpha = k$, then

$$\int_\theta^\infty x^{k-1-\alpha}\,\mathrm{d}x = [\log\{x\}]_{x=\theta}^{x=\infty} = \infty,$$

so that $\mathbb{E}[(X+\theta)^k] = \infty$. On the other hand, if $\alpha > k$, then

$$\int_\theta^\infty x^{k-1-\alpha}\,\mathrm{d}x = \frac{1}{k-\alpha}\left[x^{k-\alpha}\right]_{x=\theta}^{x=\infty} = \frac{1}{\alpha-k}\theta^{k-\alpha},$$

and by collecting terms we may conclude that

$$\mathbb{E}[(X+\theta)^k] = \frac{\alpha\theta^k}{\alpha-k}.$$

To summarize,

$$\mathbb{E}[(X+\theta)^k] = \begin{cases} \frac{\alpha\theta^k}{\alpha-k}, & \alpha > k, \\ \infty, & \alpha \le k, \end{cases}$$

so that in particular

$$\mathbb{E}[X] = \begin{cases} \frac{\theta}{\alpha-1}, & \alpha > 1, \\ \infty, & \alpha \le 1. \end{cases} \tag{2.2.1}$$

We now turn our attention to the mean excess function of $X$, which is well-defined (finite) whenever $\alpha > 1$. Note that

$$\begin{aligned} \overline{G}_d(x;\alpha,\theta) &= \frac{\overline{G}(d+x;\alpha,\theta)}{\overline{G}(d;\alpha,\theta)} \\ &= \left(\frac{\theta+d}{\theta+d+x}\right)^\alpha \\ &= \overline{G}(x;\alpha,\theta+d), \end{aligned}$$

so that by Proposition 2.2.2 and (2.2.1) it holds that

$$e(d) = \frac{\theta+d}{\alpha-1}, \qquad d \ge 0,$$

In other words, the Pareto distribution admits a linear mean excess function in correspondence with Figure 2.2.

We conclude this example by discussing so-called regularly varying distributions. Note that in the case of the Pareto distribution, we have that

$$\overline{G}(x;\alpha,\theta) = \frac{L(x)}{x^\alpha} \tag{2.2.2}$$

for some function $L : [0, \infty) \to (0, \infty)$ which satisfies the property $L(x) \sim L(cx)$ for all $c > 0$; such functions are by the way called slowly varying and include constants, logarithms, and powers of logarithms. Regularly varying distributions are exactly Pareto-like distributions in the sense that they admit the representation (2.2.2). One may show that regularly varying distributions essentially exhibit the same tail behavior as the Pareto distribution, in particular one finds that $e(d) \sim \frac{d}{\alpha-1}$ (as $d \to \infty$)[2]. The regularly varying distributions comprise some of the most heavy-tailed distributions which are actually fitted to claim size data. ◦

**Exercise 2.2.5.** Show that (2.2.2) holds with

$$L(x) = \left(\frac{x\theta}{\theta + x}\right)^{\alpha},$$

and argue that $L(x) \sim L(cx)$ for all $c > 0$. ⋄

**Example 2.2.6.** This example is dedicated to the log-normal distribution. Suppose that $Y$ follows a normal distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2$, where $\sigma > 0$. Then $X = \exp\{Y\}$ is said to follow a log-normal distribution with paramaters $\mu$ and $\sigma$. Since $\mathbb{P}(X \le x) = \mathbb{P}\Big(\frac{Y-\mu}{\sigma} \le \frac{\log\{x\}-\mu}{\sigma}\Big)$ for $x \ge 0$, the survival function of $Y$ is given by

$$\overline{G}(x; \mu, \sigma) = 1 - \Phi\left(\frac{\log\{x\} - \mu}{\sigma}\right), \qquad x \ge 0,$$

where $\Phi$ is the distribution function of $Z = \frac{Y-\mu}{\sigma}$. We are here employing the conventions $\log\{0\} = -\infty$ and $\Phi(-\infty) = 0$ so that $\overline{G}(0, \mu, \sigma) = 1$. Recall that $Z$ follows a normal distribution with mean 0 and variance 1, so that $\Phi(x) = \int_{-\infty}^{x} \phi(y)\,\mathrm{d}y$, $x \in \mathbb{R}$, for

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}, \qquad x \in \mathbb{R}.$$

We first study the moments of $X$. To this end, let $k \in \mathbb{N}$ be given. Note that

$$\mathbb{E}[X^k] = \mathbb{E}[\exp\{kY\}],$$

which corresponds to the the moment-generating function of $Y$ evaluated in $k$. Since $Y$ follows a normal distribution with mean $\mu$ and variance $\sigma^2$, we find that

$$\mathbb{E}[X^k] = \exp\left\{\mu k + \frac{1}{2}\sigma^2 k^2\right\}.$$

In particular, $\mathbb{E}[X] = \exp\{\mu + \sigma^2/2\}$.

[2]This statement is non-trivial. It is the consequence of Karamata's theorem. Jovan Karamata (1902–1967) was a Serbian mathematician who is best known for his contributions to mathematical analysis.

We now turn our attention to the mean excess function of $X$. Note that

$$\begin{aligned} e(d) &= \mathbb{E}[X - d \,|\, X > d] \\ &= \mathbb{E}[\exp\{Y\} \,|\, X > d] - d \\ &= \frac{1}{\mathbb{P}(Y > \log\{d\})} \mathbb{E}[\exp\{Y\} \mathbb{1}_{\{Y > \log\{d\}\}}] - d \\ &= \frac{1}{\overline{G}(d; \mu, \sigma)} \int_{\log\{d\}}^{\infty} \exp\{x\} \frac{1}{\sigma} \phi\Big(\frac{x - \mu}{\sigma}\Big) \,\mathrm{d}x - d. \end{aligned}$$

This follows from the events $(X > d)$ and $(Y > \log\{d\})$ being equal and from the fact that $Y$ admits the probability density function $\frac{1}{\sigma}\phi(\frac{x-\mu}{\sigma})$, $x \in \mathbb{R}$. By definition,

$$\exp\{x\} \frac{1}{\sigma} \phi\Big(\frac{x - \mu}{\sigma}\Big) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\Big\{x - \frac{(x - \mu)^2}{2\sigma^2}\Big\}.$$

Expanding the square and rearranging yields that

$$\begin{aligned} \exp\{x\} \frac{1}{\sigma} \phi\Big(\frac{x - \mu}{\sigma}\Big) &= \frac{1}{\sqrt{2\pi}} \exp\{\mu + \sigma^2/2\} \frac{1}{\sigma} \exp\Big\{-\frac{(x - (\mu + \sigma^2))^2}{2\sigma^2}\Big\} \\ &= \exp\{\mu + \sigma^2/2\} \frac{1}{\sigma} \phi\Big(\frac{x - (\mu + \sigma^2)}{\sigma}\Big). \end{aligned}$$

Collecting results we find that

$$e(d) = \frac{\exp\{\mu + \sigma^2/2\}}{\overline{G}(d; \mu, \sigma)} \int_{\log\{d\}}^{\infty} \frac{1}{\sigma} \phi\Big(\frac{x - (\mu + \sigma^2)}{\sigma}\Big) \,\mathrm{d}x - d.$$

The integrand corresponds to the probability density function of a normal distribution with mean $\mu + \sigma^2$ and variance $\sigma^2$, so that the integral equals $1 - \Phi\Big(\frac{\log\{d\} - (\mu + \sigma^2)}{\sigma}\Big) = \overline{G}(d; \mu + \sigma^2, \sigma)$. We conclude that

$$e(d) = \frac{\exp\{\mu + \sigma^2/2\} \overline{G}(d; \mu + \sigma^2, \sigma)}{\overline{G}(d; \mu, \sigma)} - d, \qquad d \geq 0.$$

One may now tediously show by writing

$$e(d) = \frac{\exp\{\mu + \sigma^2/2\} \overline{G}(d; \mu + \sigma^2, \sigma) - d\overline{G}(d; \mu, \sigma)}{\overline{G}(d; \mu, \sigma)}$$

and using L'Hôpital's rule repeatedly that $e(d) \sim \frac{\sigma^2 d}{\log\{d\} - \mu}$ (as $d \to \infty$). In other words, the mean excess function goes to infinity as the threshold goes to infinity at a rate proportional to $d/\log\{d\}$. In comparison, the Pareto distribution goes to infinity somewhat quicker, namely at a rate proportional to $d$, confer with Example 2.2.4. So while the log-normal distribution might still be considered quite heavy-tailed, it is not Pareto-like (regularly varying) and thus not among the most heavy-tailed distributions. ∘

## 2.3 Frequency risk

In the previous subsection, we discussed severity risk, which deals with uncertainty concerning the size of insurance claims. This subsection is instead devoted to frequency risk, which deals with uncertainty concerning the number of the claims within the contractual period.

Denote by $N$ a non-negative discrete random variable describing the number of claims within some contractual period, say e.g. one year. We also use the term claim count for $N$. Let $p$ denote the probability mass function of $N$, whose domain we may take to be $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$, when $p(n) = \mathbb{P}(N = n)$ for all $n \in \mathbb{N}_0$.

It is often empirically observed that $\mathrm{Var}[N] > \mathbb{E}[N]$, a phenomena closely related to that of *overdispersion*, which describes the situation where data exhibits greater variability than predicted by some more or less implicit mean-variance relationship. Popular claim size distributions include the Poisson distribution and the negative binomial distribution, which are discussed in the examples below. We shall see that only the negative binomial distribution is able to capture overdispersion.

**Example 2.3.1.** The non-negative discrete random variable $N$ is said to follow a Poisson distribution with mean $\lambda > 0$ if

$$p(n) = \frac{\lambda^n}{n!} \exp\{-\lambda\}, \qquad n \in \mathbb{N}_0.$$

We also write $N \sim \mathrm{Poisson}(\lambda)$. The moment-generating function $M$ of $N$ is by definition given by

$$M(t) = \mathbb{E}[\exp\{tN\}], \qquad t \in \mathbb{R}.$$

Recall that

$$\sum_{n=0}^{\infty} \frac{a^n}{n!} = \exp\{a\}, \qquad a \in \mathbb{R}. \tag{2.3.1}$$

Straightforward calculations, reproduced below, and an application of (2.3.1) yield

$$\begin{aligned} M(t) &= \sum_{n=0}^{\infty} \exp\{tn\} \frac{\lambda^n}{n!} \exp\{-\lambda\} \\ &= \exp\{-\lambda\} \sum_{n=0}^{\infty} \frac{(\lambda \exp\{t\})^n}{n!} \\ &= \exp\{-\lambda\} \exp\{\lambda \exp\{t\}\} \\ &= \exp\{\lambda(\exp\{t\} - 1)\}. \end{aligned}$$

In particular,

$$\begin{aligned}
\mathbb{E}[N] &= \frac{\mathrm{d}}{\mathrm{d}t} M(t)\Big|_{t=0} \\
&= \lambda \exp\{\lambda(\exp\{t\} - 1) + t\}\Big|_{t=0} \\
&= \lambda, \\
\mathbb{E}[N^2] &= \frac{\mathrm{d}^2}{\mathrm{d}t^2} M(t)\Big|_{t=0} \\
&= \Big[\lambda \exp\{\lambda(\exp\{t\} - 1) + t\} + \lambda^2 \exp\{\lambda(\exp\{t\} - 1) + 2t\}\Big]_{t=0} \\
&= \lambda + \lambda^2.
\end{aligned}$$

We conclude that

$$\mathrm{Var}[N] = \mathbb{E}[N^2] - \mathbb{E}[N]^2 = \lambda + \lambda^2 - \lambda^2 = \lambda = \mathbb{E}[N],$$

so the Poisson distribution is actually unable to capture overdispersion, since its mean $\lambda > 0$ equals its variance. ◦

**Example 2.3.2.** The non-negative discrete random variable $N$ is said to follow a negative binomial distribution with parameters $\alpha > 0$ and $p \in (0, 1)$ if

$$p(n) = \frac{\Gamma(n + \alpha)}{n!\Gamma(\alpha)} p^{\alpha}(1 - p)^n, \qquad n \in \mathbb{N}_0.$$

We also write $N \sim \mathrm{NegBin}(\alpha, p)$. Here the so-called Gamma function $\Gamma$ is defined according to

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} \exp\{-x\}\, \mathrm{d}x, \qquad \alpha > 0,$$

and it is possible to show that $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$ for $\alpha > 0$, that is the Gamma function generalized the factorial.

When $\alpha = 1$, we find that

$$p(n) = p(1 - p)^n,$$

so that $N$ follows a geometric distribution with mean $\frac{1-p}{p}$. According to Exercise X, we have

$$\begin{aligned}
\mathbb{E}[N] &= \alpha \frac{1 - p}{p}, \\
\mathrm{Var}[N] &= \alpha \frac{1 - p}{p^2}.
\end{aligned}$$

Note that this yields

$$\mathrm{Var}[N] = \frac{1}{p}\mathbb{E}[N] > \mathbb{E}[N],$$

so at the cost of introducing an additional parameter compared to the Poisson distribution, the negative binomial distribution is able to capture overdispersion. ◦

We now turn our attention to what is known as *thinning*. In actuarial practice, it often happens that the claim count $N$ within a contractual period is not fully observable. If the contract stipulates a deductible, then the insured is only likely to report claims of size above the deductible. If $N$ describes the total number of claims and $N^\star$ describes the number of reported claims, then $N^\star$ is said to be a *thinned* version of $N$. Even if the claim count $N$ within a contractual period is fully observable, the insurer may still take interest in only a subsample of claims. As an example, suppose the contractual period hitherto has been one year, but the insurer now in addition wants to offer a coverage with a shorter contractual period of say one month. If $N$ describes the number of claims within the one year period and $N^\star$ describes the number of claims within the shorter one month period, then $N^\star$ is also said to be a thinned version of $N$.

In the following, we are going to focus on the case

$$N^\star = \sum_{n=1}^{N} I_n, \tag{2.3.2}$$

where $(I_n)_{n\in\mathbb{N}}$ is a sequence of independent and identically distributed random variables which are assumed independent of $N$ and required to satisfy $\mathbb{P}(I_1 = 1) = 1 - \mathbb{P}(I_1 = 0) =: p \in (0, 1]$, and where we impose the convention $\sum_{n=1}^{0} a_n = 0$. This is also known as binomial subsampling, since the conditional distribution of $N^\star$ given $N$ can be seen to correspond to a binomial distribution with number of trials $N$ and success probability $p$.

Consider again the example with two different contractual periods. If we disregard potential seasonality, then we can expect the claim occurrences to be uniformly distributed throughout the contractual period, and thus (2.3.2) offers an appropriate relationship between the number of claims within differing contractual periods as long as we set $p$ equal to the factor corresponding to the relative difference in period lengths; in the above example with contractual periods of one year and one month, respectively, we would thus set $p = 1/12$.

The example concerning deductibles is more involved. Let $(X_n)_{n\in\mathbb{N}}$ be the claim sizes, and assume they are independent and identically distributed as well as independent of the claim count $N$. Let $d \geq 0$ be some fixed deductible. Set $I_n = \mathbb{1}_{\{X_n > d\}}$, $n \in \mathbb{N}$. If only claims of size above $d$ are reported, then the number of reported claims is exactly

$$N^\star = \sum_{n=1}^{N} I_n,$$

which is (2.3.2) with $p = \mathbb{P}(I_1 = 1) = \mathbb{P}(X_1 > d) = \overline{G}(d)$; here $\overline{G}$ is the survival function of $X_1$. The sequence $(I_n)_{n\in\mathbb{N}}$ consists of independent and identically distributed random variables which are independent of $N$ since this was the case for the sequence

$(X_n)_{n\in\mathbb{N}}$. In the above, we implicitly assumed $\overline{G}(d) > 0$ to ensure $p \in (0,1]$. The alternative case of $\overline{G}(d) = 0$ is not relevant in practice, since it presumes that claim sizes are unable to exceed the deductible (with probability one).

We conclude this section, and thus the chapter on basic insurance risks, by showing that the Poisson distribution is closed under thinning.

**Proposition 2.3.3.** *Suppose that $N \sim$ Poisson$(\lambda)$, $\lambda > 0$. Then $N^\star \sim$ Poisson$(\lambda^\star)$ with $\lambda^\star = p\lambda$.*

*Proof.* Note that $N^\star$ takes values in $\mathbb{N}_0$ since this is the case also for $N$. Fix $n \in \mathbb{N}_0$. By the law of iterated expectations, it holds that

$$\begin{aligned}\mathbb{P}(N^\star = n) &= \mathbb{E}[\mathbb{P}(N^\star = n \,|\, N)] \\ &= \sum_{k=0}^{\infty} \mathbb{P}(N^\star = n \,|\, N = k)\mathbb{P}(N = k) \\ &= \sum_{k=n}^{\infty} \mathbb{P}(N^\star = n \,|\, N = k)\mathbb{P}(N = k).\end{aligned}$$

In the last line we have used that $N^\star \le N$ by definition. Now, since the sequence $(I_n)_{n\in\mathbb{N}}$ is independent of $N$, we find that

$$\begin{aligned}\mathbb{P}(N^\star = n \,|\, N = k) &= \mathbb{P}\left(\sum_{\ell=1}^{k} I_\ell = n\right) \\ &= \frac{k!}{n!(k-n)!}p^n(1-p)^{k-n},\end{aligned}$$

where the last line rests on the fact that $\sum_{\ell=1}^{k} I_\ell$ follows a binomial distribution with number of trials $k$ and success probability $p$. Recall that since $N \sim \text{Poisson}(\lambda)$, we have

$$\mathbb{P}(N = k) = \frac{\lambda^k}{k!}\exp\{-\lambda\}.$$

Collecting results then yields

$$\begin{aligned}\mathbb{P}(N^\star = n) &= \sum_{k=n}^{\infty} \frac{k!}{n!(k-n)!}p^n(1-p)^{k-n}\frac{\lambda^k}{k!}\exp\{-\lambda\} \\ &= \frac{(p\lambda)^n}{n!}\exp\{-p\lambda\}\sum_{k=n}^{\infty}\frac{\left((1-p)\lambda\right)^{k-n}}{(k-n)!}\exp\{-(1-p)\lambda\} \\ &= \frac{(p\lambda)^n}{n!}\exp\{-p\lambda\}\sum_{k=0}^{\infty}\frac{\left((1-p)\lambda\right)^{k}}{k!}\exp\{-(1-p)\lambda\}.\end{aligned}$$

Note that

$$\frac{((1-p)\lambda)^k}{k!}\exp\{-(1-p)\lambda\}, \qquad k \in \mathbb{N}_0,$$

corresponds to the probability mass function of the Poisson distribution with mean $(1-p)\lambda$, so that

$$\sum_{k=0}^{\infty}\frac{((1-p)\lambda)^k}{k!}\exp\{-(1-p)\lambda\} = 1,$$

which implies

$$\mathbb{P}(N^\star = n) = \frac{(p\lambda)^n}{n!}\exp\{-p\lambda\}.$$

We conclude that $N^\star$ follows a Poisson distribution with mean $p\lambda$ as desired. □

## 2.4 Exercises

**Exercise I.** Consider the survival function $\overline{F}(t) = \left(1 - \frac{t}{110}\right)^{\frac{3}{4}}$, $0 \le t \le 110$.

a) Compute the probability that a newborn survives until age 30 and the probability that a newborn dies between ages 60 and 70.

b) Compute the probability that an insured of age 65 survives until age 80.

c) Compute the mortality rate at age 75.

**Exercise II.** Consider the constant mortality rate $\mu \equiv 0.004$.

a) Compute $\overline{F}_{10}(30)$ and $F_5(15)$. Interpret your results.

b) Compute an age $x \ge 0$ such that a newborn has probability 0.9 of surviving until age $x$.

**Exercise III.** Consider the survival function $\overline{F}(t) = \exp\{-\beta t^\alpha\}$, $t \ge 0$, where $\alpha > 0$ and $\beta > 0$. Compute values of $\alpha$ and $\beta$ such that $\mu(40) = 4\mu(10)$ and $\mu(25) = 0.001$.

**Exercise IV.** The Weibull distribution (or mortality law) is characterized by the mortality rate

$$\mu(t) = \beta\alpha^{-\beta}t^{\beta-1}, \qquad t \ge 0,$$

for $\alpha > 0$ and $\beta > 0$. We also write $T \sim \text{Weibull}(\alpha, \beta)$.

a) Show that $\overline{F}(t) = \exp\left\{-\left(\frac{t}{\alpha}\right)^\beta\right\}$.

b) Show that $\mu$ is increasing for $\beta > 1$, decreasing for $\beta < 1$, and that the Weibull distribution equals the exponential distribution with mean $\alpha$ for $\beta = 1$.

c) Calculate $\overline{F}_x(t)$.

**Exercise V.** The Gompertz-Makeham distribution (or mortality law) is characterized by the mortality rate

$$\mu(t) = \alpha + \beta \exp\{\gamma t\}, \qquad t \geq 0,$$

for $\alpha > 0$, $\beta > 0$, and $\gamma \in \mathbb{R}$.

a) Show that if $\gamma > 0$, then $\mu$ is a strictly increasing function.

b) Show that $\overline{F}(t) = \exp\Big\{-\alpha t - \frac{\beta}{\gamma}\left(\exp\{\gamma t\} - 1\right)\Big\}$.

c) Calculate $\overline{F}_x(t)$.

**Exercise VI.** Let $x \geq 0$ and $t \geq 0$ be given. Prove that

$$\mathbb{P}(T_x > s \,|\, T_x > t) = \mathbb{P}(T_{x+t} > s - t) = \overline{F}_{x+t}(s - t), \qquad s \geq t.$$

**Exercise VII.** Let $X$ and $Y$ be non-negative random variables, and suppose that $X$ and $Y$ are independent. Let $x \geq 0$ and $t \geq 0$ be given. In this exercise we are going to show that

$$\mathbb{P}(X > t \,|\, X + Y > x) \geq \mathbb{P}(X > t) \tag{2.4.1}$$

whenever the left-hand side is well-defined, that is whenever $\mathbb{P}(X + Y > x) > 0$.

a) Let $\overline{F}$ denote the survival function of $X$. Using the law of iterated expectations, show that

$$\mathbb{P}(X > t \,|\, X + Y > x) = \frac{\mathbb{E}[\overline{F}(\max\{t, x - Y\})]}{\mathbb{E}[\overline{F}(x - Y)]}.$$

b) Argue that $\overline{F}(\max\{t, x - Y\}) \geq \overline{F}(t)\overline{F}(x - Y)$ and use this to establish (2.4.1).

c) Construct a counterexample to (2.4.1) in case $X$ and $Y$ are not independent.

One may actually extend (2.4.1) quite considerably. It is for example possible to show that

$$\mathbb{P}(aX + bY > t \,|\, X + Y > x) \geq \mathbb{P}(aX + bY > t) \tag{2.4.2}$$

for $x \geq 0$, $t \geq 0$, $a \geq 0$, and $b \geq 0$. One can also further generalize (2.4.2) from two to multiple independent random variables. To prove (2.4.2), one can proceed as above while using the fact that $\mathrm{Cov}[u(X), v(X)] \geq 0$ whenever both real-valued functions $u$ and $v$ are non-decreasing and as long as the covariance is well-defined[3].

[3]This fact is quite elegantly proven by letting $X_1$ and $X_2$ be two independent and identically distributed copies of $X$, by noting that $(u(X_1) - u(X_2))\,(v(X_1) - v(X_2)) \geq 0$ since $g$ and $h$ are non-decreasing, and by then taking the expectation.

**Exercise VIII.** One way to categorize distributions with survival functions that admit decay at a rate between power decay and exponential decay is to study whether the following relation holds:

$$\forall x \geq 0 : \overline{G}_d(x) \to 1 \qquad (d \to \infty). \tag{2.4.3}$$

Recall that

$$\overline{G}_d(x) = \frac{\overline{G}(d+x)}{\overline{G}(d)} = \frac{\mathbb{P}(X > d+x)}{\mathbb{P}(X > d)},$$

so (2.4.3) reads $\mathbb{P}(X > d + x) \sim \mathbb{P}(X > d)$ (as $d \to \infty$) for any $x \geq 0$. The interpretation goes as follows. Once $X$ has exceeded a high threshold $d$, it is very likely to exceed an ever higher threshold $d + x$. This is a weaker notion of heavy-tailedness than that of regular variation introduced in Example 2.2.4.

a) Suppose that $X \sim \text{Weibull}(\alpha, \beta)$, $\alpha > 0$, $\beta > 0$, confer with Exercise IV. Investigate the condition (2.4.3) for $\beta \geq 1$ and $\beta < 1$.

b) What does your solution to (a) tell you about the condition (2.4.3) when $X$ follows an exponential distribution?

c) Suppose instead that $X \sim \text{Pareto}(\alpha, \theta)$, $\alpha > 0$, $\theta > 0$, confer also with Example 2.2.4. Show that (2.4.3) holds.

Actually, one may show that all regularly varying distributions satisfy the condition (2.4.3). One may also show that the Weibull distribution is not regularly varying in the sense of (2.2.2). Thus the condition (2.4.3) is indeed the weaker notion of heavy-tailedness (compared to regular variation).

**Exercise IX.** Suppose that $X \sim \text{Gamma}(\alpha, \beta)$, $\alpha > 0$, $\beta > 0$, that is suppose that $X$ follows a Gamma distribution of shape $\alpha$ and rate $\beta$, so that the survival function takes the form

$$\overline{G}(x; \alpha, \beta) = \int_x^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} \exp\{-\beta y\} \, \mathrm{d}y.$$

Here the so-called Gamma function $\Gamma$ is given by

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \exp\{-x\} \, \mathrm{d}x, \qquad \alpha > 0,$$

and it is possible to show that $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$ for $\alpha > 0$, so that the Gamma function is a generalization of the factorial.

a) Show that

$$\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp\{-\beta x\}, \qquad x \geq 0,$$

is actually a valid probability density function, that is show that

$$\int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp\{-\beta x\} \, \mathrm{d}x = 1.$$

b) Let $c > 0$. Show that the random variable $Y = cX$ follows a Gamma distribution of shape $\alpha$ and rate $\frac{\beta}{c}$.

c) Show either by direct calculation of by deriving the relevant moment-generating function that

$$\mathbb{E}[X^\gamma] = \frac{\Gamma(\gamma+\alpha)}{\beta^\gamma \Gamma(\alpha)}, \qquad \gamma \in \mathbb{N}.$$

Use this to conclude that $\mathbb{E}[X] = \frac{\alpha}{\beta}$ and $\text{Var}[X] = \frac{\alpha}{\beta^2}$.

d) Show that

$$e(d) = \frac{\alpha}{\beta} \frac{\overline{G}(d;\alpha+1,\beta)}{\overline{G}(d;\alpha,\beta)} - d, \qquad d \geq 0.$$

e) Show that $e(d) \to \frac{1}{\beta}$ as $d \to \infty$. Discuss how your results relate to Figure 2.2.

**Exercise X.** This exercise is devoted to the negative binomial distribution and to the fact that it may be cast as a Poisson-Gamma mixture distribution. Suppose that $X \sim \text{Gamma}(\alpha, \beta)$, $\alpha > 0$, $\beta > 0$, and suppose that $N$ conditionally on $X$ follows a Poisson distribution with mean $X$. Finally, let $p = \frac{\beta}{1+\beta}$.

a) Show that $N \sim \text{NegBin}(\alpha, p)$.

b) Using the law of iterated expectations, show that

$$\mathbb{E}[N] = \alpha \frac{1-p}{p}.$$

c) Using the law of total variance, show that

$$\text{Var}[N] = \alpha \frac{1-p}{p^2}.$$

**Exercise XI.** Let $M$ be a non-negative discrete random variable with probability mass function $p$. We say that $N$ is a zero-modified version of $M$ if

$$N = IM,$$

where $I$ is independent of $M$ and $\mathbb{P}(I=1) = 1 - \mathbb{P}(I=0) = q$ for some $q \in (0,1)$.

a) Express $\mathbb{E}[N^k]$, $k \in \mathbb{N}$, in terms of $\mathbb{E}[M^k]$, $k \in \mathbb{N}$. What is $\text{Var}[N]$?

b) Determine the probability mass function $\mathbb{P}(N = n)$, $n \in \mathbb{N}_0$, of $N$ in terms of the probability mass function $p$ of $M$.

c) The zero-modified Poisson distribution corresponds to the zero-modified version of $M$ with $M \sim \text{Poisson}(\lambda)$, $\lambda > 0$. Show that the zero-modified Poisson distribution, contrary to the ordinary Poisson distribution, is able to capture overdispersion.

**Exercise XII.** Let $N$ be a non-negative discrete random variable, and let $M$ denote the moment-generating function of $N$:

$$M(t) = \mathbb{E}[\exp\{tN\}], \qquad t \in \mathbb{R}.$$

a) Suppose that $N$ follows a binomial distribution with number of trials $n \in \mathbb{N}$ and success probability $p \in (0, 1]$. Show that

$$M(t) = (p \exp\{t\} + 1 - p)^n, \qquad t \in \mathbb{R}.$$

Hint: It might be helpful to recall the binomial theorem:

$$(x + y)^n = \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} x^{n-k} y^k.$$

b) Suppose that $N \sim \text{NegBin}(1, 1 - p)$, $p \in (0, 1)$. Show that

$$M(t) = \frac{1 - p}{1 - p \exp\{t\}}, \qquad t < -\log\{p\}.$$

Hint: It might be helpful to recall that whenever $r \in (-1, 1)$, it holds that

$$\sum_{n=0}^{\infty} r^n = \frac{1}{1 - r}.$$

**Exercise XIII.** Let $N \sim \text{NegBin}(1, 1 - q)$, $q \in (0, 1)$, and let $(I_n)_{n \in \mathbb{N}}$ be a sequence of independent and identically distributed random variables assumed independent of $N$ and required to satisfy $\mathbb{P}(I_1 = 1) = 1 - \mathbb{P}(I_1 = 0) =: p \in (0, 1]$. Consider the thinned version $N^\star$ of $N$ given by

$$N^\star = \sum_{n=1}^{N} I_n.$$

Show that $N^\star \sim \text{NegBin}(1, 1 - q^\star)$ with $q^\star = \frac{pq}{pq + 1 - q}$.

Hint: Apply the the results of Exercise XII in a suitable manner. It might be helpful to recall that the moment-generating function of $N^\star$ determines the distribution of $N^\star$.

# Chapter 3

# Valuation of insurance liabilities

In the previous chapter, we studied a selection of basic insurance risks, including mortality risk. Insurance is all about the transfer of such risk, or uncertainty, from the insured to the insurer, in the sense that the insured pays some *premium* in order to receive certain *benefits*, which are more or less complicated functions of the underlying risk factors. The term premium (in Danish: præmie) is actuarial lingo. It does not refer to a prize but rather to an admission fee typically payed in advance, and actuaries tend to denote the premium, or premium amount, by the Greek letter $\pi$.

From the point of view of the insurer, the benefit received by the insured is a liability. The *valuation* of insurance liabilities consists of two aspects: pricing, that is the determination of reasonable premiums, and reserving, that is the determination of accounting entries that from a legal standpoint properly reflect the liabilities of the insurer. Generally speaking, valuation is closely related to quantitative risk management in general. The latter is a broad topic which lies outside the scope of these lecture notes. In the present chapter, we simply aim at giving a short introduction to the two core aspects of valuation: actuarial principles and the time value of money.

The chapter is structured as follows. In Section 3.1, we introduce and study some actuarial valuation principles. For long-term contracts, the so-called time value of money plays an important role. This is the focus of Section 3.2. We conclude the chapter by discussing in Section 3.3 valuation of certain random payments streams appearing first and foremost in life insurance.

## 3.1 Actuarial valuation principles

This section is devoted to actuarial valuation principles. In the following, we consider insurance benefits $Y$; this is a non-negative random variable, describing in monetary units the benefits received by the insured – from the point of view of the insurer, the benefits constitute losses or liabilities. Examples include:

- $Y = \mathbb{1}_{\{T_x > \eta\}}$ with $T_x$ the residual lifetime of an insured of age $x$, see also Section 2.1. The benefits consist of one monetary unit if the insured remains alive at time $\eta$. This type of contract is termed pure endowment insurance (in Danish: ren oplevelesforsikring).
- $Y = (X - d)\mathbb{1}_{\{X > d\}}$ with $X$ some claim size and $d$ some deductible, see also Section 2.2. The benefits consist of the excess $X - d$ so long as the claim size actually exceeds the deductible.

For the pure endowment insurance, payout actually occurs at time $\eta$, which may be decades into the future, so one should take the time value of money (interest) into account; this crucial aspect, which is related to financial mathematics, is postponed to future sections. In the current section, we focus solely on classic actuarial aspects in relation to the basic insurance risks of the previous chapter. The main questions are: For given insurance benefits, what should naturally be required from premiums and which general premium formulas fulfill these requirements?

In the following, premiums are denoted by the Greek letter $\pi$. The simplest principle is the so-called *equivalence principle*, stating that

$$\pi(Y) = \mathbb{E}[Y], \tag{3.1.1}$$

where we are implicitly assuming that $Y$ follows a distribution with finite mean. The premium determined via (3.1.1) is also called the *equivalence premium*, and it is based off the Law of Large Numbers. In the following, it is helpful to think of the insured as being part of an infinitely large insurance portfolio of identical insured with identical insurance contracts. If the insured are independent, and if we denote by $Y_1, Y_2, \dots$ the corresponding benefits, which are thus independent and identically distributed replicates of $Y$, then the Law of Large Numbers states that

$$\frac{1}{n}\sum_{k=1}^{n} Y_k \to \mathbb{E}[Y] \qquad (n \to \infty). \tag{3.1.2}$$

Here the convergence may be understood in two distinct ways, either as convergence in probability, that is

$$\forall \epsilon > 0 : \lim_{n\to\infty} \mathbb{P}\left( \left| \mathbb{E}[Y] - \frac{1}{n}\sum_{k=1}^{n} Y_k \right| \geq \epsilon \right) = 0, \tag{3.1.3}$$

or almost sure convergence, that is

$$\mathbb{P}\left( \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} Y_k = \mathbb{E}[Y] \right) = 1.$$

It turns out that almost sure convergence implies convergence in probability, so the latter identity is a stronger result than the former and thus also known as the Strong Law of Large Numbers.

From the above results, the equivalence principle would appear to be quite reasonable. But insurers, just as insured, are typically risk averse, meaning they prefer certainty over uncertainty. The relation given by (3.1.2) is on the form of a limit which is of course never achieved in practice, and the expectation of $Y$ is typically unknown and would have to be estimated from data. In combination, this entails additional unsystematic and systematic risk, confer also with the discussion concerning mortality risk in the beginning of Section 2.1, which is not captured by the equivalence premium. By charging only the equivalence premium, the insurer adopts this risk without any reward: What incentive would it have to do this? Nonetheless, the equivalence principle in particular and expectation operators in general, including conditional expectations, play a very prominent role especially in life insurance and this for a multitude of reasons, some of which closely relate to regulatory requirements. Just to exemplify, the Solvency II regulatory framework of the European Union, which has been in force since 2016, states that "The best estimate shall correspond to the probability-weighted average (...)". Further details lie beyond the scope of this presentation.

Before we discuss more advanced premium principles, we introduce some desirable properties that premium principles, for various reasons, should satisfy. Several of the properties are related to the so-called *principle of no arbitrage* from finance, which basically states that it should not be possible to forge risk-free gains.

Generally speaking, a premium principle $\pi$ is a functional which maps a non-negative random variable to $[0, \infty)$. The exact domain of $\pi$, which we denote by $\mathcal{Y}$ may vary, also depending on the context. For the equivalence principle, the domain consists of all non-negative random variables with finite mean.

Definition 3.1.1 provides an inexhaustive list of desirable properties for premium principles, some of which are consistent with the principle of no arbitrage and some which are, in general, not. One finds many other important properties in the literature, such as law invariance, convexity, and independent or comonotonic additivity. Further details lie beyond the scope of these lecture notes.

**Definition 3.1.1.** *A premium principle $\pi$ is said to*

- *induce a* ***non-negative loading*** *if $\pi(Y) \geq \mathbb{E}[Y]$ for all $Y \in \mathcal{Y}$,*
- *be* ***monotonic*** *if $\pi(Y_1) \leq \pi(Y_2)$ for all $Y_1, Y_2 \in \mathcal{Y}$ satisfying $Y_1 \leq Y_2$,*
- *be* ***translation invariant*** *if $\pi(Y + c) = \pi(Y) + c$ for all $Y \in \mathcal{Y}$ and $c > 0$,*
- *be* ***subadditive*** *if $\pi(Y_1 + Y_2) \leq \pi(Y_1) + \pi(Y_2)$ for all $Y_1, Y_2 \in \mathcal{Y}$.*

The desirability of monotonicity and translation invariance are quite self-explanatory. Furthermore, if the premium principle $\pi$ is not monotonic or translation invariant, then one may forge an arbitrage, that is a risk-free gain. Say $Y_1 \leq Y_2$ but $\pi(Y_2) < \pi(Y_1)$. Then by selling $Y_1$ at the price $\pi(Y_1)$ and buying $Y_2$ at the price $\pi(Y_2)$, one obtains the net gain $\pi(Y_1) - Y_1 - \pi(Y_2) + Y_2 = (\pi(Y_1) - \pi(Y_2)) + (Y_2 - Y_1) > 0$.

*Remark* 3.1.2 (Non-negative loading). Let $Y_1, Y_2, \ldots$ be independent and identically distributed replicates of $Y$. Suppose that $\pi(Y) < \mathbb{E}[Y]$, that is suppose $\pi$ does not induce a non-negative loading. Letting $\epsilon = \mathbb{E}[Y] - \pi(Y) > 0$, we find that (3.1.3) reads

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\mathbb{E}[Y] - \frac{1}{n}\sum_{k=1}^{n} Y_k\right| \geq \mathbb{E}[Y] - \pi(Y)\right) = 0,$$

so that

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\mathbb{E}[Y] - \frac{1}{n}\sum_{k=1}^{n} Y_k\right| < \mathbb{E}[Y] - \pi(Y)\right) = 1,$$

which implies

$$\lim_{n\to\infty} \mathbb{P}\left(\mathbb{E}[Y] - \frac{1}{n}\sum_{k=1}^{n} Y_k < \mathbb{E}[Y] - \pi(Y)\right) = 1.$$

Canceling terms and rearranging yields

$$\lim_{n\to\infty} \mathbb{P}\left(\left(\sum_{k=1}^{n} Y_k\right) - n\pi(Y) > 0\right) = 1.$$

In other words, the total net loss of the insurer given by the total benefits $\sum_{k=1}^{n} Y_k$ subtracted total premiums $n\pi(Y)$ becomes positive with probability one as the size of the insurance portfolio $n$ grows to infinity. This is unsatisfactory for the insurer and thus helps to explain why it is desirable to require premium principles to induce a non-negative loading. $\triangledown$

*Remark* 3.1.3 (Subadditivity and additivity). Given two benefits $Y_1$ and $Y_2$, one may desire the premium for the aggregated benefit $Y_1 + Y_2$ to not be larger than the premium for $Y_1$ added the premium for $Y_2$, that is

$$\pi(Y_1 + Y_2) \leq \pi(Y_1) + \pi(Y_2). \tag{3.1.4}$$

If this was not the case, insured could obtain premium discounts by splitting up contracts. We could even go one step further and consider additivity, where, obviously, the inequality of (3.1.4) is an equality. Unless additivity is fulfilled, it is possible to forge an arbitrage as follows. Sell $Y_1$ at the price $\pi(Y_1)$, sell $Y_2$ at the price $\pi(Y_2)$, and buy $Y_1 + Y_2$ at the price $\pi(Y_1 + Y_2)$. If $\pi(Y_1 + Y_2) < \pi(Y_1) + \pi(Y_2)$, then the net gain reads

$$\pi(Y_1) - Y_1 + \pi(Y_2) - Y_2 - \pi(Y_1 + Y_2) + (Y_1 + Y_2) = \pi(Y_1) + \pi(Y_2) - \pi(Y_1 + Y_2) > 0.$$

So only additivity, and not subadditivity, is consistent with the principle of no arbitrage. $\triangledown$

We now discuss in more detail two common premium principles, namely the *expected value principle* and the *standard deviation principle*. We should like to stress that there exists a plethora of more or less popular premium principles, including the *variance principle* (which is the subject of Exercise XIV), the Esscher principle, and the *quantile principle*. The latter is of the form

$$\pi(Y) = \inf\{y \in [0, \infty) : \mathbb{P}(Y > y) \leq q\}, \qquad q \in (0, 1).$$

The diversity in premium principles reflects the difficulty in proposing a straightforward principle which is both easily computable and possesses desirable properties such as those found in Definition 3.1.1.

**Example 3.1.4** (Expected value principle)**.** The *expected value principle* is given by

$$\pi(Y) = (1 + \alpha)\mathbb{E}[Y], \qquad Y \in \mathcal{Y}, \tag{3.1.5}$$

where $\alpha > 0$ is a loading factor, and where $\mathcal{Y}$ consists of all non-negative random variables with finite mean.

Obviously, the expected value principle induces a non-negative loading. Furthermore, if $Y_1 \leq Y_2$, then the random variable $Y_2 - Y_1$ is non-negative, which implies

$$\mathbb{E}[Y_2 - Y_1] \geq 0,$$

so that

$$\mathbb{E}[Y_1] \leq \mathbb{E}[Y_2],$$

which again entails

$$\pi(Y_1) = (1 + \alpha)\mathbb{E}[Y_1] \leq (1 + \alpha)\mathbb{E}[Y_2] = \pi(Y_2).$$

We conclude that the expected value principle is also monotonic. Finally,

$$\pi(Y_1 + Y_2) = (1 + \alpha)\mathbb{E}[Y_1 + Y_2] = (1 + \alpha)\mathbb{E}[Y_1] + (1 + \alpha)\mathbb{E}[Y_2] = \pi(Y_1) + \pi(Y_2),$$

which shows that the expected value principle is additive and thus in particular subadditive.

The above properties also hold if instead $\alpha = 0$, in which case (3.1.5) corresponds to the equivalence principle. Since

$$\mathbb{E}[Y + c] = \mathbb{E}[Y] + c, \qquad c \in \mathbb{R},$$

we find that the equivalence principle is also translation invariant. This is, however, not the case for the expected value principle, since

$$(1 + \alpha)\mathbb{E}[Y + c] = (1 + \alpha)\mathbb{E}[Y] + (1 + \alpha)c > (1 + \alpha)\mathbb{E}[Y] + c$$

whenever $c \neq 0$ and $\alpha > 0$. ∘

**Example 3.1.5** (Standard deviation principle)**.** The *standard deviation principle* is given by

$$\pi(Y) = \mathbb{E}[Y] + \alpha\sqrt{\mathrm{Var}[Y]}, \qquad Y \in \mathcal{Y},$$

where $\alpha > 0$ is a loading factor, and where $\mathcal{Y}$ consists of all non-negative random variables with finite variance.

Obviously, the standard deviation principle induces non-negative loading. Since

$$\mathbb{E}[Y + c] = \mathbb{E}[Y] + c \quad \text{and} \quad \mathrm{Var}[Y + c] = \mathrm{Var}[Y]$$

for all $c \in \mathbb{R}$, we may also conclude that the standard deviation principle is translation invariant. The principle is not, on the other hand, monotonic. This follows from a proof by contradiction. Note that if the principle $\pi$ is monotonic, then $Y \leq c$ for some $c > 0$ implies

$$\pi(Y) \leq \pi(c),$$

so that if the principle is also translation invariant, which ensures $\pi(c) = c$, we would have

$$\pi(Y) \leq c.$$

If we produce a counterexample yielding $\pi(Y) > c$, we would arrive at a contradiction from which we may conclude that the standard deviation principle cannot be both monotonic and translation invariant. This in turn would allow us to conclude that the standard deviation principle is not monotonic, since we have already shown that it is translation invariant. To this end, suppose that

$$\mathbb{P}(Y = 1) = 1 - \mathbb{P}(Y = 0) = p$$

for some $p \in (0, 1)$. Then straightforward calculations yield

$$\pi(Y) = p + \alpha\sqrt{p(1-p)}.$$

Note that $Y \leq 1$. To produce a counterexample, we thus have to show that we may find a $p \in (0, 1)$ for which $\pi(Y) > 1$. To this end, cast $h : [0, 1] \to \mathbb{R}$ according to

$$h(p) = p + \alpha\sqrt{p(1-p)}.$$

Note that $h(1) = 1$ and that

$$\frac{\mathrm{d}}{\mathrm{d}p} h(p) = 1 + \frac{1}{2}\frac{\alpha(1-2p)}{\sqrt{p(1-p)}}$$

for $p \in (0, 1)$, which yields

$$\lim_{p \uparrow 1} \frac{\mathrm{d}}{\mathrm{d}p} h(p) = -\infty.$$

In particular, there exists a $p \in (0,1)$ such that $h(p) > h(1) = 1$. Since $h(p) = \pi(Y)$, this choice of $p$ provides the desired counterexample. All in all we conclude that the standard deviation principle is not monotonic.

We conclude the example by verifying that the standard deviation principle is subadditive (but not in general additive). Recall that

$$\mathrm{Var}[Y_1 + Y_2] = \mathrm{Var}[Y_1] + \mathrm{Var}[Y_2] + 2\mathrm{Cov}[Y_1, Y_2],$$

and that

$$\mathrm{Cov}[Y_1, Y_2] \leq \sqrt{\mathrm{Var}[Y_1]}\sqrt{\mathrm{Var}[Y_2]}.$$

In combination, we then find

$$\begin{aligned}\sqrt{\mathrm{Var}[Y_1 + Y_2]} &= \sqrt{\mathrm{Var}[Y_1] + \mathrm{Var}[Y_2] + 2\mathrm{Cov}[Y_1, Y_2]} \\ &\leq \sqrt{\mathrm{Var}[Y_1] + \mathrm{Var}[Y_2] + 2\sqrt{\mathrm{Var}[Y_1]}\sqrt{\mathrm{Var}[Y_2]}} \\ &= \sqrt{\left(\sqrt{\mathrm{Var}[Y_1]} + \sqrt{\mathrm{Var}[Y_2]}\right)^2} \\ &= \sqrt{\mathrm{Var}[Y_1]} + \sqrt{\mathrm{Var}[Y_2]}.\end{aligned}$$

In particular,

$$\begin{aligned}\pi(Y_1 + Y_2) &= \mathbb{E}[Y_1 + Y_2] + \alpha\sqrt{\mathrm{Var}[Y_1 + Y_2]} \\ &\leq \mathbb{E}[Y_1] + \mathbb{E}[Y_2] + \alpha\left(\sqrt{\mathrm{Var}[Y_1]} + \sqrt{\mathrm{Var}[Y_2]}\right) \\ &= \pi(Y_1) + \pi(Y_2),\end{aligned}$$

which establishes subadditivity. ∘

There is a close relationship between actuarial valuation principles, which we have briefly discussed here, and so-called risk measures. In the last decade, risk measures have been subject to an extensive debate in academia as well as in the industry. The discussions are still ongoing and remain as relevant as ever; they go far beyond premium calculations and deal with important questions such as

- When insurers present a snapshot of their business to their stakeholders and the financial world, which information is relevant, important, and/or sufficient?
- When the insurers defend their state of financial health to the supervisory authorities, which information is relevant, important, and/or sufficient?

An important risk measure, which also plays a major role in the aforementioned Solvency II regulatory framework of the European Union, is the so-called *Value-at-Risk* (abbreviated VaR). It is very similar to the quantile principle, which we briefly mentioned earlier. While quite simple, it is a fundamental tool for modern risk managers, who (if not already sooner, then at the very least after the financial crisis of 2007-2008) have become aware of the fact that key risks may be non-Gaussian, in which case standard deviations do not paint a full picture of the potential downsides of risk-taking.

## 3.2 Time value of money

In especially life insurance, long-term contracts are prevalent. The classic example is a pension scheme, which may consist of periodic premium payments until retirement followed by periodic benefits until death. Here, the first premium payment may easily occur forty years in advance, and one should suspect such significant time lag to possibly have an impact on the valuation of the contract. In life insurance in particular, it is quite common to use the nomenclature contribution rather than premium payment. In the following, the two terms are used interchangeably.

In this section, we study the value of one monetary unit paid at one point in time measured at another point in time and discuss the associated time value of money. Our main goal is to give a precise notion to the value of a stream of deterministic payments.

The value of one monetary unit in time measured at another point in time is driven by the supply and demand for capital. Financial markets act as intermediaries between parties who at some point in time earn more money than they spend and parties who earn less money than they spend. In its most narrow sense, the time value of money is the widely accepted conjecture that it is advantageous to receive a benefit today rather than later due to its earning potential. So in particular, it helps explain why interest is paid or earned.

In especially life insurance, payments do not occur only on an annual basis. As an example, pension contributions and benefits typically occur on a monthly basis. This encourages us to develop an interest theory in continuous time. Later, we shall even go one step further and approximate for example monthly payments by payment rates, since this leads to more mathematically tractable results.

The value of one monetary unit at time $s \geq 0$ measured at time $t \geq 0$ is denoted $v(t,s)$. If $t < s$, then $v(t,s)$ is called the $(s-t)$-*discount factor*, while if $t > s$ then $v(t,s)$ is called the $(t-s)$-*interest factor*; they are naturally assumed strictly positive. If interest accumulates periodically, say annually at a constant rate $i > -1$, then we should have

$$v(n,0) = (1+i)^n \qquad \text{and} \qquad v(0,n) = (1+i)^{-n}$$

for $n \in \mathbb{N}_0$, where the identity $v(n,0) = \frac{1}{v(0,n)}$ relates to the principle of no arbitrage, in the sense that interest rates for bank deposits and loans/debt are identical; here

$v(n,0)$ could for example be interpreted as the bank balance at time $n$ if you deposit one monetary unit at time 0. In more general fashion, and also in accordance with the principle of no arbitrage, we shall in the following require that

$$v(t,s) = v(t,u)v(u,s), \qquad t,u,s \geq 0. \tag{3.2.1}$$

If this requirement is not imposed, one may forge an arbitrage. For example, suppose one financial institution offers you an annual interest rate of 2% on both bank deposits and loans/debt, that is

$$v(1,0) = 1.02 \qquad \text{and} \qquad v(0,1) = \frac{1}{1.02},$$

while another financial institution offers you a biannual interest rate of 1% on both bank deposits and loans/debt, that is

$$v(1/2,0) = v(1,1/2) = 1.01 \qquad \text{and} \qquad v(1/2,1) = v(0,1/2) = \frac{1}{1.01}.$$

In this case, we have that

$$v(1,0) = 1.02 < 1.201 = 1.01 \cdot 1.01 = v(1/2,0)v(1,1/2),$$

and thus a risk-free gain may be forged as follows: Initially, borrow one monetary unit from the first financial institution. Deposit this monetary unit into the second financial institution for half a year and then again for another half a year. Afterwards, withdraw the accumulated deposit and use it to repay the initial loan. Finally, enjoy the net gain

$$v(1/2,0)v(1,1/2) - v(1,0) = 1.201 - 1.02 = 0.001 > 0.$$

We now turn to the implications of requiring (3.2.1). It implies with $u = t$ that $v(t,s) = v(t,t)v(t,s)$, so that

$$v(t,t) = 1, \qquad t \geq 0,$$

since we assumed the interest and discount factors to be strictly positive. Furthermore, letting instead $s = t$ and using $v(t,t) = 1$, we find that

$$v(t,u) = \frac{1}{v(u,t)}, \qquad t,u \geq 0,$$

so that the discount factor is the inverse of the interest factor. Finally, set $t = 0$ and use the above identity to obtain

$$v(u,s) = \frac{v(0,s)}{v(0,u)}, \qquad u,s \geq 0.$$

In other words, the one dimensional strictly positive *discount function* $v$ defined (with a slight abuse of notation) by

$$v(t) = v(0,t), \qquad t \geq 0,$$

fully determines the interest and discount factors.

**Example 3.2.1** (Constant force of interest)**.** Suppose that $v(t,s) = \exp\{r(t-s)\}$, $t, s \geq 0$, for some so-called constant force of interest $r \in \mathbb{R}$. It is straightforward to verify that (3.2.1) holds:

$$v(t,u)v(u,s) = \exp\{r(t-u) + r(u-s)\} = \exp\{r(t-s)\} = v(t,s), \qquad t, u, s \geq 0.$$

Similarly, the discount function $v$ given by

$$v(t) = \exp\{-rt\}, \qquad t \geq 0,$$

determines the interest and discount factors via

$$v(t,s) = \frac{v(s)}{v(t)}, \qquad t, s \geq 0.$$

There is a clear connection between constant annual interest rates and constant forces of interest, in the sense that the later provides a continuous generalization of the former, see also Exercise XVI. In particular, with $i = \exp\{r\} - 1$, we find that

$$v(n,0) = \exp\{rn\} = (1+i)^n \qquad \text{and} \qquad v(0,n) = \exp\{-rn\} = (1+i)^{-n}$$

for $n \in \mathbb{N}_0$. ○

In the following, we impose additional structure on the discount function $v$ in the form of the below assumption, which is much in the spirit of Assumption 2.1.2 from Chapter 2.

**Assumption 3.2.2.** *We assume that*

$$v(t) = \exp\left\{-\int_0^t r(u)\,\mathrm{d}u\right\}, \qquad t \geq 0,$$

*for some continuous function $r : [0,\infty) \to \mathbb{R}$ denoted the* ***force of interest****.*

The continuous function $r$ which is central to Assumption 3.2.2 is also known as the *instantaneous interest rate*; here the term instantaneous is used to distinguish it from say the annual interest rate, but it is also common to forgo the term instantaneous.

*Remark* 3.2.3. The interpretation of $r$ as the force of interest stems from the fact that

$$\frac{\mathrm{d}}{\mathrm{d}s} v(s,t)\bigg|_{s=t} = r(t)v(s,t)|_{s=t} = r(t), \qquad t \geq 0,$$

where the first identity follows from the First Fundamental Theorem of Calculus. ▽

**Exercise 3.2.4.** Verify that Assumption 3.2.2 is consistent with the previous assumptions of this section. That is, with $v(t) = e^{-\int_0^t r(s)\,\mathrm{d}s}$, $t \geq 0$, for some continuous force of interest $r : [0,\infty) \to \mathbb{R}$ and $v(t,s) = \frac{v(s)}{v(t)}$, $t, s \geq 0$, show that $v(t,s) > 0$ and that $v(t,s) = v(t,u)v(u,s)$ for all $t, u, s \geq 0$. ◇

Note that the discount functions of Example 3.2.1 satisfy Assumption 3.2.2; they simply correspond to the special case where the force of interest is constant.

We now turn our attention to streams of payments and the value of streams of payments at different points in time. A *payment stream* $B$ is a collection of (suitably regular) real-valued random variables $(B(t))_{t\geq 0}$ such that $B(t)$ describes the accumulated payments from time zero up to and including time $t$, $t \geq 0$. Payments may be both positive and negative; we adopt the convention that contributions (premiums) are negative and benefits are positive.

For now, we disregard any type of randomness in the payments; random payment streams are instead the focal point of the subsequent section. Consequently, we may think of $B$ simply as a function from $[0, \infty)$ to $\mathbb{R}$. To keep things general yet relatively simple, we restrict ourselves to following types of payments:

- Initial payments $b_0$. In an actuarial context, we typically have $b_0 \leq 0$ since premiums are to be paid in advance.
- Real-valued discrete payments $(b_n)_{n\in\mathbb{N}}$ at time points $0 < t_1 < t_2 < \cdots$ satisfying that $\sum_{n=1}^{\infty} \mathbb{1}_{\{t\leq t_n\}} b_n < \infty$ for all $t \geq 0$.
- Piecewise continuous payment rates $b : [0, \infty) \to \mathbb{R}$, in the sense that the payment $b(t + h/2)h$ occurs between time $t$ and time $t + h$ (as $h \downarrow 0$).

In other words, we suppose that

$$B(t) = b_0 + \int_0^t b(s)\,\mathrm{d}s + \sum_{n=1}^{\infty} \mathbb{1}_{\{t_n\leq t\}} b_n, \qquad t \geq 0. \tag{3.2.2}$$

Payments are actually discrete in nature, confer with Example 3.2.5 below, so it could suffice to only consider an initial payment and discrete payments. But small discrete payments occurring relatively frequently may be approximated by a payment rate, which is convenient since payment rates typically lead to more mathematically tractable results than discrete payments; the link between discrete payments and payment rates is also explored in Example 3.2.6 below.

**Example 3.2.5** (Discrete annuity certain)**.** An annuity is a sequence of payments with a fixed frequency, and the term annuity originally referred to annual payments. An annuity with a fixed number of payments is called an annuity certain, while an annuity whose number of payments depend on some event is called a contingent annuity. The following is a list of some common annuities certain.

- Let $t_n = n$ and $b_n = c$ for $n = 1, \ldots, k$, $k \in \mathbb{N}$, and zero otherwise, so that (3.2.2) reads

$$B(t) = \sum_{n=1}^{k} \mathbb{1}_{\{n\leq t\}} c, \qquad t \geq 0,$$

which yields

$$B(t) = c \min\{n, k\}, \qquad t \in [n, n+1), n \in \mathbb{N}_0.$$

This annuity certain is said to be immediate, since the payments are made in arrears, and it is said to be level, since all payments are equal (to $c$).

- Let $b_0 = c$, $t_n = n$, and $b_n = c$ for $n = 1, \dots, k-1$, $k \in \mathbb{N}$, and zero otherwise, so that (3.2.2) reads

$$B(t) = c + \sum_{n=1}^{k-1} \mathbb{1}_{\{n \le t\}} c, \qquad t \ge 0,$$

which yields

$$B(t) = c \min\{n+1, k\}, \qquad t \in [n, n+1), n \in \mathbb{N}_0.$$

This annuity certain is an annuity due, since the payments are not made in arrears, and it is said to be level, since all payments are equal (to $c$).

- Let $t_n = n/m$ and $b_n = c/m$ for $n = 1, \dots, km$, $k, m \in \mathbb{N}$, and zero otherwise, so that (3.2.2) reads

$$B(t) = \sum_{n=1}^{km} \mathbb{1}_{\{n/m \le t\}} c/m, \qquad t \ge 0, \tag{3.2.3}$$

which yields

$$B(t) = c/m \min\{n, km\}, \qquad t \in [n, n+1), n \in \mathbb{N}_0.$$

This annuity certain is also immediate and level, but payments occur at an increased frequency. Instead of one payment of $c$ per period, we now have $m$ payments of $c/m$ per period. If we think of the former as yearly payments, then the case $m = 12$ corresponds to monthly payments.

An example of a contingent annuity is a life annuity (in Danish: livrente), where payments are contingent on the insured staying alive. Life annuities are examples of random payment streams and thus discussed in more detail in the next section. ∘

**Example 3.2.6** (Continuous annuity certain)**.** By letting $m \to \infty$ in (3.2.3) we obtain the payment stream

$$B(t) = \int_0^{\min\{t,k\}} c \, \mathrm{d}s = c \min\{t, k\}, \qquad t \ge 0, \tag{3.2.4}$$

which is also known as a continuous annuity. It relates to (3.2.2) via $b(t) = c\mathbb{1}_{\{t \le k\}}$, $t \ge 0$. An annuity is said to be deferred if the payment period does not commence immediately. For example

$$b(t) = c\mathbb{1}_{\{u < t \le u+s\}}, \qquad t \ge 0,$$

correspond to a deferred continuous annuity where payments first commence at time $u \geq 0$ (and continue until time $u + s$, $s \geq 0$). ◦

We now turn our attention to valuation of payment streams. It is important to note that the value of the accumulated payments $B(t)$ measured at time zero in general is not simply $v(0,t)B(t)$, since the payments involved occur at different points in time between time zero and time $t \geq 0$. To take this into account, we introduce the notion of *present value* of a payment stream. Please note that this notion still applies even if the payment rate and discrete payments are random, confer with Section 3.3.

**Definition 3.2.7.** *The* ***present value*** *of the payment stream $B$ is*

$$PV(t) = \int_t^\infty v(t,s)b(s)\,\mathrm{d}s + \sum_{n=1}^{\infty} \mathbb{1}_{\{t<t_n\}} v(t,t_n)b_n, \qquad t \geq 0, \tag{3.2.5}$$

*whenever it is well-defined, that is finite.*

*Remark* 3.2.8. At time $t \geq 0$, the present value provides the accumulated value of all future payments at time $t$, where the future payments are all those payments that lie strictly after time $t$. In particular, the initial payment $b_0$ is of no importance.

The expression given by (3.2.5) is a consequence of the following considerations. Since the $n$'th discrete payment $b_n$ occurs exactly at time $t_n$, it is quite straightforward to argue that its value at time $t$ is $v(t,t_n)b_n$. Similarly, recall that the payment rate $b$ exactly yields a payment of $b(s+h/2)h$ between time $s$ and time $s+h$ (as $h \downarrow 0$), so that the value of this payment at time $t$ lies between $v(t,s)b(s+h/2)h$ and $v(t,s+h)b(s+h/2)h$ (as $h \downarrow 0$), where the difference between $v(t,s)$ and $v(t,s+h)$ vanishes as $h \downarrow 0$. In conjunction, this yields (3.2.5).

Due to Assumption 3.2.2, we may also cast (3.2.5) as

$$\mathrm{PV}(t) = \int_t^\infty \exp\left\{-\int_t^s r(u)\,\mathrm{d}u\right\} b(s)\,\mathrm{d}s + \sum_{n=1}^{\infty} \mathbb{1}_{\{t<t_n\}} \exp\left\{-\int_t^{t_n} r(u)\,\mathrm{d}u\right\} b_n,$$

for $t \geq 0$, whenever the integral and sum are well-defined, that is finite. ▽

If we think of $Y = \mathrm{PV}(0)$ as some (time-independent) future benefits in the spirit of Section 3.1 and suppose $Y \geq 0$, then it would be natural to determine $b_0$, which we could interpret as the negative of a positive initial premium, according to $b_0 = -\pi(Y)$ for some premium principle $\pi$. In this case, since $B$ and thus $Y$ is deterministic, translation invariance of $\pi$ would be equivalent to the requirement that

$$\mathrm{PV}(0) = -b_0, \tag{3.2.6}$$

where we have implicitly supposed $\pi(0) = 0$, which could be taken as another desirable property for premium principles. Actually, it would follow from another reasonable property for premium principles, namely *no rip-off*, which is said to mean

$$\pi(Y) \leq \inf\{0 \leq p < \infty : \mathbb{P}(Y \leq p) = 1\} \text{ for all } Y \in \mathcal{Y}.$$

The identity (3.2.6) is one of multiple motivations for the following fairness criterion, which applies to deterministic payment streams.

**Definition 3.2.9.** *The payment stream $B$ is said to be **fair** if*

$$PV(0-) := PV(0) + b_0 = 0. \tag{3.2.7}$$

*Remark* 3.2.10. Due to Assumption 3.2.2, we may also cast (3.2.7) as

$$b_0 + \int_0^\infty \exp\Big\{-\int_0^s r(u)\,\mathrm{d}u\Big\}\,b(s)\,\mathrm{d}s + \sum_{n=1}^\infty \exp\Big\{-\int_0^{t_n} r(u)\,\mathrm{d}u\Big\}\,b_n = 0.$$

We conclude this remark by recollecting a somewhat classic motivation for introducing the fairness condition (3.2.7). It is of course closely related to the initial motivation based on translation invariance of premium principles, in the sense that both motivations have their roots in the principle of no arbitrage.

Think of the payment stream $B$ as a financial contract between two parties, where negative payments correspond to a transaction from the first party to the second party, and positive payment correspond to a transaction from the second party to the first party. Suppose for simplicity that there are no transactions after time $t = \eta > 0$, that is, $B(t) = B(\eta)$ for $t > \eta$. Taking the time value of money into account, the net gain of the contract at time $t = \eta$ from the point of view of the first party is

$$b_0 v(\eta, 0) + \int_0^\eta v(\eta, t) b(t)\,\mathrm{d}t + \sum_{n=1}^\infty \mathbb{1}_{\{t_n \leq \eta\}} v(\eta, t_n) b_n, \tag{3.2.8}$$

while the net gain at time $t = \eta$ from the point of view of the second party is

$$-b_0 v(\eta, 0) - \int_0^\eta v(\eta, t) b(t)\,\mathrm{d}t - \sum_{n=1}^\infty \mathbb{1}_{\{t_n \leq \eta\}} v(\eta, t_n) b_n. \tag{3.2.9}$$

If (3.2.8) is negative, then the first party would feel cheated by the second party. On the other hand, if (3.2.9) is negative, then the second party would feel cheated by the first party. All in all, the contract appears to only be fair if

$$b_0 v(\eta, 0) + \int_0^\eta v(\eta, t) b(t)\,\mathrm{d}t + \sum_{n=1}^\infty \mathbb{1}_{\{t_n \leq \eta\}} v(\eta, t_n) b_n = 0,$$

which corresponds to (3.2.7) after multiplication with $v(0, \eta)$ (whenever $B(t) = B(\eta)$ for $t > \eta$). $\triangledown$

We have now achieved what we set out to do in the beginning of this section, namely to give a precise notion to the value of a stream of deterministic payments. In combination with the study of premium principles conducted in Section 3.1, we are thus ready to study in more detail valuation of streams of random payments, which lies at the center of attention of the next section. In preparation, we conclude this section by calculating the present value for two types of annuities certain.

**Example 3.2.11** (Annuities certain, continued)**.** This example builds on Example 3.2.5 and Example 3.2.6. Throughout it, we assume that the force of interest $r : [0,\infty) \to \mathbb{R}$ is constant and denote its value by $r$. Additionally, we set $i = \exp\{r\} - 1 > -1$.

Let $t_n = n$ and $b_n = c$ for $n = 1, \dots, k$, $k \in \mathbb{N}$, and zero otherwise, so that

$$B(t) = \sum_{n=1}^{k} \mathbb{1}_{\{n \le t\}} c, \qquad t \ge 0.$$

According to Definition 3.2.7, we have that

$$\mathrm{PV}(t) = c \sum_{n=1}^{k} \mathbb{1}_{\{t<n\}} \exp\{-r(n-t)\}, \qquad t \ge 0.$$

In particular, for $n = 0, \dots, k-1$,

$$\mathrm{PV}(n) = c \sum_{\ell=n+1}^{k} \exp\{-r(\ell - n)\} = c \sum_{\ell=1}^{k-n} \left(\frac{1}{1+i}\right)^{\ell},$$

so that

$$\mathrm{PV}(0-) = c \frac{1}{1+i} \sum_{\ell=0}^{k-1} \left(\frac{1}{1+i}\right)^{\ell}.$$

We recognize this as the partial sum of a geometric series, so that if $i \neq 0$, that is $r \neq 0$, one can show that

$$\mathrm{PV}(0-) = c \frac{1}{1+i} \frac{1 - \left(\frac{1}{1+i}\right)^{k}}{1 - \frac{1}{1+i}} = c \frac{1 - \frac{1}{(1+i)^k}}{i},$$

which for $i > 0$, that is $r > 0$, converges to $c\frac{1}{i}$ as $k$ goes to infinity.

Let instead $b(t) = c\mathbb{1}_{\{t \le s\}}$, $t \ge 0$, for some $s \ge 0$, so that

$$B(t) = \int_0^{\min\{t,s\}} c \, \mathrm{d}u = c \min\{t, s\}, \qquad t \ge 0.$$

According to Definition 3.2.7, we have that

$$\mathrm{PV}(t) = c \int_t^{\max\{t,s\}} \exp\{-r(u-t)\} \, \mathrm{d}u, \qquad t \ge 0,$$

so that in particular,

$$\mathrm{PV}(0-) = c\int_0^s \exp\{-ru\}\,\mathrm{d}u = c\frac{1-\exp\{-rs\}}{r}$$

whenever $r \neq 0$. Whenever $r > 0$, the above quantity converges to $c\frac{1}{r}$ as $s$ goes to infinity. ○

## 3.3 Valuation of random payment streams

In this section, we turn our attention to random payment streams. We focus on payments that are contingent on the survival of the insured, so that the randomness stems from the residual lifetime $T_x$ for some $x \geq 0$, confer with Section 2.1, and adopt the assumptions of Section 2.1 and Section 3.2.

In an attempt to strike an appropriate balance between mathematical tractability and practical relevance, we throughout this section assume that the random payment stream of interest $B$ is of the form

$$B(t) = b_0 + B^{\mathrm{a}}(t) + B^{\mathrm{d}}(t) + B^{\mathrm{ad}}(t), \qquad t \geq 0, \tag{3.3.1}$$

for some initial deterministic payment $b_0 \in \mathbb{R}$ and payments

$$\begin{aligned} B^{\mathrm{a}}(t) &= \int_0^t \mathbb{1}_{\{T_x > s\}} b^{\mathrm{a}}(s)\,\mathrm{d}s + \sum_{n=1}^{\infty} \mathbb{1}_{\{t_n \leq t\}} \mathbb{1}_{\{T_x > t_n\}} b_n^{\mathrm{a}}, \\ B^{\mathrm{d}}(t) &= \int_0^t \mathbb{1}_{\{T_x \leq s\}} b^{\mathrm{d}}(s)\,\mathrm{d}s + \sum_{n=1}^{\infty} \mathbb{1}_{\{t_n \leq t\}} \mathbb{1}_{\{T_x \leq t_n\}} b_n^{\mathrm{d}}, \\ B^{\mathrm{ad}}(t) &= \mathbb{1}_{\{T_x \leq t\}} b^{\mathrm{ad}}(T_x), \end{aligned} \tag{3.3.2}$$

for $t \geq 0$. Here the deterministic payment rates $b^{\mathrm{a}} : [0,\infty) \to \mathbb{R}$ and $b^{\mathrm{d}} : [0,\infty) \to \mathbb{R}$ and the deterministic payment upon death $b^{\mathrm{ad}} : [0,\infty) \to \mathbb{R}$ are assumed piecewise continuous, and the deterministic (real-valued) discrete payments $(b_n^{\mathrm{a}})_{n\in\mathbb{N}}$ and $(b_n^{\mathrm{d}})_{n\in\mathbb{N}}$ are required to satisfy $\sum_{n=1}^{\infty} \mathbb{1}_{\{t \leq t_n\}} b_n^{\mathrm{a}} < \infty$ and $\sum_{n=1}^{\infty} \mathbb{1}_{\{t \leq t_n\}} b_n^{\mathrm{d}} < \infty$ for all $t \geq 0$.

Note that if $b^{\mathrm{a}} = b^{\mathrm{d}}$ and $b^{\mathrm{ad}}(t) = 0$, $t \geq 0$, then the payments do not depend the residual lifetime $T_x$. In this case, we actually recover (3.2.2), which shows that the deterministic payment streams considered in Section 3.2 are simply special cases of the random payment streams given by (3.3.1)–(3.3.2).

Payments may be both positive and negative; similarly to Section 3.2, we adopt the convention that contributions (premiums) are negative and benefits are positive.

The payment stream $B$ may seem somewhat involved, but the interpretation is actually quite straightforward. The initial payment is $b_0$. Then, as long as the insured is alive, payments occur according to the payment rate $b^{\mathrm{a}}$ and the discrete payments $(b_n^{\mathrm{a}})_{n\in\mathbb{N}}$. Upon death, the payment $b^{\mathrm{ad}}(T_x)$ occurs, where $T_x$ is the time of death. Afterwards, payments occur according to the payment rate $b^{\mathrm{d}}$ and the discrete payments $(b^{\mathrm{d}}_n)_{n\in\mathbb{N}}$. The superscripts 'a' and 'd' are used to indicate whether the payments occur if the insured is 'alive' or 'dead', respectively.

While easy to interpret, the payment stream $B$ is sufficiently general to capture a wide range of life insurance contracts, confer with the examples below. In particular, the methods and results we develop in this section are relevant concerning valuation of life insurance liabilities in general.

**Example 3.3.1** (Pure endowment insurance)**.** A pure endowment insurance (in Danish: ren oplevelsesforsikring) pays out a certain amount, say $c$, at a certain point in time, say $\eta \geq 0$, if the insured remains alive at that point in time. It corresponds to the random payment stream

$$B(t) = c\mathbb{1}_{\{\eta\leq t\}}\mathbb{1}_{\{T_x>\eta\}}, \qquad t \geq 0,$$

which is of the form (3.3.1)–(3.3.2) with $t_1 = \eta$ and $b^{\mathrm{a}}_1 = c$. ◦

**Example 3.3.2** (Term insurance)**.** A term insurance (in Danish: ren dødsfaldsdækning) pays out a certain amount, say $c$, when the insured dies. It corresponds to the random payment stream

$$B(t) = c\mathbb{1}_{\{T_x\leq t\}}, \qquad t \geq 0,$$

which is of the form (3.3.1)–(3.3.2) with $b^{\mathrm{ad}}(t) = c$, $t \geq 0$. ◦

**Example 3.3.3** (Continuous life annuity)**.** A continuous life annuity (in Danish: kontinuert livrente) pays out at a certain rate, say $c$, as long as the insured is alive. It corresponds to the random payment stream

$$B(t) = c\int_0^t \mathbb{1}_{\{T_x>s\}}\,\mathrm{d}s = c\min\{t, T_x\}, \qquad t \geq 0,$$

which is of the form (3.3.1)–(3.3.2) with $b^{\mathrm{a}}(t) = c$, $t \geq 0$.

If the continuous life annuity is temporary (in Danish: ophørende), then payments cease after a certain point in time, say $\eta$, even if the insured is still alive. The corresponding random payment stream reads

$$B(t) = c\int_0^t \mathbb{1}_{\{T_x>s\}}\mathbb{1}_{\{s\leq\eta\}}\,\mathrm{d}s = c\min\{\min\{t, T_x\}, \eta\}, \qquad t \geq 0,$$

which is of the form (3.3.1)–(3.3.2) with $b^{\mathrm{a}}(t) = c\mathbb{1}_{\{t\leq\eta\}}$, $t \geq 0$.

A continuous life annuity may also be deferred (in Danish: opsat), in the sense that payments do not commence immediately but first at some later point in time, say $\eta$. The corresponding random payment stream reads

$$B(t) = c\int_0^t \mathbb{1}_{\{T_x>s\}}\mathbb{1}_{\{\eta<s\}}\,\mathrm{d}s = c\min\{\max\{t-\eta, 0\}, T_x-\eta\}, \qquad t \geq 0,$$

which is of the form (3.3.1)–(3.3.2) with $b^{\mathrm{a}}(t) = c\mathbb{1}_{\{\eta<t\}}$, $t \geq 0$.

Broadly speaking, a classic pension scheme is simply a combination of a temporary life annuity and a deferred life annuity in the following sense. Until retirement, say at time $\eta$, the insured pays contributions at some rate, say $c_1 < 0$. Following retirement, the insured receives benefits at some rate, say $c_2 > 0$, as long as the insured is alive. The corresponding random payment stream reads

$$B(t) = \int_0^t \mathbb{1}_{\{T_x>s\}} \left(\mathbb{1}_{\{s\leq\eta\}}c_1 + \mathbb{1}_{\{\eta<s\}}c_2\right) \mathrm{d}s, \qquad t \geq 0,$$

which is of the form (3.3.1)–(3.3.2) with $b^{\mathrm{a}}(t) = c_1\mathbb{1}_{\{t\leq\eta\}} + c_2\mathbb{1}_{\{\eta<t\}}$, $t \geq 0$. ∘

The following definition is in the spirit of Definition 3.2.7.

**Definition 3.3.4.** *The* ***present value*** *of the random payment stream $B$ is*

$$\begin{aligned} PV(t) = &\int_t^\infty v(t,s)\mathbb{1}_{\{T_x>s\}}b^a(s)\,\mathrm{d}s + \sum_{n=1}^\infty \mathbb{1}_{\{t<t_n\}}v(t,t_n)\mathbb{1}_{\{T_x>t_n\}}b_n^a \\ &+ \int_t^\infty v(t,s)\mathbb{1}_{\{T_x\leq s\}}b^d(s)\,\mathrm{d}s + \sum_{n=1}^\infty \mathbb{1}_{\{t<t_n\}}v(t,t_n)\mathbb{1}_{\{T_x\leq t_n\}}b_n^d \\ &+ \mathbb{1}_{\{t<T_x\}}v(t,T_x)b^{ad}(T_x) \end{aligned} \tag{3.3.3}$$

*for $t \geq 0$, whenever it is well-defined, that is finite.*

*Remark* 3.3.5. Due to Assumption 3.2.2, we may replace $v(t,s)$ by

$$\exp\left\{-\int_t^s r(u)\,\mathrm{d}u\right\}$$

for $t, s \geq 0$, and we shall do so whenever it is convenient. ▽

Does the fairness criterion of the previous section remain relevant for random payment streams? It took the form of

$$\mathrm{PV}(0-) := \mathrm{PV}(0) + b_0 = 0,$$

confer with (3.2.7). Actually, insurance contracts do not and should not satisfy this fairness criterion, since it would entail the contract to be equally attractive to the insured and the insurer regardless of the realization of $T_x$, that is regardless of when the insured dies. This is not in the spirit of insurance, which is about the transfer of risk, in this case mortality risk, from the insured to the insurer. Instead, insurance contracts should be settled in accordance with some premium principle, confer with Section 3.1, while also taking into account the time value of money, confer with Section 3.2. For various reasons, least not regulatory, the equivalence principle, which was introduced in Section 3.1, plays a fundamental role especially in life insurance. In the context of streams of payments and since $b_0$ is deterministic, the corresponding criterion reads

$$\mathbb{E}[\mathrm{PV}(0-)] = \mathbb{E}[\mathrm{PV}(0)] + b_0 = 0, \tag{3.3.4}$$

and the payment stream $B$ is said to satisfy the equivalence principle if (3.3.4) holds.

*Remark* 3.3.6. It is of course viable to use a different premium principle than the equivalence principle, also regarding the valuation of random payment streams in the context of life insurance. But in comparison to Section 3.1 some additional care has to be taken since PV(0) in our context typically does not consist solely of benefits but also of contributions. ▽

To cast (3.2.7) more explicitly, it is helpful to recall the shorthand notation

$$p_x^{\mathrm{aa}}(t,s) := \mathbb{P}(T_x > s \,|\, T_x > t) \quad \text{and} \quad p_x^{\mathrm{ad}}(t,s) := \mathbb{P}(T_x \leq s \,|\, T_x > t)$$

for $0 \leq t \leq s$, confer with (2.1.11). We are now ready to state and prove the main result of this section.

**Proposition 3.3.7.** *It holds that*

$$\begin{aligned}\mathbb{E}[PV(0)] = &\int_0^\infty v(0,s)p_x^{aa}(0,s)\big(b^a(s) + \mu(x+s)b^{ad}(s)\big)\,\mathrm{d}s + \sum_{n=1}^\infty v(0,t_n)p_x^{aa}(0,t_n)b_n^a \\ &+ \int_0^\infty v(0,s)p_x^{ad}(0,s)b^d(s)\,\mathrm{d}s + \sum_{n=1}^\infty v(0,t_n)p_x^{ad}(0,t_n)b_n^d\end{aligned} \tag{3.3.5}$$

*so long as* $\mathbb{E}[|PV(0)|] < \infty$.

*Remark* 3.3.8. Recall that $v(0,t) = v(t)$, $t \geq 0$, by definition of the latter. Thus we may write $v(t)$ in place of $v(0,t)$. Furthermore, due to Assumption 3.2.2, we may even write

$$\exp\Big\{-\int_0^t r(s)\,\mathrm{d}s\Big\}$$

in place of $v(t)$. Similarly, in accordance with (2.1.12), we may write

$$\exp\Big\{-\int_0^t \mu(x+s)\,\mathrm{d}s\Big\} \quad \text{and} \quad \Big(1 - \exp\Big\{-\int_0^t \mu(x+s)\,\mathrm{d}s\Big\}\Big)$$

in place of $p_x^{\mathrm{aa}}(0,t)$ and $p_x^{\mathrm{ad}}(0,t)$, respectively. Actuaries are keen on casting results in terms of the force of interest $r$ and the mortality rate $\mu$. There are multiple reasons for this preference, including the fact that the force of interest and mortality rate turn out to be fairly fundamental quantites from a statistical point of view. ▽

*Proof.* Note that since $\mathbb{P}(T_x > 0) = 1$, we have that

$$\mathbb{E}[\mathbb{1}_{\{T_x>s\}}] = p_x^{\mathrm{aa}}(0,s) \quad \text{and} \quad \mathbb{E}[\mathbb{1}_{\{T_x\leq s\}}] = p_x^{\mathrm{ad}}(0,s).$$

Starting from (3.3.3), taking expectations, and, broadly speaking, interchanging the order of integration and summation, we find that

$$\begin{aligned}\mathbb{E}[\mathrm{PV}(0)] = \int_0^\infty & v(0,s)p_x^{\mathrm{aa}}(0,s)b^{\mathrm{a}}(s)\,\mathrm{d}s + \sum_{n=1}^\infty v(0,t_n)p_x^{\mathrm{aa}}(0,t_n)b_n^{\mathrm{a}} \\ &+ \int_0^\infty v(0,s)p_x^{\mathrm{ad}}(0,s)b^{\mathrm{d}}(s)\,\mathrm{d}s + \sum_{n=1}^\infty v(0,t_n)p_x^{\mathrm{ad}}(0,t_n)b_n^{\mathrm{d}} \\ &+ \mathbb{E}[v(0,T_x)b^{\mathrm{ad}}(T_x)].\end{aligned}$$

To complete the proof, we thus only have to show that

$$\mathbb{E}[v(0,T_x)b^{\mathrm{ad}}(T_x)] = \int_0^\infty v(0,s)p_x^{\mathrm{aa}}(0,s)\mu(x+s)b^{\mathrm{ad}}(s)\,\mathrm{d}s.$$

Since $T_x$ admits the probability density function $f_x$, we find that

$$\mathbb{E}[v(0,T_x)b^{\mathrm{ad}}(T_x)] = \int_0^\infty v(0,s)b^{\mathrm{ad}}(s)f_x(s)\,\mathrm{d}s.$$

According to Lemma 2.1.17, it holds that

$$f_x(s) = \mu(x+s)p_x^{\mathrm{aa}}(0,s),$$

so that in turn

$$\mathbb{E}[v(0,T_x)b^{\mathrm{ad}}(T_x)] = \int_0^\infty v(0,s)\mu(x+s)p_x^{\mathrm{aa}}(0,s)b^{\mathrm{ad}}(s)\,\mathrm{d}s$$

as desired. □

Equation (3.3.5) may look quite menacing at first. The following examples are intended to illuminate the general result by recasting and simplifying it for some specific instances of payment streams.

**Example 3.3.9** (Pure endowment insurance, continued)**.** Suppose that

$$B(t) = b_0 + c\mathbb{1}_{\{\eta\leq t\}}\mathbb{1}_{\{T_x>\eta\}}, \qquad t \geq 0,$$

which basically corresponds to (3.3.1)–(3.3.2) with $t_1 = \eta$ and $b_1^{\mathrm{a}} = c$, see also Example 3.3.1.

According to Definition (3.3.4) and Proposition 3.3.7, we have that

$$\begin{aligned}\mathrm{PV}(t) &= c\mathbb{1}_{\{t<\eta\}}v(t,\eta)\mathbb{1}_{\{T_x>\eta\}}, \qquad t\geq 0,\\ \mathbb{E}[\mathrm{PV}(0)] &= c\,v(0,\eta)p_x^{\mathrm{aa}}(0,\eta)\\ &= c\exp\Big\{-\int_0^\eta (r(s)+\mu(x+s))\,\mathrm{d}s\Big\},\end{aligned}$$

where the last line follows from Assumption 3.2.2 and (2.1.12). Consequently, the equivalence principle is satisfied by $B$ if and only if

$$b_0 = -c \exp\left\{-\int_0^\eta (r(s) + \mu(x+s))\,\mathrm{d}s\right\},$$

confer with (3.3.4). ◦

**Example 3.3.10** (Term insurance, continued)**.** Suppose that

$$B(t) = b_0 + c\mathbb{1}_{\{T_x \leq t\}}, \qquad t \geq 0,$$

which basically corresponds to (3.3.1)–(3.3.2) with $b^{\mathrm{ad}}(t) = c$, $t \geq 0$, see also Example 3.3.2.

According to Definition (3.3.4) and Proposition 3.3.7, we have that

$$\begin{aligned}
\mathrm{PV}(t) &= c\mathbb{1}_{\{t < T_x\}} v(t, T_x), \qquad t \geq 0, \\
\mathbb{E}[\mathrm{PV}(0)] &= c \int_0^\infty v(0,s) p_x^{\mathrm{aa}}(0,s) \mu(x+s)\,\mathrm{d}s \\
&= c \int_0^\infty \exp\left\{-\int_0^s (r(u) + \mu(x+u))\,\mathrm{d}u\right\} \mu(x+s)\,\mathrm{d}s,
\end{aligned}$$

where the last line follows from Assumption 3.2.2 and (2.1.12). Consequently, the equivalence principle is satisfied by $B$ if and only if

$$b_0 = -c \int_0^\infty \exp\left\{-\int_0^s (r(u) + \mu(x+u))\,\mathrm{d}u\right\} \mu(x+s)\,\mathrm{d}s,$$

confer with (3.3.4). ◦

**Example 3.3.11** (Continuous life annuity, continued)**.** Suppose that

$$B(t) = b_0 + c \int_0^t \mathbb{1}_{\{T_x > s\}}\,\mathrm{d}s, \qquad t \geq 0,$$

which basically corresponds to (3.3.1)–(3.3.2) with $b^{\mathrm{a}}(t) = c$, $t \geq 0$, see also Example 3.3.3.

According to Definition (3.3.4) and Proposition 3.3.7, we have that

$$\begin{aligned}
\mathrm{PV}(t) &= c \int_t^\infty v(t,s) \mathbb{1}_{\{T_x > s\}}\,\mathrm{d}s, \qquad t \geq 0, \\
\mathbb{E}[\mathrm{PV}(0)] &= c \int_0^\infty v(0,s) p_x^{\mathrm{aa}}(0,s)\,\mathrm{d}s \\
&= c \int_0^\infty \exp\left\{-\int_0^s (r(u) + \mu(x+u))\,\mathrm{d}u\right\} \mathrm{d}s,
\end{aligned}$$

where the last line follows from Assumption 3.2.2 and (2.1.12). Consequently, the equivalence principle is satisfied by $B$ if and only if

$$b_0 = -c \int_0^\infty \exp\Big\{ - \int_0^s (r(u) + \mu(x+u))\,\mathrm{d}u \Big\}\,\mathrm{d}s,$$

confer with (3.3.4). ∘

In this section, we have only considered random payment streams on the form of (3.3.1)–(3.3.2). We should like to stress that the concept of present value as well as the equivalence principle carry over essentially ad verbatim to other types of random payment streams, confer also with Exercise XXIII.

In general, one has to resort to numerical methods to compute $\mathbb{E}[PV(0)]$ based on (3.3.5). If the discount factors $v(0,t)$, $t \geq 0$, are explicitly given, one may employ the following numerical forward method. First, compute $p_x^{\mathrm{aa}}(0,t)$, $t \geq 0$, by solving the differential equation

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t} p_x^{\mathrm{aa}}(0,t) = -\mu(x+t) p_x^{\mathrm{aa}}(0,t), \qquad t > 0, \\ &p_x^{\mathrm{aa}}(0,0) = 1, \end{aligned} \tag{3.3.6}$$

numerically and determine $p_x^{\mathrm{ad}}(0,t)$, $t \geq 0$, residually according to $p_x^{\mathrm{ad}} = 1 - p_x^{\mathrm{aa}}$. Next, compute the so-called discounted expected cash flows

$$\begin{aligned} &v(0,t)\left(p_x^{\mathrm{aa}}(0,t)\big(b^{\mathrm{a}}(t) + \mu(x+t) b^{\mathrm{ad}}(t)\big) + p_x^{\mathrm{ad}}(0,t) b^{\mathrm{d}}(t)\right), && t \geq 0, \\ &v(0,t_n)\left(p_x^{\mathrm{aa}}(0,t_n) b_n^{\mathrm{a}} + p_x^{\mathrm{ad}}(0,t_n) b_n^{\mathrm{d}}\right), && n \in \mathbb{N}, \end{aligned}$$

by addition and multiplication. Finally, compute $\mathbb{E}[\mathrm{PV}(0)]$ via numerical integration and summation of these discounted expected cash flows. Computational aspects are discussed in a bit more detail in Section 4.1, while a proof of (3.3.6) may be found in Exercise XXI.

We have throughout this and the previous section assumed the force of interest to be known, which is quite unreasonable. Fortunately, in conjunction actuarial and financial mathematics offer ideas, methods, and results that allow actuaries at least in theory to cope also with uncertain interest rates. In practice, insurers (must) use discount factors published by national and transnational supervisory authorities, who employ complex models in combination with vast amounts of market data to produce discounts factors that to a certain degree reflect the true, but unknown, time value of money.

## 3.4 Exercises

**Exercise XIV.** This exercise is devoted to the *variance principle*, which is given by

$$\pi(Y) = \mathbb{E}[Y] + \alpha \mathrm{Var}[Y], \qquad Y \in \mathcal{Y},$$

where $\alpha > 0$ is a loading factor, and where $\mathcal{Y}$ consists of all non-negative random variables with finite variance.

a) Argue that the variance principle induces non-negative loading and that it is translation invariant.

b) Following the arguments of Example 3.1.5, show that the variance principle is not monotonic.

   Hint: To produce a counterexample, study the case of $\mathbb{P}(Y = k) = 1 - \mathbb{P}(Y = 0) = \frac{1}{2}$ for $k \in \mathbb{N}$.

c) Show that

$$\pi(Y/2) + \pi(Y/2) < \pi(Y),$$

   and use this to conclude that the variance principle is not subadditive.

Actually, we may show that $n\pi(Y/n) < \pi(Y)$ for any $n \in \mathbb{N}$. Note that

$$n\pi(Y/n) = \mathbb{E}[Y] + \frac{\alpha}{n}\mathrm{Var}[Y],$$

which converges to the equivalence premium $\mathbb{E}[Y]$ as $n$ goes to infinity. Thus under the variance principle, one may essentially circumvent the loading factor by splitting up the contract into a multitude of smaller parts.

**Exercise XV.** Benefits $Y$ are split up into two parts $\beta Y$ and $(1 - \beta)Y$, where $\beta \in [0, 1]$. The first part is transfered to one insurer using the loading factor $\alpha_1 > 0$ for premium calculation, while the second part is transferred to another insurer using the loading factor $\alpha_2 > 0$ for premium calculation. Suppose $\alpha_2 > \alpha_1$. Which choice of $\beta$ leads to the lowest combined premium in case of (a) the expected value principle, (b) the standard deviation principle, and (c) the variance principle? For the definition of the latter premium principle, see Exercise XIV.

**Exercise XVI.** We claimed in Example 3.2.1 that there exists a close link between constant forces of interest and constant annual interest rates, in the sense that the former provides some sort of generalization of the latter to allow for continuous accumulation of interest. The purpose of this exercise is to verify that claim.

Let $i > -1$ be some constant annual interest rate, so that

$$v(n, 0) = (1 + i)^n \qquad \text{and} \qquad v(0, n) = (1 + i)^{-n}$$

for $n \in \mathbb{N}_0$. We are going to investigate what happens if we let interest accumulate more frequently than annually. To this end, suppose that interest is also allowed to accumulate $m$ times per year at rate $i^{(m)}/m$ for some $m \in \mathbb{N}$.

a) Argue that the principle of no arbitrage entails that

$$1 + i = \left(1 + \frac{i^{(m)}}{m}\right)^m,$$

so that $i^{(m)} = m\big((1+i)^{1/m} - 1\big)$.

b) Show that

$$\lim_{m\to\infty} i^{(m)} = r, \tag{3.4.1}$$

where $r = \log\{1+i\} \in \mathbb{R}$.

If $t \geq 0$ is rational, then we know that there exists $n \in \mathbb{N}_0$, $m \in \mathbb{N}$, and $k \in \mathbb{N}$ with $k \leq m$ such that $t = n + k/m$. Then, if interest is allowed to accumulate $m$ times per year at rate $i^{(m)}/m$, we may write

$$v(t,0) = \left(1 + \frac{i^{(m)}}{m}\right)^{m(n+k/m)} = (1+i)^n \left(1 + \frac{i^{(m)}}{m}\right)^k, \tag{3.4.2}$$

$$v(0,t) = \left(1 + \frac{i^{(m)}}{m}\right)^{-m(n+k/m)} = (1+i)^{-n} \left(1 + \frac{i^{(m)}}{m}\right)^{-k}. \tag{3.4.3}$$

If instead $t \geq 0$ is irrational, then we may approximate it by a sequence of rational numbers. In general, for any $t \geq 0$ there exists an $n \in \mathbb{N}_0$ and a sequence $(k(m))_{m\in\mathbb{N}}$ with $k(m) \leq m$ so that $t = \lim_{m\to\infty} n + k(m)/m$. Consequently, if interest is allowed to accumulate more and more quickly, we should be able to write

$$v(t,0) = \lim_{m\to\infty} \left(1 + \frac{i^{(m)}}{m}\right)^{m(n+k(m)/m)}, \tag{3.4.4}$$

$$v(0,t) = \lim_{m\to\infty} \left(1 + \frac{i^{(m)}}{m}\right)^{-m(n+k(m)/m)} \tag{3.4.5}$$

in correspondence with (3.4.2)–(3.4.3). Now, due to (3.4.1), since $t = \lim_{m\to\infty} n + k(m)/m$, and since one may show that

$$\lim_{x\to\infty} (1 + y/x)^x = \exp\{y\}, \qquad y \in \mathbb{R},$$

we find by evaluating the limits in (3.4.4)–(3.4.5) that

$$v(t,0) = \exp\{r\}^t = \exp\{rt\},$$
$$v(0,t) = \exp\{r\}^{-t} = \exp\{-rt\}.$$

From this we conclude that the discount function

$$v(t) = \exp\{-rt\}, \qquad t \geq 0,$$

of Example 3.2.1 generalizes the constant annual interest rate $i = \exp\{r\} - 1$ to allow for accumulation of interest in continuous time and that in a manner which is consistent with the principle of no arbitrage.

**Exercise XVII.** Assume that the force of interest $r : [0, \infty) \to \mathbb{R}$ is constant and denote its value by $r$. Set $i = \exp\{r\} - 1 > -1$. Let $b_0 = c$, $t_n = n$, and $b_n = c$ for $n = 1, \dots, k-1$, where $k \in \mathbb{N}$, and zero otherwise, so that

$$B(t) = c + \sum_{n=1}^{k-1} \mathbb{1}_{\{n \leq t\}} c, \qquad t \geq 0.$$

a) Show that if $i \neq 0$, that is $r \neq 0$, then

$$\mathrm{PV}(0-) = c(1+i)\frac{1 - \frac{1}{(1+i)^k}}{i}$$

b) Let $i > 0$. Conclude that the quantity from a) converges to $c\frac{1+i}{i}$ as $k$ goes to infinity. How does this compare to the results of Example 3.2.11?

Hint: It may be useful to recall that for $x > -1$ it holds that

$$\frac{x}{1+x} \leq \log\{1+x\} \leq x.$$

**Exercise XVIII.** Assume that the force of interest $r : [0, \infty) \to \mathbb{R}$ is constant and denote its value by $r$. Suppose $r \neq 0$. Set $i = \exp\{r\} - 1$. Let

$$b(t) = c_2 \mathbb{1}_{\{s_1 < t \leq s_1 + s_2\}} - c_1 \mathbb{1}_{\{t \leq s_1\}}, \qquad t \geq 0,$$

so that

$$B(t) = \int_{s_1}^{\min\{\max\{t, s_1\}, s_1 + s_2\}} c_2 \,\mathrm{d}s - \int_0^{\min\{t, s_1\}} c_1 \,\mathrm{d}s, \qquad t \geq 0.$$

a) Show that

$$\mathrm{PV}(0-) = c_2 \frac{\exp\{-rs_1\} - \exp\{-r(s_1 + s_2)\}}{r} - c_1 \frac{1 - \exp\{-rs_1\}}{r}.$$

b) Determine $c_2$ in terms of $c_1$ such that the payment stream is fair.

c) Let $s_1 = 30$, $s_2 = 20$, $c_1 = 1$, and $r = \log\{1 + 0.02\}$. Compute $c_2$ such that the payment stream is fair.

**Exercise XIX.** You are the owner of a bond with a face value of 1000 monetary units, which pays 20 monetary units every half year for ten years. Concurrent with the last payment you also receive its face value. The first payment is to occur three months from now.

a) Provide an expression for the payment stream generated by the bond.

b) Suppose the annual interest rate is constant and equals 3%, that is $i = 0.03$. What is the bond worth today? What would it be worth three months from now?

**Exercise XX.** You borrow 1000 monetary units. You begin repaying the loan one year from now, and you intend to have fully repaid it 30 years from now. The first ten years you repay $\frac{1}{3}a$ annually, the next ten years you repay $\frac{2}{3}a$ annually, and the last ten years you repay $a$ annually. We suppose that $a$ is determined such that the loan is exactly repaid after your 30'th payment.

a) Draw up an expression for the payment stream corresponding to the repayment scheme.

b) Suppose the annual interest rate is constant and equals 4%, that is $i = 0.04$. What is $a$? Do you owe more or less than 1000 monetary units after your tenth repayment?

**Exercise XXI.** Let $x \geq 0$ and $t \geq 0$ be given. Prove that $p_x^{\mathrm{aa}}(t,s)$, $s \geq t$, given by (2.1.11) solves the differential equation

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}s} p_x^{\mathrm{aa}}(t,s) = -\mu(x+s) p_x^{\mathrm{aa}}(t,s), \qquad s > t, \\ &p_x^{\mathrm{aa}}(t,t) = 1. \end{aligned}$$

Compare this result with (3.3.6).

**Exercise XXII.** Consider the random payment stream $B$ described at the end of Example 3.3.3, that is suppose

$$B(t) = \int_0^t \mathbb{1}_{\{T_x > s\}} \left( \mathbb{1}_{\{s \leq \eta\}} c_1 + \mathbb{1}_{\{\eta < s\}} c_2 \right) \mathrm{d}s, \qquad t \geq 0,$$

which corresponds to (3.3.1)–(3.3.2) with $b^{\mathrm{a}}(t) = c_1 \mathbb{1}_{\{t \leq \eta\}} + c_2 \mathbb{1}_{\{\eta < t\}}$, $t \geq 0$.

a) Explain in your own words the primary characteristics of this random payment stream. What kind of life insurance contract does it correspond to? Why may an insured be interested in this type of contract?

b) Argue that

$$\mathrm{PV}(t) = \int_t^\infty v(t,s) \mathbb{1}_{\{T_x > s\}} \left( \mathbb{1}_{\{s \leq \eta\}} c_1 + \mathbb{1}_{\{\eta < s\}} c_2 \right) \mathrm{d}s, \qquad t \geq 0,$$

and show that

$$\begin{aligned} \mathbb{E}[\mathrm{PV}(0)] = {} & c_1 \int_0^\eta v(0,s) p_x^{\mathrm{aa}}(0,s) \, \mathrm{d}s \\ & + c_2 v(0,\eta) p_x^{\mathrm{aa}}(0,\eta) \int_0^\infty v(\eta, \eta+s) p_{x+\eta}^{\mathrm{aa}}(0,s) \, \mathrm{d}s. \end{aligned}$$

Hint: Recall that $v(0,s) = v(0,t)v(t,s)$ for $t, s \geq 0$. It might also be helpful to consult Lemma 2.1.16.

Let $B_1$, $B_2$, and $B_3$ be given by

$$B_1(t) = \int_0^t \mathbb{1}_{\{T_x > s\}} \mathbb{1}_{\{s \le \eta\}} \mathrm{d}s, \qquad t \ge 0,$$

$$B_2(t) = \int_0^t \mathbb{1}_{\{T_{x+\eta} > s\}} \mathrm{d}s, \qquad t \ge 0,$$

$$B_3(t) = \int_0^t \mathbb{1}_{\{T_x > s\}} \mathrm{d}s, \qquad t \ge 0,$$

and denote the corresponding present values by $\mathrm{PV}_1$, $\mathrm{PV}_2$, and $\mathrm{PV}_3$, respectively. Consider the identities

$$\mathbb{E}[\mathrm{PV}(0)] \overset{\star}{=} c_1 \mathbb{E}[\mathrm{PV}_1(0)] + c_2 v(0,\eta) p_x^{\mathrm{aa}}(0,\eta) \mathbb{E}[\mathrm{PV}_2(0)],$$

$$\mathbb{E}[\mathrm{PV}(0)] \overset{\star\star}{=} c_1 \mathbb{E}[\mathrm{PV}_1(0)] + c_2 v(0,\eta) p_x^{\mathrm{aa}}(0,\eta) \mathbb{E}[\mathrm{PV}_3(0)].$$

c) What do you have to assume for $(\star)$ to be true? What do you additionally have to assume for $(\star\star)$ to be true?

d) Determine $c_1$ in terms of $c_2$ such that the random payment stream $B$ satisfies the equivalence principle.

**Exercise XXIII.** Let $T_x^{♂}$ and $T_y^{♀}$ be two independent residual lifetimes with mortality rates $\mu^{♂}$ and $\mu^{♀}$, respectively. Consider the random payment streams $B_1$ and $B_2$ given by

$$B_1(t) = c_1 \mathbb{1}_{\{\eta \le t\}} \mathbb{1}_{\{T_y^{♀} > \eta\}} \mathbb{1}_{\{T_x^{♂} \le \eta\}}, \qquad t \ge 0,$$

$$B_2(t) = c_2 \int_0^t \mathbb{1}_{\{T_y^{♀} > s\}} \mathbb{1}_{\{T_x^{♂} > s\}} \mathbb{1}_{\{s \le \eta\}} \, \mathrm{d}s, \qquad t \ge 0,$$

and denote their corresponding present values by $\mathrm{PV}_1$ and $\mathrm{PV}_2$, respectively.

a) Explain in your own words the primary characteristics of these random payment streams. What kind of life insurance contracts do they correspond to? Why may a policyholder or an insured be interested in these types of contracts?

b) Argue that

$$\mathrm{PV}_1(t) = c_1 \mathbb{1}_{\{t < \eta\}} v(t,\eta) \mathbb{1}_{\{T_y^{♀} > \eta\}} \mathbb{1}_{\{T_x^{♂} \le \eta\}}, \qquad t \ge 0,$$

$$\mathrm{PV}_2(t) = c_2 \mathbb{1}_{\{t < \eta\}} \int_t^\eta v(t,s) \mathbb{1}_{\{T_y^{♀} > s\}} \mathbb{1}_{\{T_x^{♂} > s\}} \, \mathrm{d}s, \qquad t \ge 0.$$

c) Show that

$$\mathbb{E}[\mathrm{PV}_1(0)] = c_1 \exp\Big\{ -\int_0^\eta \Big( r(s) + \mu^{♀}(y+s) \Big) \, \mathrm{d}s \Big\} \Big( 1 - \exp\Big\{ -\int_0^\eta \mu^{♂}(x+s) \, \mathrm{d}s \Big\} \Big),$$

$$\mathbb{E}[\mathrm{PV}_2(0)] = c_2 \int_0^\eta \exp\Big\{ -\int_0^s \Big( r(u) + \mu^{♀}(y+u) + \mu^{♂}(x+u) \Big) \, \mathrm{d}u \Big\} \, \mathrm{d}s.$$

d) Determine $c_2$ in terms of $c_1$ such that the random payment stream $B := B_1 + B_2$ satisfies the equivalence principle.

# Chapter 4

# Selection of advanced topics

In this chapter, we discuss a selection of slightly more advanced topics in actuarial mathematics – building on the concepts, methods, and results we have already developed in the previous chapters. Each section deals with a single topic, and the sections can be read independently from one another.

In Section 4.1, we focus on some more specific random payment streams and derive a differential equation for the corresponding so-called prospective reserve, which, among other things, provides an alternative computational scheme to the one developed in Section 3.3. This differential equation, known as *Thiele's differential equation*, dates back to 1875 and is named afters its discoverer or inventor, the Dane T.N. Thiele[4]. Today, after decades of actuarial research, Thiele's name is reserved not only for his own contribution but attributed to a wide range of results that generalize Thiele's original differential equation in various directions. These results have played and continue to play a fundamental role in the mathematics of life insurance. In Section 4.1, though, we restrict the study to Thiele's original contribution only.

While Section 4.1 relates to life insurance and takes as its starting point Section 2.1 and Section 3.2–3.3, Section 4.2 focuses on aspects closer to non-life insurance and builds primarily on Section 2.2–2.3 and Section 3.1. The starting point of Section 4.2 is the observation that to to apply sophisticated premium principles, one typically has to be able to compute the distribution function of the benefits of interest. To this end, we introduce and study the *collective risk model*, where the benefits are modeled by a compound distribution, namely the aggregate claim distribution. It turns out that one can only get so far via an analytical approach, so we conclude the section by discussing the so-called normal power approximation.

[4]Thorvald Nicolai Thiele (1838–1910) was a Danish astronomer, mathematician, statistician, and actuary who is known for his contributions to mathematical statistics and actuarial science as well as his unique abilities as an initiator. The Danish Mathematical Society and the Danish Society of Actuaries were both founded on his initiative in 1873 and 1901, respectively. Furthermore, he was the founder and a member of management of the first Danish life insurance company Hafnia from 1872 until his death.

## 4.1 Thiele's differential equation

We adopt the setting and assumptions of Section 3.3, but specialize to a payment stream $B$ of the form

$$B(t) = b_0 + \int_0^t \mathbb{1}_{\{T_x > s\}} b^{\mathrm{a}}(s)\,\mathrm{d}s + \mathbb{1}_{\{T_x \leq t\}} b^{\mathrm{ad}}(T_x), \qquad t \geq 0, \tag{4.1.1}$$

with corresponding present value PV given by

$$\mathrm{PV}(t) = \int_t^\infty v(t,s) \mathbb{1}_{\{T_x > s\}} b^{\mathrm{a}}(s)\,\mathrm{d}s + \mathbb{1}_{\{t < T_x\}} v(t, T_x) b^{\mathrm{ad}}(T_x), \qquad t \geq 0, \tag{4.1.2}$$

which we assume to be well-defined, that is finite, confer with Definition 3.3.4. This payment stream consists of the initial payment $b_0$, the payment rate $b^{\mathrm{a}}$ as long as the insured is alive, and the payment $b^{\mathrm{ad}}$ upon death of the insured.

In Section 3.3, we took interest in the expected present value at time zero, that is $\mathbb{E}[\mathrm{PV}(0)]$, and we developed the criterion

$$\mathbb{E}[\mathrm{PV}(0-)] = \mathbb{E}[\mathrm{PV}(0)] + b_0 = 0.$$

This criterion allows one to balance the level of benefits and the level of contributions (premiums) such that the life insurance contract is (in a certain sense) fair at initialization, confer with Example 3.3.9–3.3.11 and Exercise XXII–XXIII. Consequently, this criterion is important in the context of pricing. But it is also important to the insurer to be able to determine, as time passes, accounting entries that properly reflect the liabilities associated with the insurance contract. In theory as well as in practice, the conditional expectation of the present value at time $t$ given the information available to the insurer at time $t$ plays a prominent role, confer also with the discussion between (3.1.3) and Definition 3.1.1. This is the starting point for the following definition.

**Definition 4.1.1.** *The* ***prospective reserve*** *of the random payment stream $B$ is*

$$V^a(t) = \mathbb{E}[PV(t) \mid T_x > t], \qquad t \geq 0,$$

*whenever it is well-defined, that is whenever $\mathbb{E}[|PV(t)|] < \infty$, $t \geq 0$.*

*Remark* 4.1.2 (State-wise prospective reserves). We could also have introduced the quantity

$$V^{\mathrm{d}}(t) := \mathbb{E}[\mathrm{PV}(t) \mid T_x \leq t], \qquad t \geq 0,$$

but since (4.1.1) entails that there are no payments when the insured is dead, that is when $T_x \leq t$, it holds that $V^{\mathrm{d}}(t) = 0$ for $t > 0$ (while $V^{\mathrm{d}}(0)$ is actually not well-defined, since $\mathbb{P}(T_x \leq 0) = 0$).

For more general random payment streams, say payment streams of the form (3.3.1)–(3.3.2), where $V^{\mathrm{d}}$ is non-trivial, one would usually save the nomenclature prospective reserve for the collection of random variables $V(t)$, $t \geq 0$, given by

$$V(t) = \mathbb{E}[\mathrm{PV}(t) \,|\, \mathbb{1}_{\{T_x > t\}}], \qquad t \geq 0,$$

and the quantities $V^{\mathrm{a}}(t)$ and $V^{\mathrm{d}}(t)$, $t \geq 0$, would be called *state-wise prospective reserves.* ▽

*Remark* 4.1.3 (Equivalence principle). Note that since $\mathbb{P}(T_x > 0) = 1$, it holds that $V^{\mathrm{a}}(0) = \mathbb{E}[\mathrm{PV}(0)]$. Consequently, we may recast the equivalence principle of (3.3.4) as $V^{\mathrm{a}}(0) + b_0 = 0$. ▽

The following result is in line with Proposition 3.3.7.

**Proposition 4.1.4.** *It holds that*

$$\begin{aligned} V^{a}(t) &= \int_t^\infty v(t,s) p_x^{aa}(t,s) \big(b^a(s) + \mu(x+s) b^{ad}(s)\big) \,\mathrm{d}s \\ &= \int_t^\infty \exp\Big\{ -\int_t^s \big(r(u) + \mu(x+u)\big) \,\mathrm{d}u \Big\} \big(b^a(s) + \mu(x+s) b^{ad}(s)\big) \,\mathrm{d}s \end{aligned} \tag{4.1.3}$$

*for* $t \geq 0$.

*Proof.* The proof is similar to the proof of Proposition 3.3.7. Note that

$$\mathbb{E}[\mathbb{1}_{\{T_x > s\}} \,|\, T_x > t] = p_x^{\mathrm{aa}}(t,s), \qquad 0 \leq t \leq s.$$

Starting from (4.1.2), taking conditional expectations, and, broadly speaking, interchanging the order of integration, we find that

$$V^{\mathrm{a}}(t) = \int_t^\infty v(t,s) p_x^{\mathrm{aa}}(t,s) b^{\mathrm{a}}(s) \,\mathrm{d}s + \mathbb{E}[v(t,T_x) b^{\mathrm{ad}}(t_x) \,|\, T_x > t].$$

Thus to prove the first line of (4.1.3) it suffices to show that

$$\mathbb{E}[v(t,T_x) b^{\mathrm{ad}}(T_x) \,|\, T_x > t] = \int_t^\infty v(t,s) p_x^{\mathrm{aa}}(t,s) \mu(x+s) b^{\mathrm{ad}}(s) \,\mathrm{d}s.$$

According to Lemma 2.1.16, the conditional distribution of $T_x$ given the event $(T_x > t)$ corresponds to the distribution of $T_{x+t} + t$, which yields

$$\begin{aligned} \mathbb{E}[v(t,T_x) b^{\mathrm{ad}}(T_x) \,|\, T_x > t] &= \mathbb{E}[v(t, T_{x+t} + t) b^{\mathrm{ad}}(T_{x+t} + t)] \\ &= \int_0^\infty v(t, u+t) b^{\mathrm{ad}}(u+t) f_{x+t}(u) \,\mathrm{d}u \\ &= \int_t^\infty v(t,s) b^{\mathrm{ad}}(s) f_{x+t}(s-t) \,\mathrm{d}s, \end{aligned}$$

where the last line follows from the substitution $s = u + t$. Furthermore, from Lemma 2.1.17 we obtain that

$$f_{x+t}(s-t) = \mu(x+s) p_x^{\mathrm{aa}}(t,s).$$

We conclude that

$$\mathbb{E}[v(t,T_x) b^{\mathrm{ad}}(T_x) \,|\, T_x > t] = \int_t^\infty v(t,s) b^{\mathrm{ad}}(s) \mu(x+s) p_x^{\mathrm{aa}}(t,s)\,\mathrm{d}s,$$

which completes the proof of the first line of (4.1.3). The second line follows from Assumption 3.2.2 and (2.1.12). □

In general, one must resort to numerical methods to compute $V^{\mathrm{a}}(t)$ based on (4.1.3). If the discount factors $v(t,s)$, $s \geq t$, are explicitly given, one may employ the following numerical forward method. First, compute $p_x^{\mathrm{aa}}(t,s)$, $s \geq t$, by solving the differential equation

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}s} p_x^{\mathrm{aa}}(t,s) = -\mu(x+s) p_x^{\mathrm{aa}}(t,s), \qquad s > t, \\ &p_x^{\mathrm{aa}}(t,t) = 1, \end{aligned}$$

numerically, confer with Exercise XXI. Next, compute the so-called discounted expected cash flow

$$v(t,s) p_x^{\mathrm{aa}}(t,s) \left(b^{\mathrm{a}}(s) + \mu(x+s) b^{\mathrm{ad}}(s)\right), \qquad s \geq t,$$

by addition and multiplication. Finally, compute $V^{\mathrm{a}}(t)$ via numerical integration of the discounted expected cash flow.

Obviously, this numerical method has a finite runtime only if computations are halted at some some finite cutoff point $\eta \in (t,\infty)$. In this case, instead of computing the indefinite integral $V^{\mathrm{a}}(t)$, one computes the definite integral

$$\int_t^\eta v(t,s) p_x^{\mathrm{aa}}(t,s) \big(b^{\mathrm{a}}(s) + \mu(x+s) b^{\mathrm{ad}}(s)\big)\,\mathrm{d}s,$$

which causes a theoretical error of

$$V^{\mathrm{a}}(t) - \int_t^\eta v(t,s) p_x^{\mathrm{aa}}(t,s) \big(b^{\mathrm{a}}(s) + \mu(x+s) b^{\mathrm{ad}}(s)\big)\,\mathrm{d}s = v(t,\eta) p_x^{\mathrm{aa}}(t,\eta) V^{\mathrm{a}}(\eta),$$

see also Exercise XXV for further details.

The above forward method is particularly useful if one wants to compute $V^{\mathrm{a}}(t)$ for some fixed $t \geq 0$ and various discount factors $v(t,s)$, $s \geq t$, given a priori. If one wants instead to compute $V^{\mathrm{a}}(t)$ for all $t \geq 0$ and some fixed force of interest $r$, then the forward method is not particularly useful, since one would have to run it once for every $t \geq 0$. But do not despair: This is the hour when Thiele's differential equation enters the stage.

**Theorem 4.1.5** (Thiele's differential equation)**.** *Assume that* $b^a : [0,\infty) \to \mathbb{R}$ *and* $b^{ad} : [0,\infty) \to \mathbb{R}$ *are continuous. It then holds that*

$$\frac{\mathrm{d}}{\mathrm{d}t} V^a(t) = r(t)V^a(t) - b^a(t) - \mu(x+t)\big(b^{ad}(t) - V^a(t)\big), \qquad t > 0. \tag{4.1.4}$$

*Proof.* Introduce the auxiliary function $W^{\mathrm{a}}$ according to

$$W^{\mathrm{a}}(t) = \exp\Big\{-\int_0^t (r(u) + \mu(x+u))\,\mathrm{d}u\Big\} V^{\mathrm{a}}(t), \qquad t \geq 0. \tag{4.1.5}$$

Due to (4.1.3), it holds that

$$W^{\mathrm{a}}(t) = \int_t^\infty \exp\Big\{-\int_0^s (r(u) + \mu(x+u))\,\mathrm{d}u\Big\}\big(b^{\mathrm{a}}(s) + \mu(x+s)b^{\mathrm{ad}}(s)\big)\,\mathrm{d}s.$$

Since $b^{\mathrm{a}}$ and $b^{\mathrm{ad}}$ are assumed continuous, we may apply the First Fundamental Theorem of Calculus to obtain that

$$\frac{\mathrm{d}}{\mathrm{d}t} W^{\mathrm{a}}(t) = -\exp\Big\{-\int_0^t (r(u) + \mu(x+u))\,\mathrm{d}u\Big\}\big(b^{\mathrm{a}}(t) + \mu(x+t)b^{\mathrm{ad}}(t)\big) \tag{4.1.6}$$

for $t > 0$. On the other hand, by differentiation based on (4.1.5), we find that

$$\frac{\mathrm{d}}{\mathrm{d}t} W^{\mathrm{a}}(t) = \exp\Big\{-\int_0^t (r(u) + \mu(x+u))\,\mathrm{d}u\Big\}\Big(-(r(t) + \mu(x+t))V^{\mathrm{a}}(t) + \frac{\mathrm{d}}{\mathrm{d}t} V^{\mathrm{a}}(t)\Big)$$

for $t > 0$, so that

$$\exp\Big\{\int_0^t (r(u) + \mu(x+u))\,\mathrm{d}u\Big\}\frac{\mathrm{d}}{\mathrm{d}t} W^{\mathrm{a}}(t) = -(r(t) + \mu(x+t))V^{\mathrm{a}}(t) + \frac{\mathrm{d}}{\mathrm{d}t} V^{\mathrm{a}}(t).$$

From (4.1.6) it then follows that

$$-\big(b^{\mathrm{a}}(t) + \mu(x+t)b^{\mathrm{ad}}(t)\big) = -(r(t) + \mu(x+t))V^{\mathrm{a}}(t) + \frac{\mathrm{d}}{\mathrm{d}t} V^{\mathrm{a}}(t), \qquad t > 0.$$

Rearranging terms completes the proof. □

*Remark* 4.1.6 (Discontinuities). Thiele's differential equation is often presented with the continuity conditions on $b^{\mathrm{a}}$ and $b^{\mathrm{ad}}$ hidden. But be careful: as usual, the devil lies in the details. If $b^{\mathrm{a}}$ and $b^{\mathrm{ad}}$ are discontinuous in the point $\tau > 0$, then (4.1.4) still holds but only on $(0,\tau) \cup (\tau,\infty)$ rather than $(0,\infty)$, confer with the application of the First Fundamental Theorem of Calculus in the proof of Theorem 4.1.5. This could appear to be an insignificant difference, and that is indeed the case from a purely theoretical-mathematical point of view. On the other hand, it turns out to actually be of some importance concerning the development and the subsequent implementation of numerical methods.

To gain a better understanding of Thiele's differential equation, it is helpful to think of the insured as being part of an infinitely large insurance portfolio of identical insured with identical insurance contracts, so that the prospective reserve may be interpreted as the the fund per surviving insured, and to slightly rearrange (4.1.4), so that

$$\frac{\mathrm{d}}{\mathrm{d}t}V^{\mathrm{a}}(t) = \big(-b^{\mathrm{a}}(t) - \mu(x+t)b^{\mathrm{ad}}(t)\big) + r(t)V^{\mathrm{a}}(t) + \mu(x+t)V^{\mathrm{a}}(t), \qquad t > 0. \quad (4.1.7)$$

Recall the convention that payments to the insured from the insurer (benefits) are positive while payments from the insured to the insurer (contributions/premiums) are negative, and suppose for simplicity that $b^{\mathrm{ad}} > 0$ and $b^{\mathrm{a}} < 0$, so that the insured pays premiums while alive and receives a benefit upon death. Since $\mu(x+t)$ is the infinitesimal probability that the insured dies at time $t$ conditional on the insured being alive until time $t$, we may interpret $\mu(x+t)b^{\mathrm{ad}}(t)$ as the expected infinitesimal benefit at time $t$. In similar fashion, $\mu(x+t)V^{\mathrm{a}}(t)$ may be thought of as the expected inheritance from those who die at time $t$. According to (4.1.7), the fund per surviving insured at time $t$ then develops per unit of time as follows. It is increased by the excess of premiums over expected benefits, $\big(-b^{\mathrm{a}}(t) - \mu(x+t)b^{\mathrm{ad}}(t)\big)$, by the interest earned, $r(t)V^{\mathrm{a}}(t)$, and by the fund inherited from those who die, $\mu(x+t)V^{\mathrm{a}}(t)$.

While it is outside the scope of this presentation to discuss in depths the significance of Thiele's differential equation in the mathematics of life insurance, ranging from contract design to risk management, it should be mentioned that (4.1.4) already by itself provides a key decomposition of the change in prospective reserve in terms of the underlying risks, namely interest (first term) and mortality (third term).

We conclude this section by returning to the discussion concerning the computation of prospective reserves, which was also what originally led us to introduce Thiele's differential equation.

If $B$ satisfies the equivalence principle, then $V^{\mathrm{a}}(0) = -b_0$, which would provide an initial condition for Thiele's differential equation, so that one may develop a numerical forward method. If $B$ does not satisfy the equivalence principle, or if one wants to determine $b_0$ so that the equivalence principle is satisfied, Thiele's differential equation may at first seem to not be useful, since it in general lacks a boundary condition. But this is actually not the case, as we now show.

Suppose that there exists a maximal contract time $\tau > 0$ such that $b^{\mathrm{a}}(t) = 0$ and $b^{\mathrm{ad}}(t) = 0$ for $t \geq \tau$. In this case, $V^{\mathrm{a}}(t) = 0$ for $t \geq \tau$, and to compute $V^{\mathrm{a}}(t)$ for $0 \leq t \leq \tau$, one may simply solve Thiele's differential equation numerically, starting from the terminal condition $V^{\mathrm{a}}(\tau) = 0$. It is important to note that this numerical backward method yields not only $V(t)$ for some fixed $t \geq 0$, as was the case for the numerical forward method introduced between Proposition 4.1.4 and Theorem 4.1.5, but actually $V(t)$ for all $t \geq 0$.

Even if a maximal contract time $\tau > 0$ does not necessarily exist, one may still develop a numerical backward method based on Thiele's differential equation. To this end, introduce the auxiliary quantities $V^{\mathrm{a}}(t,\eta)$, $0 \leq t \leq \eta$, given by

$$V^{\mathrm{a}}(t,\eta) = \int_t^\eta \exp\Big\{-\int_t^s (r(u) + \mu(x+u))\,\mathrm{d}u\Big\}\big(b^{\mathrm{a}}(s) + \mu(x+s)b^{\mathrm{ad}}(s)\big)\,\mathrm{d}s.$$

Note that if a maximal contract time $\tau > 0$ exists, then $V^{\mathrm{a}}(t,\tau) = V^{\mathrm{a}}(t)$ for $0 \leq t \leq \tau$.

Rather than computing the indefinite integrals $V^{\mathrm{a}}(t)$, $t \geq 0$, one could compute the definite integrals $V^{\mathrm{a}}(t,\eta)$, $0 \leq t \leq \eta$, for some fixed $\eta \gg 0$; here $x \gg y$ means that $x$ is much greater than $y$. This idea is identical to the cutoff method for the numerical forward method discussed between Proposition 4.1.4 and Theorem 4.1.5. It results in the theoretical errors

$$V^{\mathrm{a}}(t) - V^{\mathrm{a}}(t,\eta) = \exp\Big\{-\int_t^\eta (r(u) + \mu(x+u))\,\mathrm{d}u\Big\}V^{\mathrm{a}}(\eta), \qquad 0 \leq t \leq \eta, \quad (4.1.8)$$

see also Exercise XXV for further details.

To actually compute $V^{\mathrm{a}}(t,\eta)$, $0 \leq t \leq \eta$, we note that

$$\frac{\mathrm{d}}{\mathrm{d}t}V^{\mathrm{a}}(t,\eta) = r(t)V^{\mathrm{a}}(t,\eta) - b^{\mathrm{a}}(t) - \mu(x+t)\big(b^{\mathrm{ad}}(t) - V^{\mathrm{a}}(t,\eta)\big), \quad (4.1.9)$$

for $0 < t < \eta$, which we recognize as Thiele's differential equation, but restricted to $(0,\eta)$ and with $V^{\mathrm{a}}(t,\eta)$ in place of $V^{\mathrm{a}}(t)$. To derive (4.1.9), differentiate based on (4.1.8), apply Thiele's differential equation, and rearrange terms, confer also with Exercise XXV. Since $V^{\mathrm{a}}(\eta,\eta) = 0$ by definition, it follows that one may compute $V^{\mathrm{a}}(t,\eta)$, $0 \leq t \leq \eta$, simply by solving (4.1.9) numerically, starting from the terminal condition $V^{\mathrm{a}}(\eta,\eta) = 0$.

## 4.2 Collective risk model

In this section, we study a benefit $Y$ on the form

$$Y = \sum_{k=1}^{N} X_k \quad (4.2.1)$$

for independent and identically distributed claims sizes $X_1, X_2, \ldots$ and a claim count $N$ assumed to be independent of the claim sizes. This is known as the *collective risk model*, and $Y$ is said to follow a compound distribution, namely the aggregate claims distribution, so named since $Y$ is exactly that: the aggregated claims.

Since we may have zero claims, corresponding to the event $(N = 0)$, equation (4.2.1) relies on the convention that $\sum_{k=1}^{0} a_k = 0$, which is used throughout this section.

The distribution function of $X_1$ is denoted $G$, while the probability mass function of $N$ is denoted $p$. Since $X_1, X_2, \ldots$ are identically distributed, the $k$'th variable $X_k$

for any $k \in \mathbb{N}$ also has distribution function $G$. For more details concerning claim size and claim count distributions we refer to Section 2.2–2.3. The claim count $N$ may, depending on the application, refer both to the number of claims associated with a single insurance contract within some contractual period or the total number of claims of some insurance portfolio within a fixed time period, say one year. In the context of pricing, the former view is prevalent, while the latter view is prevalent in the context of reserving.

The assumption that $X_1, X_2, \ldots$ be independent means that we do not learn anything about the size of the $k$'th claim by knowing the claim sizes $X_j$, $j \neq k$. Similarly, the assumption that $N$ be independent of $X_1, X_2, \ldots$ means that we do not learn anything about the severity of claims just from knowing something about the number of claims, or, conversely, that we do not learn anything about the number of claims just from knowing something about their severity. The collective risk model is convenient, since the independence assumption between claim sizes and claim numbers allows one to take a marginal approach to the modeling and estimation of severity and frequency risk.

Before we discuss the model in more detail, we first discuss how it interacts with thinning. To this end suppose that not every claim is reported or of interest, and let $I_k$ indicate whether this is the case or not for the $k$'th claim. Between Example 2.3.2 and Proposition 2.3.3, we discussed why it might be relevant to include such features into the model. With the features in place, the benefit now reads

$$Y = \sum_{k=1}^{N} I_k X_k$$

where $N, I_1, X_1, I_2, X_2, \ldots$ are independent, $I_1, I_2, \ldots$ are identically distributed with $\mathbb{P}(I_1 = 1) = 1 - \mathbb{P}(I_1 = 0) =: q \in (0, 1]$, and $X_1, X_2, \ldots$ are identically distributed. Since we are assuming that $I_1, I_2, \ldots$ and $X_1, X_2, \ldots$ are independent, it should be noted that we rule out the case of $I_k = \mathbb{1}_{\{X_k > d\}}$ for some fixed deductible $d \geq 0$.

Let $N^\star$ denote the thinned version of $N$ given by

$$N^\star = \sum_{k=1}^{N} I_k.$$

Note that $N^\star$ is independent of $X_1, X_2, \ldots$ by definition. It is tempting to suggest the identity

$$Y \stackrel{?}{=} \sum_{k=1}^{N^\star} X_k, \tag{4.2.2}$$

which is of the form (4.2.1) but with the thinned version $N^\star$ of $N$ in place of $N$. Tempting as it is, this identity is false due to subtle differences in indexing. To see this, note that on the event $(N = 2, I_1 = 0, I_2 = 1)$ we have that $Y$ equals $X_2$ while (4.2.2) yields $X_1$. On the other hand, it is actually possible to show that

$$\mathbb{P}(Y \leq y) = \mathbb{P}\left(\sum_{k=1}^{N^\star} X_k \leq y\right), \qquad y \geq 0, \tag{4.2.3}$$

confer with Exercise XXVI. We conclude that casting the benefit as $\sum_{k=1}^{N^\star} X_k$ is valid in a distributional sense. This demonstrates how the collective risk model in its original formulation of (4.2.1), to which we now return, interacts quite naturally with thinning.

In Section 3.1, we discussed actuarial valuation in general and introduced various premium principles, including the expected value principle and the standard deviation principle. In Exercise XIV, we also took a closer look at the variance principle. To apply these principles, it is vital that the expected value and variance of the benefit $Y$ may be determined with ease. This is indeed the case in the collective risk model, since according to Exercise XXVII, it holds that

$$\mathbb{E}[Y] = \mathbb{E}[X_1]\mathbb{E}[N], \tag{4.2.4}$$

$$\mathrm{Var}[Y] = \mathbb{E}[N]\mathrm{Var}[X_1] + \mathbb{E}[X_1]^2\mathrm{Var}[N], \tag{4.2.5}$$

whenever the expected values and variances of $X_1$ and $N$, respectively, are finite.

To apply more sophisticated premium principles, say for example the quantile principle, it is necessary to determine not only the expected value and variance of $Y$ but actually the full distribution of $Y$ in terms of $X_1$ and $N$. To this end, denote by $F$ the distribution function of $Y$, and denote for $n \in \mathbb{N}$ by $G^{*n}$ the distribution function of $S_n := X_1 + X_2 + \cdots + X_n$. We are now ready to state the main result of this section.

**Proposition 4.2.1.** *It holds that*

$$F(y) = p(0) + \sum_{n=1}^{\infty} G^{*n}(y)p(n), \qquad y \geq 0. \tag{4.2.6}$$

*Proof.* Let $y \geq 0$. It follows from the law of iterated expectations that

$$\begin{aligned} F(y) &= \mathbb{E}[\mathbb{P}(Y \leq y \,|\, N)] \\ &= \mathbb{E}\left[\mathbb{P}\left(\sum_{k=1}^{N} X_k \leq y \,\middle|\, N\right)\right] \\ &= \sum_{n=0}^{\infty} \mathbb{P}(S_n \leq y)p(n) \\ &= p(0) + \sum_{n=1}^{\infty} G^{*n}(y)p(n), \end{aligned}$$

since the claim sizes are assumed to be independent of the claim count. □

The expression (4.2.6) is as 'good as it gets' from an analytical point of view unless further assumptions are imposed. So in general, an approximation is required in order to evaluate this compound distribution. Simulation is, as usual, a straightforward but also computationally very demanding option. Alternatively, for the so-called $(a, b, 0)$-class of discrete distributions, which consists of the Poisson, binomial, and negative binomial distributions, a less computationally demanding option is to discretize the claim size distribution and applying what is known as Panjer's recursion. We do not explore any of these methods here, though, but instead conclude the section by stating the normal power approximation for computation of quantiles.

For $q \in (0,1)$ let $u_q$ denote the $q$-quantile of $F$, that is let

$$u_q = \inf\{y \geq 0 : \mathbb{P}(Y > y) \leq 1 - q\}.$$

Furthermore, let $z_q$ denote the $q$-quantile of the standard normal approximation. Suppose that $\mathbb{E}[Y^3] < \infty$. The normal power approximation then states that

$$u_q \approx \mathbb{E}[Y] + z_q\sqrt{\mathrm{Var}[Y]} + \frac{1}{6}(z_q^2 - 1)\frac{\mathbb{E}[(Y - \mathbb{E}[Y])^3]}{\mathrm{Var}[Y]}.$$

It is outside the scope of this presentation to discuss when and why this approximation is reasonable, although we should state that the condition $\mathbb{E}[Y^3] < \infty$ already indicates that it may not be that viable for heavy-tailed claims.

The normal power approximation is particularly convenient in connection with the quantile principle. Recall that this premium principle reads

$$\pi(Y) = \inf\{y \in [0,\infty) : \mathbb{P}(Y > y) \leq q\}$$

for some $q \in (0,1)$. Consequently, the normal power approximation yields

$$\pi(Y) \approx \mathbb{E}[Y] + z_q\sqrt{\mathrm{Var}[Y]} + \frac{1}{6}(z_q^2 - 1)\frac{\mathbb{E}[(Y - \mathbb{E}[Y])^3]}{\mathrm{Var}[Y]}.$$

The quantiles of the standard normal approximation are tabulated and readily available, while formulas for the expected value and variance of $Y$ were given in (4.2.4)–(4.2.5). It is also possible to derive a formula for the third central moment of $Y$, for example by utilizing moment-generating functions, but since the calculations become quite tedious, we leave it be.

## 4.3 Exercises

**Exercise XXIV.** Let $T_x^{♂}$ and $T_y^{♀}$ be two independent residual lifetimes with mortality rates $\mu^{♂}$ and $\mu^{♀}$, respectively. Consider for some continuous payment rate $c : [0,\infty) \to \mathbb{R}$ and some maximal contract time $\eta > 0$ the random payment stream $B$ given by

$$B(t) = \int_0^t \mathbb{1}_{\{T_y^{♀} > s\}} \mathbb{1}_{\{T_x^{♂} > s\}} \mathbb{1}_{\{s \leq \eta\}} c(s) \, \mathrm{d}s, \qquad t \geq 0.$$

The corresponding present value PV takes the form

$$\mathrm{PV}(t) = \mathbb{1}_{\{t<\eta\}} \int_t^\eta \exp\Big\{-\int_t^s r(u)\,\mathrm{d}u\Big\}\,\mathbb{1}_{\{T_y^{♀}>s\}}\mathbb{1}_{\{T_x^{♂}>s\}} c(s)\,\mathrm{d}s, \qquad t \geq 0.$$

Let $V^{\mathrm{aa}}$ be given by

$$V^{\mathrm{aa}}(t) = \mathbb{E}[\mathrm{PV}(t)\,|\,T_x^{♂} > t, T_y^{♂} > t], \qquad t \geq 0.$$

In this exercise, you are going to show that

$$\frac{\mathrm{d}}{\mathrm{d}t}V^{\mathrm{aa}}(t) = \Big(r(t) + \mu^{♂}(x+t) + \mu^{♀}(y+t)\Big)V^{\mathrm{aa}}(t) - c(t), \qquad 0 < t < \eta, \qquad (4.3.1)$$

in two distinct ways.

a) Show by direct calculation in the spirit of the proof of Proposition 4.1.4 that

$$V^{\mathrm{aa}}(t) = \int_t^\eta \exp\Big\{-\int_t^s \Big(r(u) + \mu^{♂}(x+u) + \mu^{♀}(y+u)\Big)\,\mathrm{d}u\Big\}c(s)\,\mathrm{d}s \qquad (4.3.2)$$

for $0 \leq t \leq \eta$.

b) By applying the techniques from the proof of Theorem 4.1.5 to (4.3.2), show that (4.3.1) indeed holds.

Let the survival time $T$ be given by $T = \min\{T_x^{♂}, T_y^{♀}\}$.

c) Prove that $T$ admits the mortality rate $\mu(t) := \mu^{♂}(x+t) + \mu^{♀}(y+t)$, $t \geq 0$. Hint: Show for example that

$$\mathbb{P}(T > t) = \exp\Big\{-\int_0^t \mu^{♂}(x+s) + \mu^{♀}(y+s)\,\mathrm{d}s\Big\}, \qquad t \geq 0.$$

d) Explain why

$$B(t) = \int_0^t \mathbb{1}_{\{T>s\}}\mathbb{1}_{\{s\leq\eta\}}c(s)\,\mathrm{d}s, \qquad t \geq 0,$$

and use this in conjunction with your result from c) and (4.1.4) to once more verify that (4.3.1) indeed holds.

**Exercise XXV.** In Section 4.1, we developed numerical schemes for the computation of the prospective reserve. In this exercise, we fill in some of the gaps and take a closer look at the theoretical errors that arise in case no maximal contract time exists.

Let the random payment stream $B$ be of the form (4.1.1). Introduce auxiliary quantities $V^{\mathrm{a}}(t,\eta)$, $0 \leq t \leq \eta$, according to

$$V^{\mathrm{a}}(t,\eta) = \int_t^\eta \exp\Big\{-\int_t^s (r(u) + \mu(x+u))\,\mathrm{d}u\Big\}\big(b^{\mathrm{a}}(s) + \mu(x+s)b^{\mathrm{ad}}(s)\big)\,\mathrm{d}s.$$

One of the main ideas behind the numerical schemes developed in Section 4.1 is to compute $V^{\mathrm{a}}(t,\eta)$, $0 \leq t \leq \eta$, for some $\eta \gg 0$ in place of $V^{\mathrm{a}}(t)$, $t \geq 0$. This of course results in a theoretical error, which we now study in more detail.

a) Show that the theoretical errors are given by

$$V^{\mathrm{a}}(t) - V^{\mathrm{a}}(t,\eta) = \exp\left\{-\int_t^\eta (r(u) + \mu(x+u))\,\mathrm{d}u\right\} V^{\mathrm{a}}(\eta) \tag{4.3.3}$$

for $0 \le t \le \eta$.

b) Suppose that $b^{\mathrm{a}}$ and $b^{\mathrm{ad}}$ are continuous. By differentiation based on (4.3.3) followed by an application of Theorem 4.1.5, show that

$$\frac{\mathrm{d}}{\mathrm{d}t} V^{\mathrm{a}}(t,\eta) = r(t)V^{\mathrm{a}}(t,\eta) - b^{\mathrm{a}}(t) - \mu(x+t)\big(b^{\mathrm{ad}}(t) - V^{\mathrm{a}}(t,\eta)\big)$$

for $0 < t < \eta$.

Suppose that long term the force of interest stabilizes around some level $\alpha > 0$, that is suppose $r(t) \approx \alpha$ for $t \ge \eta$. Furthermore, suppose that $\mu(x+t) \approx \beta > 0$ for $t \ge \eta$, which is consistent with the hypothesis that mortality rates flatten for exceptionally high ages, say for supercentenarians. Finally, suppose that $b^{\mathrm{a}}(t) = c \in \mathbb{R}$ and $b^{\mathrm{ad}}(t) = 0$ for $t \ge \eta$.

c) Show that

$$V^{\mathrm{a}}(\eta) \approx \frac{c}{\alpha + \beta}.$$

d) Suppose further that $r \ge 0$. Argue that for any $t \ge 0$ it holds that

$$V^{\mathrm{a}}(t) - V^{\mathrm{a}}(t,\eta) \to 0 \qquad (\text{as } \eta \to \infty).$$

Hint: Recall that

$$\exp\left\{-\int_t^\eta \mu(x+u)\,\mathrm{d}u\right\} = p_x^{\mathrm{aa}}(t,\eta) = \overline{F}_{x+t}(\eta - t),$$

confer with Lemma 2.1.16. What happens with $\overline{F}_{x+t}(\eta - t)$ when $\eta \to \infty$?

**Exercise XXVI.** In this exercise, you are going to prove (4.2.3). To this end, consider a benefit $Y$ on the form

$$Y = \sum_{k=1}^{N} I_k X_k$$

where $N, I_1, X_1, I_2, X_2, \ldots$ are independent, $I_1, I_2, \ldots$ are identically distributed with $\mathbb{P}(I_1 = 1) = 1 - \mathbb{P}(I_1 = 0) =: q \in (0,1]$, and $X_1, X_2, \ldots$ are identically distributed. Furthermore, let $N^\star$ denote the thinned version of $N$ given by

$$N^\star = \sum_{k=1}^{N} I_k,$$

and denote for $n \in \mathbb{N}$ by $G^{*n}$ the distribution function of $S_n := X_1 + X_2 + \cdots X_n$.

a) Let $n \in \mathbb{N}$. Show that

$$\mathbb{P}(Y \leq y \,|\, N^\star = n) = G^{*n}(y), \qquad y \geq 0.$$

What happens if $n = 0$?

b) Let $n \in \mathbb{N}$. Show that

$$\mathbb{P}\left(\sum_{k=1}^{N^\star} X_k \leq y \,\middle|\, N^\star = n\right) = G^{*n}(y), \qquad y \geq 0.$$

What happens if $n = 0$?

c) By applying the law of iterated expectations, establish (4.2.3).

**Exercise XXVII.** In this exercise, you are going to prove (4.2.4)–(4.2.5). To this end, consider a benefit on the form

$$Y = \sum_{k=1}^{N} X_k$$

for independent and identically distributed claims sizes $X_1, X_2, \ldots$ and a claim count $N$ assumed to be independent of the claim sizes.

a) Suppose that the expected values of $X_1$ and $N$ are finite. By applying the law of iterated expectations, show that

$$\mathbb{E}[Y] = \mathbb{E}[X_1]\mathbb{E}[N].$$

b) Suppose that $X_1$ and $N$ have finite variance. By applying the law of iterated variances, show that

$$\mathrm{Var}[Y] = \mathbb{E}[N]\mathrm{Var}[X_1] + \mathbb{E}[X_1]^2\mathrm{Var}[N].$$

Hint: It may be useful to recall that when $X_1, X_2, \ldots, X_n$, $n \in \mathbb{N}$, are assumed to be independent and identically distributed, then

$$\mathrm{Var}\left[\sum_{k=1}^{n} X_k\right] = \sum_{k=1}^{n} \mathrm{Var}[X_k] = n\mathrm{Var}[X_1].$$

**Exercise XXVIII.** Consider a benefit $Y$ on the form

$$Y = \sum_{k=1}^{N} X_k$$

for independent and identically distributed claims sizes $X_1, X_2, \ldots$ and a claim count $N$ assumed to be independent of the claim sizes. Suppose that $N \sim \text{Poisson}(\lambda)$, $\lambda > 0$, and that $\mathbb{E}[X_1^3] < \infty$.

a) Show that $\mathrm{Var}[Y] = \lambda \mathbb{E}[X_1^2]$.

It is possible to show that $\mathbb{E}[(Y - \mathbb{E}[Y])^3] = \lambda \mathbb{E}[X_1^3]$. In the following, you may take this result for granted.

b) Suppose that $\lambda = 50$ and that $X_1$ follows a log-normal distribution with parameters $\mu = 2$ and $\sigma = 1$, confer also with Example 2.2.6. Find an approximation to the 95%-quantile of the distribution function of $Y$.

(Hint: You may have to use that the 95%-quantile of the standard normal distribution is approximately equal to 1.645.)

# Index