Preface to the Second Edition

When I took office, only high energy physicists had ever heard of what is called the Worldwide Web . . . Now even my cat has its own page — Bill Clinton, 1996

This book gives a comprehensive survey of non-life insurance mathematics. Originally written for use with the actuarial science programs at the Universities of Amsterdam and Leuven, it is now in use at many other universities, as well as for the non-academic actuarial education program organized by the Dutch Actuarial Society. It provides a link to the further theoretical study of actuarial science. The methods presented can not only be used in non-life insurance, but are also effective in other branches of actuarial science, as well as, of course, in actuarial practice.

Apart from the standard theory, this text contains methods directly relevant for actuarial practice, for example the rating of automobile insurance policies, premium principles and risk measures, and IBNR models. Also, the important actuarial statistical tool of the Generalized Linear Models is studied. These models provide extra possibilities beyond ordinary linear models and regression that are the statistical tools of choice for econometricians. Furthermore, a short introduction is given to credibility theory. Another topic which always has enjoyed the attention of risk theoreticians is the study of ordering of risks. The book reflects the state of the art in actuarial risk theory; many results presented were published in the actuarial literature only recently.

In this second edition of the book, we have aimed to make the theory even more directly applicable by using the software R. It provides an implementation of the language S, not unlike S-Plus. It is not just a set of statistical routines but a full-fledged object oriented programming language. Other software may provide similar capabilities, but the great advantage of R is that it is open source, hence available to everyone free of charge. This is why we feel justified in imposing it on the users of this book as a de facto standard. On the internet, a lot of documentation about R can be found. In an Appendix, we give some examples of use of R. After a general introduction, explaining how it works, we study a problem from risk management, trying to forecast the future behavior of stock prices with a simple model, based on stock prices of three recent years. Next, we show how to use R to generate pseudo-random datasets that resemble what might be encountered in actuarial practice.

Models and paradigms studied

The time aspect is essential in many models of life insurance. Between paying premiums and collecting the resulting pension, decades may elapse. This time element is less prominent in non-life insurance. Here, however, the statistical models are generally more involved. The topics in the first five chapters of this textbook are basic for non-life actuarial science. The remaining chapters contain short introductions to other topics traditionally regarded as non-life actuarial science.

1. *The expected utility model*

The very existence of insurers can be explained by the expected utility model. In this model, an insured is a risk averse and rational decision maker, who by virtue of Jensen's inequality is ready to pay more than the expected value of his claims just to be in a secure financial position. The mechanism through which decisions are taken under uncertainty is not by direct comparison of the expected payoffs of decisions, but rather of the expected utilities associated with these payoffs.

2. *The individual risk model*

In the individual risk model, as well as in the collective risk model below, the total claims on a portfolio of insurance contracts is the random variable of interest. We want to compute, for example, the probability that a certain capital will be sufficient to pay these claims, or the value-at-risk at level 99.5% associated with the portfolio, being the 99.5% quantile of its cumulative distribution function (cdf). The total claims is modeled as the sum of all claims on the policies, which are assumed independent. Such claims cannot always be modeled as purely discrete random variables, nor as purely continuous ones, and we use a notation, involving Stieltjes integrals and differentials, encompassing both these as special cases.

The individual model, though the most realistic possible, is not always very convenient, because the available dataset is not in any way condensed. The obvious technique to use in this model is convolution, but it is generally quite awkward. Using transforms like the moment generating function sometimes helps. The Fast Fourier Transform (FFT) technique gives a fast way to compute a distribution from its characteristic function. It can easily be implemented in R.

We also present approximations based on fitting moments of the distribution. The Central Limit Theorem, fitting two moments, is not sufficiently accurate in the important right-hand tail of the distribution. So we also look at some methods using three moments: the translated gamma and the normal power approximation.

3. *Collective risk models*

A model that is often used to approximate the individual model is the collective risk model. In this model, an insurance portfolio is regarded as a process that produces claims over time. The sizes of these claims are taken to be independent, identically distributed random variables, independent also of the number of claims generated. This makes the total claims the sum of a random number of iid individual claim amounts. Usually one assumes additionally that the number of claims is a Poisson variate with the right mean, or allows for some overdispersion by taking a negative binomial claim number. For the cdf of the individual claims, one takes an average of the cdfs of the individual policies. This leads to a close fitting and computationally tractable model. Several techniques, including Panjer's recursion formula, to compute the cdf of the total claims modeled this way are presented.

For some purposes it is convenient to replace the observed claim severity distribution by a parametric loss distribution. Families that may be considered are for example the gamma and the lognormal distributions. We present a number of such distributions, and also demonstrate how to estimate the parameters from data. Further, we show how to generate pseudo-random samples from these distributions, beyond the standard facilities offered by R.

4. *The ruin model*

The ruin model describes the stability of an insurer. Starting from capital u at time $t = 0$, his capital is assumed to increase linearly in time by fixed annual premiums, but it decreases with a jump whenever a claim occurs. Ruin occurs when the capital is negative at some point in time. The probability that this ever happens, under the assumption that the annual premium as well as the claim generating process remain unchanged, is a good indication of whether the insurer's assets match his liabilities sufficiently. If not, one may take out more reinsurance, raise the premiums or increase the initial capital.

Analytical methods to compute ruin probabilities exist only for claims distributions that are mixtures and combinations of exponential distributions. Algorithms exist for discrete distributions with not too many mass points. Also, tight upper and lower bounds can be derived. Instead of looking at the ruin probability $\psi(u)$ with initial capital u, often one just considers an upper bound e^{-Ru} for it (Lundberg), where the number R is the so-called adjustment coefficient and depends on the claim size distribution and the safety loading contained in the premium.

Computing a ruin probability assumes the portfolio to be unchanged eternally. Moreover, it considers just the insurance risk, not the financial risk. Therefore not much weight should be attached to its precise value beyond, say, the first relevant decimal. Though some claim that survival probabilities are 'the goal of risk theory', many actuarial practitioners are of the opinion that ruin theory, however topical still in academic circles, is of no significance to them. Nonetheless, we recommend to study at least the first three sections of Chapter 4, which contain the description of the Poisson process as well as some key results. A simple proof is provided for Lundberg's exponential upper bound, as well as a derivation of the ruin probability in case of exponential claim sizes.

5. *Premium principles and risk measures*

Assuming that the cdf of a risk is known, or at least some characteristics of it like mean and variance, a premium principle assigns to the risk a real number used as a financial compensation for the one who takes over this risk. Note that we study only risk premiums, disregarding surcharges for costs incurred by the insurance company. By the law of large numbers, to avoid eventual ruin the total premium should be at least equal to the expected total claims, but additionally, there has to be a loading in the premium to compensate the insurer for making available his risk carrying capacity. From this loading, the insurer has to build a reservoir to draw upon in adverse times, so as to avoid getting in ruin. We present a number of premium principles, together with the most important properties that characterize premium principles. The choice of a premium principle depends heavily on the importance attached to such properties. There is no premium principle that is uniformly best.

Risk measures also attach a real number to some risky situation. Examples are premiums, infinite ruin probabilities, one-year probabilities of insolvency, the required capital to be able to pay all claims with a prescribed probability, the expected value of the shortfall of claims over available capital, and more.

6. *Bonus-malus systems*
With some types of insurance, notably car insurance, charging a premium based exclusively on factors known a priori is insufficient. To incorporate the effect of risk factors of which the use as rating factors is inappropriate, such as race or quite often sex of the policy holder, and also of non-observable factors, such as state of health, reflexes and accident proneness, many countries apply an experience rating system. Such systems on the one hand use premiums based on a priori factors such as type of coverage and list-price or weight of a car, on the other hand they adjust these premiums by using a bonus-malus system, where one gets more discount after a claim-free year, but pays more after filing one or more claims. In this way, premiums are charged that reflect the exact driving capabilities of the driver better. The situation can be modeled as a Markov chain.

The quality of a bonus-malus system is determined by the degree in which the premium paid is in proportion to the risk. The Loimaranta efficiency equals the elasticity of the mean premium against the expected number of claims. Finding it involves computing eigenvectors of the Markov matrix of transition probabilities. R provides tools to do this.

7. *Ordering of risks*
It is the very essence of the actuary's profession to be able to express preferences between random future gains or losses. Therefore, stochastic ordering is a vital part of his education and of his toolbox. Sometimes it happens that for two losses X and Y, it is known that every sensible decision maker prefers losing X, because Y is in a sense 'larger' than X. It may also happen that only the smaller group of all risk averse decision makers agree about which risk to prefer. In this case, risk Y may be larger than X, or merely more 'spread', which also makes a risk less attractive. When we interpret 'more spread' as having thicker tails of the cumulative distribution function, we get a method of ordering risks that has many appealing properties. For example, the preferred loss also outperforms the other one as regards zero utility premiums, ruin probabilities, and stop-loss premiums for compound distributions with these risks as individual terms. It can be shown that the collective model of Chapter 3 is more spread than the individual model it approximates, hence using the collective model, in most cases, leads to more conservative decisions regarding premiums to be asked, reserves to be held, and values-at-risk. Also, we can prove

that the stop-loss insurance, demonstrated to be optimal as regards the variance of the retained risk in Chapter 1, is also preferable, other things being equal, in the eyes of all risk averse decision makers.

Sometimes, stop-loss premiums have to be set under incomplete information. We give a method to compute the maximal possible stop-loss premium assuming that the mean, the variance and an upper bound for a risk are known.

In the individual and the collective model, as well as in ruin models, we assume that the claim sizes are stochastically independent non-negative random variables. Sometimes this assumption is not fulfilled, for example there is an obvious dependence between the mortality risks of a married couple, between the earthquake risks of neighboring houses, and between consecutive payments resulting from a life insurance policy, not only if the payments stop or start in case of death, but also in case of a random force of interest. We give a short introduction to the risk ordering that applies for this case. It turns out that stop-loss premiums for a sum of random variables with an unknown joint distribution but fixed marginals are maximal if these variables are as dependent as the marginal distributions allow, making it impossible that the outcome of one is 'hedged' by another.

In finance, frequently one has to determine the distribution of the sum of dependent lognormal random variables. We apply the theory of ordering of risks and comonotonicity to give bounds for that distribution.

We also give a short introduction in the theory of ordering of multivariate risks. One might say that two randoms variables are more related than another pair with the same marginals if their correlation is higher. But a more robust criterion is to restrict this to the case that their joint cdf is uniformly larger. In that case it can be proved that the sum of these random variables is larger in stop-loss order. There are bounds for joints cdfs dating back to Fréchet in the 1950's and Höffding in the 1940's. For a random pair (X,Y), the copula is the joint cdf of the *ranks* $F_X(X)$ and $F_Y(Y)$. Using the smallest and the largest copula, it is possible to construct random pairs with arbitrary prescribed marginals and (rank) correlations.

8. *Credibility theory*
The claims experience on a policy may vary by two different causes. The first is the quality of the risk, expressed through a risk parameter. This represents the average annual claims in the hypothetical situation that the policy is monitored without change over a very long period of time. The other is the purely random good and bad luck of the policyholder that results in yearly deviations from the risk parameter. Credibility theory assumes that the risk quality is a drawing from a certain structure distribution, and that conditionally given the risk quality, the actual claims experience is a sample from a distribution having the risk quality as its mean value. The predictor for next year's experience that is linear in the claims experience and optimal in the sense of least squares turns out to be a weighted average of the claims experience of the individual contract and the experience for the whole portfolio. The weight factor is the credibility attached to the individual experience, hence it is called the credibility factor, and the resulting premiums are called credibility pre-

miums. As a special case, we study a bonus-malus system for car insurance based on a Poisson-gamma mixture model.

Credibility theory is actually a Bayesian inference method. Both credibility and generalized linear models (see below) are in fact special cases of so-called Generalized Linear Mixed Models (GLMM), and the `R` function `glmm` is able to deal with both the random and the fixed parameters in these models.

9. *Generalized linear models*

Many problems in actuarial statistics are Generalized Linear Models (GLM). Instead of assuming a normally distributed error term, other types of randomness are allowed as well, such as Poisson, gamma and binomial. Also, the expected values of the dependent variables need not be linear in the regressors. They may also be some function of a linear form of the covariates, for example the logarithm leading to the multiplicative models that are appropriate in many insurance situations.

This way, one can for example tackle the problem of estimating the reserve to be kept for IBNR claims, see below. But one can also easily estimate the premiums to be charged for drivers from region i in bonus class j with car weight w.

In credibility models, there are random group effects, but in GLMs the effects are fixed, though unknown. The `glmm` function in `R` can handle a multitude of models, including those with both random and fixed effects.

10. *IBNR techniques*

An important statistical problem for the practicing actuary is the forecasting of the total of the claims that are Incurred, But Not Reported, hence the acronym IBNR, or not fully settled. Most techniques to determine estimates for this total are based on so-called run-off triangles, in which claim totals are grouped by year of origin and development year. Many traditional actuarial reserving methods turn out to be maximum likelihood estimations in special cases of GLMs.

We describe the workings of the ubiquitous chain ladder method to predict future losses, as well as, briefly, the Bornhuetter-Ferguson method, which aims to incorporate actuarial knowledge about the portfolio. We also show how these methods can be implemented in `R`, using the `glm` function. In this same framework, many extensions and variants of the chain ladder method can easily be introduced. England and Verrall have proposed methods to describe the prediction error with the chain ladder method, both an analytical estimate of the variance and a bootstrapping method to obtain an estimate for the predictive distribution. We describe an `R` implementation of these methods.

11. *More on GLMs*

For the second edition, we extended the material in virtually all chapters, mostly involving the use of `R`, but we also add some more material on GLMs. We briefly recapitulate the Gauss-Markov theory of ordinary linear models found in many other texts on statistics and econometrics, and explain how the algorithm by Nelder and Wedderburn works, showing how it can be implemented in `R`. We also study the stochastic component of a GLM, stating that the observations are independent random variables with a distribution in a subclass of the exponential family. The well-known normal, Poisson and gamma families have a variance proportional to μ^p for $p = 0, 1, 2$, respectively, where μ is the mean (heteroskedasticity). The so-called Tweedie class contains random variables, in fact compound Poisson–gamma risks, having variance proportional to μ^p for some $p \in (1, 2)$. These mean-variance relations are interesting for actuarial purposes. Extensions to R, contributed by Dunn and Smyth provide routines computing cdf, inverse cdf, pdf and random drawings of such random variables, as well as to estimate GLMs with Tweedie distributed risks.

Educational aspects

As this text has been in use for a long time now at the University of Amsterdam and elsewhere, we could draw upon a long series of exams, resulting in long lists of exercises. Also, many examples are given, making this book well-suited as a textbook. Some less elementary exercises have been marked by [♠], and these might be skipped.

The required mathematical background is on a level such as acquired in the first stage of a bachelors program in quantitative economics (econometrics or actuarial science), or mathematical statistics. To indicate the level of what is needed, the book by Bain and Engelhardt (1992) is a good example. So the book can be used either in the final year of such a bachelors program, or in a subsequent masters program in either actuarial science proper or in quantitative financial economics with a strong insurance component. To make the book accessible to non-actuaries, notation and jargon from life insurance mathematics is avoided. Therefore also students in applied mathematics or statistics with an interest in the stochastic aspects of insurance will be able to study from this book. To give an idea of the mathematical rigor and statistical sophistication at which we aimed, let us remark that moment generating functions are used routinely, while characteristic functions and measure theory are avoided in general. Prior experience with regression models, though helpful, is not required.

As a service to the student help is offered, in Appendix B, with many of the exercises. It takes the form of either a final answer to check one's work, or a useful hint. There is an extensive index, and the tables that might be needed on an exam are printed in the back. The list of references is not a thorough justification with bibliographical data on every result used, but more a collection of useful books and papers containing more details on the topics studied, and suggesting further reading.

Ample attention is given to exact computing techniques, and the possibilities that R provides, but also to old fashioned approximation methods like the Central Limit Theorem (CLT). The CLT itself is generally too crude for insurance applications, but slight refinements of it are not only fast, but also often prove to be surprisingly accurate. Moreover they provide solutions of a parametric nature such that one does not have to recalculate everything after a minor change in the data. Also, we want to stress that 'exact' methods are as exact as their input. The order of magnitude of errors resulting from inaccurate input is often much greater than the one caused by using an approximation method.

The notation used in this book conforms to what is usual in mathematical statistics as well as non-life insurance mathematics. See for example the book by Bowers et al. (1986, 1997), the non-life part of which is similar in design to the first part of this book. In particular, random variables are capitalized, though not all capitals actually denote random variables.

Acknowledgments
First and most of all, the authors would like to thank David Vyncke for all the work he did on the first edition of this book. Many others have provided useful input. We thank (in pseudo-random order) Hans Gerber, Elias Shiu, Angela van Heerwaarden, Dennis Dannenburg, Richard Verrall, Klaus Schmidt, Bjørn Sundt, Vsevolod Malinovskii, who translated this textbook into Russian, Qihe Tang, who translated it into Chinese in a joint project with Taizhong Hu and Shixue Cheng, Roger Laeven, Ruud Koning, Julien Tomas, Katrien Antonio as well as Steven Vanduffel for their comments, and Vincent Goulet for helping us with some R-problems. We also acknowledge the numerous comments of the users, students and teachers alike.

Jan Dhaene and Marc Goovaerts acknowledge the support of the Fortis Chair on Financial and Actuarial Risk Management.

World wide web support
The authors would like to keep in touch with the users of this text. On the internet page *http://www1.fee.uva.nl/ke/act/people/kaas/ModernART.htm* we maintain a list of all typos that have been found so far, and indicate how teachers may obtain solutions to the exercises as well as the slides used at the University of Amsterdam for courses based on this book.

To save users a lot of typing, and typos, this site also provides the R commands used for the examples in the book.

Amsterdam,	*Rob Kaas*
Leuven,	*Marc Goovaerts*
Louvain-la-Neuve,	*Jan Dhaene*
May 18, 2008	*Michel Denuit*

Chapter 2
The individual risk model

If the automobile had followed the same development cycle as the computer, a Rolls-Royce would today cost $100, get a million miles per gallon, and explode once a year, killing everyone inside — Robert X. Cringely

2.1 Introduction

In this chapter we focus on the distribution function of the total claim amount S for the portfolio of an insurer. We are not merely interested in the expected value and the variance of the insurer's random capital, but we also want to know the probability that the amounts paid exceed a fixed threshold. The distribution of the total claim amount S is also necessary to be able to apply the utility theory of the previous chapter. To determine the value-at-risk at, say, the 99.5% level, we need also good approximations for the inverse of the cdf, especially in the far tail. In this chapter we deal with models that still recognize the individual, usually different, policies. As is done often in non-life insurance mathematics, the time aspect will be ignored. This aspect is nevertheless important in disability and long term care insurance. For this reason, these types of insurance are sometimes considered life insurances.

In the insurance practice, risks usually cannot be modeled by purely discrete random variables, nor by purely continuous random variables. For example, in liability insurance a whole range of positive amounts can be paid out, each of them with a very small probability. There are two exceptions: the probability of having no claim, that is, claim size 0, is quite large, and the probability of a claim size that equals the maximum sum insured, implying a loss exceeding that threshold, is also not negligible. For expectations of such mixed random variables, we use the Riemann-Stieltjes integral as a notation, without going too deeply into its mathematical aspects. A simple and flexible model that produces random variables of this type is a mixture model, also called an 'urn-of-urns' model. Depending on the outcome of one drawing, resulting in one of the events 'no claim or maximum claim' or 'other claim', a second drawing is done from either a discrete distribution, producing zero or the maximal claim amount, or a continuous distribution. In the sequel, we present some examples of mixed models for the claim amount per policy.

Assuming that the risks in a portfolio are independent random variables, the distribution of their sum can be calculated by making use of convolution. Even with the computers of today, it turns out that this technique is quite laborious, so there is a need for other methods. One of the alternative methods is to make use of moment generating functions (mgf) or of related transforms like characteristic functions, probability generating functions (pgf) and cumulant generating functions (cgf). Sometimes it is possible to recognize the mgf of a sum of independent random variables and consequently identify the distribution function. And in some cases we can fruitfully employ a technique called the *Fast Fourier Transform* to reconstruct the density from a transform.

A totally different approach is to compute approximations of the distribution of the total claim amount S. If we consider S as the sum of a 'large' number of random variables, we could, by virtue of the Central Limit Theorem, approximate its distribution by a normal distribution with the same mean and variance as S. We will show that this approximation usually is not satisfactory for the insurance practice. Especially in the tails, there is a need for more refined approximations that explicitly recognize the substantial probability of large claims. More technically, the third central moment of S is usually greater than 0, while for the normal distribution it equals 0. We present an approximation based on a translated gamma random variable, as well as the normal power (NP) approximation. The quality of these approximations is similar. The latter can be calculated directly by means of a $\mathrm{N}(0,1)$ table, the former requires using a computer.

Another way to approximate the individual risk model is to use the collective risk models described in the next chapter.

2.2 Mixed distributions and risks

In this section, we discuss some examples of insurance risks, that is, the claims on an insurance policy. First, we have to slightly extend the set of distribution functions we consider, because purely discrete random variables and purely continuous random variables both turn out to be inadequate for modeling the risks.

From the theory of probability, we know that every function $F(\cdot)$ that satisfies

$$\begin{aligned} &F(-\infty) = 0; \quad F(+\infty) = 1 \\ &F(\cdot) \text{ is non-decreasing and right-continuous} \end{aligned} \tag{2.1}$$

is a cumulative distribution function (cdf) of some random variable, for example of $F^{-1}(U)$ with $U \sim$ uniform(0,1), see Section 3.9.1 and Definition 5.6.1. If $F(\cdot)$ is a step function, that is, a function that is constant outside a denumerable set of discontinuities (steps), then $F(\cdot)$ and any random variable X with $F(x) = \Pr[X \le x]$ are called *discrete*. The associated probability density function (pdf) represents the height of the step at x, so

$$f(x) = F(x) - F(x-0) = \Pr[X = x] \quad \text{for all } x \in (-\infty, \infty). \tag{2.2}$$

Here, $F(x-0)$ is shorthand for $\lim_{\varepsilon\downarrow 0} F(x-\varepsilon)$; $F(x+0) = F(x)$ holds because of right-continuity. For all x, we have $f(x) \geq 0$, and $\sum_x f(x) = 1$ where the sum is taken over the denumerable set of all x with $f(x) > 0$.

Another special case is when $F(\cdot)$ is *absolutely continuous*. This means that if $f(x) = F'(x)$, then

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t. \tag{2.3}$$

In this case $f(\cdot)$ is called the probability density function, too. Again, $f(x) \geq 0$ for all x, while now $\int f(x)\,\mathrm{d}x = 1$. Note that, just as is customary in mathematical statistics, this notation without integration limits represents the *definite* integral of $f(x)$ over the interval $(-\infty,\infty)$, and not just an arbitrary antiderivative, that is, any function having $f(x)$ as its derivative.

In statistics, almost without exception random variables are either discrete or continuous, but this is definitely not the case in insurance. Many distribution functions to model insurance payments have continuously increasing parts, but also some positive steps. Let Z represent the payment on some contract. There are three possibilities:

1. The contract is claim-free, hence $Z = 0$.
2. The contract generates a claim that is larger than the maximum sum insured, say M. Then, $Z = M$.
3. The contract generates a 'normal' claim, hence $0 < Z < M$.

Apparently, the cdf of Z has steps in 0 and in M. For the part in-between we could use a discrete distribution, since the payment will be some integer multiple of the monetary unit. This would produce a very large set of possible values, each of them with a very small probability, so using a continuous cdf seems more convenient. In this way, a cdf arises that is neither purely discrete, nor purely continuous. In Figure 2.2 a diagram of a mixed continuous/discrete cdf is given, see also Exercise 1.4.1.

The following urn-of-urns model allows us to construct a random variable with a distribution that is a mixture of a discrete and a continuous distribution. Let I be an *indicator random variable*, with values $I = 1$ or $I = 0$, where $I = 1$ indicates that some event has occurred. Suppose that the probability of the event is $q = \Pr[I = 1]$, $0 \leq q \leq 1$. If $I = 1$, in the second stage the claim Z is drawn from the distribution of X, if $I = 0$, then from Y. This means that

$$Z = IX + (1-I)Y. \tag{2.4}$$

If $I = 1$ then Z can be replaced by X, if $I = 0$ it can be replaced by Y. Note that we may act as if not just I and X,Y are independent, but in fact the triple (X,Y,I); only the conditional distributions of $X \mid I = 1$ and of $Y \mid I = 0$ are relevant, so we can take for example $\Pr[X \leq x \mid I = 0] = \Pr[X \leq x \mid I = 1]$ just as well. Hence, the cdf of Z can be written as

$$\begin{aligned} F(z) &= \Pr[Z \le z] \\ &= \Pr[Z \le z, I = 1] + \Pr[Z \le z, I = 0] \\ &= \Pr[X \le z, I = 1] + \Pr[Y \le z, I = 0] \\ &= q \Pr[X \le z] + (1-q) \Pr[Y \le z]. \end{aligned} \tag{2.5}$$

Now, let X be a discrete random variable and Y a continuous random variable. From (2.5) we get

$$F(z) - F(z-0) = q \Pr[X = z] \quad \text{and} \quad F'(z) = (1-q) \frac{\mathrm{d}}{\mathrm{d}z} \Pr[Y \le z]. \tag{2.6}$$

This construction yields a cdf $F(z)$ with steps where $\Pr[X = z] > 0$, but it is not a step function, since $F'(z) > 0$ on the support of Y.

To calculate the moments of Z, the moment generating function $\mathrm{E}[\mathrm{e}^{tZ}]$ and the stop-loss premiums $\mathrm{E}[(Z-d)_+]$, we have to calculate the expectations of functions of Z. For that purpose, we use the iterative formula of conditional expectations, also known as the law of total expectation, the law of iterated expectations, the tower rule, or the smoothing theorem:

$$\mathrm{E}[W] = \mathrm{E}[\mathrm{E}[W \mid V]]. \tag{2.7}$$

We apply this formula with $W = g(Z)$ for an appropriate function $g(\cdot)$ and replace V by I. Then, introducing $h(i) = \mathrm{E}[g(Z) \mid I = i]$, we get, using (2.6) at the end:

$$\begin{aligned} \mathrm{E}[g(Z)] &= \mathrm{E}[\mathrm{E}[g(Z) \mid I]] = qh(1) + (1-q)h(0) = \mathrm{E}[h(I)] \\ &= q\mathrm{E}[g(Z) \mid I = 1] + (1-q)\mathrm{E}[g(Z) \mid I = 0] \\ &= q\mathrm{E}[g(X) \mid I = 1] + (1-q)\mathrm{E}[g(Y) \mid I = 0] \\ &= q\mathrm{E}[g(X)] + (1-q)\mathrm{E}[g(Y)] \\ &= q \sum_z g(z) \Pr[X = z] + (1-q) \int_{-\infty}^{\infty} g(z) \frac{\mathrm{d}}{\mathrm{d}z} \Pr[Y \le z] \, \mathrm{d}z \\ &= \sum_z g(z)[F(z) - F(z-0)] + \int_{-\infty}^{\infty} g(z) F'(z) \, \mathrm{d}z. \end{aligned} \tag{2.8}$$

Remark 2.2.1 (Riemann-Stieltjes integrals)

The result in (2.8), consisting of a sum and an ordinary Riemann integral, can be written as a right hand Riemann-Stieltjes integral:

$$\mathrm{E}[g(Z)] = \int_{-\infty}^{\infty} g(z) \, \mathrm{d}F(z). \tag{2.9}$$

The integrator is the differential $\mathrm{d}F(z) = F_Z(z) - F_Z(z - \mathrm{d}z)$. It replaces the probability of z, that is, the height of the step at z if there is one, or $F'(z)\,\mathrm{d}z$ if there is no step at z. Here, $\mathrm{d}z$ denotes a positive infinitesimally small number. Note that the cdf $F(z) = \Pr[Z \le z]$ is continuous from the right. In life insurance mathematics theory, Riemann-Stieltjes integrals were used as a tool to describe situations in which it is vital which value of the integrand should be taken: the limit from the right, the limit from the left, or the actual function value. Actuarial practitioners have not adopted this convention. We avoid this problem altogether by considering continuous integrands only. ∇

Remark 2.2.2 (Mixed random variables and mixed distributions)
We can summarize the above as follows: a mixed continuous/discrete cdf $F_Z(z) = \Pr[Z \leq z]$ arises when a mixture of random variables

$$Z = IX + (1-I)Y \tag{2.10}$$

is used, where X is a discrete random variable, Y is a continuous random variable and I is a Bernoulli(q) random variable, with X, Y and I independent. The cdf of Z is again a mixture, that is, a convex combination, of the cdfs of X and Y, see (2.5):

$$F_Z(z) = qF_X(z) + (1-q)F_Y(z) \tag{2.11}$$

For expectations of functions $g(\cdot)$ of Z we get the same mixture of expectations of $\mathrm{E}[g(X)]$ and $\mathrm{E}[g(Y)]$, see (2.8):

$$\mathrm{E}[g(Z)] = q\mathrm{E}[g(X)] + (1-q)\mathrm{E}[g(Y)]. \tag{2.12}$$

It is important to make a distinction between the urn-of-urns model (2.10) leading to a convex combination of *cdfs*, and a convex combination of *random variables* $T = qX + (1-q)Y$. Although (2.12) is valid for $T = Z$ in case $g(z) = z$, the random variable T does not have (2.11) as its cdf. See also Exercises 2.2.8 and 2.2.9. ∇

Example 2.2.3 (Insurance against bicycle theft)
We consider an insurance policy against bicycle theft that pays b in case the bicycle is stolen, upon which event the policy ends. Obviously, the number of payments is 0 or 1 and the amount is known in advance, just as with life insurance policies. Assume that the probability of theft is q and let $X = Ib$ denote the claim payment, where I is a Bernoulli(q) distributed indicator random variable, with $I = 1$ if the bicycle is stolen, $I = 0$ if not. In analogy to (2.4), we can rewrite X as $X = Ib + (1-I)0$. The distribution and the moments of X can be obtained from those of I:

$$\begin{aligned} &\Pr[X=b] = \Pr[I=1] = q; &&\Pr[X=0] = \Pr[I=0] = 1-q; \\ &\mathrm{E}[X] = b\mathrm{E}[I] = bq; &&\mathrm{Var}[X] = b^2\mathrm{Var}[I] = b^2q(1-q). \end{aligned} \tag{2.13}$$

Now suppose that only half the amount is paid out in case the bicycle was not locked. Some bicycle theft insurance policies have a restriction like this. Insurers check this by requiring that all the original keys have to be handed over in the event of a claim. Then, $X = IB$, where B represents the random payment. Assuming that the probabilities of a claim $X = 400$ and $X = 200$ are 0.05 and 0.15, we get

$$\Pr[I=1, B=400] = 0.05; \quad \Pr[I=1, B=200] = 0.15. \tag{2.14}$$

Hence, $\Pr[I=1] = 0.2$ and consequently $\Pr[I=0] = 0.8$. Also,

$$\Pr[B=400\,|\,I=1]=\frac{\Pr[B=400,I=1]}{\Pr[I=1]}=0.25. \tag{2.15}$$

This represents the conditional probability that the bicycle was locked given the fact that it was stolen. ∇

Example 2.2.4 (Exponential claim size, if there is a claim)
Suppose that risk X is distributed as follows:

1. $\Pr[X=0]=\frac{1}{2}$;
2. $\Pr[X\in[x,x+\mathrm{d}x)]=\frac{1}{2}\beta\mathrm{e}^{-\beta x}\mathrm{d}x \quad$ for $\beta=0.1, x>0$,

where $\mathrm{d}x$ denotes a positive infinitesimal number. What is the expected value of X, and what is the maximum premium for X that someone with an exponential utility function with risk aversion $\alpha=0.01$ is willing to pay?

The random variable X is not continuous, because the cdf of X has a step in 0. It is also not a discrete random variable, since the cdf is not a step function; its derivative, which in terms of infinitesimal numbers equals $\Pr[x\le X<x+\mathrm{d}x]/\mathrm{d}x$, is positive for $x>0$. We can calculate the expectations of functions of X by dealing with the steps in the cdf separately, see (2.9). This leads to

$$\mathrm{E}[X]=\int_{-\infty}^{\infty}x\,\mathrm{d}F_X(x)=0\,\mathrm{d}F_X(0)+\int_0^{\infty}xF_X'(x)\,\mathrm{d}x=\tfrac{1}{2}\int_0^{\infty}x\beta\mathrm{e}^{-\beta x}\mathrm{d}x=5. \tag{2.16}$$

If the utility function of the insured is exponential with parameter $\alpha=0.01$, then (1.21) yields for the maximum premium P^+:

$$\begin{aligned}P^+&=\frac{1}{\alpha}\log(\mathrm{m}_X(\alpha))=\frac{1}{\alpha}\log\left(\mathrm{e}^0\mathrm{d}F_X(0)+\tfrac{1}{2}\int_0^{\infty}\mathrm{e}^{\alpha x}\beta\mathrm{e}^{-\beta x}\mathrm{d}x\right)\\&=\frac{1}{\alpha}\log\left(\frac{1}{2}+\frac{1}{2}\frac{\beta}{\beta-\alpha}\right)=100\log\left(\frac{19}{18}\right)\approx 5.4.\end{aligned} \tag{2.17}$$

This same result can of course be obtained by writing X as in (2.10). ∇

Example 2.2.5 (Liability insurance with a maximum coverage)
Consider an insurance policy against a liability loss S. We want to determine the expected value, the variance and the distribution function of the payment X on this policy, when there is a deductible of 100 and a maximum payment of 1000. In other words, if $S\le 100$ then $X=0$, if $S\ge 1100$ then $X=1000$, otherwise $X=S-100$. The probability of a positive claim ($S>100$) is 10% and the probability of a large loss ($S\ge 1100$) is 2%. Given $100<S<1100$, S has a uniform$(100,1100)$ distribution. Again, we write $X=IB$ where I denotes the number of payments, 0 or 1, and B represents the amount paid, if any. Therefore,

$$\begin{aligned}&\Pr[B=1000\,|\,I=1]=0.2;\\&\Pr[B\in(x,x+\mathrm{d}x)\,|\,I=1]=c\,\mathrm{d}x \quad \text{for } 0<x<1000.\end{aligned} \tag{2.18}$$

Integrating the latter probability over $x\in(0,1000)$ yields 0.8, so $c=0.0008$.

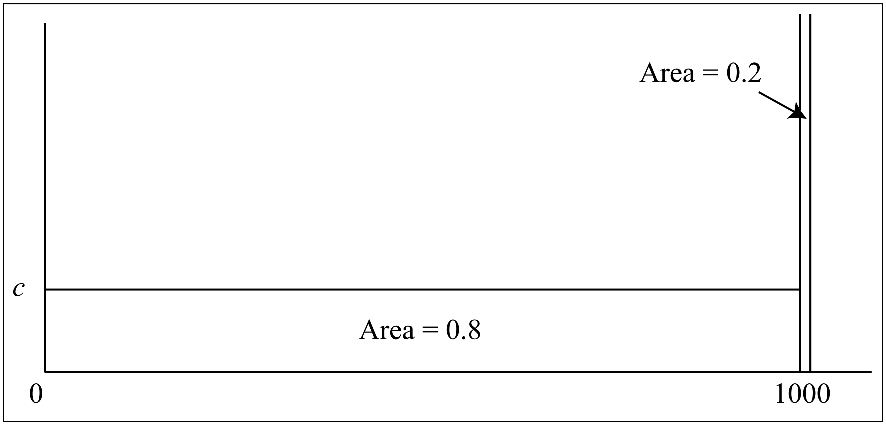

Fig. 2.1 'Probability density function' of B given $I = 1$ in Example 2.2.5.

The conditional distribution function of B, given $I = 1$, is neither discrete, nor continuous. In Figure 2.1 we attempt to depict a pdf by representing the probability mass at 1000 by a bar of infinitesimal width and infinite height such that the area equals 0.2. In actual fact we have plotted $f(\cdot)$, where $f(x) = 0.0008$ on $(0, 1000)$ and $f(x) = 0.2/\varepsilon$ on $(1000, 1000+\varepsilon)$ with $\varepsilon > 0$ very small.

For the cdf F of X we have

$$\begin{aligned} F(x) &= \Pr[X \le x] = \Pr[IB \le x] \\ &= \Pr[IB \le x, I = 0] + \Pr[IB \le x, I = 1] \\ &= \Pr[IB \le x \,|\, I = 0]\Pr[I = 0] + \Pr[IB \le x \,|\, I = 1]\Pr[I = 1] \end{aligned} \tag{2.19}$$

which yields

$$F(x) = \begin{cases} 0 \times 0.9 + 0 \times 0.1 = 0 & \text{for } x < 0 \\ 1 \times 0.9 + 1 \times 0.1 = 1 & \text{for } x \ge 1000 \\ 1 \times 0.9 + c\,x \times 0.1 & \text{for } 0 \le x < 1000. \end{cases} \tag{2.20}$$

A graph of the cdf F is shown in Figure 2.2. For the differential ('density') of F, we have

$$\mathrm{d}F(x) = \begin{cases} 0.9 & \text{for } x = 0 \\ 0.02 & \text{for } x = 1000 \\ 0 & \text{for } x < 0 \text{ or } x > 1000 \\ 0.00008\,\mathrm{d}x & \text{for } 0 < x < 1000. \end{cases} \tag{2.21}$$

The moments of X can be calculated by using this differential. ∇

The variance of risks of the form IB can be calculated through the conditional distribution of B, given I, by use of the well-known *variance decomposition rule*, see (2.7), which is also known as the law of total variance:

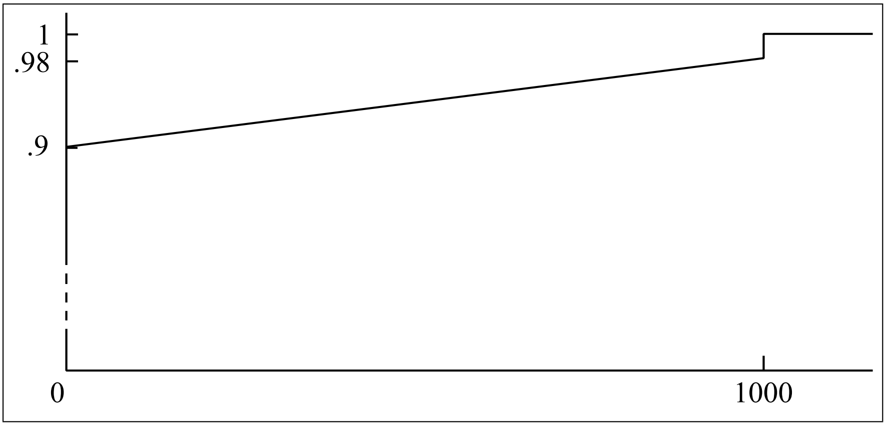

Fig. 2.2 Cumulative distribution function F of X in Example 2.2.5.

$$\operatorname{Var}[W] = \operatorname{Var}[\operatorname{E}[W \mid V]] + \operatorname{E}[\operatorname{Var}[W \mid V]]. \tag{2.22}$$

In statistics, the first term is the component of the variance of W, not explained by knowledge of V; the second is the explained component of the variance. The conditional distribution of $B \mid I = 0$ is irrelevant, so for convenience, we let it be equal to the one of $B \mid I = 1$, meaning that we take I and B to be independent. Then, letting $q = \Pr[I = 1]$, $\mu = \operatorname{E}[B]$ and $\sigma^2 = \operatorname{Var}[B]$, we have $\operatorname{E}[X \mid I = 1] = \mu$ and $\operatorname{E}[X \mid I = 0] = 0$. Therefore, $\operatorname{E}[X \mid I = i] = \mu i$ for both values $i = 0, 1$, and analogously, $\operatorname{Var}[X \mid I = i] = \sigma^2 i$. Hence,

$$\operatorname{E}[X \mid I] \equiv \mu I \quad \text{and} \quad \operatorname{Var}[X \mid I] \equiv \sigma^2 I, \tag{2.23}$$

from which it follows that

$$\begin{aligned} \operatorname{E}[X] &= \operatorname{E}[\operatorname{E}[X \mid I]] = \operatorname{E}[\mu I] = \mu q; \\ \operatorname{Var}[X] &= \operatorname{Var}[\operatorname{E}[X \mid I]] + \operatorname{E}[\operatorname{Var}[X \mid I]] = \operatorname{Var}[\mu I] + \operatorname{E}[\sigma^2 I] \\ &= \mu^2 q(1-q) + \sigma^2 q. \end{aligned} \tag{2.24}$$

Notice that a continuous cdf F is not necessarily absolutely continuous in the sense of (2.3), as is demonstrated by the following example.

Example 2.2.6 ([♠] The Cantor cdf; continuous but not absolutely continuous)
Let $X_1, X_2, \ldots$ be an infinite sequence of independent Bernoulli(1/2) random variables. Define the following random variable:

$$W = \sum_{i=1}^{\infty} \frac{2X_i}{3^i} = \frac{2}{3}X_1 + \frac{1}{3}\sum_{i=1}^{\infty} \frac{2X_{i+1}}{3^i} \tag{2.25}$$

Then the possible values of W are, in the ternary system, $0.d_1 d_2 d_3 \ldots$ with $d_i \in \{0, 2\}$ for all $i = 1, 2, \ldots$, and with $d_i = 2$ occurring if $X_i = 1$. Obviously, all of these values have zero probability as they correspond to *all* X_i having specific outcomes, so F_W is continuous.

Also, all intervals of real numbers in $(0,1)$ having a ternary digit $d_i = 1$ on some place $i = 1,2,\dots,n$ are not possible values of W, hence F_W is constant on the union B_n of all those intervals. But it is easy to see that the total length of these intervals tends to 1 as $n \to \infty$.

So we have constructed a continuous cdf F_W, known as the *Cantor distribution function*, that is constant except on a set of length 0 (known as the *Cantor set*). The cdf F_W cannot be equal to the integral over its derivative, since this is zero almost everywhere with respect to the Lebesgue measure ('interval length'). So though F_W is continuous, it is not absolutely continuous as in (2.3). ∇

2.3 Convolution

In the individual risk model we are interested in the distribution of the total S of the claims on a number of policies, with

$$S = X_1 + X_2 + \cdots + X_n, \tag{2.26}$$

where X_i, $i = 1,2,\dots,n$, denotes the payment on policy i. The risks X_i are assumed to be independent random variables. If this assumption is violated for some risks, for example in case of fire insurance policies on different floors of the same building, then these risks could be combined into one term in (2.26).

The operation 'convolution' calculates the distribution function of $X+Y$ from the cdfs of two independent random variables X and Y as follows:

$$\begin{aligned} F_{X+Y}(s) &= \Pr[X+Y \le s] \\ &= \int_{-\infty}^{\infty} \Pr[X+Y \le s \,|\, X = x]\,\mathrm{d}F_X(x) \\ &= \int_{-\infty}^{\infty} \Pr[Y \le s-x \,|\, X = x]\,\mathrm{d}F_X(x) \\ &= \int_{-\infty}^{\infty} \Pr[Y \le s-x]\,\mathrm{d}F_X(x) \\ &= \int_{-\infty}^{\infty} F_Y(s-x)\,\mathrm{d}F_X(x) =: F_X * F_Y(s). \end{aligned} \tag{2.27}$$

The cdf $F_X * F_Y(\cdot)$ is called the convolution of the cdfs $F_X(\cdot)$ and $F_Y(\cdot)$. For the density function we use the same notation. If X and Y are discrete random variables, we find for the cdf of $X+Y$ and the corresponding density

$$F_X * F_Y(s) = \sum_x F_Y(s-x)\,f_X(x) \quad \text{and} \quad f_X * f_Y(s) = \sum_x f_Y(s-x)\,f_X(x), \tag{2.28}$$

where the sum is taken over all x with $f_X(x) > 0$. If X and Y are continuous random variables, then

$$F_X * F_Y(s) = \int_{-\infty}^{\infty} F_Y(s-x)\, f_X(x)\, \mathrm{d}x \tag{2.29}$$

and, taking the derivative under the integral sign to find the density,

$$f_X * f_Y(s) = \int_{-\infty}^{\infty} f_Y(s-x)\, f_X(x)\, \mathrm{d}x. \tag{2.30}$$

Since $X+Y \equiv Y+X$, the convolution operator $*$ is *commutative*: $F_X * F_Y$ is identical to $F_Y * F_X$. Also, it is *associative*, since for the cdf of $X+Y+Z$, it does not matter in which order we do the convolutions, therefore

$$(F_X * F_Y) * F_Z \equiv F_X * (F_Y * F_Z) \equiv F_X * F_Y * F_Z. \tag{2.31}$$

For the sum of n independent and identically distributed random variables with marginal cdf F, the cdf is the n-fold convolution power of F, which we write as

$$F * F * \cdots * F =: F^{*n}. \tag{2.32}$$

Example 2.3.1 (Convolution of two uniform distributions)
Suppose that $X \sim$ uniform(0,1) and $Y \sim$ uniform(0,2) are independent. What is the cdf of $X+Y$?

The indicator function of a set A is defined as follows:

$$I_A(x) = \begin{cases} 1 \text{ if } & x \in A \\ 0 \text{ if } & x \notin A. \end{cases} \tag{2.33}$$

Indicator functions provide us with a concise notation for functions that are defined differently on some intervals. For all x, the cdf of X can be written as

$$F_X(x) = x I_{[0,1)}(x) + I_{[1,\infty)}(x), \tag{2.34}$$

while $F_Y'(y) = \frac{1}{2} I_{[0,2)}(y)$ for all y, which leads to the differential

$$\mathrm{d}F_Y(y) = \tfrac{1}{2} I_{[0,2)}(y)\, \mathrm{d}y. \tag{2.35}$$

The convolution formula (2.27), applied to $Y+X$ rather than $X+Y$, then yields

$$F_{Y+X}(s) = \int_{-\infty}^{\infty} F_X(s-y)\, \mathrm{d}F_Y(y) = \int_0^2 F_X(s-y) \tfrac{1}{2}\, \mathrm{d}y, \ s \ge 0. \tag{2.36}$$

The interval of interest is $0 \le s < 3$. Subdividing it into $[0,1)$, $[1,2)$ and $[2,3)$ yields

$$\begin{aligned} F_{X+Y}(s) = & \left\{ \int_0^s (s-y)\tfrac{1}{2}\,\mathrm{d}y \right\} I_{[0,1)}(s) \\ & + \left\{ \int_0^{s-1} \tfrac{1}{2}\,\mathrm{d}y + \int_{s-1}^{s} (s-y)\tfrac{1}{2}\,\mathrm{d}y \right\} I_{[1,2)}(s) \\ & + \left\{ \int_0^{s-1} \tfrac{1}{2}\,\mathrm{d}y + \int_{s-1}^{2} (s-y)\tfrac{1}{2}\,\mathrm{d}y \right\} I_{[2,3)}(s) \\ = & \tfrac{1}{4}s^2 I_{[0,1)}(s) + \tfrac{1}{4}(2s-1)I_{[1,2)}(s) + [1-\tfrac{1}{4}(3-s)^2]I_{[2,3)}(s). \end{aligned} \tag{2.37}$$

Notice that $X+Y$ is symmetric around $s = 1.5$. Although this problem could be solved graphically by calculating the probabilities by means of areas, see Exercise 2.3.5, the above derivation provides an excellent illustration that, even in simple cases, convolution can be a laborious process. ∇

Example 2.3.2 (Convolution of discrete distributions)
Let $f_1(x) = \frac{1}{4}, \frac{1}{2}, \frac{1}{4}$ for $x = 0,1,2$, $f_2(x) = \frac{1}{2}, \frac{1}{2}$ for $x = 0,2$ and $f_3(x) = \frac{1}{4}, \frac{1}{2}, \frac{1}{4}$ for $x = 0,2,4$. Let f_{1+2} denote the convolution of f_1 and f_2 and let f_{1+2+3} denote the convolution of f_1, f_2 and f_3. To calculate F_{1+2+3}, we need to compute the values as shown in Table 2.1. In the discrete case, too, convolution is clearly a laborious exercise. Note that the more often we have $f_i(x) \neq 0$, the more calculations need to be done. ∇

Table 2.1 Convolution computations for Example 2.3.2

x	$f_1(x)$	$*$ $f_2(x)$	$=$ $f_{1+2}(x)$	$*$ $f_3(x)$	$=$ $f_{1+2+3}(x)$	$\Rightarrow$ $F_{1+2+3}(x)$
0	1/4	1/2	1/8	1/4	1/32	1/32
1	1/2	0	2/8	0	2/32	3/32
2	1/4	1/2	2/8	1/2	4/32	7/32
3	0	0	2/8	0	6/32	13/32
4	0	0	1/8	1/4	6/32	19/32
5	0	0	0	0	6/32	25/32
6	0	0	0	0	4/32	29/32
7	0	0	0	0	2/32	31/32
8	0	0	0	0	1/32	32/32

Example 2.3.3 (Convolution of iid uniform distributions)
Let X_i, $i = 1,2,\ldots,n$, be independent and identically uniform$(0,1)$ distributed. By using the convolution formula and induction, it can be shown that for all $x > 0$, the pdf of $S = X_1 + \cdots + X_n$ equals

$$f_S(x) = \frac{1}{(n-1)!}\sum_{h=0}^{[x]} \binom{n}{h} (-1)^h (x-h)^{n-1} \tag{2.38}$$

where $[x]$ denotes the integer part of x. See also Exercise 2.3.4. ∇

Example 2.3.4 (Convolution of Poisson distributions)
Let $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\mu)$ be independent random variables. From (2.28) we have, for $s = 0, 1, 2, \ldots$,

$$\begin{aligned} f_{X+Y}(s) &= \sum_{x=0}^{s} f_Y(s-x) f_X(x) = \frac{e^{-\mu-\lambda}}{s!} \sum_{x=0}^{s} \binom{s}{x} \mu^{s-x} \lambda^x \\ &= e^{-(\lambda+\mu)} \frac{(\lambda+\mu)^s}{s!}, \end{aligned} \tag{2.39}$$

where the last equality is the binomial theorem. Hence, $X+Y$ is $\text{Poisson}(\lambda+\mu)$ distributed. For a different proof, see Exercise 2.4.2. ∇

2.4 Transforms

Determining the distribution of the sum of independent random variables can often be made easier by using transforms of the cdf. The *moment generating function (mgf)* suits our purposes best. For a non-negative random variable X, it is defined as

$$\mathrm{m}_X(t) = \mathrm{E}\left[e^{tX}\right], \quad -\infty < t < h, \tag{2.40}$$

for some h. The mgf is going to be used especially in an interval around 0, which requires $h > 0$ to hold. Note that this is the case only for light-tailed risks, of which exponential moments $\mathrm{E}[e^{\varepsilon x}]$ for some $\varepsilon > 0$ exist.

If X and Y are independent, then

$$\mathrm{m}_{X+Y}(t) = \mathrm{E}\left[e^{t(X+Y)}\right] = \mathrm{E}\left[e^{tX}\right]\mathrm{E}\left[e^{tY}\right] = \mathrm{m}_X(t)\mathrm{m}_Y(t). \tag{2.41}$$

So, the convolution of cdfs corresponds to simply multiplying the mgfs. Note that the mgf-transform is one-to-one, so every cdf has exactly one mgf. Also, it is continuous, in the sense that the mgf of the limit of a series of cdfs is the limit of the mgfs. See Exercises 2.4.12 and 2.4.13.

For random variables with a heavy tail, such as the Pareto distributions, the mgf does not exist. The *characteristic function*, however, always exists. It is defined as:

$$\phi_X(t) = \mathrm{E}\left[e^{\mathrm{i}tX}\right] = \mathrm{E}\left[\cos(tX) + \mathrm{i}\sin(tX)\right], \quad -\infty < t < \infty. \tag{2.42}$$

A disadvantage of the characteristic function is the need to work with complex numbers, although applying the same function formula derived for real t to imaginary t as well produces the correct results most of the time, resulting for example in the $\mathrm{N}(0,2)$ distribution with mgf $\exp(t^2)$ having $\exp((\mathrm{i}t)^2) = \exp(-t^2)$ as its characteristic function.

As their name indicates, moment generating functions can be used to generate moments of random variables. The usual series expansion of e^x yields

$$m_X(t) = E[e^{tX}] = \sum_{k=0}^{\infty} \frac{E[X^k]t^k}{k!}, \tag{2.43}$$

so the k-th moment of X equals

$$E[X^k] = \left.\frac{d^k}{dt^k} m_X(t)\right|_{t=0}. \tag{2.44}$$

Moments can also be generated from the characteristic function in similar fashion.

The *probability generating function (pgf)* is reserved for random variables with natural numbers as values:

$$g_X(t) = E[t^X] = \sum_{k=0}^{\infty} t^k \Pr[X = k]. \tag{2.45}$$

So, the probabilities $\Pr[X = k]$ in (2.45) are just the coefficients in the series expansion of the pgf. The series (2.45) converges absolutely if $|t| \leq 1$.

The *cumulant generating function (*cgf*)* is convenient for calculating the third central moment; it is defined as:

$$\kappa_X(t) = \log m_X(t). \tag{2.46}$$

Differentiating (2.46) three times and setting $t = 0$, one sees that the coefficients of $t^k/k!$ for $k = 1,2,3$ are $E[X]$, $\mathrm{Var}[X]$ and $E[(X - E[X])^3]$. The quantities generated this way are the *cumulants* of X, and they are denoted by κ_k, $k = 1,2,\ldots$ One may also proceed as follows: let μ_k denote $E[X^k]$ and let, as usual, the 'big O notation' $O(t^k)$ denote 'terms of order t to the power k or higher'. Then

$$m_X(t) = 1 + \mu_1 t + \tfrac{1}{2}\mu_2 t^2 + \tfrac{1}{6}\mu_3 t^3 + O(t^4), \tag{2.47}$$

which, using $\log(1+z) = z - \frac{1}{2}z^2 + \frac{1}{3}z^3 + O(z^4)$, yields

$$\begin{aligned}
\log m_X(t) &= \log\left(1 + \mu_1 t + \tfrac{1}{2}\mu_2 t^2 + \tfrac{1}{6}\mu_3 t^3 + O(t^4)\right) \\
&= \mu_1 t + \tfrac{1}{2}\mu_2 t^2 + \tfrac{1}{6}\mu_3 t^3 + O(t^4) \\
&\quad - \tfrac{1}{2}\left\{\mu_1^2 t^2 + \mu_1\mu_2 t^3 + O(t^4)\right\} \\
&\quad + \tfrac{1}{3}\left\{\mu_1^3 t^3 + O(t^4)\right\} + O(t^4) \\
&= \mu_1 t + \tfrac{1}{2}(\mu_2 - \mu_1^2)t^2 + \tfrac{1}{6}(\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3)t^3 + O(t^4) \\
&= E[X]t + \mathrm{Var}[X]\,\tfrac{1}{2}t^2 + E[(X - E[X])^3]\,\tfrac{1}{6}t^3 + O(t^4).
\end{aligned} \tag{2.48}$$

The *skewness* of a random variable X is defined as the following dimension-free quantity:

$$\gamma_X = \frac{\kappa_3}{\sigma^3} = \frac{E[(X-\mu)^3]}{\sigma^3}, \tag{2.49}$$

with $\mu = \mathrm{E}[X]$ and $\sigma^2 = \mathrm{Var}[X]$. If $\gamma_X > 0$, large values of $X - \mu$ are likely to occur, hence the (right) tail of the cdf is heavy. A negative skewness $\gamma_X < 0$ indicates a heavy left tail. If X is symmetric then $\gamma_X = 0$, but having zero skewness is not sufficient for symmetry. For some counterexamples, see the exercises.

The cumulant generating function, the probability generating function, the characteristic function and the moment generating function are related by

$$\kappa_X(t) = \log \mathrm{m}_X(t); \quad g_X(t) = \mathrm{m}_X(\log t); \quad \phi_X(t) = \mathrm{m}_X(\mathrm{i}t). \tag{2.50}$$

In Exercise 2.4.14 the reader is asked to examine the last of these equalities. Often the mgf can be extended to the whole complex plane in a natural way. The mgf operates on the real axis, the characteristic function on the imaginary axis.

2.5 Approximations

A well-known method to approximate a cdf is based on the Central Limit Theorem (CLT). We study this approximation as well as two more accurate ones that involve three moments rather than just two.

2.5.1 Normal approximation

Next to the Law of Large Numbers, the Central Limit Theorem is the most important theorem in statistics. It states that by adding up a large number of independent random variables, we get a normally distributed random variable in the limit. In its simplest form, the Central Limit Theorem (CLT) is as follows:

Theorem 2.5.1 (Central Limit Theorem)
If $X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables with mean μ and variance $\sigma^2 < \infty$, then

$$\lim_{n\to\infty} \Pr\left[\sum_{i=1}^{n} X_i \leq n\mu + x\sigma\sqrt{n}\right] = \Phi(x). \tag{2.51}$$

Proof. We restrict ourselves to proving the convergence of the sequence of cgfs. Let $S^* = (X_1 + \cdots + X_n - n\mu)/\sigma\sqrt{n}$, then for $n \to \infty$ and for all t:

$$\begin{aligned}
\log \mathrm{m}_{S^*}(t) &= -\frac{\sqrt{n}\mu t}{\sigma} + n\left\{\log \mathrm{m}_X\left(\frac{t}{\sigma\sqrt{n}}\right)\right\} \\
&= -\frac{\sqrt{n}\mu t}{\sigma} + n\left\{\mu\left(\frac{t}{\sigma\sqrt{n}}\right) + \tfrac{1}{2}\sigma^2\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + O\left(\left(\frac{1}{\sqrt{n}}\right)^3\right)\right\} \\
&= \tfrac{1}{2}t^2 + O\left(\frac{1}{\sqrt{n}}\right),
\end{aligned} \tag{2.52}$$

which converges to the cgf of the $\mathrm{N}(0,1)$ distribution, with mgf $\exp(\frac{1}{2}t^2)$. As a consequence, the cdf of S^* converges to the standard normal cdf Φ. ∇

As a result, if the summands are *independent* and have *finite* variance, we can approximate the cdf of $S = X_1 + \cdots + X_n$ by

$$F_S(s) \approx \Phi\left(s; \sum_{i=1}^{n} \mathrm{E}[X_i], \sum_{i=1}^{n} \mathrm{Var}[X_i]\right). \tag{2.53}$$

This approximation can safely be used if n is 'large'. But it is difficult to define 'large', as is shown in the following examples.

Example 2.5.2 (Generating approximately normal random deviates fast)
If pseudo-random numbers can be generated fast (using bit-manipulations), but computing logarithms and the inverse normal cdf takes a lot of time, approximately $\mathrm{N}(0,1)$ distributed pseudo-random drawings numbers can conveniently be produced by adding up twelve uniform$(0,1)$ numbers and subtracting 6 from their sum. This technique is based on the CLT with $n = 12$. Comparing this cdf with the normal cdf, using (2.38), yields a maximum difference of 0.002. Hence, the CLT performs quite well in this case. See also Exercise 2.4.5. ∇

Example 2.5.3 (Illustrating the various approximations)
Suppose that $n = 1000$ young men take out a life insurance policy for a period of one year. The probability of dying within this year is 0.001 for everyone and the payment for every death is 1. We want to calculate the probability that the total payment is at least 4. This total payment is binomial$(1000, 0.001)$ distributed and since $n = 1000$ is large and $p = 0.001$ is small, we will approximate this probability by a Poisson(np) distribution. Calculating the probability at $3 + \frac{1}{2}$ instead of at 4, applying a continuity correction needed later on, we find

$$\Pr[S \geq 3.5] = 1 - \mathrm{e}^{-1} - \mathrm{e}^{-1} - \tfrac{1}{2}\mathrm{e}^{-1} - \tfrac{1}{6}\mathrm{e}^{-1} = 0.01899. \tag{2.54}$$

Note that the exact binomial probability is 0.01893. Although n is much larger than in the previous example, the CLT gives a poor approximation: with $\mu = \mathrm{E}[S] = 1$ and $\sigma^2 = \mathrm{Var}[S] = 1$, we find

$$\Pr[S \geq 3.5] = \Pr\left[\frac{S-\mu}{\sigma} \geq \frac{3.5-\mu}{\sigma}\right] \approx 1 - \Phi(2.5) = 0.0062. \tag{2.55}$$

The CLT approximation is not very good because of the extreme skewness of the terms X_i and the resulting skewness of S, which is $\gamma_S = 1$. In the previous example, we started from symmetric terms, leading to a higher order of convergence, as can be seen from derivation (2.52). ∇

As an alternative for the CLT, we give two more refined approximations: the translated gamma approximation and the normal power approximation (NP). In numerical examples, they turn out to be much more accurate than the CLT approximation. As regards the quality of the approximations, there is not much to choose between the two. Their inaccuracies are minor compared with the errors that result from the lack of precision in the estimates of the first three moments that are involved.

2.5.2 Translated gamma approximation

Most total claim distributions are skewed to the right (skewness $\gamma > 0$), have a non-negative support and are unimodal. So they have roughly the shape of a gamma distribution. To gain more flexibility, apart from the usual parameters α and β we allow a shift over a distance x_0. Hence, we approximate the cdf of S by the cdf of $Z + x_0$, where $Z \sim \text{gamma}(\alpha, \beta)$ (see Table A). We choose α, β and x_0 in such a way that the approximating random variable has the same first three moments as S.

The translated gamma approximation can then be formulated as follows:

$$F_S(s) \approx G(s - x_0; \alpha, \beta),$$
$$\text{where } G(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)} \int_0^x y^{\alpha-1} \beta^\alpha e^{-\beta y} dy, \quad x \geq 0. \tag{2.56}$$

Here $G(x; \alpha, \beta)$ is the gamma cdf. We choose α, β and x_0 such that the first three moments are the same, hence $\mu = x_0 + \frac{\alpha}{\beta}$, $\sigma^2 = \frac{\alpha}{\beta^2}$ and $\gamma = \frac{2}{\sqrt{\alpha}}$ (see Table A), so

$$\alpha = \frac{4}{\gamma^2}, \quad \beta = \frac{2}{\gamma\sigma} \quad \text{and} \quad x_0 = \mu - \frac{2\sigma}{\gamma}. \tag{2.57}$$

It is required that the skewness γ is strictly positive. In the limit $\gamma \downarrow 0$, the normal approximation appears. Note that if the first three moments of the cdf $F(\cdot)$ are equal to those of $G(\cdot)$, by partial integration it can be shown that the same holds for $\int_0^\infty x^j [1 - F(x)]\, dx$, $j = 0, 1, 2$. This leaves little room for these cdfs to be very different from each other.

Example 2.5.4 (Illustrating the various approximations, continued)
If $S \sim$ Poisson(1), we have $\mu = \sigma = \gamma = 1$, and (2.57) yields $\alpha = 4$, $\beta = 2$ and $x_0 = -1$. Hence, $\Pr[S \geq 3.5] \approx 1 - G(3.5 - (-1); 4, 2) = 0.0212$. This value is much closer to the exact value than the CLT approximation. ∇

The translated gamma approximation leads to quite simple formulas to approximate the moments of a stop-loss claim $(S - d)_+$ or of the retained loss $S - (S - d)_+$. To evaluate the gamma cdf is easy in R, and in spreadsheet programs the gamma distribution is also included, although the accuracy sometimes leaves much to be desired. Note that in many applications, for example MS Excel, the parameter β should be replaced by $1/\beta$. In R, specify $\beta = 2$ by using `rate=2`, or by `scale=1/2`.

Example 2.5.5 (Translated gamma approximation)
A total claim amount S has expected value 10000, standard deviation 1000 and skewness 1. From (2.57) we have $\alpha = 4$, $\beta = 0.002$ and $x_0 = 8000$. Hence,

$$\Pr[S > 13000] \approx 1 - G(13000 - 8000; 4, 0.002) = 0.010. \tag{2.58}$$

The regular CLT approximation is much smaller: 0.0013. Using the inverse of the gamma distribution function, the value-at-risk on a 95% level is found by reversing the computation (2.58), resulting in 11875. ∇

2.5.3 NP approximation

Another approximation that uses three moments of the approximated random variable is the Normal Power approximation. It goes as follows.

If $\mathrm{E}[S] = \mu$, $\mathrm{Var}[S] = \sigma^2$ and $\gamma_S = \gamma$, then, for $s \geq 1$,

$$\Pr\left[\frac{S-\mu}{\sigma} \leq s + \frac{\gamma}{6}(s^2 - 1)\right] \approx \Phi(s) \tag{2.59}$$

or, equivalently, for $x \geq 1$,

$$\Pr\left[\frac{S-\mu}{\sigma} \leq x\right] \approx \Phi\left(\sqrt{\frac{9}{\gamma^2} + \frac{6x}{\gamma} + 1} - \frac{3}{\gamma}\right). \tag{2.60}$$

The second formula can be used to approximate the cdf of S, the first produces approximate quantiles. If $s < 1$ (or $x < 1$), the correction term is negative, which implies that the CLT gives more conservative results.

Example 2.5.6 (Illustrating the various approximations, continued)
If $S \sim$ Poisson(1), then the NP approximation yields $\Pr[S \geq 3.5] \approx 1 - \Phi(2) = 0.0228$. Again, this is a better result than the CLT approximation.

The R-calls needed to produce all the numerical values are the following:

```
x <- 3.5; mu <- 1; sig <- 1; gam <- 1; z <- (x-mu)/sig
1-pbinom(x, 1000, 0.001)                          ##  0.01892683
1-ppois(x,1)                                      ##  0.01898816
1-pnorm(z)                                        ##  0.00620967
1-pnorm(sqrt(9/gam^2 + 6*z/gam + 1) - 3/gam)  ##  0.02275013
1-pgamma(x-(mu-2*sig/gam), 4/gam^2, 2/gam/sig)##  0.02122649
```

Equations (2.53), (2.60) and (2.56)–(2.57) were used. ∇

Example 2.5.7 (Recalculating Example 2.5.5 by the NP approximation)
We apply (2.59) to determine the capital that covers loss S with probability 95%:

$$\Pr\left[\frac{S-\mu}{\sigma} \leq s + \frac{\gamma}{6}(s^2 - 1)\right] \approx \Phi(s) = 0.95 \quad \text{if } s = 1.645, \tag{2.61}$$

hence for the desired 95% quantile of S we find

$$\mathrm{E}[S] + \sigma_S\left(1.645 + \frac{\gamma}{6}(1.645^2 - 1)\right) = \mathrm{E}[S] + 1.929\sigma_S = 11929. \tag{2.62}$$

To determine the probability that capital 13000 will be insufficient to cover the losses S, we apply (2.60) with $\mu = 10000$, $\sigma = 1000$ and $\gamma = 1$:

$$\begin{aligned}\Pr[S > 13000] &= \Pr\left[\frac{S-\mu}{\sigma} > 3\right] \approx 1 - \Phi(\sqrt{9+6\times 3+1} - 3)\\ &= 1 - \Phi(2.29) = 0.011.\end{aligned} \tag{2.63}$$

Note that the translated gamma approximation gave 0.010, against only 0.0013 for the CLT. ∇

Remark 2.5.8 (Justifying the NP approximation)
For $U \sim \mathrm{N}(0,1)$ consider the random variable $Y = U + \frac{\gamma}{6}(U^2 - 1)$. It is easy to verify that (see Exercise 2.5.21), writing $w(x) = \sqrt{\left(\frac{9}{\gamma^2} + \frac{6x}{\gamma} + 1\right)_+}$, we have

$$F_Y(x) = \Phi\Big(+w(x) - \frac{3}{\gamma}\Big) - \Phi\Big(-w(x) - \frac{3}{\gamma}\Big) \approx \Phi\Big(w(x) - \frac{3}{\gamma}\Big). \tag{2.64}$$

The term $\Phi(-w(x) - 3/\gamma)$ accounts for small U leading to large Y. It is generally negligible, and vanishes as $\gamma \downarrow 0$.

Also, using $\mathrm{E}[U^6] = 15$, $\mathrm{E}[U^4] = 3$ and $\mathrm{E}[U^2] = 1$, for small γ one can prove

$$\mathrm{E}[Y] = 0; \quad \mathrm{E}[Y^2] = 1 + O(\gamma^2); \quad \mathrm{E}[Y^3] = \gamma\big(1 + O(\gamma^2)\big). \tag{2.65}$$

Therefore, the first three moments of $\frac{S-\mu}{\sigma}$ and Y as defined above are alike. This, with (2.64), justifies the use of formula (2.60) to approximate the cdf of $\frac{S-\mu}{\sigma}$. ∇

Remark 2.5.9 ([♠] Deriving NP using the Edgeworth expansion)
Formula (2.59) can be derived by the use of a certain expansion for the cdf, though not in a mathematically rigorous way. Define $Z = (S - \mathrm{E}[S])/\sqrt{\mathrm{Var}[S]}$, and let $\gamma = \mathrm{E}[Z^3]$ be the skewness of S (and Z). For the cgf of Z we have

$$\log \mathrm{m}_Z(t) = \tfrac{1}{2}t^2 + \tfrac{1}{6}\gamma t^3 + \ldots, \tag{2.66}$$

hence

$$\mathrm{m}_Z(t) = \mathrm{e}^{t^2/2} \cdot \exp\left\{\tfrac{1}{6}\gamma t^3 + \ldots\right\} = \mathrm{e}^{t^2/2} \cdot \Big(1 + \tfrac{1}{6}\gamma t^3 + \ldots\Big). \tag{2.67}$$

The 'mgf' (generalized to functions that are not a density) of $\varphi^{(3)}(x)$, with $\varphi(x)$ the $\mathrm{N}(0,1)$ density, can be found by partial integration:

$$\begin{aligned}\int_{-\infty}^{\infty} \mathrm{e}^{tx}\varphi^{(3)}(x)\,\mathrm{d}x &= \mathrm{e}^{tx}\varphi^{(2)}(x)\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} t\mathrm{e}^{tx}\varphi^{(2)}(x)\,\mathrm{d}x\\ &= 0 - 0 + \int_{-\infty}^{\infty} t^2\mathrm{e}^{tx}\varphi^{(1)}(x)\,\mathrm{d}x\\ &= 0 - 0 + 0 - \int_{-\infty}^{\infty} t^3\mathrm{e}^{tx}\varphi(x)\,\mathrm{d}x = -t^3\mathrm{e}^{t^2/2}.\end{aligned} \tag{2.68}$$

Therefore we recognize the cdf corresponding to mgf (2.67) as:

$$F_Z(x) = \Phi(x) - \tfrac{1}{6}\gamma\Phi^{(3)}(x) + \dots \tag{2.69}$$

Formula (2.69) is called the *Edgeworth expansion* for F_Z; leaving out the dots gives an *Edgeworth approximation* for it. There is no guarantee that the latter is an increasing function. To derive the NP approximation formula (2.59) from it, we try to find a correction $\delta = \delta(s)$ to the argument s such that

$$F_Z(s+\delta) \approx \Phi(s). \tag{2.70}$$

That means that we have to find a zero for the auxiliary function $g(\delta)$ defined by

$$g(\delta) = \Phi(s) - \left\{\Phi(s+\delta) - \tfrac{1}{6}\gamma\Phi^{(3)}(s+\delta)\right\}. \tag{2.71}$$

Using a Taylor expansion $g(\delta) \approx g(0) + \delta g'(0)$ we may conclude that $g(\delta) = 0$ for $\delta \approx -g(0)/g'(0)$, so

$$\delta \approx \frac{-\frac{1}{6}\gamma\Phi^{(3)}(s)}{-\Phi'(s) + \frac{1}{6}\gamma\Phi^{(4)}(s)} = \frac{-\frac{1}{6}\gamma(s^2-1)\varphi(s)}{\left(-1 + \frac{1}{6}\gamma(-s^3+3s)\right)\varphi(s)}. \tag{2.72}$$

Since the skewness γ is of order $\lambda^{-1/2}$, see for example (2.48), therefore small for large portfolios, we drop the term with γ in the denominator of (2.72), leading to

$$F_Z(s+\delta) \approx \Phi(s) \qquad \text{when} \qquad \delta = \tfrac{1}{6}\gamma(s^2-1). \tag{2.73}$$

This is precisely the NP approximation (2.59) given earlier.

The dots in formula (2.69) denote the inverse mgf-transform of the dots in (2.67). It is not possible to show that the terms replaced by dots in this formula are small, let alone their absolute sum. So it is an exaggeration to say that the approximations obtained this way, dropping terms of a possibly divergent series and then using an approximate inversion, are justified by theoretical arguments. ∇

2.6 Application: optimal reinsurance

An insurer is looking for an optimal reinsurance for a portfolio consisting of 20 000 one-year life insurance policies that are grouped as follows:

Insured amount b_k	Number of policies n_k
1	10 000
2	5 000
3	5 000

The probability of dying within one year is $q_k = 0.01$ for each insured, and the policies are independent. The insurer wants to optimize the probability of being able to meet his financial obligations by choosing the best retention, which is the maximum payment per policy. The remaining part of a claim is paid by the reinsurer. For example, if the retention is 1.6 and someone with insured amount 2 dies, then the insurer pays 1.6, the reinsurer pays 0.4. After collecting the premiums, the insurer holds a capital B from which he has to pay the claims and the reinsurance premium. This premium is assumed to be 120% of the net premium.

First, we set the retention equal to 2. From the point of view of the insurer, the policies are then distributed as follows:

Insured amount b_k	Number of policies n_k
1	10 000
2	10 000

The expected value and the variance of the insurer's total claim amount S are equal to

$$\begin{aligned} \mathrm{E}[S] &= n_1 b_1 q_1 + n_2 b_2 q_2 \\ &= 10\,000 \times 1 \times 0.01 + 10\,000 \times 2 \times 0.01 = 300, \\ \mathrm{Var}[S] &= n_1 b_1^2 q_1 (1 - q_1) + n_2 b_2^2 q_2 (1 - q_2) \\ &= 10\,000 \times 1 \times 0.01 \times 0.99 + 10\,000 \times 4 \times 0.01 \times 0.99 = 495. \end{aligned} \tag{2.74}$$

By applying the CLT, we get for the probability that the costs S plus the reinsurance premium $1.2 \times 0.01 \times 5\,000 \times 1 = 60$ exceed the available capital B:

$$\Pr[S + 60 > B] = \Pr\left[\frac{S - \mathrm{E}[S]}{\sigma_S} > \frac{B - 360}{\sqrt{495}}\right] \approx 1 - \Phi\left(\frac{B - 360}{\sqrt{495}}\right). \tag{2.75}$$

We leave it to the reader to determine this same probability for retentions between 2 and 3, as well as to determine which retention for a given B leads to the largest probability of survival. See the exercises with this section.

2.7 Exercises

Section 2.2

1. Determine the expected value and the variance of $X = IB$ if the claim probability equals 0.1. First, assume that B equals 5 with probability 1. Then, let $B \sim \text{uniform}(0, 10)$.
2. Throw a true die and let X denote the outcome. Then, toss a coin X times. Let Y denote the number of heads obtained. What are the expected value and the variance of Y?

3. In Example 2.2.4, plot the cdf of X. Also determine, with the help of the obtained differential, the premium the insured is willing to pay for being insured against an inflated loss $1.1X$. Do the same by writing $X = IB$. Has the zero utility premium followed inflation exactly?
4. Calculate $\mathrm{E}[X]$, $\mathrm{Var}[X]$ and the moment generating function $m_X(t)$ in Example 2.2.5 with the help of the differential. Also plot the 'density'.
5. If $X = IB$, what is $m_X(t)$?
6. Consider the following cdf F: $F(x) = \begin{cases} 0 & \text{for } x < 2, \\ \frac{x}{4} & \text{for } 2 \le x < 4, \\ 1 & \text{for } 4 \le x. \end{cases}$
   Determine independent random variables I, X and Y such that $Z = IX + (1-I)Y$ has cdf F, $I \sim$ Bernoulli, X is a discrete and Y a continuous random variable.
7. The differential of cdf F is $\mathrm{d}F(x) = \begin{cases} \mathrm{d}x/3 & \text{for } 0 < x < 1 \text{ and } 2 < x < 3, \\ \frac{1}{6} & \text{for } x \in \{1,2\}, \\ 0 & \text{elsewhere.} \end{cases}$
   Find a discrete cdf G, a continuous cdf H and a real constant c with the property that $F(x) = cG(x) + (1-c)H(x)$ for all x.
8. Suppose that $T = qX + (1-q)Y$ and $Z = IX + (1-I)Y$ with $I \sim$ Bernoulli(q) and I, X and Y independent. Compare $\mathrm{E}[T^k]$ with $\mathrm{E}[Z^k]$, $k = 1,2$.
9. In the previous exercise, assume additionally that X and Y are independent $\mathrm{N}(0,1)$. What distributions do T and Z have?
10. [♠] In Example 2.2.6, show that $\mathrm{E}[W] = \frac{1}{2}$ and $\mathrm{Var}[W] = \frac{1}{8}$.
    Also show that $m_W(t) = e^{t/2} \prod_{i=1}^{\infty} \cosh(t/3^i)$. Recall that $\cosh(t) = (e^t + e^{-t})/2$.

Section 2.3

1. Calculate $\Pr[S = s]$ for $s = 0,1,\ldots,6$ when $S = X_1 + 2X_2 + 3X_3$ and $X_j \sim$ Poisson(j).
2. Determine the number of multiplications of non-zero numbers that are needed for the calculation of all probabilities $f_{1+2+3}(x)$ in Example 2.3.2. How many multiplications are needed to calculate $F_{1+\cdots+n}(x)$, $x = 0,\ldots,4n-4$ if $f_k = f_3$ for $k = 4,\ldots,n$?
3. Prove by convolution that the sum of two independent normal random variables, see Table A, has a normal distribution.
4. [♠] Verify the expression (2.38) in Example 2.3.3 for $n = 1,2,3$ by using convolution. Determine $F_S(x)$ for these values of n. Using induction, verify (2.38) for arbitrary n.
5. Assume that $X \sim$ uniform$(0,3)$ and $Y \sim$ uniform$(-1,1)$. Calculate $F_{X+Y}(z)$ graphically by using the area of the sets $\{(x,y) \,|\, x+y \le z, x \in (0,3) \text{ and } y \in (-1,1)\}$.

Section 2.4

1. Determine the cdf of $S = X_1 + X_2$ where the X_k are independent and exponential(k) distributed. Do this both by convolution and by calculating the mgf and identifying the corresponding density using the method of partial fractions.
2. Same as Example 2.3.4, but now by making use of the mgfs.
3. What is the fourth cumulant κ_4 in terms of the central moments?

4. Prove that cumulants actually cumulate in the following sense: if X and Y are independent, then the kth cumulant of $X+Y$ equals the sum of the kth cumulants of X and Y.
5. Prove that the sum of twelve independent uniform(0,1) random variables has variance 1 and expected value 6. Determine κ_3 and κ_4.
   Plot the difference between the cdf of this random variable and the $\mathrm{N}(6,1)$ cdf, using the expression for $F_S(x)$ found in Exercise 2.3.4.
6. Determine the skewness of a Poisson(μ) distribution.
7. Determine the skewness of a gamma(α,β) distribution.
8. If S is symmetric, then $\gamma_S = 0$. Prove this, but also, for $S = X_1 + X_2 + X_3$ with $X_1 \sim$ Bernoulli(0.4), $X_2 \sim$ Bernoulli(0.7) and $X_3 \sim$ Bernoulli(p), all independent, calculate the value of p such that S has skewness $\gamma_S = 0$, and verify that S is not symmetric.
9. Determine the skewness of a risk of the form Ib where $I \sim$ Bernoulli(q) and b is a fixed amount. For which values of q and b is the skewness equal to zero, and for which of these values is I actually symmetric?
10. Determine the pgf of the binomial, the Poisson and the negative binomial distribution, see Table A.
11. Determine the cgf and the cumulants of the following distributions: Poisson, binomial, normal and gamma.
12. Show that X and Y are equal in distribution if they have the same support $\{0,1,\dots,n\}$ and the same pgf. If $X_1, X_2, \dots$ are risks, again with range $\{0,1,\dots,n\}$, such that the pgfs of X_i converge to the pgf of Y for each argument t when $i \to \infty$, verify that also $\Pr[X_i = x] \to \Pr[Y = x]$ for all x.
13. Show that X and Y are equal in distribution if they have the same support $\{0,\delta,2\delta,\dots,n\delta\}$ for some $\delta > 0$ and moreover, they have the same mgf.
14. Examine the equality $\phi_X(t) = m_X(it)$ from (2.50), for the special case that $X \sim \text{exponential}(1)$. Show that the characteristic function is real-valued if X is symmetric around 0.
15. Show that the skewness of $Z = X + 2Y$ is 0 if $X \sim \text{binomial}(8,p)$ and $Y \sim \text{Bernoulli}(1-p)$. For which values of p is Z symmetric?
16. For which values of δ is the skewness of $X - \delta Y$ equal to 0, if $X \sim \text{gamma}(2,1)$ and $Y \sim \text{exponential}(1)$?
17. Can the pgf of a random variable be used to generate moments? Can the mgf of an integer-valued random variable be used to generate probabilities?

Section 2.5

1. What happens if we replace the argument 3.5 in Example 2.5.3 by $3-0$, $3+0$, $4-0$ and $4+0$? Is a correction for continuity needed here?
2. Prove that both versions of the NP approximation are equivalent.
3. If $Y \sim \text{gamma}(\alpha,\beta)$ and $\gamma_Y = \frac{2}{\sqrt{\alpha}} \le 4$, then $\sqrt{4\beta Y} - \sqrt{4\alpha - 1} \overset{\approx}{\sim} \mathrm{N}(0,1)$. See ex. 2.5.14 for a comparison of the first four moments. So approximating a translated gamma approximation with parameters α, β and x_0, we also have $\Pr[S \le s] \approx \Phi\big(\sqrt{4\beta(s-x_0)} - \sqrt{4\alpha-1}\big)$.
   Show $\Pr[S \le s] \approx \Phi\left(\sqrt{\frac{8}{\gamma}\frac{s-\mu}{\sigma} + \frac{16}{\gamma^2}} - \sqrt{\frac{16}{\gamma^2} - 1}\right)$ if $\alpha = \frac{4}{\gamma^2}, \beta = \frac{2}{\gamma\sigma}, x_0 = \mu - \frac{2\sigma}{\gamma}$.
   Inversely, show $\Pr\left[S \le x_0 + \frac{1}{4\beta}(y + \sqrt{4\alpha-1})^2\right] \approx 1 - \varepsilon$ if $\Phi(y) = 1 - \varepsilon$,
   as well as $\Pr\left[\frac{S-\mu}{\sigma} \le y + \frac{\gamma}{8}(y^2-1) + y(\sqrt{1-\gamma^2/16} - 1)\right] \approx \Phi(y)$.

4. Show that the translated gamma approximation as well as the NP approximation result in the normal approximation (CLT) if μ and σ^2 are fixed and $\gamma \downarrow 0$.
5. Approximate the critical values of a χ^2_{18} distribution for $\varepsilon = 0.05, 0.1, 0.5, 0.9, 0.95$ with the NP approximation and compare the results with the exact values.
6. In the previous exercise, what is the result if the translated gamma approximation is used?
7. Use the identity 'having to wait longer than x for the nth event' $\equiv$ 'at most $n-1$ events occur in $(0,x)$' in a Poisson process to prove that $\Pr[Z > x] = \Pr[N < n]$ if $Z \sim$ gamma(n,1) and $N \sim$ Poisson(x). How can this fact be used to calculate the translated gamma approximation?
8. Compare the exact critical values of a χ^2_{18} distribution for $\varepsilon = 0.05, 0.1, 0.5, 0.9, 0.95$ with the approximations obtained in exercise 2.5.3.
9. An insurer's portfolio contains 2 000 one-year life insurance policies. Half of them are characterized by a payment $b_1 = 1$ and a probability of dying within 1 year of $q_1 = 1\%$. For the other half, we have $b_2 = 2$ and $q_2 = 5\%$. Use the CLT to determine the minimum safety loading, as a percentage, to be added to the net premium to ensure that the probability that the total payment exceeds the total premium income is at most 5%.
10. As the previous exercise, but now using the NP approximation. Employ the fact that the third cumulant of the total payment equals the sum of the third cumulants of the risks.
11. Show that the right hand side of (2.60) is well-defined for all $x \geq -1$. What are the minimum and the maximum values? Is the function increasing? What happens if $x = 1$?
12. Suppose that X has expected value $\mu = 1000$ and standard deviation $\sigma = 2000$. Determine the skewness γ if (i) X is normal, (ii) $X/\phi \sim$ Poisson(μ/ϕ), (iii) $X \sim$ gamma(α, β), (iv) $X \sim$ inverse Gaussian(α, β) or (v) $X \sim$ lognormal(ν, τ^2). Show that the skewness is infinite if (vi) $X \sim$ Pareto. See also Table A.
13. A portfolio consists of two types of contracts. For type k, $k = 1,2$, the claim probability is q_k and the number of policies is n_k. If there is a claim, then its size is x with probability $p_k(x)$:

	n_k	q_k	$p_k(1)$	$p_k(2)$	$p_k(3)$
Type 1	1000	0.01	0.5	0	0.5
Type 2	2000	0.02	0.5	0.5	0

   Assume that the contracts are independent. Let S_k denote the total claim amount of the contracts of type k and let $S = S_1 + S_2$. Calculate the expected value and the variance of a contract of type k, $k = 1,2$. Then, calculate the expected value and the variance of S. Use the CLT to determine the minimum capital that covers all claims with probability 95%.
14. [♠] Let $U \sim$ gamma(α,1), $Y \sim$ N($\sqrt{4\alpha - 1}$,1) and $T = \sqrt{4U}$. Show that $\mathrm{E}[U^t] = \Gamma(\alpha + t)/\Gamma(\alpha)$, $t > 0$. Then show that $\mathrm{E}[Y^j] \approx \mathrm{E}[T^j]$, $j = 1,3$, by applying $\Gamma(\alpha + 1/2)/\Gamma(\alpha) \approx \sqrt{\alpha - 1/4}$ and $\alpha\Gamma(\alpha) = \Gamma(\alpha + 1)$. Also, show that $\mathrm{E}[Y^2] = \mathrm{E}[T^2]$ and $\mathrm{E}[Y^4] = \mathrm{E}[T^4] - 2$.
15. [♠] A justification for the 'correction for continuity', see Example 2.5.3, used to approximate cdfs of integer valued random variables by continuous ones, goes as follows. Let G be the continuous cdf of some non-negative random variable, and construct cdf H by $H(k + \varepsilon) = G(k + 0.5), k = 0,1,2,\ldots, 0 \leq \varepsilon < 1$. Using the *midpoint rule* with intervals of length 1 to approximate the right hand side of (1.33) at $d = 0$, show that the means of G and H are about equal. Conclude that if G is a continuous cdf that is a plausible candidate for approximating the discrete cdf F and has the same mean as F, by taking $F(x) := G(x + 0.5)$ one gets an approximation with the proper mean value. [Taking $F(x) = G(x)$ instead, one gets a mean that is about $\mu + 0.5$ instead of μ. Thus very roughly speaking, each tail probability of the sum approximating (1.33) will be too big by a factor $\frac{1}{2\mu}$.]
16. To get a feel for the approximation error as opposed to the error caused by errors in the estimates of μ, σ and γ needed for the NP approximation and the gamma approximation, recal-

culate Example 2.5.5 if the following parameters are changed: (i) $\mu = 10100$ (ii) $\sigma = 1020$ (iii) $\mu = 10100$ and $\sigma = 1020$ (iv) $\gamma = 1.03$. Assume that the remaining parameters are as they were in Example 2.5.5.

17. The function `pNormalPower`, when implemented carelessly, sometimes produces the value `NaN` (not a number). Why and when could that happen? Build in a test to cope with this situation more elegantly.
18. Compare the results of the translated gamma approximation with an exact Poisson(1) distribution using the calls `pTransGam(0:10,1,1,1)` and `ppois(0:10,1)`.
    To see the effect of applying a correction for continuity, compare also with the result of `pTransGam(0:10+0.5,1,1,1)`.
19. Repeat the previous exercise, but now for the Normal Power approximation.
20. Note that we have prefixed the (approximate) cdfs with `p`, as is customary in `R`. Now write quantile functions `qTransGam` and `qNormalPower`, and do some testing.
21. Prove (2.64) and (2.65).

Section 2.6

1. In the situation of Section 2.6, calculate the probability that B will be insufficient for retentions $d \in [2,3]$. Give numerical results for $d = 2$ and $d = 3$ if $B = 405$.
2. Determine the retention $d \in [2,3]$ that minimizes this probability for $B = 405$. Which retention is optimal if $B = 404$?
3. Calculate the probability that B will be insufficient if $d = 2$ by using the NP approximation.