Chapter 3
Collective risk models

Any sufficiently advanced technology is indistinguishable from magic — A.C. Clarke's third law of prediction, 1973

3.1 Introduction

In this chapter, we introduce collective risk models. Just as in Chapter 2, we calculate the distribution of the total claim amount, but now we regard the portfolio as a collective that produces a random number N of claims in a certain time period. We write

$$S = X_1 + X_2 + \cdots + X_N, \tag{3.1}$$

where X_i is the ith claim. Obviously, the total claims $S = 0$ if $N = 0$. The terms of S in (3.1) correspond to actual claims; in (2.26), there are many terms equal to zero, corresponding to the policies that do not produce a claim. We assume that the individual claims X_i are independent and identically distributed, and also that N and all X_i are independent. In the special case that N is Poisson distributed, S has a *compound Poisson distribution*. If N has a (negative) binomial distribution, then S has a *compound (negative) binomial distribution*.

In collective models, some policy information is ignored. If a portfolio contains only one policy that could generate a high claim amount, this amount will appear at most once in the individual model (2.26). In the collective model (3.1), however, it could occur several times. Moreover, in collective models we require the claim number N and the claim amounts X_i to be independent. This makes it somewhat less appropriate to model a car insurance portfolio, since for example bad weather conditions will cause a lot of small claim amounts. In practice, however, the influence of these phenomena appears to be small.

A collective risk model turns out to be both computationally efficient and rather close to reality. We give some algorithms to calculate the distribution of (3.1). An obvious but laborious method is convolution, conditioning on $N = n$ for all n. We also discuss the sparse vector algorithm. This can be used if $N \sim$ Poisson, and is based on the fact that the frequencies of the claim amounts can be proved to be independent Poisson random variables. For a larger class of distributions, we can use Panjer's recursion, which expresses the probability of $S = s$ recursively in terms

of the probabilities of $S = k$, $k = 0, 1, \ldots, s-1$. Another approach is to use the *Fast Fourier Transform* to invert the characteristic function.

We can express the moments of S in terms of those of N and X_i. With this information we can again approximate the distribution of S using the CLT if $\mathrm{E}[N]$ is large, as well as by the translated gamma approximation and the normal power approximation (NP) from the previous chapter.

Next, we look for appropriate distributions for N and X_i such that the collective model fits closely to a given individual model. It will turn out that the Poisson distribution and the negative binomial distribution are often appropriate choices for N. We will show some relevant relations between these distributions. We will also discuss some special properties of the compound Poisson distributions. Many parametric distributions are suitable to model insurance losses. We study their properties, including how to estimate the parameters by maximum likelihood and how to simulate random drawings from them.

Stop-loss insurance policies are not only in use for reinsurance treaties, but also for insuring absence due to illness, or if there is a deductible. We give a number of techniques to calculate stop-loss premiums for discrete distributions, but also for several continuous distributions. With the help of the approximations for distribution functions introduced in Chapter 2, we can also approximate stop-loss premiums.

3.2 Compound distributions

Assume that S is a compound random variable such as in (3.1), with terms X_i distributed as X. Further use the following notation:

$$\mu_k = \mathrm{E}[X^k], \qquad P(x) = \Pr[X \le x], \qquad F(s) = \Pr[S \le s]. \tag{3.2}$$

We can then calculate the expected value of S by using the conditional distribution of S, given N. First, we use the condition $N = n$ to substitute outcome n for the random variable N on the left of the conditioning bar below. Next, we use the independence of X_i and N to dispose of the condition $N = n$. This gives the following computation:

$$\begin{aligned}
\mathrm{E}[S] &= \mathrm{E}\big[\mathrm{E}[S\,|\,N]\big] = \sum_{n=0}^{\infty} \mathrm{E}[X_1 + \cdots + X_N \,|\, N = n] \Pr[N = n] \\
&= \sum_{n=0}^{\infty} \mathrm{E}[X_1 + \cdots + X_n \,|\, N = n] \Pr[N = n] \\
&= \sum_{n=0}^{\infty} \mathrm{E}[X_1 + \cdots + X_n] \Pr[N = n] \\
&= \sum_{n=0}^{\infty} n\mu_1 \Pr[N = n] = \mu_1 \mathrm{E}[N].
\end{aligned} \tag{3.3}$$

Note that the expected claim total equals expected claim number times expected claim size.

The variance can be determined with the variance decomposition rule (2.22):

$$\begin{aligned}\mathrm{Var}[S] &= \mathrm{E}\big[\mathrm{Var}[S\,|\,N]\big] + \mathrm{Var}\big[\mathrm{E}[S\,|\,N]\big] \\ &= \mathrm{E}\big[N\mathrm{Var}[X]\big] + \mathrm{Var}[N\mu_1] \\ &= \mathrm{E}[N]\mathrm{Var}[X] + \mu_1^2\mathrm{Var}[N].\end{aligned} \tag{3.4}$$

The same technique as used in (3.3) yields for the mgf:

$$\begin{aligned}\mathrm{m}_S(t) &= \mathrm{E}\big[\mathrm{E}[\mathrm{e}^{tS}\,|\,N]\big] \\ &= \sum_{n=0}^{\infty} \mathrm{E}\big[\mathrm{e}^{t(X_1+\cdots+X_N)}\,|\,N=n\big]\Pr[N=n] \\ &= \sum_{n=0}^{\infty} \mathrm{E}\big[\mathrm{e}^{t(X_1+\cdots+X_n)}\big]\Pr[N=n] \\ &= \sum_{n=0}^{\infty} \big\{\mathrm{m}_X(t)\big\}^n \Pr[N=n] = \mathrm{E}\big[(\mathrm{e}^{\log \mathrm{m}_X(t)})^N\big] \\ &= \mathrm{m}_N(\log \mathrm{m}_X(t)).\end{aligned} \tag{3.5}$$

Example 3.2.1 (A compound distribution with closed form cdf)
Let $N \sim$ geometric(p), $0 < p < 1$, and $X \sim$ exponential(1). What is the cdf of S?

Write $q = 1 - p$. First, we compute the mgf of S, and then we try to identify it. For $q\mathrm{e}^t < 1$, which means $t < -\log q$, we have

$$\mathrm{m}_N(t) = \sum_{n=0}^{\infty} \mathrm{e}^{nt} pq^n = \frac{p}{1 - q\mathrm{e}^t}. \tag{3.6}$$

Since $X \sim$ exponential(1), so $\mathrm{m}_X(t) = (1-t)^{-1}$, (3.5) yields

$$\mathrm{m}_S(t) = \mathrm{m}_N(\log \mathrm{m}_X(t)) = \frac{p}{1 - q\mathrm{m}_X(t)} = p + q\frac{p}{p-t}, \tag{3.7}$$

so the mgf of S is a mixture of the mgfs of the constant 0 and of the exponential(p) distribution. Because of the one-to-one correspondence of cdfs and mgfs, we may conclude that the cdf of S is the same mixture:

$$F(x) = p + q(1 - \mathrm{e}^{-px}) = 1 - q\mathrm{e}^{-px} \quad \text{for } x \geq 0. \tag{3.8}$$

This is a distribution with a jump of size p in 0, exponential otherwise. ∇

This example is unique in the sense that it presents the only non-trivial compound distribution with a closed form for the cdf.

3.2.1 Convolution formula for a compound cdf

The conditional distribution of S, given $N = n$, allows us to calculate F:

$$F(x) = \Pr[S \le x] = \sum_{n=0}^{\infty} \Pr[X_1 + \cdots + X_N \le x \,|\, N = n] \Pr[N = n], \tag{3.9}$$

so

$$F(x) = \sum_{n=0}^{\infty} P^{*n}(x) \Pr[N = n], \quad f(x) = \sum_{n=0}^{\infty} p^{*n}(x) \Pr[N = n]. \tag{3.10}$$

These expressions are the *convolution formulas* for a compound cdf.

Example 3.2.2 (Application of the convolution formula)
Let $\Pr[N = j-1] = j/10$ for $j = 1,2,3,4$, and let $p(1) = 0.4$, $p(2) = 0.6$. By using (3.10), $F(x)$ can be calculated as follows:

x	$p^{*0}(x)$	$p^{*1}(x)$	$p^{*2}(x)$	$p^{*3}(x)$		$f(x)$		$F(x)$
0	1					0.1000		0.1000
1		0.4				0.0800		0.1800
2		0.6	0.16			0.1680		0.3480
3			0.48	0.064		0.1696		0.5176
4			0.36	0.288		:		:
5				0.432		:		:
:				:		:		:
	$\updownarrow \times$	$+\updownarrow \times$	$+\updownarrow \times$	$+\updownarrow \times$	$=$	$\uparrow$	$\Rightarrow$	$\uparrow$
$\Pr[N = n]$	0.1	0.2	0.3	0.4				

The probabilities $\Pr[N = n]$ in the bottom row are multiplied by the numbers in a higher row. Then, the sum of these results is put in the corresponding row in the column $f(x)$. For example: $0.2 \times 0.6 + 0.3 \times 0.16 = 0.168$. ∇

Note that if we attempt convolution in case of arbitrary discrete claim sizes rather than integer-valued ones such as here, the number of possible values and the required number of computations increase exponentially.

Example 3.2.3 (Compound distributions, exponential claim amounts)
From expression (3.10) for $F(x)$, we see that it is convenient to choose the distribution of X in such a way that the n-fold convolution is easy to calculate. This is the case for the normal and the gamma distribution: the sum of n independent $\mathrm{N}(\mu, \sigma^2)$ random variables is $\mathrm{N}(n\mu, n\sigma^2)$, while the sum of n gamma(α, β) random variables is a gamma$(n\alpha, \beta)$ random variable.

Suppose the claim amounts have an exponential(1) distribution, which is the same as gamma(α, β) with $\alpha = \beta = 1$. In Poisson waiting time processes, see also Exercise 2.5.7 and Chapter 4, the probability of waiting at least a time x for the

n-th event, which is at the same time the probability that at most $n-1$ events have occurred at time x, is a Poisson(x) probability. Hence we have

$$1 - P^{*n}(x) = \int_x^\infty y^{n-1} \frac{\mathrm{e}^{-y}}{(n-1)!}\,\mathrm{d}y = \mathrm{e}^{-x} \sum_{i=0}^{n-1} \frac{x^i}{i!}. \tag{3.11}$$

This can also be proved with partial integration or by comparing the derivatives, see Exercise 3.2.7. So, for $x > 0$,

$$1 - F(x) = \sum_{n=1}^{\infty} \Pr[N = n]\, \mathrm{e}^{-x} \sum_{i=0}^{n-1} \frac{x^i}{i!}. \tag{3.12}$$

We can stop the outer summation as soon as $\Pr[N \geq n]$ is smaller than the required precision; also, two successive inner sums differ by the final term only, which implies that a single summation suffices. ∇

Computing the distribution of the total claims is much easier if the terms are integer-valued, so we will often approximate X by rounding it to the nearest multiples of some discretization width.

3.3 Distributions for the number of claims

In practice, we will not have a lot of relevant data at our disposal to choose a distribution for N. To describe 'rare events', the Poisson distribution, having only one parameter to be estimated, is always the first choice. Also, its use can be justified if the underlying process can be described as a Poisson process, see Chapter 4. It is well-known that the expected value and the variance of a Poisson(λ) distribution are both equal to λ. If $\mathrm{Var}[N]/\mathrm{E}[N] > 1$, that is, there is *overdispersion*, one may use the negative binomial distribution instead. We consider two models in which the latter distribution is derived as a generalization of a Poisson distribution.

Example 3.3.1 (Poisson distribution, uncertainty about the parameter)
Assume that some car driver causes a Poisson(λ) distributed number of accidents in one year. The parameter λ is unknown and different for every driver. We assume that λ is the outcome of a random variable Λ. Then the conditional distribution of the number of accidents N in one year, given $\Lambda = \lambda$, is Poisson(λ). What is the marginal distribution of N?

Let $U(\lambda) = \Pr[\Lambda \leq \lambda]$ denote the distribution function of Λ. Then we can write the marginal probabilities of event $N = n$ as

$$\Pr[N = n] = \int_0^\infty \Pr[N = n \,|\, \Lambda = \lambda]\,\mathrm{d}U(\lambda) = \int_0^\infty \mathrm{e}^{-\lambda} \frac{\lambda^n}{n!}\,\mathrm{d}U(\lambda), \tag{3.13}$$

while for the unconditional mean and variance of N we have

$$\begin{aligned} \mathrm{E}[N] &= \mathrm{E}\big[\mathrm{E}[N\,|\,\Lambda]\big] = \mathrm{E}[\Lambda]; \\ \mathrm{Var}[N] &= \mathrm{E}\big[\mathrm{Var}[N\,|\,\Lambda]\big] + \mathrm{Var}\big[\mathrm{E}[N\,|\,\Lambda]\big] = \mathrm{E}[\Lambda] + \mathrm{Var}[\Lambda] \geq \mathrm{E}[N]. \end{aligned} \tag{3.14}$$

Now assume additionally that $\Lambda \sim \text{gamma}(\alpha, \beta)$, then, writing $p = \beta/(\beta+1)$,

$$\begin{aligned} \mathrm{m}_N(t) &= \mathrm{E}\big[\mathrm{E}[\mathrm{e}^{tN}\,|\,\Lambda]\big] = \mathrm{E}\big[\exp\{\Lambda(\mathrm{e}^t - 1)\}\big] = \mathrm{m}_\Lambda(\mathrm{e}^t - 1) \\ &= \left(\frac{\beta}{\beta - (\mathrm{e}^t - 1)}\right)^\alpha = \left(\frac{p}{1-(1-p)\mathrm{e}^t}\right)^\alpha, \end{aligned} \tag{3.15}$$

which from Table A we recognize as the mgf of a negative binomial$(\alpha, \beta/(\beta+1))$ distribution. It can be shown that the overdispersion $\mathrm{Var}[N]/\mathrm{E}[N]$ is $1/p = 1 + 1/\beta$.

Obviously, the value of Λ for a particular driver is a non-observable random variable. It is the 'long run claim frequency', the value to which the observed average number of accidents in a year would converge if the driver could be observed for a very long time, during which his claims pattern does not change. The distribution of Λ is called the structure distribution, see also Chapter 8. ∇

Example 3.3.2 (Compound negative binomial is also compound Poisson)
At some intersection there are N traffic accidents with casualties in a year. There are L_i casualties in the ith accident, so $S = L_1 + L_2 + \cdots + L_N$ is the total number of casualties. Now assume $N \sim \text{Poisson}(\lambda)$ and $L_i \sim \text{logarithmic}(c)$ with $0 < c < 1$, so

$$\Pr[L_i = k] = \frac{c^k}{k\,h(c)}, \qquad k = 1, 2, \ldots \tag{3.16}$$

The division by the function $h(\cdot)$ serves to make the sum of the probabilities equal to 1, so from the usual series expansion of $\log(1+x)$, this function is $h(c) = -\log(1-c)$, hence the name logarithmic distribution. What is the distribution of S?

The mgf of the terms L_i is given by

$$\mathrm{m}_L(t) = \sum_{k=1}^{\infty} \frac{\mathrm{e}^{tk} c^k}{k\,h(c)} = \frac{h(c\,\mathrm{e}^t)}{h(c)}. \tag{3.17}$$

Then, for the mgf of S, we get

$$\begin{aligned} \mathrm{m}_S(t) &= \mathrm{m}_N(\log \mathrm{m}_L(t)) = \exp\lambda(\mathrm{m}_L(t) - 1) \\ &= \big(\exp\{h(c\,\mathrm{e}^t) - h(c)\}\big)^{\lambda/h(c)} = \left(\frac{1-c}{1-c\,\mathrm{e}^t}\right)^{\lambda/h(c)}, \end{aligned} \tag{3.18}$$

which, see again Table A, we recognize as the mgf of a negative binomial distribution with parameters $\lambda/h(c) = -\lambda/\log(1-c)$ and $1-c$.

On the one hand, the total payment Z for the casualties has a compound Poisson distribution since it is the sum of a Poisson(λ) number of payments per fatal accident (*cumulation*). On the other hand, summing over the casualties leads to a compound negative binomial distribution. It can be shown that if S_2 is compound negative binomial with parameters r and $p = 1 - q$ and claims distribution $P_2(\cdot)$, then S_2

has the same distribution as S_1, where S_1 is compound Poisson distributed with parameter λ and claims distribution $P_1(\cdot)$ given by:

$$\lambda = rh(q) \quad \text{and} \quad P_1(x) = \sum_{k=1}^{\infty} \frac{q^k}{kh(q)} P_2^{*k}(x). \tag{3.19}$$

In this way, any compound negative binomial distribution can be written as a compound Poisson distribution. ∇

Remark 3.3.3 (Compound Poisson distributions in probability theory)
The compound Poisson distributions are also object of study in probability theory. If we extend this class with its limits, to which the gamma and the normal distribution belong, then we have just the class of infinitely divisible distributions. This class consists of the random variables X with the property that for each n, a sequence of iid random variables $X_1, X_2, \ldots, X_n$ exists with $X \sim X_1 + X_2 + \cdots + X_n$. ∇

3.4 Properties of compound Poisson distributions

In this section we prove some important theorems on compound Poisson distributions and use them to construct a better algorithm to calculate $F(\cdot)$ than given by (3.10). First, we show that the class of compound Poisson distributions is closed under convolution.

Theorem 3.4.1 (Sum of compound Poisson r.v.'s is compound Poisson)
If $S_1, S_2, \ldots, S_m$ are independent compound Poisson random variables with Poisson parameter λ_i and claims distribution P_i, $i = 1, 2, \ldots, m$, then $S = S_1 + S_2 + \cdots + S_m$ is compound Poisson distributed with specifications

$$\lambda = \sum_{i=1}^{m} \lambda_i \quad \text{and} \quad P(x) = \sum_{i=1}^{m} \frac{\lambda_i}{\lambda} P_i(x). \tag{3.20}$$

Proof. Let m_i be the mgf of P_i. Then S has the following mgf:

$$\mathrm{m}_S(t) = \prod_{i=1}^{m} \exp\left\{\lambda_i \big[m_i(t) - 1\big]\right\} = \exp\lambda \left\{\sum_{i=1}^{m} \frac{\lambda_i}{\lambda} m_i(t) - 1\right\}. \tag{3.21}$$

So S is a compound Poisson random variable with specifications (3.20). ∇

Consequently, the total result of m independent compound Poisson portfolios is again compound Poisson distributed. The same holds if we observe the same portfolio in m years, assuming that the annual results are independent.

A special case is when the S_i have fixed claims x_i, hence $S_i = x_i N_i$ with $N_i \sim$ Poisson(λ_i). Assume the x_i to be all different. We get the random variable

$$S = x_1 N_1 + x_2 N_2 + \cdots + x_m N_m, \tag{3.22}$$

which by Theorem 3.4.1 is compound Poisson with specifications:

$$\lambda = \lambda_1 + \cdots + \lambda_m \quad \text{and} \quad p(x_i) = \frac{\lambda_i}{\lambda}, i = 1, \ldots, m. \tag{3.23}$$

We can also prove the reverse statement, as follows:

Theorem 3.4.2 (Frequencies of claim sizes are independent Poisson)
Assume that S is compound Poisson distributed with parameter λ and with discrete claims distribution

$$\pi_i = p(x_i) = \Pr[X = x_i], \quad i = 1, 2, \ldots, m. \tag{3.24}$$

Suppose S is written as (3.22), where N_i denotes the frequency of the claim amount x_i, that is, the number of terms in S with value x_i. Then $N_1, \ldots, N_m$ are independent Poisson($\lambda \pi_i$) random variables, $i = 1, \ldots, m$.

Proof. Let $N = N_1 + \cdots + N_m$ and $n = n_1 + \cdots + n_m$. Conditionally on $N = n$, we have $N_1, \ldots, N_m \sim$ Multinomial$(n, \pi_1, \ldots, \pi_m)$. Hence,

$$\begin{aligned} &\Pr[N_1 = n_1, \ldots, N_m = n_m] \\ &\quad = \Pr[N_1 = n_1, \ldots, N_m = n_m \,|\, N = n] \Pr[N = n] \\ &\quad = \frac{n!}{n_1! n_2! \ldots n_m!} \pi_1^{n_1} \pi_2^{n_2} \ldots \pi_m^{n_m} \mathrm{e}^{-\lambda} \frac{\lambda^n}{n!} \\ &\quad = \prod_{i=1}^{m} \mathrm{e}^{-\lambda \pi_i} \frac{(\lambda \pi_i)^{n_i}}{n_i!}. \end{aligned} \tag{3.25}$$

By summing over all n_i, $i \neq k$, we see that N_k is marginally Poisson($\lambda \pi_k$) distributed. The N_i are independent since $\Pr[N_1 = n_1, \ldots, N_m = n_m]$ is the product of the marginal probabilities of $N_i = n_i$. ∇

Example 3.4.3 (Application: sparse vector algorithm)
If the claims X are integer-valued and non-negative, we can calculate the compound Poisson cdf F in an efficient way. We explain this by an example. Let $\lambda = 4$ and $\Pr[X = 1, 2, 3] = \frac{1}{4}, \frac{1}{2}, \frac{1}{4}$. Then, gathering together terms as we did in (3.22), we can write S as $S = 1N_1 + 2N_2 + 3N_3$ and calculate the distribution of S by convolution. We can compute $f(x) = \Pr[S = x]$ as follows:

x	$\Pr[N_1 = x]$ $(\mathrm{e}^{-1}\times)$	$*$	$\Pr[2N_2 = x]$ $(\mathrm{e}^{-2}\times)$	$=$	$\Pr[N_1 + 2N_2 = x]$ $(\mathrm{e}^{-3}\times)$	$*$	$\Pr[3N_3 = x]$ $(\mathrm{e}^{-1}\times)$	$=$	$\Pr[S = x]$ $(\mathrm{e}^{-4}\times)$
0	1		1		1		1		1
1	1		–		1		–		1
2	1/2		2		5/2		–		5/2
3	1/6		–		13/6		1		19/6
4	1/24		2		:		–		:
:	:		:		:		:		:
	↑		↑				↑		
	$1/x!$		$2^{x/2}/(x/2)!$				$1/(x/3)!$		

The density of the total amount of the claims of size $1,2,\ldots,j-1$ is convoluted with the one of jN_j. In the column with probabilities of jN_j, only the rows $0,j,2j,\ldots$ are filled, which is why this algorithm is called a 'sparse vector' algorithm. These probabilities are Poisson($\lambda\pi_j$) probabilities.

Implementing the Sparse vector algorithm in R is easy, since convolution of two vectors can be handled by the `convolve` function. It employs a technique called the *Fast Fourier Transform (FFT)*, the workings of which are explained in Section 3.6. An R-implementation of the sparse vector algorithm is as follows:

```
SparseVec <- function (freq)
{if (any(freq<0)) stop("negative frequency")
 M <- length(freq)
 mu <- sum((1:M)*freq); sigma2 <- sum((1:M)^2*freq)
 ##mean and variance of the compound r.v.; see (3.4)
 MM <- ceiling(mu + 10 * sqrt(sigma2)) + 6
 fs <- dpois(0:(MM-1), freq[1])  ##density of S_1 = 1*N_1
 for (j in 2:M)
 {MMM <- trunc((MM-1)/j)
  fj <- rep(0, MM)  ##construct the density of j*N_j
  fj[(0:MMM)*j+1] <- dpois(0:MMM, freq[j])
  fs <- convolve(fs, rev(fj), type="o") }
 ##fs is the density of S_j = 1*N_1 + ... + j*N_j, j=2..M
 return(fs)  }
f <- SparseVec(c(1,2,1)); f[1:7] * exp(4)
```

The last line reproduces the first seven numbers in the last column of the table in Example 3.4.3. The argument `freq` contains the expected frequencies λp_j of each claim amount $j = 1,2,\ldots$, which should of course be non-negative. The vector length `MM` is taken to be the mean plus 10 standard deviations plus 7, ensuring that `sum(fs[1:MM])` will always be virtually equal to 1. The vector `fs` is initialized to the density of $1N_1$, and convoluted with the one of jN_j in step j, $j = 2,\ldots,m$. Note that it is required that the second vector given as an argument to `convolve` is reversed and that the type given is `"o"`, short for `"open"`. The function result returned is the probability distribution of the compound random variable.

The algorithm given is fast since it uses the efficient FFT technique to do the convolutions. It is, however, not a proper sparse vector algorithm, since the fact that the vector `fj` has zeros at places that are non-multiples of `j` is never used. It can be shown that for large m and n, to compute probabilities of $0,\ldots,n$ with a maximal claim amount m by `convolve` takes $O(mn\log n)$ operations, while the sparse vector algorithm needs $O(n^2\log m)$ (see Exercise 3.5.7). ∇

3.5 Panjer's recursion

In 1981, Panjer described a method to calculate the probabilities $f(x)$ recursively. In fact, the method can be traced back to as early as Euler. As a result of Panjer's publication, a lot of other articles have appeared in the actuarial literature covering similar recursion relations. The recursion relation described by Panjer is as follows:

Theorem 3.5.1 (Panjer's recursion)
Consider a compound distribution with integer-valued non-negative claims with pdf $p(x)$, $x = 0,1,2,\ldots$, for which, for some real a and b, the probability q_n of having n claims satisfies the following recursion relation

$$q_n = \Big(a + \frac{b}{n}\Big) q_{n-1}, \quad n = 1,2,\ldots \tag{3.26}$$

Then the following relations for the probability of a total claim equal to s hold:

$$\begin{aligned} f(0) &= \begin{cases} \Pr[N=0] & \text{if} \quad p(0) = 0; \\ \mathrm{m}_N\big(\log p(0)\big) & \text{if} \quad p(0) > 0; \end{cases} \\ f(s) &= \frac{1}{1-ap(0)} \sum_{h=1}^{s} \Big(a + \frac{bh}{s}\Big) p(h) f(s-h), \quad s = 1,2,\ldots \end{aligned} \tag{3.27}$$

Proof. From $\Pr[S=0] = \sum_{n=0}^{\infty} \Pr[N=n] p^n(0)$ we get the starting value $f(0)$. Write $T_k = X_1 + \cdots + X_k$. First, note that because of symmetry:

$$\mathrm{E}\Big[a + \frac{bX_1}{s} \,\Big|\, T_k = s\Big] = a + \frac{b}{k}. \tag{3.28}$$

This expectation can also be determined in the following way:

$$\begin{aligned} \mathrm{E}\Big[a + \frac{bX_1}{s} \,\Big|\, T_k = s\Big] &= \sum_{h=0}^{s} \Big(a + \frac{bh}{s}\Big) \Pr[X_1 = h \,|\, T_k = s] \\ &= \sum_{h=0}^{s} \Big(a + \frac{bh}{s}\Big) \frac{\Pr[X_1 = h] \Pr[T_k - X_1 = s-h]}{\Pr[T_k = s]}. \end{aligned} \tag{3.29}$$

Because of (3.26) and the previous two equalities, we have, for $s = 1,2,\ldots$,

$$\begin{aligned} f(s) &= \sum_{k=1}^{\infty} q_k \Pr[T_k = s] = \sum_{k=1}^{\infty} q_{k-1} \Big(a + \frac{b}{k}\Big) \Pr[T_k = s] \\ &= \sum_{k=1}^{\infty} q_{k-1} \sum_{h=0}^{s} \Big(a + \frac{bh}{s}\Big) \Pr[X_1 = h] \Pr[T_k - X_1 = s-h] \\ &= \sum_{h=0}^{s} \Big(a + \frac{bh}{s}\Big) \Pr[X_1 = h] \sum_{k=1}^{\infty} q_{k-1} \Pr[T_k - X_1 = s-h] \\ &= \sum_{h=0}^{s} \Big(a + \frac{bh}{s}\Big) p(h) f(s-h) \\ &= ap(0) f(s) + \sum_{h=1}^{s} \Big(a + \frac{bh}{s}\Big) p(h) f(s-h), \end{aligned} \tag{3.30}$$

from which the second relation of (3.27) follows immediately. ∇

Example 3.5.2 (Distributions suitable for Panjer's recursion)
Only the following distributions satisfy relation (3.26):

1. Poisson(λ) with $a = 0$ and $b = \lambda \geq 0$; in this case, (3.27) simplifies to:

$$\begin{aligned} f(0) &= \mathrm{e}^{-\lambda(1-p(0))}; \\ f(s) &= \frac{1}{s}\sum_{h=1}^{s} \lambda h p(h) f(s-h); \end{aligned} \tag{3.31}$$

2. Negative binomial(r, p) with $p = 1-a$ and $r = 1+\frac{b}{a}$; so $0 < a < 1$ and $a+b > 0$;
3. Binomial(k, p) with $p = \frac{a}{a-1}$ and $k = -\frac{b+a}{a}$; so $a < 0$, $b = -a(k+1)$.

If $a+b = 0$, then $q_0 = 1$ and $q_j = 0$ for $j = 1,2,\ldots$, so we get a Poisson(0) distribution. For other values of a and b than the ones used above, $q_n = \big(a+\frac{b}{n}\big)q_{n-1}$ for all $n = 1,2,\ldots$ cannot hold for a probability distribution:

- $q_0 \leq 0$ is not feasible, so assume $q_0 > 0$;
- $a+b < 0$ results in $q_1 < 0$;
- $a < 0$ and $b \neq a(n+1)$ for all n also results in negative probabilities;
- if $a \geq 1$ and $a+b > 0$, then $nq_n = \big((n-1)a+a+b\big)q_{n-1} > (n-1)q_{n-1}$ from (3.26), so $q_n > q_1/n$, $n = 1,2,\ldots$ and consequently $\sum_n q_n = \infty$.

By allowing (3.26) to hold only for $n \geq 2$, hence admitting an arbitrary probability of no claims, we can find similar recursions for a larger group of counting distributions, which includes the logarithmic distributions. See Exercise 3.5.14 and 3.5.15. ∇

Example 3.5.3 (Example 3.4.3 solved by Panjer's recursion)
As in Example 3.4.3, consider a compound Poisson distribution with $\lambda = 4$ and $\Pr[X = 1,2,3] = \frac{1}{4}, \frac{1}{2}, \frac{1}{4}$. Then (3.31) simplifies to

$$f(s) = \frac{1}{s}\big[f(s-1)+4f(s-2)+3f(s-3)\big], \quad s = 1,2,\ldots, \tag{3.32}$$

and the starting value is $f(0) = \mathrm{e}^{-4} \approx 0.0183$. We have

$$\begin{aligned} f(1) &= f(0) = \mathrm{e}^{-4}, \\ f(2) &= \tfrac{1}{2}\big[f(1)+4f(0)\big] = \tfrac{5}{2}\mathrm{e}^{-4}, \\ f(3) &= \tfrac{1}{3}\big[f(2)+4f(1)+3f(0)\big] = \tfrac{19}{6}\mathrm{e}^{-4}, \end{aligned} \tag{3.33}$$

and so on. ∇

Example 3.5.4 (Panjer's recursion and stop-loss premiums)
For an integer-valued S, we can write the stop-loss premium in an integer retention d as follows, see Section 1.4:

$$\mathrm{E}[(S-d)_+] = \sum_{x=d}^{\infty}(x-d)f(x) = \sum_{x=d}^{\infty}[1-F(x)]. \tag{3.34}$$

The stop-loss premium is piecewise linear in the retention on the intervals where the cdf remains constant, since for the right hand derivative we have by (1.38):

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{E}[(S-t)_+] = F(t) - 1. \tag{3.35}$$

So the stop-loss premiums for non-integer d follow by linear interpolation.

With Panjer's recursion the stop-loss premiums can be calculated recursively, too, since from the last relation in (3.34), we have for integer d

$$\pi(d) := \mathrm{E}[(S-d)_+] = \pi(d-1) - [1 - F(d-1)]. \tag{3.36}$$

As an example, take $S \sim$ compound Poisson(1) with $p(1) = p(2) = \frac{1}{2}$. Then, Panjer's recursion relation (3.31) simplifies to

$$f(x) = \frac{1}{x}\left[\tfrac{1}{2}f(x-1) + f(x-2)\right], \quad x = 1, 2, \ldots \tag{3.37}$$

with starting values

$$f(0) = \mathrm{e}^{-1} \approx 0.368, \quad F(0) = f(0), \quad \pi(0) = \mathrm{E}[S] = \lambda\mu_1 = \frac{3}{2}. \tag{3.38}$$

This leads to the following calculations:

x	$f(x) = (3.37)$	$F(x) = F(x-1) + f(x)$	$\pi(x) = \pi(x-1) - 1 + F(x-1)$
0	0.368	0.368	1.500
1	0.184	0.552	0.868
2	0.230	0.782	0.420
3	0.100	0.881	0.201
4	0.070	0.951	0.083
5	0.027	0.978	0.034

The advantage of computing the cdf and the stop-loss premiums simultaneously with the recursion is that there is no need to store the whole array of n values of $f(x)$, which might make a difference if the maximum claim m is much smaller than n. When in `R` the $f(x)$ values are stored in `f`, `cumsum(f)` produces the values of the cdf. For how to compute the successive stop-loss premiums, see Exercise 3.5.15. ∇

Remark 3.5.5 (Proof of Panjer's recursion through pgfs)
Panjer's recursion can also be derived from the probability generating functions. For the compound Poisson distribution, this goes as follows. First write

$$\frac{\mathrm{d}g_S(t)}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\sum_{s=0}^{\infty} t^s \Pr[S=s] = \sum_{s=1}^{\infty} s t^{s-1} \Pr[S=s]. \tag{3.39}$$

Just as in (3.5), we have

$$g_S(t) = g_N(g_X(t)) = \exp\lambda(g_X(t) - 1), \tag{3.40}$$

so the derivative also equals $g'_S(t) = \lambda g_S(t) g'_X(t)$. For other distributions, similar expressions can be derived from (3.26). Now for $g_S(\cdot)$ and $g'_X(\cdot)$, substitute their series expansions:

$$\begin{aligned}
\lambda g_S(t) g'_X(t) &= \lambda \left(\sum_{s=0}^{\infty} t^s \Pr[S=s] \right) \left(\sum_{x=1}^{\infty} x t^{x-1} \Pr[X=x] \right) \\
&= \sum_{x=1}^{\infty} \sum_{s=0}^{\infty} \lambda x t^{s+x-1} \Pr[S=s] \Pr[X=x] \\
&= \sum_{x=1}^{\infty} \sum_{v=x}^{\infty} \lambda x t^{v-1} \Pr[S=v-x] \Pr[X=x] \\
&= \sum_{v=1}^{\infty} \sum_{x=1}^{v} \lambda x t^{v-1} \Pr[S=v-x] \Pr[X=x].
\end{aligned} \tag{3.41}$$

Comparing the coefficients of t^{s-1} in (3.39) and (3.41) yields

$$s \Pr[S=s] = \sum_{x=1}^{s} \lambda x \Pr[S=s-x] \Pr[X=x]. \tag{3.42}$$

Dividing by s, one sees that this relation is equivalent with Panjer's recursion relation for the Poisson case (3.31). ∇

Remark 3.5.6 (Convolution using Panjer's recursion)
How can we calculate the n-fold convolution of a distribution on $0, 1, 2, \ldots$ with Panjer's recursion?

Assume that $p(0) > 0$. If we replace X_i by $I_i Y_i$ where $\Pr[I_i = 1] = \Pr[X_i > 0] =: p$ and $Y_i \sim X_i \,|\, X_i > 0$, then $\sum_i X_i$ has the same distribution as $\sum_i I_i Y_i$, which gives us a compound binomial distribution with $p < 1$ as required in Example 3.5.2. Another method is to take limits for $p \uparrow 1$ in (3.27) for those values of a and b that produce a binomial(n, p) distribution. ∇

Remark 3.5.7 (Implementing Panjer's recursion)
To compute the sum in the Panjer recursion (3.31), we might use R's loop mechanisms, but because R is an interpreted language it is worthwhile to 'vectorize' the computations, replacing the innermost loop by a call of `sum`. For this, note that to compute $f(s)$, the terms to be added in (3.31) are λ/s times the products of successive elements of the three vectors $(1, \ldots, m)$, $(p(1), \ldots, p(m))$ and $(f(s-1), \ldots, f(s-m))$. Here $m = \min\{s, r\}$ with r the maximal index for which $p_r > 0$. The second vector is the head of the p-vector, the third is the reverse of the tail part of the f-vector. An R program to implement Panjer's recursion and to reproduce the results of Example 3.5.3 is as follows:

```
Panjer.Poisson <- function (p, lambda)
{ if (sum(p)>1||any(p<0)) stop("p parameter not a density")
  if (lambda * sum(p) > 727) stop("Underflow")
```

```
    cumul <- f <- exp(-lambda * sum(p))
    r <- length(p)
    s <- 0
    repeat
    { s <- s+1
      m <- min(s, r)
      last <- lambda / s * sum(1:m * head(p,m) * rev(tail(f,m)))
      f <- c(f,last)
      cumul <- cumul + last
      if (cumul > 0.99999999) break  }
      return(f)  }
Panjer.Poisson(c(0.25,0.5,0.25), 4) * exp(4)
```

The parameter `p` must contain the values of $p(1), p(2), \ldots$, and it is checked if this, combined with $p(0) = 1 - \sum_h p(h)$, is indeed a density.

The parameter `lambda` representing λ should not be too big; in a standard Windows system problems arise if $\lambda(1-p(0)) > 727$ holds, because in that case $f(0)$ is too small. `R` uses *double* precision (64-bit reals), but in programming environments employing *extended* precision (80-bit reals), one can easily cope with portfolios having $\lambda \approx 11340$, starting from $\Pr[S=0] \approx 10^{-5000}$. In some older languages, a 48-bit real data type was used, leading to underflow already for $\lambda(1-p(0)) \geq 88$. So for a portfolio of n life insurance policies with probabilities of claim equal to 0.5%, the calculation of $\Pr[S=0]$ already experienced underflow for $n = 17600$. This underflow problem cannot be easily resolved in `R` itself, but it is possible to call compiled external code, using extended precision calculations, from `R`. This also reduces the running time, and in case of compound binomial probabilities, it might help remedy the numerical instability sometimes encountered.

The result of calling `pp <- Panjer.Poisson(...)` is a vector `pp` of probabilities $f(0), f(1), \ldots, f(n)$ with the upper bound n such that $f(0)+f(1)+\cdots+f(n) > 1-10^{-8}$. Recall that in `R`, all arrays start with index `1`, so `pp[1]` stores $f(0)$, and so on. ∇

3.6 Compound distributions and the Fast Fourier Transform

Another method to compute the probabilities of a compound distribution is based on inversion of the characteristic function. Let $X_1, X_2, \cdots \sim X$ be random variables with values in $\{0,1,2,\ldots\}$, independent of each other as well as of the claim number N. Let $S = X_1 + \cdots + X_N$, and denote the probabilities by

$$p_h = \Pr[X=h]; \qquad q_n = \Pr[N=n]; \qquad f_s = \Pr[S=s]. \tag{3.43}$$

Now let m be a number sufficiently large to let $\Pr[S \leq m] \approx 1$. For the Poisson and (negative) binomial case, Panjer's recursion requires $O(m^2)$ steps to compute $f_0, \ldots, f_m$, or $O(m)$ if X is bounded. In this section we will introduce the Fast Fourier Transform method and show that it requires $O(m \log m)$ steps, for all compound

distributions with a easily computable expression for the generating function of the number of claims.

The characteristic function of the compound random variable S is $\phi_S(t) \stackrel{\text{def}}{=} \mathrm{E}[\mathrm{e}^{\mathrm{i}tS}]$. Along the lines of (3.5) it can be proved that $\phi_S(t) = g_N(\phi_X(t))$, with g_N the pgf of N. The probabilities f_s can be found back from the characteristic function as follows:

$$f_s = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{e}^{-\mathrm{i}ts}\phi_S(t)\,\mathrm{d}t. \tag{3.44}$$

To prove this inversion formula is easy:

$$\frac{1}{2\pi}\int_0^{2\pi} \mathrm{e}^{-\mathrm{i}ts}\phi_S(t)\,\mathrm{d}t = \frac{1}{2\pi}\int_0^{2\pi} \Big\{f_s + \sum_{k\neq s} f_k \mathrm{e}^{-\mathrm{i}t(k-s)}\Big\}\,\mathrm{d}t = f_s + 0, \tag{3.45}$$

since for all $k \neq s$, as is readily verified by substituting $u = t(k-s)$,

$$\int_0^{2\pi} \mathrm{e}^{-\mathrm{i}t(k-s)}\mathrm{d}t = \int_0^{2\pi} \big\{\cos(-t(k-s)) + \mathrm{i}\sin(-t(k-s))\big\}\,\mathrm{d}t = 0. \tag{3.46}$$

Applying the trapezoidal rule to (3.44) with intervals of length $2\pi/n$, we see that an approximation for f_s is given by

$$f_s \approx r_s := \frac{1}{n}\sum_{h=0}^{n-1} \mathrm{e}^{-\mathrm{i}2\pi sh/n}\phi_S(2\pi h/n). \tag{3.47}$$

Note that (3.47) applies only if $s = 0,1,\ldots,n-1$; for example for $s = n, 2n, \ldots$ we get the same approximation as for $s = 0$.

Now introduce the *discrete Fourier Transform* of a vector $\vec{f} = (f_0,\ldots,f_{n-1})$ as the vector $\vec{y} = \mathbf{T}^-\vec{f}$, with the matrix $\mathbf{T}^-$ defined as follows:

$$T^-_{jk} = \mathrm{e}^{-\mathrm{i}2\pi jk/n}, \qquad j,k = 0,1,\ldots,n-1. \tag{3.48}$$

This means that every element t occurring in $\mathbf{T}^-$ is a unit root, with $t^n = 1$. Also define $\mathbf{T}^+$ in the same way, but with a plus sign in the exponent. Then approximation (3.47) can be written as

$$\vec{f} \approx \frac{1}{n}\mathbf{T}^- g_N\big(\mathbf{T}^+\vec{p}\big). \tag{3.49}$$

All this would not be very useful but for two things. First, it is possible to compute approximation (3.49) very fast by an algorithm called the *Fast Fourier Transform*. It takes time $O(n\log n)$ and memory $O(n)$ only. If implemented naively, it would require $O(n^3)$ operations and $O(n^2)$ memory. Second, using another interpretation for the right hand side of (3.47), we will show how to make the error of the approximation negligible by taking n large enough.

The matrices $\mathbf{T}^+$ and $\mathbf{T}^-$ are in fact each other's inverse in the sense that $\frac{1}{n}\mathbf{T}^- = (\mathbf{T}^+)^{-1}$, because for fixed j,k, writing $\omega = \mathrm{e}^{\mathrm{i}2\pi(j-k)/n}$, we get

$$(\mathbf{T}^+\mathbf{T}^-)_{jk} = \sum_{h=0}^{n-1} \mathrm{e}^{+\mathrm{i}2\pi jh/n}\mathrm{e}^{-\mathrm{i}2\pi hk/n} = \sum_{h=0}^{n-1} \omega^h = \begin{cases} n & \text{if } j=k \\ \frac{1-\omega^n}{1-\omega} = 0 & \text{if } j \neq k. \end{cases} \tag{3.50}$$

If $z = x + \mathrm{i}y = r\mathrm{e}^{\mathrm{i}\phi}$, then $\overline{z} = x - \mathrm{i}y = r\mathrm{e}^{-\mathrm{i}\phi}$ is its complex conjugate. So

$$(\mathbf{T}^+\vec{g})_j = \sum_{h=0}^{n-1} \mathrm{e}^{+\mathrm{i}2\pi jh/n} g_h = \sum_{h=0}^{n-1} \overline{\mathrm{e}^{-\mathrm{i}2\pi jh/n}\,\overline{g_h}} = \overline{(\mathbf{T}^-\overline{\vec{g}})_j}. \tag{3.51}$$

Therefore the inverse operation of an FFT can be handled by the same algorithm, apart from taking complex conjugates and a division by the length of the vector.

To show that a Fast Fourier Transform algorithm can be constructed taking only time $O(n\log n)$, we use a 'divide and conquer' approach. Assume n even, then for $m = 0, 1, \ldots, n-1$, substituting $k = 2h + j$:

$$y_m = \sum_{h=0}^{n-1} \mathrm{e}^{\mathrm{i}2\pi mk/n} g_k = \sum_{j=0}^{1} \underbrace{\sum_{h=0}^{n/2-1} \mathrm{e}^{\mathrm{i}2\pi m2h/n} g_{2h+j}} \tag{3.52}$$

For $j = 0$, the underbraced sum involves an FFT of length $\frac{n}{2}$ on $(g_0, g_2, \ldots, g_{n-2})$, for $j = 1$, on $(g_1, g_3, \ldots, g_{n-1})$. Therefore, using an induction assumption, to compute the FFT of length n takes time $2 \times \alpha\frac{1}{2}n\log\frac{1}{2}n + \beta n$ for some α and β, since two FFT's of length $n/2$ must be computed, plus a summation over j for each n. This adds up to $\alpha n\log n + n(\beta - \alpha\log 2)$, which is less than $\alpha n\log n$ in total provided $\alpha > \beta/\log 2$ is taken. Iterating this proves that FFT can be done using only $O(n\log n)$ operations.

If $\mathrm{Re}(g_j) = \Pr[Z = j], \mathrm{Im}(g_j) = 0$, and $\Pr[Z \in \{0, 1, \ldots, n-1\}] = 1$ for a random variable Z, then for the characteristic function $\phi_Z(t) = \mathrm{E}[\mathrm{e}^{\mathrm{i}tZ}]$ we have

$$\phi_Z(2\pi j/n) = \sum_{k=0}^{n-1} \mathrm{e}^{\mathrm{i}2\pi jk/n} g_k = (\mathbf{T}^+\vec{g})_j, \quad j = 0, 1, \ldots, n-1. \tag{3.53}$$

Since Z is integer-valued, $\phi_Z(t + 2\pi) = \mathrm{E}[\mathrm{e}^{\mathrm{i}tZ}\mathrm{e}^{\mathrm{i}2\pi Z}] = \phi_Z(t)$ for all real t, so ϕ_Z is periodical with period 2π.

By the above, for a random variable Z with support $\{0, 1, \ldots, n-1\}$ we have

$$y_j = \phi_Z(2\pi j/n) \quad \Longrightarrow \quad \mathbf{T}^+\vec{g} = \vec{y} \quad \Longrightarrow \quad \vec{g} = \frac{1}{n}\mathbf{T}^-\vec{y}. \tag{3.54}$$

To apply this to $S = X_1 + \cdots + X_N$, write $Z \equiv S \bmod n$: the remainder when S is divided by n. Then Z has support $\{0, 1, \ldots, n-1\}$, and $\Pr[Z \in \{S, S \pm n, S \pm 2n, \ldots\}] = 1$. Therefore their discrete Fourier transforms coincide:

$$\phi_S(2\pi j/n) = \phi_Z(2\pi j/n) \quad \text{for all integer } j. \tag{3.55}$$

For large n, $\Pr[Z=k] \approx \Pr[S=k]$. To compute the characteristic function of S from transforms of N and X, use the relation $\phi_S(t) = g_N(\phi_X(t))$, see also (3.5).

We summarize the discussion above in the following theorem.

Theorem 3.6.1 (Computing compound cdfs by FFT)
If S has a compound distribution with specifications (3.43), we can approximate the probabilities f_s of $S=s$, $s=0,\ldots,n-1$ by the exact probabilities r_s of $Z=s$, where $Z \equiv S \bmod n$. These probabilities can be computed, in time $O(n\log n)$, as:

$$\vec{f} \approx \vec{r} := \frac{1}{n}\mathbf{T}^{-} g_N(\mathbf{T}^{+}\vec{q}). \tag{3.56}$$

The error $r_s - f_s = \Pr[S \in \{s+n, s+2n, \ldots\}]$, $s=0,\ldots,n-1$. ∇

Example 3.6.2 (Example 3.5.4 using FFT)
R has a built-in function `fft` to do the calculations. Using it, the probabilities of a compound Poisson($\lambda=1$) random variable with claims distribution $\Pr[X=1] = \Pr[X=2] = \frac{1}{2}$ (see Example 3.5.4) can be reproduced as follows:

```
n <- 64; p <- rep(0, n); p[2:3] <- 0.5; lab <- 1
f <- Re(fft(exp(lab*(fft(p)-1)), inverse=TRUE))/n
```

Note that the R-function does not automatically do the division by `n`. Also note that we need to pad the `p` vector with enough zeros to let the resulting total probability of $S \in \{0,1,\ldots,n-1\}$ be near enough to one. For the Fast Fourier Transform to live up to its name, the number of elements in the vector should preferably be a power of two, but in any case have a lot of factors. ∇

In case $N \sim$ binomial or negative binomial, all one has to do is plug in the appropriate generating function. Also, for example a logarithmic number of claims can be handled easily. Moreover, the FFT-technique can be adapted to deal with negative claim amounts.

When the number n of conceivable total claim sizes is large, Panjer's recursion requires time $O(n^2)$ if the individual claim sizes are unbounded. In that case, FFT provides an easy to use and fast alternative, not quite exact but with a controllable error, and taking only $O(n\log n)$ time. For Panjer's recursion, one would typically have to compute the probabilities up to either the retention d, or to a certain quantile like the 75% quantile, but with FFT, it is mandatory to take n large enough to let $\Pr[S>n]$ be negligible. This does not make a lot of difference, asymptotically.

Note that for FFT, no for-loops were needed such as with Panjer's recursion. Therefore using FFT in many cases will be a lot faster than a recursive method.

3.7 Approximations for compound distributions

In the previous chapter, approximations were given that were refinements of the CLT, in which the distribution of a sum of a large number of random variables is

approximated by a normal distribution. These approximations can also be used if the number of terms in a sum is a random variable with large values. For example, for the compound Poisson distribution with large λ we have the following counterpart of the CLT; similar results can be derived for other compound distributions.

Theorem 3.7.1 (CLT for compound Poisson distributions)
Let S be compound Poisson distributed with parameter λ and general claims cdf $P(\cdot)$ with finite variance. Then, with $\mu = \mathrm{E}[S]$ and $\sigma^2 = \mathrm{Var}[S]$,

$$\lim_{\lambda\to\infty} \Pr\left[\frac{S-\mu}{\sigma} \le x\right] = \Phi(x). \tag{3.57}$$

Proof. If $N_1, N_2, \ldots$ is a series of independent Poisson(1) random variables and if X_{ij}, $i = 1,2,\ldots$, $j = 1,2,\ldots$ are independent random variables with cdf $P(\cdot)$, then for integer-valued λ, we have

$$S \sim \sum_{j=1}^{\lambda} \sum_{i=1}^{N_j} X_{ij}, \quad \text{since} \quad \sum_{j=1}^{\lambda} N_j \sim N. \tag{3.58}$$

As S in (3.58) is the sum of λ independent and identically distributed random variables, the CLT can be applied directly. Note that taking λ to be an integer presents no loss of generality, since the influence of the fractional part vanishes for large λ.

In this proof, we have reduced the situation to the Central Limit theorem. A proof along the lines of the one of Theorem 2.5.1 is asked in Exercise 3.7.3. ∇

To use the CLT, translated gamma approximation and normal power approximation (NP) one needs the cumulants of S. Again, let μ_k denote the kth moment of the claims distribution. Then, for the compound Poisson distribution, we have

$$\kappa_S(t) = \lambda\,(\mathrm{m}_X(t) - 1) = \lambda \sum_{k=1}^{\infty} \mu_k \frac{t^k}{k!}. \tag{3.59}$$

From (2.46) we know that the coefficients of $\frac{t^k}{k!}$ are the cumulants. Hence mean, variance and third central moment of a compound Poisson(λ, x) random variable with *raw moments* $\mu_j = \mathrm{E}[X^j]$ are given by

$$\mathrm{E}[S] = \lambda\mu_1, \quad \mathrm{Var}[S] = \lambda\mu_2 \quad \text{and} \quad \mathrm{E}[(S - \mathrm{E}[S])^3] = \lambda\mu_3. \tag{3.60}$$

The skewness is proportional to $\lambda^{-1/2}$:

$$\gamma_S = \frac{\mu_3}{\mu_2^{3/2}\sqrt{\lambda}}. \tag{3.61}$$

Remark 3.7.2 (Asymptotics and underflow)
There are certain situations in which one would have to resort to approximations. First of all, if the calculation time is uncomfortably long: for the calculation of $f(s)$

in (3.31) for large s, we need a lot of multiplications, see Exercise 3.5.4. Second, the recursion might not 'get off the ground'; see Remark 3.5.7.

Fortunately, the approximations improve with increasing λ; they are asymptotically exact, since in the limit they coincide with the usual normal approximation based on the CLT. ∇

3.8 Individual and collective risk model

In the preceding sections we have shown that replacing the individual model by the collective risk model has distinct computational advantages. In this section we focus on the question which collective model should be chosen. We consider a situation from life insurance, but the same situation occurs in non-life insurance, for example when fines are imposed (malus) if an employee gets disabled.

Consider n one-year life insurance policies. At death, which happens with probability q_i, the claim amount on policy i is b_i, assumed positive, otherwise it is 0. We want to approximate the total amount of the claims on all policies using a collective model. For that purpose, we replace the I_i payments of size b_i for policy i, where $I_i \sim$ Bernoulli(q_i), by a Poisson(λ_i) distributed number of payments b_i. Instead of the cdf of the total payment in the individual model

$$\widetilde{S} = \sum_{i=1}^{n} I_i b_i, \quad \text{with} \quad \Pr[I_i = 1] = q_i = 1 - \Pr[I_i = 0], \tag{3.62}$$

we consider the cdf of the following approximating random variable:

$$S = \sum_{i=1}^{n} Y_i, \quad \text{with} \quad Y_i = N_i b_i = \sum_{j=1}^{N_i} b_i \quad \text{and} \quad N_i \sim \text{Poisson}(\lambda_i). \tag{3.63}$$

If we choose $\lambda_i = q_i$, the expected number of payments for policy i is equal in both models. To stay on the safe side, we could also choose $\lambda_i = -\log(1-q_i) > q_i$. With this choice, the probability of 0 claims on policy i is equal in both the collective and the individual model. This way, we incorporate implicit margins by using a larger total claim size than the original one. See also Section 7.4.1 and Remark 3.8.2.

Although (3.63) still has the form of an individual model, S is a compound Poisson distributed random variable because of Theorem 3.4.1, so it is indeed a collective model as in (3.1). The specifications are:

$$\lambda = \sum_{i=1}^{n} \lambda_i \quad \text{and} \quad P(x) = \sum_{i=1}^{n} \frac{\lambda_i}{\lambda} I_{[b_i,\infty)}(x), \tag{3.64}$$

with the indicator function $I_A(x) = 1$ if $x \in A$ and 0 otherwise. From this it is clear that the expected numbers of payments are equal if $\lambda_i = q_i$ is taken:

$$\lambda = \sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} q_i. \tag{3.65}$$

Also, by (3.62) and (3.63), the expectations of $\widetilde{S}$ and S are then equal:

$$\mathrm{E}[\widetilde{S}] = \sum_{i=1}^{n} q_i b_i = \mathrm{E}[S]. \tag{3.66}$$

For the variances of S and $\widetilde{S}$ we have

$$\mathrm{Var}[S] = \sum_{i=1}^{n} q_i b_i^2; \qquad \mathrm{Var}[\widetilde{S}] = \sum_{i=1}^{n} q_i(1-q_i) b_i^2 = \mathrm{Var}[S] - \sum_{i=1}^{n} (q_i b_i)^2. \tag{3.67}$$

We see that S has a larger variance. If $\lambda_i = q_i$ then using a collective model results in risk averse decision makers tending to take more conservative decisions, see further Chapter 7. Also notice that the smaller $\sum_{i=1}^{n} (q_i b_i)^2$ is, the less the collective model will differ from the individual model.

Remark 3.8.1 (*The* collective model)
By *the* collective model for a portfolio, we mean a compound Poisson distribution as in (3.64) with $\lambda_i = q_i$. We also call it the *canonical* collective approximation.

In Exercise 3.8.3 we show that in the situation (3.62), *the* collective model can be obtained as well by replacing each claim X_i by a Poisson(1) number of independent claims with the same distribution as X_i. We can also do this if the random variables X_i are more general than those in (3.62). For example, assume that contract i produces claims $b_0 = 0, b_1, b_2, \ldots, b_n$ with probabilities $p_0, p_1, \ldots, p_n$. Since X_i equals exactly one of these values, we can write

$$X_i \equiv I_0 b_0 + I_1 b_1 + \cdots + I_n b_n, \tag{3.68}$$

with $I_j = 1$ if $X_i = b_j$, zero otherwise. So $\Pr[I_j = 1] = p_j$ for the marginal distributions of I_j, and their joint distribution is such that $I_0 + I_1 + \cdots + I_n \equiv 1$. One can show that if we choose the canonical collective model, we actually replace X_i by the compound Poisson distributed random variable Y_i, with

$$Y_i = N_0 b_0 + N_1 b_1 + \cdots + N_n b_n, \tag{3.69}$$

where the N_j are independent Poisson(p_j) random variables. In this way, the expected frequencies of all claim sizes remain unchanged. ∇

Remark 3.8.2 (Model for an *open* portfolio)
The second proposed model with $\lambda_i = -\log(1-q_i)$ can be used to model an open portfolio, with entries and exits not on renewal times. Assume that in a certain policy the waiting time W until death has an exponential(β) distribution. For the probability of no claims to be $1-q$, we must have $\Pr[W > 1] = 1-q$, so $\beta = -\log(1-q)$. Now assume that, at the moment of death, each time we replace this policy by an identical one. Thus, we have indeed an open model for our portfolio. The waiting

times until death are always exponentially(β) distributed. But from the theory of Poisson processes, see also Exercise 2.5.7, we know that the number of deaths before time 1 is Poisson(β) distributed. In this model, we in fact replace, for each i, the ith policy by a Poisson($-\log(1-q_i)$) distributed number of copies. Since $I_i \sim \min\{N_i, 1\}$, the *open* collective model we get this way is a safe approximation to the individual model, as it allows for more claims per policy than one. See also Section 7.4.1. ∇

Remark 3.8.3 (Negative risk amounts)
If we assume that the b_i are positive integers, then we can quickly calculate the probabilities for S, and consequently quickly approximate those for $\tilde{S}$, with Panjer's recursion. But if the b_i can be negative as well as positive, we cannot use this recursion. In that case, we can split up S in two parts $S = S^+ - S^-$ where S^+ is the sum of the terms Y_i in (3.63) with $b_i \geq 0$. By Theorem 3.4.2, S^+ and S^- are independent compound Poisson random variables with non-negative terms. The cdf of S can then be found by convolution of those of S^+ and S^-.

To find the stop-loss premium $\mathrm{E}[(S-d)_+]$ for only one value of d, the convolution of S^+ and S^- is not needed. Conditioning on the total S^- of the negative claims, we can rewrite the stop-loss premium as follows:

$$\mathrm{E}[(S-d)_+] = \sum_{x\geq 0} \mathrm{E}[(S^+ - (x+d))_+]\,\Pr[S^- = x]. \tag{3.70}$$

To calculate this we only need the stop-loss premiums of S^+, which follow as a by-product of Panjer's recursion, see Example 3.5.4. Then the desired stop-loss premium can be calculated with a simple summation. For the convolution, a double summation is necessary, or it could be handled through the use of `convolve`.

Note that the FFT-technique, see Example 3.6.2, is not restricted to non-negative claim amounts. ∇

3.9 Loss distributions: properties, estimation, sampling

In a compound model for losses, we have to specify both the claim number distribution and the claim severity distribution. For the former we often take the Poisson distribution, such as in the canonical or the open collective model, or when the assumptions of a Poisson process apply, see Chapter 4. In case of overdispersion, due to parameter uncertainty or cumulation of events, see Examples 3.3.1 and 3.3.2, we might use the negative binomial distribution. For some purposes, for example to compute premium reductions in case of a deductible, it is convenient to use a parametric distribution that fits the observed severity distribution well. Depending on the type of insurance at hand, candidates may vary from light tailed (for example the Gaussian distribution) to very heavy-tailed (Pareto). In this section we will present some severity distributions, explain their properties and suggest when to use them. We use maximum likelihood to estimate parameters. Often it is useful to generate

pseudo-random samples from the loss distribution, for example if we want to compute the financial consequences of applying some risk management instrument like a complicated reinsurance scheme.

3.9.1 Techniques to generate pseudo-random samples

For many distributions, pseudo-random samples may be drawn by using standard `R` functions. They often use the *inversion method*, also known as the probability integral transform. It is based on the fact that if $U \sim \text{uniform}(0,1)$, then $F^{-1}(U) \sim F$ because $\Pr[F^{-1}(U) \leq x] = \Pr[U \leq F(x)] = F(x)$. For example the function `rnorm` in its standard mode applies the inverse normal cdf `qnorm` to results of `runif`, see also Appendix A.

The function `runif` to generate uniform pseudo-random numbers in `R` is state-of-the-art. Its default method is Mersenne-Twister, described in `R`'s help-files as a "twisted generalized feedback shift register algorithm with period $2^{19937}-1$ and equidistribution in 623 consecutive dimensions (over the whole period)".

Another sampling method is the *rejection method*. Suppose that it is hard to sample from density $f(\cdot)$, but that an easier to handle distribution $g(\cdot)$ exists satisfying $f(x) \leq kg(x)\ \forall x$ for some appropriate bound $k \geq 1$. We get a random outcome from $f(\cdot)$ by sampling a point uniformly from the area below the graph of $f(x)$, and taking its x-coordinate; the probability of an outcome x or less is then just $F(x)$. John von Neumann's (1951) idea was not to do this directly, but to sample a random point below the graph of $kg(x)$, drawing its x-coordinate using, in most cases, the inversion method, and then its y-coordinate uniformly from $(0, kg(x))$. The x-coordinate is accepted as an outcome from $f(\cdot)$ if this random point happens to be under $f(x)$ as well. If `rg()` produces a random drawing from $g(x)$ and $f(x)/g(x) \leq k$, an `R`-program to draw x randomly from $f(\cdot)$ could be as follows:

```
repeat {x <- rg(); if (runif(1) < f(x)/k/g(x)) break}
```

The number of points rejected is a geometric$(1/k)$ random variable, so the smaller k, the faster the random number generation.

In many cases, we can construct drawings from given distributions by using the fact that they are a simple transformation of other random variables for which a standard `R`-function `r...` exists to produce pseudo-random values. See Exercise 3.9.24 for some applications of this, including sampling from (log)normal, Pareto, Erlang, Weibull or Gompertz distributions.

In mixed models, one may first draw from the distribution of the conditioning random variable (structure variable), and next from the conditional distribution of the random variable of interest.

When the cdf $F(x) = G(x)H(x)$ for simple G and H, or the survival function is $1-F(x) = (1-G(x))(1-H(x))$, a random value from F is produced by taking the maximum (minimum) of independent random variables distributed as G and

H. See Exercise 3.9.17 where this device is employed to sample from a Makeham distributed lifetime random variable.

3.9.2 Techniques to compute ML-estimates

We aim to obtain estimates for $\alpha, \beta, \ldots$ by maximizing the (log-)likelihood

$$\ell(\alpha, \beta, \ldots; \vec{y}) = \log \prod f_{Y_i}(y_i; \alpha, \beta, \ldots). \tag{3.71}$$

A a first step, we inspect the *normal equations* $\partial \ell / \partial \alpha = 0$, $\partial \ell / \partial \beta = 0, \ldots$ to see if they admit an explicit solution. This is the case for normal, lognormal, Poisson and inverse Gaussian samples. For Pareto, the optimal solution is explicit but it is not the solution to the normal equations. It may also happen that the normal equations provide a partial solution, often in the sense that the optimal parameters must be related in such a way that the fitted mean coincides with the sample mean. The advantage of this is that it reduces the dimension of the maximization problem to be solved; one of the parameters may be substituted away. This occurs with the negative binomial and the gamma distributions. If only one normal equation remains to be solved, this can be done using the `R` function `uniroot`, which is both reliable and fast. Alternatively, the optimization can be done using `optimize` to search an interval for a minimum or maximum of a real function. If needed, one can use `optim` for optimization in more dimensions.

3.9.3 Poisson claim number distribution

Not just by the Poisson process often underlying rare events, see Chapter 4, but also by the mere fact that it has only one parameter to estimate, the Poisson distribution is an attractive candidate for the claim number of the compound total loss. For a Poisson(λ) sample $Y_1 = y_1, \ldots, Y_n = y_n$, the loglikelihood is

$$\ell(\lambda; \vec{y}) = \log \prod f_{Y_i}(y_i; \lambda) = -n\lambda + \sum y_i \log \lambda - \sum \log n_i!, \tag{3.72}$$

which gives $\widehat{\lambda} = \overline{Y}$ as the maximum likelihood estimator of λ.

In insurance situations, often the numbers of claims pertain to policies that were not in force during a full calendar year, but only a known fraction of it. We denote this exposure for policy i by w_i. In that case, it follows from the properties of a Poisson process, see also Chapter 4, that the number of claims for policy i has a Poisson(λw_i) distribution. Therefore the loglikelihood is

$$\begin{aligned} \ell(\lambda; \vec{y}, \vec{w}) &= \log \prod f_{Y_i}(y_i; \lambda w_i) \\ &= -\sum \lambda w_i + \sum y_i \log \lambda + \sum y_i \log w_i - \sum \log n_i!, \end{aligned} \tag{3.73}$$

so in this case $\widehat{\lambda} = \sum Y_i / \sum w_i$ is the maximum likelihood estimator of λ, that is, the number of claims divided by the total exposure.

If we consider Y_i/w_i, the number of claims per unit of exposure, one claim in a contract that was insured for only nine months counts as $\frac{4}{3}$. Then $\widehat{\lambda}$ is the *weighted* average of these quantities, with weights $w_i / \sum w_i$. In practice, however, often simply a straight average of the number of claims per unit of exposure is taken.

Random sampling from the Poisson distribution is achieved by calling `rpois`. See Exercise 3.9.17 for a way to do this random sampling by the use of (3.11), which relates numbers of events to exponential waiting times.

3.9.4 Negative binomial claim number distribution

We have seen in Section 3.3 that sometimes it is proper to use a claim number distribution more spread than the Poisson, the variance of which equals the mean. In fact, both parameter uncertainty and cumulation in a Poisson process may lead to a negative binomial(r,p) claim number, see Examples 3.3.1 and 3.3.2, with overdispersion factor $\mathrm{Var}[N]/\mathrm{E}[N] = 1/p$. On the basis of a sample of outcomes $Y_1 = y_1, \ldots, Y_n = y_n$, we want to estimate the parameters r, p by maximum likelihood. In Table A one sees that the corresponding density equals

$$f_Y(y;\, r,p) = \binom{r+y-1}{y} p^r (1-p)^y, \quad y = 0,1,\ldots \tag{3.74}$$

For non-integer r, the binomial coefficient is defined using gamma-functions, with $x! = \Gamma(x+1)$ for all real $x > 0$.

Now, just as in Theorem 3.4.2, let N_j, $j = 0,1,\ldots$, count the number of times a sample element Y_i in the sample equals the value j. Observe that $\sum N_j \equiv n$ holds, while $\sum Y_i \equiv \sum j N_j$. In the loglikelihood of the total sample we find logarithms of the factorials in the density (3.74), as well as of the factorials arising because the places i in the sample where $Y_i = j$ occurs may be chosen arbitrarily, see (3.25). The latter are constant with respect to r and p, so we ignore them, and get

$$\begin{aligned} \ell(r,p;\, \vec{y}) &= \log \prod_{j=0}^{\infty} \{f_Y(j;\, r,p)\}^{n_j} + \ldots \\ &= \sum_j n_j \{\log(r+j-1) + \cdots + \log r - \log j!\} \\ &\quad + rn \log p + \sum_j j n_j \log(1-p) + \ldots \end{aligned} \tag{3.75}$$

Note that the outcomes $y_1, \ldots, y_n$ do not carry more information about the parameters r, p than do their frequencies $n_0, n_1, \ldots$ Apart from the order in which the iid observations occurred, the sample can be reconstructed from the frequencies. The conditional joint density of the sample, given the frequencies, no longer depends on

the parameters r and p, or in other words, these frequencies are sufficient statistics. The iterative computation of the ML-estimates using the frequencies is of course faster for large integer-valued samples.

We first look at the partial derivative $\partial\ell/\partial p$. It must be zero for those r, p for which the likelihood is maximal:

$$0 = \frac{\partial \ell}{\partial p} = \frac{rn}{p} - \frac{\sum j n_j}{1-p} \iff \frac{r(1-p)}{p} = \overline{y} \iff p = \frac{r}{r+\overline{y}}, \tag{3.76}$$

This equation expresses that the mean of the ML-estimated density equals the observed sample mean.

The second ML-equation $\partial\ell/\partial r = 0$ results in

$$0 = \sum_{j=1}^{\infty} n_j \left(\frac{1}{r} + \cdots + \frac{1}{r+j-1} \right) + n \log p. \tag{3.77}$$

From this, no explicit expression for the ML-estimators can be derived. But substituting $p = r/(r+\overline{y})$ from (3.76) into (3.77) results in a one-dimensional equation for r that can be solved numerically using `R`. One can also use `R`'s function `optimize` to do the maximization over r, or simply use `optim`; see below.

About generating negative binomial(r, p) random samples, see Table A, we first remark that in elementary probability texts, such random variables are introduced as the number of failures before the rth success in a sequence of Bernoulli trials, or sometimes as the number of trials needed, successes and failures combined. This requires r to be integer. For the general case, instead of the standard function `rnbinom` we can use the mixed model of Example 3.3.1 to generate negative binomial(r, p) outcomes, by first drawing the parameter from a suitable gamma distribution and next drawing from a Poisson distribution with that parameter. The following `R` commands draw a negative binomial sample, count the frequencies of each outcome and compare these with the theoretical frequencies:

```
set.seed(1); n <- 2000; r <- 2; p <- 0.5
hh <- rpois(n, rgamma(n,r,p/(1-p)))
n.j <- tabulate(1+hh); j <- 0:max(hh)
rbind(n.j, round(dnbinom(j,r,p)*n))
```

In the first line, we initialize the random number generator so as to be able to reproduce the results. The second line draws a sample using the method suggested. The function `tabulate` counts the frequencies of the numbers $1, 2, \ldots$ in its argument. We added 1 to each element of the sample to include the frequency of 0 as well. Running this script one sees that the observed frequencies match the theoretical ones, computed using `dnbinom`, quite closely.

To get initial estimates for the parameters, we use the method of moments:

$$\frac{r(1-p)}{p} = \overline{y}; \quad \frac{r(1-p)}{p^2} = \overline{y^2} - \overline{y}^2. \tag{3.78}$$

Solving these equations for r and p is done by dividing the first by the second:

```
y.bar <- sum(j*n.j/n); y2.bar <- sum(j^2*n.j/n)
p0 <- y.bar/(y2.bar-y.bar^2); r0 <- p0 * y.bar/(1-p0)
```

For how to solve the second ML-equation (3.77) using `uniroot` and inserting (3.76), see Exercise 3.9.26. To maximize the loglikelihood using the `optimize` function of `R`, the loglikelihood is computed by specifying `log=T` in the density and taking a weighted sum with as weights the frequencies n_j for all $j = 0, 1, \ldots$:

```
g <- function (r) {sum(dnbinom(j,r,r/(r+y.bar),log=T)*n.j)}
r <- optimize(g, c(r0/2, 2*r0), max=T, tol=1e-12)$maximum
p <- r/(r+y.bar)
```

We get $\widehat{r} = 1.919248$ and $\widehat{p} = 0.4883274$.

By using the general `R` function `optim`, we no not have to rely on the fact that in the optimum, $p = r/(r+\bar{y})$ must hold because of (3.76):

```
h <- function (x) {-sum(dnbinom(j,x[1],x[2],log=T)*n.j)}
optim(c(r0,p0), h, control=list(reltol=1e-14))
```

The first argument of `optim` is a vector of starting values, the second the function `h(x)` with `x` the vector of parameters. The first element of `x` represents r, the second is p. The `control` argument contains a list of refinements of the optimization process; to get the exact same results as before, we have set the relative tolerance to 10^{-14}. Note the minus-sign in the definition of `h`, needed because the standard mode of `optim` is to minimize a function.

3.9.5 Gamma claim severity distributions

The gamma(α, β) distribution can be used to model non-negative losses if the tail of the cdf is not too 'heavy', such as in motor insurance for damage to the own vehicle. Density, cdf, quantiles and random deviates for this distribution are given by the standard `R` functions `dgamma`, `pgamma`, `qgamma` and `rgamma`, respectively. Note that the parameter α corresponds to the `shape` parameter of these functions, while β corresponds to `rate`, $1/\beta$ to `scale`. See Table A.

To find maximum likelihood estimators for the parameters α, β on the basis of a random sample $Y_1 = y_1, \ldots, Y_n = y_n$ from a gamma(α, β) distribution, proceed as follows. The loglikelihood $\ell(\alpha, \beta)$ of the parameters is given by

$$\ell(\alpha, \beta; \vec{y}) = \log \prod_{i=1}^{n} f_Y(y_i;\, \alpha, \beta). \tag{3.79}$$

Filling in the gamma density $f_Y(y) = \frac{1}{\Gamma(\alpha)} \beta^{\alpha} y^{\alpha-1} \mathrm{e}^{-\beta y}$, we get

$$\ell(\alpha, \beta) = n\alpha \log \beta - n \log \Gamma(\alpha) + (\alpha - 1) \log \prod y_i - \beta \sum y_i. \tag{3.80}$$

One of the ML-equations ensures again that the fitted mean is the observed mean:

$$\frac{\partial \ell}{\partial \beta} = \frac{n\alpha}{\beta} - \sum y_i = 0 \iff \widehat{\beta} = \frac{\widehat{\alpha}}{\overline{y}}. \tag{3.81}$$

Writing $\overline{\log y}$ for the mean of the logarithms of the observations, we see from the other one that there is no explicit solution:

$$\begin{aligned} \frac{\partial \ell}{\partial \alpha} = n\log\beta - n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} + \sum \log y_i = 0 \iff \\ \log\widehat{\alpha} - \frac{\Gamma'(\widehat{\alpha})}{\Gamma(\widehat{\alpha})} - \log\overline{y} + \overline{\log y} = 0. \end{aligned} \tag{3.82}$$

Using the R-function `digamma()` to compute the digamma (psi) function defined as $\Gamma'(\alpha)/\Gamma(\alpha)$, the solution can be found like this:

```
set.seed(2525); y <- rgamma(2000, shape=5, rate=1)
aux <- log(mean(y)) - mean(log(y))
f <- function(x) log(x) - digamma(x) - aux
alpha <- uniroot(f, c(1e-8,1e8))$root  ##  5.049
beta <- alpha/mean(y)                  ##  1.024
```

The interval $(10^{-8}, 10^{8})$ in which a zero for `f` above is sought covers skewnesses from $+2\times10^{-4}$ to $+2\times10^{4}$. The function `f(x)` decreases from `f(0)` $= +\infty$ to `f(`∞`)` < 0. This is because $\log(x) - \Gamma'(x)/\Gamma(x)$ decreases to 0, and, as is proved in Exercise 9.3.12, `aux` is strictly positive unless all y_i are equal.

Just as for the negative binomial distribution, optimal $\widehat{\alpha}$ and $\widehat{\beta}$ can also be found by using `optimize` and $\widehat{\beta} = \widehat{\alpha}/\overline{y}$, or `optim`.

3.9.6 Inverse Gaussian claim severity distributions

A distribution that sometimes appears in the actuarial literature, for several purposes, is the inverse Gaussian (IG). Its properties resemble those of the gamma and lognormal distributions. Its name derives from the inverse relationship that exists between the cumulant generating functions of these distributions and those of Gaussian distributions, see (3.85) below. Various parameterizations are in use. Just like with the gamma distribution, we will use a shape parameter α and a scale parameter β. See also Table A.

The probability density function of the IG distribution is:

$$f(x;\alpha,\beta) = \frac{\alpha}{\sqrt{2\pi\beta}} x^{-\frac{3}{2}} \mathrm{e}^{-\frac{(\alpha-\beta x)^2}{2\beta x}}, \quad x > 0. \tag{3.83}$$

The main reason the IG distribution has never gained much popularity is because it is not easy to manage mathematically. Indeed to prove that the density integrates to one is not at all trivial without knowing the corresponding cdf, which is:

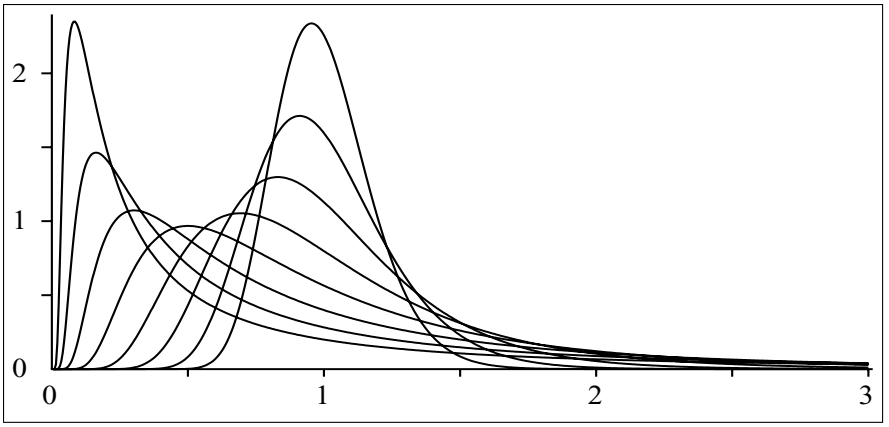

Fig. 3.1 Inverse Gaussian densities for $\alpha = \beta = \frac{1}{4}, \frac{1}{2}$,1,2,4,8,16,32 (tops from left to right).

$$F(x;\alpha,\beta) = \Phi\Big(\frac{-\alpha}{\sqrt{\beta x}} + \sqrt{\beta x}\Big) + \mathrm{e}^{2\alpha}\Phi\Big(\frac{-\alpha}{\sqrt{\beta x}} - \sqrt{\beta x}\Big), \quad x > 0. \tag{3.84}$$

This function has $\lim_{x\downarrow 0} F(x;\alpha,\beta) = 0$, $\lim_{x\to\infty} F(x;\alpha,\beta) = 1$ and its derivative is (3.83), which is non-negative on $(0,\infty)$. So (3.84) is an absolutely continuous cdf, and (3.83) is its density. Detailed proofs are asked in Exercise 3.9.3.

Using the fact that (3.83) is a density to compute the resulting integral, we can prove that the mgf equals

$$m(t;\alpha,\beta) = \exp\Big\{\alpha\Big[1 - \sqrt{1 - 2t/\beta}\Big]\Big\}, \quad t \le \frac{\beta}{2}. \tag{3.85}$$

Notice that the mgf is finite for $t = \beta/2$, but not for $t > \beta/2$.

The special case with $\alpha = \beta$ is also known as the Wald distribution. From the mgf one easily sees that β is indeed a scale parameter, since βX is inverse Gaussian$(\alpha,1)$ if $X \sim$ inverse Gaussian(α,β). We also see that adding two independent inverse Gaussian distributed random variables, with parameters α_1,β and α_2,β, yields an inverse Gaussian random variable with parameters $\alpha_1+\alpha_2,\beta$. The expected value and the variance are α/β and α/β^2 respectively, just as for the gamma distribution. The easiest way to show this is by taking a series expansion of the cgf. The skewness is $3/\sqrt{\alpha}$, as opposed to $2/\sqrt{\alpha}$ for a gamma distribution with the same mean and variance. The flexibility of the inverse Gaussian distributions, from very skew to almost normal, is illustrated in Figure 3.1. All depicted distributions have the same mean $\alpha/\beta = 1$, and a strictly positive mode.

For the inverse Gaussian(α,β) distribution, the loglikelihood is

$$\ell(\alpha,\beta;\vec{y}) = \sum\left(\log\frac{\alpha}{\sqrt{2\pi\beta}} - \frac{3}{2}\log y_i - \frac{\alpha^2}{2\beta y_i} + \alpha - \frac{\beta y_i}{2}\right). \tag{3.86}$$

Setting the partial derivatives equal to zero gives

$$\begin{aligned}\frac{\partial \ell}{\partial \alpha} &= \sum\left(\frac{1}{\alpha} - \frac{\alpha}{\beta y_i} + 1\right) = 0;\\ \frac{\partial \ell}{\partial \beta} &= \sum\left(\frac{-1}{2\beta} + \frac{\alpha^2}{2\beta^2 y_i} - \frac{y_i}{2}\right) = 0.\end{aligned} \tag{3.87}$$

Writing $\overline{1/y}$ for the average of the numbers $1/y_i$, we can rewrite this as

$$\frac{1}{\alpha} + 1 = \frac{\alpha}{\beta}\,\overline{1/y} \quad \text{and} \quad \frac{1}{\alpha} + \frac{\beta}{\alpha}\overline{y} = \frac{\alpha}{\beta}\,\overline{1/y}, \tag{3.88}$$

so we get very simple explicit expressions for the ML parameter estimates:

$$\widehat{\alpha} = \frac{1}{\overline{y}\,\overline{1/y} - 1} \quad \text{and} \quad \widehat{\beta} = \frac{\widehat{\alpha}}{\overline{y}}. \tag{3.89}$$

The second equation ensures that, again, the mean of the fitted distribution equals the sample mean. It is an easy exercise to show that in case of a sample $Y_1, \ldots, Y_n$, the quantities $\sum Y_i$ and $\sum 1/Y_i$ are sufficient statistics.

In the package `statmod` extending `R` one finds functions `d/p/q/rinvgauss` for density, cdf, quantiles and random number generation with this distribution. Similar functions can be found in `SuppDist` (then with capital `G`). The parameters used there are not the same we use; see also Exercise 3.9.12. They are the mean $\mu = \alpha/\beta$ and a precision parameter λ. The latter is taken in such a way that the variance α/β^2 equals μ^3/λ, therefore $\lambda = \alpha^2/\beta$. Conversely, $\alpha = \lambda/\mu$ and $\beta = \lambda/\mu^2$. Generating a random sample and estimating the parameters from it by maximum likelihood, using (3.89), then goes as follows:

```
library(statmod); set.seed(2525)
y <- rinvgauss(2000, mu=5, lambda=3)
alpha <- 1/(mean(y)*mean(1/y)-1); beta <- alpha/mean(y)
```

We get $\widehat{\alpha} = 0.626$; $\widehat{\beta} = 0.128$. So, $\widehat{\mu} = \widehat{\alpha}/\widehat{\beta} = 4.89$ and $\widehat{\lambda} = \widehat{\alpha}^2/\widehat{\beta} = 3.06$. The true values $\mu = \alpha/\beta = 5$, $\lambda = \alpha^2/\beta = 3$ give $\alpha = 0.6$, $\beta = 0.12$. In Exercise 3.9.12, the reader is asked to verify that we always get feasible estimates $\widehat{\alpha} > 0$ and $\widehat{\beta} > 0$ this way, in other words, that $\overline{y}\,\overline{1/y} > 1$ must hold in (3.89).

3.9.7 Mixtures/combinations of exponential distributions

Another useful class of parametric claim severity distributions, especially in the context of ruin theory (Chapter 4), consists of mixtures/combinations of exponential distributions. A *mixture* arises if the parameter of an exponential distribution is a random variable that is α with probability q and β with probability $1-q$. The density is then given by

$$p(x) = q\alpha e^{-\alpha x} + (1-q)\beta e^{-\beta x}, \quad x > 0. \tag{3.90}$$

For each q with $0 \le q \le 1$, the function $p(\cdot)$ is a probability density function. But also for $q < 0$ or $q > 1$, $p(\cdot)$ in (3.90) is sometimes a pdf. Since $\int p(x)\,dx = 1$ always holds, we only have to check if $p(x) \ge 0$ for all x. From Exercise 3.9.4, we learn that it suffices to check $p(0) \ge 0$ and $p(\infty) \ge 0$. Assuming $\alpha < \beta$, this holds if $1 < q \le \beta/(\beta-\alpha)$, and in this case (3.90) is called a *combination* of exponential distributions.

An example of a proper combination of exponential distributions is given by

$$p(x) = 2(e^{-x} - e^{-2x}) = 2 \times 1e^{-1x} - 1 \times 2e^{-2x}, \tag{3.91}$$

which has $q = 2$, $\alpha = 1$ and $\beta = 2$. A second example is the function

$$p(x) = \frac{4}{3}(e^{-x} - \frac{1}{2}e^{-2x}) = \frac{4}{3} \times 1e^{-1x} - \frac{1}{3} \times 2e^{-2x}. \tag{3.92}$$

If $X \sim$ exponential(α) and $Y \sim$ exponential(β), with $\alpha \neq \beta$, then

$$m_{X+Y}(t) = \frac{\alpha\beta}{(\alpha-t)(\beta-t)} = \frac{\beta}{\beta-\alpha}\frac{\alpha}{\alpha-t} - \frac{\alpha}{\beta-\alpha}\frac{\beta}{\beta-t}. \tag{3.93}$$

This is the mgf of density (3.90) with $q = \frac{\beta}{\beta-\alpha}$. So a sum of independent exponential random variables has a combination of exponential distributions as its density. The reverse is not always true: (3.91) is the pdf of the convolution of an exponential(1) and an exponential(2) distribution, since $q = \beta/(\beta-\alpha) = 2$, but the pdf (3.92) cannot be written as such a convolution.

If $\alpha \uparrow \beta$, then $\beta/(\beta-\alpha) \to \infty$, and $X+Y$ tends to a gamma(2,β) random variable. Hence, the gamma distributions with $r = 2$ are limits of densities that are combinations of exponential distributions, and the same holds for all gamma distributions with an integer shape parameter (so-called Erlang distributions).

A mixture of exponential distributions with parameters $0 < \alpha < \beta$ and $0 \le q \le 1$ can be generated using the urn-of-urns model $Z = IX/\alpha + (1-I)Y/\beta$, with X, Y and I independent, X and $Y \sim$ exponential(1) and $I \sim$ Bernoulli(q). There is also a two-stage model that produces *all* random variables with pdf (3.90). For this, let $I \sim$ Bernoulli(γ) with $0 \le \gamma \le 1$, and let $0 < \alpha < \beta$. Then

$$Z = I\frac{X}{\alpha} + \frac{Y}{\beta} \tag{3.94}$$

has as its mgf

$$m_Z(t) = \left(1 - \gamma + \gamma\frac{\alpha}{\alpha-t}\right)\frac{\beta}{\beta-t} = \frac{\alpha\beta - t\beta(1-\gamma)}{(\alpha-t)(\beta-t)}. \tag{3.95}$$

To show that this is the mgf of a combination or a mixture of exponential distributions, it suffices to find q, using partial fractions, such that (3.95) equals the mgf of

(3.90), which is

$$q\frac{\alpha}{\alpha-t}+(1-q)\frac{\beta}{\beta-t}. \tag{3.96}$$

Comparing (3.95) and (3.96) we see that $q\alpha+(1-q)\beta=\beta(1-\gamma)$, hence

$$q=\frac{\beta\gamma}{\beta-\alpha}. \tag{3.97}$$

Since $0<\alpha<\beta$, we have that $0\le q\le 1$ if $0\le\gamma\le 1-\alpha/\beta$, and then Z is mixture of exponential distributions. If $1-\alpha/\beta<\gamma\le 1$, then $q>1$, and Z is a combination of exponential distributions.

The loss Z in (3.94) can be viewed as the result of an experiment where one suffers a loss Y/β in any case and where it is decided by a trial with probability γ of success whether one loses an additional amount X/α. Another interpretation is that the loss is drawn from either Y/β or $X/\alpha+Y/\beta$, since $Z=I(X/\alpha+Y/\beta)+(1-I)Y/\beta$. If $\gamma=1$, again a sum of two exponential distributions arises.

Writing R-functions to compute the cdf $P(x)$ and the density $p(x)$ is trivial; quantiles $x=P^{-1}(u)$ with $0<u<1$ follow by (numerical) inversion of the cdf, hence solving $P(x)=u$ for x. For this, one may call the R-function `uniroot`. But to generate random values, simply use model (3.94):

```
set.seed(1); n <- 2000; q <- 1.5; alpha <- 1; beta <- 2
gam <- (beta-alpha)/beta * q
y <- rbinom(n,1,gam) * rexp(n)/alpha + rexp(n)/beta
```

One way to estimate the parameters of the joint likelihood $\prod p(y_i;q,\alpha,\beta)$, with $p(x)$ as in (3.90), is by using the method of moments. Not for every combination of first three sample moments feasible parameters of (3.90) can be found leading to the right mean, variance and skewness; for details, consult Babier and Chan (1992), and see also Example 4.9.1. ML-optimization requires a three-dimensional maximization that cannot be easily reduced in dimension such as the ones we encountered before. But it is easy to simply let `optim` do the work, for example by:

```
f <- function (y, q, alpha, beta){
  q * alpha * exp(-alpha*y) + (1-q) * beta * exp(-beta*y)}
h <- function (x) {-sum(log(f(y, x[1], x[2], x[3])))}
optim(c(0.8, 0.9, 1.8), h)
```

The resulting parameter estimates are $\widehat{q}=1.285$, $\widehat{\alpha}=0.941$, $\widehat{\beta}=2.362$.

3.9.8 Lognormal claim severities

Using that $X\sim\text{lognormal}(\mu,\sigma^2)$ if and only if $\log X\sim\text{N}(\mu,\sigma^2)$, for the cdf and the density of this claim severity distribution, by the chain rule we get

$$F_X(x)=\Phi(\log x;\mu,\sigma^2)\quad\text{and}\quad f_X(x)=\frac{1}{x}\varphi(\log x;\mu,\sigma^2), \tag{3.98}$$

with φ and Φ the normal density and cdf. See also Table A.

It is easy to write R functions `d/p/q/rlnorm` based on these relations with the normal distribution, but such functions are also included in the package `stats`.

To compare the fatness of the tail with the one of the inverse Gaussian(α,β) distribution with the same mean α/β and variance α/β^2, note that the lognormal skewness is $\gamma = (3+1/\alpha)/\sqrt{\alpha}$, see Table A, while the inverse Gaussian skewness is $\gamma = 3/\sqrt{\alpha}$. The lognormal distribution is suitable as a severity distribution for branches with moderately heavy-tailed claims, like fire insurance.

Maximum likelihood estimation in case of a lognormal(μ,σ^2) random sample is simple by reducing it to the normal case. If the sample is $Y_1 = y_1,\ldots,Y_n = y_n$, let $X_i = \log Y_i$, then $X_i \sim \mathrm{N}(\mu,\sigma^2)$, so as is well-known, $\widehat{\mu} = \overline{X}$ and $\widehat{\sigma}^2 = \overline{(X-\widehat{\mu})^2}$.

3.9.9 Pareto claim severities

The Pareto(α,x_0) distribution, see Table A, can be used for branches with high probability of large claims, notably liability insurance. In Exercise 3.9.1 it is proved that $Y \sim$ Pareto(α,x_0) is equivalent to $\log(Y/x_0) \sim$ exponential(α). With this property, it is easy to write the function `rpareto`; the other ones are trivial.

To compute ML-estimates for a Pareto(α,x_0) sample is slightly different from what we saw before, because in this case the optimal estimates cannot be produced by simply solving the ML equations. The loglikelihood with a Pareto(α,x_0) sample $Y_1 = y_1,\ldots,Y_n = y_n$ is, if $y_{(1)} := \min(y_i)$ denotes the sample minimum:

$$\ell(\alpha,x_0;\vec{y}) = \begin{cases} n\log\alpha + \alpha n\log x_0 - (\alpha+1)\sum\log y_i & \text{if } x_0 \le y_{(1)}; \\ -\infty & \text{if } x_0 > y_{(1)}. \end{cases} \tag{3.99}$$

For each choice of α, we have $\ell(\alpha,x_0) \le \ell(\alpha,y_{(1)})$, so the ML estimate for x_0 must be $\widehat{x_0} = y_{(1)}$. Further,

$$\begin{aligned} \frac{\partial\ell(\alpha,\widehat{x_0})}{\partial\alpha} = 0 &\iff \frac{n}{\alpha} + n\log\widehat{x_0} - \sum\log y_i = 0 \\ &\iff \widehat{\alpha} = \Big(\frac{1}{n}\sum\log(y_i/\widehat{x_0})\Big)^{-1}. \end{aligned} \tag{3.100}$$

The shape of the ML estimate $\widehat{\alpha}$ is not surprising, since the transformed sample $X_i = \log(Y_i/x_0)$ is distributed as exponential(α), for which case the ML estimate of α is known to be $1/\overline{X}$. For the same reason, a random Pareto(α,x_0) sample can be generated by multiplying the exponents of an exponential(α) random sample by x_0. So to draw a sample and find ML estimates of α,x_0, do

```
set.seed(2525); x0 <- 100; alpha <- 2; n <- 2000
y <- x0*exp(rexp(n)/alpha)
x0.hat <- min(y); alpha.hat <- 1/mean(log(y/x0.hat))
```

The resulting ML-estimates are $\widehat{x}_0 = 100.024$ and $\widehat{\alpha} = 2.0014$. Determining estimates for Pareto samples by the method of moments might present problems since the population moments $\mathrm{E}[Y^j]$ only exist for powers $j < \alpha$.

3.10 Stop-loss insurance and approximations

The payment by a reinsurer in case of a stop-loss reinsurance with retention d for a loss S is equal to $(S-d)_+$. In this section we look for analytical expressions for the net stop-loss premium for some distributions. Note that expressions for stop-loss premiums can also be used to calculate net excess of loss premiums.

If $\pi(d)$ denotes the stop-loss premium for a loss with cdf $F(\cdot)$ as a function of d, then $\pi'(d+0) = F(d) - 1$. This fact can be used to verify the expressions for stop-loss premiums. For the necessary integrations, we often use partial integration.

Example 3.10.1 (Stop-loss premiums for the normal distribution)
If $X \sim \mathrm{N}(\mu, \sigma^2)$, what is the stop-loss premium for X if the retention is d?

As always for non-standard normal distributions, it is convenient to consider the case $\mu = 0$ and $\sigma^2 = 1$ first, and then use the fact that if $U \sim \mathrm{N}(0,1)$, then $X = \sigma U + \mu \sim \mathrm{N}(\mu, \sigma^2)$. The required stop-loss premium follows from

$$\mathrm{E}[(X-d)_+] = \mathrm{E}[(\sigma U + \mu - d)_+] = \sigma \mathrm{E}\left[\left(U - \frac{d-\mu}{\sigma}\right)_+\right]. \tag{3.101}$$

Since $\varphi'(u) = -u\varphi(u)$, we have the following relation

$$\int_t^\infty u\varphi(u)\,\mathrm{d}u = \int_t^\infty [-\varphi'(u)]\,\mathrm{d}u = \varphi(t). \tag{3.102}$$

It immediately follows that

$$\pi(t) = \mathrm{E}[(U-t)_+] = \varphi(t) - t[1 - \Phi(t)], \tag{3.103}$$

and hence

$$\mathrm{E}[(X-d)_+] = \sigma\varphi\left(\frac{d-\mu}{\sigma}\right) - (d-\mu)\left[1 - \Phi\left(\frac{d-\mu}{\sigma}\right)\right]. \tag{3.104}$$

For a table with a number of stop-loss premiums for the standard normal distribution, we refer to Example 3.10.5 below. See also Table C at the end of this book. ∇

Example 3.10.2 (Gamma distribution)
Another distribution that has a rather simple expression for the stop-loss premium is the gamma distribution. If $S \sim \text{gamma}(\alpha, \beta)$ and $G(\cdot; \alpha, \beta)$ denotes the cdf of S, then it can be shown that

$$\mathrm{E}[(S-d)_+] = \frac{\alpha}{\beta}[1 - G(d; \alpha+1, \beta)] - d[1 - G(d; \alpha, \beta)]. \tag{3.105}$$

We can also derive expressions for the higher moments of the stop-loss payment $\mathrm{E}[(S-d)_+^k]$, $k = 2,3,\ldots$. Even the mgf can be calculated analogously, and consequently also exponential premiums for the stop-loss payment. ∇

Remark 3.10.3 (Moments of the retained loss)
Since either $S \le d$, so $(S-d)_+ = 0$, or $S > d$, so $(S-d)_+ = S-d$, the following equivalence holds in general:

$$[(S-d)-(S-d)_+][(S-d)_+] \equiv 0. \tag{3.106}$$

With this, we can derive the moments of the retained loss $S-(S-d)_+$ from those of the stop-loss payment, using the equivalence

$$\begin{aligned} \{S-d\}^k &\equiv \{[S-d-(S-d)_+]+(S-d)_+\}^k \\ &\equiv \{S-d-(S-d)_+\}^k + \{(S-d)_+\}^k. \end{aligned} \tag{3.107}$$

This holds since, due to (3.106), the remaining terms in the binomial expansion vanish. ∇

In this way, if the loss approximately follows a translated gamma distribution, one can approximate the expected value, the variance and the skewness of the retained loss. See Exercise 3.10.4.

Example 3.10.4 (Stop-loss premiums approximated by NP)
The probabilities of $X > y$ for some random variable can be approximated quite well with the NP approximation. Is it possible to derive an approximation for the stop-loss premium for X too?

Define the following auxiliary functions for $u \ge 1$ and $y \ge 1$:

$$q(u) = u + \frac{\gamma}{6}(u^2-1) \quad \text{and} \quad w(y) = \sqrt{\frac{9}{\gamma^2} + \frac{6y}{\gamma} + 1} - \frac{3}{\gamma}. \tag{3.108}$$

From Section 2.5 we recall that $w(q(u)) = u$ and $q(w(y)) = y$. Furthermore, $q(\cdot)$ and $w(\cdot)$ are monotonically increasing, and $q(u) \ge y$ if and only if $w(y) \le u$. Let Z be a random variable with expected value 0, standard deviation 1 and skewness $\gamma > 0$. We will derive the stop-loss premiums of random variables X with $\mathrm{E}[X] = \mu$, $\mathrm{Var}[X] = \sigma^2$ and skewness γ from those of Z with the help of (3.101).

The NP approximation (2.59) states that

$$\Pr[Z > q(u)] = \Pr[w(Z) > u] \approx 1 - \Phi(u) \quad \text{if } u \ge 1. \tag{3.109}$$

Assume that $U \sim \mathrm{N}(0,1)$ and define $V = q(U)$ if $U \ge 1$, $V = 1$ otherwise, so $V = q(\max\{U,1\})$. Then

$$\Pr[V > q(u)] = \Pr[U > u] = 1 - \Phi(u), \quad u \ge 1. \tag{3.110}$$

Hence

$$\Pr[Z > y] \approx \Pr[V > y] = 1 - \Phi(w(y)), \quad y \ge 1. \tag{3.111}$$

The stop-loss premium of Z in $d > 1$ can be approximated through the stop-loss premium of V, since

$$\begin{aligned}\int_d^\infty \Pr[Z > y]\,\mathrm{d}y &\approx \int_d^\infty \Pr[V > y]\,\mathrm{d}y = \mathrm{E}[(V-d)_+] \\ &= \int_{-\infty}^{\infty} (q(\max\{u,1\}) - d)_+ \varphi(u)\,\mathrm{d}u \\ &= \int_{w(d)}^{\infty} (q(u) - d)\varphi(u)\,\mathrm{d}u.\end{aligned} \tag{3.112}$$

To calculate this integral, we use that $\frac{\mathrm{d}}{\mathrm{d}u}[u\varphi(u)] = (1-u^2)\varphi(u)$, so

$$\int_t^\infty [u^2 - 1]\varphi(u)\,\mathrm{d}u = t\varphi(t). \tag{3.113}$$

Substituting (3.102) and (3.113) and the function $q(\cdot)$ into (3.112) yields

$$\begin{aligned}\mathrm{E}[(Z-d)_+] &\approx \int_{w(d)}^\infty \left(u + \frac{\gamma}{6}(u^2-1) - d\right)\varphi(u)\,\mathrm{d}u \\ &= \varphi(w(d)) + \frac{\gamma}{6} w(d)\varphi(w(d)) - d[1 - \Phi(w(d))]\end{aligned} \tag{3.114}$$

as an approximation for the net stop-loss premium for any risk Z with mean 0, variance 1 and skewness γ. ∇

Example 3.10.5 (Comparing various approximations of stop-loss premiums) What are approximately the stop-loss premiums for X with $\mathrm{E}[X] = \mu = 0$, $\mathrm{Var}[X] = \sigma^2 = 1$ and skewness $\gamma = 0, 1, 2$, for retentions $d = 0, \frac{1}{2}, \ldots, 4$?

To get NP-approximations we apply formula (3.104) if $\gamma = 0$, (3.114) otherwise. The parameters of a translated gamma distributed random variable with expected value 0, variance 1 and skewness γ are $\alpha = 4/\gamma^2$, $\beta = 2/\gamma$ and $x_0 = -2/\gamma$. For $\gamma \downarrow 0$, (3.105) yields the stop-loss premiums for a $\mathrm{N}(0,1)$ distribution. All gamma stop-loss premiums are somewhat smaller than the NP approximated ones.

The NP-approximation (3.114) yields plausible results for $d = 0$, but the results in Table 3.1 for gamma are surprising, in that the stop-loss premiums decrease with increasing skewness. From (3.116) below, it immediately follows that if all stop-loss premiums for one distribution are larger than those of another distribution with the same expected value, then the former has a larger variance. Since in this case the variances are equal, as well as larger stop-loss premiums of the translated gamma, there have to be smaller ones. With NP, lower stop-loss premiums for higher skewness might occur, for example, in case $d < 0$. Note that the translated gamma approximation gives the stop-loss premium for a risk with the right expected value and variance. On the other hand, NP gives approximate stop-loss premiums for a random variable with almost the same moments. Obviously, random variables exist having the NP tail probabilities in the area $d \in (0, \infty)$, as well as the correct first three moments.

Table 3.1 Approximate stop-loss premiums at various retentions for a standardized random variable with skewness 0, 1 and 2, using the CLT, the NP and the gamma approximations.

	CLT	Normal Power		gamma	
d	$\gamma=0$	$\gamma=1$	$\gamma=2$	$\gamma=1$	$\gamma=2$
0.0	.3989	.4044	.4195	.3907	.3679
0.5	.1978	.2294	.2642	.2184	.2231
1.0	.0833	.1236	.1640	.1165	.1353
1.5	.0293	.0637	.1005	.0598	.0821
2.0	.0085	.0316	.0609	.0297	.0498
2.5	.0020	.0151	.0365	.0144	.0302
3.0	.0004	.0070	.0217	.0068	.0183
3.5	.0001	.0032	.0128	.0032	.0111
4.0	.0000	.0014	.0075	.0015	.0067

For arbitrary μ and σ, simply use a translation like the one given in (3.101). In that case, first determine $d=(t-\mu)/\sigma$, then multiply the corresponding stop-loss premium in the above table by σ, and if necessary, use interpolation. ∇

3.10.1 Comparing stop-loss premiums in case of unequal variances

In this subsection we compare the stop-loss premiums of two risks with equal expected value, but with unequal variance. It is impossible to formulate an exact general rule, but we can state some useful approximating results.

Just as one gets the expected value by integrating the distribution function over $(0,\infty)$, one can in turn integrate the stop-loss premiums. In Exercise 3.10.1, the reader is invited to prove that, if $U\geq 0$ with probability 1,

$$\frac{1}{2}\mathrm{Var}[U]=\int_0^\infty \big\{\mathrm{E}[(U-t)_+]-(\mu-t)_+\big\}\,\mathrm{d}t. \tag{3.115}$$

The integrand in this equation is always non-negative. From (3.115), it follows that if U and W are risks with equal expectation μ, then

$$\int_0^\infty \big\{\mathrm{E}[(U-t)_+]-\mathrm{E}[(W-t)_+]\big\}\,\mathrm{d}t=\frac{1}{2}\big\{\mathrm{Var}[U]-\mathrm{Var}[W]\big\}. \tag{3.116}$$

By approximating the integral in (3.116) with the trapezoidal rule with interval width 1, see also (3.47), we can say the following about the total of all differences in the stop-loss premiums of U and W (notice that we do not use absolute values):

$$\sum_{i=1}^{\infty}\big\{\mathrm{E}[(U-i)_+]-\mathrm{E}[(W-i)_+]\big\}\approx\frac{1}{2}\big\{\mathrm{Var}[U]-\mathrm{Var}[W]\big\}. \tag{3.117}$$

So, if we replace the actual stop-loss premiums of U by those of W, then (3.117) provides an approximation for the total error in all integer-valued arguments. In Chapter 7 we examine conditions for $\mathrm{E}[(U-d)_+] \geq \mathrm{E}[(W-d)_+]$ to hold for all d. If that is the case, then all terms in (3.117) are positive and consequently, the maximum error in all of these terms will be less than the right-hand side.

If two integrands are approximately proportional, their ratio is about equal to the ratio of the corresponding integrals. So from (3.115) we get:

$$\frac{\mathrm{E}[(U-t)_+]-(\mu-t)_+}{\mathrm{E}[(W-t)_+]-(\mu-t)_+} \approx \frac{\mathrm{Var}[U]}{\mathrm{Var}[W]}. \tag{3.118}$$

The approximation is exact if $\mu = \mathrm{E}[U]$ and $W = (1-I)\mu + IU$ with $I \sim \text{Bernoulli}(\alpha)$ independent of U and $\alpha = \mathrm{Var}[W]/\mathrm{Var}[U]$, see Exercise 3.10.2.

If $t \geq \mu$, then $(\mu-t)_+ = 0$, so the approximation (3.118) simplifies to the following rule of thumb:

Rule of thumb 3.10.6 (Ratio of stop-loss premiums)
For retentions t larger than the expectation $\mu = \mathrm{E}[U] = \mathrm{E}[W]$, we have for the stop-loss premiums of risks U and W:

$$\frac{\mathrm{E}[(U-t)_+]}{\mathrm{E}[(W-t)_+]} \approx \frac{\mathrm{Var}[U]}{\mathrm{Var}[W]}. \tag{3.119}$$

This rule works best for intermediate values of t, see below. ∇

Example 3.10.7 ('Undefined wife')
Exercise 3.7.4 deals with the situation where it is unknown for which of the insureds a widow's benefit might have to be paid. If the frequency of being married is 80%, we can either multiply all risk amounts by 0.8 and leave the probability of dying within one year as it is, or we can multiply the mortality probability by 0.8 and leave the payment as it is. We derived that the resulting variance of the total claim amount in the former case is approximately 80% of the variance in the latter case. So, if we use the former method to calculate the stop-loss premiums instead of the correct method, then the resulting stop-loss premiums for retentions that are larger than the expected claim cost are approximately 20% too small. ∇

Example 3.10.8 (Numerical evaluation of the Rule of thumb)
We calculated the stop-loss premiums for a $\mathrm{N}(0,1.01)$ and a $\mathrm{N}(0,1.25)$ distribution at retentions $d = 0, \frac{1}{2}, 1, \ldots, 3$, to compare them with those of a $\mathrm{N}(0,1)$ distribution. According to Rule of thumb 3.10.6, these should be 1.01 and 1.25 times as big respectively. Table 3.2 gives the factor by which that factor should be multiplied to get the real error. For example, for $d = 0$ the quotient $\pi(d;0,1.01)/\pi(d;0,1)$ equals 1.005 instead of 1.01, so the error is only 50% of the one predicted by the Rule of thumb. As can be seen, the Rule of thumb correction factor is too large for retentions close to the expected value, too small for large retentions and approximately correct for retentions equal to the expected value plus 0.6 standard deviation. The Rule of thumb correction factor has a large error for retentions in the far tail where the stop-

Table 3.2 Factors by which the $N(0,1.01)$ and $N(0,1.25)$ stop-loss premiums deviate from those of $N(0,1)$, expressed in terms of the Rule of thumb correction factor

d	$\pi(d;0,1)$	Correction factors	
		$1+0.01\times$	$1+0.25\times$
0.0	0.39894	0.50	0.47
0.5	0.19780	0.89	0.85
1.0	0.08332	1.45	1.45
1.5	0.02931	2.22	2.35
2.0	0.00849	3.20	3.73
2.5	0.00200	4.43	5.84
3.0	0.00038	5.92	9.10

loss premiums of the distribution with the smaller variance are negligible but those of the distribution with the larger variance are not. ∇

3.11 Exercises

Section 3.2

1. Calculate (3.3), (3.4) and (3.5) in case N has the following distribution: a) Poisson(λ), b) binomial(n,p) and c) negative binomial(r,p).
2. Give the counterpart of (3.5) for the cumulant generating function.
3. Assume that the number of eggs in a bird's nest is a Poisson(λ) distributed random variable, and that the probability that a female hatches out equals p. Determine the distribution of the number of female hatchlings in a bird's nest.
4. Let S be compound Poisson distributed with $\lambda = 2$ and $p(x) = x/10$, $x = 1,2,3,4$. Apply (3.10) to calculate the probabilities of $S = s$ for $s \leq 4$.
5. Complete the table in Example 3.2.2 for $x = 0,\ldots,6$. Determine the expected value and the variance of N, X and S.
6. Determine the expected value and the variance of S, where S is defined as in Example 3.2.2, except that N is Poisson distributed with $\lambda = 2$.
7. Prove relation (3.11) by partial integration. Do the same by differentiating both sides of the equation and examining one value, either $x = 0$ or $x \to \infty$.

Section 3.3

1. Show that the Poisson distribution also arises as the limit of the negative binomial(r,p) distribution if $r \to \infty$ and $p \to 1$ such that $r(1-p) = \lambda$ remains constant.
2. Under which circumstances does the usual Poisson distribution arise instead of the negative binomial in Examples 3.3.1 and 3.3.2?
3. [♠] Prove (3.19).

Section 3.4

1. The same as Exercise 3.2.4, but now with the sparse vector algorithm.
2. What happens with (3.23) if some x_i are equal in (3.22)?
3. Assume that S_1 is compound Poisson with $\lambda_1 = 4$ and claims $p_1(j) = \frac{1}{4}, j = 0,1,2,3$, and S_2 is also compound Poisson with $\lambda_2 = 2$ and $p_2(j) = \frac{1}{2}, j = 2,4$. If S_1 and S_2 are independent, then what is the distribution of $S_1 + S_2$?
4. In Exercise 3.2.3, prove that the number of males is independent of the number of females.
5. Let N_j, $j = 1,2$, denote the number of claims of size j in Example 3.2.2. Are N_1 and N_2 independent?
6. Assume that S is compound Poisson distributed with parameter λ and with discrete claims distribution $p(x)$, $x > 0$. Consider S_0, a compound Poisson distribution with parameter $\lambda_0 = \lambda/\alpha$ for some α with $0 < \alpha < 1$, and with claims distribution $p_0(x)$ where $p_0(0) = 1 - \alpha$ and $p_0(x) = \alpha p(x)$ for $x > 0$. Prove that S and S_0 have the same distribution by comparing their mgfs. Also show that $S \sim S_0$ holds because the frequencies of the claim amounts $x \neq 0$ in (3.22) have the same distribution.
7. How many multiplications *with non-zero numbers* does the sparse vector algorithm of Example 3.4.3 take to compute all probabilities $\Pr[S = x]$, $x = 0,1,\ldots,n-1$? Assume the claim sizes to be bounded by T, and remember that $1 + 1/2 + 1/3 + \cdots + 1/T \approx \log T + 0.5772$ (the Euler-Mascheroni constant).
8. Redo Exercise 3.4.1 using R.

Section 3.5

1. The same as Exercise 3.2.4, but now with Panjer's recursion relation.
2. The same as Exercise 3.4.6, first part, but now by proving with induction that Panjer's recursion yields the same probabilities $f(s)$.
3. Verify Example 3.5.2.
4. In case of a compound Poisson distribution for which the claims have mass points $1,2,\ldots,m$, determine how many multiplications have to be done to calculate the probability $F(t)$ using Panjer's recursion. Distinguish the cases $m < t$ and $m \geq t$.
5. Prove that $\mathrm{E}[N] = (a+b)/(1-a)$ if $q_n = \Pr[N = n]$ satisfies (3.26).
6. In Example 3.5.4, determine the retention d for which $\pi(d) = 0.3$.
7. Let N_1, N_2 and N_3 be independent and Poisson(1) distributed. For the retention $d = 2.5$, determine $\mathrm{E}[(N_1 + 2N_2 + 3N_3 - d)_+]$.
8. Assume that S_1 is compound Poisson distributed with parameter $\lambda = 2$ and claim sizes $p(1) = p(3) = \frac{1}{2}$. Let $S_2 = S_1 + N$, where N is Poisson(1) distributed and independent of S_1. Determine the mgf of S_2. What is the corresponding distribution? Determine $\Pr[S_2 \leq 2.4]$.
9. Determine the parameters of an integer-valued compound Poisson distributed Z if for some $\alpha > 0$, Panjer's recursion relation equals $\Pr[Z = s] = f(s) = \frac{\alpha}{s}[f(s-1) + 2f(s-2)]$, $s = 1,2,3,\ldots$ [Don't forget the case $p(0) \neq 0$!]
10. Assume that S is compound Poisson distributed with parameter $\lambda = 3$, $p(1) = \frac{5}{6}$ and $p(2) = \frac{1}{6}$. Calculate $f(x)$, $F(x)$ and $\pi(x)$ for $x = 0,1,2,\ldots$. Also calculate $\pi(2.5)$.
11. Derive formulas from (3.34) for the stop-loss premium that only use $f(0), f(1), \ldots, f(d-1)$ and $F(0), F(1), \ldots, F(d-1)$ respectively.

12. Give a formula, analogous to (3.36), to calculate $\mathrm{E}[(S-d)_+^2]$.
13. [♠] Write functions `Panjer.NegBin` and `Panjer.Bin` to handle (negative) binomial claim numbers.
14. Let $M \sim$ negative binomial(r,p) for some r,p with $0<p<1$. Which distribution does L have, if $\Pr[L=m] = \lim_{r\to 0}\Pr[M=m\,|\,M>0]$, $m=1,2,\ldots$? Show that, for this random variable, relation (3.26) holds for all $n=\mathbf{2},3,\ldots$
15. [♠] By a slight extension of the proof of Theorem 3.5.1, it can be shown that if (3.26) holds for all $n=\mathbf{2},3,\ldots$, the probability of a claim total s satisfies the following relation:

$$f(s) = \frac{[q_1-(a+b)q_0]f(s)+\sum_{h=1}^{s}(a+\frac{bh}{s})p(h)f(s-h)}{1-ap(0)}, \quad s=1,2,\ldots$$

Here q_0 is the arbitrary amount of probability at zero given to the frequency distribution. If $q_1=(a+b)q_0$, the first term in the numerator vanishes, resulting in (3.27). The class of counting distributions satisfying (3.26) for $n>k$ is known as the (a,b,k) class, $k=0,1,\ldots$. The $(a,b,1)$ class includes zero-modified or zero-truncated distributions. By the previous exercise, the logarithmic distribution is in the $(a,b,1)$ class, but not in the $(a,b,0)$ class.
Write a Panjering function for the $(a,b,1)$ class.

16. Using the R- functions `cumsum` and `rev`, show how the vector of stop-loss premiums at retention $d=0,1,\ldots$ can be obtained from a vector of probabilities `p` with `p[1]`$=\Pr[S=0]$, `p[2]`$=\Pr[S=1]$, and so on. Use relation (3.34).

Section 3.6

1. Describe what is produced by the R-calls:

```
y <- rep(0,64); y[2:7] <- 1/6; Re(fft(fft(y)^10,T))/64
```

2. Use the `fft` function to compute the stop-loss premium at retention $d=10$ for a compound distribution with claim sizes 1 and 3 each with probability $\frac{1}{2}$ and as claim number a logarithmic random variable L with $\mathrm{E}[L]=3$, see Example 3.3.2. The same when $\Pr[L=0,1,3,7,8,9,10] = .3,.2,.1,.1,.1,.1,.1$.

Section 3.7

1. Assume that S is compound Poisson distributed with parameter $\lambda=12$ and uniform$(0,1)$ distributed claims. Approximate $\Pr[S<10]$ with the CLT approximation, the translated gamma approximation and the NP approximation.
2. Assume that S is compound Poisson distributed with parameter $\lambda=10$ and $\chi^2(4)$ distributed claims. Approximate the distribution function of S with the translated gamma approximation. With the NP approximation, estimate the quantile s such that $F_S(s)\approx 0.95$, as well as the probability $F_S(\mathrm{E}[S]+3\sqrt{\mathrm{Var}[S]})$.
3. Prove Theorem 3.7.1 by proving, just as in the proof of Theorem 3.5.1, that as $\lambda\to\infty$, the mgf of a standardized compound Poisson(λ,X) random variable $(S-\mu_S)/\sigma_S$ converges to $\mathrm{e}^{t^2/2}$ for all t.

Section 3.8

1. Show that $\lambda_j = -\log(1-q_j)$ yields both a larger expectation and a larger variance of S in (3.63) than $\lambda_j = q_j$ does. For both cases, compare $\Pr[I_i = j]$ and $\Pr[N_i = j]$, $j = 0,1,2,\ldots$ in (3.62) and (3.63), as well as the cdfs of I_i and N_i.
2. Consider a portfolio of 100 one-year life insurance policies that are evenly divided between the insured amounts 1 and 2 and probabilities of dying within this year 0.01 and 0.02. Determine the expectation and the variance of the total claims $\widetilde{S}$. Choose an appropriate compound Poisson distribution S to approximate $\widetilde{S}$ and compare the expectations and the variances. Determine for both S and $\widetilde{S}$ the parameters of a suitable approximating translated gamma distribution.
3. Show, by comparing the respective mgfs, that the following representations of *the* collective model are equivalent:
   1. The compound Poisson distribution specified in (3.64) with $\lambda_i = q_i$.
   2. The random variable $\sum_i N_i b_i$ from (3.63) with $\lambda_i = q_i$.
   3. The random variable $Z_1 + \cdots + Z_n$ where the Z_i are compound Poisson distributed with claim number parameter 1 and claims distribution equal to those of $I_i b_i$.
   4. The compound Poisson distribution with parameter $\lambda = n$ and claims distribution $Q(x) = \frac{1}{n}\sum_j \Pr[X_j \le x]$. [Hence $Q(\cdot)$ is the arithmetic mean of the cdfs of the claims. It can be interpreted as the cdf of a claim from a *randomly* chosen policy, where each policy has probability $\frac{1}{n}$.]
4. In a portfolio of n one-year life insurance policies for men, the probability of dying in this year equals q_i for the ith policyholder. In case of death, an amount b_i has to be paid out, but only if it turns out that the policy holder leaves a widow behind. This information is not known to the insurer in advance ('undefined wife'), but it is known that this probability equals 80% for each policy. In this situation, we can approximate the individual model by a collective one in two ways: by replacing the insured amount for policy i by $0.8b_i$, or by replacing the claim probability for policy i by $0.8q_i$. Which method is correct? Determine the variance of the total claims for both methods. Show how we can proceed in both cases, if we have a program at our disposal that calculates stop-loss premiums from a mortality table and an input file containing the sex, the age and the risk amount.
5. [♠] At what value of x in (3.70) may we stop the summation if an absolute precision ε is required?
6. Consider a portfolio with 2 classes of policies. Class i contains 1000 policies with claim size $b_i = i$ and claim probability 0.01, for $i = 1,2$. Let B_i denote the number of claims in class i. Write the total claims S as $S = B_1 + 2B_2$ and let $N = B_1 + B_2$ denote the number of claims. Consider the compound binomial distributed random variable $T = X_1 + X_2 + \cdots + X_N$ with $\Pr[X_i = 1] = \Pr[X_i = 2] = 1/2$. Compare S and T as regards the maximum value, the expected value, the variance, the claim number distribution and the distribution. Do the same for B_1 and $B_2 \sim$Poisson(10).
7. Consider an excess of loss reinsurance on some portfolio. In case of a claim x, the reinsurer pays out an amount $h(x) = (x-\beta)_+$. The claims process is compound Poisson with claim number parameter 10 and uniform(1000,2000) distributed claim sizes. For $\beta \in [1000, 2000]$, determine the distribution of the total amount to be paid out by the reinsurer in a year.
8. Consider two portfolios P_1 and P_2 with the following characteristics:

	Risk amount	Number of policies	Claim probability
P_1	z_1	n_1	q_1
	z_2	n_2	q_2
P_2	z_1	$2n_1$	$\frac{1}{2}q_1$
	z_2	$2n_2$	$\frac{1}{2}q_2$

For the *individual* risk models for P_1 and P_2, determine the difference of the variance of the total claims amount. Check if the *collective* approximation of P_1 equals the one of P_2, both constructed with the recommended methods.

9. A certain portfolio contains two types of contracts. For type k, $k = 1,2$, the claim probability equals q_k and the number of policies equals n_k. If there is a claim, then with probability $p_k(x)$ it equals x, as follows:

	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 all policies are independent. Construct a collective model T to approximate the total claims. Make sure that both the expected number of positive claims and the expected total claims agree. Give the simplest form of Panjer's recursion relation in this case; also give a starting value. With the help of T, approximate the capital that is required to cover all claims in this portfolio with probability 95%. Use an approximation based on three moments, and compare the results with those of Exercise 2.5.13.

10. Consider a portfolio containing n contracts that all produce a claim 1 with probability q. What is the distribution of the total claims according to the individual model, *the* collective model and the *open* collective model? If $n \to \infty$, with q fixed, does the individual model S converge to the collective model T, in the sense that the difference of the probabilities $\Pr[(S - \mathrm{E}[S])/\sqrt{\mathrm{Var}[S]} \le x] - \Pr[(T - \mathrm{E}[S])/\sqrt{\mathrm{Var}[S]} \le x]$ converges to 0?

Section 3.9

1. Determine the mean and the variance of the lognormal and the Pareto distribution, see also Table A. Proceed as follows: $Y \sim \text{lognormal}(\mu, \sigma^2)$ means $\log Y \sim \mathrm{N}(\mu, \sigma^2)$; if $Y \sim \text{Pareto}(\alpha, x_0)$, then $Y/x_0 \sim \text{Pareto}(\alpha, 1)$ and $\log(Y/x_0) \sim \text{exponential}(\alpha)$.
2. Determine which parameters of the distributions in this section are scale parameters, in the sense that λX, or more general $f(\lambda)X$ for some function f, has a distribution that does not depend on λ. Show that neither the skewness γ_X nor the coefficient of variation σ_X/μ_X depend on such parameters. Determine these two quantities for the given distributions.
3. [♠] Prove that the expression in (3.84) is indeed a cdf that is 0 in $x = 0$, tends to 1 for $x \to \infty$ and has a positive derivative (3.83). Also verify that (3.85) is the mgf, and confirm the other statements about the inverse Gaussian distributions.
4. Show that the given conditions on q in (3.90) are sufficient for $p(\cdot)$ to be a pdf.
5. Determine the cdf $\Pr[Z \le d]$ and the stop-loss premium $\mathrm{E}[(Z-d)_+]$ for a mixture or combination Z of exponential distributions as in (3.90). Also determine the conditional distribution of $Z - z$, given $Z > z$.
6. Determine the mode of mixtures and combinations of exponential distributions. Also determine the mode and the median of the lognormal distribution.

7. Determine the mode of the inverse Gaussian(α,α) distribution. For the parameter values of Figure 3.1, use your computer to determine the median of this distribution.
8. Write R-functions `d/p/q/rCombExp` for mixtures/combinations of exponential distributions.
9. If $\mathrm{E}[X]=1$ and $\sqrt{\mathrm{Var}[X]}/\mathrm{E}[X]=2$ is the coefficient of variation, find the 97.5% quantile if $X\sim$ gamma, inverse Gaussian, lognormal, and normal.
10. [♠] A small insurer has three branches: Fire, Liability and Auto. The total claims on the branches are all independent compound Poisson random variables. The specifications are:

    - Fire: $\lambda_1=10$ is the mean claim number; claim sizes are inverse Gaussian with mean $\mu_1=10000$ and coefficient of variation $\sigma_1/\mu_1=0.5$;
    - Liability: $\lambda_2=3$; Pareto claim sizes with mean 30000 and coefficient of variation 0.5;
    - Auto: $\lambda_3=500$; gamma claim sizes with mean 1000 and coefficient of variation 0.3.

    Using the Normal Power approximation, compute the Value-at-Risk (quantiles s such that $\Pr[S>s]=q$) for the combined portfolio at the levels $q=0.5,0.75,0.9,0.95,0.99$.
    The premium income $P=P_1+P_2+P_3$ contains a safety loading 30% for Fire, 50% for Liability and 10% for Auto. It is invested in such a way that its value after a year is lognormal with mean $1.1P$ and standard deviation $0.3P$. Find the probability that the invested capital is enough to cover the claims. Find the initial capital that is needed to ensure that with probability 99%, the insurer does not get broke by the claims incurred and/or the investment losses.
11. [♠] In the situation of Remark 3.8.3, find the probability vectors of a compound Poisson(50, $p(7)=p(10)=p(19)=1/3$) distribution for S^+ and compound Poisson(3, $p(1)=p(4)=1/2$) for S^-. Use either FFT or Panjer's recursion. Then, compute the stop-loss premiums at $d=\lfloor \mathrm{E}[S]+k\sigma_S\rfloor$ for $k=1,2,3$ using (3.70). Also, compute the whole probability vector of $S=S^+-S^-$ using `convolve`. Finally, use the fact that S itself is also a compound Poisson random variable (with possible claim sizes $-1,-4,+7,+10,+19$) to compute the vector of probabilities $\Pr[S=s]$ using the FFT-technique.
12. Prove that (3.89) produces positive estimates $\widehat{\alpha}$ and $\widehat{\beta}$. Hint: apply Jensen's inequality to the r.v. Z with $\Pr[Z=y_i]=\frac{1}{n}$, $i=1,\ldots,n$.
13. Prove that `aux > 0`, so $\overline{\log y}<\log\bar{y}$, holds when computing ML-estimates for a gamma sample using the R program involving `uniroot` given.
14. [♠] For range $y>0$, shape parameter $\alpha>0$ and scale parameter $\beta>0$, a Weibull(α,β) random variable Y has density $f(y;\alpha,\beta)=\alpha\beta\,(\beta y)^{\alpha-1}\exp\big(-(\beta y)^{\alpha}\big)$. Find the cdf of Y. Show that $X=(\beta Y)^{\alpha}$ has an exponential(1) distribution. Also show that the ML equations for estimating α and β using a sample $Y_1=y_1,\ldots,Y_n=y_n$ can be written as $g(\alpha):=1/\alpha+\overline{\log y}-\overline{y^{\alpha}\log y}/\overline{y^{\alpha}}=0$ and $\beta=(1/\overline{y^{\alpha}})^{1/\alpha}$. Using `rweibull`, generate a Weibull sample of size 2000 with parameters $\alpha=5$ and $\beta=1$, and compute the ML estimates for the parameters from this sample.
15. Write a function `rTransGam` to generate random deviates from translated gamma distributions. Draw a large sample and verify if the first three sample moments are 'close' to the theoretical values.
    Identify the problem with programming `rNormalPower`.
16. In life actuarial science, it is sometimes useful to be able to generate samples from Gompertz and Makeham lifetime distributions. The mortality rate of Makeham's law equals

$$\mu_X(x)\overset{\text{def}}{=}\frac{f_X(x)}{1-F_X(x)}=-\frac{\mathrm{d}}{\mathrm{d}x}\log(1-F_X(x))=a+bc^x,$$

    while Gompertz' law is the special case with $a=0$. Assuming $c>1$, the second component of the mortality rate increases exponentially with age; the first is a constant. Since the minimum $X=\min(Y,Z)$ of two independent random variables is easily seen to have $\mu_X(x)=\mu_Y(x)+$

$\mu_Z(x)$ as its mortality rate, under Makeham's law people either die from 'senescence' after time $Y \sim$ Gompertz(b,c), or from some 'accident' occurring independently after time $Z \sim$ exponential(a), whichever happens first. In the parameterization given, X has the following survival function:

$$1 - F_X(x) = \Pr[Z > x] \Pr[Y > x] = \exp\Big(-ax - \frac{b}{\log c}(c^x - 1)\Big), \quad x > 0.$$

If $Y \sim$ Gompertz(b,c), show that $c^Y - 1$ is exponential$(\frac{b}{\log c})$. Use this fact to write a function `rGompertz` to generate random deviates from the Gompertz distribution. Use it to write a function `rMakeham`. Verify if your results make sense using, e.g., $a = 0.0005$, $b = 0.00007$, $c = 1.1$.

17. Use relation (3.11) to construct an R function to do one drawing from a Poisson(λ) distribution.
18. [♠] If U_1, U_2 are independent uniform$(0,1)$ random variables, show that $\tan(\pi(U_1 - \frac{1}{2})) \sim$ Cauchy$(0,1)$. Also show that $\Phi^{-1}(U_1)/\Phi^{-1}(U_2)$ has this same distribution. [Because it has fat tails and an easy form, this distribution is often used to find a majorizing function in the rejection method.]
19. Apply the rejection method of Section 3.9.1 to construct an R-function to draw from the triangular distribution $f(x) = x$ on $(0,1)$, $2 - x$ on $(1,2)$. Use the uniform$(0,2)$ density to get an upper bound for $f(x)$.
    Also, draw from $f(x)$ by using the fact that $U_1 + U_2 \sim f(x)$ if the U_i are iid uniform$(0,1)$.
20. The rejection method can also applied to draw from integer-valued random variables. In fact, to generate values from N, generate a value from $N+U$ and round down. Here $U \sim$ uniform$(0,1)$, independent of N. Sketch the density of $N+U$. Apply to a binomial$(3,p)$ distribution.
21. What is produced by `sum(runif(10)<1/3)`? What by `sum(runif(12))-6`?
22. [♠] For the inverse Gaussian distributions, show that $\sum Y_i$ and $\sum \frac{1}{Y_i}$ are jointly sufficient. Recall that this may be proved by using the factorization criterion, that is, by showing that g and h exist such that the joint density of $Y_1, \ldots, Y_n$ can be factorized as $f_{Y_1,\ldots,Y_n}(y_1,\ldots,y_n;\alpha,\beta) = g\big(\sum y_i, \sum \frac{1}{y_i};\alpha,\beta\big)\, h(y_1,\ldots,y_n)$.
23. [♠] Let $X \sim$ gamma$(\alpha+1,1)$, $\alpha > 0$, and $U \sim$ uniform$(0,1)$ be independent. Prove that $\mathrm{m}_{\log X}(t) = \frac{\Gamma(\alpha+t+1)}{\Gamma(\alpha+1)}$ and $\mathrm{m}_{\frac{1}{\alpha}\log U}(t) = \frac{1}{1+t/\alpha}$. What is the distribution of $X U^{1/\alpha}$?
24. Let $U \sim$ uniform$(0,1)$. Name the distribution of the following transformed random variables:

    a) $B = I_{(0,p)}(U)$ and $N = \sum_1^n B_i$ with $B_i \sim B$ iid (see Exercise 3.9.21);
    b) $X = \Phi^{-1}(U)$ and $\mu + \sigma X$;
    c) $\mathrm{e}^{\mu+\sigma X}$;
    d) $Y = -\log U$ and Y/β;
    e) $\sum_1^n Y_i/\beta$ with $Y_i \sim Y$ iid;
    f) $\mathrm{e}^Y (= 1/U)$ and $x_0 \mathrm{e}^{Y/\beta}$;
    g) $Y^{1/\alpha}/\beta$ (see Exercise 3.9.14);
    h) $\log\big(1 + Y\log(c)/b\big)/\log(c)$ (see Exercise 3.9.16).

25. [♠] Assume we have observations of the number of claims $N_i = n_i$, $i = 1,\ldots,n$ in a portfolio of risks. It is known that, conditionally given that the value of a structure variable $\Lambda_i = \lambda_i$, this number has a Poisson$(\lambda_i w_i)$ distribution, where the $w_i \in [0,1]$ are known exposures. Also assume $\Lambda_i \sim$ gamma(α,β). Find the marginal distribution of the N_i. Draw a random sample of size $n = 10000$ from the N_i, assuming the w_i are 0.6 for the first 2000 policies, 1.0 for the remainder. Take α, β such that the mean number of claims per unit of exposure is 0.1, and the coefficient of variation of the risk parameters $\sigma_\Lambda/\mu_\Lambda = 1.3$. From this sample, find the ML-estimates for α, β.

26. [♠] Solve the second ML-equation (3.77) using `uniroot` and inserting (3.76). For this, write a function of x delivering the sum of $n_1, n_2, \ldots$ times the cumulative sums of $\frac{1}{x}, \frac{1}{x+1}, \ldots$, less $n \log \frac{x}{x+\bar{y}}$.
27. [♠] How does one generate pseudo-random samples from the following distributions: Student, chi-square, beta, Fisher-Snedecor? Use their well-known relations to normal random samples.
28. Write a function to generate pseudo-random samples from the logistic distribution, see Example 11.5.6.

Section 3.10

1. Assume that X is normally distributed with expectation 10000 and standard deviation 1000. Determine the stop-loss premium for a retention 13000. Do the same for a random variable Y that has the same first two moments as X, but skewness 1.
2. Show that $\mathrm{E}[(S-d)_+] = \mathrm{E}[S] - d + \int_0^d (d-x)\,\mathrm{d}F(x) = \mathrm{E}[S] - \int_0^d [1-F(x)]\,\mathrm{d}x$.
3. If $X \sim \mathrm{N}(\mu, \sigma^2)$, show that $\int_\mu^\infty \mathrm{E}[(X-t)_+]\,\mathrm{d}t = \frac{1}{4}\sigma^2$ and determine $\mathrm{E}[(X-\mu)_+]$.
4. Verify (3.105). Also verify (3.106) and (3.107), and show how these can be used to approximate the variance of the retained loss.
5. Give an expression for the net premium if the number of claims is Poisson(λ) distributed and the claim size is Pareto distributed. Assume that there is a deductible d.
6. [♠] Let $X \sim$ lognormal(μ, σ^2). Show that for $d > 0$, the stop-loss premium is

$$\mathrm{E}[(X-d)_+] = \mathrm{e}^{\mu+\sigma^2/2}\Phi\Big(\frac{-\log d + \mu + \sigma^2}{\sigma}\Big) - d\Phi\Big(\frac{-\log d + \mu}{\sigma}\Big).$$

   Compare your result with the Black-Scholes option pricing formula, and explain.
7. In Table 3.1, does using linear interpolation to calculate the stop-loss premium in for example $d = 0.4$ for one of the given values for γ yield a result that is too high or too low?
8. Assume that N is an integer-valued risk with $\mathrm{E}[(N-d)_+] = \mathrm{E}[(U-d)_+]$ for $d = 0, 1, 2, \ldots$, where $U \sim \mathrm{N}(0,1)$. Determine $\Pr[N=1]$.
9. Let $\pi(t) = \mathrm{E}[(U-t)_+]$ denote the stop-loss premium for $U \sim \mathrm{N}(0,1)$ and retention t, $-\infty < t < \infty$. Show that $\pi(-t)$, $t \geq 0$ satisfies $\pi(-t) = t + \pi(+t)$. Sketch $\pi(t)$.
10. In Sections 3.9 and 3.10, the retention is written as $\mu + k\sigma$, so it is expressed in terms of a number of standard deviations above the expected loss. However, in the insurance practice, the retention is always expressed as a percentage of the expected loss. Consider two companies for which the risk of absence due to illness is to be covered by stop-loss insurance. This risk is compound Poisson distributed with parameter λ_i and exponentially distributed individual losses X with $\mathrm{E}[X] = 1000$. Company 1 is small: $\lambda_1 = 3$; company 2 is large: $\lambda_2 = 300$. What are the net stop-loss premiums for both companies in case the retention d equals 80%, 100% and 120% of the expected loss respectively? Express these amounts as a percentage of the expected loss and use the normal approximation.
11. For the normal, lognormal and gamma distributions, as well as mixtures/combinations of exponentials, write functions like `slnorm` giving the stop-loss premiums of the corresponding random variables.
    Also, give a function yielding an approximate stop-loss premium for an r.v. having mean μ, variance σ^2 and skewness γ, based on the Normal Power approximation, see (3.114). Do the same for the translated gamma approximation, see (2.57) and (3.105).
12. Using R, verify Tables 3.1 and 3.2.

13. Prove (3.115) and (3.116) and verify that the integrand in (3.115) is non-negative.
14. Show that (3.118) is exact if $W = (1-I)\mu + IU$ with $\mu = \mathrm{E}[U]$ and $I \sim$ Bernoulli(α), for $\alpha = \mathrm{Var}[W]/\mathrm{Var}[U]$.
15. Verify Rule of thumb 3.10.6 for the case $U \sim$ Poisson(1) and $V \sim$ binomial(10, $\frac{1}{10}$).
16. Assume that $X_1, X_2, \ldots$ are independent and identically distributed risks that represent the loss on a portfolio in consecutive years. We could insure these risks with separate stop-loss contracts for one year with a retention d, but we could also consider only one contract for the whole period of n years with a retention nd. Show that $\mathrm{E}[(X_1 - d)_+] + \cdots + \mathrm{E}[(X_n - d)_+] \geq \mathrm{E}[(X_1 + \cdots + X_n - nd)_+]$. If $d \geq \mathrm{E}[X_i]$, examine how the total net stop-loss premium for the one-year contracts $\mathrm{E}[(X_1 - d)_+]$ relates to the stop-loss premium for the n-year period $\mathrm{E}[(X_1 + \cdots + X_n - nd)_+]$.
17. Let $B_1 \sim$ binomial(4,0.05), $B_2 \sim$ binomial(2,0.1), $S = B_1 + B_2$ and $T \sim$ Poisson(0.4). For the retentions $d = \frac{1}{2}, 1, \frac{3}{2}$, use the Rule of thumb 3.10.6 and discuss the results.
18. Derive (3.117) from the trapezoidal rule $\int_0^\infty f(x)\,\mathrm{d}x \approx \frac{1}{2\delta} \sum_1^\infty [f(i\delta) + f((i-1)\delta)]$ with interval width $\delta = 1$.