Modelling Dependent Credit Risks
with Extensions of CreditRisk$^+$
and Application to Operational Risk
(Lecture Notes)

Prof. Dr. Uwe Schmock
PRisMa Lab
Financial and Actuarial Mathematics (FAM)
Institute for Stochastics and Mathematical Methods in Economics
Vienna University of Technology
Wiedner Hauptstraße 8–10/E105-1
A-1040 Vienna, Austria
2004–2021

schmock@fam.tuwien.ac.at

Version$^1$ of Notes: April 22, 2021, 12:50
(Incomplete revision)

$^1$Updates on fam.tuwien.ac.at/~schmock/notes/ExtensionsCreditRiskPlus.pdf
## Contents

1 Introduction 5

2 Bernoulli Models for Credit Defaults 5
   2.1 Notation and Basic Bernoulli Model 5
   2.2 General Bernoulli Mixture Model 7
      2.2.1 Assumptions on the Random Default Probabilities 7
      2.2.2 Number of Default Events, Expected Value and Variance 8
   2.3 Uniform Bernoulli Mixture Model 9
      2.3.1 Beta-Binomial Mixture Model 10
      2.3.2 Biased Measure and the Beta Distribution 13
   2.4 One-Factor Bernoulli Mixture Model 14

3 Poisson Models for Credit Defaults 15
   3.1 Elementary Properties of the Poisson Distribution 15
   3.2 Calibration of the Poisson Distribution 17
   3.3 Metrics for Spaces of Probability Measures 18
   3.4 Poisson Approximation 25
      3.4.1 Results Using an Elementary Coupling Method 25
      3.4.2 Proof by the Stein–Chen Method for the Total Variation 31
      3.4.3 Proof by the Stein–Chen Method for the Wasserstein Metric 35
   3.5 Multivariate Poisson Distribution 36
   3.6 General Multivariate Poisson Mixture Model 38
      3.6.1 Expected Values, Variances, and Individual Covariances 39
      3.6.2 One-Factor Poisson Mixture Model 42
      3.6.3 Uniform Poisson Mixture Model 42

4 Generating Functions, Mixed and Compound Distributions 43
   4.1 Probability-Generating Functions 43
      4.1.1 Examples 43
      4.1.2 Basic Properties and Calculation of Moments 45
   4.2 Application to the General Multivariate Poisson Mixture Model 52
   4.3 Properties of the Gamma Distribution 53
      4.3.1 Moments of the Gamma Distribution 54
      4.3.2 Biased Gamma Distribution 54
   4.4 Gamma-Mixed Poisson Distribution 55
   4.5 Generating Function of Compound Distributions 57
   4.6 Some Compound Distributions Arising from the Multivariate
      Bernoulli Distribution 62
      4.6.1 Multivariate Logarithmic Distribution 63
1 Introduction

Credit risk models can be roughly divided into three classes:

- Actuarial models,
- Structural or asset value models,
- Reduced form or intensity-based models.

These lecture notes concentrate on actuarial models, starting from Bernoulli models and -- justified by the Poisson approximation -- progressing to Poisson models for credit risks. Considerable effort is made to discuss extensions of CreditRisk+, which are also extensions of the collective model used in actuarial science. The presented algorithm for the calculation of the portfolio loss distribution, based on variations of Panjer’s recursion, offers a flexible tool to aggregate risks and to determine popular values to quantify risk, like value-at-risk or expected shortfall. The algorithm is recursive and numerically stable, avoiding Monte Carlo methods completely.

2 Bernoulli Models for Credit Defaults

Parts of Sections 2 and 3 are inspired by the corresponding presentation in Bluhm, Overbeck and Wagner [9].

2.1 Notation and Basic Bernoulli Model

First of all we have to introduce some notation: Let $m$ be the number of individual obligors/counterparties/credit risks and $(N_1, \ldots, N_m)$ be a random vector of Bernoulli default indicators, i.e. binary values

$$N_i = \begin{cases} 1 & \text{if obligor } i \text{ defaults (within one period),} \\ 0 & \text{otherwise,} \end{cases}$$

giving the number of defaults. Furthermore, let

$$p_i := P[N_i = 1] \in [0,1] \quad (2.1)$$

denote the probability of default$^2$ of obligor $i \in \{1, \ldots, m\}$ within a certain period (usually one year) and

$$N := \sum_{i=1}^{m} N_i \quad (2.2)$$

$^1$Named after Jacob Bernoulli (also known as James or Jacques, 1655–1705 according to the Gregorian calendar). His main work, the *Ars conjectandi*, was published in 1713, eight years after his death, by his nephew, Nicolaus Bernoulli.

$^2$Determining reliable values for $p_1, \ldots, p_m$ in practice can be a challenging task.
be the random variable representing the total number of defaults. Obviously
\[ \mathbb{E}[N_i] = p_i \]  
(2.3)
and, using \( N_i^2 = N_i \),
\[ \text{Var}(N_i) = \mathbb{E}[N_i^2] - (\mathbb{E}[N_i])^2 = p_i(1 - p_i). \]  
(2.4)
The expected number of defaults (within one period) is given by
\[ \mathbb{E}[N] = \sum_{i=1}^{m} \mathbb{E}[N_i] = \sum_{i=1}^{m} p_i, \]  
(2.5)
where the expected value from (2.3) is used.

If the \( N_1, \ldots, N_m \) are uncorrelated, meaning that
\[ \text{Cov}(N_i, N_j) = \mathbb{E}[(N_i - \mathbb{E}[N_i])(N_j - \mathbb{E}[N_j])] = 0 \]
for all \( i, j \in \{1, \ldots, m\} \) with \( i \neq j \), then the variance of \( N \) is
\[ \text{Var}(N) = \sum_{i=1}^{m} \text{Var}(N_i) = \sum_{i=1}^{m} p_i(1 - p_i); \]  
(2.6)
see (2.18) and Exercise 2.3 for a more general formula.

The probability of exactly \( n \in \{0, 1, \ldots, m\} \) defaults is the sum over the probabilities of all possible subsets of \( n \) obligors defaulting during the period, i.e.
\[ P[N = n] = \sum_{|I| = n} \prod_{i \in I} p_i \prod_{i \in \{1, \ldots, m\} \setminus I} (1 - p_i). \]  
(2.7)
Moreover, if the \( N_1, \ldots, N_m \) are independent (which is a strong assumption), then
\[ P[N = n] = \binom{m}{n} p^n (1 - p)^{m-n}, \quad n \in \{0, 1, \ldots, m\}, \]  
(2.8)
For \( n = 100 \) defaults in a portfolio of \( m = 1000 \) obligors, assuming pairwise different \( p_1, \ldots, p_m \), this gives in general
\[ \binom{1000}{100} \approx 6.4 \times 10^{139} \]
terms, which is impossible to calculate explicitly using a computer. This illustrates the need for simplifying assumptions and suitable approximations.

In the special case of equal default probabilities for all obligors, i.e.
\[ p_1 = \cdots = p_m =: p, \]
the distribution in (2.8) simplifies to
\[ P[N = n] = \binom{m}{n} p^n (1 - p)^{m-n}, \quad n \in \{0, 1, \ldots, m\}, \]  
(2.9)
which is the binomial distribution \( \text{Bin}(m, p) \) for \( m \in \mathbb{N}_0 \) independent trials with success probability \( p \in [0, 1] \). In Section 2.3 and in the context of uniform portfolios, we will encounter the case of equal default probabilities again.

In practice, \( N_1, \ldots, N_m \) usually are dependent on each other!
2.2 General Bernoulli Mixture Model

In the introduction above, all the default probabilities were constant numbers. Taking the step to the general Bernoulli mixture model, we will introduce random probabilities of default. This generalization is natural, as the default probabilities affecting the obligors in the coming period are not exactly known today. The uncertainty is expressed by introducing a distribution for them as follows.

Let $P_1, \ldots, P_m$ be $[0,1]$-valued random variables with a joint distribution $F$ on $[0,1]^m$. We will denote this fact by writing $(P_1, \ldots, P_m) \sim F$.

2.2.1 Assumptions on the Random Default Probabilities

At this point no specific distribution is assumed for $F$. Only some general assumptions are made. The first, and a quite natural one, is that $P_i$ completely describes the conditional default probability of obligor $i \in \{1, \ldots, m\}$, i.e.

$$\mathbb{P}[N_i = 1 | P_1, \ldots, P_m] \overset{a.s.}= \mathbb{P}[N_i = 1 | P_i] \overset{a.s.}= P_i. \quad (2.10)$$

The second assumption states that the default numbers $N_1, \ldots, N_m$ are conditionally independent given $(P_1, \ldots, P_m)$. In other words: If the default probabilities are known, then the individual defaults are independent. Formally, for all $n_1, \ldots, n_m \in \{0,1\}$, the joint conditional probabilities satisfy

$$\mathbb{P}[N_1 = n_1, \ldots, N_m = n_m | P_1, \ldots, P_m] \overset{a.s.}= \prod_{i=1}^m \mathbb{P}[N_i = n_i | P_1, \ldots, P_m] \overset{a.s.}= \prod_{i=1}^m P_i^{n_i} (1 - P_i)^{1-n_i}, \quad (2.11)$$

where we used (2.10), the convention $0^0 := 1$ and

$$P_i^{n_i} (1 - P_i)^{1-n_i} = \begin{cases} P_i, & \text{if } n_i = 1, \\ 1 - P_i, & \text{if } n_i = 0, \end{cases}$$

for the last equation in (2.11). Note that, for every $i \in \{1, \ldots, m\}$,

$$\sum_{n_i \in \{0,1\}} P_i^{n_i} (1 - P_i)^{1-n_i} = 1. \quad (2.12)$$

In the unconditional case, the joint distribution is obtained by integration of (2.11) over all possible values of $(P_1, \ldots, P_m)$ with respect to the distribution $F$, or formally

$$\mathbb{P}[N_1 = n_1, \ldots, N_m = n_m] = \mathbb{E} \left[ \prod_{i=1}^m P_i^{n_i} (1 - P_i)^{1-n_i} \right] = \int_{[0,1]^m} \prod_{i=1}^m P_i^{n_i} (1 - P_i)^{1-n_i} F(dp_1, \ldots, dp_m). \quad (2.13)$$
If \( I \subset \{1, \ldots, m\} \) is any subset of obligors, then iterative summation over all \( n_i \in \{0, 1\} \) with \( i \in \{1, \ldots, m\} \setminus I \) using (2.12) implies that

\[
\mathbb{P}[N_i = n_i \text{ for all } i \in I] = \mathbb{E}\left[ \prod_{i \in I} P_i^{n_i} (1 - P_i)^{1-n_i} \right]. \tag{2.14}
\]

**Exercise 2.1** (Conditional expectation involving independent random variables).

Let \((\Omega, \mathcal{A}, \mathbb{P})\) be a probability space, \(B \subset \mathcal{A}\) a sub-\(\sigma\)-algebra, \((S_1, \mathcal{S}_1)\) and \((S_2, \mathcal{S}_2)\) measurable spaces, \(X: \Omega \to S_1\) and \(Y: \Omega \to S_2\) random variables, and \(F: S_1 \times S_2 \to \mathbb{R}\) an \(S_1 \otimes S_2\)-measurable function, which is bounded or non-negative. Suppose that \(X\) is \(B\)-measurable and \(Y\) is independent of \(B\). Prove that

\[
\mathbb{E}[F(X,Y) | B] \underset{\text{a.s.}}{=} H(X), \tag{2.15}
\]

where \(H(x) := \mathbb{E}[F(x,Y)]\) for all \(x \in S_1\).

**Hint:** Show that the set

\[
\mathcal{F} := \{ F: S_1 \times S_2 \to \mathbb{R} \mid F \text{ is bounded and } S_1 \otimes S_2\text{-measurable satisfying (2.15)} \}
\]

contains all \(F\) of the form \(F(x,y) = 1_A(x)1_B(y)\) with \(A \in \mathcal{S}_1\) and \(B \in \mathcal{S}_2\). Show that the monotone class theorem is applicable.

**Exercise 2.2** (Explicit construction of the general Bernoulli mixture model).

Consider a \([0, 1]^m\)-valued random vector \((P_1, \ldots, P_m)\) and let \(U_1, \ldots, U_m\) be independent random variables, uniformly distributed on \([0, 1]\), and independent of \((P_1, \ldots, P_m)\). Define, for every obligor \(i \in \{1, \ldots, m\}\),

\[
N_i = 1_{[0,P_i]}(U_i) = \begin{cases} 
1 & \text{if } U_i \leq P_i, \\
0 & \text{if } U_i > P_i.
\end{cases}
\]

Use Exercise 2.1 to show that \(N_1, \ldots, N_m\) satisfy (2.10) and (2.11).

**2.2.2 Number of Default Events, Expected Value and Variance**

With the assumptions (2.10) and (2.11) above, it is possible to deduce the expectation and the variance of the total number of default events from the respective properties of the individual random default probabilities. For every obligor \(i \in \{1, \ldots, m\}\),

\[
\mathbb{E}[N_i] = \mathbb{P}[N_i = 1] = \mathbb{E}[\mathbb{P}[N_i = 1 | P_1, \ldots, P_m]] = \mathbb{E}[P_i] \tag{2.16}
\]

by (2.10), where we also used a defining property of conditional expectation, or more directly by (2.14) with \(I = \{i\}\) and \(n_i = 1\). Using (2.2), we obtain for the expected number of defaults, cf. (2.5),

\[
\mathbb{E}[N] = \sum_{i=1}^m \mathbb{E}[N_i] = \sum_{i=1}^m \mathbb{E}[P_i]. \tag{2.17}
\]

8
For the variance, first note that by the general formula for sums of square-integrable random variables,
\[
\text{Var}(N) = \sum_{i=1}^{m} \text{Var}(N_i) + \sum_{i,j=1, i \neq j}^{m} \text{Cov}(N_i, N_j). \tag{2.18}
\]
Using \(N_i^2 = N_i\) for \(\{0, 1\}\)-valued random variables, we obtain in a similar way as in (2.4) for the variance
\[
\text{Var}(N_i) = \mathbb{E}[N_i^2] - (\mathbb{E}[N_i])^2 = \mathbb{E}[N_i] - (\mathbb{E}[N_i])^2 = \mathbb{E}[P_i] (1 - \mathbb{E}[P_i]) \tag{2.19}
\]
for every \(i \in \{1, \ldots, m\}\), where we used (2.16) for the last equality. Next we compute the covariance. From (2.14) we get for \(i \neq j\) in \(\{1, \ldots, m\}\)
\[
\mathbb{E}[N_i N_j] = \mathbb{P}[N_i = 1, N_j = 1] = \mathbb{E}[P_i P_j], \tag{2.20}
\]
hence with (2.16)
\[
\text{Cov}(N_i, N_j) = \mathbb{E}[N_i N_j] - \mathbb{E}[N_i] \mathbb{E}[N_j] \\
= \mathbb{E}[P_i P_j] - \mathbb{E}[P_i] \mathbb{E}[P_j] \\
= \text{Cov}(P_i, P_j). \tag{2.21}
\]
Equations (2.18), (2.19) and (2.21) together yield the variance
\[
\text{Var}(N) = \sum_{i=1}^{m} \mathbb{E}[P_i] (1 - \mathbb{E}[P_i]) + \sum_{i,j=1, i \neq j}^{m} \text{Cov}(P_i, P_j). \tag{2.22}
\]

**Exercise 2.3.** Prove (2.18) for square-integrable random variables \(N_1, \ldots, N_m\).

### 2.3 Uniform Bernoulli Mixture Model

A uniform Bernoulli mixture model is defined as a special case of the general Bernoulli mixture model, where the default probabilities of all obligors are equal (but possibly random), i.e.,
\[
P_1 = P_2 = \cdots = P_m =: P,
\]
where \(P\) is a \([0, 1]\)-valued random variable, whose distribution function we denote by \(F\). The mixing random variable \(P\) can be viewed as a macroeconomic variable driving the default probabilities.

Then, for \(n_1, \ldots, n_m \in \{0, 1\}\) and \(n := n_1 + \cdots + n_m\) the total number of defaults, it follows from (2.13) that
\[
\mathbb{P}[N_1 = n_1, \ldots, N_m = n_m] = \int_{0}^{1} p^n (1 - p)^{m-n} F(dp). \tag{2.23}
\]
Without knowing which obligor defaults (like above), the probability just for \( n \in \{0, \ldots, m\} \) defaults is given by

\[
P[N = n] = \mathbb{E}[\mathbb{P}[N = n \mid P]]
\]

\[
= \mathbb{E}\left( \binom{m}{n} P^n (1 - P)^{m-n} \right)
\]

\[
= \binomial{m}{n} \int_0^1 p^n(1-p)^{m-n} F(dp),
\]

where \( \binomial{m}{n} \) is the usual binomial coefficient describing the number of \( m \)-tuples \((n_1, \ldots, n_m) \in \{0,1\}^m \) with sum \( n \), see (2.9).

In the case of such a uniform portfolio, the expectation in (2.17) reduces to

\[
\mathbb{E}[N] = m \mathbb{E}[P]
\]

and the variance of the total number of defaults can be computed using (2.22). For \( i \neq j \) in \( \{1, \ldots, m\} \) we have \( \text{Cov}(P_i, P_j) = \text{Var}(P) \geq 0 \) and therefore

\[
\text{Var}(N) = m \mathbb{E}[P](1 - \mathbb{E}[P]) + m(m - 1) \text{Var}(P).
\]

The variance therefore is comprised of a binomial part \( m \mathbb{E}[P](1 - \mathbb{E}[P]) \) with success probability \( \mathbb{E}[P] \) and a non-negative additional variance term. In other words, using the uniform Bernoulli mixture model can only increase the variance of the total number of defaults.

A special case of the uniform Bernoulli mixture model is given by the extreme assumption that \( P \) is itself a Bernoulli random variable. Then, either no or all obligors default.

### 2.3.1 Beta-Binomial Mixture Model

Let us consider a more interesting class of distributions on the unit interval \([0, 1]\). Recall that the [gamma function](#gamma-function) is defined by

\[
\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} \, dx, \quad \alpha > 0.
\]

(2.27)

By partial integration,

\[
\alpha \Gamma(\alpha) = \Gamma(\alpha + 1), \quad \alpha > 0,
\]

(2.28)

which is the functional equation of the gamma function. Iterated application of (2.28) yields

\[
\Gamma(\alpha + n) = \Gamma(\alpha) \prod_{i=0}^{n-1} (\alpha + i), \quad \alpha > 0, \, n \in \mathbb{N}_0.
\]

(2.29)

Since \( \Gamma(1) = 1 \) by (2.27), the case \( \alpha = 1 \) shows that \( \Gamma(n + 1) = n! \) for all \( n \in \mathbb{N}_0 \).

---

3The gamma function is actually a [meromorphic function](#meromorphic-function) on the complex plane \( \mathbb{C} \) with poles at 0 and the negative integers, but this will not be used in the following.
Exercise 2.4 (Multivariate beta function). For integer dimension $d \geq 2$ define the open standard orthogonal $(d - 1)$-dimensional simplex (also called lower simplex in the open unit cube) by

$$\Delta_{d-1} = \{(x_1, \ldots, x_{d-1}) \in (0,1)^{d-1} \mid x_1 + \cdots + x_{d-1} < 1\}.$$ 

Show by direct calculation for the multivariate beta function that

$$B(\alpha_1, \ldots, \alpha_d) := \int_{\Delta_{d-1}} \left(\prod_{i=1}^{d-1} x_i^{\alpha_i-1}\right) (1 - x_1 - \cdots - x_{d-1})^{\alpha_d-1} \, dx_1 \cdots dx_{d-1}$$

$$= \frac{\prod_{i=1}^{d} \Gamma(\alpha_i)}{\Gamma(\alpha_1 + \cdots + \alpha_d)}, \quad \alpha_1, \ldots, \alpha_d > 0,$$

which in the case $d = 2$ simplifies to

$$B(\alpha, \beta) := \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} \, dx = \frac{\Gamma(\alpha) \Gamma(\beta)}{\Gamma(\alpha + \beta)}, \quad \alpha, \beta > 0.$$ 

Definition 2.5 (Beta distribution). A density of the beta distribution with real shape parameters $\alpha, \beta > 0$ is given by

$$f_{\alpha,\beta}(p) = \begin{cases} 
\frac{\beta^\alpha(1-p)^{\beta-1}}{B(\alpha,\beta)} & \text{for } p \in (0,1), \\
0 & \text{for } p \in \mathbb{R} \setminus (0,1),
\end{cases}$$

where $B(\alpha, \beta)$ denotes the beta function, see (2.31). For a random variable $P$ with a beta distribution, we use the notation $P \sim \text{Beta}(\alpha, \beta)$.

If the mixing random variable $P$ has a beta distribution, then we can calculate the distribution of the number of defaults more explicitly. From (2.24) we get, for every $n \in \{0,1,\ldots,m\}$,

$$\mathbb{P}[N = n] = \binom{m}{n} \int_0^1 p^n (1-p)^{m-n} \frac{\beta^\alpha(1-p)^{\beta-1}}{B(\alpha,\beta)} \, dp$$

$$= \binom{m}{n} \frac{1}{B(\alpha,\beta)} \int_0^1 p^{\alpha+n-1}(1-p)^{\beta+m-n-1} \, dp$$

$$= \binom{m}{n} \frac{B(\alpha+n,\beta+m-n)}{B(\alpha,\beta)},$$

which is called the beta-binomial distribution with shape parameters $\alpha, \beta > 0$ and $m \in \mathbb{N}_0$ trials. We will use the notation BetaBin($\alpha, \beta, m$).

The proof of Lemma 4.22 below contains a probabilistic argument for the case $d = 2$. 

---

4The proof of Lemma 4.22 below contains a probabilistic argument for the case $d = 2$. 

11
Exercise 2.6 (Moments of the beta distribution). Let \( P \sim \text{Beta}(\alpha, \beta) \) with \( \alpha, \beta > 0 \). Show that
\[
\mathbb{E}[P^\gamma (1 - P)^\delta] = \frac{B(\alpha + \gamma, \beta + \delta)}{B(\alpha, \beta)}, \quad \gamma > -\alpha, \; \delta > -\beta,
\]
and, using the relation (2.31) for the beta function and the functional equation (2.29) of the gamma function, conclude that
\[
\mathbb{E}[P^n] = \prod_{i=0}^{n-1} \frac{\alpha + i}{\alpha + \beta + i}, \quad n \in \mathbb{N},
\]
in particular
\[
\mathbb{E}[P] = \frac{\alpha}{\alpha + \beta} \quad \text{and} \quad \text{Var}(P) = \frac{\alpha \beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.
\]
\[
(2.35)
\]

Exercise 2.7 (Computation of the beta-binomial distribution). Using the relation (2.31) for the beta function and the functional equation (2.28) of the gamma function, show that the beta-binomial distribution (2.33) can be computed in an elementary way by
\[
P[N = n] = \left( \prod_{i=0}^{n-1} \frac{\alpha + i}{\alpha + \beta + i} \right) \left( \prod_{i=0}^{m-n-1} \frac{\beta + i}{\alpha + \beta + i} \right) \prod_{i=0}^{m-1} \frac{i + 1}{\alpha + \beta + i},
\]
for every \( n \in \{0, \ldots, m\} \), and conclude that it can also be calculated recursively from the initial value
\[
P[N = 0] = \prod_{i=0}^{m-1} \frac{\beta + i}{\alpha + \beta + i}
\]
and the recursion formula
\[
P[N = n] = \frac{(\alpha + n - 1)(m - n + 1)}{n(\beta + m - n)} P[N = n - 1], \quad n \in \{1, \ldots, m\},
\]
in a numerically stable way, because only differences of integers are calculated.

Exercise 2.8 (Factorial moments of the beta-binomial distribution). Let \( N \) have a beta-binomial distribution with shape parameters \( \alpha, \beta > 0 \) and \( m \in \mathbb{N} \) trials. Show that, for every \( l \in \{0, \ldots, m\} \), the \( l \)-th factorial moment is given by
\[
\mathbb{E}[l! (N - k)^{l-1}] = \frac{B(\alpha + l, \beta)}{B(\alpha, \beta)} \prod_{k=0}^{l-1} (m - k),
\]
and conclude from (2.36) using \( N^2 = N + N(N-1) \), as well as from (2.25), (2.26) and (2.35), that
\[
\mathbb{E}[N] = \frac{\alpha m}{\alpha + \beta} \quad \text{and} \quad \text{Var}(N) = \frac{\alpha \beta m (\alpha + \beta + m)}{(\alpha + \beta)^2(\alpha + \beta + 1)}.
\]
\[
(2.36)
\]
Exercise 2.9 (Calculating moments from factorial moments). Using the convention $x^0 = 1$, show that in the polynomial ring $R[x]$ over a commutative ring $R$ (with 1),

$$x^n = \sum_{l=0}^{n} \binom{n}{l} \prod_{k=0}^{l-1} (x - k), \quad n \in \mathbb{N}_0,$$

(2.37)

where $\binom{n}{l}$ denotes the Stirling number of the second kind defined recursively by

$$\binom{n+1}{l} = \binom{n}{l-1} + l \binom{n}{l}, \quad l \in \mathbb{N} \text{ and } n \in \mathbb{N}_0,$$

(2.38)

with initial conditions $\binom{0}{0} = 1$, $\binom{n}{0} := 0$ and $\binom{0}{l} := 0$ for $l, n \in \mathbb{N}$. Conclude that, for every $\mathbb{N}_0$-valued random variable $N$, the moments can be calculated from the factorial moments by the formula

$$\mathbb{E}[N^n] = \sum_{l=0}^{n} \binom{n}{l} \mathbb{E}\left[\prod_{k=0}^{l-1} (N - k)\right], \quad n \in \mathbb{N}_0.$$

(2.39)

Show that (2.39) is also true for $\mathbb{C}$-valued random variables, provided the absolute factorial moments for the right-hand side of (2.39) are finite or the absolute $n$th moment for the left-hand side is finite. Explain how (2.39) can be applied to random $\mathbb{C}^{d \times d}$-matrices and see Exercise 4.12 for the multivariate extension.

Hint: Show for all $l, n \in \mathbb{N}$ that $\binom{n}{l} = 0$ if $l > n$ and $\binom{n}{l} = 1$ if $l = n$. Use $x = (x - l) + l$ to prove (2.37).

2.3.2 Biased Measure and the Beta Distribution

Definition 2.10 (Biased probability measure). Let $\Lambda$ be a $[0, \infty)$-valued random variable on a probability space $(\Omega, \mathcal{F}, P)$ such that $0 < \mathbb{E}[\Lambda] < \infty$. Then the $\Lambda$-biased probability measure $P_\Lambda$ on $(\Omega, \mathcal{F})$ is defined by

$$P_\Lambda[A] = \frac{\mathbb{E}[\Lambda A]}{\mathbb{E}[\Lambda]}, \quad A \in \mathcal{F}. \quad (2.40)$$

Some distributions just change their distributional parameters under suitable biasing, an example is the beta distribution.

Lemma 2.11. Assume that $P \sim \text{Beta}(\alpha, \beta)$ with parameters $\alpha, \beta > 0$ and that $\gamma \in (-\alpha, \infty)$ and $\delta \in (-\beta, \infty)$. Then $P_{P^{\gamma}(1-P)^{\delta}} = \text{Beta}(\alpha + \gamma, \beta + \delta)$, that means the distribution of $P$ under the $P^{\gamma}(1-P)^{\delta}$-biased probability measure $P_{P^{\gamma}(1-P)^{\delta}}$ given by Definition 2.10 is the Beta($\alpha + \gamma, \beta + \delta$) distribution.

---

The Stirling number of the second kind $\binom{n}{l}$ gives the number of ways to partition a set of $n \in \mathbb{N}$ elements into $l \in \{1, \ldots, n\}$ non-empty subsets: Obviously $\binom{1}{1} = 1$. To explain the recursion formula (2.38) by induction, you can add $\{n+1\}$ as a new subset to the partition of $\{1, \ldots, n\}$ into $l-1$ subsets, or you can put $n+1$ into one of the $l$ existing sets of the partition.
Proof. By (2.34) and (2.40), a density of the $P^\gamma(1 - P)^\delta$-biased probability measure $P_{P^\gamma(1 - P)^\delta}$ is given by

$$
\frac{dP_{P^\gamma(1 - P)^\delta}}{dP} = \frac{B(\alpha, \beta)}{B(\alpha + \gamma, \beta + \delta)} P^\gamma(1 - P)^\delta.
$$

Let $\mu$ denote the Lebesgue measure on $\mathbb{R}$. Using the density $f_{\alpha, \beta}$ from (2.32) shows that, for $\mu$-almost all $p \in (0, 1)$,

$$
\frac{d(P_{P^\gamma(1 - P)^\delta} - 1)}{d\mu}(p) = \frac{B(\alpha, \beta)}{B(\alpha + \gamma, \beta + \delta)} P^\gamma(1 - P)^\delta \cdot f_{\alpha, \beta}(p)
$$

which by (2.32) gives a density of the Beta($\alpha + \gamma, \beta + \delta$) distribution.

## 2.4 One-Factor Bernoulli Mixture Model

We now introduce a version of the Bernoulli mixture model, which is more restrictive than the general one from Subsection 2.2 in the sense that there is only one (macroeconomic) random variable driving the default probabilities. However, it’s more general than the uniform Bernoulli mixture model of Subsection 2.3, because the individual obligors have susceptibilities $p_1, \ldots, p_m$, which don’t need to be equal.

**Definition 2.12 (One-factor Bernoulli mixture model).** Consider Bernoulli random variables $N_1, \ldots, N_m$. Let $\Lambda$ be a $[0, \infty)$-valued random variable such that $0 < \mathbb{E}[\Lambda] < \infty$. If there exist $p_1, \ldots, p_m \in [0, 1]$ such that

$$
\mathbb{P}[N_i = 1 | \Lambda] \overset{a.s.}{=} p_i \Lambda, \quad i \in \{1, \ldots, m\},
$$

and if $N_1, \ldots, N_m$ are conditionally independent given $\Lambda$, i.e.,

$$
\mathbb{P}[N_1 = n_1, \ldots, N_m = n_m | \Lambda] \overset{a.s.}{=} \prod_{i=1}^m \mathbb{P}[N_i = n_i | \Lambda]
$$

for all $n_1, \ldots, n_m \in \{0, 1\}$, then we call $(N_1, \ldots, N_m, \Lambda)$ a one-factor Bernoulli mixture model with conditional success probabilities $p_1, \ldots, p_m$. If $p_1 = \cdots = p_m$, then we call the model homogeneous.

Condition (2.41) implies that $\max\{p_1, \ldots, p_m\} \Lambda \leq 1 \ \mathbb{P}$-almost surely. Furthermore, $\mathbb{P}[N_i = 1] = \mathbb{E}[\mathbb{P}[N_i = 1 | \Lambda]] = p_i \mathbb{E}[\Lambda]$. Hence in the case $\mathbb{E}[\Lambda] = 1$, the parameters $p_1, \ldots, p_m$ are the individual default probabilities within the next period as introduced in (2.1).

14
Remark 2.13 (Discussion of expectation and variance). Let \((N_1, \ldots, N_m, \Lambda)\) be a one-factor Bernoulli mixture model with conditional success probabilities \(p_1, \ldots, p_m\), let \(N = N_1 + \cdots + N_m\) denote the number of defaults, and define \(\lambda = p_1 + \cdots + p_m\). Then (2.41) implies that
\[
\mathbb{E}[N | \Lambda] \overset{a.s.}{=} (p_1 + \cdots + p_m)\lambda,
\]
hence \(\mathbb{E}[N] = \lambda \mathbb{E}[\Lambda]\). For the variance we see from (2.22) that
\[
\text{Var}(N) = \sum_{i=1}^{m} p_i \mathbb{E}[\Lambda] (1 - p_i \mathbb{E}[\Lambda]) + \sum_{i,j=1}^{m} \text{Cov}(p_i \Lambda, p_j \Lambda) = \lambda \mathbb{E}[\Lambda] - \lambda^2 \left(\mathbb{E}[\Lambda]\right)^2 + (\lambda^2 - \lambda_2) \text{Var}(\Lambda),
\]
which can be smaller or larger than \(\mathbb{E}[N] = \lambda \mathbb{E}[\Lambda]\) depending on \((\lambda^2 - \lambda_2) \text{Var}(\Lambda)\). If \(\lambda^2 = \lambda_2\), then at most one of \(p_1, \ldots, p_m\) is non-zero, and we exclude this uninteresting case of a single Bernoulli random variable in the remaining discussion. Hence \(p_{b} := (\lambda^2 - \lambda_2) / \lambda^2\) defines a strictly positive probability. If, for a given mean \(\mu > 0\), the conditional success probabilities satisfy \(p_i \leq p_b / \mu\) for every \(i \in \{1, \ldots, m\}\), then there exists a random variable \(\Lambda\) with \(\mathbb{E}[\Lambda] = \mu\) and \(p_i \Lambda \leq 1\) for all \(i \in \{1, \ldots, m\}\) satisfying
\[
\text{Var}(\Lambda) = \frac{\lambda_2}{\lambda^2 - \lambda_2} \left(\mathbb{E}[\Lambda]\right)^2;
\]
a simple (but extreme) example is a random variable \(\Lambda\) with \(\mathbb{P}[\Lambda = 0] = 1 - p_b\) and \(\mathbb{P}[\Lambda = \mu / p_b] = p_b\), because \(\mathbb{E}[\Lambda] = \mu\) and \(\mathbb{E}[\Lambda^2] = \mu^2 / p_b\), hence
\[
\text{Var}(\Lambda) = \mathbb{E}[\Lambda^2] - (\mathbb{E}[\Lambda])^2 = \left(1 - \frac{\mu}{p_b}\right) \mu^2 = \frac{\lambda_2}{\lambda^2 - \lambda_2} \mu^2.
\]
In the case (2.45), the expectation and the variance of \(N\) agree, see (2.44).

3 Poisson Models for Credit Defaults

For the application of Poisson models to describe defaults in credit portfolios, it is necessary to look at some of the basic properties of the Poisson distribution.

3.1 Elementary Properties of the Poisson Distribution

Definition 3.1 (Poisson distribution). An \(\mathbb{N}_0\)-valued random variable \(N\) has a Poisson distribution with parameter \(\lambda \geq 0\) if
\[
\mathbb{P}[N = n] = \frac{\lambda^n}{n!} e^{-\lambda}, \quad n \in \mathbb{N}_0,
\]
where we use the convention \(0^0 := 1\). We will use the notation \(N \sim \text{Poisson}(\lambda)\).

\(^6\)Named after the French mathematician Siméon Denis Poisson (1781–1840).
In a credit risk context, if \( N \) describes the number of defaults of an obligor within one period, then mainly the events \( N = 0 \) or \( N = 1 \) are of practical interest. The event \( N = 2 \) would correspond to a default of the obligor after recapitalization, and in principle recapitalization and subsequent default could happen several times.

First we consider moments. Suppose \( N \sim \text{Poisson}(\lambda) \) and \( l \in \mathbb{N}_0 \). Then, by the power series of the exponential function, the \( l \)-th factorial moment of the Poisson distribution is given by

\[
E\left[ \prod_{k=0}^{l-1} (N - k) \right] = \sum_{n=l}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} = \lambda^l e^{-\lambda} \sum_{n=l}^{\infty} \frac{\lambda^{n-l}}{(n-l)!} = \lambda^l. \tag{3.2}
\]

For \( l = 1 \) this gives the expected value

\[
E[N] = \lambda. \tag{3.3}
\]

Using \( N^2 = N + N(N - 1) \) and (3.2) for \( l = 2 \), the variance can be calculated according to

\[
\text{Var}(N) = E[N^2] - (E[N])^2 = E[N] + E[N(N - 1)] - (E[N])^2 = \lambda + \lambda^2 - \lambda^2 = \lambda. \tag{3.4}
\]

To calculate higher moments of \( N \), use (2.39) from Exercise 2.9.

Another very important feature of Poisson distributed random variables is their summation property: The sum of independent Poisson distributed random variables is again a Poisson distributed random variable with parameter given by the sum of the respective parameters.

**Lemma 3.2** (Summation property of the Poisson distribution). If \( N_1, \ldots, N_k \) are independent with \( N_i \sim \text{Poisson}(\lambda_i) \) for all \( i \in \{1, \ldots, k\} \), then

\[
N := \sum_{i=1}^{k} N_i \sim \text{Poisson}(\lambda_1 + \cdots + \lambda_k). \tag{3.5}
\]

We give a direct proof below, for a short one using probability-generating functions, see (4.28). For the multivariate generalization, see Lemma 3.41.

**Proof of Lemma 3.2**. For the proof, we first consider the case \( k = 2 \), i.e., the sum of two independent Poisson distributed random variables.

Let \( X \sim \text{Poisson}(\lambda) \) and \( Y \sim \text{Poisson}(\mu) \) be independent and let \( n \in \mathbb{N}_0 \). Then, by considering all possibilities to get the sum \( n \),

\[
P[X + Y = n] = \sum_{l=0}^{n} P[X = n - l, Y = l] = \sum_{l=0}^{n} e^{-\lambda} \frac{\lambda^{n-l}}{(n-l)!} e^{-\mu} \frac{\mu^l}{l!} = e^{-(\lambda+\mu)} \frac{1}{n!} \sum_{l=0}^{n} \binom{n}{l} \lambda^{n-l} \mu^l, \tag{3.6}
\]
where we used the factorial definition of the binomial coefficient and the binomial theorem at the end. Hence $X + Y \sim \text{Poisson}(\lambda + \mu)$. The rest of the proof follows by mathematical induction on the number $k$ of random variables.

Remark 3.3 (**Infinite divisibility** of the Poisson distribution). Lemma 3.2 implies that, for every $\lambda \geq 0$, the Poisson distribution $\text{Poisson}(\lambda)$ is infinitely divisible, because for every $k \in \mathbb{N}$ the distribution of $N_1 + \cdots + N_k$ is $\text{Poisson}(\lambda)$, when $N_1, \ldots, N_k$ are independent with $N_i \sim \text{Poisson}(\lambda/k)$ for every $i \in \{1, \ldots, k\}$.

Remark 3.4 (**Raikov’s theorem**). The summation property in Lemma 3.2 characterizes the Poisson distribution in the following sense: Given $k \in \mathbb{N}$ independent, real-valued random variables $N_1, \ldots, N_k$ such that $N_1 + \cdots + N_k \sim \text{Poisson}(\lambda)$, then there exist $a_1, \ldots, a_k \in \mathbb{R}$ and $\lambda_1, \ldots, \lambda_k \in [0, \lambda]$ with $a_1 + \cdots + a_k = 0$ and $\lambda_1 + \cdots + \lambda_k = \lambda$ such that $N_i' := N_i + a_i \sim \text{Poisson}(\lambda_i)$ for every $i \in \{1, \ldots, k\}$. If, in addition, $N_1, \ldots, N_k$ are assumed to be non-negative, then $a_1 = \cdots = a_k = 0$ and $N_i \sim \text{Poisson}(\lambda_i)$ for every $i \in \{1, \ldots, k\}$. This general case of Raikov’s theorem follows from the case $k = 2$ by induction. The proof for $k = 2$ uses the Hadamard factorization theorem from complex analysis, hence we omit the more involved part of the proof here.

### 3.2 Calibration of the Poisson Distribution

There are at least three choices of calibration for the Poisson parameter $\lambda$ of a random variable $N \sim \text{Poisson}(\lambda)$:

(a) Given $p \in [0, 1)$, choose $\lambda \in [0, \infty)$ so that

$$1 - e^{-\lambda} = \mathbb{P}[N \geq 1] = p,$$

or equivalently, using the Taylor expansion,

$$\lambda = -\log(1 - p) = \sum_{n=1}^{\infty} \frac{p^n}{n} = p + \frac{1}{2}p^2 + \frac{1}{3}p^3 + \cdots. \quad (3.8)$$

In this way the probability of no default coincides with the one in the Bernoulli model.

(b) Given $p \in [0, 1]$, choose $\lambda \in [0, 1]$ so that the expected number of defaults fits with the one in the Bernoulli model, i.e.

$$\lambda = \mathbb{E}[N] = p, \quad (3.9)$$

where (2.3) for the expectation of a Bernoulli random variable and (3.3) for the expectation of $N$ are used.

(c) Given $p \in [0, 1]$, choose $\lambda \in [0, 1/4]$ so that the variance of the number of defaults equals the corresponding variance in the Bernoulli model, i.e.

$$\lambda = \text{Var}(N) = p(1 - p), \quad (3.10)$$

where (2.4) for the variance of a Bernoulli random variable and (3.4) for the variance of $N$ are used.
Note that, using the expansion (3.8), the results of the three calibration methods (3.7), (3.9) and (3.10) are ordered in the sense that $-\log(1-p) \geq p \geq p(1-p)$ for $p \in [0,1)$ with equality only for $p = 0$. For small $p$ the expansion (3.8) justifies the approximations

$$-\log(1-p) \approx p \approx p(1-p),$$

hence the three methods above give very similar results for small $p$. For $p$ close to 1, the three methods give quite different results, and the “good” one depends on the purpose; in most cases the calibration (3.9) will be the appropriate one.

### 3.3 Metrics for Spaces of Probability Measures

To quantify the quality of the Poisson approximation in the next section, we need a way to measure the distance between probability measures. To this end, let $(S, \mathcal{S})$ denote a measurable space, $\mathcal{M}_1(S, \mathcal{S})$ the set of all probability measures on $(S, \mathcal{S})$, and $\mathcal{F}$ a non-empty set of real-valued, measurable functions on $(S, \mathcal{S})$. When it is clear from the context, we will suppress the $\sigma$-algebra in the notation.

Define the set

$$\mathcal{M}_1^\mathcal{F}(S) = \{ \mu \in \mathcal{M}_1(S) \mid \int_S |f| \, d\mu < \infty \text{ for all } f \in \mathcal{F} \} \quad (3.11)$$

defines an $\mathbb{R}_+$-valued pseudometric on $\mathcal{M}_1^\mathcal{F}(S)$, meaning that $d_\mathcal{F}$ is non-negative, symmetric, and satisfies the triangle inequality. However, $d_\mathcal{F}(\mu, \nu) = 0$ does not need to imply $\mu = \nu$. To ensure that $d_\mathcal{F}(\mu, \nu) = 0$ actually implies that $\mu = \nu$, it suffices that $\mathcal{F}$ separates the probability measures in $\mathcal{M}_1^\mathcal{F}(S)$, meaning that for every choice of $\mu, \nu \in \mathcal{M}_1^\mathcal{F}(S)$ with $\mu \neq \nu$ there exists an $f \in \mathcal{F}$ such that $\int_S f \, d\mu \neq \int_S f \, d\nu$.

**Remark 3.5.** Note that the supremum in (3.12) can result in $d_\mathcal{F}(\mu, \nu) = \infty$, which normally is not an allowed value for a metric or a pseudometric. This already happens with $S = \{0,1\}$ and $\mathcal{F}$ the set for bounded functions on $S$, just take $\mu = \delta_0$, $\nu = \delta_1$ and $f_n(x) = nx$ for $n \in \mathbb{N}$ and $x \in S$. This problem can be rectified by choosing a real number $r > 0$ and considering the bounded (pseudo-)metric $d'_\mathcal{F}(\mu, \nu) := \min\{r, d_\mathcal{F}(\mu, \nu)\}$. However, in the first two examples we consider, the functions in $\mathcal{F}$ are bounded by 1, and in the third example of the Wasserstein metric for probability measures on a metric space $(S, d)$ (see Definition 3.14 below), this problem does not occur, see Remark 3.15.

---

We will mainly need $S = \mathbb{N}_0$ and $S = \mathbb{R}$ with $\mathcal{S}$ denoting the set $\mathcal{P}(\mathbb{N}_0)$ of all subsets of $\mathbb{N}_0$ or the Borel $\sigma$-algebra $\mathcal{B}_\mathbb{R}$ on $\mathbb{R}$, respectively.
Remark 3.6. If, for every $f \in \mathcal{F}$, there exists a constant $c_f \in \mathbb{R}$ such that $c_f - f$ is also in $\mathcal{F}$, then
\[ \int_S (c_f - f) \, d\mu - \int_S (c_f - f) \, d\nu = \int_S f \, d\nu - \int_S f \, d\mu, \quad f \in \mathcal{F}, \]
because $\mu$ and $\nu$ are probability measures, hence we can omit the absolute value in the definition (3.12) of $d_F$.

We will consider three different choices for $\mathcal{F}$, giving rise to three different metrics. The first one arises from the set $\mathcal{F}_{\mathrm{TV}} = \{ 1_A \mid A \in \mathcal{S} \}$ of all indicator functions, which has the property discussed in Remark 3.6 with $c_f = 1$, and which by definition separates the probability measures in $\mathcal{M}_1(S)$.

**Definition 3.7** (Total variation metric). The total variation metric $d_{\mathrm{TV}}$ on the set $\mathcal{M}_1(S)$ of all probability measures on the measurable space $(S, \mathcal{S})$ is defined by
\[ d_{\mathrm{TV}}(\mu, \nu) = \sup_{A \in \mathcal{S}} (\mu(A) - \nu(A)), \quad \mu, \nu \in \mathcal{M}_1(S). \]

**Remark 3.8.** Note that $d_{\mathrm{TV}}(\mu, \nu) \leq 1$ for all $\mu, \nu \in \mathcal{M}_1(S)$. If $\mu$ and $\nu$ are mutually singular, then $d_{\mathrm{TV}}(\mu, \nu) = 1$. The reverse direction is also true and follows from Exercise [3.19] below.

For many applications, in particular when proving convergence of the distributions of $\mathbb{R}^d$-valued random variables, the total variation metric is too strong. Therefore, in the case $S = \mathbb{R}^d$ with Borel $\sigma$-algebra $\mathcal{B}_{\mathbb{R}^d}$, we consider the collection
\[ \mathcal{F}_{\mathrm{KS}} = \{ 1_{(-\infty,a_1] \times \cdots \times (-\infty,a_d]} \mid (a_1, \ldots, a_d) \in \mathbb{R}^d \}. \]
Since the distribution function $F_\mu$ of a probability measure $\mu$ on $\mathbb{R}^d$, defined by $F_\mu(a_1, \ldots, a_d) = \mu((-\infty, a_1] \times \cdots \times (-\infty, a_d])$ for all $(a_1, \ldots, a_d) \in \mathbb{R}^d$, uniquely determines $\mu$, the collection $\mathcal{F}_{\mathrm{KS}}$ separates the probability measures on $\mathbb{R}^d$.

**Definition 3.9** (Kolmogorov–Smirnov metric). The Kolmogorov–Smirnov metric $d_{\mathrm{KS}}$ – sometimes just called Kolmogorov metric – on the set $\mathcal{M}_1(\mathbb{R}^d)$ of all probability measures on $\mathbb{R}^d$ is defined by
\[ d_{\mathrm{KS}}(\mu, \nu) = \sup_{a \in \mathbb{R}^d} |F_\mu(a) - F_\nu(a)| = \|F_\mu - F_\nu\|_\infty, \quad \mu, \nu \in \mathcal{M}_1(\mathbb{R}^d), \]
where $F_\mu$ and $F_\nu$ denote the distribution functions of $\mu$ and $\nu$, respectively.

---

8There are other notions of “distances” for probability measures like the Hellinger metric, the $p$th Wasserstein metric for $p > 1$, the Levy–Prokhorov metric, metricizing the so-called weak topology, the Kullback–Leibler divergence (which is not a metric), and so on. For connections to optimal transport, see the textbooks by C. Villani [33, 34].

9For a proof, show that $\mathcal{E} := \{ (-\infty, a_1] \times \cdots \times (-\infty, a_d] \mid (a_1, \ldots, a_d) \in \mathbb{R}^d \}$ is intersection-stable and generates $\mathcal{B}_{\mathbb{R}^d}$. Then consider for $\mu$ and $\tilde{\mu}$ with $F_\mu = F_{\tilde{\mu}}$ the set $\mathcal{D} := \{ A \in \mathcal{B}_{\mathbb{R}^d} \mid \mu(A) = \tilde{\mu}(A) \}$ and apply Dynkin’s lemma to conclude that $\mu = \tilde{\mu}$.

10Named after Andrey Kolmogorov (1903–1987) and Nikolai Smirnov (1900–1966), because the metric appears in the test statistic in their Kolmogorov–Smirnov test.
Remark 3.10. For probability measures \( \mu \) and \( \nu \) on \( \mathbb{R}^d \), it follows from \( \mathcal{F}_{KS} \subset \mathcal{F}_{TV} \) that
\[
d_{KS}(\mu, \nu) \leq d_{TV}(\mu, \nu).
\] (3.13)

The Kolmogorov–Smirnov metric is useful to obtain estimates for quantiles and value-at-risk, see Lemma 7.7 below. Remark 3.10 implies that \( d_{TV} \) generates a (not necessarily strictly) finer topology on \( \mathcal{M}_1(\mathbb{R}^d) \) and that convergence with respect to \( d_{TV} \) implies convergence with respect to \( d_{KS} \). The following example shows that the converse is not true in general, hence the metrics \( d_{TV} \) and \( d_{KS} \) generate different topologies on \( \mathcal{M}_1(\mathbb{R}^d) \).

Example 3.11. Let \( \mu \) denote the uniform distribution on \([0, 1]\) and define \( \mu_n = (1/n) \sum_{i=1}^n \delta_{i/n} \). Then \( \mu([1/n, \ldots, n/n]) = 0 \) and \( \mu_n([1/n, \ldots, n/n]) = 1 \), hence \( d_{TV}(\mu, \mu_n) = 1 \) by Remark 3.8 while \( d_{KS}(\mu, \mu_n) \leq 1/n \) for all \( n \in \mathbb{N} \).

The next example shows that weak convergence does not imply convergence in the Kolmogorov–Smirnov metric.

Example 3.12. Consider the probability measures \( \mu = \delta_0 \) and \( \mu_n = \delta_{1/n} \) on \( \mathbb{R} \). Then \( \mu((-\infty, 0]) = 1 \) and \( \mu_n((-\infty, 0]) = 0 \), hence \( d_{KS}(\mu, \mu_n) = 1 \) for every \( n \in \mathbb{N} \). On the other hand, \( \int_S f \, d\mu_n = f(1/n) \to f(0) = \int_S f \, d\mu \) as \( n \to \infty \) for every bounded and continuous function \( f: \mathbb{R} \to \mathbb{R} \), which means weak convergence of \( (\mu_n)_{n \in \mathbb{N}} \) to \( \mu \).

For the last one of the three metrics, consider a metric space \((S, d)\) with Borel \( \sigma \)-algebra \( \mathcal{S} \) and let \( \mathcal{F}_W \) denote the set of all functions \( f: S \to \mathbb{R} \), which are Lipschitz continuous with constant at most 1, i.e.,
\[
|f(x) - f(y)| \leq d(x, y), \quad x, y \in S.
\]
Note that \( \mathcal{F}_W \) has the property discussed in Remark 3.6 with \( c_f = 0 \). Define \( \mathcal{M}_1^F(S) \) according to (3.11).

Exercise 3.13. Let \((S, d)\) be a metric space. Show that already the bounded functions in \( \mathcal{F}_W \) separate the probability measures in \( \mathcal{M}_1(S) \).

Hint: Consider \( f_A(n) = (1 - n \operatorname{dist}(A, x))^+ \) for closed \( A \subset S \) and \( n \in \mathbb{N} \).

Definition 3.14 (Wasserstein metric). Let \((S, d)\) be metric space with Borel \( \sigma \)-algebra \( \mathcal{S} \). The Wasserstein metric\[d_{W}(\mu, \nu) = \sup_{f \in \mathcal{F}_W} \left( \int_S f \, d\mu - \int_S f \, d\nu \right), \quad \mu, \nu \in \mathcal{M}_1^F(S). \] (3.14)

Remark 3.15 (The Wasserstein metric is well-defined on \( \mathcal{M}_1^F(S) \)). Consider a point \( x_0 \in S \) and two probability measures \( \mu, \nu \in \mathcal{M}_1^F(S) \). Then, for every function \( f: S \to \mathbb{R} \) having Lipschitz constant
\[
\operatorname{Lip}(f) := \sup_{x, y \in S} \frac{|f(x) - f(y)|}{d(x, y)} < \infty,
\] (3.15)
\[\]Named after the Russian mathematician Leonid Vaserestein, most English-language publications use the German spelling Wasserstein. The metric is also known as Dudley, Fortet–Mourier, and Kantorovich \( D_{1,1} \) metric.
the expectations \( \int_S f \, d\mu \) and \( \int_S f \, d\nu \) are well-defined, because \( |f(x)| \leq |f(x_0)| + \operatorname{Lip}(f) d(x,x_0) \) for all \( x \in S \), and the function \( S \ni x \mapsto d(x,x_0) \in \mathbb{R} \) is in \( \mathcal{F}_W \) by the reverse triangle inequality. Furthermore,

\[
\left| \int_S f \, d\mu - \int_S f \, d\nu \right| = \left| \int_S (f(x) - f(x_0)) \mu(dx) - \int_S (f(x) - f(x_0)) \nu(dx) \right|
\leq \operatorname{Lip}(f) \left( \int_S d(x,x_0) \mu(dx) + \int_S d(x,x_0) \nu(dx) \right),
\]

which in particular implies that \( d_W(\mu,\nu) \) in (3.14) is finite, cf. Remark 3.5.

**Remark 3.16 (Bounds for the Wasserstein metric).** Consider two probability measures \( \mu,\nu \in \mathcal{M}_1^W(S) \). Let \((X,Y)\) be an \((S \times S)\)-valued random variable, defined on a probability space \((\Omega,\mathcal{A},\mathbb{P})\), such that \( \mathcal{L}(X) = \mu \) and \( \mathcal{L}(Y) = \nu \). Suppose the function \( f: S \to \mathbb{R} \) has Lipschitz constant \( \operatorname{Lip}(f) < \infty \). If \( \operatorname{Lip}(f) = 0 \), then \( f \) is constant. If \( \operatorname{Lip}(f) > 0 \), then the function \( f/\operatorname{Lip}(f) \) has Lipschitz constant 1. Hence Definition 3.14 implies the lower bound

\[
|\mathbb{E}[f(X)] - \mathbb{E}[f(Y)]| \leq \operatorname{Lip}(f) d_W(\mu,\nu),
\]

which will be used in Lemma 7.25 below to estimate differences of expected shortfalls. If the metric space \((S,d)\) is separable (and equipped with the Borel \( \sigma \)-algebra \( S \) as before), then the metric \( d: S \times S \to [0,\infty) \) is \( S \otimes S \)-measurable, hence \( d(X,Y) \) is a random variable, see [47, Corollary 15.75]. Then by (3.15), for every function \( f: S \to \mathbb{R} \) with Lipschitz constant \( \operatorname{Lip}(f) < \infty \),

\[
|\mathbb{E}[f(X)] - \mathbb{E}[f(Y)]| \leq \mathbb{E}[|f(X) - f(Y)|] \leq \operatorname{Lip}(f) \mathbb{E}[d(X,Y)],
\]

and taking the supremum in (3.17) over all functions \( f \) with \( \operatorname{Lip}(f) \leq 1 \),

\[
d_W(\mu,\nu) = \sup_{f \in \mathcal{F}_W} \mathbb{E}[|f(X)| - |f(Y)|] \leq \mathbb{E}[d(X,Y)].
\]

To obtain a good upper bound, we can optimize the right-hand side of (3.18) with respect to the dependence of \( X \) and \( Y \).

The next example shows that weak convergence in general does not imply convergence in the Wasserstein metric, because there are unbounded functions in \( \mathcal{F}_W \). See Exercise 3.22 below for a proper characterization in case of a normed vector space.

**Example 3.17.** Define the probability measures \( \mu = \delta_0 \) and \( \mu_n = (1 - 1/n)\delta_0 + (1/n)\delta_n \) on \( \mathbb{R} \). Using the function \( \mathbb{R} \ni x \mapsto |x| \), it follows from Definition 3.14 that \( d_W(\mu,\mu_n) \geq 1 \) for all \( n \in \mathbb{N} \). On the other hand, \( |\int_S f \, d\mu - \int_S f \, d\mu_n| = |f(0) - f(n)| / n \leq 2\|f\|_\infty / n \to 0 \) as \( n \to \infty \), for every bounded and continuous function \( f: \mathbb{R} \to \mathbb{R} \), which verifies weak convergence.

**Lemma 3.18 (Total variation and Wasserstein on \( \mathcal{M}_1(\mathbb{N}_0) \)).** Let \( S \neq \emptyset \) be a finite or countable infinite set. Then, for all \( \mu,\nu \in \mathcal{M}_1(S,\mathcal{P}(S)) \):

21
(a) A set $A \subset S$ satisfies $d_{TV}(\mu, \nu) = \mu(A) - \nu(A)$ if and only if $A \subset \{n \in S \mid \mu(\{n\}) \geq \nu(\{n\})\}$ and $A^c \subset \{n \in S \mid \mu(\{n\}) \leq \nu(\{n\})\}$.

(b) $d_{TV}(\mu, \nu) = \frac{1}{2} \sum_{n \in S} |\mu(\{n\}) - \nu(\{n\})|.$

(c) Let $S \subset \mathbb{Z}$ with the usual distance. If $\mu$ and $\nu$ have finite expectation, i.e.

$$\sum_{n \in S} |n| \mu(\{n\}) < \infty \quad \text{and} \quad \sum_{n \in S} |n| \nu(\{n\}) < \infty,$$

(3.19)

then $d_{TV}(\mu, \nu) \leq d_{W}(\mu, \nu)$.

For $S \subset \mathbb{Z}$ the Wasserstein distance $d_{W}(\mu, \nu)$ between the probability measures $\mu$ and $\nu$ takes into account not only the amounts by which their individual probabilities differ, as in the total variation distance $d_{TV}(\mu, \nu)$, but also how far apart the differences occur, which explains the inequality in part (c) above.

**Proof of Lemma 3.18.** (a), (b) Let $e_n := \mu(\{n\}) - \nu(\{n\})$ denote the approximation error for $n \in S$. Then, for every $A \subset S$,

$$\frac{1}{2} \sum_{n \in S} |e_n| \geq \frac{1}{2} \sum_{n \in A} e_n - \frac{1}{2} \sum_{n \in S \setminus A} e_n = \sum_{n \in A} e_n - \frac{1}{2} \sum_{n \in S} e_n = \mu(A) - \nu(A),$$

where the inequality is an equality if and only if $|e_n| = e_n$ for every $n \in A$ and $|e_n| = -e_n$ for every $n \in S \setminus A$.

(c) Due to (3.19), it follows as in Remark 3.15 that the Wasserstein distance $d_{W}(\mu, \nu)$ is well-defined. Given a set $A \subset S$, the indicator function $1_A: S \rightarrow \mathbb{R}$ is Lipschitz continuous on $S \subset \mathbb{Z}$ with constant at most 1, hence (c) follows from the Definitions 3.7 and 3.14.

**Exercise 3.19** (Representation of the total variation distance with densities). Let $(S, \mathcal{S})$ be a measurable space and consider $\mu, \nu \in \mathcal{M}_1(S, \mathcal{S})$. Let $\lambda$ be a non-negative $\sigma$-finite measure on $(S, \mathcal{S})$ such that $\mu \ll \lambda$ and $\nu \ll \lambda$ (such a measure always exists, take $\lambda = \mu + \nu$, for example, or the counting measure when $S$ is countable). By the Radon–Nikodým theorem there exist corresponding probability densities $f = d\mu/d\lambda$ and $g = d\nu/d\lambda$.

(a) Generalize Lemma 3.18(a) by proving that a set $A \subset S$ satisfies $d_{TV}(\mu, \nu) = \mu(A) - \nu(A)$ if and only if there exists a set $N \subset S$ with $\lambda(N) = 0$ such that $A \setminus N \subset \{x \in S \mid f(x) \geq g(x)\}$ and $A^c \setminus N \subset \{x \in S \mid f(x) \leq g(x)\}$.

(b) Generalize Lemma 3.18(b) by proving that $d_{TV}(\mu, \nu) = \frac{1}{2} \|f - g\|_{L^1(\lambda)}$.

(c) Derive from part (b) that $d_{TV}(\mu, \nu) = 1 - \|\min\{f, g\}\|_{L^1(\lambda)}$ and compare with Remark 3.8.

**Hint:** (c) Verify and use $\frac{1}{2}|a - b| = \frac{1}{2}(a + b) - \min\{a, b\}$ for all $a, b \in \mathbb{R}$. 22
Exercise 3.20 (Total variation norm). Let \((S, \mathcal{S})\) be a measurable space and consider the set \(\mathcal{M}(S, S)\) of all \(\mathbb{R}\)-valued (or \(\mathbb{C}\)-valued) measures on \((S, \mathcal{S})\). Let \(\mathcal{D}\) be a measure-determining subset of \(\mathcal{S}\), meaning that \(\mu(A) = 0\) for all \(A \in \mathcal{D}\) is only possible if \(\mu \in \mathcal{M}(S, S)\) is the zero measure, i.e. \(\mu(A) = 0\) for all \(A \in \mathcal{S}\). Prove:

(a) \(\|\mu\|_D := \sup_{A \in \mathcal{D}} |\mu(A)|\) for \(\mu \in \mathcal{M}(S, S)\) defines a norm.

Hint: Measures on \(\sigma\)-algebras with values in \(\mathbb{R}\) or \(\mathbb{C}\) are always bounded, see e.g. [43, Theorem 6.4].

For \(\mathcal{D} = \mathcal{S}\) this is the total variation norm \(\|\cdot\|_{TV}\). In particular, \((\mathcal{M}(S, S), \|\cdot\|_D)\) is a normed vector space. Prove in addition:

(b) \((\mathcal{M}(S, S), \|\cdot\|_{TV})\) is a Banach space.

Hint: When showing completeness, \(\sigma\)-additivity of the limiting candidate \(\mu\) has to be shown. For this purpose, given a sequence \((A_k)_{k \in \mathbb{N}} \subset \mathcal{S}\) of disjoint sets and \(\varepsilon > 0\), show that there exists \(m_\varepsilon \in \mathbb{N}\) such that \(|\mu(\bigcup_{k \in \mathbb{N}} A_k) - \sum_{k=1}^{m} \mu(A_k)| \leq \varepsilon\) for all \(m \geq m_\varepsilon\).

(c) If \(\mathcal{D}' \subset \mathcal{S}\) with \(\mathcal{D}' \supset \mathcal{D}\), then \(\|\mu\|_\mathcal{D} \leq \|\mu\|_{\mathcal{D}'}\) for all \(\mu \in \mathcal{M}(S, S)\).

(d) \(\mathcal{D} = \{\mathbb{N}\} \cup \{\{k\} : k \in \mathbb{N}\}\) is measure-determining for \(\mathcal{P}(\mathbb{N})\), but the normed space \((\mathcal{M}(\mathbb{N}, \mathcal{P}(\mathbb{N})), \|\cdot\|_D)\) is not complete.

Hint: For \(n \in \mathbb{N}\) consider the discrete uniform probability distribution \(\mu_n\) on \([1, \ldots, n]\).

(e) Explain where the proof of \(\sigma\)-additivity for a limiting candidate \(\mu\) in item \([\text{i}]\) goes wrong when the sequence \((A_k)_{k \in \mathbb{N}}\) with \(A_k = \{k\}\) in the setting of \([\text{d}]\) is considered.\[12\]

Exercise 3.21 (Scaling property of the Wasserstein metric). Let \((S, \|\cdot\|)\) denote a normed vector space over \(\mathbb{K} = \mathbb{R}\) or \(\mathbb{K} = \mathbb{C}\). Let \(X\) and \(Y\) be \(S\)-valued random vectors with \(\mathbb{E}[\|X\|] < \infty\) and \(\mathbb{E}[\|Y\|] < \infty\). Prove that, for every \(c \in \mathbb{K} \setminus \{0\}\),

\[d_W(\mathcal{L}(cX), \mathcal{L}(cY)) = |c| d_W(\mathcal{L}(X), \mathcal{L}(Y)).\]

Hint: For \(f: S \to \mathbb{R}\) with \(\text{Lip}(f) \leq 1\) consider \(f_c(x) := \frac{1}{|c|} f(cx)\) for \(x \in S\).

Exercise 3.22 (Characterization of convergence in the Wasserstein metric). Let \((S, \|\cdot\|)\) be a normed real or complex vector space, \((X_n)_{n \in \mathbb{N}}\) a sequence of \(S\)-valued random vectors with \(\mathbb{E}[\|X_n\|] < \infty\) for every \(n \in \mathbb{N}\), and \(\mu \in \mathcal{M}_1(S)\).

(a) Prove that \([\text{i}]\) implies \([\text{ii}]\):

(i) \(\int_S \|x\| \mu(dx) < \infty\) and \(d_W(\mathcal{L}(X_n), \mu) \to 0\) as \(n \to \infty\).

---

12To learn how to use Zorn’s lemma to produce non-trivial \(\{0, 1\}\)-valued additive set functions on \((\mathbb{N}, \mathcal{P}(\mathbb{N}))\), which are not \(\sigma\)-additive, see [44] Chapter V, Sections 10, Problems 34–41.
(ii) The set \( \{X_n\}_{n \in \mathbb{N}} \) is uniformly integrable, i.e.
\[
\lim_{c \to \infty} \sup_{n \in \mathbb{N}} \mathbb{E}[\|X_n\|1_{\{\|X_n\| > c\}}] = 0,
\]
and converges weakly to \( \mu \).

Due to the claimed uniformity in \( f \in \mathcal{F}_W \), the reverse implication is more involved than just (e) and (f) and outlined here with stronger assumptions on \( (S, \|\cdot\|) \), see (\textsuperscript{ii}) below. The proof is divided into several steps.

(b) For \( C \subset S \) let \( f: C \to \mathbb{R} \) denote a function with \( \text{Lip}(f) \leq 1 \). Show that
\[
g(x) := \inf_{z \in C} (f(z) + \|z - x\|)
\]
for all \( x \in S \) is in \( \mathcal{F}_W \) and extends \( f \).

Always assume (ii) for the following steps.

(c) Prove that \( \int_S |x| \mu(dx) < \infty \).

(d) Prove for each \( f \in \mathcal{F}_W \) that
\[
\lim_{n \to \infty} \mathbb{E}[f(X_n)] = \int_S f \, d\mu.
\]
It remains to show that the convergence in (d) is uniform in \( f \in \mathcal{F}_W \).

(e) For each \( b > 0 \) define \( \mathcal{F}_{W,b} := \{ f \in \mathcal{F}_W \mid \|f\|_\infty \leq b \} \) and assume that
\[
d_{W,b}(\mathcal{L}(X_n), \mu) := \sup_{f \in \mathcal{F}_{W,b}} \left( \mathbb{E}[f(X_n)] - \int_S f \, d\mu \right) \to 0 \quad \text{as } n \to \infty.
\]
Prove that \( d_W(\mathcal{L}(X_n), \mu) \to 0 \) as \( n \to \infty \).

(f) When \( (S, \|\cdot\|) \) is a separable Banach space, prove that the assumption in (e) is satisfied.

Hints: \([\text{a}]\) You may use that weak convergence of probability measures on metric spaces is determined by all integrals over bounded Lipschitz continuous functions, see [17, Chapter 3, Theorem 3.1, proof of (c) implies (d)]. For \( b > 0 \) the Lipschitz continuous function \( h_b \), defined by \( h_b(x) = \max\{0, \|x\| - \max\{0, b(b - \|x\|)\}\} \) for all \( x \in S \), can be useful. \([\text{b}]\) Uniform integrability implies boundedness in \( L^1(P) \), i.e. \( \sup_{n \in \mathbb{N}} \mathbb{E}[\|X_n\|] < \infty \). Define \( f_n(x) = \min\{\|x\|, n\} \) for all \( x \in S \), use the monotone convergence theorem. \([\text{c}]\) Uniform integrability implies boundedness in \( L^1(P) \), i.e. \( \sup_{n \in \mathbb{N}} \mathbb{E}[\|X_n\|] < \infty \). Define \( f_n(x) = \min\{\|x\|, n\} \) for all \( x \in S \), use the monotone convergence theorem. \([\text{d}]\) Restrict to \( f \in \mathcal{F}_{W,b}^0 := \{ f \in \mathcal{F}_W \mid f(0) = 0 \} \), then use \( |f(x)| \leq \|x\| \) for \( x \in S \), part (c) and uniform integrability of \( \{X_n\}_{n \in \mathbb{N}} \). \([\text{e}]\) Similar to (d). \([\text{f}]\) For a proof by contradiction, assume that there are \( b, \varepsilon > 0 \) such that, after passing to a subsequence if necessary, \( d_{W,b}(\mathcal{L}(X_n), \mu) \geq 8\varepsilon \) for every \( n \in \mathbb{N} \). Then there exists a sequence \( (f_n)_{n \in \mathbb{N}} \) in \( \mathcal{F}_{W,b} \) such that \( \mathbb{E}[f_n(X_n)] - \int_S f_n \, d\mu \geq 7\varepsilon \) for each \( n \in \mathbb{N} \). By Prokhorov’s theorem (see e.g. [17, Chapter 3, Theorem 2.2 in combination with Theorem 3.1]), there exists a compact subset \( C \) of \( S \) such that \( \mu(C) \geq 1 - \varepsilon/b \) and \( \mathbb{P}[X_n \in C] \geq 1 - \varepsilon/b \) for every \( n \in \mathbb{N} \). By the Arzela–Ascoli theorem there exists a subsequence \( (f_{n_k})_{k \in \mathbb{N}} \) converging uniformly to a function \( f: C \to \mathbb{R} \). Show that \( \|f\|_\infty \leq b \) and \( \text{Lip}(f) \leq 1 \). Apply (f), define \( h(x) = \min\{b, \max\{-b, g(x)\}\} \) for \( x \in S \), verify
that \( h(x) = f(x) \) for all \( x \in C \), and \( h \in \mathcal{F}_{\mathcal{W},b} \) as well as \( \| f_n - h \|_{\infty} \leq 2b \). Take \( k_\varepsilon \) so large that \( |f_{nk}(x) - h(x)| \leq \varepsilon \) for all \( x \in C \) and \( k \geq k_\varepsilon \). Then
\[
\mathbb{E}[f_{nk}(X_{nk})] - \int_S f_{nk} \, d\mu \leq 6\varepsilon + \mathbb{E}[h(X_{nk})] - \int_S h \, d\mu, \quad k \geq k_\varepsilon,
\]
and the assumption of weak convergence can be applied to \( h \).

### 3.4 Poisson Approximation

In this section we show that the distribution of a sum of independent Bernoulli random variables can be approximated by a Poisson distribution. The quality of the approximation is measured by the total variation metric \( d_{TV} \) of probability distributions as well as the Wasserstein metric \( d_{W} \), see Definitions 3.7 and 3.14, respectively.

**Theorem 3.23.** Let \( X_1, \ldots, X_m \) be independent Bernoulli random variables. Then \( W := X_1 + \cdots + X_m \) is the random variable counting the number of ones. Define \( p_i = \mathbb{P}[X_i = 1] \) and \( \lambda = \mathbb{E}[W] = p_1 + \cdots + p_m \). Then
\[
d_{TV}(\text{Poisson}(\lambda), \mathcal{L}(W)) \leq 1 - \frac{e^{-\lambda}}{\lambda} \sum_{i=1}^{m} p_i^2,
\]
(cf. Barbour and Hall [4], with the understanding that the fraction on the right-hand side is one for \( \lambda = 0 \) (apply L’Hôpital’s rule for \( \lambda \downarrow 0 \)). In addition,
\[
d_{W}(\text{Poisson}(\lambda), \mathcal{L}(W)) \leq \min \left\{ 1, \frac{4}{3} \sqrt{\frac{2}{e\lambda}} \right\} \sum_{i=1}^{m} p_i^2.
\]

**Remark 3.24.** Since \( e^{-\lambda} > 0 \) and \( 1 - e^{-\lambda} \leq \lambda \), we have the upper bound
\[
1 - \frac{e^{-\lambda}}{\lambda} \leq \min \left\{ 1, \frac{1}{\lambda} \right\}, \quad \lambda > 0,
\]
which is illustrated in Figure 3.1.

**Remark 3.25.** In the Theorem 3.23, the Poisson parameter \( \lambda \) is chosen such that the expectations of \( W \) and \( N \sim \text{Poisson}(\lambda) \) agree, cf. (3.3). This corresponds to the calibration method (3.9). If \( p_1, \ldots, p_m \) are small, then the estimate (3.20) can be improved by using the calibration method of (3.7) to obtain the bound (3.32) from Exercise 3.31, see also Remark 3.32 and Table 3.2.

#### 3.4.1 Results Using an Elementary Coupling Method

In this subsection we want to prove a weaker version of (3.21), namely the estimate
\[
d_{W}(\text{Poisson}(\lambda), \mathcal{L}(W)) \leq \sum_{i=1}^{m} p_i^2.
\]
Figure 3.1: The factor $[0, \infty) \ni \lambda \mapsto (1 - e^{-\lambda})/\lambda$ in (3.20) and its upper bound $\lambda \mapsto \min\{1, 1/\lambda\}$ from (3.22). The upper line is the factor from (3.21) with a kink at $\lambda \approx 1.144$.

which by Lemma 3.18 also implies

$$d_{TV}(\text{Poisson}(\lambda), \mathcal{L}(W)) \leq \sum_{i=1}^{m} p_i^2,$$

(3.24)

which is (3.20) without the factor $(1 - e^{-\lambda})/\lambda$, see Le Cam [35]. This can be done using the so-called coupling method (cf. Lindvall [37] for a textbook presentation).

Example 3.26 (Comparison of upper bounds). To see that the difference between the estimates (3.20) and (3.24) can be substantial, consider the case $p_1 = \cdots = p_m = 1/\sqrt{m}$. Then the right-hand side of (3.24) is 1 and therefore useless (cf. Remark 3.8), while the right-hand side of (3.20) is smaller than $1/\sqrt{m}$, which is small for large $m \in \mathbb{N}$, think of $m = 10^6$, and see Table 3.1 for some specific values.

Proof of (3.23) using the coupling method. Since the estimate (3.23) concerns only the distribution of $W$, we may define this random variable in a suitable way as long as it satisfies the distributional assumption. For every $i \in \{1, \ldots, m\}$ define the sample space $\Omega_i = \{-1\} \cup \mathbb{N}_0$ and the probability measure

$$\mathbb{P}_i(\{n\}) := \begin{cases} 1 - p_i & \text{for } n = 0, \\ p_i^n e^{-p_i}/n! & \text{for } n \in \mathbb{N}, \\ e^{-p_i} - (1 - p_i) \geq 0 & \text{for } n = -1. \end{cases}$$

Define the product space $\Omega = \Omega_1 \times \cdots \times \Omega_m$ together with the product measure $\mathbb{P} = \mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_m$. In addition, for all $i \in \{1, \ldots, m\}$ and $\omega = \langle \omega_1, \ldots, \omega_m \rangle \in \Omega$,
Table 3.1: Quality of Poisson approximation. For various $m \in \mathbb{N}$ the second column gives the Kolmogorov–Smirnov distance, cf. Definition 3.9, of the binomial distribution $\text{Bin}(m, 1/\sqrt{m})$ and the Poisson distribution $\text{Poisson}(\sqrt{m})$, while the third column gives the total variation distance. The fourth column gives the upper bound (3.20) from Theorem 3.23, which is proved by the Stein–Chen method and results in $(1 - \exp(-\sqrt{m}))/\sqrt{m}$ in this example. The fifth column gives the total variation distance as a percentage of the upper bound in the fourth column. The elementary coupling bound (3.24) always gives 1 in this example and is not shown; instead the last column shows the slightly improved bound (3.32) when $\text{Poisson}(-m \log(1 - 1/\sqrt{m}))$ is used for the approximation (which is not applicable for $m = 1$). It converges to 1/2.

<table>
<thead>
<tr>
<th>$m$</th>
<th>K.-S. dist.</th>
<th>total var.</th>
<th>Bound (3.20)</th>
<th>Percentage</th>
<th>Bound (3.32)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.367879</td>
<td>0.632121</td>
<td>0.632121</td>
<td>100.00 %</td>
<td>n.a.</td>
</tr>
<tr>
<td>2</td>
<td>0.169948</td>
<td>0.327278</td>
<td>0.535197</td>
<td>61.15 %</td>
<td>1.50785</td>
</tr>
<tr>
<td>3</td>
<td>0.101422</td>
<td>0.199464</td>
<td>0.475205</td>
<td>41.97 %</td>
<td>1.1253</td>
</tr>
<tr>
<td>4</td>
<td>0.093506</td>
<td>0.144796</td>
<td>0.399416</td>
<td>36.25 %</td>
<td>0.82583</td>
</tr>
<tr>
<td>5</td>
<td>0.085456</td>
<td>0.134432</td>
<td>0.332656</td>
<td>32.73 %</td>
<td>0.76131</td>
</tr>
<tr>
<td>6</td>
<td>0.077148</td>
<td>0.1098</td>
<td>0.298242</td>
<td>30.15 %</td>
<td>0.72250</td>
</tr>
<tr>
<td>8</td>
<td>0.062349</td>
<td>0.0586</td>
<td>0.221053</td>
<td>27.33 %</td>
<td>0.64058</td>
</tr>
<tr>
<td>10</td>
<td>0.050520</td>
<td>0.0470</td>
<td>0.21053</td>
<td>25.93 %</td>
<td>0.58123</td>
</tr>
<tr>
<td>20</td>
<td>0.034107</td>
<td>0.0343</td>
<td>0.151307</td>
<td>24.68 %</td>
<td>0.51628</td>
</tr>
<tr>
<td>50</td>
<td>0.014170</td>
<td>0.0107</td>
<td>0.060421</td>
<td>23.23 %</td>
<td>0.5504</td>
</tr>
<tr>
<td>200</td>
<td>0.009583</td>
<td>0.0065</td>
<td>0.037777</td>
<td>21.44 %</td>
<td>0.5378</td>
</tr>
<tr>
<td>500</td>
<td>0.005915</td>
<td>0.0038</td>
<td>0.011105</td>
<td>20.38 %</td>
<td>0.5232</td>
</tr>
<tr>
<td>1000</td>
<td>0.004131</td>
<td>0.0026</td>
<td>0.007805</td>
<td>20.04 %</td>
<td>0.5128</td>
</tr>
<tr>
<td>2000</td>
<td>0.003880</td>
<td>0.0021</td>
<td>0.005481</td>
<td>19.50 %</td>
<td>0.5114</td>
</tr>
<tr>
<td>5000</td>
<td>0.001792</td>
<td>0.0012</td>
<td>0.003450</td>
<td>19.50 %</td>
<td>0.5076</td>
</tr>
<tr>
<td>$10^4$</td>
<td>0.001258</td>
<td>0.0009</td>
<td>0.002435</td>
<td>19.50 %</td>
<td>0.5050</td>
</tr>
</tbody>
</table>

For various $m \in \mathbb{N}$ the second column gives the Kolmogorov–Smirnov distance, cf. Definition 3.9, of the binomial distribution $\text{Bin}(m, 1/\sqrt{m})$ and the Poisson distribution $\text{Poisson}(\sqrt{m})$, while the third column gives the total variation distance. The fourth column gives the upper bound (3.20) from Theorem 3.23, which is proved by the Stein–Chen method and results in $(1 - \exp(-\sqrt{m}))/\sqrt{m}$ in this example. The fifth column gives the total variation distance as a percentage of the upper bound in the fourth column. The elementary coupling bound (3.24) always gives 1 in this example and is not shown; instead the last column shows the slightly improved bound (3.32) when $\text{Poisson}(-m \log(1 - 1/\sqrt{m}))$ is used for the approximation (which is not applicable for $m = 1$). It converges to 1/2.

Define

$$N_i(\omega) = \begin{cases} 0 & \text{if } \omega_i \in \{-1, 0\}, \\ \omega_i & \text{if } \omega_i \geq 1. \end{cases}$$

and

$$X_i(\omega) = \begin{cases} 0 & \text{if } \omega_i = 0, \\ 1 & \text{otherwise.} \end{cases}$$

With these definitions, $N_1, \ldots, N_m$ are independent and so are $X_1, \ldots, X_m$. Furthermore, $\mathbb{P}[X_i = 1] = p_i$ and $N_i \sim \text{Poisson}(p_i)$. However, note that $N_i$ and $X_i$ are coupled and strongly dependent, in particular $X_i = 0$ implies $N_i = 0$ and $N_i \geq 1$ implies $X_i = 1$. As shown in Lemma 3.2, the sum of independent Poisson distributed random variables is again Poisson distributed. Therefore

$$N := N_1 + \cdots + N_m \sim \text{Poisson}(\lambda).$$
Table 3.2: Quality of Poisson approximation as in Table 3.1, but here the binomial distribution $\text{Bin}(m, 1/m)$ is approximated by the Poisson distribution $\text{Poisson}(1)$. In this example, the elementary coupling bound \((3.24)\) always gives $1/m$ and is greater than \((3.20)\) by the factor $1/(1 - e^{-1}) \approx 1.58198$; it is not shown here. The last column shows the improved bound \((3.32)\) for $m \geq 2$, when $\text{Poisson}(−m \log(1 − 1/m))$ is used for the approximation. For $m \geq 5$ this gives a better approximation than \((3.20)\), but the expectations of the two distributions do not agree.

All together we now have the means to derive the upper estimate \((3.23)\). Using the upper bound \((3.18)\) and the triangle inequality,

\[
d_W(\mathcal{L}(N), \mathcal{L}(W)) \leq E[|N - W|] \leq \sum_{i=1}^{m} E[|N_i - X_i|]. \tag{3.25}
\]

By considering the cases $X_i = 0$ and $X_i = 1$,

\[|N_i - X_i| = N_i - X_i + 2 \cdot 1_{\{N_i = 0, X_i = 1\}}.\]

Since $E[N_i] = p_i$ and $P[N_i = 0, X_i = 1] = P_i(\{-1\}) = e^{-p_i} + p_i - 1$, it follows that

\[E[|N_i - X_i|] = 2(e^{-p_i} + p_i - 1), \quad i \in \{1, \ldots, m\}. \tag{3.26}\]

Note that the function $f: [0, \infty) \to \mathbb{R}$ with $f(x) := 2(e^{-x} + x - 1)$ satisfies
Figure 3.2: Comparison of the individual terms in (3.24), (3.28) and (3.29), namely $[0, 1] \ni p \mapsto p^2$ (upper curve), $p \mapsto 2(e^{-p} + p - 1)$ and $p \mapsto p(1 - e^{-p})$ (lower curve).

$f(0) = f'(0) = 0$, hence by applying the fundamental theorem of calculus twice,

$$f(x) = \int_0^x f'(y) \, dy = \int_0^x \int_0^y f''(z) \, dz \, dy \leq \int_0^x 2y \, dy = x^2$$

(3.27)

for all $x \in [0, \infty)$. Combining (3.25), (3.26) and applying (3.27) gives (3.23).

Remark 3.27. By omitting the application of (3.27) in the above proof, we obtain the slightly better estimate

$$d_W(\text{Poisson}(\lambda), \mathcal{L}(W)) \leq 2 \sum_{i=1}^{m} (e^{-p_i} + p_i - 1),$$

(3.28)

which by Lemma 3.18(c) implies the same estimate for $d_{TV}(\text{Poisson}(\lambda), \mathcal{L}(W))$.

An additional slight improvement, see Figure 3.2, namely

$$d_{TV}(\text{Poisson}(\lambda), \mathcal{L}(W)) \leq \sum_{i=1}^{m} p_i(1 - e^{-p_i}),$$

(3.29)

is derived below by estimating total variation distance directly. Note that for $m = 1$, estimate (3.29) agrees with (3.20).

To derive (3.29), define $A = \{ n \in \mathbb{N}_0 \mid \mathbb{P}[N = n] > \mathbb{P}[W = n] \}$. By Lemma...
Combining (3.30) and (3.31), the estimate (3.29) follows.

Exercise 3.28. Prove directly that the right-hand side of (3.29) is indeed smaller than the right-hand side of (3.28). Hint: Use the method from (3.27).

Exercise 3.29. Let $(S,d)$ be a separable metric space and let $X$ and $Y$ be two $S$-valued random variables, defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Prove that $d_{\text{TV}}(\mathcal{L}(X), \mathcal{L}(Y)) \leq \mathbb{P}[X \neq Y]$. Hint: See Exercise 3.19(a) and (3.30).

Remark 3.30. Since $\mathbb{P}[N_i = 0] > \mathbb{P}[X_i = 0]$ for every $i \in \{1, \ldots, m\}$ with $p_i > 0$ in the above coupling proofs, there is a trade-off for the large values $N_i \geq 2$, for example on $\{N_1 = 2, N_2 = 0, X_1 = X_2 = 1\}$ we have $N_1 + N_2 = X_1 + X_2$. The last estimates in (3.29) and (3.30) do not take this cancellation effect of individual approximation errors into account, hence there is room for improvement. The Stein–Chen method used below does this in an ingenious way, see Example 3.26 for a comparison.

Exercise 3.31. Let $X_1, \ldots, X_m$ be independent Bernoulli random variables with $p_i := \mathbb{P}[X_i = 1] \in (0,1)$ for all $i \in \{1, \ldots, m\}$. Define $W = X_1 + \cdots + X_m$ and $\lambda = \lambda_1 + \cdots + \lambda_m$, where $\lambda_i = -\log(1-p_i)$. Use the coupling method to prove

$$d_{\text{TV}}(\text{Poisson}(\lambda), \mathcal{L}(W)) \leq \frac{1}{2} \sum_{i=1}^{m} \lambda_i^2.$$  (3.32)

Hint: If $N_i$ has a Poisson distribution with parameter $\lambda_i$, define $X_i = 0$ if $N_i = 0$, and $X_i = 1$ otherwise. Use the steps in (3.30). You may use the estimate

$$e^x - 1 - x = \sum_{n=2}^{\infty} \frac{x^n}{n!} - \sum_{n=2}^{\infty} \frac{2}{n(n-1)} \frac{x^{n-2}}{(n-2)!} \leq \frac{1}{2} x^2 e^x, \quad x \geq 0.$$
Remark 3.32. If $p_1, \ldots, p_m$ and their sum $p_1 + \cdots + p_m$ are small, then the approximation used in Exercise 3.31 gives the (easily obtainable) upper bound for the approximation error, which can be as small as about half the size of the one in Theorem 3.23. To be specific, consider the example $p_i := 1 - e^{-1/m^2}$ for all $i \in \{1, \ldots, m\}$. Then the right-hand side of (3.32) gives $1/(2m^3)$. With $\lambda := p_1 + \cdots + p_m = m(1 - e^{-1/m^2})$ the right-hand side of (3.20) yields, for large $m \in \mathbb{N}$,

$$
\frac{1 - e^{-\lambda}}{\lambda} m(1 - e^{-1/m^2})^2 = (1 - e^{-\lambda})(1 - e^{-1/m^2}) = \frac{1}{m^3} - \frac{1}{2m^4} + O\left(\frac{1}{m^5}\right)
$$

by using the Taylor expansion of $\mathbb{R} \ni x \mapsto (1 - \exp(-1 - e^{-x^2}))(1 - e^{-x^2})$ at $x_0 = 0$, evaluated for $x = 1/m$. For another illustration, see Table 3.2.

Exercise 3.33 (Normal approximation). Using a computer and suitable software of your choice, compute similarly to Table 3.1 the Kolmogorov–Smirnov distance between the binomial distribution $\text{Bin}(m, 1/\sqrt{m})$ and the normal distribution $\mathcal{N}(\sqrt{m}, \sqrt{m} - 1)$ with expectation $\sqrt{m}$ and variance $\sqrt{m} - 1$ for various values of $m \in \mathbb{N}$. Compare with the upper bound given by the Berry–Esseen theorem. Why is the total variation distance not useful in this context?

3.4.2 Proof by the Stein–Chen Method for the Total Variation

Let $N \sim \text{Poisson}(\lambda)$ with $\lambda \geq 0$. Then, using (3.1),

$$
\lambda \mathbb{P}[N = n - 1] = \frac{\lambda^n}{(n-1)!} e^{-\lambda} = n \mathbb{P}[N = n], \quad n \in \mathbb{N}, \quad (3.33)
$$

and this recursion relation uniquely determines the Poisson distribution with parameter $\lambda$: If $N$ is $\mathbb{N}_0$-valued, then (3.33) implies by induction that

$$
\mathbb{P}[N = n] = \frac{\lambda^n}{n!} \mathbb{P}[N = 0], \quad n \in \mathbb{N}_0,
$$

and $\mathbb{P}[N = 0] = e^{-\lambda}$ gives the right starting value to obtain a probability distribution. The recursion (3.33) implies that, for every function $g: \mathbb{N}_0 \to \mathbb{R}$ which is bounded below,

$$
\lambda \mathbb{E}[g(N + 1)] = \sum_{n=1}^{\infty} \lambda g(n) \mathbb{P}[N = n - 1] = \sum_{n=1}^{\infty} ng(n) \mathbb{P}[N = n] = \mathbb{E}[Ng(N)] . \quad (3.34)
$$

Relation (3.34) applied to the functions $g_n = 1_{\{n\}}$ for $n \in \mathbb{N}$ reduces to (3.33), hence (3.34) also uniquely determines the Poisson distribution with parameter $\lambda \geq 0$. Therefore, if $\mathcal{L}(N) \neq \text{Poisson}(\lambda)$ for an $\mathbb{N}_0$-valued random variable $N$, then equality in (3.34) is violated for at least one bounded $g: \mathbb{N}_0 \to \mathbb{R}$.

The recursion relation (3.33) also shows that the Poisson distribution with parameter $\lambda \geq 0$ agrees with the Panjer(0, $\lambda$, 0) distribution, see Example 5.17 below.
Exercise 3.34 (Characterization of the Poisson distribution). Let $Z$ be a $[0, \infty)$-valued random variable satisfying $\lambda \mathbb{E}[g(Z + 1)] = \mathbb{E}[Zg(Z)]$ for all indicator functions $g$ of Borel subsets of $[0, \infty)$. Prove that $\mathcal{L}(Z) = \text{Poisson}(\lambda)$. Hint: Consider $1_{(n, n+1)}$ for $n \in \mathbb{N}_0$.

The idea of the **Stein–Chen method** is to measure the distance of a distribution on $\mathbb{N}_0$, in our case $\mathcal{L}(W)$ with $W$ as in Theorem 3.23, to the Poisson distribution with parameter $\lambda \geq 0$ by the amount

$$\lambda \mathbb{E}[g(W + 1)] - \mathbb{E}[Wg(W)] \tag{3.35}$$

of inequality in (3.34), for a specific function $g$ or a suitable collection of them.

If $\lambda = 0$, then $p_1 = \cdots = p_m = 0$, and $N = W = 0$ almost surely, hence (3.20) and (3.21) hold and we may assume $\lambda > 0$ in the following.

According to Lemma 3.18(a) the set $A := \{ n \in \mathbb{N}_0 \mid \mathbb{P}[W = n] > \mathbb{P}[N = n] \}$ satisfies

$$d_{\text{TV}}(\mathcal{L}(W), \mathcal{L}(N)) = \mathbb{P}[W \in A] - \mathbb{P}[N \in A]. \tag{3.36}$$

Since $\mathbb{P}[W = n] = 0$ for all $n > m$, it follows that $A \subset \{0, 1, \ldots, m\}$ is finite. Define $f : \mathbb{N}_0 \to [-1, 1]$ by

$$f = 1_A - \mathbb{P}[N \in A]. \tag{3.37}$$

Note that

$$\mathbb{E}[f(W)] = \mathbb{P}[W \in A] - \mathbb{P}[N \in A] \tag{3.38}$$

is the right-hand side of (3.36), for which we want to obtain an upper estimate. The next aim is to find a function $g$ to express $\mathbb{E}[f(W)]$ from (3.38) by (3.35). We do this more general, not just for the function $f$ from (3.37), because we also want to use the result for the Wasserstein metric in Subsection 3.4.3 below.

**Lemma 3.35 (Solution of the Stein equation).** Let $f : \mathbb{N}_0 \to \mathbb{R}$ be a function and $\lambda > 0$. Then the function $g : \mathbb{N}_0 \to \mathbb{R}$ given by $g(0) = 0$ (or any other value) and

$$g(l + 1) = \frac{l!}{\lambda^{l+1}} \sum_{n=0}^{l} \frac{\lambda^n}{n!} f(n), \quad l \in \mathbb{N}_0, \tag{3.39}$$

solves the so-called Stein equation for the Poisson distribution with parameter $\lambda$, i.e.

$$f(l) = \lambda g(l + 1) - lg(l), \quad l \in \mathbb{N}_0. \tag{3.40}$$

**Proof.** By direct inspection of (3.39) for $l = 0$, we get that $\lambda g(1) = f(0)$, which is (3.40) for $l = 0$. For every $l \in \mathbb{N}$,

$$\lambda g(l + 1) - lg(l) = \frac{l!}{\lambda^l} \sum_{n=0}^{l} \frac{\lambda^n}{n!} f(n) - \frac{l(l-1)!}{\lambda^l} \sum_{n=0}^{l-1} \frac{\lambda^n}{n!} f(n) = f(l). \quad \square$$

---

14Named after Charles M. Stein (1920–2016) and his former Ph.D. student Louis H. Y. Chen, Emeritus Professor at the National University of Singapore.
Exercise 3.36. In the setting of Lemma 3.35 let $N \sim \text{Poisson}(\lambda)$ and show that
\[
g(l+1) = \frac{\mathbb{E}[f(N)1_{N \leq l}]}{\lambda \mathbb{P}[N = l]}, \quad l \in \mathbb{N}_0.
\]  
(3.41)

In addition, if $f$ has a finite Lipschitz constant and $\mathbb{E}[f(N)] = 0$, prove that $g$ is bounded.

Since $W$ takes values in the finite set $\{0, \ldots, m\}$, the expectations $\mathbb{E}[g(W+1)]$, $\mathbb{E}[Wg(W)]$ and $\mathbb{E}[f(W)]$ are well defined and the Stein equation (3.40) implies that
\[
\mathbb{E}[f(W)] = \lambda \mathbb{E}[g(W+1)] - \mathbb{E}[Wg(W)].
\]  
(3.42)

We are now prepared for the main probabilistic argument of the proof, which is valid not just for the function $g$ arising from the specific $f$ given by (3.37).

Lemma 3.37. For every function $g: \mathbb{N}_0 \to \mathbb{R}$,
\[
\lambda \mathbb{E}[g(W+1)] - \mathbb{E}[Wg(W)] \leq \max_{l \in \{1, \ldots, m\}} \Delta g(l) \sum_{i=1}^{m} p_i^2
\]  
(3.43)

with forward difference $\Delta g(l) := g(l+1) - g(l)$ for all $l \in \mathbb{N}$.

Proof. Using that $\lambda = p_1 + \cdots + p_m$ and $W = X_1 + \cdots + X_m$, we obtain for the left-hand side of (3.43) that
\[
\lambda \mathbb{E}[g(W+1)] - \mathbb{E}[Wg(W)] = \sum_{i=1}^{m} (p_i \mathbb{E}[g(W+1)] - \mathbb{E}[X_i g(W)])
\]
Define $W_i = W - X_i$ for every $i \in \{1, \ldots, m\}$. By splitting $\mathbb{E}[X_i g(W)]$ into the two cases $X_i = 1$ and $X_i = 0$, noting that $X_i g(W) = 0$ for $X_i = 0$, and using the independence of $W_i$ and $X_i$, we obtain that
\[
\mathbb{E}[X_i g(W)] = \sum_{j \in \{0,1\}} \mathbb{E}[X_i g(W_i + X_i)1_{\{X_i = j\}}] = \mathbb{E}[g(W_i + 1)] p_i.
\]
Repeating this reasoning and noting that $W_i$ takes values in $\{0, \ldots, m - 1\}$,
\[
\lambda \mathbb{E}[g(W+1)] - \mathbb{E}[Wg(W)] = \sum_{i=1}^{m} p_i \mathbb{E}[g(W_i + X_i + 1) - g(W_i + 1)]
\]
\[
= \mathbb{E}[(g(W_i + X_i + 1) - g(W_i + 1))1_{\{X_i = 1\}}]
\]
\[
= \mathbb{E}[g(W_i + 2) - g(W_i + 1)] p_i \text{ by indep. of } W_i \text{ and } X_i
\]
\[
= \mathbb{E}[\Delta g(W_i + 1)] p_i
\]
\[
\leq \max_{l \in \{1, \ldots, m\}} \Delta g(l) \sum_{i=1}^{m} p_i^2.
\]  
(3.43)

Combining (3.36), (3.38), (3.42) and (3.43), we just need the result of the next lemma to obtain (3.20).
Figure 3.3: The function $N_0 \ni l \mapsto g_n(l)$ from (3.45) for $\lambda = 5$ and $n = 4$. The increments of this Stein solution are estimated by (3.46).

Lemma 3.38. For the function $f = 1_A - P[N \in A]$ defined in (3.37), the solution $g$ of the Stein equation (3.40) given by Lemma 3.35 satisfies $\Delta g(l) \leq (1 - e^{-\lambda})/\lambda$ for all $l \in \mathbb{N}$ (with equality for $A = \{1\}$ and $l = 1$).

Proof. For every $n \in N_0$ define the function

$$f_n(l) = \mathbb{1}_{\{n\}}(l) - P[N = n], \quad l \in N_0.$$  

(3.44)

Due to Lemma 3.35 and (3.41), a corresponding solution $g_n: N_0 \to \mathbb{R}$ of the Stein equation (3.40) is given by $g_n(0) = 0$ and, for every $l \in \mathbb{N}$,

$$g_n(l + 1) = \frac{E[f_n(N)\mathbb{1}_{\{N \leq l\}}]}{\lambda P[N = l]} = \frac{\mathbb{1}_{\{n,n+1,...\}}(l) - P[N \leq l]}{\lambda P[N = l]} P[N = n],$$

(3.45)

because $E[\mathbb{1}_{\{n\}}(N)\mathbb{1}_{\{N \leq l\}}] = \mathbb{1}_{\{n,n+1,...\}}(l) P[N = n]$. Since $A \subset \{0,\ldots,m\}$ is finite, $f = \sum_{n \in A} f_n$. Since the Stein equation (3.40) is linear with respect to the functions, it follows that $g := \sum_{n \in A} g_n$ is a corresponding solution for $f$ and $\Delta g = \sum_{n \in A} \Delta g_n$ with forward difference $\Delta g_n(l) := g_n(l+1) - g_n(l)$ for all $l \in N_0$. Hence it suffices to show that, see Figure 3.3,

$$\Delta g_n(l) \leq \begin{cases} 
(1 - e^{-\lambda})/\lambda & \text{for } l = n \in \mathbb{N}, \\
0 & \text{for } l \in \mathbb{N} \text{ and } n \in N_0 \text{ with } l \neq n.
\end{cases}$$

(3.46)

Using (3.45) and the recursion formula

$$\lambda P[N = l - 1] = \frac{\lambda^l}{(l-1)!} e^{-\lambda} = l P[N = l], \quad l \in \mathbb{N},$$

(3.47)
see (3.33), we see that for \( l = n \in \mathbb{N} \),

\[
g_n(n + 1) - g_n(n) = \frac{1 - \mathbb{P}[N \leq n]}{\lambda} + \frac{\mathbb{P}[N \leq n - 1]}{\lambda \mathbb{P}[N = n - 1]} \mathbb{P}[N = n] = \frac{\mathbb{P}[N \geq n + 1]}{\lambda} + \frac{\mathbb{P}[N \leq n - 1]}{n} \leq \frac{\mathbb{P}[N \geq 1]}{\lambda} = 1 - e^{-\lambda}
\]

with equality for \( l = n = 1 \). For \( l \in \mathbb{N} \) and \( n \in \mathbb{N}_0 \) with \( l < n \) we get from (3.45)

\[
g_n(l + 1) - g_n(l) = \left( -\frac{\mathbb{P}[N \leq l]}{\mathbb{P}[N = l]} + \frac{\mathbb{P}[N \leq l - 1]}{\mathbb{P}[N = l - 1]} \right) \frac{\mathbb{P}[N = n]}{\lambda}.
\]

The term in parentheses is negative, because, by the recursion formula (3.47)

\[
\mathbb{P}[N \leq l - 1] = \sum_{k=1}^{l} \frac{\lambda \mathbb{P}[N = k - 1]}{\lambda \mathbb{P}[N = l - 1]} = \sum_{k=1}^{l} k \mathbb{P}[N = k] \frac{\mathbb{P}[N = l]}{\mathbb{P}[N = l - 1]} < \frac{\mathbb{P}[N \leq l]}{\mathbb{P}[N = l]}.
\]

For \( l \in \mathbb{N} \) and \( n \in \mathbb{N}_0 \) with \( l > n \) we get from (3.45)

\[
g_n(l + 1) - g_n(l) = \left( -\frac{\mathbb{P}[N \geq l + 1]}{\mathbb{P}[N = l]} + \frac{\mathbb{P}[N \geq l]}{\mathbb{P}[N = l - 1]} \right) \frac{\mathbb{P}[N = n]}{\lambda}.
\]

Again, the term in parentheses is negative, because, using (3.47),

\[
\frac{\mathbb{P}[N \geq l]}{\mathbb{P}[N = l - 1]} = \sum_{k=l+1}^{\infty} \frac{\lambda \mathbb{P}[N = k - 1]}{\lambda \mathbb{P}[N = l - 1]} = \sum_{k=l+1}^{\infty} k \mathbb{P}[N = k] \frac{\mathbb{P}[N = l]}{\mathbb{P}[N = l - 1]} > \frac{\mathbb{P}[N \geq l + 1]}{\mathbb{P}[N = l]}.
\]

Therefore, the estimate (3.46) for \( \Delta g_n \) is proved. \( \square \)

### 3.4.3 Proof by the Stein–Chen Method for the Wasserstein Metric

To prove the Poisson approximation for \( W \) in the Wasserstein metric, i.e. (3.21), we can follow the strategy used in the previous subsection. Let \( N \sim \text{Poisson}(\lambda) \) and let \( f: \mathbb{N}_0 \to \mathbb{R} \) have Lipschitz constant at most 1. By subtracting the constant \( \mathbb{E}[f(N)] \) from \( f \) if necessary, we may assume that \( \mathbb{E}[f(N)] = 0 \). By Lemma 3.35 the corresponding solution \( g \) of the Stein equation is given by (3.39), and Lemma 3.37 applies to \( g \). In view of the definition of the Wasserstein metric in (3.14), all we need for (3.21) is the following lemma.

**Lemma 3.39.** Let \( f: \mathbb{N}_0 \to \mathbb{R} \) have Lipschitz constant at most 1 and satisfy \( \mathbb{E}[f(N)] = 0 \). Then the corresponding solution \( g \) of the Stein equation for the Poisson distribution with parameter \( \lambda > 0 \) satisfies

\[
\Delta g \leq \min \left\{ 1, \frac{4}{3} \sqrt{\frac{2}{e \lambda}} \right\}.
\]
Proof. See [5, Remark 1.1.6] or, for a more explicit presentation, [6, Eq. (1.4) in Theorem 1.1]. Note that, according to Exercise 3.36, the solution is bounded. 

For more details and further applications of the Stein–Chen method, see e.g. the textbook by Barbour, Holst and Janson [5] or the lecture notes by Eichelsbacher [15]. For the application of Stein’s method for the normal approximation, see e.g. the textbook by Chen, Goldstein and Shao [10].

3.5 Multivariate Poisson Distribution

The multivariate generalization of the Poisson distribution is motivated by common Poisson shock models [36]; with different notation it is also given in [51, Chapter 20.1]. It will easily allow us to model joint defaults of obligors.

Definition 3.40 (Multivariate Poisson distribution). Let \( m \in \mathbb{N} \), consider a collection \( G \subset \mathcal{P}(\{1, \ldots, m\}) \) of subsets of \( \{1, \ldots, m\} \) with \( \emptyset \not\in G \), and Poisson parameters \( \lambda = (\lambda_g)_{g \in G} \in [0, \infty)^G \). Let \( (N_g)_{g \in G} \) be independent random variables with \( N_g \sim \text{Poisson}(\lambda_g) \) for every \( g \in G \). Then the distribution of the \( \mathbb{N}_0^m \)-valued random vector

\[
N = \sum_{g \in G} c_g N_g, \tag{3.48}
\]

where the vector \( c_g = (c_{g,1}, \ldots, c_{g,m})^\top \in \{0,1\}^m \) is given by

\[
c_{g,i} = 1_g(i) = \begin{cases} 1 & \text{if } i \in g, \\ 0 & \text{if } i \not\in g, \end{cases} \tag{3.49}
\]

is called the \( m \)-variate Poisson distribution \( \text{MPoisson}(G, \lambda, m) \) on \( \mathbb{N}_0^m \).

In the credit risk interpretation, the obligors in the group \( g \subset \{1, \ldots, m\} \) default together with Poisson intensity \( \lambda_g \), independent of the other groups in \( G \). An empty group of obligors cannot cause any default, for this reason we excluded \( \emptyset \) from \( G \). For practical applications we should assume that \( \{1, \ldots, m\} \subset \bigcup_{g \in G} g \), because otherwise there would exist obligors who can never default. If \( G = \emptyset \), then (3.48) is an empty sum and \( \text{MPoisson}(G, \lambda, m) \) is interpreted as the degenerate distribution concentrated at the origin \( 0 \in \mathbb{N}_0^m \). If \( m = 1 \) and \( G = \{g\} \) with \( g = \{1\} \), then \( \text{MPoisson}(G, \lambda, m) \) coincides with \( \text{Poisson}(\lambda_g) \). It might be tempting to choose \( G = \mathcal{P}(\{1, \ldots, m\}) \setminus \{\emptyset\} \) for greatest generality, but then there are \( 2^m - 1 \) Poisson parameters \( \lambda_g \) for \( g \in G \), which already for \( m = 1000 \) obligors are far too many to yield a practically useful model.

The next result is the multivariate generalization of Lemma 3.2.

---

15 We consider \([0, \infty)^G\) as the set of all functions \( \lambda : G \to [0, \infty) \), where the image of \( g \in G \) is denoted by \( \lambda_g \), hence \( \lambda \) can be represented by the “tuple” \( (\lambda_g)_{g \in G} \). With this interpretation, the \( d \)-fold Cartesian products \( \mathbb{R}^d \) and \( \mathbb{N}_0^d \) are short-hand versions of \( \mathbb{R}^{\{1, \ldots, d\}} \) and \( \mathbb{N}_0^{\{1, \ldots, d\}} \).

16 The vector \( c_g \) points to a corner of the \( m \)-dimensional hypercube.
Lemma 3.41 (Summation property of the multivariate Poisson distribution). If \( N_1, \ldots, N_k \) are independent with \( N_i \sim \text{MPoisson}(G_i, \lambda(i), m) \) for all \( i \in \{1, \ldots, k\} \) with \( \lambda(i) = (\lambda_g(i))_{g \in G_i} \), according to Definition 3.40, then

\[
N := \sum_{i=1}^{k} N_i \sim \text{MPoisson}(G, \lambda, m),
\]

where \( G = \bigcup_{i=1}^{k} G_i \) and \( \lambda = (\lambda_g)_{g \in G} \) is given by

\[
\lambda_g = \sum_{i=1}^{k} \lambda^{(i)}_g, \quad g \in G.
\]

Exercise 3.42. Use Lemma 3.2 and Definition 3.40 to prove Lemma 3.41.

Remark 3.43 (Infinite divisibility of the multivariate Poisson distribution). Lemma 3.41 implies that the multivariate Poisson distribution \( \text{MPoisson}(G, \lambda, m) \) with \( \lambda = (\lambda_g)_{g \in G} \) is infinitely divisible, because for every \( k \in \mathbb{N} \) the distribution of \( N_1 + \cdots + N_k \) is \( \text{MPoisson}(G, \lambda, m) \), when \( N_1, \ldots, N_k \) are independent with \( N_i \sim \text{MPoisson}(G, \lambda^{(k)}_i, m) \) for every \( i \in \{1, \ldots, k\} \), where \( \lambda^{(k)}_i = (\lambda_g(i)/k)_{g \in G} \).

Lemma 3.44 (Moments of the multivariate Poisson distribution). Assume that \( N = (N_1, \ldots, N_m)^\top \sim \text{MPoisson}(G, \lambda, m) \). Then, with the notation from Definition 3.40, for all \( i, j \in \{1, \ldots, m\} \),

\[
E[N_i] = \sum_{g \in G} \lambda_g \quad \text{(3.50)}
\]

and for the components of the covariance matrix of \( N, \)

\[
\text{Cov}(N_i, N_j) = \sum_{g \in G} \lambda_g. \quad \text{(3.51)}
\]

Proof. Equation (3.50) follows from (3.48), (3.49) and (3.3). Similarly, using the bi-linearity of the covariance and the independence of \((N_g)_{g \in G}, \)

\[
\sum_{g \in G} \sum_{g' \in G} \text{Cov}(N_g, N_{g'}) = \sum_{g \in G} \text{Var}(N_g) = 0 \text{ if } g \neq g'.
\]

Using (3.4), the result (3.51) follows.

Remark 3.45. Note that by (3.51) the components of a multivariate Poisson distribution can only have a non-negative covariance.

Lemma 3.46 (Multivariate Poisson distribution with independent components). Assume that \( N = (N_1, \ldots, N_m)^\top \sim \text{MPoisson}(G, \lambda, m) \) and \( m \geq 2 \). Then, with the notation from Definition 3.40, the following properties are equivalent:

37
(a) The components \(N_1, \ldots, N_m\) are independent.

(b) \(\text{Cov}(N_i, N_j) = 0\) for all \(i, j \in \{1, \ldots, m\}\) with \(i \neq j\).

(c) \(\lambda_g = 0\) for all \(g \in G\) with \(|g| \geq 2\).

Proof. Note that (a) implies (b), which in turn implies (c) via (3.51). If (c) holds, then \(N_g \overset{\text{a.s.}}{=} 0\) for all \(g \in G\) with \(|g| \geq 2\), hence

\[
\begin{pmatrix}
N_1 \\
\vdots \\
N_m
\end{pmatrix} \overset{\text{a.s.}}{=} \sum_{\{i\} \in G} c_{\{i\}} N_{\{i\}}
\]

by (3.48), hence \(N_i \overset{\text{a.s.}}{=} N_{\{i\}}\) if \(\{i\} \in G\) and \(N_i \overset{\text{a.s.}}{=} 0\) otherwise. Since \((N_g)_{g \in G, |g| = 1}\) are independent by Definition 3.40, part (a) follows.

3.6 General Multivariate Poisson Mixture Model

Following the mixture approach outlined in Section 2.2 for Bernoulli default indicators, this section generalizes the multivariate Poisson distribution discussed in the previous section by introducing random Poisson intensities \((\Lambda_g)_{g \in G}\) for all the groups of obligors defaulting together.

Formally, \((\Lambda_g)_{g \in G}\) is a collection of \([0, \infty)\)-valued random variables, which may even be dependent. Similar assumptions as in Section 2.2.1 are made for the intensities, namely

\[
P[N_g = n_g | (\Lambda_h)_{h \in G}] \overset{\text{a.s.}}{=} P[N_g = n_g | \Lambda_g] \overset{\text{a.s.}}{=} e^{-\Lambda_g} \frac{\Lambda_g^{n_g}}{n_g!}
\]

for every \(g \in G\) and \(n_g \in \mathbb{N}_0\), cf. (2.10), and the conditional independence of \((N_g)_{g \in G}\) given \((\Lambda_g)_{g \in G}\), i.e., for all \(n_g \in \mathbb{N}_0\) for \(g \in G\),

\[
P[N_g = n_g \text{ for all } g \in G | (\Lambda_h)_{h \in G}] \overset{\text{a.s.}}{=} \prod_{g \in G} P[N_g = n_g | (\Lambda_h)_{h \in G}]
\]

by (3.52), cf. (2.11). The unconditional joint distribution of \((N_g)_{g \in G}\) can be obtained by integrating over the random intensities, i.e.

\[
P[N_g = n_g \text{ for all } g \in G] = E \left[ \prod_{g \in G} \frac{\Lambda_g^{n_g}}{n_g!} e^{-\Lambda_g} \right].
\]

Exercise 3.47 (Explicit construction of the general multivariate Poisson mixture model). Consider a \([0, \infty)^G\)-valued random vector \(\Lambda' = (\Lambda'_g)_{g \in G}\) on a probability space \((\Omega', \mathcal{A}', \mathbb{P}')\). Define \(\Omega = \Omega' \times \mathbb{N}_0^G\) and \(\mathcal{A} = \mathcal{A}' \otimes \mathcal{P}(\mathbb{N}_0^G)\).
(a) Show that \( K : [0, \infty)^G \times \mathcal{P}(\mathbb{N}_0^G) \to [0, 1] \) with
\[
K(\lambda, B) := \sum_{(n_g)_{g \in G} \in B} \prod_{g \in G} \frac{\lambda_g^{n_g}}{n_g!} e^{-\lambda_g}
\]
for all \( \lambda = (\lambda_g)_{g \in G} \in [0, \infty)^G \) and \( B \subset \mathbb{N}_0^G \) is a well-defined stochastic transition kernel. Hint: (3.55) can be expressed as \( K(\lambda, \cdot) = \bigotimes_{g \in G} \text{Poisson}(\lambda_g) \).

(b) Show that a well-defined probability measure \( \mathbb{P} \) on the product space \((\Omega, \mathcal{A})\) is uniquely determined by
\[
\mathbb{P}[A \times B] = \mathbb{E}_{\mathbb{P}'}[1_A K(\Lambda', B)], \quad A \in \mathcal{A}', \; B \subset \mathbb{N}_0^G.
\]
Hint: Consider \( \mathbb{P}' \otimes \nu \) on \((\Omega, \mathcal{A})\), where \( \nu \) is the counting measure on \( \mathbb{N}_0^G \), and consider the product in (3.55) as probability density. Alternatively, apply [32, Corollary 14.23].

(c) For every \( g \in G \) define \( \Lambda_g(\omega) = \Lambda'_g(\omega') \) and \( N_g(\omega) = n_g \) for all \( \omega = (\omega', (n_h)_{h \in G}) \in \Omega \). Prove that (3.52) and (3.53) are satisfied. Hint: Use (3.56) and the hint for (a).

3.6.1 Expected Values, Variances, and Individual Covariances

Again, the expected number of defaults can be deduced from the properties of the underlying random intensities \((\Lambda_g)_{g \in G}\). From (3.3), (3.4) and (3.52) we obtain that \( \mathbb{E}[N_g | \Lambda_g] \overset{\text{a.s.}}{=} \Lambda_g \) and \( \text{Var}(N_g | \Lambda_g) \overset{\text{a.s.}}{=} \Lambda_g \) for every \( g \in G \). For the numbers \( N_1, \ldots, N_m \) of default events of the individual obligors \( 1, \ldots, m \), we have the representation
\[
\begin{pmatrix}
N_1 \\
\vdots \\
N_m
\end{pmatrix} = \sum_{g \in G} c_g N_g
\]
from (3.48), hence
\[
\begin{pmatrix}
\mathbb{E}[N_1] \\
\vdots \\
\mathbb{E}[N_m]
\end{pmatrix} = \sum_{g \in G} c_g \mathbb{E}[\mathbb{E}[N_g | \Lambda_g]] = \sum_{g \in G} c_g \mathbb{E}[\Lambda_g],
\]
or, written out componentwise,
\[
\mathbb{E}[N_i] = \sum_{g \in G} c_g \mathbb{E}[\Lambda_g], \quad i \in \{1, \ldots, m\}.
\]
Note that the sum of all ones in the vector \( c_g \) gives the number \(|g|\) of obligors defaulting together when the group \( g \) defaults. Hence, using (3.57),
\[
N := N_1 + \cdots + N_m = \sum_{i=1}^m \sum_{g \in G} c_{g,i} N_g = \sum_{g \in G} |g| N_g
\]
is the random variable representing the overall number of default events in the credit portfolio. Similarly, using (3.58),

$$\mathbb{E}[N] = \sum_{i=1}^{m} \mathbb{E}[N_i] = \sum_{g \in G} |g| \mathbb{E}[\Lambda_g].$$

To calculate the variances and covariances of $N_1, \ldots, N_m$, we start with a general formula, which is helpful in particular for mixture models. We will apply (3.61) with $X = N_g$ and the sub-$\sigma$-algebra $\mathcal{B} = \sigma(\Lambda_g)$ containing all the information about $\Lambda_g$.

**Lemma 3.48.** Let $X$ and $Y$ be square-integrable $\mathbb{R}^c$- and $\mathbb{R}^d$-valued random vectors, respectively, on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and $\mathcal{B} \subset \mathcal{A}$ a sub-$\sigma$-algebra.

The covariance matrix of size $c \times d$ satisfies

$$\text{Cov}(X, Y) = \mathbb{E}[\text{Cov}(X, Y|\mathcal{B})] + \text{Cov}(\mathbb{E}[X|\mathcal{B}], \mathbb{E}[Y|\mathcal{B}]),$$

where expectations are taken componentwise. If $c = d = 1$ and $X = Y$, then (3.60) reduces to

$$\text{Var}(X) = \mathbb{E}[\text{Var}(X|\mathcal{B})] + \text{Var}(\mathbb{E}[X|\mathcal{B}]).$$

**Proof.** The formula for the variance follows from the one for the covariance matrix. It therefore suffices to prove (3.60). We view $X$ and $Y$ as column vectors. Using the definition of the covariance matrix, adding and subtracting conditional expectations, we get that

$$\text{Cov}(X, Y) = \mathbb{E}\left[ (X - \mathbb{E}[X])(Y - \mathbb{E}[Y])^\top \right]
= \mathbb{E}\left[ ((X - \mathbb{E}[X|\mathcal{B}]) + (\mathbb{E}[X|\mathcal{B}] - \mathbb{E}[X]))(Y - \mathbb{E}[Y|\mathcal{B}]) + (E[Y|\mathcal{B}] - \mathbb{E}[Y])^\top \right].$$

Expanding the product, inserting conditional expectations given $\mathcal{B}$ in the first three terms and using properties of conditional expectation,

$$\text{Cov}(X, Y) = \mathbb{E}\left[ (X - \mathbb{E}[X|\mathcal{B}]) (Y - \mathbb{E}[Y|\mathcal{B}])^\top | \mathcal{B} \right] \quad \overset{\Delta}{=} \quad \text{Cov}(X|\mathcal{B})$$
$$+ \mathbb{E}\left[ (\mathbb{E}[X|\mathcal{B}] - \mathbb{E}[X]) (Y - \mathbb{E}[Y|\mathcal{B}])^\top | \mathcal{B} \right] \quad \overset{\Delta}{=} \quad 0$$
$$+ \mathbb{E}\left[ (X - \mathbb{E}[X|\mathcal{B}]) (\mathbb{E}[Y|\mathcal{B}] - \mathbb{E}[Y])^\top | \mathcal{B} \right] \quad \overset{\Delta}{=} \quad 0$$
$$+ \mathbb{E}\left[ (\mathbb{E}[X|\mathcal{B}] - \mathbb{E}[X]) (\mathbb{E}[Y|\mathcal{B}] - \mathbb{E}[Y])^\top \right].$$

$\square$

40
Corollary 3.49. Let $A$, $B$ be random matrices and $X$, $Y$ random vectors of compatible sizes such that $AX$ and $BY$ are well-defined. Assume that $AX$, $BY$, $X$ and $Y$ are square-integrable. If $(A,B)$ and $(X,Y)$ are independent, then

$$\text{Cov}(AX, BY) = \mathbb{E}[A \text{Cov}(X, Y) B^\top] + \text{Cov}(\mathbb{E}[X], B \mathbb{E}[Y]).$$

Proof. We apply (3.60) from Lemma 3.48 with $B = \sigma(A,B)$. Since $A$ and $B$ are $\mathcal{B}$-measurable, $\mathbb{E}[AX | \mathcal{B}] \stackrel{a.s.}{=} A \mathbb{E}[X | \mathcal{B}]$ and $\mathbb{E}[BY | \mathcal{B}] \stackrel{a.s.}{=} B \mathbb{E}[Y | \mathcal{B}]$ as well as

$$\text{Cov}(AX, BY | \mathcal{B}) \stackrel{a.s.}{=} \mathbb{E}[(AX - \mathbb{E}[AX | \mathcal{B}](BY - \mathbb{E}[BY | \mathcal{B}])^\top | \mathcal{B}]$$

$$\stackrel{a.s.}{=} A \mathbb{E}[(X - \mathbb{E}[X | \mathcal{B}])(Y - \mathbb{E}[Y | \mathcal{B}])^\top | \mathcal{B}] B^\top$$

Due to the assumed independence, it follows that $\mathbb{E}[X | \mathcal{B}] \stackrel{a.s.}{=} \mathbb{E}[X]$ and $\mathbb{E}[Y | \mathcal{B}] \stackrel{a.s.}{=} \mathbb{E}[Y]$ as well as $\text{Cov}(X, Y | \mathcal{B}) \stackrel{a.s.}{=} \text{Cov}(X, Y)$. $\square$

Now we are in a position to calculate the variances and covariances of $N_1, \ldots, N_m$ as well as the variance of $N = N_1 + \cdots + N_m$, provided that these default numbers have a finite expectation. Using (3.61) from Lemma 3.48 and (3.52),

$$\text{Var}(N_g) = \mathbb{E} \left[ \text{Var}(N_g | \Lambda_g) \right] + \text{Var} \left( \mathbb{E}[N_g | \Lambda_g] \right) = \mathbb{E}[\Lambda_g] + \text{Var}(\Lambda_g) \quad (3.62)$$

for every $g \in G$. By the conditional independence of $N_g$ and $N_h$, see (3.53), and (3.52),

$$\mathbb{E}[N_g N_h] = \mathbb{E} \left[ \mathbb{E}[N_g N_h | (\Lambda_{g'})_{g' \in G}] \right] = \mathbb{E} \left[ \mathbb{E}[N_g | \Lambda_g] \mathbb{E}[N_h | \Lambda_h] \right] = \mathbb{E}[\Lambda_g \Lambda_h]$$

for all $g, h \in G$ with $g \neq h$, hence

$$\text{Cov}(N_g, N_h) = \mathbb{E}[N_g N_h] - \mathbb{E}[N_g] \mathbb{E}[N_h]$$

$$= \mathbb{E}[\Lambda_g \Lambda_h] - \mathbb{E}[\Lambda_g] \mathbb{E}[\Lambda_h]$$

$$= \text{Cov}(\Lambda_g, \Lambda_h). \quad (3.63)$$

Using the representation (3.57), in particular $N_i = \sum_{g \in G,i \in g} N_g$ and $N_j = \sum_{h \in G,j \in h} N_h$, it follows with the linearity of covariance in both arguments that

$$\text{Cov}(N_i, N_j) = \sum_{g,h \in G, i \in g,j \in h} \text{Cov}(N_g, N_h)$$

$$= \sum_{g \in G} \text{Var}(N_g) + \sum_{g,h \in G, g \neq h \atop i \in g,j \in h} \text{Cov}(N_g, N_h)$$

for all obligors $i,j \in \{1, \ldots, m\}$. By inserting (3.62) and (3.63),

$$\text{Cov}(N_i, N_j) = \sum_{g \in G \atop i,j \in g} \left( \mathbb{E}[\Lambda_g] + \text{Var}(\Lambda_g) \right) + \sum_{g,h \in G, g \neq h \atop i \in g,j \in h} \text{Cov}(\Lambda_g, \Lambda_h).$$
For the case \( i = j \), we obtain that
\[
\text{Var}(N_i) = \sum_{g \in G} \left( \mathbb{E}[\Lambda_g] + \text{Var}(\Lambda_g) + \sum_{h \in G \setminus \{g\}} \text{Cov}(\Lambda_g, \Lambda_h) \right), \quad i \in \{1, \ldots, m\}.
\]

Using the representation (3.59) and formula (2.18), it follows for the total number of defaults in the portfolio that
\[
\text{Var}(N) = \sum_{g \in G} |g|^2 \text{Var}(N_g) + \sum_{g,h \in G \setminus \{g\}} |h| \text{Cov}(N_g, N_h);
\]
rearranging and inserting (3.62) and (3.63), it follows that
\[
\text{Var}(N) = \sum_{g \in G} |g| \left( |g| (\mathbb{E}[\Lambda_g] + \text{Var}(\Lambda_g)) + \sum_{h \in G \setminus \{g\}} |h| \text{Cov}(\Lambda_g, \Lambda_h) \right).
\]

**Exercise 3.50.** Rederive (2.21) using (3.60) and the conditional independence formulated in (2.11).

### 3.6.2 One-Factor Poisson Mixture Model

As a special case of the general multivariate Poisson mixture model, assume that \( G = \{\{1\}, \ldots, \{m\}\} \), that there exists a single \([0, \infty)\)-valued random variable \( \Lambda \), let \( F \) denote its distribution function, and assume that there are parameters \( \mu_1, \ldots, \mu_m \geq 0 \) such that \( \Lambda_{\{i\}} = \mu_i \Lambda \) for all \( i \in \{1, \ldots, m\} \). Then \( N_i = N_{\{i\}} \) by (3.57) for all \( i \in \{1, \ldots, m\} \) and (3.54) simplifies, i.e., for all \( n_1, \ldots, n_m \in \mathbb{N}_0 \),
\[
\mathbb{P}[N_1 = n_1, \ldots, N_m = n_m] = \left( \prod_{i=1}^m \frac{\mu_i^{n_i}}{n_i!} \right) \mathbb{E}[\Lambda^{n_1+\cdots+n_m}] e^{-\mu \Lambda} \tag{3.64}
\]
with \( \mu := \mu_1 + \cdots + \mu_m \).

Since \( N_1, \ldots, N_m \) are conditionally independent given \( \Lambda \), the summation property of the Poisson distribution, see Lemma 3.2, implies that the conditional distribution of the sum \( N = N_1 + \cdots + N_m \) given \( \Lambda \) is almost surely Poisson(\( \mu \Lambda \)). Hence, for all \( n \in \mathbb{N}_0 \),
\[
\mathbb{P}[N = n] = \int_0^\infty \mathbb{P}[N = n | \Lambda = \lambda] F(d\lambda) = \int_0^\infty \frac{(\mu \lambda)^n}{n!} e^{-\mu \lambda} F(d\lambda), \tag{3.65}
\]

### 3.6.3 Uniform Poisson Mixture Model

To model a uniform portfolio, we may consider the one-factor Poisson mixture model of Subsection 3.6.2 with \( \mu_1 = \cdots = \mu_m = 1 \), hence \( \mu = m \). Then (3.64) simplifies, i.e., for all \( n_1, \ldots, n_m \in \mathbb{N}_0 \),
\[
\mathbb{P}[N_1 = n_1, \ldots, N_m = n_m] = \int_0^\infty \frac{\lambda^{n_1+\cdots+n_m}}{n_1! \cdots n_m!} e^{-\mu \lambda} F(d\lambda),
\]
and (3.65) holds with \( \mu = m \).
4 Generating Functions, Mixed and Compound Distributions

4.1 Probability-Generating Functions

Probability-generating functions are a powerful tool when working with \( N_0 \)-valued or, more generally, \( N_0^d \)-valued random variables. Especially, as will be shown, a probability-generating function uniquely determines a probability distribution on \( N_0^d \) and vice versa.

Usually, the distribution of the sum of two independent random variables is expressed as convolution of their distributions. In the context of probability-generating functions, it is simply the distribution uniquely determined as the product of the two probability-generating functions, see (4.27) below. In the following we will use some multi-index notation, which we will introduce when convenient.

**Definition 4.1.** For a random vector \( X = (X_1, \ldots, X_d) : \Omega \rightarrow N_0^d \) define the probability-generating function\(^\text{17}\) of its distribution by

\[
\varphi_X(s) := E \left[ \prod_{i=1}^{d} s^{X_i} \right] = \sum_{n=(n_1, \ldots, n_d) \in N_0^d} \left( \prod_{i=1}^{d} s^{|n_i|} \right) \mathbb{P}[X = n], \tag{4.1}
\]

where the series is absolutely convergent at least for all \( s = (s_1, \ldots, s_d) \in \mathbb{C}^d \) with \( \|s\|_\infty := \max\{ |s_1|, \ldots, |s_d| \} \leq 1 \), so the generating function is defined at least on the \( d \)-fold Cartesian product of the closed unit disk of \( \mathbb{C} \). The probability-generating function actually belongs to the probability distribution \( \mathcal{L}(X) \) and not to the random vector \( X \) itself, but we will avoid the more clumsy notation \( \varphi_{\mathcal{L}(X)} \).

**Example 4.1 (Bernoulli distribution).** Let the random variable \( B \) take values in \( \{0, 1\} \), where \( p := \mathbb{P}[B = 1] \). Then \( B \) is said to have a Bernoulli distribution with success probability \( p \in [0, 1] \). Considering this distribution as a special case of the binomial distribution, we write \( B \sim \text{Bin}(1, p) \). Its probability-generating function is given by

\[
\varphi_B(s) := \mathbb{E}[s^B] = \mathbb{P}[B = 0] + \mathbb{P}[B = 1] s = (1 - p) + ps = 1 + p(s - 1), \quad s \in \mathbb{C}. \tag{4.2}
\]

**Example 4.3 (Poisson distribution).** For a random variable \( N \sim \text{Poisson}(\lambda) \) with parameter \( \lambda \geq 0 \), the probability-generating function is given by

\[
\varphi_N(s) := \mathbb{E}[s^N] = \sum_{n=0}^{\infty} s^n \frac{\lambda^n}{n!} e^{-\lambda} = e^{\lambda s} e^{-\lambda} = e^{\lambda(s-1)}, \quad s \in \mathbb{C}. \tag{4.3}
\]

\(^{17}\)The factorial moment generating function \( s \mapsto \mathbb{E}[s^X] \), defined at least for all \( s = (s_1, \ldots, s_d) \in \mathbb{C}^d \) with \( |s_i| = 1 \) for all \( i \in \{1, \ldots, d\} \), extends the notion of the probability-generating function to \( \mathbb{R}^d \)-valued random variables, but we will not need this extension. However, we will use the moment-generating property of the probability-generating function, see (4.21) below.
Example 4.4 (Univariate logarithmic distribution). Consider an \( \mathbb{N} \)-valued random variable \( N \) with univariate logarithmic distribution \( \log(\rho) \) with parameter \( \rho \in [0, 1) \), i.e.,

\[
\mathbb{P}[N = n] = \frac{\rho^{n-1}}{c(\rho) n}, \quad n \in \mathbb{N},
\]

with normalising factor\(^{18}\)

\[
c(\rho) := \sum_{n \in \mathbb{N}} \frac{\rho^{n-1}}{n} = \begin{cases} \frac{-\log(1-\rho)}{p} & \text{if } p \in (0, 1), \\ 1 & \text{if } p = 0, \end{cases}
\]

see the Taylor series \(^{38}\). Using this Taylor series again, we see that

\[
\varphi_N(s) \overset{[4.1]}{=} \frac{s}{c(\rho)} \sum_{n \in \mathbb{N}} \frac{(ps)^{n-1}}{n} = c(p) = \begin{cases} \frac{\log(1-p)}{\log(1-p)} & \text{if } p \in (0, 1), \\ s & \text{if } p = 0, \end{cases}
\]

defined for all \( s \in \mathbb{C} \) with \( p|s| < 1 \), is the probability-generating function of \( N \). If \( p \) is small, then the calculation of \( \log(1 - p) \) leads to the cancellation of significant digits. Therefore, if for example \( p \leq 0.1 \) and an \( l \)-digit precision is desired, then it is numerically more stable to add the first \( l \) terms of the power series in \( [4.5] \) defining \( c(p) \) than to use the formula of the right-hand side of \( [4.5] \). The same advice applies to \( [4.6] \) when \( p|s| \) is small. For more information about the univariate logarithmic distribution see \([30, \text{Chapter 7}]\), for the multivariate version see Definition 4.36 below.

Example 4.5 (Multivariate Bernoulli distribution). For \( d \in \mathbb{N} \) consider a random vector \( B = (B_1, \ldots, B_d) \) with a multivariate Bernoulli distribution with parameter vector \( \rho = (\rho_1, \ldots, \rho_d) \in [0, 1]^d \) satisfying \( \rho_1 + \cdots + \rho_d = 1 \), i.e.

\[
\mathbb{P}[B = e_i] = \rho_i, \quad i \in \{1, \ldots, d\},
\]

where \( e_i = (0, \ldots, 0, 1, 0, \ldots, 0) \in \{0, 1\}^d \) denotes the \( i \)th unit vector with the digit 1 at position \( i \). It is also called categorical distribution on the finite set \( \{e_1, \ldots, e_d\} \). In analogy to Example 4.2 we will consider this distribution as a special case of the multinomial distribution, see Example 4.16 below, and write \( B \sim \text{Multinomial}(\rho) \). Its probability-generating function is given by

\[
\varphi_B(s) \overset{[4.1]}{=} \sum_{i=1}^{d} s_i \mathbb{P}[B = e_i] = \sum_{i=1}^{d} \rho_i s_i, \quad s = (s_1, \ldots, s_d) \in \mathbb{C}^d.
\]

Note that \( B_i \) is \{0, 1\}-valued and \( \mathbb{P}[B_i = 1] = \rho_i \), hence \( B_i \sim \text{Bin}(1, \rho_i) \) for every component \( i \in \{1, \ldots, d\} \) of \( B = (B_1, \ldots, B_d) \), in particular \( \mathbb{E}[B] = \rho \) and \( \text{Var}(B_i) = \rho_i(1 - \rho_i) \). Since \( \|e_i\|_1 = 1 \) for every \( i \in \{1, \ldots, d\} \), it follows that

\[
\|B\|_1 = \sum_{i=1}^{d} B_i \equiv 1.
\]

\(^{18}\) The function \( c \) is the hypergeometric function \( _2F_1(1, 1; 2; \cdot) \) and also the derivative of the dilogarithm \( \text{Li}_2 \).
The multivariate Bernoulli distribution has the aggregation property

\[ (B_1, \ldots, B_i, B_{i+1} + \cdots + B_d) \sim \text{Multinomial}(1, p_1, \ldots, p_i, p_{i+1} + \cdots + p_d) \] (4.9)

for every \( i \in \{1, \ldots, d-1\} \), and the permutation property

\[ (B_{\sigma(1)}, \ldots, B_{\sigma(d)}) \sim \text{Multinomial}(1, p_{\sigma(1)}, \ldots, p_{\sigma(d)}) \] (4.10)

for every permutation \( \sigma \) of \( \{1, \ldots, d\} \). Properties (4.8), (4.9) and (4.10) will imply corresponding properties for compound distributions involving the multivariate Bernoulli distribution, see Exercises 4.17, 4.37 and 4.42 below. If \( d \geq 2 \), then exactly one of the components of \( B \) attains the value 1, all others are zero, hence for all \( i, j \in \{1, \ldots, d\} \) with \( i \neq j \),

\[ \text{Cov}(B_i, B_j) = \mathbb{E}[B_i B_j] - \mathbb{E}[B_i] \mathbb{E}[B_j] = -p_i p_j, \] (4.11)

which implies dependence unless \( p_i = 0 \) or \( p_j = 0 \). For the generalizations of the properties (4.8), (4.9), (4.10) and (4.11) to the general multinomial distribution, see Exercise 4.17 below.

### 4.1.2 Basic Properties and Calculation of Moments

Some of the basic properties of probability-generating functions of the distributions of \( \mathbb{N}_0^d \)-valued random vectors \( X = (X_1, \ldots, X_d) \) following directly from (4.1) are

\[ \varphi_X(0, \ldots, 0) = \mathbb{P}[X = 0], \] (4.12)

\[ \varphi_X(1, \ldots, 1) = \sum_{n \in \mathbb{N}_0^d} \mathbb{P}[X = n] = 1 \] (4.13)

and

\[ \varphi_X^{(n)}(0, \ldots, 0) = n_1! \cdots n_d! \mathbb{P}[X = n], \quad n = (n_1, \ldots, n_d) \in \mathbb{N}_0^d, \] (4.14)

where \( \varphi_X^{(n)} = \varphi_X^{(n_1 \ldots n_d)} \) means \( n_i \) partial derivatives with respect to the \( i \)-th variable iteratively for all \( i \in \{1, \ldots, d\} \). Because of (4.1) and (4.14), \( \varphi_X \) uniquely determines the distribution of \( X \) and vice versa. It follows from the power series representation in (4.1) that \( [0,1]^d \ni (s_1, \ldots, s_d) \mapsto \varphi_X(s_1, \ldots, s_d) \) is monotonically increasing separately in every argument.

The probability-generating function \( \varphi_X \) contains the information about all distributions arising from \( X \) by a linear transformation with coefficients in \( \mathbb{N}_0 \).

**Lemma 4.6.** Let \( X = (X_1, \ldots, X_d)^T \) be an \( \mathbb{N}_0^d \)-valued random vector with probability-generating function \( \varphi_X \) and let \( A = (a_{i,j})_{i \in \{1, \ldots, c\}, j \in \{1, \ldots, d\} } \in \mathbb{N}_0^{c \times d} \) be a matrix. Then the probability-generating function of the distribution of the random vector \( AX \) satisfies

\[ \varphi_{AX}(s_1, \ldots, s_c) = \varphi_X(t_1, \ldots, t_d) \quad \text{with} \quad t_j := \prod_{i=1}^c s_i^{a_{i,j}}, \quad j \in \{1, \ldots, d\}, \] (4.15)

at least for every \( s = (s_1, \ldots, s_c) \in \mathbb{C}^c \) with \( \|s\|_{\infty} \leq 1 \).
Proof. Using the definitions in (4.1) and (4.15),

$$\varphi_{AX}(s_1, \ldots, s_c) = \mathbb{E} \left[ \prod_{i=1}^{c} s_i^{\sum_{j=1}^{d} a_{i,j} X_j} \right] = \mathbb{E} \left[ \prod_{j=1}^{d} \left( \prod_{i=1}^{c} s_i a_{i,j} X_j \right) \right] = \varphi_X(t_1, \ldots, t_d).$$

Example 4.7. Let us rewrite (4.15) for three special cases.

(a) For the first $c$-dimensional marginal distribution with $c \in \{1, \ldots, d\}$,

$$\varphi_{(X_1, \ldots, X_c)}(s_1, \ldots, s_c) = \varphi_X(s_1, \ldots, s_c, 1, \ldots, 1),$$

because $a_{i,j} = \delta_{i,j}$ for $i \in \{1, \ldots, c\}$ and $j \in \{1, \ldots, d\}$, i.e.

$$A = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots & \vdots & & \vdots \\ \vdots & \ddots & \ddots & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{pmatrix}.$$

(b) Sum of all the $d$ components of $X$,

$$\varphi_{X_1 + \cdots + X_d}(s_1) = \varphi_X(s_1, \ldots, s_1),$$

because $A = (1, \ldots, 1) \in \mathbb{N}_0^{1 \times d}$ is actually a row vector.

(c) Addition of the last $d - c + 1$ components of $X$, for every $c \in \{2, \ldots, d\}$,

$$\varphi_{(X_1, \ldots, X_{c-1}, X_c + \cdots + X_d)}(s_1, \ldots, s_c) = \varphi_X(s_1, \ldots, s_{c-1}, s_c, \ldots, s_c),$$

because

$$A = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots & \vdots & & \vdots \\ \vdots & \ddots & \ddots & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 1 \end{pmatrix}.$$

This observation can be used to prove the aggregation property for several multi-dimensional distributions discussed below.

(d) For every permutation $\sigma$ of $\{1, \ldots, d\}$, with $\sigma^{-1}$ denoting the inverse permutation,

$$\varphi_{(X_{\sigma(1)}, \ldots, X_{\sigma(d)})}(s_1, \ldots, s_d) = \varphi_X(s_{\sigma^{-1}(1)}, \ldots, s_{\sigma^{-1}(d)}),$$

because $a_{i,j} = \delta_{\sigma(i),j} = \delta_{i,\sigma^{-1}(j)}$ for all $i, j \in \{1, \ldots, d\}$. 
Example 4.8 (Multivariate Bernoulli distribution revisited). Assume that the random vector $B = (B_1, \ldots, B_d)$ with $d \geq 2$ has a multivariate Bernoulli distribution, i.e. $B \sim \text{Multinomial}(1, p_1, \ldots, p_d)$ as in Example 4.5. Using the probability-generating function from (4.2) and (4.7),

$$
\varphi_{B_i}(s_i) = p_is_i + (1 - p_i) = p_is_i + \sum_{j=1}^{d} p_j = \varphi_B(1, \ldots, 1, s_i, 1, \ldots, 1), \quad s_i \in \mathbb{C},
$$

for every $i \in \{1, \ldots, d\}$, which illustrates (4.16). See Remark 4.43 below for higher-dimensional marginal distributions of Multinomial$(1, p_1, \ldots, p_d)$.

Information about the multivariate factorial moments of the $\mathbb{N}_0^d$-valued $X$ can also be obtained in a simple manner. Let us first consider component $i \in \{1, \ldots, d\}$. At least for all $s = (s_1, \ldots, s_d) \in \mathbb{C}^d$ with $\|s\|_\infty \leq 1$ and $|s_i| < 1$,

$$
\frac{\partial}{\partial s_i} \varphi_X(s) = \mathbb{E}\left[ s_1^{X_1} \cdots s_{i-1}^{X_{i-1}} X_i s_i^{X_i - 1} s_{i+1}^{X_{i+1}} \cdots s_d^{X_d} \right]
$$

and

$$
\frac{\partial^2}{\partial s_i^2} \varphi_X(s) = \mathbb{E}\left[ s_1^{X_1} \cdots s_{i-1}^{X_{i-1}} X_i(X_i - 1) s_i^{X_i - 2} s_{i+1}^{X_{i+1}} \cdots s_d^{X_d} \right].
$$

More generally, taking partial differentiation with respect all $d$ variables into account, at least for all $s = (s_1, \ldots, s_d) \in \mathbb{C}^d$ with $\|s\|_\infty < 1$,

$$
\varphi_X^{(n)}(s) = \mathbb{E}\left[ \prod_{i=1}^{d} \left( s_i^{X_i - n_i} \prod_{l=0}^{n_i-1} (X_i - l_i) \right) \right], \quad n = (n_1, \ldots, n_d) \in \mathbb{N}_0^d.
$$

It follows from (4.1) that $\varphi_X$ and its derivatives are monotonically increasing on $[0, 1]^d$. By monotone convergence for the left-hand side limits at the $i$th position, for every $i \in \{1, \ldots, d\}$,

$$
\frac{\partial}{\partial s_i} \varphi_X(1, \ldots, 1, s_i, 1, \ldots, 1) \bigg|_{s_i = 1^-} = \mathbb{E}[X_i] \quad (4.19)
$$

and

$$
\frac{\partial^2}{\partial s_i^2} \varphi_X(1, \ldots, 1, s_i, 1, \ldots, 1) \bigg|_{s_i = 1^-} = \mathbb{E}[X_i(X_i - 1)], \quad (4.20)
$$

and generally for the multivariate factorial moments,

$$
\varphi_X^{(n)}(1, \ldots, 1) = \mathbb{E}\left[ \prod_{i=1}^{d} \prod_{l_i=0}^{n_i-1} (X_i - l_i) \right], \quad n = (n_1, \ldots, n_d) \in \mathbb{N}_0^d, \quad (4.21)
$$

where the precaution with the left-hand side limit is unnecessary for those $i \in \{1, \ldots, d\}$ which satisfy $n_i = 0$. The precaution is also unnecessary when there exists an $r > 1$ such that the power series in (4.1) converges for all $s \in \mathbb{C}^d$ with $\|s\|_\infty < r$. It follows from a proposition on doubly monotone arrays [57].
Section A5.1] or an iterated application of the monotone convergence theorem that \( \varphi_X^{(n)}(1,\ldots,1-1) \) does not depend on the order in which the left-hand side limits are taken. As the next example shows, these left-hand side limits can be infinite, which is also the reason for calculating partial derivatives only in the interior of the domain of definition.

**Example 4.9 (A distribution on \( \mathbb{N} \) with infinite expectation).** Consider an \( \mathbb{N} \)-valued random variable \( X \) with

\[
\mathbb{P}[X = n] = \frac{1}{n(n+1)} = \frac{1}{n} - \frac{1}{n+1}, \quad n \in \mathbb{N}.
\]

Since \( \mathbb{P}[X \in \{1,\ldots,k\}] = 1 - \frac{1}{k+1} \to 1 \) as \( k \to \infty \), this is indeed a probability distribution. Its probability-generating function \( \varphi_X \) satisfies

\[
\varphi_X(s) = \left( \sum_{n=1}^{\infty} \frac{s^n}{n(n+1)} \right)' = \sum_{n=1}^{\infty} \frac{s^{n-1}}{n+1}, \quad |s| < 1.
\]

Comparison with the harmonic series and application of the monotone convergence theorem (or Abel’s theorem for power series) shows that \( E[X] = \varphi_X'(1-) = \infty \).

**Remark 4.10 (Variances and Covariances).** Consider an \( \mathbb{N}_0^d \)-valued random vector \( X \). For every component \( i \in \{1,\ldots,d\} \) with \( E[X_i] < \infty \), we can use

\[
\text{Var}(X_i) = E[X_i^2] - (E[X_i])^2 = E[X_i(X_i - 1)] - E[X_i](E[X_i] - 1)
\]

as well as (4.19) and (4.20) to calculate the variance. For \( i,j \in \{1,\ldots,d\} \) with \( i \neq j \), a special case of (4.21) is

\[
\frac{\partial^2 \varphi_X}{\partial s_i \partial s_j}(1,\ldots,1,1-,\ldots,1,1-,\ldots,1) = E[X_iX_j],
\]

where the left-hand side limits are considered for the \( i \)th and \( j \)th argument. Therefore, if \( E[X_i] < \infty \) and \( E[X_j] < \infty \), then we can use

\[
\text{Cov}(X_i, X_j) = E[X_iX_j] - E[X_i]E[X_j]
\]

together with (4.19) and (4.23) to calculate the covariance of \( X_i \) and \( X_j \), which is allowed to be infinite here.

**Exercise 4.11** (Factorial moments and variance of the univariate logarithmic distribution). Suppose that \( N \sim \text{Log}(p) \) with \( p \in [0,1) \), see Example 4.4. Show that

\[
E \left[ \prod_{l=0}^{n-1} (N - l) \right] = \frac{(n-1)!p^{n-1}}{c(p)(1-p)^n}, \quad n \in \mathbb{N},
\]

and

\[
\text{Var}(N) = \frac{c(p) - 1}{c^2(p)(1-p)^2}
\]

with \( c(p) \) given by (4.5). For the multivariate case, see Exercise 4.37.
Exercise 4.12 (Calculating mixed moments from multivariate factorial moments). Extending Exercise 2.9 to the multivariate case, show that in the polynomial ring $R[x_1, \ldots, x_d]$ of $d$ variables over a commutative ring $R$ (with 1),

$$
x^n = \sum_{l \in \mathbb{N}_0^d} \prod_{i=1}^d \left( \prod_{k_i=0}^{l_i-1} (x_i - k_i) \right), \quad n = (n_1, \ldots, n_d) \in \mathbb{N}_0^d,
$$

where $x = (x_1, \ldots, x_d)$ and the inequality $l \leq n$ is understood componentwise. Conclude that, for every $\mathbb{N}_0^d$-valued random vector $N = (N_1, \ldots, N_d)$, the mixed moments can be calculated from the multivariate factorial moments given in (4.21) by the formula

$$
\mathbb{E}[N^n] = \sum_{l \in \mathbb{N}_0^d} \left( \prod_{i=1}^d \binom{n_i}{l_i} \right) \mathbb{E} \left[ \prod_{i=1}^d \prod_{k_i=0}^{l_i-1} (N_i - k_i) \right], \quad n = (n_1, \ldots, n_d) \in \mathbb{N}_0^d,
$$

and that the formula is also true for $\mathbb{C}^d$-valued random vectors, provided the absolute multivariate factorial moments for the right-hand side are finite.

Now the multiplication theorem of probability-generating functions mentioned above. Its proof is so simple that we include it in the statement of the theorem.

**Theorem 4.13.** Suppose that $X, Y : \Omega \to \mathbb{N}_0^d$ are independent. Then, using multi-index notation,

$$
\varphi_{X+Y}(s) = \mathbb{E}[s^{X+Y}] = \mathbb{E}[s^X] \mathbb{E}[s^Y] = \varphi_X(s) \varphi_Y(s)
$$

at least for all $s \in \mathbb{C}^d$ with $\|s\|_{\infty} \leq 1$.

An application of this formula provides a very short proof of the Poisson summation theorem given in Lemma 3.2.

**Alternative proof of Lemma 3.2.** Let $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\mu)$ be independent. Using (4.27) first and then (4.3),

$$
\varphi_{X+Y}(s) = \varphi_X(s) \varphi_Y(s) = e^{\lambda(s-1)} e^{\mu(s-1)} = e^{(\lambda+\mu)(s-1)}, \quad s \in \mathbb{C}.
$$

Hence $X + Y \sim \text{Poisson}(\lambda + \mu)$, because $\varphi_{X+Y}$ uniquely determines the distribution of $X + Y$, see (4.14). □

**Example 4.14** (Binomial distribution). Let the random variable $N \sim \text{Bin}(m, p)$ describe the number of successes in $m \in \mathbb{N}$ independent Bernoulli trials with success probability $p \in [0, 1]$, meaning that $N = B_1 + \cdots + B_m$ with independent Bernoulli random variables $B_1, \ldots, B_m$. By (4.2), for every $i \in \{1, \ldots, m\}$,

$$
\varphi_{B_i}(s) = 1 + p(s-1), \quad s \in \mathbb{C},
$$

hence the multiplication theorem of probability-generating functions, see (4.27), implies that

$$
\varphi_N(s) = \prod_{i=1}^m \varphi_{B_i}(s) = (1 + p(s-1))^m, \quad s \in \mathbb{C}.
$$

49
Example 4.15 (Multivariate Poisson distribution). Assume that \( N \) has the multivariate Poisson distribution \( \text{MPoisson}(G, (\lambda_g)_{g \in G}, m) \) as in Definition 3.40. By the representation (3.48), using multi-index notation, the probability-generating function is given by

\[
\varphi_N(s) = \mathbb{E}[s^N] = \mathbb{E} \left[ \prod_{g \in G} (s^{s^g})^{N_g} \right], \quad s \in \mathbb{C}^m,
\]

(4.30)

where \( s^g = \prod_{i \in g} s_i \) by (3.49). Using the independence of \((N_g)_{g \in G}\) and the multiplication theorem (4.27) of probability-generating functions,

\[
\varphi_N(s) = \prod_{g \in G} \mathbb{E}[(s^{s^g})^{N_g}], \quad s \in \mathbb{C}^m.
\]

Finally, using the probability-generating function of \( \text{Poisson}(\lambda_g) \) for every \( g \in G \), see Example 4.3,

\[
\varphi_N(s) = \prod_{g \in G} \exp(\lambda_g(s^{s^g} - 1)) = \exp\left( \sum_{g \in G} \lambda_g(s^{s^g} - 1) \right), \quad s \in \mathbb{C}^m.
\]

(4.31)

Example 4.16 (Multinomial distribution). Given a dimension \( d \in \mathbb{N} \), let \( B_1, \ldots, B_m \) be \( m \in \mathbb{N} \) independent \( d \)-dimensional random vectors, each one having a multivariate Bernoulli distribution with probability vector \( p = (p_1, \ldots, p_d) \in [0, 1]^d \) satisfying \( p_1 + \cdots + p_d = 1 \), see Example 4.5, i.e. \( B_j \sim \text{Multinomial}(1, p) \) for each \( j \in \{1, \ldots, m\} \). We can interpret \( B_j \) as describing the result of the \( j \)th trial, which can have \( d \) different outcomes. Then the \( i \)th component \( N_i \) of \( N := B_1 + \cdots + B_m \) describes the number of outcomes of type \( i \) in a sequence of \( m \) independent trials, for every \( i \in \{1, \ldots, d\} \). By definition, \( N \) has a multinomial distribution, which we denote by \( \text{Multinomial}(m, p_1, \ldots, p_d) \) or \( \text{Multinomial}(m, p) \) for short. By (4.7), the probability-generating function of \( B_j \) is given by

\[
\varphi_{B_j}(s) = \sum_{i=1}^d p_i s_i, \quad s = (s_1, \ldots, s_d) \in \mathbb{C}^d,
\]

for every \( j \in \{1, \ldots, m\} \), hence the multiplication theorem of probability-generating functions, see (4.27), implies that

\[
\varphi_N(s) = \prod_{j=1}^m \varphi_{B_j}(s) = \left( \sum_{i=1}^d p_i s_i \right)^m, \quad s = (s_1, \ldots, s_d) \in \mathbb{C}^d.
\]

(4.32)

Either by using (4.14) to derive the probability mass function from \( \varphi_N \), or by using the multinomial theorem to expand \( \varphi_N(s) = (p_1 s_1 + \cdots + p_d s_d)^m \), it follows that

\[
\mathbb{P}[N = (n_1, \ldots, n_d)] = m! \prod_{i=1}^d \frac{p_i^{n_i}}{n_i!} = \binom{m}{n_1, \ldots, n_d} p^n
\]

(4.33)
in multi-index notation for all \(n = (n_1, \ldots, n_d) \in \mathbb{N}_0^d\) with \(n_1 + \cdots + n_d = m\). Note that
\[
\frac{m!}{n_1! \cdots n_d!} = \binom{m}{n_1, \ldots, n_d}
\]
is the multinomial coefficient.

**Exercise 4.17.** Let \(N = (N_1, \ldots, N_d) \sim \text{Multinomial}(m, p_1, \ldots, p_d)\) with parameters \(m \in \mathbb{N}\) and \(p = (p_1, \ldots, p_d) \in [0, 1]^d\) satisfying \(p_1 + \cdots + p_d = 1\). Show the following:

(a) \(N_1 + \cdots + N_d \equiv m\).

(b) One-dimensional marginal distributions: \(N_i \sim \text{Bin}(m, p_i)\) hence \(\mathbb{E}[N] = mp_i\) and \(\text{Var}(N_i) = mp_i(1-p_i)\) for every \(i \in \{1, \ldots, d\}\). (See Remark 4.43 below for higher-dimensional marginal distributions.)

(c) Aggregation property: For every \(i \in \{1, \ldots, d-1\}\),
\[(N_1, \ldots, N_i, N_{i+1} + \cdots + N_d) \sim \text{Multinomial}(m, p_1, \ldots, p_i, p_{i+1} + \cdots + p_d).\]

(d) Permutation property: For every permutation \(\sigma\) of \(\{1, \ldots, d\}\),
\[(N_{\sigma(1)}, \ldots, N_{\sigma(d)}) \sim \text{Multinomial}(m, p_{\sigma(1)}, \ldots, p_{\sigma(d)}).\]

(e) Covariances: \(\text{Cov}(N_i, N_j) = -mp_ip_j\) for all \(i, j \in \{1, \ldots, d\}\) with \(i \neq j\).

**Lemma 4.18** (Summation property of the multinomial distribution). Let \(d, k \in \mathbb{N}\), \(m_1, \ldots, m_k \in \mathbb{N}_0\) and \(p_1, \ldots, p_d \in [0, 1]\) with \(p_1 + \cdots + p_d = 1\). If \(N_1, \ldots, N_k\) are independent with \(N_i \sim \text{Multinomial}(m_i, p_1, \ldots, p_d)\) for every \(i \in \{1, \ldots, k\}\), then
\[N := \sum_{i=1}^{k} N_i \sim \text{Multinomial}(m_1 + \cdots + m_k, p_1, \ldots, p_d).\]

**Exercise 4.19.** Prove Lemma 4.18 using (4.32).

**Remark 4.20** (Summation property of the binomial distribution). Using Lemma 4.18 for \(d = 2\) and looking at the one-dimensional marginal distribution (cf. Exercise 4.17(b)), we obtain the summation property of the binomial distribution. Of course, this also follows directly using (4.29).

**Remark 4.21.** The following observation uses generating functions to make the Poisson approximation of Theorem 3.23 plausible. Let \(\varphi_{B_i}\) denote the probability-generating function of the Bernoulli random variable \(B_i\) of obligor \(i \in \{1, \ldots, m\}\), indicating a default with probability \(p_i\). As in (4.2),
\[\varphi_{B_i}(s) = 1 + p_i(s - 1), \quad s \in \mathbb{C}.\]

We denote the number of defaults in the whole portfolio by \(W = B_1 + \cdots + B_m\) and the corresponding generating function by \(\varphi_W\). If we assume the defaults
of the obligors to be independent, then \( \varphi_W(s) = \prod_{i=1}^{m} \varphi_{B_i}(s) \). Using the linear approximation \( 1 + x \approx e^x \) for \(|x|\) small, we get

\[
\varphi_W(s) = \prod_{i=1}^{m} (1 + p_i(s - 1)) \approx \prod_{i=1}^{m} e^{p_i(s-1)} = e^{\lambda(s-1)}, \quad s \in \mathbb{C},
\]

with \( \lambda := p_1 + \cdots + p_m \), which according to (4.3) is the probability-generating function of \( N \sim \text{Poisson}(\lambda) \).

### 4.2 Application to the General Multivariate Poisson Mixture Model

After this excursion to probability-generating functions, let us reconsider the general multivariate Poisson mixture model introduced in Subsection 3.6. We start by considering the random vector \((N_1, \ldots, N_m)^\top\) of defaults of the individual \( m \) obligors and use the representation (3.57) and the calculation (4.30) to see that the corresponding probability-generating function is given by

\[
\varphi(N_1, \ldots, N_m)(s) = \mathbb{E} \left[ \prod_{g \in G} (s^g)^{N_g} \right] \quad \text{with} \quad s^g = \prod_{i \in g} s_i
\]

at least for all \( s = (s_1, \ldots, s_m) \in \mathbb{C}^m \) with \( \|s\|_{\infty} \leq 1 \). By conditioning on the random intensities \((\Lambda_h)_{h \in G}\) and using conditional independence, see (3.53), as well as (3.52), it follows that

\[
\varphi(N_1, \ldots, N_m)(s) = \mathbb{E} \left[ \mathbb{E} \left[ \prod_{g \in G} (s^g)^{N_g} \bigg| (\Lambda_h)_{h \in G} \right] \right] = \mathbb{E} \left[ \exp \left( \sum_{g \in G} \Lambda_g (s^g - 1) \right) \right].
\]

For the number of defaults \( N = N_1 + \cdots + N_m \) in the portfolio as considered in (3.59), Example 4.7(b) yields

\[
\varphi_N(s) = \varphi(N_1, \ldots, N_m)(s, \ldots, s) = \mathbb{E} \left[ \exp \left( \sum_{g \in G} \Lambda_g (s^g| - 1) \right) \right],
\]

at least for all \( s \in \mathbb{C} \) with \( |s| \leq 1 \). In the case \( G = \{1, \ldots, m\} \), writing \( \Lambda_i := \Lambda_{\{i\}} \) for each \( i \in \{1, \ldots, m\} \), the last result simplifies to

\[
\varphi_N(s) = \mathbb{E}[e^{(\Lambda_1 + \cdots + \Lambda_m)(s-1)}]
\]

and, provided \( \Lambda_1, \ldots, \Lambda_m \) are independent, to

\[
\varphi_N(s) = \prod_{i=1}^{m} \mathbb{E}[e^{\Lambda_i(s-1)}].
\]
4.3 Properties of the Gamma Distribution

Until now, no assumption was made about the distribution of any $\Lambda_i$. In this section we will consider only one factor $\Lambda$. An arbitrary, but well-accepted choice for mathematical convenience, is the gamma distribution. Therefore, suppose $\Lambda$ to be gamma-distributed (notation $\Lambda \sim \text{Gamma}(\alpha, \beta)$) with shape parameter $\alpha > 0$ and rate (or inverse scale) parameter $\beta > 0$, i.e., $\Lambda$ has a density

$$f(\lambda) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta \lambda} & \text{for } \lambda > 0, \\ 0 & \text{for } \lambda \leq 0, \end{cases}$$

(4.34)

where $\Gamma$ denotes the gamma function given in (2.27). The integral substitution $x = \beta \lambda$ shows that $f$ is indeed a probability density.

Note that Gamma$(1, \beta)$ is the exponential distribution with rate parameter $\beta > 0$, whereas Gamma$(n, \beta)$ with general $n \in \mathbb{N}$ is called Erlang distribution. Furthermore, Gamma$(\frac{n}{2}, \frac{1}{2})$ is called $\chi^2$-distribution with $n \in \mathbb{N}$ degrees of freedom.

The next lemma shows that for every inverse scale parameter $\beta > 0$ the gamma distributions $\{\text{Gamma}(\alpha, \beta)\}_{\alpha > 0}$ form a semigroup under convolution. It also shows that the gamma distribution is infinitely divisible.

Lemma 4.22 (Summation property of the gamma distribution). Let $k \in \mathbb{N}$ and $\alpha_1, \ldots, \alpha_k, \beta > 0$. If $\Lambda_1, \ldots, \Lambda_k$ are independent random variables with $\Lambda_i \sim \text{Gamma}(\alpha_i, \beta)$ for every $i \in \{1, \ldots, k\}$, then

$$\sum_{i=1}^{k} \Lambda_i \sim \text{Gamma}(\alpha_1 + \cdots + \alpha_k, \beta).$$

Proof. The lemma follows by induction as soon as it is proved for $k = 2$. Let $f_1$ and $f_2$ be densities according to (4.34) for $\Lambda_1 \sim \text{Gamma}(\alpha_1, \beta)$ and $\Lambda_2 \sim \text{Gamma}(\alpha_2, \beta)$, respectively. Due to independence of $\Lambda_1$ and $\Lambda_2$, a density $f$ for $\Lambda := \Lambda_1 + \Lambda_2$ is given by the convolution, i.e., for all $\lambda > 0$,

$$f(\lambda) = \int_{0}^{\lambda} f_1(\mu)f_2(\lambda - \mu) \, d\mu$$

$$= \int_{0}^{\lambda} \frac{\beta^{\alpha_1}}{\Gamma(\alpha_1)} \mu^{\alpha_1-1} e^{-\beta \mu} \frac{\beta^{\alpha_2}}{\Gamma(\alpha_2)} (\lambda - \mu)^{\alpha_2-1} e^{-\beta (\lambda - \mu)} \, d\mu.$$

Rearranging, defining $\alpha = \alpha_1 + \alpha_2$, and using the substitution $\mu = \lambda x$ yields

$$f(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta \lambda} \cdot \frac{\Gamma(\alpha)}{\Gamma(\alpha_1) \Gamma(\alpha_2)} \int_{0}^{1} x^{\alpha_1-1}(1-x)^{\alpha_2-1} \, dx, \quad \lambda > 0,$$

where the remaining constant needs to equal 1, because both sides are probability distributions. As a side effect, this calculation evaluates the beta function $B(\alpha_1, \alpha_2)$, see Exercise 2.4 and (2.31).
4.3.1 Moments of the Gamma Distribution

For $\gamma \in (-\alpha, \infty)$ and $z \in (-\infty, \beta)$, we can generally compute

$$
E[\Lambda^\gamma e^{\Lambda z}] = \int_0^\infty \lambda^\gamma e^{\lambda z} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta \lambda} d\lambda \\
= \frac{\Gamma(\alpha + \gamma)}{\Gamma(\alpha)} \frac{\beta^\alpha}{(\beta - z)^{\alpha+\gamma}} \int_0^\infty \frac{(\beta - z)^{\alpha+\gamma}}{\Gamma(\alpha + \gamma)} \lambda^{\alpha+\gamma-1} e^{-(\beta - z)\lambda} d\lambda \\
= \frac{\Gamma(\alpha + \gamma)}{\beta^\gamma \Gamma(\alpha)} (1 - z/\beta)^{-(\alpha+\gamma)}.
$$

(4.35)

For $z = 0$, the calculation (4.35) gives all the moments

$$
E[\Lambda^\gamma] = \frac{\Gamma(\alpha + \gamma)}{\beta^\gamma \Gamma(\alpha)}, \quad \gamma \in (-\alpha, \infty),
$$

(4.36)

in particular, using the functional equation (2.28) for the gamma function,

$$
E[\Lambda] = \frac{\Gamma(\alpha + 1)}{\beta \Gamma(\alpha)} = \frac{\alpha}{\beta},
$$

(4.37)

$$
E[\Lambda^2] = \frac{\Gamma(\alpha + 2)}{\beta^2 \Gamma(\alpha)} = \frac{\alpha(\alpha + 1)}{\beta^2},
$$

$$
\text{Var}(\Lambda) = E[\Lambda^2] - (E[\Lambda])^2 = \frac{\alpha}{\beta^2}.
$$

(4.38)

For $\gamma = 0$, the calculation (4.35) gives the exponential moments and the moment-generating function

$$
E[e^{\Lambda z}] = (1 - z/\beta)^{-\alpha}, \quad z \in (-\infty, \beta),
$$

(4.39)

and the Laplace transform

$$
E[e^{-s\Lambda}] = (1 + s/\beta)^{-\alpha}, \quad s \in (-\beta, \infty).
$$

Given $\gamma \in (-\alpha, \infty)$, let $\Lambda' \sim \text{Gamma}(\alpha + \gamma, \beta)$, where the shape parameter is shifted by $\gamma$. Then (4.35), (4.36) and (4.39) imply the peculiar relation

$$
E[\Lambda^\gamma e^{\Lambda z}] = E[\Lambda^\gamma] (1 - z/\beta)^{-(\alpha+\gamma)} = E[\Lambda^\gamma] E[e^{\Lambda' z}], \quad z \in (-\infty, \beta),
$$

(4.40)

which we will use to derive (6.90) below.

4.3.2 Biased Gamma Distribution

The following lemma makes clear that the peculiar relation (4.40) is the consequence of a more general observation, which is very similar to Lemma 2.11 for the beta distribution.
Lemma 4.23. Assume that $\Lambda \sim \text{Gamma}(\alpha, \beta)$ with parameters $\alpha, \beta > 0$ and that $\gamma \in (-\infty, \alpha)$ and $\delta \in (-\infty, \beta)$. Then $\mathbb{P}_{\Lambda \sim \text{Gamma}(\alpha, \beta)} \eta = \text{Gamma}(\alpha + \gamma, \beta + \delta)$, that means the distribution of $\Lambda$ under the $\Lambda^\gamma e^{-\delta \Lambda}$-biased probability measure $\mathbb{P}_{\Lambda \sim \text{Gamma}(\alpha, \beta)}$ given by Definition 2.16 is the Gamma$(\alpha + \gamma, \beta + \delta)$ distribution.

Proof. By (4.35) and (2.40), a density of the $\Lambda^\gamma e^{-\delta \Lambda}$-biased probability measure $\mathbb{P}_{\Lambda \sim \text{Gamma}(\alpha, \beta)}$ is given by

$$\frac{d\mathbb{P}_{\Lambda \sim \text{Gamma}(\alpha, \beta)}^{\sim < \delta \Lambda}}{d\mathbb{P}} = \frac{\beta^\gamma}{\Gamma(\alpha + \gamma)} \left(1 + \frac{\delta}{\beta}\right)^{\alpha + \gamma} \Lambda^\gamma e^{-\delta \Lambda} = \frac{(\beta + \delta)^{\alpha + \gamma}}{\beta^\alpha \Gamma(\alpha + \gamma)} \Lambda^\gamma e^{-\delta \Lambda}.$$

Let $\mu$ denote the Lebesgue measure on $\mathbb{R}$. Using the density $f$ from (4.34) shows that, for $\mu$-almost all $\lambda > 0$,

$$\frac{d\mathbb{P}_{\Lambda \sim \text{Gamma}(\alpha, \beta)}^{\sim < \delta \Lambda}}{d\mu} (\lambda) = \frac{d\mathbb{P}_{\Lambda \sim \text{Gamma}(\alpha, \beta)}^{\sim < \delta \Lambda}}{d\mathbb{P}} (\lambda) \cdot \frac{d\mathbb{P}^{\sim < \delta \Lambda}}{d\mu} (\lambda)$$

$$= \frac{(\beta + \delta)^{\alpha + \gamma}}{\beta^\alpha \Gamma(\alpha + \gamma)} \Lambda^\gamma e^{-\delta \Lambda} \cdot f(\lambda)$$

$$= \frac{(\beta + \delta)^{\alpha + \gamma}}{\Gamma(\alpha + \gamma)} \Lambda^{\alpha + \gamma - 1} e^{-(\beta + \delta) \lambda},$$

which by (4.34) gives a density of the Gamma$(\alpha + \gamma, \beta + \delta)$ distribution. $\square$

### 4.4 Gamma-Mixed Poisson Distribution

In continuation of the investigation of $N$ from above, assume that the conditional distribution of $N$ given $\Lambda$ is Poisson$(\Lambda)$, notation $\mathcal{L}(N|\Lambda) \overset{\text{a.s.}}{=} \text{Poisson}(\Lambda)$, meaning that

$$\mathbb{P}[N = n|\Lambda] \overset{\text{a.s.}}{=} \frac{\Lambda^n}{n!} e^{-\Lambda}, \quad n \in \mathbb{N}_0. \quad (4.41)$$

Combining (4.41) and (4.35) with $z = -1$, the unconditional distribution of $N$ is

$$\mathbb{P}[N = n] = \mathbb{E} [\mathbb{P}[N = n|\Lambda]] = \frac{1}{n!} \mathbb{E} [\Lambda^n e^{-\Lambda}] = \frac{1}{n! \Gamma(\alpha)} \frac{\Gamma(\alpha + n)}{\beta^n (1 + 1/\beta)^{\alpha + n}}$$

for all $n \in \mathbb{N}_0$. Using the functional equation (2.28) of the gamma function and the abbreviations

$$p = \frac{1}{1 + \beta} \in (0, 1) \quad \text{and} \quad q = 1 - p = \frac{\beta}{1 + \beta}, \quad (4.42)$$

we get

$$\mathbb{P}[N = n] = \frac{\Gamma(\alpha + n)}{n! \Gamma(\alpha)} p^n q^\alpha = \binom{\alpha + n - 1}{n} p^n q^\alpha, \quad n \in \mathbb{N}_0, \quad (4.43)$$

which is called the negative binomial distribution\footnote{The term $\binom{\alpha + n - 1}{n} p^n q^\alpha$ in (4.43) shows up when considering the binomial series for $(1 - p)^{-\alpha}$, see (5.20) and (5.21), which is a negative power. This might motivate the name.}. We will use the notation $N \sim \text{NegBin}(\alpha, p)$. We will interpret $\text{NegBin}(0, p)$ with $p \in [0, 1)$ and
NegBin($\alpha, 0$) with $\alpha \geq 0$ as the degenerate distribution concentrated in 0. Note that the right-hand sides of (4.44), (4.45), (4.46) and (4.47) below, hence also (4.48) and (4.49) are correct for these cases.

If $\alpha \in \mathbb{N}$, then (4.43) gives the probability of exactly $n \in \mathbb{N}_0$ successes before the $\alpha$-th failure in a sequence of independent Bernoulli experiments with success probability $p$. For $\alpha = 1$, the negative binomial distribution (4.43) reduces to the geometric distribution with parameter $p \in [0, 1)$.

Let us calculate the expectation, the variance and the probability-generating function of $N$. Since $L(N|\Lambda) \overset{a.s.}{=} \text{Poisson}(\Lambda)$, we have

$$
\mathbb{E}[N] = \mathbb{E}[\mathbb{E}[N|\Lambda]] = \mathbb{E}[\Lambda] = \frac{\alpha p}{1-p} \tag{4.44}
$$

by (4.37) and the substitution $\beta = \frac{1-p}{p}$ arising from (4.42). Using Lemma 3.48 as well as (4.37) for the mean and (4.38) for the variance of $\Lambda$, we obtain

$$
\text{Var}(N) = \mathbb{E}[\text{Var}(N|\Lambda)] + \text{Var}(\mathbb{E}[N|\Lambda]) = \mathbb{E}[\Lambda] + \text{Var}(\Lambda) = \frac{\alpha p}{1-p} + \frac{\alpha}{\beta^2} = \frac{\alpha p}{1-p} \tag{4.45}
$$

where we used (4.42) and $\beta = \frac{1-p}{p}$ for the last equation. It remains to calculate the corresponding probability-generating function. Using (4.43) and expanding the fraction by $(1-ps)^{\alpha}$, it follows that

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

$$
= \frac{q^\alpha}{(1-ps)^\alpha} \sum_{n=0}^{\infty} \binom{\alpha + n - 1}{n} (1-ps)^\alpha (ps)^n = \left( \frac{q}{1-ps} \right)^\alpha \tag{4.46}
$$

for all real $s \geq 0$ with $ps < 1$, hence for all $s \in \mathbb{C}$ with $p|s| < 1$. Alternatively, using $L(N|\Lambda) \overset{a.s.}{=} \text{Poisson}(\Lambda)$ and the generating function (4.3) of the Poisson distribution,

$$
\varphi_{N|\Lambda}(s) := \mathbb{E}[s^N|\Lambda] \overset{a.s.}{=} e^{\Lambda(s-1)}, \quad s \in \mathbb{C},
$$

as well as the exponential moments (4.39) of $\Lambda \sim \text{Gamma}(\alpha, \beta)$,

$$
\varphi_N(s) = \mathbb{E}[\mathbb{E}[s^N|\Lambda]] = \mathbb{E}[e^{\Lambda(s-1)}] = \left( 1 - \frac{s - 1}{\beta} \right)^{-\alpha} = \left( \frac{\beta}{1 + \beta - s} \right)^\alpha \tag{4.47}
$$

for all $s \in \mathbb{C}$ with $p|s| < 1$. Since

$$
\varphi_N^{(n)}(s) = \frac{p^n q^\alpha}{(1-ps)^{\alpha+n}} \prod_{l=0}^{n-1} (\alpha + l), \quad n \in \mathbb{N}, \tag{4.48}
$$

56
it follows via (4.21) for the factorial moments of the negative binomial distribution that

\[ E \left[ \prod_{l=0}^{n-1} (N - l) \right] = \frac{p^n}{q^n} \prod_{l=0}^{n-1} (\alpha + l), \quad n \in \mathbb{N}. \] (4.49)

Here is the analogue of the Poisson and gamma summation properties given in Lemma 3.2 and Lemma 4.22 respectively, transferred to independent random variables with a negative binomial distribution (see Lemma 4.40 below for a multi-dimensional generalization):

**Lemma 4.24** (Summation property of the negative binomial distribution). Let \( k \in \mathbb{N} \) and \( \alpha_1, \ldots, \alpha_k \geq 0 \) as well as \( p \in [0, 1) \). If \( N_1, \ldots, N_k \) are independent with \( N_i \sim \text{NegBin}(\alpha_i, p) \) for every \( i \in \{1, \ldots, k\} \), then

\[ N := \sum_{i=1}^{k} N_i \sim \text{NegBin}(\alpha_1 + \cdots + \alpha_k, p). \]

**Proof.** By independence, cf. (4.27), and generating function from (4.47),

\[ \varphi_N(s) = \prod_{i=1}^{k} \varphi_{N_i}(s) = \prod_{i=1}^{k} \left( \frac{q}{1 - ps} \right)^{\alpha_i} = \left( \frac{q}{1 - ps} \right)^{\alpha_1 + \cdots + \alpha_k} \] (4.50)

for all \( s \in \mathbb{C} \) satisfying \( p|s| < 1 \). Therefore, \( N \sim \text{NegBin}(\alpha, p) \) with \( \alpha = \alpha_1 + \cdots + \alpha_k \), because the probability-generating function uniquely determines the distribution, cf. (4.14).

**Exercise 4.25.** Give a more probabilistic proof of Lemma 4.24 by considering the negative binomial distribution as a gamma-mixed Poisson distribution and using Lemma 3.2 and Lemma 4.22.

### 4.5 Generating Function of Compound Distributions

Assume that \( N \) is \( \mathbb{N}_0 \)-valued and that \((X_n)_{n \in \mathbb{N}}\) is a sequence of \( \mathbb{N}_0 \)-valued, independent, identically distributed random variables, which is independent of \( N \). To characterize the distribution of the \( \mathbb{N}_0 \)-valued random sum

\[ S := \sum_{n=1}^{N} X_n, \] (4.51)

we compute its generating function \( \varphi_S \). Using the multi-index notation as in Example 4.1, the dominated convergence theorem, the independence of the sum \( X_1 + \cdots + X_n \) from the event \( \{N = n\} \) as well as the i.i.d. assumption for the
sequence \((X_n)_{n \in \mathbb{N}}\),
\[
\varphi_S(s) = \mathbb{E}[s^{X_1 + \cdots + X_N}] = \sum_{n=0}^{\infty} \mathbb{E}[s^{X_1 + \cdots + X_n}1_{\{N=n\}}] \\
= \sum_{k=0}^{\infty} \mathbb{E}[s^{X_1 + \cdots + X_n}1_{\{N=n\}}] \mathbb{P}[N=n] \\
= (\mathbb{E}[s^{X_1}])^n = (\varphi_X(s))^n \\
= \varphi_N(\varphi_X(s)),
\] (4.52)

where is calculation is valid for all \(s \in \mathbb{C}^d\) such that the power series defining \(\varphi_X(s)\) is absolutely convergent and such that the power series defining \(\varphi_N\) converges at \(|\varphi_X(s)|\). This is the case at least for all \(s \in \mathbb{C}^d\) with \(||s||_\infty \leq 1\); note that \(|\varphi_X(s)| \leq 1\) for these \(s\).

**Example 4.26** (Pairwise independence is not enough for (4.52)). We emphasise that the i.i.d. sequence \((X_n)_{n \in \mathbb{N}}\) should be independent of \(N\); the independence of \(X_n\) and \(N\) for every \(n \in \mathbb{N}\), that means pairwise independence, is not enough for (4.52).

For a counterexample, consider an i.i.d. sequence \((X_n)_{n \in \mathbb{N}}\) with \(X_1 \sim \text{Bin}(1, \frac{1}{2})\), hence \(\varphi_X(s) = \frac{1}{2}(1 + s)\) for \(s \in \mathbb{C}\) by (4.2). Define \(N = 2 - ((X_1 + X_2) \mod 2)\). Then \(\mathbb{P}[N=1] = \mathbb{P}[N=2] = \frac{1}{2}\) and
\[
\mathbb{P}[N=1, X_i = j] = \mathbb{P}[X_i = j, X_{3-i} = 1 - j] = \frac{1}{4}
\]
as well as
\[
\mathbb{P}[N=2, X_i = j] = \mathbb{P}[X_1 = j, X_2 = j] = \frac{1}{4}
\]
for all \(i \in \{1, 2\}\) and \(j \in \{0, 1\}\), hence \(N\) and \(X_i\) are independent for every \(i \in \{1, 2\}\). Note that \(\varphi_N(s) = \frac{1}{2}s + \frac{1}{4}s^2\) and
\[
\varphi_N(\varphi_X(s)) = \frac{1}{4}(1 + s) + \frac{1}{8}(1 + s)^2 = \frac{3}{8} + \frac{1}{2}s + \frac{1}{8}s^2, \quad s \in \mathbb{C}. \quad (4.53)
\]

However, for the compound sum \(S\) given by (4.51), we have that \(\{S = 0\} = \{X_1 = 0\}\), \(\{S = 1\} = \{X_1 = 1, X_2 = 0\}\) and \(\{S = 2\} = \{X_1 = 1, X_2 = 1\}\), hence
\[
\varphi_S(s) = \frac{1}{2} + \frac{1}{4}s + \frac{1}{4}s^2, \quad s \in \mathbb{C},
\]

which differs from (4.53), hence (4.52) does not hold in this case.

Let 
\[Q = (q_{ij})_{i,j \in \mathbb{N}_0} \text{ with } q_{ij} := \mathbb{P}[X_1 = i, X_2 = j]\] denote the distribution of \(X_1\). If \(N \sim \text{Poisson}(\lambda)\) with \(\lambda \geq 0\), then the random sum \(S\) in (4.51) has a so-called compound Poisson distribution and we use the notation \(S \sim \text{CPoisson}(\lambda, Q)\). Since \(\varphi_N(s) = e^{\lambda(s-1)}\) for all \(s \in \mathbb{C}\) by (4.3), the calculation in (4.52) implies that
\[
\varphi_S(s) = \exp(\lambda(\varphi_X(s) - 1)) \quad (4.54)
\]
for all \(s \in \mathbb{C}^d\) for which the power series defining \(\varphi_X(s)\) converges, which is the case at least when \(||s||_\infty \leq 1\).

Similarly, if \(N \sim \text{NegBin}(\alpha, p)\) with \(\alpha \geq 0\) and \(p \in [0, 1]\), then \(S\) from (4.51) has a so-called compound negative binomial distribution and we use the notation
\( S \sim \text{CNegBin}(\alpha, p, Q). \) Since \( \varphi_N(s) = q^\alpha/(1-ps)^\alpha \) with \( q := 1 - p \) for all \( s \in \mathbb{C} \) with \( p|s| < 1 \) by (4.47), the calculation in (4.52) implies that

\[
\varphi_S(s) = \left( \frac{q}{1 - p \varphi_X(s)} \right)^\alpha
\]

for all \( s \in \mathbb{C}^d \) for which the power series defining \( \varphi_X(s) \) is absolutely convergent and for which \( p|\varphi_X(s)| < 1 \), which is the case at least when \( ||s||_\infty \leq 1 \).

Let us look at a prominent example and its credit risk interpretation.

**Example 4.27 (Negative binomial distribution as compound Poisson distribution).** Let \((X_n)_{n \in \mathbb{N}}\) denote i.i.d. random variables, where \( X_1 \sim \text{Log}(p) \) has a logarithmic distribution with parameter \( p \in (0, 1) \), cf. Example 4.4. Recall (4.6) to see that

\[
\varphi_{X_1}(s) = \frac{\log(1 - ps)}{\log(1 - p)}, \quad |s| < 1/p.
\]

According to (4.54), the compound Poisson sum \( S \) has the generating function

\[
\varphi_S(s) = \exp\left( \lambda \left( \frac{\log(1 - ps)}{\log(1 - p)} - 1 \right) \right) = \left( \frac{1 - p}{1 - ps} \right)^\alpha, \quad |s| < 1/p,
\]

with

\[
\alpha := -\frac{\lambda}{\log(1 - p)} \geq 0,
\]

which according to (4.47) is the probability-generating function of a negative binomial distribution, hence \( \text{CPoisson}(\lambda, \text{Log}(p)) = \text{NegBin}(\alpha, p) \).

**Remark 4.28.** As a historical remark, note that the result of Example 4.27 can be traced back at least to H. Ammeter\(^{20}\) [2]. At [2, top of page 183] he makes the Ansatz to write the characteristic function of a compound negative binomial distribution as a characteristic function of a compound Poisson distribution. He uses \( h_0 \) and \( P/(h_0 + P) \) for our parameters \( \alpha \) and \( p \) to specify \( \text{NegBin}(\alpha, p) \), hence \( P \) is the expectation of the distribution, cf. (4.44). At the bottom of the page, he obtains the logarithmic distribution with parameter \( \frac{1}{1+\chi} \) where \( \chi = P/h_0 \), which is our parameter \( p \), and also the Poisson intensity \( \frac{P}{\chi} \log(1 + \chi) \), which simplifies to \(-\alpha \log(1 - p)\) in our notation and agrees with (4.56).

**Remark 4.29 (Interpretation of the negative binomial distribution as a model for dependent defaults).** Motivated by the Poisson approximation discussed in Section 3.4, we can model the number of defaults in a credit portfolio during one period by \( N \sim \text{Poisson}(\lambda) \) with \( \lambda > 0 \) and visualize \( N \) as the number of events of a homogeneous Poisson process of intensity \( \lambda \) (see [39, Section 2.1]) during \([0, 1]\). To reflect the imprecise knowledge of the rate parameter \( \lambda \), we can model it by a

\(^{20}\) Prof. Dr. h.c. Hans A. Ammeter (1912–1986), president of the Schweizerische Lebensversicherungs- und Rentenanstalt (now Swiss Life) from 1973 to 1978.
random factor $\Lambda \sim \text{Gamma}(\alpha, \beta)$ with $\alpha, \beta > 0$ such that $E[\Lambda] = 1$ and express the uncertainty by $\text{Var}(\Lambda) = \sigma^2 > 0$. We assume that $\mathcal{L}(N|\Lambda) \overset{a.s.}{=} \text{Poisson}(\lambda \Lambda)$, which implies that $E[N] = E[E[N|\Lambda]] = E[\lambda \Lambda] = \lambda$. Since $E[\Lambda] = \alpha/\beta$ and $\text{Var}(\Lambda) = \alpha/\beta^2$ by (4.37) and (4.38), this means $\alpha = \beta = 1/\sigma^2$. Then $\Lambda \sim \text{Gamma}(\alpha, \beta/\lambda) = \text{Gamma}(1/\sigma^2, 1/(\lambda \sigma^2))$, hence $N \sim \text{NegBin}(1/\sigma^2, p)$ with

$$p = \frac{1}{1 + \beta/\lambda} = \frac{1}{1 + 1/(\lambda \sigma^2)} = \frac{\lambda \sigma^2}{1 + \lambda \sigma^2}$$

as shown in Section 4.3 and we can visualize $N$ as the number of events of a mixed Poisson process of random intensity $\lambda \Lambda$ during $[0, 1]$ (see [39, Section 2.3], it is also a special version of a Cox process). Example 4.27 offers another interpretation of the distribution of $N$: We can consider a compound Poisson process with reduced intensity

$$\lambda' = -\alpha \log(1 - p) = -\frac{1}{\sigma^2} \log \frac{1}{1 + \lambda \sigma^2} = \lambda f(\lambda \sigma^2), \quad (4.57)$$

see (4.56), where

$$f(x) := \frac{1}{x} \log(1 + x), \quad x > 0, \quad (4.58)$$

see Figure 4.1. At the $i$th event of the Poisson process, there are one or several joint defaults given by $X_i \sim \text{Log}(p)$ with

$$E[X_i] = \frac{p}{(1 - p) \log(1 - p)} = \frac{\lambda \sigma^2}{\log(1 + \lambda \sigma^2)} = \frac{1}{f(\lambda \sigma^2)}, \quad i \in \mathbb{N}, \quad (4.59)$$
see (4.25). This leads to the same distribution of the number of defaults during 
\([0, 1]\), namely 
\[N \sim \text{CPoisson}(\lambda', \log(p)).\]

As a corollary to the summation property of the Poisson distribution (Lemma 3.2) and the negative binomial distribution (Lemma 4.24), we get the corresponding property for the compound distributions.

**Corollary 4.30.** Let \(Q, Q_1, \ldots, Q_k\) be probability distributions on \(\mathbb{N}_0^d\) and let \(S_1, \ldots, S_k\) be independent.

(a) Let \(\lambda_1, \ldots, \lambda_k \geq 0\). If \(S_i \sim \text{CPoisson}(\lambda_i, Q_i)\) for every \(i \in \{1, \ldots, k\}\), then
\[S_1 + \cdots + S_k \sim \text{CPoisson}(\lambda_1 + \cdots + \lambda_k, Q), \]
if \(Q\) satisfies\(^{21}\) \((\lambda_1 + \cdots + \lambda_k)Q = \lambda_1 Q_1 + \cdots + \lambda_k Q_k\).

(b) Let \(\alpha_1, \ldots, \alpha_k \geq 0\) and \(p \in [0, 1)\). If \(S_i \sim \text{CNegBin}(\alpha_i, p, Q)\) for every \(i \in \{1, \ldots, k\}\), then
\[S_1 + \cdots + S_k \sim \text{CNegBin}(\alpha_1 + \cdots + \alpha_k, p, Q).\]

**Exercise 4.31.** Prove Corollary 4.30. Hint: Use probability-generating functions, (4.28), (4.50), (4.52), (4.54) and (4.55).

**Remark 4.32.** The definitions and Corollary 4.30 can be extended to probability distributions \(Q, Q_1, \ldots, Q_k\) on \(\mathbb{R}^d\). In this case the proof can be done using characteristic functions.

**Lemma 4.33** (Representation of the multivariate Poisson distribution as compound Poisson distribution). Given \(\text{MPoisson}(G, (\lambda_g)_{g \in G}, m)\) as in Definition 3.40, define the total intensity by \(\lambda = \sum_{g \in G} \lambda_g\) and let \(\mu\) be a probability measure on \(\{0, 1\}^m\) satisfying \(\lambda \mu = \sum_{g \in G} \lambda_g \delta_{c_g}\), where \(\delta_{c_g}\) denotes the Dirac measure concentrated in \(c_g \in \{0, 1\}^m\) given by (3.49). Then \(\text{MPoisson}(G, (\lambda_g)_{g \in G}, m) = \text{CPoisson}(\lambda, \mu).\)

**Proof.** The probability-generating function of the Dirac measure \(\delta_{c_g}\) is given in multi-index notation by \(\varphi_{\delta_{c_g}}(s) = s^{c_g}\) for all \(s \in \mathbb{C}^m\), hence
\[\lambda \varphi_{\mu}(s) = \sum_{g \in G} \lambda_g s^{c_g}, \quad s \in \mathbb{C}^m.\]

Therefore, using (4.54), the probability-generating function \(\varphi\) of \(\text{CPoisson}(\lambda, \mu)\) is given by
\[\varphi(s) = \exp(\lambda(\varphi_{\mu}(s) - 1)) = \exp\left(\sum_{g \in G} \lambda_g (s^{c_g} - 1)\right), \quad s \in \mathbb{C}^m,\]
which agrees with the probability-generating function (4.31) of the multivariate Poisson distribution \(\text{MPoisson}(G, (\lambda_g)_{g \in G}, m)\).

---

\(^{21}\)If \(\lambda_1 + \cdots + \lambda_k > 0\), then the degenerate case in excluded and \(Q\) is uniquely determined.
4.6 Some Compound Distributions

Arising from the Multivariate Bernoulli Distribution

Throughout this subsection, let \((B_m)_{m \in \mathbb{N}}\) denote i.i.d. multivariate Bernoulli random variables with \(B_1 \sim \text{Multinomial}(1, \tilde{p}_1, \ldots, \tilde{p}_d)\), where \(\tilde{p}_1, \ldots, \tilde{p}_d \in [0, 1] \) with \(\tilde{p}_1 + \cdots + \tilde{p}_d = 1\), see Example 4.5. Then \(\varphi_{B_i}(s) = \sum_{i=1}^{d} \tilde{p}_is_i\) for all \(s = (s_1, \ldots, s_d) \in \mathbb{C}^d\) by (4.7). Furthermore, let \(M\) be an \(\mathbb{N}_0\)-valued random variable, independent of \((B_m)_{m \in \mathbb{N}}\), and consider the random sum

\[
N = (N_1, \ldots, N_d) = \sum_{m=1}^{M} B_m. \tag{4.60}
\]

Remark 4.34 (Covariation of components). Suppose that \(\text{Var}(M) < \infty\). Using the representation from (4.60) and Lemma 3.48 applied with \(B = \sigma(M)\),

\[
\text{Cov}(N_i, N_j) = \text{Cov} \left( \mathbb{E}[N_i | M], \mathbb{E}[N_j | M] \right) + \mathbb{E} \left[ \text{Cov}(N_i, N_j | M) \right] = \tilde{p}_i \tilde{p}_j \left( \text{Var}(M) - \mathbb{E}[M] \right).
\]

Hence the sign of the covariation of two components can vary depending on the expectation and the variance of the distribution of \(M\). It vanishes for \(M \sim \text{Poisson}(\lambda)\) due to (3.3) and (3.4): Example 4.35 below shows that there is even independence in this case. For \(M \sim \text{Log}(p)\) the sign depends on the value of \(p \in (0, 1)\), see Exercise 4.41(b) below.

Example 4.35. (Compound Poisson) Let \(M \sim \text{Poisson}(\lambda)\) with \(\lambda \geq 0\). Then (4.7) and (4.54) imply for the random sum (4.60) that

\[
\varphi_M(s) = \exp \left( \lambda \left( \sum_{i=1}^{d} \tilde{p}_i s_i - 1 \right) \right) = \prod_{i=1}^{d} \exp \left( \lambda \tilde{p}_i (s_i - 1) \right) \tag{4.61}
\]

for all \(s = (s_1, \ldots, s_d) \in \mathbb{C}^d\), hence the components of \(N\) are independent and satisfy \(N_i \sim \text{Poisson}(\lambda \tilde{p}_i)\) for every \(i \in \{1, \ldots, d\}\). This independence may come as a surprise, because the components of the multivariate Bernoulli distributed summands are dependent. However, this independence is a special feature of the Poisson distribution, it is lost if, for example, the logarithmic distribution (see Subsection 4.6.1) or the negative binomial distribution (see Subsection 4.6.2) is considered for \(M\).

If \(\mathbb{P}[M = m] = 1\) for an \(m \in \mathbb{N}\), then \(N \sim \text{Multinomial}(m, \tilde{p}_1, \ldots, \tilde{p}_d)\) for the random variable in (4.60), see Example 4.16. More generally, given \((n_1, \ldots, n_d) \in \mathbb{N}_0^d\), define \(m = n_1 + \cdots + n_d \in \mathbb{N}_0\). Then \(N = (n_1, \ldots, n_d)\) is only possible when \(M = m\), hence by independence

\[
\mathbb{P}[N = (n_1, \ldots, n_d)] = \mathbb{P}[M = m] \mathbb{P}[B_1 + \cdots + B_m = (n_1, \ldots, n_d) | M = m]
\]

or

\[
= \mathbb{P}[M = m] \mathbb{P}[B_1 + \cdots + B_m = (n_1, \ldots, n_d)].
\]

62
Since \( B_1 + \cdots + B_m \sim \text{Multinomial}(m, \tilde{\pi}_1, \ldots, \tilde{\pi}_d) \), it follows from (4.33) that

\[
\mathbb{P}[N = (n_1, \ldots, n_d)] = \mathbb{P}[M = m] \cdot m! \prod_{i=1}^{d} \frac{p_i^{n_i}}{n_i!},
\]

(4.62)

In the next subsections, we will look at three additional interesting examples for the distribution of \( M \), namely the logarithmic distribution, the negative binomial distribution and the binomial distribution. Of course, additional choices are possible, like the extended negative binomial distribution (see Example 5.22), the extended logarithmic distribution (see Example 5.23) and truncations of these distribution (see Definition 5.3).

4.6.1 Multivariate Logarithmic Distribution

Consider \( M \sim \text{Log}(\bar{p}) \) with \( \bar{p} \in (0, 1) \), cf. Example 4.4. It follows from (4.4) and (4.62) that, for every \((n_1, \ldots, n_d) \in \mathbb{N}_0^d \setminus \{(0, \ldots, 0)\} \),

\[
\mathbb{P}[N = (n_1, \ldots, n_d)] = \frac{p^{m-1}}{c(p)} m \cdot m! \prod_{i=1}^{d} \frac{p_i^{n_i}}{n_i!} = \frac{(m-1)!}{c(p)} \prod_{i=1}^{d} \frac{p_i^{n_i}}{n_i!},
\]

with \( m := n_1 + \cdots + n_d \) and \( p_i := \tilde{\pi}_i \) for \( i \in \{1, \ldots, d\} \). This motivates the following definition:

**Definition 4.36 (Multivariate logarithmic distribution).** A random vector \( N = (N_1, \ldots, N_d) \) of dimension \( d \in \mathbb{N} \) is said to have the multivariate logarithmic distribution \( M\text{Log}(p_1, \ldots, p_d) \) with parameters \( p_1, \ldots, p_d \in [0, 1) \) satisfying \( 0 < p := p_1 + \cdots + p_d < 1 \), if

\[
\mathbb{P}[N = (n_1, \ldots, n_d)] = \frac{(n_1 + \cdots + n_d - 1)!}{c(p)p} \prod_{i=1}^{d} \frac{p_i^{n_i}}{n_i!},
\]

(4.63)

for all \((n_1, \ldots, n_d) \in \mathbb{N}_0^d \setminus \{(0, \ldots, 0)\} \) with normalising factor, cf. (4.5),

\[
c(p) := -\log(1 - p).
\]

For \( d = 1 \), Definition 4.36 reduces to the univariate logarithmic distribution given in Example 4.4 which is well-defined also for \( p = 0 \).

With

\[
\varphi_M(s) = \frac{\log(1 - ps)}{\log(1 - p)}, \quad |s| < 1/p,
\]

given by (4.6) and \( \varphi_B(s) = \sum_{i=1}^{d} \tilde{\pi}_i s_i \) for \( s = (s_1, \ldots, s_d) \in \mathbb{C}^d \) given by (4.7), it follows from (4.52) for the probability-generating function of \( N \) that

\[
\varphi_N(s) = \frac{\log(1 - \sum_{i=1}^{d} p_i s_i)}{\log(1 - p)}
\]

(4.64)

for all \( s = (s_1, \ldots, s_d) \in \mathbb{C}^d \) with \( |\sum_{i=1}^{d} p_i s_i| < 1 \), which is certainly the case if \( \|s\|_{\infty} < 1/p \).
Exercise 4.37 (Properties of the multivariate logarithmic distribution). Assume that $N = (N_1, \ldots, N_d) \sim \text{MLog}(p_1, \ldots, p_d)$ with $p_1, \ldots, p_d \in [0, 1)$ satisfying $0 < p := p_1 + \cdots + p_d < 1$, cf. Definition 4.36. Show:

(a) Factorial moments and variances: For every $(n_1, \ldots, n_d) \in \mathbb{N}_0^d \setminus \{0\}$,

$$
\mathbb{E} \left[ \prod_{i=1}^d \prod_{l=0}^{n_i-1} (N_i - l_i) \right] = -\frac{(n_1 + \cdots + n_d - 1)! \log(1 - p)}{\prod_{i=1}^d \left( \frac{p_i}{1 - p} \right)^{n_i}},
$$

and for every component $i \in \{1, \ldots, d\}$,

$$
\text{Var}(N_i) = -p_i \left( \frac{p_i}{1 - p} \right)^{n_i} \frac{1 - p_i}{(1 - p)^2 \log^2(1 - p)}.
$$

In the case $d = 1$, these results coincide with (4.25) and (4.26), respectively.

(b) Covariances: For every $i, j \in \{1, \ldots, d\}$ with $i \neq j$,

$$
\text{Cov}(N_i, N_j) = -p_i p_j \frac{1 + \log(1 - p)}{(1 - p)^2 \log^2(1 - p)}.
$$

hence $\text{Cov}(N_i, N_j) \geq 0$ for $p \geq 1 - \frac{1}{e} \approx 0.6321$ and negative otherwise.

(c) Permutation property: For every permutation $\sigma$ of $\{1, \ldots, d\}$,

$$(N_{\sigma(1)}, \ldots, N_{\sigma(d)}) \sim \text{MLog}(p_{\sigma(1)}, \ldots, p_{\sigma(d)}).$$

(d) Aggregation property: For every $i \in \{1, \ldots, d-1\}$,

$$(N_1, \ldots, N_i, N_{i+1} + \cdots + N_d) \sim \text{MLog}(p_1, \ldots, p_i, p_{i+1} + \cdots + p_d).$$

(e) $N_1 + \cdots + N_d \sim \text{Log}(p)$.

Remark 4.38. Parts (a) and (b) of Exercise 4.37 can be solved using probability-generating functions, cf. (4.19), (4.20), (4.22), (4.23), (4.24) and (4.64), or they can be solved using the representation (4.60) together with Lemma 3.48 and results for the multinomial distribution and the univariate logarithmic distribution, see Exercises 4.17 and 4.11, respectively.

4.6.2 Negative Multinomial Distribution

Let $M \sim \text{NegBin}(\alpha, p)$ with $\alpha > 0$ and $p \in [0, 1)$, see (4.43). It follows from (4.43) and (4.62) that, for every $(n_1, \ldots, n_d) \in \mathbb{N}_0^d$,

$$
\mathbb{P}[N = (n_1, \ldots, n_d)] = \frac{\Gamma(\alpha + m)}{m! \Gamma(\alpha)} p^m q^a \cdot \prod_{i=1}^d \frac{p_i^{n_i}}{n_i!} = \frac{\Gamma(\alpha + m)}{\Gamma(\alpha)} q^a \prod_{i=1}^d \frac{p_i^{n_i}}{n_i!},
$$

with $m := n_1 + \cdots + n_d$ and $p_i := p \tilde{p}_i$ for $i \in \{1, \ldots, d\}$. This motivates the following definition:
Definition 4.39 (Negative multinomial distribution). A random vector \( N = (N_1, \ldots, N_d) \) of dimension \( d \in \mathbb{N} \) is said to have the **negative multinomial distribution** \( \text{NegMult}(\alpha, p_1, \ldots, p_d) \) with shape parameter \( \alpha > 0 \) and success probabilities \( p_1, \ldots, p_d \in [0, 1) \) satisfying \( q := 1 - (p_1 + \cdots + p_d) \in (0, 1] \), if

\[
\mathbb{P}[N = (n_1, \ldots, n_d)] = \frac{\Gamma(\alpha + n_1 + \cdots + n_d)}{\Gamma(\alpha)} q^{\alpha} \prod_{i=1}^{d} \frac{p_i^{n_i}}{n_i!}
\]  

(4.66)

for all \((n_1, \ldots, n_d) \in \mathbb{N}_0^d\). We interpret \( \text{NegMult}(0, p_1, \ldots, p_d) \) as the degenerate distribution concentrated in \((0, \ldots, 0) \in \mathbb{N}_0^d\).

For \( d = 1 \), Definition 4.39 reduces to the negative binomial distribution given by (4.43).

For \( \alpha \in \mathbb{N} \) the negative multinomial distribution has a combinatorial interpretation: Consider the \( d \) components as mutually different types of successes, which occur with probabilities \( p_1, \ldots, p_d \), and let \( q \) denote the probability of failure. Using the functional equation of the gamma function, (4.66) can be rewritten with a multinomial coefficient as

\[
\mathbb{P}[N = (n_1, \ldots, n_d)] = \binom{\alpha - 1 + n_1 + \cdots + n_d}{n_1, \ldots, n_d} q^{\alpha} \prod_{i=1}^{d} p_i^{n_i}
\]  

(4.67)

for \((n_1, \ldots, n_d) \in \mathbb{N}_0^d\), and the product in (4.67) could be written using multi-index notation. In a sequence of independent trials, (4.67) gives the probability of \( n_1, \ldots, n_d \in \mathbb{N}_0 \) successes of types \( 1, \ldots, d \) before the \( \alpha \)th failure happens.

With \( \varphi_M(s) = \left( \frac{q}{1 - ps} \right)^\alpha \), \( s \in \mathbb{C} \) with \( p|s| < 1 \), given by (4.40) and \( \varphi_B(s) = \sum_{i=1}^{d} \tilde{p}_i s_i \) for \( s = (s_1, \ldots, s_d) \in \mathbb{C}^d \) given by (4.7), it follows from (4.52) for the probability-generating function of \( N \) that

\[
\varphi_N(s) = \left( \frac{q}{1 - \sum_{i=1}^{d} p_i s_i} \right)^\alpha
\]  

(4.68)

for all \( s = (s_1, \ldots, s_d) \in \mathbb{C}^d \) with \( \left| \sum_{i=1}^{d} p_i s_i \right| < 1 \), which is certainly the case if \( (p_1 + \cdots + p_d)|s|_\infty < 1 \). Note that the calculation leading to (4.68) is correct for \( p_1 = \cdots = p_d = 0 \), and the result (4.68) is also correct for \( \alpha = 0 \).

Here is the multi-dimensional generalization of Lemma 4.24, which also implies that the negative multinomial distribution is infinitely divisible:

**Lemma 4.40.** Let \( k \in \mathbb{N}, \alpha_1, \ldots, \alpha_k \geq 0 \) and \( p_1, \ldots, p_d \in [0, 1) \) with \( p_1 + \cdots + p_d < 1 \). If \( N_1, \ldots, N_k \) are independent with \( N_i \sim \text{NegMult}(\alpha_i, p_1, \ldots, p_d) \) for every \( i \in \{1, \ldots, k\} \), then

\[
N := \sum_{i=1}^{k} N_i \sim \text{NegMult}(\alpha_1 + \cdots + \alpha_k, p_1, \ldots, p_d).
\]  

(4.69)

Exercise 4.42 (Properties of the negative multinomial distribution). Assume that \( N = (N_1, \ldots, N_d) \sim \text{NegMult}(\alpha, p_1, \ldots, p_d) \) with \( \alpha \geq 0 \) and \( p_1, \ldots, p_d \in [0, 1) \) satisfying \( q := 1 - (p_1 + \cdots + p_d) \in (0, 1] \), cf. Definition 4.39. Show:

(a) Factorial moments and variances: For every \( (n_1, \ldots, n_d) \in \mathbb{N}_0^d \),
\[
E \left[ \prod_{i=1}^{d} \prod_{l=0}^{n_i-1} (N_i - l) \right] = \left( \frac{n_1 + \cdots + n_d - 1}{\prod_{l=0}^{d}} \right)^{\alpha + l} \prod_{i=1}^{d} \left( \frac{p_i}{q} \right)^{n_i},
\]
and for every component \( i \in \{1, \ldots, d\} \),
\[
\text{Var}(N_i) = \frac{\alpha p_i (p_i + q)}{q^2}.
\]
In the case \( d = 1 \), these results coincide with (4.49) and (4.45), respectively.

(b) Covariances: For every \( i, j \in \{1, \ldots, d\} \) with \( i \neq j \),
\[
\text{Cov}(N_i, N_j) = \frac{\alpha p_i p_j}{q^2}.
\]

(c) Permutation property: For every permutation \( \sigma \) of \( \{1, \ldots, d\} \),
\[
(N_{\sigma(1)}, \ldots, N_{\sigma(d)}) \sim \text{NegMult}(\alpha, p_{\sigma(1)}, \ldots, p_{\sigma(d)}).
\]

(d) Aggregation property: For every \( i \in \{1, \ldots, d-1\} \),
\[
(N_1, \ldots, N_i, N_{i+1} + \cdots + N_d) \sim \text{NegMult}(\alpha, p_1, \ldots, p_{i+1} + \cdots + p_d).
\]

(e) \( N_1 + \cdots + N_d \sim \text{NegBin}(\alpha, p_1 + \cdots + p_d) \).

(f) Marginal distributions: For every \( i \in \{1, \ldots, d\} \),
\[
(N_1, \ldots, N_i) \sim \text{NegMult}\left( \alpha, \frac{p_1}{1 - p_{i+1} - \cdots - p_d}, \ldots, \frac{p_i}{1 - p_{i+1} - \cdots - p_d} \right),
\]
in particular \( N_i \sim \text{NegBin}(\alpha, \frac{p_i}{p_i + q}) \).

4.6.3 Multivariate Binomial Distribution

Let \( M \sim \text{Bin}(m, p) \) with \( m \in \mathbb{N}_0 \) and \( p \in [0, 1] \). It follows from (2.9) and (4.62) that, for every \( (n_1, \ldots, n_d) \in \mathbb{N}_0^d \) with \( l := n_1 + \cdots + n_d \leq m \),
\[
\mathbb{P}[N = (n_1, \ldots, n_d)] = \binom{m}{l} (1 - p)^{m-l} \prod_{i=1}^{d} \frac{p_i^{n_i}}{n_i!}
\]
\[
= \frac{m!}{(m-l)!} (1 - p)^{m-l} \prod_{i=1}^{d} \frac{p_i^{n_i}}{n_i!},
\]
Remark 4.43 (Relation to multinomial distribution). Note that the multivariate binomial distribution is not a new distribution, but already contained in the multinomial distribution (see Example 4.16) by looking at marginals, more precisely, if \((N_1, \ldots, N_d) \sim \text{MBin}(m, p_1, \ldots, p_d)\) with \(m \in \mathbb{N}_0\) and \(p_1, \ldots, p_d \in [0, 1]\) satisfying \(p_1 + \cdots + p_d \leq 1\). For \(d = 1\), this coincides with the binomial distribution, compare (2.9) with (4.70). If \(p = 1\), hence \(p_1 + \cdots + p_d = 1\), then \(\text{MBin}(m, p_1, \ldots, p_d) = \text{Multinomial}(m, p_1, \ldots, p_d)\).

With \(\varphi_M(s) = (1 + p(s - 1))^m\) for \(s \in \mathbb{C}\) as in (4.29) and \(\varphi_B(s) = \sum_{i=1}^d p_i s_i\) for \(s = (s_1, \ldots, s_d) \in \mathbb{C}^d\) given by (4.7), it follows from (4.52) for the probability-generating function of \(N\) that

\[
\varphi_N(s) = \left(1 + \sum_{i=1}^d p_i(s_i - 1)\right)^m, \quad s = (s_1, \ldots, s_d) \in \mathbb{C}^d. \tag{4.71}
\]

Note that the multivariate binomial distribution is precisely, if \((N_1, \ldots, N_d) \sim \text{MBin}(m, p_1, \ldots, p_d)\) with \(m \in \mathbb{N}_0\) and \(p_1, \ldots, p_d \in [0, 1]\) satisfying \(p_1 + \cdots + p_d \leq 1\), then it follows from (4.33) and (4.70) that

\[
(N_1, \ldots, N_d, m - (N_1 + \cdots + N_d)) \sim \text{Multinomial}(m, p_1, \ldots, p_d, 1 - (p_1 + \cdots + p_d)). \tag{4.72}
\]

The other way round, if \((N_1, \ldots, N_d) \sim \text{Multinomial}(m, p_1, \ldots, p_d)\) with \(m \in \mathbb{N}_0\) and \(p_1, \ldots, p_d \in [0, 1]\) satisfying \(p_1 + \cdots + p_d = 1\), then, using the aggregation property of the multinomial distribution from Exercise 4.17, (c) and (7.2),

\[
(N_1, \ldots, N_i) \sim \text{MBin}(m, p_1, p_i) \tag{4.73}
\]

for every \(i \in \{1, \ldots, d\}\). Of course, (4.72) and (4.73) can also be proved by applying (4.16) to the probability generating functions (4.32) and (4.68).

Due to Remark 4.43, the multivariate binomial distribution inherits many properties of the multinomial distribution given in Exercise 4.17.

Exercise 4.44. Let \(N = (N_1, \ldots, N_d) \sim \text{MBin}(m, p_1, \ldots, p_d)\) with parameters \(m \in \mathbb{N}_0\) and \(p_1, \ldots, p_d \in [0, 1]\) satisfying \(p_1 + \cdots + p_d \leq 1\). Show the following:

(a) \(N_1 + \cdots + N_d \sim \text{Bin}(m, p_1 + \cdots + p_d)\).

(b) Aggregation property: For every \(i \in \{1, \ldots, d-1\}\),

\[
(N_1, \ldots, N_i, N_{i+1} + \cdots + N_d) \sim \text{MBin}(m, p_1, \ldots, p_i, p_{i+1} + \cdots + p_d).
\]

(c) Marginal distributions: For every \(i \in \{1, \ldots, d\}\),

\[
(N_1, \ldots, N_i) \sim \text{MBin}(m, p_1, \ldots, p_i).
\]

(d) Permutation property: For every permutation \(\sigma\) of \(\{1, \ldots, d\}\),

\[
(N_{\sigma(1)}, \ldots, N_{\sigma(d)}) \sim \text{MBin}(m, p_{\sigma(1)}, \ldots, p_{\sigma(d)}).
\]
(e) Expectations and variances: \( \mathbb{E}[N_i] = mp_i \) and \( \text{Var}(N_i) = mp_i(1 - p_i) \) for every \( i \in \{1, \ldots, d\} \).

(f) Covariances: \( \text{Cov}(N_i, N_j) = -mp_ip_j \) for all \( i, j \in \{1, \ldots, d\} \) with \( i \neq j \).

**Lemma 4.45** (Summation property of the multivariate binomial distribution). Let \( k \in \mathbb{N}, m_1, \ldots, m_k \in \mathbb{N}_0 \) and \( p_1, \ldots, p_d \in [0, 1] \) with \( p_1 + \cdots + p_d \leq 1 \). If \( N_1, \ldots, N_k \) are independent with \( N_i \sim \text{MBin}(m_i, p_1, \ldots, p_d) \) for every \( i \in \{1, \ldots, k\} \), then

\[
N := \sum_{i=1}^{k} N_i \sim \text{MBin}(m_1 + \cdots + m_k, p_1, \ldots, p_d).
\] (4.74)

**Exercise 4.46.** Prove Lemma 4.45 (using (4.68) or Lemma 4.18 and (4.72)).

### 4.7 Conditional Compound Distributions

In the next step we look at the case, where \( N \) is conditionally Poisson-distributed, namely \( \mathcal{L}(N|\Lambda) \overset{a.s.}{=} \text{Poisson}(\Lambda) \) for a non-negative random variable \( \Lambda \). To compute the generating function of the random sum \( S \) given in (4.51), conditioned on \( \Lambda \), first note that

\[
\varphi_{N|\Lambda}(s) := \mathbb{E}[s^N | \Lambda] \overset{a.s.}{=} \exp(\Lambda(s - 1)), \quad s \in \mathbb{C},
\] (4.75)

by (4.3). Assume that the i.i.d. sequence \( (X_n)_{n \in \mathbb{N}} \) is not only independent of \( N \), but even independent of \( (\Lambda, N) \). Then, for every \( n \in \mathbb{N}_0 \), using multi-index notation and the multiplication theorem for probability-generating functions,

\[
\mathbb{1}_{\{N=n\}} \mathbb{E}[s^{X_1+\cdots+X_N} | \Lambda, N] \overset{a.s.}{=} \mathbb{1}_{\{N=n\}} \mathbb{E}[s^{X_1+\cdots+X_n} | \Lambda, N] \\
\overset{a.s.}{=} \mathbb{1}_{\{N=n\}} \mathbb{E}[s^{X_1+\cdots+X_n}] \\
= \mathbb{1}_{\{N=n\}} (\varphi_{X_1}(s))^n,
\]

hence

\[
\mathbb{E}[s^{X_1+\cdots+X_N} | \Lambda, N] \overset{a.s.}{=} (\varphi_{X_1}(s))^N
\] (4.76)

for all \( s \in \mathbb{C}^d \) for which the power series defining \( \varphi_{X_1}(s) \) converges, which is the case at least for all \( s \in \mathbb{C}^d \) with \( \|s\|_{\infty} \leq 1 \). Hence for these \( s \in \mathbb{C}^d \), by using the tower property of conditional expectation, (4.76) and (4.75),

\[
\varphi_{S|\Lambda}(s) := \mathbb{E}[s^{X_1+\cdots+X_N} | \Lambda] \\
\overset{a.s.}{=} \mathbb{E}[\mathbb{E}[s^{X_1+\cdots+X_N} | \Lambda, N] | \Lambda] \\
\overset{a.s.}{=} \varphi_{N|\Lambda}(\varphi_{X_1}(s)) \\
\overset{a.s.}{=} \exp(\Lambda(\varphi_{X_1}(s) - 1)),
\] (4.77)

and therefore

\[
\varphi_S(s) = \mathbb{E}[\varphi_{S|\Lambda}(s)] = \mathbb{E}[\exp(\Lambda(\varphi_{X_1}(s) - 1))]
\] (4.78)
at least for all \( s \in \mathbb{C}^d \) with \( \|s\|_\infty \leq 1 \), which generalizes (4.54).

If \( \Lambda \sim \Gamma(\alpha, \beta) \) with \( \alpha, \beta > 0 \), then \( N \sim \text{NegBin}(\alpha, p) \) with \( p = \frac{1}{1+\beta} \) by (4.43), hence \( S \sim \text{CNegBin}(\alpha, p, Q) \), where \( Q \) denotes the distribution of \( X_1 \), and the probability-generating function of \( S \) is given by (4.55). Evaluating the right-hand side of (4.78) using the exponential moment of \( \Lambda \) given by (4.39) and \( \beta = \frac{1-p}{p} \) leads to

\[
\varphi_S(s) = \varphi_N(\varphi_{X_1}(s)) = \left(1 - \frac{\varphi_{X_1}(s) - 1}{\beta}\right)^{-\alpha} = \left(\frac{1 - p}{1 - p\varphi_{X_1}(s)}\right)^\alpha
\]

(4.79)

at least for all \( s \in \mathbb{C}^d \) with \( \|s\|_\infty \leq 1 \), which agrees with (4.55).

**Exercise 4.47.** Let \((X_{m,n})_{m,n\in\mathbb{N}}\) denote a family of \( \mathbb{N}_0^d \)-valued i.i.d. random variables. Given \( \alpha, \beta, \lambda > 0 \), let \( \Lambda \sim \text{Gamma}(\alpha, \beta) \) and \( \mathcal{L}(N|\Lambda) \overset{a.s.}{=} \text{Poisson}(\lambda \Lambda) \). Assume that \((X_{m,n})_{m,n\in\mathbb{N}}\) and \((\Lambda, N)\) are independent. Define \( p = \lambda/(\beta + \lambda) \) and \( \mu = -\alpha \log(1-p) \). Let \((N_m)_{m\in\mathbb{N}}\) be an i.i.d. sequence with \( N_1 \sim \text{Log}(p) \) and let \( M \sim \text{Poisson}(\mu) \). Assume the \( M \), the sequence \((N_m)_{m\in\mathbb{N}}\) and the double-indexed sequence \((X_{m,n})_{m,n\in\mathbb{N}}\) are independent.

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

(b) Show by calculating the probability-generating functions of

\[
S := \sum_{n=1}^N X_{1,n} \quad \text{and} \quad S' := \sum_{m=1}^M \sum_{n=1}^{N_m} X_{m,n},
\]

that they have the same distribution, which is then a compound negative binomial as well as a compound Poisson distribution.

(c) Assume that \( \mathbb{E}[\Lambda] = 1 \) and \( \text{Var}(\Lambda) = \sigma^2 > 0 \). Determine \( \alpha, \beta > 0 \) and conclude that

\[
p = \frac{\lambda \sigma^2}{1 + \lambda \sigma^2}
\]

and

\[
\mu = -\frac{1}{\sigma^2} \log \left(1 - \frac{\lambda \sigma^2}{1 + \lambda \sigma^2}\right) = \frac{\lambda}{1 + \lambda \sigma^2} \left(1 + \sum_{n=2}^\infty \frac{1}{n} \left(\frac{\lambda \sigma^2}{1 + \lambda \sigma^2}\right)^{n-1}\right)
\]

**4.7.1 Expectation, Variance and Covariance**

Assume that \( N \) is \( \mathbb{N}_0 \)-valued and that \((X_n)_{n\in\mathbb{N}}\) is a sequence of \( \mathbb{N}_0^d \)-valued, independent, identically distributed random vectors \( X_n = (X_{n,1}, \ldots, X_{n,d}) \), which is independent of \( N \). We want to calculate the expectations, variances and covariances of the components \((S_1, \ldots, S_d)\) of the random sum \( S := X_1 + \cdots + X_N \) considered in (4.51).

Given \( k \in \mathbb{N}_0 \) with \( \mathbb{P}[N = k] > 0 \), we use the independence of the sum \( X_1 + \cdots + X_k \) from the event \( \{N = k\} \) as well as the i.i.d. assumption for \((X_n)_{n\in\mathbb{N}}\) to get in the case \( \mathbb{E}[\|X_1\|] < \infty \) that

\[
\mathbb{E}[S | N = k] = \mathbb{E}[X_1 + \cdots + X_k | N = k] = k \mathbb{E}[X_1],
\]
and in the case \(\mathbb{E}[\|X_1\|^2] < \infty\) that, for all \(i, j \in \{1, \ldots, d\}\),

\[
\Cov(S_i, S_j | N = k) = \Cov(X_{1,i} + \cdots + X_{k,i}, X_{1,j} + \cdots + X_{k,j} | N = k) \\
= \Cov(X_{1,i} + \cdots + X_{k,i}, X_{1,j} + \cdots + X_{k,j}) \\
= k \Cov(X_{1,i}, X_{1,j}).
\]

These two results can be rewritten as

\[
\mathbb{E}[S | N] \overset{a.s.}{=} N \mathbb{E}[X_1] \tag{4.80}
\]

and

\[
\Cov(S_i, S_j | N) \overset{a.s.}{=} N \Cov(X_{1,i}, X_{1,j}),
\]

where the last equation gives the conditional variance if we choose \(i = j\). Therefore, if \(N\) and \(X_1\) are integrable, we get as special case of Wald’s equation

\[
\mathbb{E}[S] = \mathbb{E}[\mathbb{E}[S | N]] = \mathbb{E}[N] \mathbb{E}[X_1] \tag{4.81}
\]

and, if they are square integrable, using Lemma 3.48,

\[
\Cov(S_i, S_j) = \mathbb{E}\{\Cov(S_i, S_j | N)\} + \mathbb{E}\{\Cov(S_i | N), \mathbb{E}[S_j | N]\} \\
= \mathbb{E}[N] \Cov(X_{1,i}, X_{1,j}) + \Var(N) \mathbb{E}[X_{1,i}] \mathbb{E}[X_{1,j}] \tag{4.82}
\]

which, for \(i = j\), is a special case of the Blackwell–Girshick equation.

We now specialize to the case where \(N\) is conditionally Poisson-distributed, namely \(\mathcal{L}(N | \Lambda) \overset{a.s.}{=} \text{Poisson}(\Lambda)\) for a non-negative random variable \(\Lambda\). Then the random sum \(S\) is conditionally compound Poisson given \(\Lambda\), hence, if \(\mathbb{E}[\|X_1\|] < \infty\), then taking conditional expectations in (4.80) replaces (4.81) by

\[
\mathbb{E}[S | \Lambda] \overset{a.s.}{=} \mathbb{E}[N | \Lambda] \mathbb{E}[X_1] \overset{a.s.}{=} \Lambda \mathbb{E}[X_1] \tag{4.83}
\]

and, if \(\mathbb{E}[\|X_1\|^2] < \infty\), then (4.82) turns into

\[
\Cov(S_i, S_j | \Lambda) \overset{a.s.}{=} \mathbb{E}[N | \Lambda] \Cov(X_{1,i}, X_{1,j}) + \Var(N | \Lambda) \mathbb{E}[X_{1,i}] \mathbb{E}[X_{1,j}] \\
\overset{a.s.}{=} \Lambda \mathbb{E}[X_{1,i}X_{1,j}] \tag{4.84}
\]

If \(N\) is unconditionally Poisson distributed, i.e., \(\mathcal{L}(N) = \text{Poisson}(\lambda)\), then (4.81) and (4.82) simplify to

\[
\mathbb{E}[S] = \lambda \mathbb{E}[X_1] \tag{4.85}
\]

and

\[
\Cov(S_i, S_j) = \lambda \mathbb{E}[X_{1,i}X_{1,j}] \tag{4.86}
\]
5 Recursive Algorithms and Weighted Convolutions

5.1 Panjer Distributions and Extended Panjer Recursion

As in Subsection 4.5, assume that \(N\) is \(\mathbb{N}_0\)-valued and that \((X_n)_{n \in \mathbb{N}}\) is a sequence of \(\mathbb{N}_0\)-valued, independent, identically distributed random variables, which are independent of \(N\). We want to calculate the distribution

\[
p_n := \mathbb{P}[S = n], \quad n \in \mathbb{N}_0,
\]

of the random sum \(S = X_1 + \cdots + X_N\) defined in (4.51). If the distribution

\[
q_n := \mathbb{P}[N = n], \quad n \in \mathbb{N}_0,
\]

of \(N\) satisfies the recursion formula given in Definition 5.1 below, then Theorem 5.8 shows that there is an efficient way to do this.

**Definition 5.1.** A probability distribution \((q_n)_{n \in \mathbb{N}_0}\) is called Panjer \((a,b,k)\) distribution with \(a,b \in \mathbb{R}\) and \(k \in \mathbb{N}_0\) if

\[
q_0 = q_1 = \cdots = q_{k-1} = 0 \quad \text{and} \quad q_n = \left(a + \frac{b}{n}\right) q_{n-1} \quad \text{for all } n \in \mathbb{N} \text{ with } n \geq k + 1.
\]

(5.1)

**Remark 5.2.** Given \(a,b \in \mathbb{R}\) and \(k \in \mathbb{N}_0\), the linearity of (5.1) implies that there exists at most one probability distribution \((q_n)_{n \in \mathbb{N}_0}\) satisfying Definition 5.1, because there can be at most one \(q_k \in [0,1]\) such that \(\sum_{n=k}^{\infty} q_n = 1\).

**Definition 5.3** (Truncation). Let \((q_n)_{n \in \mathbb{N}_0}\) be a probability distribution and \(l \in \mathbb{N}_0\) such that there is mass at \(l\) or above, meaning that \(\sum_{n=l}^{\infty} q_n > 0\). Then the \(l\)-truncated probability distribution \((\bar{q}_n)_{n \in \mathbb{N}_0}\) of \((q_n)_{n \in \mathbb{N}_0}\) is defined by

\[
\bar{q}_0 = \cdots = \bar{q}_{l-1} := 0 \quad \text{and} \quad \bar{q}_n := \frac{q_n}{1 - \sum_{j=0}^{l-1} q_j}, \quad n \geq l.
\]

(5.2)

**Lemma 5.4.** Suppose \((q_n)_{n \in \mathbb{N}_0}\) is the Panjer \((a,b,k)\) distribution and \(l \geq k\) is an integer such that there is mass at \(l\) or above. Then the \(l\)-truncation of \((q_n)_{n \in \mathbb{N}_0}\) is the Panjer \((a,b,l)\) distribution.

**Exercise 5.5.** Prove Lemma 5.4 using the linearity of the recursion equation (5.1)

**Remark 5.6.** All probability distributions satisfying Definition 5.1 were identified by Sundt and Jewell [50] for the case \(k = 0\), Willmot [58] for the case \(k = 1\), and finally Hess, Liewald and Schmidt [28] for general \(k \in \mathbb{N}_0\). The Panjer distributions are the following:

(a) Poisson distribution (cf. Example 5.17),

(b) Negative binomial distribution (cf. Example 5.18),
(c) Binomial distribution (cf. Example 5.20),
(d) Logarithmic distribution (cf. Example 5.21),
(e) Extended negative binomial distribution (cf. Example 5.22),
(f) Extended logarithmic distribution (cf. Example 5.23),
(g) All truncations of these distributions (cf. Definition 5.3 and Lemma 5.4).

Exercise 5.7. Prove that the only non-degenerate probability distributions in the class \{Panjer(a,b,0) \mid a, b \in \mathbb{R}\} are the Poisson, binomial, and the negative binomial distributions.

The following theorem combines results of Panjer [41] and Hess, Liewald and Schmidt [28] with the multivariate extension of Sundt [49]. For \(j, n \in \mathbb{N}_0^d\) we write \(j \leq n\) if this is true for all \(d\) components, and we write \(j < n\) if \(j \leq n\) and \(j \neq n\), meaning that there is strict inequality for at least one component. Note that \(\leq\) then a partial order on \(\mathbb{N}_0^d\). We write \(\langle \cdot, \cdot \rangle\) for the standard inner product in \(\mathbb{R}^d\).

**Theorem 5.8** (Multivariate extended Panjer recursion). Assume that the probability distribution \((q_n)_{n \in \mathbb{N}_0}\) of \(N\) is the Panjer \((a,b,k)\) distribution and that \(a \mathbb{P}[X_1 = 0] \neq 1\). Then the distribution \((p_n)_{n \in \mathbb{N}_0^d}\) of the random sum \(S\) defined in (4.51) can be calculated by

\[
p_0 = \varphi_N(\mathbb{P}[X_1 = 0]) = \begin{cases} q_0 & \text{if } \mathbb{P}[X_1 = 0] = 0, \\ \mathbb{E}[(\mathbb{P}[X_1 = 0])^N] & \text{otherwise,} \end{cases}
\]

(5.3)

where \(\varphi_N\) is the probability-generating function of \(N\), and the recursion formula

\[
p_n = \frac{1}{1 - a \mathbb{P}[X_1 = 0]} \left( \mathbb{P}[S_k = n] q_k + \sum_{j \in \mathbb{N}_0^d_0 < j \leq n} \left( a + b(\langle c_n, j \rangle) \right) \mathbb{P}[X_1 = j] p_{n-j} \right)
\]

(5.4)

for all \(n \in \mathbb{N}_0^d \setminus \{0\}\), where \(S_k := X_1 + \cdots + X_k\) and \(c_n \in \mathbb{R}^d\) is chosen such that \(\langle c_n, n \rangle \neq 0\); the vector \(c_n := (1, \ldots, 1)\) works in every case.

**Proof.** Theorem 5.8 is a corollary of Theorem 5.26(a) below, hence its proof is given just after the statement of Theorem 5.26.

**Remark 5.9** (Technical assumption). Of the Panjer distributions given in Remark 5.6, only the uninteresting case \(\mathbb{P}[X_1 = 0] = 1\) with \(N \sim \text{ExtLog}(k,1)\), cf. Example 5.23 or one of its truncations, see Lemma 5.4, violates the technical assumption \(a \mathbb{P}[X_1 = 0] \neq 1\). Obviously, \(p_n = 0\) for all \(n \in \mathbb{N}_0^d \setminus \{0\}\) in these cases.

**Remark 5.10** (Computational speed-up for small support of \(L(X_1)\)). For \(n = (n_1, \ldots, n_d) \in \mathbb{N}_0^d \setminus \{0\}\), the number of terms in (5.4) is \((n_1 + 1) \cdots (n_d + 1) - 1\), which may limit the practical applicability of the recursion to small dimension \(d\). A remarkable speed-up is possible if the support of the distribution of \(X_1\)
is concentrated on just a few points of \( \mathbb{N}_0^d \), let’s us write \( S_X = \{ n \in \mathbb{N}_0^d \setminus \{0\} \mid \mathbb{P}[X_1 = n] > 0 \} \) for this support without the origin of \( \mathbb{N}_0^d \). Then the sum in (5.4) runs over all \( j \in S_X \) satisfying \( j \leq n \), i.e.

\[
j \in S_n(X) := S_X \cap \prod_{i=1}^d \{0, \ldots, n_i\},
\]

and their cardinalities satisfies \(|S_n(X)| \leq \min\{|S_X|, (n_1 + 1) \cdots (n_d + 1) - 1\} \). If \(|S_X| < \infty\), then \(|S_X| \) is an upper bound for the number of terms which doesn’t grow with \( n \). Remark 5.11 below simplifies the computation of the individual terms.

**Remark 5.11 (Choice of \( c_n \)).** While \( c_n = (1, \ldots, 1) \) works in (5.4) in every case, there is a computational advantage in choosing \( c_n \) dependent on \( n \). To illustrate this, let us take the notation of Remark 5.10 and define \( S_{i,n}(X) = \{ j_i \mid (j_1, \ldots, j_d) \in S_n(X) \} \) for every \( i \in \{1, \ldots, d\} \). Since every \( n = (n_1, \ldots, n_d) \in S_X \) has at least one non-zero component, let’s say the \( i \)th one \( n_i \), we can then choose \( c_n = (0, \ldots, 0, 1, 0, \ldots, 0) \) with the 1 at the \( i \)th position, which simplifies \( \langle c_n, j \rangle \) and \( \langle c_n, n \rangle \) to \( j_i \) and \( n_i \), respectively, and allows us to pull out the factor from the other summations in (5.4), i.e.,

\[
\sum_{j \in S_n(X)} \left( a + \frac{b \langle c_n, j \rangle}{\langle c_n, n \rangle} \right) \mathbb{P}[X_1 = j] p_{n-j} = \sum_{i \in S_{i,n}(X)} \left( a + \frac{b l_i}{n_i} \right) \sum_{j_i = l} \mathbb{P}[X_1 = j] p_{n-j}.
\]

**Remark 5.12 (Calculation of \( \mathcal{L}(S_k) \) by convolutions).** If \( k = 0 \), then \( S_k = 0 \), hence \( \mathbb{P}[S_k = n] = 0 \) for all \( n \in \mathbb{N}_0^d \setminus \{0\} \). If \( k = 1 \), then \( S_k = X_1 \). If \( k \geq 2 \), then \( S_k = S_{k-1} + X_k \) and the distribution of \( S_k \) can be calculated recursively by convolution in a numerically stable way, i.e.,

\[
\mathbb{P}[S_k = n] = \sum_{j \in \mathbb{N}_0^d \setminus \{0\}} \mathbb{P}[S_{k-1} = n-j] \mathbb{P}[X_k = j], \quad n \in \mathbb{N}_0^d. \tag{5.5}
\]

Starting with the integer \( k \geq 4 \), there is a more efficient way to calculate the distribution of \( S_k \), similar to the [exponentiation by squaring](https://en.wikipedia.org/wiki/Exponentiation_by_squaring) or the [Russian peasant multiplication](https://en.wikipedia.org/wiki/Russian_peasant_multiplication). Given \( l, m \in \mathbb{N} \), observe that \( S_{l+m} = S_l + S_m' \) with \( S_m' := X_{l+1} + \cdots + X_{l+m} = S_m \). Therefore by convolution,

\[
\mathbb{P}[S_{l+m} = n] = \sum_{j \in \mathbb{N}_0^d \setminus \{0\}} \mathbb{P}[S_l = n-j] \mathbb{P}[S_m = j], \quad n \in \mathbb{N}_0^d. \tag{5.6}
\]

Define now \( l = \lfloor \log_2 k \rfloor \) and let \( k = \sum_{i=0}^{l} b_i 2^i \) with \( b_l = 1 \) and \( b_0, \ldots, b_{l-1} \in \{0, 1\} \) be the binary representation of \( k \). Calculate iteratively via (5.6) the distributions.
of \( S_{2i} = S_{2i-1} + S_{2i-1} \) for \( i \in \{1, 2, \ldots, l\} \), which requires \( l \) convolutions. If \( k = 2^l \), then we are done, otherwise the distribution of \( S_k \) is obtained by using (5.6) to calculate the convolution of the distributions of all those \( S_{2i} \) with \( i \in \{0, 1, \ldots, l\} \), for which \( b_i = 1 \). This requires \( b_0 + \cdots + b_{l-1} \) additional convolutions, so there are \( l + b_0 + \cdots + b_{l-1} \leq 2^l \) altogether. This is numerically more precise than the \( k - 1 \) convolutions via (5.5), because a smaller number of operations for the calculation of the distribution of \( S_k \) and, therefore, a smaller number of rounding errors to machine precision are needed. Furthermore, it can be substantially faster for large \( k \). However, due to the effect already discussed in Remark 5.10 the speed-up might not be as large as the number of convolutions suggests. As an illustration, suppose the \( X_1 \) takes only values in \( \{0, 1, \ldots, \nu\} \) \( d \)-wise independent, \( \nu \geq 1 \). Then there are at most \( (\nu+1)^d \) non-zero summands on the right-hand side of (5.5), but there can be up to \( \min(\{l\nu + 1\}^d, (\nu+1)^d) \) non-zero summands on the right-hand side of (5.6).

Instead of using iterated convolutions as explained in Remark 5.12 it is possible to use a direct recursion based on the following observation (which is well known for powers of formal power series, cf. [55], and goes back to Euler [18, Chapter 4, Section 76], see also Remark 5.24).

**Lemma 5.13.** Let \((X_i)_{i \in \mathbb{N}}\) be a sequence of \( \mathbb{N}_0^d \)-valued, independent, identically distributed random variables. For \( k \in \mathbb{N}_0 \) define \( S_k = \sum_{i=1}^k X_i \), where the empty sum is the zero vector in \( \mathbb{R}^d \). Then, for every \( k \in \mathbb{N}_0 \) and \( n \in \mathbb{N}_0^d \),

\[
\sum_{j \in \mathbb{N}_0^d, j \leq n} ((k+1)j - n) \mathbb{P}[X_1 = j] \mathbb{P}[S_k = n - j] = 0. \tag{5.7}
\]

**Proof.** For \( k = 0 \) we have that \( S_0 \equiv 0 \in \mathbb{N}_0^d \), hence \( \mathbb{P}[S_k = n - j] = 0 \) unless \( n = j \). In this case \((k+1)j - n = 0\), hence (5.7) holds for all \( n \in \mathbb{N}_0^d \).

Now fix \( k \in \mathbb{N} \) and \( n \in \mathbb{N}_0^d \). First note that \( S_{k+1} = S_k + X_{k+1} \), where \( S_k \) and \( X_{k+1} \) are independent. We can rewrite the convolution formula (5.5) in the form

\[
n \mathbb{P}[S_{k+1} = n] = \sum_{j \in \mathbb{N}_0^d, j \leq n} \mathbb{P}[X_{k+1} = j] \mathbb{P}[S_k = n - j]. \tag{5.8}
\]

Furthermore

\[
n \mathbb{P}[S_{k+1} = n] = \mathbb{E}[S_{k+1} \mathbb{1}_{\{S_{k+1} = n\}}] = \sum_{i=1}^{k+1} \mathbb{E}[X_i \mathbb{1}_{\{S_{k+1} = n\}}].
\]

Note that all terms in this sum are equal. Hence, by writing down the expectation,

\[
n \mathbb{P}[S_{k+1} = n] = (k+1) \mathbb{E}[X_{k+1} \mathbb{1}_{\{S_{k+1} = n\}}]
\]

\[
= (k+1) \sum_{j \in \mathbb{N}_0^d, j \leq n} j \mathbb{P}[X_{k+1} = j] \mathbb{P}[S_k = n - j]. \tag{5.9}
\]

Subtracting (5.8) from (5.9) yields (5.7). \( \square \)
Corollary 5.14 (Recursion for $L(S_k)$). Consider a vector $c \in [0, \infty)^d$ such that there exists a unique $m \in \mathbb{N}_0^d$ with $\mathbb{P}[X_1 = m] > 0$ satisfying $\langle c, m \rangle < \langle c, j \rangle$ for all $j \in \mathbb{N}_0^d \setminus \{m\}$ with $\mathbb{P}[X_1 = j] > 0$. Then, for every natural number $k \geq 2$, the distribution of $S_k = X_1 + \cdots + X_k$ can be calculated by

$$
\mathbb{P}[S_k = n] = \begin{cases} 
0 & \text{if } n \in \mathbb{N}_0^d \text{ with } \langle c, n \rangle \leq k\langle c, m \rangle \text{ and } n \neq km, \\
\left(\frac{\mathbb{P}[X_1 = m]}{\mathbb{P}[X_1 = m]}\right)^k & \text{if } n = km,
\end{cases}
$$

and, for every $n \in \mathbb{N}_0^d$ with $\langle c, n \rangle > k\langle c, m \rangle$, via the recursion

$$
\mathbb{P}[S_k = n] = \frac{1}{\langle c, n - km \rangle \mathbb{P}[X_1 = m]} \times \sum_{j \in N_{m,n}} \langle c, (k+1)j - m - n \rangle \mathbb{P}[X_1 = j] \mathbb{P}[S_k = m + n - j]
$$

with summation over

$$N_{m,n} := \{ j \in \mathbb{N}_0^d \mid j \leq m + n \text{ and } \langle c, m \rangle < \langle c, j \rangle \leq \langle c, n - (k-1)m \rangle \text{ and } \mathbb{P}[X_1 = j] > 0 \}.
$$

Remark 5.15 (Applicability of Corollary 5.14).

(a) Equation [5.11] is indeed a recursion, because for $\mathbb{P}[S_k = n]$ only values $\mathbb{P}[S_k = l]$ with $\langle c, l \rangle < \langle c, n \rangle$ are used, as the definition of $N_{m,n}$ shows.

(b) Contrary to the approach given in Remark 5.12, the recursion [5.11] can be numerically unstable, because already in one dimension for $m = 0$ and $n > k + 1$ the term $(k+1)j - n$ changes sign as $j$ runs from 1 to $n$. For an example, see Exercise 5.16 below.

(c) Let us show that there always exists a $c \in \mathbb{N}^d$ with the property required for Corollary 5.14. A vector $c$ with integer components is useful to avoid rounding errors in the computation of the inner product $\langle c, (k+1)j - m - n \rangle$ appearing in the recursion formula [5.11]. Naturally, small components of $c$ are preferred. Note that $\|X_1\|_1$ takes values in $\mathbb{N}_0$, hence there is a smallest $\tilde{m} \in \mathbb{N}_0$ with $\mathbb{P}[\|X_1\|_1 = \tilde{m}] > 0$. Consider the first of the following cases that applies:

(i) If there is just a single $m \in \mathbb{N}_0^d$ with $\|m\|_1 = \tilde{m}$ and $\mathbb{P}[X_1 = m] > 0$ (which is certainly the case when $d = 1$ or $\tilde{m} = 0$), then $m$ and $c := (1, \ldots, 1) \in \mathbb{N}^d$ satisfy the conditions of Corollary 5.14.

(ii) If there exists an $i \in \{1, \ldots, d\}$ such that $m = (0, \ldots, 0, \tilde{m}, 0, \ldots, 0)$ with $\tilde{m}$ at position $i$ satisfies $\mathbb{P}[X_1 = m] > 0$ (which is necessarily the case when $\tilde{m} = 1$), then take $c = (2, \ldots, 2, 1, 2, \ldots, 2)$ with the 1 at position $i$. Then $\langle c, m \rangle = \tilde{m}$ while $\langle c, j \rangle \geq \tilde{m} + 1$ for all $j \in \mathbb{N}_0^d \setminus \{m\}$ satisfying $\|j\|_1 \geq \tilde{m}$.
(iii) Suppose there exist different \( i_1, i_2 \in \{1, \ldots, d\} \) and an \( m = (m_1, \ldots, m_d) \in \mathbb{N}_0^d \) with \( \|m\|_1 = \tilde{m} \) and \( m_i = 0 \) for all \( i \in \{1, \ldots, d\} \setminus \{i_1, i_2\} \) such that \( \mathbb{P}[X_1 = m] > 0 \) (this case always applies for \( d = 2 \)). Among all these possibilities, we take \( i_1 \) and a corresponding \( m \) such that \( \tilde{m}_1 := m_{i_1} \) is maximal (and correspondingly \( \tilde{m}_2 := m_{i_2} = \tilde{m} - \tilde{m}_1 \) is minimal). Note that \( \tilde{m}_2 \geq 1 \), because otherwise case (cii) already applied. Define \( c = (c_1, \ldots, c_d) \) by \( c_{i_1} = \tilde{m}_2 + 1 \) and \( c_{i_2} = c_{i_1} + 1 \) as well as \( c_i = 2c_{i_1} \) for all remaining \( i \in \{1, \ldots, d\} \). To show that \( m \) and this \( c \) satisfies the conditions of Corollary 5.14, consider \( j \in \mathbb{N}_0^d \setminus \{m\} \) with \( \|j\|_1 = \tilde{m} \) and \( \mathbb{P}[X_1 = j] > 0 \), hence \( j \) has at least two non-zero components (otherwise case (cii) already applied). If it has exactly two, then \( \langle c, m \rangle < \langle c, j \rangle \) by the choice of \( i_1 \) and \( m \) as well as \( c_{i_1} < c_{i_2} < c_i \) for all \( i \in \{1, \ldots, d\} \setminus \{i_1, i_2\} \). If \( j \) has at least three non-zero components, then

\[
\langle c, j \rangle \geq c_{i_1}(\tilde{m} - 2) + c_{i_2} + 2c_{i_1} = c_{i_1}\tilde{m}_1 + c_{i_2}\tilde{m}_2 + \tilde{m}_2 + 2 = c_{i_1}\tilde{m}_1 + c_{i_2}\tilde{m}_2 + 2 = \langle c, m \rangle + 2 > \langle c, m \rangle.
\]

If \( j \in \mathbb{N}_0^d \) with \( \|j\|_1 \geq \tilde{m} + 1 \), then

\[
\langle c, j \rangle \geq c_{i_1}(\tilde{m} + 1) = c_{i_1}\tilde{m}_1 + c_{i_2}\tilde{m}_2 + \tilde{m}_2 + 1 = c_{i_1}\tilde{m}_1 + c_{i_2}\tilde{m}_2 + 1 > \langle c, m \rangle.
\]

(iv) We claim that in the remaining cases, we can take \( c = (c_1, \ldots, c_d) \) with \( c_i = (\tilde{m} + 1)^{d-i} + (\tilde{m} + 1)^{d-i} \) for \( i \in \{1, \ldots, d\} \). Within all \( j = (j_1, \ldots, j_d) \in \mathbb{N}_0^d \) with \( \|j\|_1 = \tilde{m} \) and \( \mathbb{P}[X_1 = j] > 0 \) there is a minimal one, which we call \( m \), with respect to lexicographical order. Then this \( m \) also uniquely minimizes \( \langle c, j \rangle \) with respect to all \( j \in \mathbb{N}_0^d \) satisfying \( \|j\|_1 = \tilde{m} \) and \( \mathbb{P}[X_1 = j] > 0 \).

**Exercise 5.16** (Complete cancellation in recursion (5.11)). Fix \( l \in \mathbb{N} \) with \( l \geq 3 \) and a probability distribution on \( \mathbb{N}_0 \) such that \( \mathbb{P}[X_1 = j] > 0 \) for \( j \in \{0, 1, l\} \) and \( \mathbb{P}[X_1 = j] = 0 \) for all other \( j \in \mathbb{N}_0 \). Show that the right-hand side of (5.11) for \( k = l - 1 \) and \( n = 2l - 1 \) contains exactly two non-zero terms of opposite sign, hence complete cancellation occurs and \( \mathbb{P}[S_{l-1} = 2l - 1] = 0 \).

**Proof of Corollary 5.14**. Since we are only interested in the distribution of the partial sums, and due to the choice of \( c \) and \( m \), we may redefine each \( X_i \) on a set of probability zero such that \( \{\langle c, X_i \rangle \leq \langle c, m \rangle\} = \{X_i = m\} \). Define \( X_i' := \langle c, X_i - m \rangle \) for \( i \in \mathbb{N} \). These are i.i.d. and \( [0, \infty) \)-valued random variables. Fix the natural number \( k \geq 2 \). Define \( S_k' = X_1' + \cdots + X_k' \). Then \( 0 \leq S_k' = \langle c, S_k - km \rangle \) and \( \{S_k' = 0\} = \{X_1' = 0, \ldots, X_k' = 0\} = \{X_1 = m, \ldots, X_k = m\} \). Using the i.i.d. assumption, (5.10) follows.
To prove (5.11) for a given \( n \in \mathbb{N}_d \) with \( \langle c, n \rangle > k \langle c, m \rangle \), rewrite (5.7) with \( m + n \) in place of \( n \). Then take the inner product with \( c \) and solve for \( \mathbb{P}[S_k = n] \), which is possible because \( \langle c, n - km \rangle \neq 0 \) and \( \mathbb{P}[X_1 = m] > 0 \) by the choice of \( c \) and \( m \). Furthermore, all remaining terms with \( \langle c, j \rangle \leq \langle c, m \rangle \) on the right-hand side of (5.11) are zero and can be omitted. Since \( \langle c, X_1 \rangle \geq \langle c, m \rangle \), it follows that \( \langle c, S_k \rangle \geq k \langle c, m \rangle \) by the above part of the proof, hence we may skip all terms on the right-hand side of (5.11) with \( \langle c, m + n - j \rangle < k \langle c, m \rangle \). Since \( j \leq m + n \), these are the ones with \( \langle c, j \rangle > \langle c, m + n \rangle - k \langle c, m \rangle = \langle c, n - (k - 1)m \rangle \). This justifies to sum only over \( j \in \mathbb{N}_{m,n} \).

Before we derive Theorem 5.8 from Theorem 5.26(a) below, let us look at several examples and keep the numerical stability for the recursion formula (5.4) in mind.

Example 5.17 (Poisson distribution). If \((q_n)_{n \in \mathbb{N}_0}\) is Poisson(\( \lambda \)) with \( \lambda \geq 0 \), then \( q_0 = e^{-\lambda} \) and

\[
q_n = \frac{\lambda^n}{n!} e^{-\lambda} = \frac{\lambda}{n} q_{n-1}, \quad n \in \mathbb{N},
\]

hence Poisson(\( \lambda \)) is the Panjer(0, \( \lambda \), 0) distribution. Using (4.3), the initial value (5.3) turns into

\[
p_0 = e^{\lambda (\mathbb{P}[X_1 = 0] - 1)}. \tag{5.12}
\]

The recursion formula (5.4) can be simplified to

\[
p_n = \frac{\lambda}{n_i} \sum_{j \in \mathbb{N}_{0}^d} j_i \mathbb{P}[X_1 = j] p_{n-j}, \tag{5.13}
\]

for every \( n = (n_1, \ldots, n_d) \in \mathbb{N}_0^d \setminus \{0\} \), where \( i \in \{1, \ldots, d\} \) is chosen such that \( n_i \neq 0 \). See Remark 5.10 to omit terms in (5.13) with value zero. The recursion (5.13) is numerically stable because only non-negative numbers are multiplied and added.

Example 5.18 (Negative binomial distribution). If \((q_n)_{n \in \mathbb{N}_0}\) is NegBin(\( \alpha, p \)) with parameters \( \alpha > 0 \) and \( p \in [0, 1) \) as specified in (4.43), then \( q_0 = q^\alpha \) and

\[
q_n = \binom{\alpha + n - 1}{n} p^n q^\alpha = \frac{\alpha + n - 1}{n} p q_{n-1}, \quad n \in \mathbb{N},
\]

with \( q := 1 - p \), hence NegBin(\( \alpha, p \)) is the Panjer(\( p, (\alpha - 1)p \), 0) distribution. Using (4.47), the initial value (5.3) turns into

\[
p_0 = \left( \frac{q}{1 - p \mathbb{P}[X_1 = 0]} \right)^\alpha. \tag{5.14}
\]

The recursion formula (5.4) can be simplified to

\[
p_n = \frac{p}{n_i (1 - p \mathbb{P}[X_1 = 0])} \sum_{j \in \mathbb{N}_{0}^d} \left( \alpha j_i + n_i - j_i \right) \mathbb{P}[X_1 = j] p_{n-j} \tag{5.15}
\]
for every \( n = (n_1, \ldots, n_d) \in \mathbb{N}_0^d \setminus \{0\} \), where \( i \in \{1, \ldots, d\} \) is chosen such that \( n_i \neq 0 \). See Remark 5.10 for the possibility to omit terms in (5.15) with value zero. The recursion (5.15) is numerically stable because \( n_i - j_i \in \mathbb{N}_0 \) (this requires proper programming, \( \alpha_j \) has to be added afterwards) and otherwise only non-negative numbers are multiplied and added to calculate the sum.

**Remark** 5.19 (Calculation of the initial value). To apply the \( d \)-dimensional extended Panjer recursion (5.4), the probability \( p_0 \) of a loss of zero is needed as starting value, see (5.3). If \( N \sim \text{Poisson}(\lambda) \) with \( \lambda \geq 0 \), then \( p_0 \) is given by (5.12). If \( N \sim \text{NegBin}(\alpha, \beta) \) with \( \alpha > 0 \) and \( \beta \in (0,1) \), then \( p_0 \) is given by (5.14). When modeling large portfolios with the collective risk model (4.51) using one of these two claim number distributions, it can happen for large \( \lambda \) or \( \alpha \), respectively, that \( p_0 \) is so small that it can only be represented as zero on a computer (numerical underflow). The recursion (5.4) then produces \( p_n = 0 \) for all \( n \in \mathbb{N}_0^d \setminus \{0\} \), which is clearly wrong. The standard solution, cf. [33, Section 6.6.2], is to perform Panjer’s recursion with the reduced parameter \( \lambda' := \lambda/2^n \) (resp. \( \alpha' := \alpha/2^n \)) instead, where \( n \in \mathbb{N} \) is chosen such that the new starting value \( p_0 \) is properly representable on the computer. Afterwards, \( n \) iterative and numerically stable convolutions are needed to calculate the original probability distribution. This approach works because for independent \( N_1, \ldots, N_{2^n} \sim \text{Poisson}(\lambda/2^n) \), we have that \( N = N_1 + \cdots + N_{2^n} \sim \text{Poisson}(\lambda) \) by Lemma 3.2 similarly for the negative binomial distribution, see Lemma 4.24. In general, this works for claim number distributions closed under convolutions.

**Example** 5.20 (Binomial distribution). Let \( (q_n)_{n \in \mathbb{N}_0} \) denote the binomial distribution \( \text{Bin}(m, p) \) with success probability \( p \in [0, 1) \) and number of trials \( m \in \mathbb{N} \). Let \( q := 1 - p \) denote the failure probability. Then, for every \( n \in \mathbb{N} \),

\[
q_n = \binom{m}{n} p^n q^{m-n} = \frac{m-n+1}{n} \binom{m}{n-1} p^{n-1} q^{m-(n-1)} = \left( -\frac{p}{q} + \frac{m+1}{q} \frac{p}{n} \right) q_{n-1},
\]

hence \( \text{Bin}(m, p) \) is the Panjer\((-p/q, (m+1)p/q, 0)\) distribution. The recursion factor \( a + b/n \) is zero for \( n = m + 1 \), giving \( q_n = 0 \) for \( n \geq m + 1 \) as expected. Using (4.29), the initial value (5.3) turns into

\[
p_0 = (1 + p(\mathbb{P}[X_1 = 0] - 1))^m. \tag{5.16}
\]

Consider Panjer’s recursion formula (5.4) for \( n = (n_1, \ldots, n_d) \in \mathbb{N}_0^d \setminus \{0\} \) with \( n_1 \geq m + 2 \) and \( n_2 = \cdots = n_d = 0 \). Without loss of generality we can take \( c_n = (1, 0, \ldots, 0) \). Then the term

\[
a + \frac{b(c_n, j)}{c_n, n} = -\frac{p}{q} \left( 1 - \frac{m+1}{n_1} j_1 \right)
\]

changes sign as \( j = (j_1, 0, \ldots, 0) \) varies between \((1, 0, \ldots, 0)\) and \((n_1, 0, \ldots, 0)\). Therefore, the recursion might not be numerically stable because cancellations...
can occur. The problem with numerical underflow during the calculation of the initial value \( p_0 \) given in (5.16) can also occur for large \( m \), cf. Remark 5.19. Since

\[
\varphi_S(s) = \varphi_N(\varphi_X(s)) = (q + p\varphi_X(s))^m = \prod_{k=1}^{l} (q + p\varphi_X(s))^{2^k}
\]

at least for all \( s \in \mathbb{C}^d \) with \( \|s\|_\infty \leq 1 \), where \( m = \sum_{k=0}^{l} b_k 2^k \) with \( b_1, \ldots, b_{l-1} \in \{0,1\} \), \( b_1 = 1 \) and \( l = \lfloor \log_2 m \rfloor \) denotes the binary representation of \( m \), we see that the distribution \((p_n)_{n \in \mathbb{N}_0} \) of \( S \) can be computed in a numerically stable way with \( b_0 + \cdots + b_{l-1} + l \leq 2l \) convolutions, see Remark 5.12.

Example 5.21 (Logarithmic distribution). If \((q_n)_{n \in \mathbb{N}_0} \) is \( \log(p) \) with \( p \in [0,1) \), cf. Example 4.4, then \( q_0 = 0 \), \( q_1 = 1/c(p) \) with \( c(p) \) defined by (4.5) and

\[
q_n = p^{n-1}/c(p)n = \frac{n-1}{n} q_{n-1}, \quad \text{for } n \in \mathbb{N}, n \geq 2,
\]

hence \( \log(p) \) is the Panjer\((p,-p,1)\) distribution. Using (4.6), the initial value (5.3) turns into

\[
p_0 = \mathbb{P}[X_1 = 0] \frac{c(p\mathbb{P}[X_1 = 0])}{c(p)} \quad (5.17)
\]

The recursion formula (5.4) simplifies to

\[
p_n = \frac{1}{1 - p \mathbb{P}[X_1 = 0]} \left( \frac{\mathbb{P}[X_1 = n]}{c(p)} + \frac{p}{n_i} \sum_{j \in \mathbb{N}_0^d \cap \{0\} \setminus j < n_i} (n_i - j_i) \mathbb{P}[X_1 = j] p_{n-j} \right) \quad (5.18)
\]

for every \( n = (n_1, \ldots, n_d) \in \mathbb{N}_0^d \setminus \{0\} \), where \( i \in \{1, \ldots, d\} \) is chosen such that \( n_i \neq 0 \). See Remark 5.10 about the possibility to omit further terms in (5.18) with value zero. The recursion (5.18) is numerically stable because \( n_i - j_i \in \mathbb{N}_0 \) and otherwise only non-negative numbers are multiplied and added inside the parenthesis to calculate the sum. For \( p = 0 \), the recursion (5.18) simplifies dramatically to \( p_n = \mathbb{P}[X_1 = n] \) for all \( n \in \mathbb{N}_0^d \setminus \{0\} \).

Example 5.22 (Extended negative binomial distribution). For parameters \( k \in \mathbb{N} \), \( \alpha \in (-k,0) \setminus \{-1,-2,\ldots,-(k-1)\} \) and \( p \in (0,1] \), define \( q_0 = \cdots = q_{k-1} = 0 \) and, using the abbreviation \( q = 1 - p \),

\[
q_n = \frac{\left(\frac{\alpha+n-1}{n}\right) p^n}{q^{-\alpha} - \sum_{j=0}^{k-1} \left(\frac{\alpha+j-1}{j}\right) p^j}, \quad \text{for } n \geq k. \quad (5.19)
\]

We will verify below that (5.19) is a well-defined probability distribution \((q_n)_{n \in \mathbb{N}_0}\), called extended negative binomial distribution, notation \( \text{ExtNegBin}(\alpha,k,p) \).

First note that the \( \alpha \)-truncation of the negative binomial distribution defined in (4.43) gives the same formula, however with \( \alpha > 0 \) and valid for \( p \in [0,1) \). A short calculation shows that the \( \alpha \)-truncation of an \( \text{ExtNegBin}(\alpha,l,p) \) distribution
with \( l \in \{1, \ldots, k-1\} \), \( \alpha \in (-l, -l+1) \) and \( p \in (0, 1) \) is also given by (5.19). Hence, for every \( k \in \mathbb{N} \), the formula (5.19) defines a probability distribution for all \( \alpha \in (-k, \infty) \setminus \{0, -1, -2, \ldots\} \) and \( p \in (0, 1) \). If \( -\alpha \in \mathbb{N}_0 \), then \((a+n^{-1}) = 0 \) for all \( n \in \mathbb{N} \) with \( n \geq 1 - \alpha \) and the binomial coefficient is of different sign for the even and the odd \( n \in \{1, \ldots, -\alpha\} \), hence (5.19) cannot define an interesting probability distribution in this case.

To verify that (5.19) defines a probability distribution, note that, for every \( n \in \mathbb{N}_0 \),

\[
\binom{\alpha + n - 1}{n} = \frac{1}{n!} \prod_{j=1}^{n} (\alpha + n - j)
= \frac{(-1)^n}{n!} \prod_{l=0}^{n-1} (-\alpha - l) = (-1)^n \binom{-\alpha}{n},
\]

and, for all integers \( n \geq k \),

\[
\binom{\alpha + n - 1}{n} = \left( \prod_{j=1}^{k} \frac{\alpha + j - 1}{j} \right) \prod_{j=k+1}^{\frac{-\alpha + n - 1}{j}} \frac{\alpha - j + 1}{j} > 0
\]
has the same sign. Using \( \log(1 + x) \leq x \) for \( x > -1 \) and noting that \( -\alpha - 1 < 0 \),

\[
\log \prod_{j=k+1}^{n} \left( 1 + \frac{\alpha - 1}{j} \right) \leq \sum_{j=k+1}^{n} \frac{\alpha - 1}{j} \leq (\alpha - 1) \int_{k+1}^{n+1} \frac{dx}{x} = \log \left( \frac{n+1}{k+1} \right) ^{\alpha - 1}
\]
for all integers \( n \geq k \). Therefore,

\[
\sum_{n=k}^{\infty} \left| \binom{\alpha + n - 1}{n} \right| \leq \left| \prod_{j=1}^{k} \frac{\alpha + j - 1}{j} \right| \sum_{n=k}^{\infty} \left( \frac{k+1}{n+1} \right) ^{1-\alpha} < \infty
\]
by the integral test for convergence because \( 1 - \alpha > 1 \) and

\[
\sum_{n=k}^{\infty} \frac{1}{(n+1)^{1-\alpha}} \leq \int_{k}^{\infty} \frac{dx}{x^{1-\alpha}} = - \frac{1}{\alpha k^{-\alpha}} < \infty.
\]

Using (5.20), we see that the binomial series

\[
(1 + x)^{-\alpha} = \sum_{n \in \mathbb{N}_0} \binom{-\alpha}{n} x^n
\]
converges absolutely for all \( x \in \mathbb{C} \) with \( |x| \leq 1 \); for \( x = -p \) we see that

\[
\sum_{n \in \mathbb{N}_0} \binom{\alpha + n - 1}{n} p^n = \sum_{n \in \mathbb{N}_0} \binom{-\alpha}{n} (-p)^n = (1 - p)^{-\alpha} = q^{-\alpha}.
\]
We conclude that the nominators in (5.19) are all of the same sign and, by (5.22),
the denominator is the sum of these. Hence $q_n > 0$ for all integers $n \geq k$ and
$\sum_{n=k}^{\infty} q_n = 1$.

Using the first equality in (5.20), we see that, for every $n \geq k + 1$,
\[
\binom{\alpha + n - 1}{n} p^n = \frac{\alpha + n - 1}{n} p \cdot \frac{p^{n-1}}{(n-1)!} \prod_{j=1}^{n-1} (\alpha + n - 1 - j) \\
= \left(1 + \frac{\alpha - 1}{n}\right) \frac{p}{1} \cdot \frac{(\alpha + n - 2)}{(n-1)} p^{n-1},
\]

hence ExtNegBin$(\alpha, k, p)$ is the Panjer$(p, (\alpha - 1)p, k)$ distribution. Consider
Panjer’s recursion formula (5.4) for $n = (n_1, \ldots, n_d) \in \mathbb{N}_0^d \setminus \{0\}$ with $n_1 > 1 - \alpha$
and $n_2 = \cdots = n_d = 0$. Without loss of generality we can take $c_n = (1, 0, \ldots, 0)$.
Then the term
\[
a + \frac{b(c_{n,j})}{c_{n,n}} = \left(1 + \frac{\alpha - 1}{n_1}\right) p
\]
changes sign as $j = (j_1, 0, \ldots, 0)$ varies between $(1, 0, \ldots, 0)$ and $(n_1, 0, \ldots, 0)$. Therefore, the recursion can be numerically unstable due to cancellations, see Remark 5.28.

To calculate the probability-generating function of a random variable $N \sim$ ExtNegBin$(\alpha, k, p)$, note that by (5.22)
\[
\sum_{n=0}^{\infty} \binom{\alpha + n - 1}{n} p^n s^n = (1 - ps)^{-\alpha} \quad \text{for } |s| \leq \frac{1}{p},
\]
therefore
\[
\varphi_N(s) = \sum_{n=0}^{\infty} q_n s^n = \frac{(1 - ps)^{-\alpha} - \sum_{j=0}^{k-1} \binom{\alpha + j - 1}{j} (ps)^j}{q^{-\alpha} - \sum_{j=0}^{k-1} \binom{\alpha + j - 1}{j} p^j} \quad \text{for } |s| \leq \frac{1}{p}. \quad (5.23)
\]

For $k = 1$, hence $\alpha \in (-1, 0)$, this simplifies to
\[
\varphi_N(s) = \frac{1 - (1 - ps)^{-\alpha}}{1 - q^{-\alpha}} \quad \text{for } |s| \leq \frac{1}{p}. \quad (5.24)
\]

**Example 5.23 (Extended logarithmic distribution).** Assume that $(q_n)_{n \in \mathbb{N}_0}$ is an
extended logarithmic distribution, notation ExtLog$(k, p)$, with parameters $k \in \mathbb{N}$, $k \geq 2$, and $p \in (0, 1]$, which means that $q_0 = \cdots = q_{k-1} = 0$ and
\[
q_n = \frac{\binom{n}{k}^{-1} p^n}{\sum_{l=k}^{\infty} \binom{l}{k}^{-1} p^l} \quad \text{for } n \geq k. \quad (5.25)
\]

Since, for every $m \in \mathbb{N}$ with $m \geq k$,
\[
\sum_{l=k}^{m} \frac{1}{l \binom{l}{k}} \leq \sum_{l=k}^{m} \frac{k!}{l(l-1)} = k! \sum_{l=k}^{m} \frac{1}{l-1} = k! \left(\frac{1}{k-1} - \frac{1}{m}\right) \leq \frac{k!}{k - 1},
\]

81
the extended logarithmic distribution is well defined for every $p \in (0,1]$. For $n \geq k + 1$ we have

$$\binom{n}{k} = \frac{n}{n-k} \binom{n-1}{k},$$

which yields

$$q_n = \frac{n-k}{n} p q_{n-1} = \left( p - \frac{kp}{n} \right) q_{n-1},$$

hence ExtLog$(k,p)$ is the Panjer$(p,-kp,k)$ distribution. Consider Panjer’s recursion formula (5.4) for $n = (n_1,\ldots,n_d) \in \mathbb{N}_0^d \setminus \{0\}$ with $n_1 \geq k + 1$ and $n_2 = \cdots = n_d = 0$. Without loss of generality we can take $c_n = (1,0,\ldots,0)$. Then the term

$$a + b\langle c_n,j \rangle \langle c_n,n \rangle = p \left( 1 - \frac{kj_1}{n_1} \right)$$

changes sign as $j = (j_1,0,\ldots,0)$ varies between $(1,0,\ldots,0)$ and $(n_1,0,\ldots,0)$. Therefore, the recursion might not be numerically stable because cancellations might occur; see Subsection 5.4 and [21, Section 5.2] for a solution of this problem.

We remark that a closed-form expression for the denominator in (5.25) is given by [21, Lemma 2.1], which makes it possible to express the probability-generating function also in closed form, cf. [21, (2.7)]

**Remark 5.24.** As a historical remark, we mention that the one-dimensional Panjer recursion for binomial, negative binomial, and extended negative binomial claim number distributions is contained in a much older result: For $\alpha \in \mathbb{R}$ and a power series $f(s) = \sum_{k=0}^{\infty} a_k s^k$ with $a_0 \neq 0$, the coefficients $(b_n)_{n \in \mathbb{N}_0}$ of the power series $f^{-\alpha}(s)$ satisfy the recursion

$$b_n = \frac{1}{n a_0} \sum_{k=1}^{n} ((1-\alpha)k-n)a_k b_{n-k}, \quad n \in \mathbb{N}. \quad (5.26)$$

Gould [25] has traced this remarkable, often rediscovered recurrence back to Euler [18, Chapter 4, Section 76]. Using the probability-generating functions of the above distributions and $\varphi_S = \varphi_N \circ \varphi_{X_1}$, the formula (5.26) applied to $f(s) = q + p \varphi_{X_1}(s)$ or $f(s) = 1 - p \varphi_{X_1}(s)$, respectively, yields recursions which indeed agree with the respective Panjer recursions.

**Exercise 5.25.** Use (4.29), (4.47) and (5.23) to verify the last statement in Remark 5.24.

### 5.2 A Generalisation of the Multivariate Panjer Recursion

The multivariate extended Panjer recursion in Theorem 5.8 is a special case of part (a) of the following theorem, which combines [21, Theorem 4.5] with the multivariate idea in [49, Theorem 1] and is of independent interest for questions of numerical stability, see Subsections 5.3 and 5.4 below.
Theorem 5.26. Fix \( l \in \mathbb{N} \). Let \((q_n)_{n \in \mathbb{N}_0}\) and \((\tilde{q}_{i,n})_{n \in \mathbb{N}_0}\) denote the probability distributions of the \(\mathbb{N}_0\)-valued random variables \(N\) and \(\tilde{N}_i\) for \(i \in \{1, \ldots, l\}\), where \((N, \tilde{N}_1, \ldots, \tilde{N}_l)\) is independent of the \(\mathbb{N}_0^l\)-valued i. i. d. sequence \((X_n)_{n \in \mathbb{N}}\). Let \((p_n)_{n \in \mathbb{N}_0^d}\) and \((\tilde{p}_{i,n})_{n \in \mathbb{N}_0^d}\) denote the probability distributions of the random sums \(S = X_1 + \cdots + X_N\) and \(\tilde{S}^{(i)} = X_1 + \cdots + X_{\tilde{N}_i}\) for \(i \in \{1, \ldots, l\}\), respectively.

(a) Assume\(^{22}\) that there exist \(k \in \mathbb{N}_0\) and \(a_1, \ldots, a_l, b_1, \ldots, b_l \in \mathbb{R}\) such that
\[
q_n = \sum_{i=1}^{l} \left( a_i + \frac{b_i}{n} \right) \tilde{q}_{i,n-i} \quad \text{for all } n \in \mathbb{N} \text{ with } n \geq k + l
\]  
(5.27)
and all probabilities not used on the right-hand side of \((5.27)\) are zero, i.e.
\[
\tilde{q}_{i,0} = \cdots = \tilde{q}_{i,k+l-i-1} = 0 \quad \text{for all } i \in \{1, \ldots, \min(l, k + l - 1)\}.
\]  
(5.28)
Then, for every \(n \in \mathbb{N}_0^l \setminus \{0\}\) and \(c_n \in \mathbb{R}^d\) with \((c_n, n) \neq 0\),
\[
p_n = \sum_{j=1}^{k+l-1} p_j[q_j + \sum_{j=1}^{l} \sum_{j \in \mathbb{N}_0^d \setminus \{0\}}} \frac{b_i(c_n, j)}{i(c_n, n)} p[S_i = j] \tilde{p}_{i,n-j},
\]  
(5.29)
and \(p_0\) is given by \((5.3)\).

(b) Assume that there exist \(v_1, \ldots, v_l \in [0, 1]\) with \(v_1 + \cdots + v_l \leq 1\) such that
\[
q_n = \sum_{i=1}^{l} v_i \tilde{q}_{i,n} \quad \text{for all } n \in \mathbb{N}_0\]  
(5.30)
Then \(p_n = \sum_{i=1}^{l} v_i \tilde{p}_{i,n}\) for all \(n \in \mathbb{N}_0^l \setminus \{0\}\), which implies the \((a)\) case with \(l = 1\) and \(\tilde{q}_{1,n} = q_n\) for all \(n \in \mathbb{N}_0\), which implies \((5.28)\) and solving \((5.29)\) for \(p_n\) yields \((5.4)\).

Proof of Theorem 5.26 (a) We extend a standard proof (cf. \[39\] Theorem 3.3.9) for the case \(k = 0\) and \(l = 1\) with the idea from \[49\] for the \(d\)-dimensional setting.

To prove the representation for the initial value given in \((5.3)\), note that \(\{S = 0, N = 0\} = \{N = 0\}\). Hence
\[
p_0 = \mathbb{P}[S = 0] = \mathbb{P}[S = 0, N = 0] + \mathbb{P}[S = 0, N \geq 1].
\]
\[
= \mathbb{P}[N=0] = q_0 = 0 \text{ if } \mathbb{P}[X_1 = 0] = 0
\]
If \(\mathbb{P}[X_1 = 0] > 0\), then we use independence of \(N\) and \(X_n\) for all \(n \in \mathbb{N}_0\) as well as the i. i. d. assumption for this sequence to obtain
\[
p_0 = q_0 + \sum_{n \in \mathbb{N}} \frac{\mathbb{P}[S = 0|N = n]}{\mathbb{P}[N = n]} \mathbb{P}[X_1 = 0, \ldots, X_n = 0] = \mathbb{E}[\left(\mathbb{P}[X_1 = 0]\right)^n].
\]
\(^{22}\)In these lecture notes, we only apply this case with \(l = 1\).
We now prove \[5.29\] for fixed \(n \in \mathbb{N}_0^d \setminus \{0\}\) and \(c \in \mathbb{R}^d\) satisfying \(\langle c, n \rangle \neq 0\). For this we need a preparation. Fix \(i \in \{1, \ldots, l\}\). For every \(m \in \mathbb{N}\) with \(m \geq i\), we use the representations \(S_m = X_1 + \cdots + X_m = S_{m-i} + S_{i,m}\) with \(S_{i,m} := X_{m-i+1} + \cdots + X_m\) and independent and identically distributed \(X_1, \ldots, X_m\). If \(\mathbb{P}[S_m = n] > 0\), then we obtain that

\[
\langle c, n \rangle = \mathbb{E}[\langle c, S_m \rangle \mid S_m = n] = \sum_{j=1}^{m} \mathbb{E}[\langle c, X_j \rangle \mid S_m = n]
\]

\[
= m \mathbb{E}[\langle c, X_m \rangle \mid S_m = n] = \frac{m}{i} \mathbb{E}[\langle c, S_{i,m} \rangle \mid S_m = n],
\]

to obtain that

\[
\left( a_i + \frac{b_i}{m} \right) = \mathbb{E}\left[ a_i + \frac{b_i \langle c, S_{i,m} \rangle}{i \langle c, n \rangle} \mid S_m = n \right]
\]

\[
= \sum_{j \in \mathbb{N}_0^d} \left( a_i + \frac{b_i \langle c, j \rangle}{i \langle c, n \rangle} \right) \mathbb{P}[S_{i,m} = j \mid S_m = n].
\]

\[
(5.31)
\]

For every \(m \geq i\) we know that \(S_{m-i}\) and \(S_{i,m}\) are independent, hence

\[
\mathbb{P}[S_{i,m} = j, S_m = n] = \mathbb{P}[S_{i,m} = j] \mathbb{P}[S_{m-i} = n - j] = \mathbb{P}[S_{i,m} = j] \mathbb{P}[S_{m-i} = n - j].
\]

\[
(5.32)
\]

We now rewrite \(p_n = \mathbb{P}[S = n]\) using \(5.27\) as follows

\[
p_n = \sum_{m=1}^{\infty} \sum_{m>0} \mathbb{P}[S_m = n | N = m] \mathbb{P}[N = m]
\]

\[
= \sum_{m=1}^{\infty} \mathbb{P}[S_m = n] q_m + \sum_{m=k+1}^{\infty} \sum_{i=1}^{l} \left( a_i + \frac{b_i}{m} \right) \mathbb{P}[S_m = n] \bar{q}_{i,m-i}
\]

\[
= \sum_{i=1}^{l} \sum_{j \in \mathbb{N}_0^d} \left( a_i + \frac{b_i \langle c, j \rangle}{i \langle c, n \rangle} \right) \mathbb{P}[S_i = j] \mathbb{P}[S_{m-i} = n - j] \bar{q}_{i,m-i}
\]

\[
= \sum_{i=1}^{l} \sum_{j \leq n} \left( a_i + \frac{b_i \langle c, j \rangle}{i \langle c, n \rangle} \right) \mathbb{P}[S_i = j] \sum_{m=k+1}^{\infty} \mathbb{P}[S_{m-i} = n - j] \bar{q}_{i,m-i},
\]

\[
(*)
\]

where the rearrangement from the first to the second line is admissible, because the series in the second line converge for every \(i \in \{1, \ldots, l\}\) and \(j \in \{0, \ldots, n\}\).
Using (5.28), the index shift \( m - i \sim m \), and similar arguments as for (5.33), we get for these series

\[
(*) = \sum_{m=i}^{\infty} \mathbb{P}[S_m - i = n - j] \tilde{q}_{i,m-i}
\]

\[
= \sum_{m=0}^{\infty} \mathbb{P}[S_m = n - j, \tilde{N}_i = m] = \mathbb{P}[\tilde{S}^{(i)} = n - j] = \tilde{p}_{i,n-j}.
\]

Substituting \((**)\) into \((*)\) and this result into (5.33) gives (5.29).

(b) Modifying the calculation in (5.33) using independence of \( \{S_m = n\} \) and \( \{N = m\} \) and the formula \( \mathbb{P}[N = m] = \sum_{i=1}^{l} \nu_i \mathbb{P}[\tilde{N}_i = m] \) for \( m \in \mathbb{N} \), we obtain

\[
p_n = \sum_{m=1}^{\infty} \frac{\mathbb{P}[S_m = n, N = m]}{\mathbb{P}[S_m = n] \mathbb{P}[N = m]} = \sum_{i=1}^{l} \nu_i \sum_{m=1}^{\infty} \frac{\mathbb{P}[S_m = n] \mathbb{P}[\tilde{N}_i = m]}{\mathbb{P}[S_m = n]} = \tilde{p}_{i,n}
\]

for every \( n \in \mathbb{N}_0^d \setminus \{0\} \).

The following corollary of Theorem 5.26[b] is useful, when only a \( k \)-truncation of a probability distribution is a Panjer \( (a,b,k) \) distribution. It is the multivariate extension of [21 Corollary 4.7].

**Corollary 5.27.** Assume that \((q_n)_{n \in \mathbb{N}_0}\) has mass at or above \( k \in \mathbb{N} \) and that \((\tilde{q}_n)_{n \in \mathbb{N}_0}\) denotes its \( k \)-truncated probability distribution according to Definition 5.3. Assume that \( N \) respectively \( \tilde{N} \) have these distributions, and that \( S = X_1 + \cdots + X_N \) and \( \tilde{S} = X_1 + \cdots + X_{\tilde{N}} \) are the corresponding random sums with distributions \((p_n)_{n \in \mathbb{N}_0^d}\) and \((\tilde{p}_n)_{n \in \mathbb{N}_0^d}\). Then \( p_0 \) is given by (5.3) and

\[
p_n = \sum_{i=1}^{k-1} \mathbb{P}[S_i = n] q_i + \left( 1 - \sum_{j=0}^{k-1} q_j \right) \tilde{p}_n, \quad n \in \mathbb{N}_0^d \setminus \{0\}.
\]  

**(5.34)**

**Proof.** Apply Theorem 5.26[b] with \( l = k \), \( \nu_i = q_i \) and \( \tilde{q}_{i,i} = 1 \) for \( i \in \{1, \ldots, k-1\} \), \( \nu_k = 1 - (q_0 + \cdots + q_{k-1}) \), \( \tilde{q}_{k,n} = \tilde{q}_n \) for all \( n \geq k \), and all other \( \tilde{q}_{i,n} = 0 \).

5.3 **Numerically Stable Algorithm for ExtNegBin**

**Remark 5.28.** As noticed in Example 5.22, the Panjer algorithm for the extended negative binomial distribution can be numerically unstable due to cancellations. To show that this is a real danger, let us consider the following example. Take \( k \in \mathbb{N} \) and \( \varepsilon, p \in (0,1) \), define \( \alpha = -k + \varepsilon \) and let \((q_n)_{n \in \mathbb{N}_0}\) denote the distribution of \( N \sim \text{ExtNegBin}(\alpha, k, p) \) given by (5.19). Choose \( l \in \mathbb{N} \) with \( l \geq 3 \) and \( \mathbb{P}[X_1 = 1] = \mathbb{P}[X_1 = l] = 1/2 \) as one-dimensional loss distribution. Note that

\[
p_k = \mathbb{P}[N = k, X_1 = \cdots = X_k = 1] = \frac{q_k}{2^k}
\]

85
and
\[ p_{k+l-1} = \sum_{j=1}^{k} \mathbb{P}[N = k, X_j = l, X_i = 1 \text{ for all } i \in \{1, \ldots, k\} \setminus \{j\}] \]
\[ + \mathbb{P}[N = k + l - 1, X_1 = \cdots = X_{k+l-1} = 1] \]
\[ = \frac{kq_k}{2^k} + \frac{q_{k+l-1}}{2^{k+l-1}}. \]

Recall from Example 5.22 that the frequency distribution ExtNegBin(\(\alpha, k, p\)) is the Panjer(\(p, p(\alpha - 1), k\)) distribution. Note that \(S_k\) takes values in the set \(\{k + j(l - 1) \mid j = 0, \ldots, k\}\), which does not contain \(k + l\), hence Panjer recursion formula (5.4) for \(p_{k+l}\) reduces to
\[ p_{k+l} = \sum_{j=1}^{k+l} p(1 + \frac{\alpha - 1}{k + l} j) \mathbb{P}[X_1 = j] p_{k+l-j}. \]

Since \(\mathbb{P}[X_1 = j] \neq 0\) only for \(j \in \{1, l\}\), this simplifies to two summands, i.e.,
\[ p_{k+l} = p\left(1 + \frac{\alpha - 1}{k + l} \right) p_{k+l-1} + p\left(1 + \frac{\alpha - 1}{k + l} \right) \frac{p_k}{2} \]
\[ = p\frac{k(l-1) + \varepsilon k}{k+l} \left(\frac{q_k}{2^{k+1}} + \frac{q_{k+l-1}}{2^{k+l}}\right) - p\frac{k(l-1) - \varepsilon l}{k+l} \frac{q_k}{2^{k+1}}, \]
hence severe cancellation occurs for \(p_{k+l}\) when \(\varepsilon\) is small and \(q_{k+l-1} \ll 2^{l-1}kq_k\).
For example, the values \(\varepsilon = 10^{-4}, k = 1, l = 5\) and \(p = 9/10\) give
\[ p_6 \approx 0.14999262 - 0.14997009 = 0.00002253, \]

hence we lose four significant digits in this case.

Following [21, Section 5.1], we now develop a numerically stable algorithm to compute the distribution of \((p_n)_{n \in \mathbb{N}_0}\) of \(S = X_1 + \cdots + X_N\), when \(N\) has an extended negative binomial distribution. The main ingredient is the following corollary of Theorem 5.26 [1] for the case \(l = 1\) (we will omit the index 1 for simplicity).

**Corollary 5.29.** For the parameters \(k \in \mathbb{N}_0, \alpha \in (-k, -k+1)\) and \(p \in (0, 1]\), with \(p \neq 1\) for \(k = 0\), let \((q_n)_{n \in \mathbb{N}_0}\) denote the ExtNegBin(\(\alpha - 1, k+1, p\)) distribution and \((\tilde{q}_n)_{n \in \mathbb{N}_0}\) the ExtNegBin(\(\alpha, k, p\)) distribution, where ExtNegBin(\(\alpha, 0, p\)) stands for the negative binomial distribution NegBin(\(\alpha, p\)). Then (5.27) holds with \(l = 1\) and \(\tilde{q}_{1,n} = \tilde{q}_n\) for \(n \geq k + 1\). The constants are given by \(a = 0\) and
\[ b = (\alpha - 1)p \frac{q^{-a} - \sum_{j=0}^{k-1} \binom{\alpha + j - 1}{j} p^j}{q^{-a} - \sum_{j=0}^{k+2} \binom{\alpha + j - 2}{j} p^j}, \tag{5.35} \]
hence (5.29) simplifies to the numerically stable weighted convolution
\[ p_n = \frac{b}{n_i} \sum_{\substack{j \in \mathbb{N}_0^i \\ j \leq n, j_i > 0}} j_i \mathbb{P}[X_1 = j] \tilde{p}_{n-j}, \tag{5.36} \]

86
for every \( n = (n_1, \ldots, n_d) \in \mathbb{N}_0^d \setminus \{0\} \), where \( i \in \{1, \ldots, d\} \) is chosen such that \( n_i \neq 0 \). The initial value \( p_0 \) is given by (5.3) with probability-generating function from (5.23) with parameters \( \alpha \) and \( k \) replaced by \( \alpha - 1 \) and \( k + 1 \), respectively.

Proof. Using (5.19), we see that, for every \( n \geq k + 1 \),

\[
\binom{(\alpha - 1) + n - 1}{n} p^n = \binom{(\alpha - 1) p}{n} \binom{(\alpha - 1)(n - 1) - 1}{n - 1} p^{n-1},
\]

hence \( q_n = b q_{n-1} / n \) and Theorem 5.26(a) is applicable.

The case \( k = 0, p = 1 \) is excluded in the preceding corollary. We cannot reduce the calculation for a claim number \( N \sim \text{ExtNegBin}(\alpha - 1, k + 1, p) \) to the one for \( N \sim \text{ExtNegBin}(\alpha, k, p) \) in this case, because the negative binomial distribution is not defined for \( p = 1 \). However, a suitable limit \( p \searrow 1 \) gives the following numerically stable procedure.

**Lemma 5.30 (Stable recursion for ExtNegBin(\( \alpha - 1, 1, 1 \))).** For \( \alpha \in (0,1) \) consider a claim number \( N \sim \text{ExtNegBin}(\alpha - 1, 1, 1) \). Then the distribution \( (p_n)_{n \in \mathbb{N}_0^d} \) of the random sum \( S = X_1 + \cdots + X_N \) can be calculated by \( p_0 = 1 - (\mathbb{P}[X_1 \geq 1])^{1-\alpha} \) and

\[
p_n = \begin{cases} \frac{1-\alpha}{n_i} \sum_{j \in \mathbb{N}_0^d, 0 < j \leq n} j_i \mathbb{P}[X_1 = j] r_{n-j} & \text{if } \mathbb{P}[X_1 \geq 1] > 0, \\ 0 & \text{if } \mathbb{P}[X_1 \geq 1] = 0, \end{cases}
\]

for every \( n = (n_1, \ldots, n_d) \in \mathbb{N}_0^d \setminus \{0\} \), where \( i \in \{1, \ldots, d\} \) is chosen such that \( n_i \neq 0 \). In the case \( \mathbb{P}[X_1 \geq 1] > 0 \) the non-negative sequence \( (r_n)_{n \in \mathbb{N}_0^d} \) is defined by \( r_0 = (\mathbb{P}[X_1 \geq 1])^{-\alpha} \) and recursively in a numerically stable way by

\[
r_n = \frac{1}{n_i \mathbb{P}[X_1 \geq 1]} \sum_{j \in \mathbb{N}_0^d, 0 < j \leq n} (\alpha j_i + n_i - j_i) \mathbb{P}[X_1 = j] r_{n-j}
\]

for every \( n = (n_1, \ldots, n_d) \in \mathbb{N}_0^d \setminus \{0\} \), where \( i \in \{1, \ldots, d\} \) is chosen such that \( n_i \neq 0 \).

Proof. It suffices to consider the non-trivial case \( \mathbb{P}[X_1 \geq 1] > 0 \). We start with \( p \in (0,1) \) and let \( (\tilde{p}_n(p))_{n \in \mathbb{N}_0^d} \) denote the distribution of \( \tilde{S} = X_1 + \cdots + X_N \), where \( \tilde{N} \sim \text{NegBin}(\alpha, p) \), and \( (p_n(p))_{n \in \mathbb{N}_0^d} \) the distribution of \( S = X_1 + \cdots + X_N \), where \( N \sim \text{ExtNegBin}(\alpha - 1, 1, p) \). Since \( \text{NegBin}(\alpha, p) \) is the Panjer(\( p, (\alpha - 1)p, 0 \)) distribution, a recursion for the auxiliary sequence

\[
r_n(p) := (1 - p)^{-\alpha} \tilde{p}_n(p), \quad n \in \mathbb{N}_0^d,
\]

follows from Panjer’s recursion (5.15) for \( (\tilde{p}_n(p))_{n \in \mathbb{N}_0^d} \), namely

\[
r_n(p) = \frac{p}{n_i(1 - p \mathbb{P}[X_1 = 0])} \sum_{j \in \mathbb{N}_0^d, 0 < j \leq n} (\alpha j_i + n_i - j_i) \mathbb{P}[X_1 = j] r_{n-j}(p)
\]

87
for every \( n = (n_1, \ldots, n_d) \in \mathbb{N}_0^d \setminus \{0\} \) and \( i \in \{1, \ldots, d\} \) satisfying \( n_i \neq 0 \) and with starting value
\[
 r_0(p) = (1 - p \mathbb{P}[X_1 = 0])^{-\alpha}
\] (5.39)
given by (5.3) with probability-generating function from (5.14). The weighted convolution (5.36) becomes
\[
p_n(p) = \frac{(1 - p)^\alpha b(p)}{n_i} \sum_{j \leq n, j_i > 0} j_i \mathbb{P}[X_1 = j] r_{n-j}(p)
\] (5.40)
for every \( n = (n_1, \ldots, n_d) \in \mathbb{N}_0^d \setminus \{0\} \) and \( i \in \{1, \ldots, d\} \) satisfying \( n_i \neq 0 \) and
\[
p_0(p) = \frac{1 - (1 - p \mathbb{P}[X_1 = 0])^{1-\alpha}}{1 - (1 - p)^{1-\alpha}}
\] (5.41)
given by (5.3) with probability-generating function from (5.24). The normalization in (5.37) is chosen so that we can take the limit \( p \uparrow 1 \) in (5.38)–(5.41), in particular \((1 - p)^\alpha b(p)\) tends to \( 1 - \alpha \). With \( r_n := \lim_{p \uparrow 1} r_n(p) \) and \( p_n := \lim_{p \uparrow 1} p_n(p) \), the lemma follows.

\textit{Algorithm 5.31.} Corollary 5.29 and Lemma 5.30 lead to the following numerically stable algorithm for the calculation of the distribution of the aggregate loss in the collective risk model \( S = X_1 + \cdots + X_N \), where \( N \sim \text{ExtNegBin}(\alpha, k, p) \) with \( k \in \mathbb{N}, \alpha \in (-k, -k + 1) \) and \( p \in (0, 1] \):

- If \( p < 1 \), perform a stable Panjer recursion according to Theorem 5.8 for \( N \sim \text{NegBin}(\alpha + k, p) \), followed by a stable weighted convolution according to Corollary 5.29 to pass to \( N \sim \text{ExtNegBin}(\alpha + k - 1, 1, p) \).

- If \( p = 1 \), use Lemma 5.30 to calculate the distribution of the compound sum \( S \) for \( N \sim \text{ExtNegBin}(\alpha + k - 1, 1, p) \).

Calculate \( k - 1 \) weighted convolutions according to (5.36) to pass iteratively to \( N \sim \text{ExtNegBin}(\alpha + k - 2, 2, p), \ldots, \) and finally to \( N \sim \text{ExtNegBin}(\alpha, k, p) \).

\textit{Remark 5.32.} Of course, compared to the ordinary (but possibly unstable) Panjer recursion of Theorem 5.8 Algorithm 5.31 increases the numerical effort by a factor of \( k + 1 \). Note that the weighted convolution in (5.36) is not a recurrence, hence unavoidable rounding errors do not propagate as in a recursive calculation.

### 5.4 Numerically Stable Algorithm for ExtLog

Similar results as in the previous subsection can be obtained for the extended logarithmic distribution.\textsuperscript{23}

\textsuperscript{23}The results of this subsection will not be used in the remaining part of lecture notes.
Corollary 5.33 ([21] Corollary 5.4). For the parameters \( k \in \mathbb{N} \) and \( p \in (0, 1] \) with \( p < 1 \) in case \( k = 1 \), let \((q_n)_{n \in \mathbb{N}_0}\) denote the ExtLog\((k+1,p)\) distribution and \((\tilde{q}_n)_{n \in \mathbb{N}_0}\) the ExtLog\((k,p)\) distribution, where ExtLog\((1,p)\) stands for Log\((p)\).

Then (5.27) holds with \( l = 1 \) (we drop this index for convenience) and \( \tilde{q}_{1,n} = \tilde{q}_n \) for \( n \geq k + 1 \). The constants are given by \( a = 0 \) and

\[
b = (k+1)p - \frac{\sum_{l=k}^{l=k+1} (l-1)^{-1} \frac{1}{p^l}}{\sum_{l=k+1}^{\infty} (l-1)^{-1} \frac{1}{p^l}}
\]

hence (5.29) simplifies to the numerically stable weighted convolution (5.30) and \( p_0 \) is given by (5.3).

Exercise 5.34. Use Theorem 5.26(a) to prove Corollary 5.33.

In the excluded case \((k, p) = (1, 1)\), we cannot reduce the calculation for \( N \sim \text{ExtLog}(2, p) \) to that for \( N \sim \text{ExtLog}(1, p) = \text{Log}(p) \), because the logarithmic distribution from Example 4.4 is not defined for \( p = 1 \). Fortunately, a similar limit consideration as for the extended negative binomial distribution works.

Lemma 5.35 (Multi-dimensional version of [21] Lemma 5.5), stable recursion for \( \text{ExtLog}(2, 1) \). Assume that \( N \sim \text{ExtLog}(2, 1) \). Then the distribution \((p_n)_{n \in \mathbb{N}_d^0}\) of the random sum \( S = X_1 + \cdots + X_N \) can be calculated by

\[
p_0 = P[X_1 = 0] + P[X_1 \geq 1] \log P[X_1 \geq 1]
\]

with the convention \( 0 \log 0 = 0 \), and

\[
p_n = \begin{cases} \frac{1}{n_1} \sum_{j \in \mathbb{N}_d^0, 0 < j \leq n} j_i P[X_1 = j] r_{n-j} & \text{if } P[X_1 \geq 1] > 0, \\ 0 & \text{if } P[X_1 \geq 1] = 0, \end{cases}
\]

for every \( n = (n_1, \ldots, n_d) \in \mathbb{N}_d^0 \setminus \{0\} \) and \( i \in \{1, \ldots, d\} \) satisfying \( n_i \neq 0 \), where for the case \( P[X_1 \geq 1] > 0 \) the non-negative sequence \((r_n)_{n \in \mathbb{N}_d^0}\) is defined by \( r_0 = -\log P[X_1 \geq 1] \) and recursively in a numerically stable way by

\[
r_n = \frac{1}{P[X_1 \geq 1]} \left( P[X_1 = n] + \frac{1}{n_i} \sum_{j \in \mathbb{N}_d^0 \setminus \{0\}} (n_i - j_i) P[X_1 = j] r_{n-j} \right)
\]

for every \( n = (n_1, \ldots, n_d) \in \mathbb{N}_d^0 \setminus \{0\} \) and \( i \in \{1, \ldots, d\} \) satisfying \( n_i \neq 0 \).

Exercise 5.36. Prove Lemma 5.35. Hints: For \( p \in (0, 1) \) consider \( \tilde{N} \sim \text{Log}(p) \), let \((\tilde{p}_n(p))_{n \in \mathbb{N}_d^0}\) denote the distribution of \( \tilde{S} = X_1 + \cdots + X_N \), and let \((p_n(p))_{n \in \mathbb{N}_d^0}\) denote the distribution of \( S = X_1 + \cdots + X_N \), where \( N \sim \text{ExtLog}(2, p) \). Define the auxiliary sequence

\[
r_n(p) := -\tilde{p}_n(p) \log(1 - p), \quad n \in \mathbb{N}_d^0,
\]

and proceed in a similar way as in the proof of Lemma 5.30. Consider the limit \( p \nrightarrow 1 \) at the end.
6 Extensions of CreditRisk$^+$

Note that the extended multi-period CreditRisk$^+$ framework presented here can also be seen as a multi-period multi-business-line extension of the collective risk model from actuarial science.

6.1 Introduction

With the tools developed above we can now introduce the CreditRisk$^+$ framework and its extensions. First some general notes:

- The original CreditRisk$^+$ framework was developed by Credit Suisse First Boston (CSFB) [11].
- It is a one-period actuarial model for the aggregation of credit risks.
- It is based on the Poisson approximation of individual defaults, utilizing a trade-off effect occurring in sums, cf. Remark 3.30.
- One of the big advantages of the model is that the probability-generating function of the loss distribution is available in closed form.
- Extending the Poisson mixture model, several independent and gamma-distributed default causes as well as deterministic exposures are taken into account.
- The model does not call for Monte Carlo methods, hence the output is completely determined by the input data without any variations due to different simulation runs.

The extensions presented here include:

- The individual exposures of obligors are allowed to be $d$-dimensional random vectors making a multi-period model possible.
- Risk groups of obligors and corresponding, possibly stochastically dependent exposures can be handled.
- Default causes don’t need to be independent, they are allowed to have a special but flexible dependence structure, given by scenarios and independent risk factors.
- The distributions of the risk factors are not restricted to gamma distributions, instead also more flexible distributions like tempered stable distributions can be used.
- At least for gamma-distributed risk factors, the risk contributions of individual obligors can be calculated.
- The probability distribution of the portfolio loss can be derived with a numerically stable algorithm, even with all the mentioned extensions.
Note that, due to stochastic exposures, the risk of a downgraded credit rating can easily be incorporated in the extended version of CreditRisk+. Using risk groups, even joint downgrades can be modelled.

**Remark 6.1 (Multi-period extension).** The extension to several periods can be used in various ways and is also applicable in actuarial mathematics.

(a) If there are \( d \) periods, it is of importance to know in which period an obligor defaults. For example, an early default might cause liquidity problems for the lender, because write off is required early. Furthermore, the size of the loss given default can depend on the time of the default, in particular when a loan or a mortgage is amortized during its life span and not at maturity.

(b) A two-period model is of interest for a portfolio of credit guarantees. Here the default probability (or intensity) only refers to defaults happening during the first period, and the first component for the losses refers to the payout during this period. The second component of the losses models the payment obligations after the first period, it would correspond to the actuarial reserves to be built up at the end of the first period.

(c) In an insurance context, the \( d \) components can represent different types of claim payments. For a portfolio of health insurance contracts, this can be costs of medical treatments and allowances for missing income of the insured. For a portfolio of personal liability or automobile collision insurances, these can be claims for bodily injuries and property damages.

(d) In the context of stochastic claims reserving (see [59] for a textbook presentation), the \( d \) periods can represent the development years. Here the default probability (or intensity) refers to the claims originating from the initial insured period; the claims may be reported at a later period and payments may be spread out during the remaining periods of the model.

### 6.2 Description of the Model

We now assemble the necessary input parameters and the notation of the extended CreditRisk+ methodology.

#### 6.2.1 Input Parameters

Our extended version of CreditRisk+ needs the following input parameters:

- The number \( m \in \mathbb{N} \) of obligors,
- the number \( d \in \mathbb{N} \) of periods,
- the basic loss units \( E_1, \ldots, E_d > 0 \) for the \( d \) periods,
- the number \( C \in \mathbb{N} \) of non-idiosyncratic default causes,
- the number \( K \in \mathbb{N} \) of independent risk factors,
• the parameters specifying the gamma distributions or the tempered stable distributions of the independent risk factors $R_1, \ldots, R_K$,

• a non-empty finite set $\mathcal{J}$ of dependence scenarios,

• a probability distribution on the set $\mathcal{J}$ of dependence scenarios,

• for each dependence scenario $j \in \mathcal{J}$ a matrix $A_j = (a_{c,k}^j)_{c \in \{0, \ldots, C\}, k \in \{0, \ldots, K\}}$ of size $(C+1) \times (K+1)$ with non-negative entries, where

\[
a_{0,k}^j = 0 \quad \text{for all } j \in \mathcal{J} \text{ and } k \in \{1, \ldots, K\}, \tag{6.1}
\]

• the collection $G$ of nonempty subsets of all obligors $\{1, \ldots, m\}$, called the risk groups, which are subject to joint defaults.

For every group $g \in G$ we need

• the $d$-period default probability $p_g \in [0, 1]$,

and then, for every dependence scenario $j \in \mathcal{J}$,

• the susceptibility $w_{0,g,j} \in [0, 1]$ to idiosyncratic default,

• the susceptibilities $w_{c,g,j} \in [0, 1]$ to default causes $c \in \{1, \ldots, C\}$,

• the multivariate probability distributions $Q_{c,g,j} = (q_{c,g,j,\mu})_{\mu \in (\mathbb{N}_0^d)^g}$ on $(\mathbb{N}_0^d)^g$ describing the stochastic losses in $d$ periods of all the obligors $i \in g$ in multiples of the basic loss units $E_1, \ldots, E_d$ in case the risk group $g$ defaults due to cause $c \in \{0, \ldots, C\}$.

**Assumption 6.2.** Every obligor $i \in \{1, \ldots, m\}$ belongs to at least one group $g \in G$. Let $G_i := \{g \in G \mid i \in g\}$ denote the set of all groups to which obligor $i \in \{1, \ldots, m\}$ belongs, by assumption $G_i \neq \emptyset$.

**Remark 6.3.** While Assumption 6.2 is not necessary for the algorithm, it is useful to check the proper set-up of the model. If an obligor is not contained in any risk group, then a default is impossible and the obligor could be left out from the credit risk model.

**Assumption 6.4.** For each group $g \in G$ and each scenario $j \in \mathcal{J}$, the susceptibilities (also called weights) exhaustively describe the default causes. That is, for every $g \in G$ and $j \in \mathcal{J}$,

\[
\sum_{c=0}^C w_{c,g,j} = 1. \tag{6.2}
\]

**Remark 6.5.** Assumption 6.4 is useful for the interpretation of the default probability $p_g$ and the default intensity $\lambda_g$ for every risk group $g \in G$ in every scenario $j \in \mathcal{J}$, but the assumption is not necessary for the algorithm itself. See also the normalization in Assumption 6.35 below.
The idea of risk groups modelling joint defaults is motivated by the common Poisson shock models discussed by Lindskog and McNeil [36]. The idea to have different scenarios comes from [40], it originates from the desire to make negatively correlated default causes possible, see Example 6.38 below.

**Remark 6.6** (Classical CreditRisk+ model). The classical CreditRisk+ model is contained in the above set-up by choosing \( G = \{\{1\}, \{2\}, \ldots, \{m\}\} \), that means the only risk groups are the individual obligors. In this case \( Q_{c,(i),j} \) denotes the univariate distribution of the stochastic loss given default of obligor \( i \in \{1, \ldots, m\} \) due to cause \( c \in \{0, \ldots, C\} \) in scenario \( j \in J \). Note also that in the classical CreditRisk+ model there is just one scenario, i.e. \( |J| = 1 \), one period, i.e. \( d = 1 \), and risk causes and risk factors are identified, which corresponds to \( A^I \) being the identity matrix. Furthermore, all loss distributions \( Q_{c,(i),j} \) are one-dimensional and degenerate, which corresponds to deterministic one-period losses given default. Therefore, the classical CreditRisk+ model doesn’t even contain the collective model from actuarial mathematics.

**Remark 6.7** (Directly dependent defaults). Suppose obligor \( i \in \{1, \ldots, m\} \) is a large factory and obligors \( i_1, \ldots, i_l \in \{1, \ldots, m\} \) are suppliers of \( i \), being economically heavily dependent on the factory. If the factory \( i \) defaults and is subsequently closed, the suppliers \( i_1, \ldots, i_l \) have a high probability to default, too. Therefore, \( \{i, i_1, \ldots, i_l\} \) is certainly a meaningful risk group. Of course, \( G \) should also contain \( \{i\} \), because \( i \) could default and subsequently be taken over by a competitor running its production in the factory. Also \( \{i_1\}, \ldots, \{i_l\} \in G \) makes sense, because every supplier can individually default due to poor management and subsequently be replaced by a competing supplier. Note that different distributions \( Q_{c,g,j} \) of the \( (N_0^d)^g \)-valued loss vectors given default due to cause \( c \in \{0, \ldots, C\} \) in scenario \( j \in J \) can be specified for the big risk group \( g = \{i, i_1, \ldots, i_l\} \) and for the individual obligors represented by \( g = \{i\} \) and \( g = \{i_1\}, \ldots, \{i_l\} \).

**Remark 6.8** (Hindering defaults, competition groups). Suppose that the obligors \( i_1, \ldots, i_l \in \{1, \ldots, m\} \) are direct competitors in the market (e.g. airline companies), and a default of one of them may hinder a default of the others during the \( d \) periods, because they can take over the market share of the defaulting obligor and are then economically better off, they may even raise prices. To include this effect in the model, define a risk group \( g = \{i_1, \ldots, i_l\} \) with a default probability \( p_g \) and choose the multivariate loss distribution \( Q_{c,g,j} = (q_{c,g,j,\mu})_{\mu \in (N_0^d)^g} \) in such a way that \( q_{c,g,j,\mu} = 0 \) for every integer vector \( \mu = (\mu_1, \ldots, \mu_l) \) where two or more of the components \( \mu_1, \ldots, \mu_l \in N_0^d \) representing the losses during the \( d \) periods are different from 0 in \( N_0^d \). This means in case of a default of risk group \( g \) due to cause \( c \in \{0, \ldots, C\} \) in scenario \( j \in J \), that only one of the obligors in the group \( g \) causes a loss, and the distribution of this loss can of course depend on the obligor, on the cause \( c \) and on the scenario \( j \).

**Remark 6.9** (Examples of default causes). Default causes make it possible to build-in joint variations of default intensities for risk groups (and individual obligors); these variations jointly improve or degrade the credit quality of these
groups/obligors. Default causes can be industry sectors, individual countries, currency regions (e.g. Euro zone), geographic regions (e.g. North Africa, Latin America), religious regions (e.g. Islamic countries), economic regions (e.g. southern Europe, petroleum exporting countries (OPEC)), or represent exposure to macroeconomic indices like exchange rates, interest rates, business cycle, unemployment rates, real estate prices, interest rate changes and divorce rates (for modelling the risk of mortages, cf. [12, 13]), and so on. Note that these default causes don’t need to be stochastically independent, this is handled separately by the dependence scenarios and the matrices $A^j$ with $j \in J$.

**Remark 6.10 (Hierarchically ordered default causes).** For a worldwide diversified credit risk portfolio, it is a good idea to start with default cause intensities ordered in a hierarchical way:

(a) Worldwide, continental or multi-national causes, like the state of the economy in developed countries, international political or military conflicts, energy prices, crises due to excessive national debt in the European Union, turmoil in arabic countries, ...

(b) Default causes for every country, modeling an economic crises, the burst of a real-estate bubble, political turmoil, civil war, transfer risk, convertibility of the local currency, international sanctions, natural or man-made disasters, ...

(c) Local, industry sector specific causes within every country, like agriculture, mining, manufacturing, transport, financial and insurance industry, etc., where the granularity depends on the individual needs.

### 6.2.2 Stochastic Rounding

While losses are certainly multiples of one cent, the computation time required for this precision normally forces us to use basic loss units $E_1, \ldots, E_d$ of a larger size like 100 000 Euro. Then, however, losses are in general not integer multiples of this quantity and some rounding is required. Deterministic rounding with the aforementioned basic loss unit would round, for example, every loss below 50 000 Euro to zero, which is certainly not acceptable since it ignores the risk. The idea of stochastic rounding is to keep at least the expected loss constant. Hence, for example, a loss of 150 000 Euro happening with probability $p$ should be turned into two losses of sizes 100 000 and 200 000 Euros, respectively, each one happening with probability $p/2$. This idea, generalized to higher dimensions and mixed moments, is the content of the next lemma.

**Lemma 6.11** (Stochastic rounding). Let $X = (X_1, \ldots, X_d)$ be an $\mathbb{R}^d$-valued random vector. Define

$$p_n = \mathbb{E} \left[ \prod_{i=1}^{d} (1 - |X_i - n_i|)^+ \right], \quad n = (n_1, \ldots, n_d) \in \mathbb{Z}^d,$$

where $x^+ := \max\{x, 0\}$ for all $x \in \mathbb{R}$. Then the following holds:
(a) \((p_n)_{n \in \mathbb{Z}^d}\) is a probability mass function.

(b) If all components of \(X\) are almost surely non-negative, then \((p_n)_{n \in \mathbb{N}_0^d}\) is a probability mass function.

Let \(Y = (Y_1, \ldots, Y_d)\) be a \(\mathbb{Z}^d\)-valued random vector with distribution \((p_n)_{n \in \mathbb{Z}^d}\) given by (6.3) and let \(I\) be a non-empty subset of \(\{1, \ldots, d\}\).

(c) Stochastic rounding commutes with taking marginal distributions, i.e., stochastic rounding of the distribution of the random vector \((X_i)_{i \in I}\) equals the distribution of \((Y_i)_{i \in I}\).

(d) If \((X_i)_{i \in I}\) are independent, then \((Y_i)_{i \in I}\) are independent.

(e) For every \(i \in I\) let \(g_i: \mathbb{R} \to \mathbb{R}\) be a function which changes sign only at integers and which is piecewise linear between the integers, i.e.

\[
\lambda g_i(k) + (1 - \lambda) g_i(k + 1) = g_i(\lambda k + (1 - \lambda)(k + 1)) \quad (6.4)
\]

for all \(k \in \mathbb{Z}\) and \(\lambda \in [0, 1]\). Then the product \(\prod_{i \in I} g_i(X_i)\) is integrable if and only if \(\prod_{i \in I} g_i(Y_i)\) is integrable and in this case

\[
\mathbb{E} \left[ \prod_{i \in I} g_i(X_i) \right] = \mathbb{E} \left[ \prod_{i \in I} g_i(Y_i) \right]. \quad (6.5)
\]

Remark 6.12. Part (e) applied to \(I = \{i\}\) with \(i \in \{1, \ldots, d\}\) and the identity function \(g_i(x) = x\) on \(\mathbb{R}\) implies that expectations are unchanged by stochastic rounding, i.e. \(\mathbb{E}[X_i] = \mathbb{E}[Y_i]\), provided at least one (and therefore both) expectations exist. For \(I = \{i, j\} \subset \{1, \ldots, d\}\) with \(i \neq j\) and \(g_j\) also the identity function, we see that \(\mathbb{E}[X_iX_j] = \mathbb{E}[Y_iY_j]\), hence \(\text{Cov}(X_i, X_j) = \text{Cov}(Y_i, Y_j)\), provided \(X_i, X_j\) and their product \(X_iX_j\) are integrable.

Proof of Lemma 6.11. For each integer \(k \in \mathbb{Z}\) define \(f_k: \mathbb{R} \to [0, 1]\) by \(f_k(x) = (1 - |x - k|)^+\) for all \(x \in \mathbb{R}\). Let \(g: \mathbb{R} \to \mathbb{R}\) be a function which is piecewise linear between the integers, cf. (6.4). For \(x \in \mathbb{R}\) define \(k_x = \lfloor x \rfloor\) and observe that \(f_k(x) = 0\) for all \(k \in \mathbb{Z} \setminus \{k_x, k_x + 1\}\). Using (6.4) for the third equality,

\[
\sum_{k \in \mathbb{Z}} f_k(x) g(k) = f_{k_x}(x) g(k_x) + f_{k_x+1}(x) g(k_x + 1)
\]

\[
= (1 - (x - k_x)) g(k_x) + (1 - (k_x + 1 - x)) g(k_x + 1) \quad (6.6)
\]

\[
= \lambda g(k_x) + (1 - \lambda) g(k_x + 1)
\]

\[
= g((1 - (x - k_x))k_x + (x - k_x)(k_x + 1)) = g(x).
\]

Note that no convergence problems arise, since at most two terms are different from zero. Using (6.6) for \(g \equiv 1\), we see that \(\{f_k\}_{k \in \mathbb{Z}}\) is a partition of unity meaning in particular that

\[
\sum_{k \in \mathbb{Z}} f_k(x) = 1, \quad x \in \mathbb{R}. \quad (6.7)
\]
(a) Using (6.7) for every dimension and expanding leads to

$$\sum_{n_1, \ldots, n_d \in \mathbb{Z}^d} \prod_{i=1}^d f_{n_i}(x_i) = \prod_{i=1}^d \sum_{n_i \in \mathbb{Z}} f_{n_i}(x_i) = 1, \quad x = (x_1, \ldots, x_d) \in \mathbb{R}^d.$$  

Hence by monotone convergence,

$$\sum_{n \in \mathbb{Z}^d} p_n = \mathbb{E} \left[ \sum_{n \in \mathbb{Z}^d} \prod_{i=1}^d f_{n_i}(X_i) \right] = 1.$$  

(b) For every $n = (n_1, \ldots, n_d) \in \mathbb{Z}^d \setminus \mathbb{N}_0^d$ there exists $i \in \{1, \ldots, d\}$ with $n_i \leq -1$, hence $f_{n_i}(X_i) \overset{\text{a.s.}}{=} 0$ and $p_n = 0$.

(c) Let $(n_i)_{i \in I} \in \mathbb{Z}^I$ and $J := \{1, \ldots, d\} \setminus I$. Using monotone convergence and factoring,

$$\mathbb{P}[Y_i = n_i \text{ for all } i \in I] = \sum_{(n_j)_{j \in J} \in \mathbb{Z}^J} \mathbb{P}[Y_1, \ldots, Y_d = (n_1, \ldots, n_d)] = \mathbb{E}[\prod_{i=1}^d f_{n_i}(X_i)]$$

$$= \mathbb{E} \left[ \left( \prod_{i \in I} f_{n_i}(X_i) \right) \prod_{j \in J} \sum_{n_j \in \mathbb{Z}} f_{n_j}(X_j) \right].$$  

(d) Let $(n_i)_{i \in I} \in \mathbb{Z}^I$. Using part (c), the independence of $(X_i)_{i \in I}$, and again part (c),

$$\mathbb{P}[Y_i = n_i \text{ for all } i \in I] = \mathbb{E} \left[ \prod_{i \in I} f_{n_i}(X_i) \right] = \prod_{i \in I} \mathbb{E}[f_{n_i}(X_i)] = \prod_{i \in I} \mathbb{P}[Y_i = n_i].$$  

(e) Note that, if the functions $g_i$ change sign only at integers, then the functions $R \ni x \mapsto |g_i(x)|$ are also piecewise linear between integers, see (6.4), and (6.6) applies to them. Since all terms are non-negative, using the monotone convergence theorem,

$$\mathbb{E} \left[ \prod_{i \in I} |g_i(Y_i)| \right] = \sum_{(n_i)_{i \in I} \in \mathbb{Z}^I} \left( \prod_{i \in I} |g_i(n_i)| \right) \mathbb{P}[Y_i = n_i \text{ for all } i \in I]$$

$$= \sum_{(n_i)_{i \in I} \in \mathbb{Z}^I} \mathbb{E} \left[ \prod_{i \in I} f_{n_i}(X_i) |g_i(n_i)| \right] = \mathbb{E} \left[ \prod_{i \in I} \sum_{n_i \in \mathbb{Z}} f_{n_i}(X_i) |g_i(n_i)| \right],$$

hence $\prod_{i \in I} g_i(Y_i)$ is integrable if and only if $\prod_{i \in I} g_i(X_i)$ is integrable. The same calculation without the absolute value, which uses the dominated convergence theorem, proves (6.5).
Example 6.13 (Stochastic rounding can change the variance). Consider a degenerate random variable $X$ with $P[X = \frac{1}{2}] = 1$, which has zero variance. Stochastic rounding produces the Bernoulli distribution $Bin(1, \frac{1}{2})$, which has variance $\frac{1}{4}$.

Example 6.14 (Stochastic rounding can change the correlation). While Lemma 6.11 guarantees that stochastic rounding preserves covariances, rounding can change the correlations. As an explicit example, consider a random vector $(X_1, X_2) = \frac{1}{2}(Z, Z)$ with $Z \sim Bin(2, \frac{1}{2})$. Then $\text{Var}(Z) = \frac{1}{2}$, hence $\text{Cov}(X_1, X_2) = \frac{1}{4} \text{Var}(Z) = \frac{1}{8}$. Since $X_1$ and $X_2$ are four neighbouring lattice points in $\mathbb{Z}$. Stochastic rounding produces one quarter of the probability of $(\frac{1}{2}, \frac{1}{2})$ and $(\frac{3}{2}, \frac{3}{2})$, respectively, which are located on a square. The components $X_1$ and $X_2$ are comonotone, or by noting that $\text{Var}(X_1) = \frac{1}{4} \text{Var}(Z) = \frac{1}{8}$, it follows that $\text{Corr}(X_1, X_2) = 1$. Stochastic rounding might change the correlations. As an explicit example, consider a random vector $(Y_1, Y_2)$ with dependent components, then stochastic rounding might produce the probability mass function $p_{(0,0)} = p_{(1,1)} = \frac{3}{8}$ and $p_{(1,0)} = p_{(0,1)} = \frac{1}{8}$.

If $(Y_1, Y_2)$ has this distribution, then $\text{Cov}(Y_1, Y_2) = \frac{1}{8}$ by explicit calculation or an application of Lemma 6.11(e). Since $Y_1, Y_2 \sim Bin(1, \frac{1}{2})$, it follows that $\text{Var}(Y_1) = \text{Var}(Y_2) = \frac{1}{4}$, hence $\text{Corr}(Y_1, Y_2) = \frac{1}{2} \neq 1$.

Example 6.15 (Stochastic rounding can create independence). If $(X_1, X_2)$ is a random vector with dependent components, then stochastic rounding might remove the dependence. If $\text{Cov}(X_1, X_2)$ is well defined, then Lemma 6.11 shows that $\text{Cov}(X_1, X_2) = 0$ is a necessary condition for this phenomenon to occur. As an example, consider a random vector $(X_1, X_2)$ taking the four values $(1, 0), (1, 1), (\frac{1}{2}, \frac{1}{2})$ and $(\frac{3}{2}, \frac{3}{2})$, respectively, which are located on a square. The components $X_1$ and $X_2$ are clearly dependent, because

$$P[X_1 = 1, X_2 = \frac{1}{2}] = 0 \neq \frac{1}{2} = P[X_1 = 1] P[X_2 = \frac{1}{2}] .$$

Stochastic rounding distributes one quarter of the probability of $(\frac{1}{2}, \frac{1}{2})$ equally to is four neighbouring lattice points in $\mathbb{Z}^2$, the same happens to the probability of $(\frac{3}{2}, \frac{3}{2})$. Hence $p_{(0,0)} = p_{(0,1)} = p_{(2,1)} = \frac{1}{16}$ and $p_{(1,0)} = p_{(1,1)} = \frac{3}{8}$. This is the product measure of $\frac{1}{8}(\delta_0 + 6\delta_1 + \delta_2)$ with $\frac{1}{2}(\delta_0 + \delta_1)$.

### 6.2.3 Derived Parameters

The following quantities are derived from the input parameters:

- The Poisson intensity $\lambda_g$ for defaults of group $g \in G$ during the $d$ periods. As explained in Section 3.2, the choices $\lambda_g = p_g$ and $\lambda_g = p_g(1 - p_g)$ as well as $\lambda_g = -\log(1 - p_g)$ in case $p_g < 1$ can be used to calibrate the model. We will use the first choice in the following.

- From the multivariate probability distribution $Q_{c,g,j}$ on $N_0^d$ of the loss during the $d$ periods due to a default of group $g \in G$ caused by $c \in \{0, \ldots, C\}$ scenario $j \in J$, the $d$-dimensional distribution $Q^g_{c,g,j} = (q^g_{c,g,j,\nu})_{\nu \in N_0^d}$ of the group loss during the $d$ periods as sum of the individual losses of all the obligors $i$ in the group $g$ is given by

$$q^g_{c,g,j,\nu} = \frac{\sum_{\mu=0}^{\nu} q_{c,g,j,\mu}}{\sum_{\mu \in \nu} q_{c,g,j,\mu}}, \quad \nu \in N_0^d, \quad (6.8)$$

see Remark 6.17 below.
• The cumulative Poisson intensity

\[ \lambda_{j,k,\nu} := \sum_{g \in G} \lambda_g \sum_{c=0}^{C} w_{c,g,j} \alpha_{c,k}^j q_{c,g,j,\nu}^s \geq 0 \]  
  (6.9)

for losses of size \( \nu \in \mathbb{N}_0^d \setminus \{0\} \) in the portfolio due to risk factor \( k \in \{1, \ldots, K\} \) or idiosyncratic risk \( k = 0 \). In the first case, due to (6.1), the term for \( c = 0 \) can be omitted in (6.9).

• The set

\[ S_{j,k} := \{ \nu \in \mathbb{N}_0^d \setminus \{0\} \mid \lambda_{j,k,\nu} > 0 \} \]  
  (6.10)

of all non-zero \( d \)-period exposure vectors with strictly positive intensity in scenario \( j \in J \) due to risk factor \( k \in \{1, \ldots, K\} \) in terms of the basic loss units \( E_1, \ldots, E_d \). This set is used in (6.87) and (6.97) below.

• The cumulative Poisson intensity for non-zero \( d \)-period loss vectors in the portfolio in scenario \( j \in J \) due to risk \( k \in \{0, 1, \ldots, K\} \), given by

\[
\bar{\lambda}_{j,k} := \sum_{g \in G} \lambda_g \sum_{c=0}^{C} w_{c,g,j} \alpha_{c,k}^j (1 - q_{c,g,j,0}^s)
\]

\[
= \sum_{\nu \in \mathbb{N}_0^d \setminus \{0\}} \sum_{g \in G} \lambda_g \sum_{c=0}^{C} w_{c,g,j} \alpha_{c,k}^j q_{c,g,j,\nu}^s = \sum_{\nu \in S_{j,k}} \lambda_{j,k,\nu} \geq 0,
\]  
  (6.11)

where we used (6.9) and (6.10) for the last equality. Due to (6.10), \( \bar{\lambda}_{j,k} = 0 \) if and only if \( S_{j,k} = \emptyset \).

• If \( \bar{\lambda}_{j,k} > 0 \) for scenario \( j \in J \) and risk \( k \in \{0, \ldots, K\} \), then we can define the \( d \)-dimensional distribution \( Q_{j,k} = (q_{j,k,\nu})_{\nu \in \mathbb{N}_0^d} \) by

\[
q_{j,k,\nu} = \begin{cases} 
\lambda_{j,k,\nu} / \bar{\lambda}_{j,k} & \text{for all } \nu \in \mathbb{N}_0^d \setminus \{0\}, \\
0 & \text{for } \nu = 0 \in \mathbb{N}_0^d.
\end{cases}
\]  
  (6.12)

It is a probability distribution due to (6.11). By (6.9) and (6.11), the distribution \( Q_{j,k} \) is a mixture distribution of the family \( \{Q_{c,g,j} \mid c \in \{0, \ldots, C\}, g \in G\} \), conditioned to be non-zero. If \( \bar{\lambda}_{j,k} = 0 \) for a scenario \( j \in J \) and a risk \( k \in \{0, \ldots, K\} \), then no non-zero \( d \)-period loss vector in this scenario due to this risk factor is possible and we define

\[
q_{j,k,\nu} = \begin{cases} 
0 & \text{for all } \nu \in \mathbb{N}_0^d \setminus \{0\}, \\
1 & \text{for } \nu = 0 \in \mathbb{N}_0^d.
\end{cases}
\]  
  (6.13)

to avoid notational complications.
Note that the algorithm in Section 6.7 uses the intensities from (6.9) and (6.11), but not the default intensities of individual groups, not the individual susceptibilities, not the matrices $A_j$ with $j \in J$, and not the individual $d$-period loss distributions. Without loss of precision, the data can be aggregated accordingly. However, for the calculation of risk contributions (cf. Lemma 7.33 below), the individual quantities are important.

### 6.2.4 Notation for the Number of Default Events

For every risk group $g \in G$ and every scenario $j \in J$ we write

- $N_{0,g,j}$ for the number of idiosyncratic defaults (during the $d$ periods),
- $N_{c,g,j}$ for the number of defaults due to cause $c \in \{1, \ldots, C\}$,
- $N_{g,j} := \sum_{c=0}^C N_{c,g,j}$ for the total number of defaults in scenario $j$.

For every obligor $i \in \{1, \ldots, m\}$ and every scenario $j \in J$ we write analogously

- $N_{0,i,j} := \sum_{g \in G_i} N_{0,g,j}$ for the number of idiosyncratic defaults,
- $N_{c,i,j} := \sum_{g \in G_i} N_{c,g,j}$ for the number of defaults caused by $c \in \{1, \ldots, C\}$,
- $N_{i,j} := \sum_{c=0}^C N_{c,i,j} = \sum_{g \in G_i} N_{g,j}$ for the total number of defaults.

It may happen that a default results in a $d$-period loss vector of size zero. Let $J$ be a random variable selecting the scenario, i.e., $J$ takes values in the set $J$. Then

- $N_{c,g} := N_{c,g,j} = \sum_{j \in J} N_{c,g,j} 1\{J=j\}$ is the number of defaults of group $g \in G$ due to cause $c \in \{0, \ldots, C\}$,
- $N_{g} := N_{g,j} = \sum_{c=0}^C N_{c,g,j}$ describes the total number of defaults of risk group $g \in G$, and
- $N_{i} := N_{i,j} = \sum_{j \in J} N_{i,j} 1\{J=j\}$ describes the total number of defaults of the individual obligor $i \in \{1, \ldots, m\}$.

### 6.2.5 Notation for Stochastic Losses

Losses are $\mathbb{N}_0^d$-multiples of the basic loss units $E_1, \ldots, E_d$. As in Subsection 6.2.4, let $J$ be a random variable selecting the scenario from $J$.

- Let $L_{c,g,i,j,n}$ denote the $\mathbb{N}_0^d$-valued loss vector attributed to obligor $i \in g$ at default number $n \in \mathbb{N}$ of risk group $g \in G$ in scenario $j \in J$ due to cause $c \in \{1, \ldots, C\}$ or due to idiosyncratic cause $c = 0$.

The $\mathbb{N}_0^d$-valued loss vector due to default number $n \in \mathbb{N}$ of group $g \in G$ in scenario $j \in J$ caused by $c \in \{1, \ldots, C\}$ or due to idiosyncratic cause $c = 0$ is defined by

\[
L_{c,g,j,n} = \sum_{i \in g} L_{c,g,i,j,n}.
\]
• The $\mathbb{N}_0^d$-valued loss vector in scenario $j \in \mathcal{J}$ due to risk group $g \in G$ and cause $c \in \{1, \ldots, C\}$ or idiosyncratic cause $c = 0$ is defined by

$$L_{c,g,j} = \sum_{n=1}^{N_{c,g,j}} L_{c,g,j,n}. \quad (6.15)$$

• The $\mathbb{N}_0^d$-valued loss vector due to risk group $g \in G$ and cause $c \in \{0, \ldots, C\}$ is defined by

$$L_c = L_{c,g,j} = \sum_{j \in \mathcal{J}} L_{c,g,j} 1\{j=j\}. \quad (6.16)$$

• The total $\mathbb{N}_0^d$-valued loss vector in scenario $j \in \mathcal{J}$ due to group $g \in G$ is given by

$$L_{g,j} := \sum_{c=0}^C L_{c,g,j}. \quad (6.17)$$

• The total $\mathbb{N}_0^d$-valued loss vector in the portfolio in scenario $j \in \mathcal{J}$ is given by

$$L_j := \sum_{g \in G} L_{g,j}. \quad (6.18)$$

• The total $\mathbb{N}_0^d$-valued loss vector in the portfolio is given by

$$L := L_J = \sum_{j \in \mathcal{J}} L_j 1\{J=j\}. \quad (6.19)$$

For the interpretation of the model and the calculation of risk contributions in Subsection 7.3 below, we will also need the following definitions of $\mathbb{N}_0^d$-valued loss vectors attributed to obligor $i \in \{1, \ldots, m\}$:

• The attributed $\mathbb{N}_0^d$-valued loss vector in scenario $j \in \mathcal{J}$ due to defaults of group $g \in G_i$ and cause $c \in \{0, \ldots, C\}$ is given by

$$L_{c,g,i,j} := \sum_{n=1}^{N_{c,g,i,j}} L_{c,g,i,j,n}. \quad (6.20)$$

• The attributed $\mathbb{N}_0^d$-valued loss vector in scenario $j \in \mathcal{J}$ due to cause $c \in \{0, \ldots, C\}$ is given by the sum over all risk groups to which obligor $i$ belongs, i.e.,

$$L_{c,i,j} := \sum_{g \in G_i} L_{c,g,i,j}. \quad (6.21)$$

• The total attributed $\mathbb{N}_0^d$-valued loss vector in scenario $j \in \mathcal{J}$ is calculated by summing over all default causes, i.e.,

$$L_{i,j} := \sum_{c=0}^C L_{c,i,j}. \quad (6.22)$$
The total attributed \( \mathbb{N}_0^d \)-valued loss vector is given by the loss in the randomly selected scenario, i.e.,

\[
L_i := L_{i,j} = \sum_{j \in J} L_{i,j} 1_{\{J=j\}}.
\] (6.23)

### 6.3 Probabilistic Assumptions

The following assumptions are made:

**Assumption 6.16** (Group losses). For every group \( g \in G \), every default cause \( c \in \{0,\ldots,C\} \) and every dependence scenario \( j \in J \), the sequence of \( (\mathbb{N}_0^d)^g \)-valued random group loss vectors \( (L_{c,g,i,j,n})_{i \in g} \) with \( n \in \mathbb{N} \) is i.i.d. and independent of all other random variables\(^{24}\) with distribution

\[
P[L_{c,g,i,j,n} = \mu_i \text{ for all } i \in g] = q_{c,g,j,\mu}, \quad \mu = (\mu_i)_{i \in g} \in (\mathbb{N}_0^d)^g.
\] (6.24)

**Remark 6.17.** From Assumption 6.16 it follows that the sequence \( (L_{c,g,j,n})_{n \in \mathbb{N}} \) of \( \mathbb{N}_0^d \)-valued loss vectors of group \( g \in G \) in scenario \( j \in J \) due to cause \( c \in \{0,\ldots,C\} \) defined in (6.14) is also i.i.d. with distribution \( Q_{c,g,j} \) given in (6.8).

More explicitly, for all \( n \in \mathbb{N} \) and \( \nu \in \mathbb{N}_0^d \),

\[
P[L_{c,g,j,n} = \nu] = \mathbb{P}\left[ \sum_{i \in g} L_{c,g,i,j,n} = \nu \right] = \sum_{\mu=(\mu_i)_{i \in g} \in (\mathbb{N}_0^d)^g} \mathbb{P}[L_{c,g,i,j,n} = \mu_i \text{ for all } i \in g] = q_{c,g,j,\nu}. \] (6.25)

Note that for the multivariate Bernoulli distribution, the multinomial distribution, the multivariate logarithmic distribution, and the negative binomial distribution, the distribution of the sum of the components is available in closed form, see (4.8), Exercise 4.17(a), Exercise 4.37(c), and Exercise 4.42(e), respectively.

**Example 6.18** (Deterministic subdivision of a loss within a risk group). Given a risk group \( g \in G \) with at least two obligors, a scenario \( j \in J \) and a default cause \( c \in \{0,\ldots,C\} \), we may want to attribute a deterministic share of the group loss to the individual obligors \( i \in g \) of the group. For this purpose, consider for every obligor \( i \in g \) a deterministic function \( h_{c,g,i,j}: \mathbb{N}_0^d \rightarrow \mathbb{N}_0^d \) such that

\[
\sum_{i \in g} h_{c,g,i,j}(\nu) = \nu, \quad \text{for all } \nu \in \mathbb{N}_0^d.
\] (6.26)

We can then divide up the \( n \)th group loss \( L_{c,g,j,n} \sim Q_{c,g,j}^a \) in a deterministic way and attribute the loss \( L_{c,g,i,j,n} = h_{c,g,i,j}(L_{c,g,j,n}) \) to obligor \( i \in g \). Due to

---

\(^{24}\)This means all other sequences of loss vectors, the scenario \( J \), the idiosyncratic default numbers \( (N_{a,g})_{a \in \mathcal{G}} \) in Assumption 6.25, the non-idiosyncratic default numbers \( (N_{c,g})_{c \in \{1,\ldots,C\}, g \in G} \) in Assumption 6.30 and the risk factors \( R_1,\ldots,R_K \) in Assumption 6.31 below.
we have \( \sum_{i \in g} L_{c,g,i,j,n} = L_{c,g,j,n} \) for every \( n \in \mathbb{N} \). For all \( n \in \mathbb{N} \) and \( \mu = (\mu_i)_{i \in g} \in (\mathbb{N}_0^d)^g \) with \( \nu := \sum_{i \in g} \mu_i \) we have that

\[
q_{c,g,j,\mu} = P[L_{c,g,i,j,n} = \mu_i \text{ for all } i \in g] = \begin{cases} 
q^s_{c,g,j,\nu} & \text{if } \mu = (h_{c,g,i,j}(\nu))_{i \in g}, \\
0, & \text{otherwise},
\end{cases}
\]

in particular the right-hand side of (6.8) only consists of a single term. If we restrict to the one-period case \( d = 1 \) and the functions \( \{h_{c,g,i,j}\}_{i \in g} \) are non-decreasing, then the attributed losses \( (L_{c,g,i,j,n})_{i \in g} \) are comonotonic. If we want to distribute the one-period loss of a group \( g = \{i_1, \ldots, i_l\} \) as uniform as possible over its members in a comonotone way, then

\[
h_{c,g,i,j}(\nu) = (\nu + k - 1)/l, \quad \text{for all } k \in \{1, \ldots, l\} \text{ and } \nu \in \mathbb{N}_0, \quad \text{(6.27)}
\]

is a possible choice.

**Remark 6.19.** Suppose that a risk group \( g \) has at least two members and that, for a specific default cause \( c \in \{0, \ldots, C\} \) and scenario \( j \in \mathcal{J} \), the individual \( \mathbb{N}_0^d \)-valued loss vectors of the obligors in \( g \) are given. If all but at most one of these losses are deterministic, then the losses are independent and the distribution of the \( (\mathbb{N}_0^d)^g \)-valued group loss vector and, therefore, the distribution \( Q_{c,g,j}^s \) from (6.8) and (6.23) are uniquely determined. If at least two individual loss vectors are non-deterministic, then their joint distribution on \( (\mathbb{N}_0^d)^g \) is not uniquely determined and can only be computed under additional assumptions. We treat the case of independent loss vectors in Example 6.20. For \( d = 1 \), we treat the case of comonotonic losses in Example 6.21 and the mixture of independent and comonotonic losses in Example 6.22. In applications, it remains to decide whether the marginal distributions of the group loss vector should equal the distributions of the loss vectors of the individual obligors and whether the additional assumption is a good approximation of economic reality.

**Example 6.20 (Independent losses within a risk group).** Given a risk group \( g \in G \) with at least two obligors, a scenario \( j \in \mathcal{J} \) and a default cause \( c \in \{0, \ldots, C\} \), we can consider independent \( \mathbb{N}_0^d \)-valued loss vectors \( (L_{c,g,i,j,n})_{i \in g} \) of the obligors in \( g \) given default of the group, with \( L_{c,g,i,j,n} \sim Q_{c,g,i,j} = (q_{c,g,i,j,\nu})_{\nu \in \mathbb{N}_0^d} \) for every \( i \in g \) and \( n \in \mathbb{N} \). In this case \( Q_{c,g,j} = (q_{c,g,j,\nu})_{\mu \in (\mathbb{N}_0^d)^g} \) is given by

\[
q_{c,g,j,\mu} = P[L_{c,g,i,j,1} = \mu_i \text{ for all } i \in g] = \prod_{i \in g} P[L_{c,g,i,j,1} = \mu_i] = q_{c,g,i,j,\mu_i}
\]

for every \( \mu = (\mu_i)_{i \in g} \in (\mathbb{N}_0^d)^g \). The distribution \( Q_{c,g,j}^s = (q^s_{c,g,j,\nu})_{\nu \in \mathbb{N}_0^d} \) from (6.23) for the group loss is then the convolution of the \( Q_{c,g,i,j} \) with \( i \in g \), explicitly

\[
q^s_{c,g,j,\nu} = \sum_{\mu = (\mu_i)_{i \in g} \in (\mathbb{N}_0^d)^g} \prod_{i \in g} q_{c,g,i,j,\mu_i}, \quad \nu \in \mathbb{N}_0^d. \quad \text{(6.29)}
\]
Example 6.21 (Comonotonic one-period losses within a risk group). Given a risk group \( g \in G \) with at least two obligors, a scenario \( j \in J \) and a default cause \( c \in \{0, \ldots, C\} \), we can consider comonotonic \( \mathbb{N}_0 \)-valued losses \( (L_{c,g,i,j,n})_{i \in g} \) of the obligors in \( g \) given default of the group, with \( L_{c,g,i,j,n} \sim Q_{c,g,i,j} = (q_{c,g,i,j,n})_{n \in \mathbb{N}_0} \) for every \( i \in g \) and \( n \in \mathbb{N} \). Let

\[
F_{c,g,i,j}(\mu_i) = \sum_{\nu=0}^{\mu_i} q_{c,g,i,j,\nu}, \quad \mu_i \in \mathbb{N}_0,
\]

denote the discrete distribution function of \( Q_{c,g,i,j} \) for \( i \in g \). In this case the distribution \( Q^c_{c,g,j} = (q^c_{c,g,j,\nu})_{\nu \in \mathbb{N}_0} \), where the superscript reminds of comonotonicity, with discrete distribution function

\[
F_{c,g,j}(\mu) = \sum_{\nu \in \mathbb{N}_0} q^c_{c,g,j,\nu'}, \quad \mu \in \mathbb{Z}^g,
\]
of the group loss vector is given recursively by

\[
q^c_{c,g,j,\nu} = \min_{i \in g} F_{c,g,i,j}(\mu_i) - \max_{i \in g} F_{c,g,j}(\mu - e_i), \quad \mu = (\mu_i)_{i \in g} \in \mathbb{N}_0^g, \quad (6.30)
\]

where \( e_i = (\delta_i,i)_{i \in g} \) with Kronecker's delta. Due to comonotonicity there is, for every \( \nu \in \mathbb{N}_0 \), at most one \( \mu_\nu = (\mu_i,\nu)_{i \in g} \in \mathbb{N}_0^g \) with \( \sum_{i \in g} \mu_i,\nu = \nu \) and \( q^c_{c,g,j,\mu_\nu} > 0 \). Hence the distribution \( Q^c_{c,g,j} = (q^c_{c,g,j,\nu})_{\nu \in \mathbb{N}_0} \), determined via (6.8), is in the comonotonic case given by

\[
q^c_{c,g,j,\nu} = \begin{cases} 
q^c_{c,g,j,\mu_\nu} & \text{if } \mu_\nu \text{ exists,} \\
0 & \text{otherwise,} 
\end{cases} \quad \nu \in \mathbb{N}_0. \quad (6.31)
\]

The discrete distribution function \( F^c_{c,g,j} \) corresponding to \( Q^c_{c,g,j} \) can be calculated recursively as follows: For each \( i \in g \) let \( \nu_{i,0} \in \mathbb{N}_0 \) denote the smallest number with \( q_{c,g,i,j,\nu_{i,0}} > 0 \). With \( \nu_0 := \sum_{i \in g} \nu_{i,0} \) define the initial terms by

\[
F^c_{c,g,j}(\nu) = \begin{cases} 
0 & \text{for } \nu \in \{0, \ldots, \nu_0 - 1\}, \\
\min_{i \in g} F_{c,g,i,j}(\nu_{i,0}) & \text{for } \nu = \nu_0.
\end{cases}
\]

For the recursion, assume that \( (\nu_{i,n})_{i \in g} \) is \( \mathbb{N}_0^g \) valued as well as \( F^c_{c,g,j} \) on \( \{0, \ldots, \nu_n\} \) are given. If \( F^c_{c,g,j}(\nu_n) = 1 \), then we can set \( F^c_{c,g,j}(\nu) = 1 \) for all \( \nu \in \mathbb{N} \) with \( \nu > \nu_n \) and we are done. Otherwise, proceed as follows: Define for every \( i \in g \)

\[
\nu_{i,n+1} = \begin{cases} 
\nu_{i,n} & \text{if } F_{c,g,i,j}(\nu_{i,n}) > F^c_{c,g,j}(\nu_n), \\
\min\{\nu \in \mathbb{N}_0 \mid \nu > \nu_{i,n}, q_{c,g,i,j,\nu} > 0\} & \text{otherwise,}
\end{cases}
\]

\[
\nu_{n+1} = \sum_{i \in g} \nu_{i,n+1}, \quad \text{and correspondingly}
\]

\[
F^c_{c,g,j}(\nu) = \begin{cases} 
F^c_{c,g,j}(\nu_{n}) & \text{for } \nu \in \{\nu_{n}+1, \ldots, \nu_{n+1}-1\}, \\
\min_{i \in g} F_{c,g,i,j}(\nu_{i,n+1}) & \text{for } \nu = \nu_{n+1}.
\end{cases}
\]
Example 6.22 (Mixture of independent and comonotonic one-period losses within a risk group). Given a risk group \( g \in G \) with at least two obligors, a scenario \( j \in \mathcal{J} \) and a default cause \( c \in \{0, \ldots, C\} \), we can consider a mixture distribution of independent and comonotonic \( N_0 \)-valued losses \((L_{c,g,i,j,n})_{i \in g}\) of the obligors in \( g \) given default of the group. Specifically, choose an \( \alpha_{c,g,j} \in [0,1] \) and define the mixed group loss distribution \( Q^m_{c,g,j} = (q^m_{c,g,j,n})_{n \in N_0^g} \) by

\[
q^m_{c,g,j,n} = \alpha_{c,g,j} q_{c,g,j,n} + (1 - \alpha_{c,g,j}) q^c_{c,g,j,n}, \quad n \in N_0^g,
\]

with \( q_{c,g,j,n} \) given by (6.28) and \( q^c_{c,g,j,n} \) given by (6.30). The distribution of the sum of all the losses in the group is then

\[
q_{s,m}^{c,g,j} = \alpha_{c,g,j} q_{s,g,j}^c + (1 - \alpha_{c,g,j}) q^{c,c}_{s,g,j}, \quad n \in N_0^g,
\]

with \( q_{s,g,j}^c \) given by (6.29) with \( d = 1 \) and \( q^{c,c}_{s,g,j} \) given by (6.31).

Remark 6.23 (Obligors with a credit guarantee). Suppose a bank, a regional authority or a country, let’s call it obligor \( a \in \{1, \ldots, m\} \), gives a credit guarantee to all obligors of a group \( g \subset \{1, \ldots, m\} \setminus \{a\} \) and possibly also issues a bond on its own. A default of institution \( a \) can cause a substantial loss, because all its credit guarantees become worthless and defaults of obligors in \( g \) cause greater losses. To model this concentration of risk, there are several options:

(a) A rough solution is to take, for every obligor \( i \in g \), every risk group \( h \in G_i \) to which \( i \) belongs, every default cause \( c \in \{0, \ldots, C\} \) and every scenario \( j \in \mathcal{J} \), as loss distribution \( Q_{c,h,j} \) a mixture of two distributions, the first corresponding to the loss given the guarantee for \( i \) is in place, and the second corresponding to the loss given the guarantor \( a \) defaulted before or together with \( i \). The weights for these mixtures have to be chosen appropriately. Note that this modelling approach can be set up such that the expected loss is the right one and the computational effort is minor. However, it can be a (rough) approximation of the loss distribution, because it can ignore a substantial part of the concentration risk arising from a default of guarantor \( a \) while taking the larger losses of the obligors in \( g \) into account without guarantor \( a \) actually defaulting.

(b) We can consider a risk group \( g(a) = \{a\} \cup g \) consisting of the guarantor \( a \) and all guarantees, because they may all default together. In the simplest case, the default intensity \( \lambda_{g(a)} \) and the susceptibilities of the risk group \( g(a) \) are those of obligor \( a \), who does not appear as a risk group of its own. Of course, a multivariate distribution \( Q_{c,g(a),j} \) on \( (N_0^g)^{g(a)} \) describing the stochastic loss of all the obligors in \( g(a) \) for scenario \( j \in \mathcal{J} \) and default cause \( c \in \{0, \ldots, C\} \) is needed. The following practical problems come to mind:

\[25\]This remark is work in progress.
• If $g$ is large, think of $|g| \geq 100$, then $Q_{c,g(a),j}$ and the corresponding sum $Q_{c,g(a),j}^k$ from (6.8) are computationally hard to calculate. A solution might be to make additional assumptions and apply the extended CreditRisk+ methodology to calculate an approximation of $Q_{c,g(a),j}$.

• It’s not apparent how to choose the susceptibilities for the risk group $g(a)$. The default causes for the guarantor $a$ might be disjoint from the default causes of the obligors in $g$, for example.

**Assumption 6.24** (Distribution of idiosyncratic default numbers). For each group $g \in G$, the number $N_{0,g}$ of idiosyncratic defaults is, conditioned on $J$, Poisson distributed according to the Poisson intensity $\lambda_g$, the susceptibility $w_{0,g,J}$ and the matrix entry $a_{0,0}^J$, i.e.,

$$\mathcal{L}(N_{0,g}|J) = \text{Poisson}(\lambda_g w_{0,g,J} a_{0,0}^J) \quad \text{for every } g \in G.$$  \hspace{1cm} (6.32)

**Assumption 6.25** (Conditional independence of idiosyncratic default numbers). Conditioned on $J$, the group default numbers $(N_{0,g})_{g \in G}$ due to idiosyncratic defaults are independent from one another and everything else\textsuperscript{26} in particular

$$\mathbb{P}[N_{0,g} = n_{0,g} \text{ for all } g \in G|J] = \prod_{g \in G} \mathbb{P}[N_{0,g} = n_{0,g}|J] = \prod_{g \in G} e^{-\lambda_g w_{0,g,J} a_{0,0}^J} \frac{\left(\lambda_g w_{0,g,J} a_{0,0}^J\right)^{n_{0,g}}}{n_{0,g}!}$$  \hspace{1cm} (6.33)

for all $n_{0,g} \in \mathbb{N}_0$, where we used (6.32) for the second equality.

**Assumption 6.26** (Structure of default cause intensities). The default cause intensities $\Lambda_1, \ldots, \Lambda_C$ are expressed in terms of the random matrix $A_J = \sum_{j \in J} A_j 1_{\{J = j\}}$ of size $(C + 1) \times (K + 1)$ and the non-negative risk factors $R_1, \ldots, R_K$ by

$$\Lambda_c = a_{c,0}^J + \sum_{k=1}^K a_{c,k}^J R_k, \quad c \in \{1, \ldots, C\}.$$  \hspace{1cm} (6.34)

**Remark 6.27** (Lower bound for default cause intensity). The scenario-dependent but otherwise constant term $a_{c,0}^J \geq 0$ in (6.34) is added so that a strictly positive lower bound for the default cause intensity $\Lambda_c$ can be put into the model despite mathematically convenient distributions (like gamma distributions) for the risk factors $R_1, \ldots, R_K$.

**Remark 6.28**. For notational convenience, we will sometimes use a constant ‘risk factor’ $R_0 \equiv 1$ and a scenario-dependent default cause intensity $\Lambda_0 = a_{0,0}^J$ for

\textsuperscript{26}This means the random loss vectors in Assumption 6.16, the non-idiosyncratic default numbers $(N_{c,g})_{c \in \{1, \ldots, C\}, g \in G}$ in Assumption 6.30 and the risk factors $R_1, \ldots, R_K$ in Assumption 6.31 below.
idiosyncratic risk, see (6.1), to write (6.34) in a more compact form or in matrix notation as
\[ \Lambda = AJR \] (6.35)
with column random vectors \( \Lambda = (\Lambda_0, \ldots, \Lambda_C)^\top \) and \( R = (R_0, \ldots, R_K)^\top \).

**Assumption 6.29** (Conditional distribution of non-idiosyncratic default numbers). For every default cause \( c \in \{1, \ldots, C\} \) and every group \( g \in G \), the non-idiosyncratic default number \( N_{c,g} \) is, conditioned on \( J, R_1, \ldots, R_K \), Poisson distributed with parameter given as product of the group default intensity \( \lambda_g \), the susceptibility \( w_{c,g,J} \), and the default cause intensity \( \Lambda_c \), this means
\[
\mathbb{P}[N_{c,g} = n | J, R_1, \ldots, R_K] \overset{\text{a.s.}}{=} e^{-\lambda_g w_{c,g,J} \Lambda_c} \frac{\lambda_g w_{c,g,J} \Lambda_c^n n!}{n!} \quad \text{by (6.36)}
\]
for all \( n \in \mathbb{N}_0 \), i.e.,
\[
\mathcal{L}(N_{c,g} | J, R_1, \ldots, R_K) \overset{\text{a.s.}}{=} \text{Poisson}(\lambda_g w_{c,g,J} \Lambda_c) \quad \text{(6.37)}
\]

**Assumption 6.30** (Conditional independence of non-idiosyncratic default numbers). Conditionally on \( J, R_1, \ldots, R_K \), the family
\[
\{N_{c,g} | c \in \{1, \ldots, C\}, g \in G\}
\]
of default numbers is independent, hence
\[
\mathbb{P}[N_{c,g} = n_{c,g} \text{ for } c \in \{1, \ldots, C\} \text{ and } g \in G | J, R_1, \ldots, R_K]
\overset{\text{a.s.}}{=} \prod_{c=1}^{C} \prod_{g \in G} \mathbb{P}[N_{c,g} = n_{c,g} | J, R_1, \ldots, R_K]
\overset{\text{a.s.}}{=} \prod_{c=1}^{C} \prod_{g \in G} e^{-\lambda_g w_{c,g,J} \Lambda_c} \frac{\lambda_g w_{c,g,J} \Lambda_c^{n_{c,g}} n!}{n!}
\text{by (6.36)}
\]
\[
\text{for all } n_{c,g} \in \mathbb{N}_0.
\]

**Assumption 6.31** (Independence of risk factors and scenario). The non-negative risk factors \( R_1, \ldots, R_K \) and the scenario variable \( J \) are stochastically independent random variables.

The independence of \( J \) and the risk factors \( R_1, \ldots, R_K \) is used for the algorithm in (6.89) below. It is also useful for calculating the moments and the covariances of the default cause intensities, as the following remark shows.

**Remark 6.32** (Expectation, variance and covariance of default cause intensities). If \( R_1, \ldots, R_K \in L^1(\mathbb{P}) \) and Assumptions 6.29 and 6.31 hold, then
\[
\mathbb{E}[\Lambda_c | J] = a_{c,0}^J + \sum_{k=1}^{K} a_{c,k}^J \mathbb{E}[R_k]
\]
(6.39)
and variance \( \sigma^2 \) for every \( c \in \{1, \ldots, C\} \). If, in addition, \( R_1, \ldots, R_K \in L^2(\mathbb{P}) \), then, for all \( c, d \in \{1, \ldots, C\} \),

\[
\text{Cov}(\Lambda_c, \Lambda_d | J) = \sum_{k,l=1}^{K} a_{c,k}^j a_{d,l}^j \text{Cov}(R_k, R_l) = \sum_{k=1}^{K} a_{c,k}^j a_{d,k}^j \text{Var}(R_k),
\]

hence, by \((3.60)\) from Lemma \(3.48\) it follows from \((6.39)\) and \((6.41)\) that

\[
\text{Cov}(\Lambda_c, \Lambda_d) = \mathbb{E}[\text{Cov}(\Lambda_c, \Lambda_d | J)] + \text{Cov}(\mathbb{E}[\Lambda_c | J], \mathbb{E}[\Lambda_d | J])
\]

\[
= \sum_{k=1}^{K} \mathbb{E}[a_{c,k}^j a_{d,k}^j] \text{Var}(R_k) + \sum_{k,l=0}^{K} \text{Cov}(a_{c,k}^j, a_{d,l}^j) \epsilon_k \epsilon_l
\]

with \( \epsilon_0 := 1 \) and \( \epsilon_k := \mathbb{E}[R_k] \) for \( k \in \{1, \ldots, K\} \).

**Remark 6.33** (Pseudo risk factors). Due to the independence of the risk factors \( R_1, \ldots, R_K \), see Assumption \(6.31\) it is not always possible to give them an economic interpretation. On the other hand, the distribution of the group losses, see Assumption \(6.16\) may vary with the default causes and might be determined by the legal contract. Therefore, it can be difficult to set up a dependence structure between the default cause intensities \( \Lambda_1, \ldots, \Lambda_C \) as in \((6.34)\) by economic considerations. A solution is the introduction of a random vector \( P = (P_0, \ldots, P_{K'})^\top \) of pseudo risk factors with an economic interpretation. Then a random matrix \( A'_j = \sum_{j' \in \mathcal{J}} A'_{j'} 1_{\{j=j'\}} \) of size \((C+1) \times (K'+1)\) with non-negative entries can be set up by economic considerations such that \( \Lambda = A'_j P \), where as before \( \Lambda = (\Lambda_0, \ldots, \Lambda_C)^\top \). The dependence of \( P_0, \ldots, P_{K'} \) can be specified by a random matrix \( A_j = \sum_{j' \in \mathcal{J}} A_{j'} 1_{\{j=j'\}} \) of size \((K'+1) \times (K'+1)\) with non-negative entries such that \( P = A_j R \), where \( R = (R_0, \ldots, R_K)^\top \) is the column vector of the independent risk factors. Then \((6.35)\) is satisfied for the matrix product

\[
A_j = A'_j \tilde{A}_j = \sum_{j' \in \mathcal{J}} A'_{j'} \tilde{A}_j 1_{\{j=j'\}}.
\]

Of course one has to make sure that the entries of the matrices \( A_j := A'_j \tilde{A}_j \) for \( j \in \mathcal{J} \) satisfy \((6.1)\); this is certainly the case if the corresponding entries of \( A'_j \) and \( \tilde{A}_j \) satisfy \((6.1)\).

**Assumption 6.34** (Gamma-distributed risk factors). The risk factors \( R_1, \ldots, R_K \) are gamma distributed random variables with expectation \( \epsilon_k := \mathbb{E}[R_k] > 0 \) and variance \( \sigma_k^2 := \text{Var}(R_k) > 0 \), i.e., with shape parameter \( \alpha_k = e_k^2/\sigma_k^2 \) and inverse scale parameter \( \beta_k = e_k/\sigma_k^2 \) for all \( k \in \{1, \ldots, K\} \) by \((4.37)\) and \((4.38)\).
Assumption 6.35 (Normalization of default causes). We assume that

$$E \left[ w_{0,g,J} a_{0,0}^J + \sum_{c=1}^{C} w_{c,g,J} \Lambda_c \right] = 1$$

(6.44)

for every group \( g \in G \).

Remark 6.36. Similar to Assumption 6.4, the preceding Assumption 6.35 is useful for the interpretation of the default probability \( p_g \) and the default intensity \( \lambda_g \) for every risk group \( g \in G \), but the assumption is not necessary for the algorithm itself.

Remark 6.37 (Sufficient conditions for Assumption 6.35). If \( E[R_k] = 1 \) for every risk \( k \in \{1, \ldots, K\} \) and \( E[A_J] \) is a stochastic matrix, then \( E[\Lambda_c] = 1 \) by (6.40) for every default cause \( c \in \{1, \ldots, C\} \). If the weights are deterministic, meaning that they do not depend on the scenario, then due to (6.1), which implies \( E[a_{0,0}^J] = 1 \) for the stochastic matrix \( E[A_J] \), and due to Assumption 6.4, the condition (6.44) is satisfied for every group \( g \in G \).

6.4 Covariance Structure of Default Cause Intensities

The following example, which is based on [44, Ex. 3.14], shows that due to the scenarios we can have negatively correlated default cause intensities and the correlation can be any value in \([-1, 0)\).

Example 6.38 (Negative correlation of default cause intensities). Let \( J \) attain the values in \( J = \{0, 1\} \) with strictly positive probability. Let \( R_1 \) and \( R_2 \) be two independent and gamma distributed random variables, independent of \( J \), with \( E[R_1] = E[R_2] = 1 \). Then Assumptions 6.31 and 6.34 are satisfied. Define

\[
A_J = \begin{pmatrix} \frac{1}{E[J]} & 0 & 0 \\ \frac{1}{E[J]} & 0 & 0 \\ 0 & 0 & \frac{1}{1-E[J]} \end{pmatrix}.
\]

Then \( \Lambda_1 = J R_1 / E[J] \) and \( \Lambda_2 = (1 - J) R_2 / E[1 - J] \) by (6.34). Since \( E[A_J] = I_3 \) is a stochastic matrix, \( E[\Lambda_1] = E[\Lambda_2] = 1 \). If the weights do not depend on the scenario \( j \in \{0, 1\} \) and satisfy Assumption 6.4, then Assumption 6.35 is satisfied, cf. Remark 6.37. Since the product \( \Lambda_1 \Lambda_2 \) contains the factor \( J(1 - J) = 0 \), we get \( \Lambda_1 \Lambda_2 = 0 \) and

\[
\text{Cov}(\Lambda_1, \Lambda_2) = -E[\Lambda_1] E[\Lambda_1] = -1.
\]

By direct computation using \( E[R_k^2] = Var(R_k) + 1 \) for \( k \in \{1, 2\} \) or by (6.42),

\[
\text{Var}(\Lambda_1) = \frac{\text{Var}(R_1) + 1}{E[J]} - 1 \quad \text{and} \quad \text{Var}(\Lambda_2) = \frac{\text{Var}(R_2) + 1}{1-E[J]} - 1.
\]

The correlation is therefore given by

\[
\text{Corr}(\Lambda_1, \Lambda_2) = \frac{\text{Cov}(\Lambda_1, \Lambda_2)}{\sqrt{\text{Var}(\Lambda_1) \text{Var}(\Lambda_2)}} = -\frac{\sqrt{E[J] E[1-J]}}{\sqrt{\text{Var}(R_1) + 1 - E[J] \sqrt{\text{Var}(R_2) + E[J]}}},
\]

108
which attains every value in $[-1, 0)$ if suitable values for $\text{Var}(R_1)$ and $\text{Var}(R_2)$ in $[0, \infty)$ are chosen. For the symmetric case $\mathbb{E}[J] = 1/2$ and $\text{Var}(R_1) = \text{Var}(R_2)$, this simplifies to

$$\text{Corr}(\Lambda_1, \Lambda_2) = -\frac{1}{1 + 2\text{Var}(R_1)}.$$  

Example 6.38 raises the question, whether every covariance structure of the default cause intensities is possible. We first characterize covariance matrices and collect some of their properties.

**Definition 6.39.** A quadratic matrix $\Sigma$ of size $d$ with real entries is called **positive semi-definite** if $\Sigma$ is symmetric and $v^\top \Sigma v \geq 0$ for all $v \in \mathbb{R}^d$.

**Remark 6.40.** If a symmetric matrix $\Sigma$ with real entries is not positive semi-definite, the $\text{R}$-command $\text{nearPD}$ can be used to calculate a corresponding approximation.

**Lemma 6.41.** (a) Let $X$ be a square-integrable $\mathbb{R}^d$-valued random vector. Then its covariance matrix $\text{Cov}(X, X)$ is positive semi-definite.

(b) Let $\Sigma$ be a positive semi-definite $d \times d$ matrix with real entries. Then there exists a square-integrable $\mathbb{R}^d$-valued random vector with $\text{Cov}(X, X) = \Sigma$.

(c) Let $X = (X_1, \ldots, X_d)^\top$ be a square-integrable $[0, \infty)^d$-valued random vector. Then $\text{Cov}(X_i, X_j) \geq -\mathbb{E}[X_i] \mathbb{E}[X_j]$ for all $i, j \in \{1, \ldots, d\}$ with $i \neq j$.

Let $\Sigma = (\Sigma_{i,j})_{i,j \in \{1, \ldots, d\}}$ be a positive semi-definite matrix with real entries.

(d) For all $i, j \in \{1, \ldots, d\}$,

$$\Sigma_{i,i} \geq 0 \quad \text{and} \quad |\Sigma_{i,j}| \leq \sqrt{\Sigma_{i,i} \Sigma_{j,j}}.$$  

(e) Let $A$ be a matrix of size $d \times k$ with real entries. Then $\Sigma' := A^\top \Sigma A$ is positive semi-definite.

(f) Assume that $\Sigma$ satisfies $\Sigma = A \Sigma' A^\top$ with a matrix $A$ of size $d \times k$ and a quadratic matrix $\Sigma'$ of size $k$, both with real entries. If $A^\top A$ is invertible, then $\Sigma'$ is positive semi-definite.

**Remark 6.42.** To see that the invertibility of $A^\top A$ in Lemma 6.41(f) is necessary, let all entries of $A$ and $\Sigma$ be zero. Then $\Sigma = A \Sigma' A^\top$ gives no information about the entries of $\Sigma'$, in particular $\Sigma' = -I_k$ is possible.

**Proof of Lemma 6.41.** (a) Note that $\text{Cov}(X, X)$ is symmetric and of size $d$ with real entries. Consider $X$ and $v \in \mathbb{R}^d$ as column vectors. Then

$$v^\top \text{Cov}(X, X) v = v^\top \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^\top] v = \mathbb{E}[v^\top (X - \mathbb{E}[X])(X - \mathbb{E}[X])^\top v] \geq 0.$$  

$$= v^\top (X - \mathbb{E}[X])$$
Then there exist an integer \( k \) such that their sizes are non-decreasing and compatible such that the product 
\[ X_d := A_{J_d} \ldots A_{J_2} X_1 \] is well defined and satisfies 
\[ \text{Cov}(X_d, X_d) = \Sigma. \]

In addition, \( \mathbb{E}[A_{J_2}], \ldots, \mathbb{E}[A_{J_d}] \) are sub-stochastic matrices (meaning that the entries in every row sum to at most 1).

This theorem and its proof are work in progress, skip it or read with caution.
Remark 6.45 (Non-uniqueness of the representation). Without further conditions, the representation in Theorem 6.44 is not unique. Already for Σ = Iₙ, where Iₙ denotes the identity matrix of size d ≥ 2, there exist several solutions: Take k = d and deterministic Aᵢⱼ = Pᵢⱼ with iⱼ ∈ {1, ..., d} for l ∈ {2, ..., d}, where Pᵢⱼ denotes the matrix permuting rows i and j, with Pᵢⱼ = Iₙ if i = j.

Proof of Theorem 6.44. We give a constructive, inductive proof of Theorem 6.44, where in each induction step several cases have to be considered.

Case 1: If d = 1, then take k = 1 and any non-negative random variable X₁,₁ with Var(X₁,₁) = Σ.

Case 2: If d ≥ 2 and Σ is a diagonal matrix with all diagonal elements different from zero, take k = d and independent and non-negative X₁,₁, ..., Xₙ,ₙ with Var(X₁,i) = Σ₁,i for all i ∈ {1, ..., d}. Furthermore, take degenerate random variables J₁ = ··· = Jₙ = 0 and deterministic A₁,₁ = ··· = Aₙ,ₙ = Iₙ.

Case 3: Suppose there exist different i, j ∈ {1, ..., d} with Σᵢ,i ≥ Σᵢ,j and |Σᵢ,j| = √Σᵢ,i − Σᵢ,j (according the Lemma 6.41[d] this certainly happens if Σ has a diagonal entry which is zero). Define the permutation matrix

\[ P = \begin{cases} Pᵢⱼ & \text{if } i \neq d \text{ and } j \neq d - 1, \\ Pᵢ₋₁,jPᵢ,j & \text{if } i = d \text{ or } j = d - 1, \end{cases} \]

which moves row i to row d − 1 and row j to row d, taking care of special cases. Then \( P⁻¹ = Pᵀ \), and Σ' := ΣPΣᵀ satisfies Σ = ΠΣP as well as \( Σ'
\]

\[ \begin{align*} Σ'_{d,1} & \geq Σ'_{d,d} \quad \text{and} \quad Σ'_{d,1} = fΣ'_{d-1,d-1} \] with factor

\[ f := \begin{cases} 0 & \text{if } Σ'_{d-1,d-1} = 0, \\ \sqrt{Σ'_{d,d}/Σ'_{d-1,d-1}} & \text{if } Σ'_{d-1,d-1} > 0 \text{ and } Σ'_{d-1,d} \geq 0, \\ -\sqrt{Σ'_{d,d}/Σ'_{d-1,d-1}} & \text{if } Σ'_{d-1,d-1} > 0 \text{ and } Σ'_{d-1,d} < 0. \end{cases} \]

Note that \( f \in [-1, 1] \) and Σ'ₙ,ₙ = f²Σ'ₙ₋₁,ₙ₋₁. We can partition Σ' as

\[ Σ' = \begin{pmatrix} Σ'_{d,d} & v \\ vᵀ & Σ'_{d,1} \end{pmatrix} \]

with column vector v = (v₁, ..., vₙ₋₂, Σ'ₙ₋₁,ₙ₋₁)ᵀ. Let u₁, ..., uₙ₋₂, Σ'ₙ₋₁,ₙ₋₁) denote the last column vector of Σ'. If d = 2, then v = fu. To prove by contradiction that v = fu also for d ≥ 3, assume that there exists an i ∈ {1, ..., d - 2} with vi ≠ fu. Define x = -(Σ'ᵢ,i + 1)/(2fuᵢ - 2vi) and z = (0, ..., 0, 1, 0, ..., 0, fu, -x)ᵀ ∈ Rᵈ with the 1 in position i. Then

\[ (Σ'z)ᵢ = \begin{cases} Σ'ᵢ,i + (fuᵢ - vᵢ)x & \text{for } j \in \{1, ..., d - 2\}, \\ uᵢ & \text{for } j = d - 1, \\ vᵢ & \text{for } j = d, \end{cases} \]
and \( z^\top \Sigma' z = \Sigma'_{i,i} + 2(fu_i - v_i)x = -1 \), which is impossible for the positive semi-definite matrix \( \Sigma' \). Due to \( v = fu \) and \( \Sigma'_{d,d} = f^2 \Sigma'_{d-1,d-1} \), it follows that

\[
\Sigma' = \begin{pmatrix} I_{d-1} \\ w^\top \end{pmatrix} \Sigma'' \begin{pmatrix} I_{d-1} \\ w \end{pmatrix},
\]

where \( w = (0, \ldots, 0, f)^\top \in \mathbb{R}^{d-1} \).

**Case 3(a):** Suppose that \( f \geq 0 \). Define \( J_d \equiv 0 \) and note that \( \Sigma = A J_d \Sigma'' A_{J_d}^\top \) with \( A_{J_d} := P^\top \begin{pmatrix} I_{d-1} \\ w^\top \end{pmatrix} = I_2 \).

Furthermore, note that \( A_{J_d}^\top A_{J_d} \) is an invertible diagonal matrix. Hence, \( \Sigma'' \) is positive semi-definite by Lemma 6.41(f) and the problem is reduced by one dimension and one risk factor.

**Case 3(b):** Suppose that \( f < 0 \).

**Case 4:** Take an \( i \in \{1, \ldots, d\} \) in the following order of priorities:

(a) All off–diagonal entries of \( \Sigma \) in row \( i \) are zero.

(b) All entries of \( \Sigma \) in row \( i \) are non-negative and the diagonal entry of every column \( j \in \{1, \ldots, d\} \setminus \{i\} \) with \( \Sigma_{i,j} > 0 \) satisfies \( \Sigma_{j,j} \leq \Sigma_{i,i} \).

(c) For every \( j \in \{1, \ldots, d\} \setminus \{i\} \) the diagonal entry satisfies \( \Sigma_{j,j} \leq \Sigma_{i,i} \).

By symmetry of \( \Sigma \), the same is true for column \( i \). We use the permutation matrix \( P = P_{d,i} \) to exchange rows \( d \) and \( i \) (hence \( P^{-1} = P^\top = P \)) and represent \( \Sigma = P \Sigma' P \) with \( \Sigma' := P \Sigma P \). Note that \( P \) is a stochastic matrix and that \( \Sigma' \) is positive semi-definite by Lemma 6.41(f). Now the last row and the last column of \( \Sigma' \) have the property (a), (b) or (c), respectively. We write

\[
\Sigma' = \begin{pmatrix} B \\ w \\ c \end{pmatrix} = \begin{pmatrix} B \\ cu^\top - cv^\top \\ c \\ cu - cv \end{pmatrix}
\]

with real square matrix \( B \) of size \( d - 1 \), constant \( c \in (0, \infty) \), and column vector \( w = (w_1, \ldots, w_{d-1})^\top \in \mathbb{R}^{d-1} \) decomposed componentwise into \( u = \max\{w/c, 0\} \) and \( v = \max\{0, -w/c\} \) in \( [0, \infty)^{d-1} \). The matrix \( B \) is positive semi-definite by Lemma 6.41(e).

**Case 4(a)** Here \( w = 0 \) and \( \Sigma' \) has block-diagonal form, hence the problem can be reduced by one dimension. Applying the theorem to the matrix \( B \) of size \( d - 1 \) yields \( k' \in \{1, \ldots, d - 1\} \), independent random variables \( J_2, \ldots, J_{d-1} \) and
X_{1,1}, \ldots, X_{1,k'}, and matrices A'_{J_2}, \ldots, A'_{J_{d-1}}. Define J_d \equiv 0 and k = k' + 1 as well as A_{J_d} = P, and take any independent, non-negative random variable X_{1,k} with \text{Var}(X_{1,k}) = c. Furthermore, define

\[ A_{J_l} = \begin{pmatrix} A'_{J_l} & 0 \\ 0 & 1 \end{pmatrix}, \quad l \in \{2, \ldots, d - 1\}. \tag{6.45} \]

**Cases 4(b) and 4(c):** Define the diagonal matrix \( \tilde{D} = \text{diag}(\tilde{D}_{1,1}, \ldots, \tilde{D}_{d-1,d-1}) \) with \( \tilde{D}_{j,j} = 1 - u_j \) for every \( j \in \{1, \ldots, d - 1\} \). Note that \( |w_j| \leq \sqrt{c\Sigma_{j,j}} \leq c \) by Lemma 6.41(f) and the choice of \( i \) satisfying [b] or [c], respectively. Since the case of equality without the absolute value was treated already, we have that \( u_j \in [0, 1) \) for every \( j \in \{1, \ldots, d - 1\} \), hence \( \tilde{D} \) is invertible. Define

\[ A = \begin{pmatrix} \tilde{D} & u \\ 0 & 1 \end{pmatrix} \quad \text{and note that} \quad A^{-1} = \begin{pmatrix} \tilde{D}^{-1} & -\tilde{D}^{-1}u \\ 0 & 1 \end{pmatrix} \]

and that \( A \) is a stochastic matrix. Hence we have the representation \( \Sigma' = A\Sigma''A^\top \) with

\[
\Sigma'' = A^{-1}\Sigma'(A^{-1})^\top = \begin{pmatrix} \tilde{D}^{-1} & -\tilde{D}^{-1}u \\ 0 & 1 \end{pmatrix} \begin{pmatrix} B & cu - cv \\ cu^\top - cv^\top & c \end{pmatrix} \begin{pmatrix} \tilde{D}^{-1} \\ 0 \\ -u^\top \tilde{D}^{-1} \\ 1 \end{pmatrix} = \begin{pmatrix} \tilde{D}^{-1}(B - cuu^\top + cuv^\top) - c\tilde{D}^{-1}v \\ cu^\top - cv^\top - u^\top \tilde{D}^{-1} \end{pmatrix} \begin{pmatrix} \tilde{D}^{-1} & 0 \\ -u^\top \tilde{D}^{-1} & 1 \end{pmatrix}.
\]

Defining \( \tilde{B} := \tilde{D}^{-1}(B - cuu^\top + cuv^\top + cuv^\top)\tilde{D}^{-1} \) and using that \( \tilde{D}v = v \), hence \( \tilde{D}^{-1}v = v \), it follows that

\[
\Sigma'' = \begin{pmatrix} \tilde{B} & -cv \\ -cv^\top & c \end{pmatrix}.
\]

By Lemma 6.41(c), the matrix \( \Sigma'' \) is positive semi-definite. By Lemma 6.41(c), the matrix \( \tilde{B} \) is positive semi-definite, too.

**Case 4(b):** Here \( v = 0 \) and \( \Sigma'' \) has block-diagonal form, hence the problem can be reduced by one dimension. Applying the theorem to the matrix \( \tilde{B} \) of size \( d - 1 \) yields \( k' \in \{1, \ldots, d - 1\} \), independent random variables \( J_2, \ldots, J_{d-1} \) and \( X_{1,1}, \ldots, X_{1,k'} \), and matrices \( A'_{J_2}, \ldots, A'_{J_{d-1}} \). Define \( J_d \equiv 0 \) and \( k = k' + 1 \) as well as the deterministic \( A_{J_d} = P\tilde{A} \), and take any independent, non-negative random variable \( X_{1,k} \) with \( \text{Var}(X_{1,k}) = c. \) Furthermore, define by \( A_{J_2}, \ldots, A_{J_{d-1}} \) by (6.45).

**Case 4(c):** It remains to treat case [c] by introducing scenarios. Let \( Y = (Y_1, \ldots, Y_d) \) be a square-integrable random vector and define \( e_d = E[Y_d] \). Let \( J \) be \( \{0, 1\} \)-valued with \( p := P[J = 1] = c/(c + e_d^2) \in (0, 1) \). Consider

\[
A_J = \begin{pmatrix} \tilde{C} & fJv \\ 0 & f(1 - J) \end{pmatrix},
\]

113
where $\tilde{C}$ denotes any invertible matrix of size $d - 1$ with non-negative entries and $f := 1/(1 - p) = (c + e_d^2)/e_d^2$ so that $\mathbb{E}[f(1 - J)] = 1$. For

$$\Sigma'' := \begin{pmatrix} \tilde{C}^{-1}(\tilde{B} - cvv^\top)(\tilde{C}^\top)^{-1} & 0 \\ 0 & 0 \end{pmatrix},$$

which is symmetric because $(C^\top)^{-1} = (C^{-1})^\top$, it follows that

$$A_J\Sigma''A_J^\top = \begin{pmatrix} \tilde{C} & fJv \\ 0 & f(1 - J) \end{pmatrix} \begin{pmatrix} \tilde{C}^{-1}(\tilde{B} - cvv^\top)(\tilde{C}^\top)^{-1} & 0 \\ 0 & 0 \end{pmatrix} A_J^\top$$

$$= \begin{pmatrix} (\tilde{B} - cvv^\top)(\tilde{C}^\top)^{-1} & 0 \\ 0 & \tilde{C}^\top & fJv^\top & f(1 - J) \end{pmatrix} = \begin{pmatrix} \tilde{B} - cvv^\top & 0 \\ 0 & 0 \end{pmatrix}.$$ 

Then

$$\text{Cov}(\mathbb{E}[Y], A_J\mathbb{E}[Y]) = \text{Cov}\left(\begin{pmatrix} v \\ -1 \end{pmatrix}, \begin{pmatrix} v \\ -1 \end{pmatrix} \right) e_d^2 f^2$$

$$= \begin{pmatrix} v \\ -1 \end{pmatrix} (v^\top - 1) e_d^2 f^2 \text{Var}(J) = \begin{pmatrix} cvv^\top & -cv \\ -cv & c \end{pmatrix},$$

because $\text{Var}(J) = p(1 - p)$ and $e_d^2 f^2 \text{Var}(J) = e_d^2 f p = c$. Therefore,

$$\mathbb{E}[A_J\Sigma''A_J^\top] + \text{Cov}(\mathbb{E}[Y], A_J\mathbb{E}[Y]) = \Sigma''.$$

Note that

$$\mathbb{E}[A_J] = \tilde{C} fJv \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \tilde{C} & cv/e_d^2 \\ 0 & 1 \end{pmatrix},$$

which can be turned into an invertible stochastic matrix by a proper choice of $\tilde{C}$ if all components of $cv/e_d^2$ are less than 1.

Since $\Sigma''$ is positive semi-definite and

$$\begin{pmatrix} I_{d-1} & v \\ -cv & c \end{pmatrix} \begin{pmatrix} \tilde{B} - cvv^\top \\ 0 \end{pmatrix} = \tilde{B} - cvv^\top,$$

it follows from Lemma 6.41(e) and (f), that the matrices $\tilde{B} - cvv^\top$ and $\tilde{\Sigma} := \tilde{C}^{-1}(\tilde{B} - cvv^\top)(\tilde{C}^\top)^{-1}$ of size $d - 1$ are also positive semi-definite, hence we can reduce the problem by one dimension. Applying the theorem to $\tilde{\Sigma}$ yields $k' \in \{1, \ldots, d - 1\}$, independent random variables $J_2, \ldots, J_{d-1}$ and $X_{1,1}, \ldots, X_{1,k'}$, and matrices $A'_J, \ldots, A'_{J_{d-1}}$. Define $J_1 = J$ and $k = k' + 1$ as well as the random matrix $A_{J_1} = PAA_J$, and take any independent, non-negative random variable $X_{1,k}$ with $\mathbb{E}[X_{1,k}] = e_d$ and $\text{Var}(X_{1,k}) = c$. Furthermore, define by $A_{J_2}, \ldots, A_{J_{d-1}}$ by (6.45).
6.5 Expectations, Variances and Covariances for Defaults

To illustrate the above assumptions, we calculate the expectations, variances and covariances of various default numbers and losses. The first three subsections apply Subsection 3.6.1 to the current model. Note that the results of Subsections 6.5.1, 6.5.2 and 6.5.3 are actually special cases of the results of Subsection 6.5.4, cf. Remark 6.51.

6.5.1 Expectation of Default Numbers

Let us start with the number of defaults

$$N_i = \sum_{g \in G_i} N_g = \sum_{g \in G_i} \sum_{c=0}^{C} N_{c,g}$$  (6.46)

of obligor $i \in \{1, \ldots, m\}$. First note that by Assumptions 6.24, 6.25, 6.29, 6.30 and the Poisson summation property (3.5) we have

$$\mathcal{L}(N_i | J, R_1, \ldots, R_K) \stackrel{a.s.}{=} \mathcal{L}\left( \sum_{g \in G_i} \left( N_{0,g} + \sum_{c=1}^{C} N_{c,g} \right) \bigg| J, R_1, \ldots, R_K \right)$$  (6.47)

where

$$\Lambda_i := \sum_{g \in G_i} \lambda_g \left( w_{0,g,J} a_{0,0}^J + \sum_{c=1}^{C} w_{c,g,J} \Lambda_c \right)$$  (6.48)

is the conditional default intensity of obligor $i$, hence

$$\mathbb{E}[N_i | J, R_1, \ldots, R_K] = \Lambda_i$$  (6.49)

by (3.3). By inserting a conditional expectation given $J, R_1, \ldots, R_K$, using (6.49) and the normalization given in Assumption 6.35

$$\mathbb{E}[N_i] = \mathbb{E}[\Lambda_i] = \sum_{g \in G_i} \lambda_g \mathbb{E} \left[ w_{0,g,J} a_{0,0}^J + \sum_{c=1}^{C} w_{c,g,J} \Lambda_c \right] = \sum_{g \in G_i} \lambda_g.$$  (6.50)

Therefore, the expected number of defaults of obligor $i$ is the sum of the default intensities of the risk groups, to which $i$ belongs.

Remark 6.46. Note that (6.50) gives the expected number of defaults of obligor $i \in \{1, \ldots, m\}$, but not every default has to lead to a credit loss, due to a sufficiently high collateral or deductible (in case of credit insurance). A corresponding remark applies to the results of Subsections 6.5.2 and 6.5.3 below.

Example 6.47. Consider a credit risk model with $m = 2$ obligors and the three risk groups $\{1\}$, $\{2\}$ and $\{1,2\}$. Assume that the one-year default intensities $\lambda_i = \mathbb{E}[N_i] > 0$ for obligors $i \in \{1,2\}$ are known. To calibrate the model, we
can take any \( \lambda_g \in [0, \min\{\lambda_1, \lambda_2\}] \) for \( g = \{1, 2\} \) and define for the remaining one-obligor risk groups \( \lambda_{\{i\}} = \lambda_i - \lambda_g \), where \( i \in \{1, 2\} \). Then (6.50) is satisfied, which shows that default intensities of risk groups with several obligors can in general not be derived from individual default intensities.

**Remark 6.48.** Suppose that in a credit risk model with \( m \geq 2 \) obligors, the individual default intensities \( \lambda_i = \mathbb{E}[N_i] \) of all obligors \( i \in \{1, \ldots, m\} \) and the default intensities \( \lambda_g \) of all groups \( g \in G \) with at least two obligors were derived by statistical estimates and expert opinions. Assuming that all one-obligor risk groups \( \{i\} \) with \( i \in \{1, \ldots, m\} \) belong to \( G \), we can then define

\[
\lambda_{\{i\}} = \lambda_i - \sum_{g \in G_i \atop g \neq \{i\}} \lambda_g, \quad i \in \{1, \ldots, m\},
\]

provided that this results in \( \lambda_{\{i\}} \geq 0 \) for every \( i \in \{1, \ldots, m\} \). Otherwise the statistical estimates and expert opinions are inconsistent.

### 6.5.2 Variance of Default Numbers

To calculate the variance of the number \( N_i \) of defaults of obligor \( i \in \{1, \ldots, m\} \), first note that \( \text{Var}(N_i \mid J, R_1, \ldots, R_K) \overset{\text{a.s.}}{=} \Lambda_i \) by (6.47), (3.3) and (3.4). Using (3.61) from Lemma 3.48 and (6.49), we obtain

\[
\text{Var}(N_i) = \mathbb{E}[\text{Var}(N_i \mid J, R_1, \ldots, R_K)] + \text{Var}(\mathbb{E}[N_i \mid J, R_1, \ldots, R_K]), \quad (6.51)
\]

which corresponds to (3.62). Using (6.50) and again (3.61) from Lemma 3.48, equation (6.51) turns into

\[
\text{Var}(N_i) = \mathbb{E}[N_i] + \mathbb{E}[\text{Var}(\Lambda_i \mid J)] + \text{Var}(\mathbb{E}[\Lambda_i \mid J]). \quad (6.52)
\]

Note that \( \text{Var}(N_i) \geq \mathbb{E}[N_i] \), because variances are non-negative. Using Assumption 6.34 about the independence of \( J, R_1, \ldots, R_K \),

\[
\mathbb{E}[\Lambda_i \mid J] = \sum_{g \in G_i} \lambda_g \left( \sum_{c=0}^{C} w_{c,g,J} a_{c,0}^J + \sum_{k=1}^{K} \mathbb{E}[R_k] \sum_{c=1}^{C} w_{c,g,J} a_{c,k}^J \right), \quad (6.53)
\]

Using Assumption 6.31 about the independence of \( J, R_1, \ldots, R_K \),

\[
\mathbb{E}[\Lambda_i \mid J] = \sum_{g \in G_i} \lambda_g \left( \sum_{c=0}^{C} w_{c,g,J} a_{c,0}^J + \sum_{k=1}^{K} \mathbb{E}[R_k] \sum_{c=1}^{C} w_{c,g,J} a_{c,k}^J \right), \quad (6.54)
\]

and

\[
\text{Var}(\Lambda_i \mid J) = \sum_{k=1}^{K} \text{Var}(R_k) \left( \sum_{g \in G_i} \lambda_g \sum_{c=1}^{C} w_{c,g,J} a_{c,k}^J \right)^2, \quad (6.55)
\]

where \( \mathbb{E}[R_k] \) and \( \text{Var}(R_k) \) are specified by Assumption 6.34.
If there is just one scenario, then $J$ and therefore $\mathbb{E}[\mathbf{A}_i \mid J]$ are constant, hence the last term $\text{Var}(\mathbb{E}[\mathbf{A}_i \mid J])$ in (6.52) is zero and $\text{Var}(\mathbf{A}_i \mid J)$ from (6.55) coincides with the term $\mathbb{E}[\text{Var}(\mathbf{A}_i \mid J)]$ in (6.52).

For the general case, note that $\text{Var}(\mathbb{E}[\mathbf{A}_i \mid J]) = \mathbb{E}[(\mathbb{E}[\mathbf{A}_i \mid J])^2] - (\mathbb{E}[\mathbf{A}_i])^2$ with $\mathbb{E}[\mathbf{A}_i]$ given by (6.50) and

$$\mathbb{E}[\mathbb{E}[\mathbf{A}_i \mid J]^2] = \sum_{j \in J} \left( \sum_{g \in G} \lambda_g \left( \sum_{c=0}^C w_{c,g,j} a_{c,0}^j + \sum_{k=1}^K \mathbb{E}[R_k] \sum_{c=1}^C w_{c,g,j} a_{c,k}^j \right) \right)^2 \mathbb{P}[J = j].$$

Taking the expectation of (6.55) shows that

$$\mathbb{E}[\text{Var}(\mathbf{A}_i \mid J)] = \sum_{k=1}^K \text{Var}(R_k) \sum_{j \in J} \left( \sum_{g \in G} \lambda_g \sum_{c=1}^C w_{c,g,j} a_{c,k}^j \right)^2 \mathbb{P}[J = j],$$

(6.57)

### 6.5.3 Covariances of Default Numbers

For obligors $i, i' \in \{1, \ldots, m\}$ with $i \neq i'$ we can calculate the covariance of $N_i$ and $N_{i'}$. By (3.60) from Lemma 3.48

$$\text{Cov}(N_i, N_{i'}) = \text{Cov}(\mathbb{E}[N_i \mid J], \mathbb{E}[N_{i'} \mid J]) + \mathbb{E}[\text{Cov}(N_i, N_{i'} \mid J)]$$

(6.58)

Using (6.46), the linearity of conditional covariance in both arguments, Assumption 6.25 and (3.60) from Lemma 3.48 we obtain

$$\text{Cov}(N_i, N_{i'} \mid J) = \sum_{g \in G \cap G_{i'}} \text{Var}(N_{0,g} \mid J) \lambda_g w_{0,g,j} a_{0,j}^j \text{ by Assumption 6.24 and 4.4}

+ \sum_{g \in G_{i}} \sum_{h \in G_{i'}} \sum_{c,d=1}^C \left( \mathbb{E}[\text{Cov}(N_{c,g}, N_{d,h} \mid J, R_1, \ldots, R_K) \mid J] \right. \left. \overset{\text{assumption 6.29}}{=} \text{Var}(N_{c,g}, N_{d,h} \mid J, \Lambda_c) \overset{\text{assumption 3.3}}{=} \lambda_g w_{c,g,j} \Lambda_c \text{ for } (c,g)=(d,h)

+ \text{Cov}(\mathbb{E}[N_{c,g} \mid J, R_1, \ldots, R_K], \mathbb{E}[N_{d,h} \mid J, R_1, \ldots, R_K]) \right) \overset{\text{assumption 6.30}}{=} \lambda_g w_{c,g,j} \Lambda_c \lambda_{h} w_{d,h,j} \Lambda_{d}$$

(6.59)

where we used Assumption 6.29 and 3.3 and 3.4 to calculate the conditional expectations and the conditional covariance. The conditional covariance of $N_{c,g}$ and $N_{d,h}$ given $J, R_1, \ldots, R_K$ vanishes if $(g,k) \neq (h,l)$ due to conditional independence formulated in Assumption 6.30. Therefore,

$$\text{Cov}(N_i, N_{i'} \mid J) = \sum_{g \in G_{i'}} \lambda_g \left( w_{0,g,j} a_{0,j}^j + \sum_{c=1}^C w_{c,g,j} \mathbb{E}[\Lambda_c \mid J] \right) \overset{\text{by 6.44}}{=} 1

+ \sum_{g \in G_{i}} \lambda_g \sum_{h \in G_{i'}} \lambda_h \sum_{c,d=1}^C w_{c,g,j} w_{d,h,j} \text{Cov}(\Lambda_c, \Lambda_d \mid J),$$

117
where the remaining covariance is given by (6.41). Substituting (6.41) into (6.59), and the result into (6.58) yields

\[
\text{Cov}(N_i, N_i') = \text{Cov}(E[N_i | J], E[N_{i'} | J]) + \sum_{g \in G_i \cap G_{i'}} \lambda_g \\
+ \sum_{g \in G_i} \lambda_g \sum_{h \in G_{i'}} \lambda_h \sum_{k=1}^{K} \text{Var}(R_k) \sum_{c,d=1}^{C} E[w_{c,g,J} w_{d,h,J} a^J_c a^J_d] ,
\]

and it follows from (6.49) and (6.54) that

\[
E[N_i | J] = E[\Delta_i | J] = \sum_{g \in G_i} \lambda_g \left( \sum_{c=0}^{C} w_{c,g} a^J_c + \sum_{k=1}^{K} E[R_k] \sum_{c=1}^{C} w_{c,g,J} a^J_c \right),
\]

and similarly for \( E[N_{i'} | J] \).

If there is just one scenario, then \( E[N_i | J] \) and \( E[N_{i'} | J] \) are deterministic and the covariance in (6.60) vanishes. Furthermore, there is no need to take the expectation on the right hand side of (6.60) and (omitting the \( J \)) we obtain

\[
\text{Cov}(N_i, N_i') = \sum_{g \in G_i \cap G_{i'}} \lambda_g \\
+ \sum_{k=1}^{K} \text{Var}(R_k) \left( \sum_{g \in G_i} \lambda_g \sum_{c=1}^{C} w_{c,g} a^J_c \right) \left( \sum_{h \in G_{i'}} \lambda_h \sum_{d=1}^{C} w_{d,h} a^J_d \right).
\]

(6.62)

**Remark 6.49.** In the classical CreditRisk\(^+\) model (cf. Remark 6.6) with only one-element risk groups, the expectation in (6.50), the variance from Subsection 6.5.2, and the covariance given in (6.62) simplify to \( E[N_i] = \lambda_i \),

\[
\text{Var}(N_i) = \lambda_i + \lambda_i^2 \sum_{k=1}^{K} w^2_{k,i} \text{Var}(R_k)
\]

and

\[
\text{Cov}(N_i, N_{i'}) = \lambda_i \lambda_{i'} \sum_{k=1}^{K} w_{k,i} w_{k,i'} \text{Var}(R_k)
\]

for all \( i, i' \in \{1, \ldots, m\} \) with \( i \neq i' \), where we used the abbreviations \( \lambda_i := \lambda_{\{i\}} \) and \( w_{k,i} := w_{k,\{i\}} \) and corresponding ones for the index \( i' \). Note that in the extended version, as shown by (6.60), contributions to the covariance can also come from the risk groups in \( G_i \cap G_{i'} \) and from the scenarios.
6.5.4 Default Losses

In this subsection, we assume that every $\mathbb{N}_0^d$-valued stochastic loss vector $L_{c,g,i,j,1}$ attributed to obligor $i \in g$ of risk group $g \in G$ in scenario $j \in J$ due to default cause $c \in \{0, \ldots, C\}$, as introduced in Subsection 6.2.5, satisfies $\mathbb{E}[\|L_{c,g,i,j,1}\|] < \infty$ and, when calculating variances and covariances, $\mathbb{E}[\|L_{c,g,i,j,1}\|^2] < \infty$.

Let us start with the calculation of the conditional expected loss vector attributed to obligor $i \in \{1, \ldots, m\}$ given the scenario $J$ and the risk factors $R_1, \ldots, R_K$.

$$L_i = \sum_{c=0}^C \sum_{g \in G_i} L_{c,g,i,J}$$

By (6.21) and (6.23),

$$\mathbb{E}[L_i | J, R_1, \ldots, R_K] \overset{\text{as}}{=} \sum_{g \in G_i} \left( \mathbb{E}[L_{0,g,i,J} | J] + \sum_{c=1}^C \mathbb{E}[L_{c,g,i,J} | J, R_1, \ldots, R_K] \right), \quad (6.65)$$

where we used Assumptions 6.16, 6.25, and (6.36) to simplify the conditional expectations. By Assumptions 6.16 and 6.24, the loss $L_{0,g,i,J}$ defined in (6.20) has a compound Poisson distribution and (4.85) implies that

$$\mathbb{E}[L_{0,g,i,J} | J] = \mathbb{E}[N_{0,g,J} | J] \mathbb{E}[L_{0,g,i,J,1} | J]. \quad (6.66)$$

By Assumptions 6.16 and 6.29, the loss $L_{g,i,k}$ due to risk factor $k \in \{1, \ldots, K\}$ has a conditional compound Poisson distribution given $\Lambda_k$, hence by (4.83)

$$\mathbb{E}[L_{g,i,k} | \Lambda_k] \overset{\text{as}}{=} \lambda_g w_{g,k} \mathbb{E}[L_{g,i,k,1}]. \quad (6.67)$$

Substitution of (6.66) and (6.67) into (6.65) yields

$$\mathbb{E}[L_i | \Lambda_1, \ldots, \Lambda_K] \overset{\text{as}}{=} \sum_{g \in G_i} \lambda_g \left( w_{g,0} \mathbb{E}[L_{g,i,0,1}] + \sum_{k=1}^K w_{g,k} \mathbb{E}[L_{g,i,k,1}] \right). \quad (6.68)$$

Since $\mathbb{E}[\Lambda_k] = 1$ by Assumption 6.34, we obtain

$$\mathbb{E}[L_i] = \sum_{g \in G_i} \lambda_g \sum_{k=0}^K w_{g,k} \mathbb{E}[L_{g,i,k,1}]. \quad (6.69)$$

Using (6.14) and (6.25), we get for the expected credit loss in the entire portfolio

$$\mathbb{E}[L] = \sum_{i=1}^m \mathbb{E}[L_i] = \sum_{g \in G} \lambda_g \sum_{k=0}^K w_{g,k} \mathbb{E}[L_{g,k,1}]. \quad (6.70)$$

\footnote{This section has to be adapted to the new notation and the generalized setting.}
Due to (6.2), the sums over the risks $k \in \{0, \ldots, K\}$ in (6.69) and (6.70) are actually convex combinations.

The next step is to calculate the conditional covariance of the losses due to obligors $i,j \in \{1, \ldots, m\}$ given the risk factors $\Lambda_1, \ldots, \Lambda_K$. Considering $i = j$, this calculation will give the conditional variance. We first rewrite $L_i$ and $L_j$ using (6.21) and (6.23). We then note that, conditioned on the risk factors $\Lambda_1, \ldots, \Lambda_K$, the family of random vectors \( \{(L_{g,i',k}, k') \in G, g \in G, k \in \{0, \ldots, K\}\} \) is independent by Assumptions 6.16, 6.25, and 6.30, hence

\[
\text{Cov}(L_i, L_j | \Lambda_1, \ldots, \Lambda_K) \approx \sum_{g \in G_i \cap G_j} \sum_{k=0}^{K} \lambda_g \sum_{g' \in G_j} w_{g,k} \text{E}[L_{g,i,k} L_{g,j,k}] = \sigma^2 \text{Var}(\Lambda_k).
\]
For $i = j$ this result simplifies to

$$\text{Var}(L_i) = \sum_{g \in G_i} \lambda_g \sum_{k=0}^K w_{g,k} \mathbb{E}[L_{g,i,k,1}^2] + \sum_{k=1}^K \left( \sum_{g \in G_i} \lambda_g w_{g,k} \mathbb{E}[L_{g,i,k,1}] \right)^2 \sigma_k^2. \quad (6.76)$$

**Remark 6.50.** In the classical CreditRisk$^+$ model (cf. Remarks 6.6 and 6.49) with only one-element risk groups, the results (6.69), (6.76) and (6.75) simplify to

$$\mathbb{E}[L_i] = \lambda_i \sum_{k=0}^K w_{i,k} \mathbb{E}[L_{i,k,1}],$$

(6.77)

$$\text{Var}(L_i) = \lambda_i \sum_{k=0}^K w_{i,k} \mathbb{E}[L_{i,k,1}^2] + \lambda_i^2 \sum_{k=1}^K \sigma_k^2 w_{i,k}^2 \mathbb{E}[L_{i,k,1}]^2 \quad (6.78)$$

and

$$\text{Cov}(L_i, L_j) = \lambda_i \lambda_j \sum_{k=1}^K \sigma_k^2 w_{i,k} w_{j,k} \mathbb{E}[L_{i,k,1}] \mathbb{E}[L_{j,k,1}] \quad (6.79)$$

for all $i, j \in \{1, \ldots, m\}$ with $i \neq j$, where we used the abbreviations $\lambda_i := \lambda_{\{i\}}$ and $w_{i,k} := w_{\{i\},k}$ as well as $L_{i,k,1} := L_{\{i\},i,k,1}$ and corresponding ones for the index $j$.

**Remark 6.51.** To see that the results of Subsections 6.5.1, 6.5.2 and 6.5.3 are actually special cases of the results of Subsection 6.5.4, define $L_{g,i,k,n} = 1$ for all risk groups $g \in G$, risks $k \in \{0, 1, \ldots, K\}$, obligors $i \in g$, and defaults $n \in \mathbb{N}$. Then (6.46) and (6.20)–(6.23) imply $N_i = L_i$ for all obligors $i \in \{1, \ldots, m\}$. Comparison shows that the expectation in (6.69) simplifies to (6.50), the variance in (6.76) simplifies to (6.51), and the covariance in (6.75) simplifies to (6.60).

### 6.5.5 Default Numbers with Non-Zero Loss

The default numbers considered in Subsections 6.5.1, 6.5.2 and 6.5.3 include defaults which lead to a loss of zero. This can actually happen in practice, for example, when the collateral is sufficient to cover the outstanding amount. The results of the previous subsection can be used to calculate the expectations, variances and covariances of the default numbers with non-zero loss. This is accomplished by using the Bernoulli random variables $L'_{g,i,k,n} := 1_{\{L_{g,i,k,n} \neq 0\}}$ instead of $L_{g,i,k,n}$.

Define for every obligor $i \in \{1, \ldots, m\}$ the number $L'_i$ of defaults with non-zero loss via (6.20), (6.21), and (6.23) using the just introduced $L'_{g,i,k,n}$. The results (6.69), (6.76) and (6.75) applied to $L'_i$ and $L'_j$ can easily be rewritten using

$$\mathbb{E}[(L'_{g,i,k,1})^2] = \mathbb{E}[L'_{g,i,k,1}] = \mathbb{P}[L_{g,i,k,1} > 0]$$

and

$$\mathbb{E}[L'_{g,i,k,1} L'_{g,j,k,1}] = \mathbb{P}[L_{g,i,k,1} > 0, L_{g,j,k,1} > 0].$$

29This section has to be adapted to the new notation and the generalized setting.

121
for all obligors $i, j \in \{1, \ldots, m\}$, risks $k \in \{0, \ldots, K\}$ and groups $g \in G_i$ and $g \in G_i \cap G_j$, respectively.

6.6 Probability-Generating Function of the Biased Loss Vector

Fix a $\gamma = (\gamma_1, \ldots, \gamma_K) \in [0, \infty)^K$ such that $0 < \mathbb{E}[R_1^{\gamma_1} \cdots R_K^{\gamma_K}] < \infty$. In this section, using multi-index notation, we calculate the coefficients of the probability-generating function of the portfolio loss vector $L$ under the $R_1^{\gamma_1} \cdots R_K^{\gamma_K}$-biased probability measure, given according to Definition 2.10 which we denote by $P_\gamma$ for short. The corresponding expectation operator is denoted by $E_\gamma$. Hence we want to calculate

$$\varphi_{L,\gamma}(s) := \sum_{\nu \in \mathbb{N}_d^0} \mathbb{P}_\gamma[L = \nu] s^\nu = \mathbb{E}_\gamma[s^L] = \frac{\mathbb{E}[R_1^{\gamma_1} \cdots R_K^{\gamma_K} s^L | J]}{\mathbb{E}[R_1^{\gamma_1} \cdots R_K^{\gamma_K}]},$$

$s \in \mathbb{C}^d, \|s\|_\infty \leq 1,$ (6.80)

of the $\mathbb{N}_d^0$-valued total loss vector $L$ given by (6.19). For $\gamma = (0, \ldots, 0)$, we will obtain the usual probability-generating function $\varphi_L$ of $L$. Let

$$L' = \sum_{c=1}^C \sum_{g \in G} L_{c,g}$$

(6.81)

denote the non-idealysycratic $\mathbb{N}_d^0$-valued portfolio loss vector. By Assumptions 6.16 and 6.25, the random vectors $(L_{0,g})_{g \in G}$ and the random vector $(L', R_1, \ldots, R_K)$ are conditionally independent given $J$. Since

$$L = L' + \sum_{g \in G} L_{0,g},$$

it therefore follows that

$$\mathbb{E}[R_1^{\gamma_1} \cdots R_K^{\gamma_K} s^L | J] = \mathbb{E}[R_1^{\gamma_1} \cdots R_K^{\gamma_K} s^{L'} | J] \prod_{g \in G} \mathbb{E}[s^{L_{0,g}} | J].$$

(6.82)

By Assumptions 6.16, 6.24 and (4.54), it follows for the compound Poisson sum $L_{0,g,j}$, defined in (6.15), of idiosyncratic loss vectors of group $g \in G$ in scenario $j \in J$, that

$$\mathbb{E}[s^{L_{0,g,j}} | J = j] = \exp(\lambda_g w_{0,g,j} a_{0,g}^j (\varphi_{L_{0,g,j},1}(s) - 1)).$$

(6.83)

Conditioning on $J, R_1, \ldots, R_K$, the sector default numbers $(N_{c,g})_{c \in \{1, \ldots, C\}, g \in G}$ are independent by Assumption 6.30 hence the random sums $(L_{c,g})_{c \in \{1, \ldots, C\}, g \in G}$ in (6.81), given by (6.16), are also conditionally independent due to Assumption 6.16. Therefore, we obtain

$$\mathbb{E}[R_1^{\gamma_1} \cdots R_K^{\gamma_K} s^{L'} | J, R_1, \ldots, R_K]$$

$$\overset{a.s.}{=} R_1^{\gamma_1} \cdots R_K^{\gamma_K} \prod_{c=1}^C \prod_{g \in G} \mathbb{E}[s^{L_{c,g}} | J, R_1, \ldots, R_K]$$

(6.84)
Due to Assumptions 6.16 and 6.29, the result (4.77) and Assumption 6.26 it follows that, for every default cause \( c \in \{1, \ldots, C\} \) and every group \( g \in G \),

\[
\mathbb{E}[s_{L,c,g} \mid J = j, R_1, \ldots, R_K] \\
\overset{a.s.}{=} \mathbb{E}[s_{L,c} \mid J = j, \Lambda_c] \\
\overset{a.s.}{=} \exp(\lambda_g w_{c,g,j} \Lambda_c(\varphi_{L,c,g,j,1}(s) - 1)) \\
= \exp(\lambda_g w_{c,g,j} (\alpha^j_c + \sum_{k=1}^K a^j_{c,k} R_k)(\varphi_{L,c,g,j,1}(s) - 1)).
\]  

(6.85)

Substitution (6.83), (6.84) and (6.85) into (6.82) and rearrangement leads to

\[
\mathbb{E}[R_1 \cdots R_K s^L \mid J = j, R_1, \ldots, R_K] \\
\overset{a.s.}{=} \exp\left(\sum_{g \in G} \lambda_g \sum_{c=0}^C w_{c,g,j} a^j_c(\varphi_{L,c,g,j,1}(s) - 1)\right) \\
\times \prod_{k=1}^K R^*_k \exp\left(\sum_{g \in G} \lambda_g \sum_{c=0}^C w_{c,g,j} a^j_c(\varphi_{L,c,g,j,1}(s) - 1)\right).
\]  

(6.86)

For every scenario \( j \in J \) and risk \( k \in \{0, \ldots, K\} \) let

\[
\varphi_{j,k}(s) = \sum_{\nu \in S_{j,k} \cup \{0\}} q_{j,k,\nu} s^\nu = \begin{cases} \\
\lambda_{j,k}^{-1} \sum_{\nu \in S_{j,k}} \lambda_{j,k,\nu} s^\nu & \text{if } \lambda_{j,k} > 0, \\
1 & \text{if } \lambda_{j,k} = 0,
\end{cases}
\]  

(6.87)

at least for all \( s \in \mathbb{C}^d \) with \( \|s\|_\infty \leq 1 \), denote the probability-generating function of the distribution \( Q_{j,k} = (q_{j,k,\nu})_{\nu \in \mathbb{N}_0^d} \) defined in (6.12) and (6.13), respectively, with the set \( S_{j,k} \) defined in (6.10). Recall that, for all default causes \( c \in \{0, \ldots, C\} \), groups \( g \in G \) and scenarios \( j \in J \),

\[
\varphi_{L,c,g,j,1}(s) = \sum_{\nu \in \mathbb{N}_0^d} s^\nu \mathbb{P}[L_{c,g,j,1} = \nu] = q^e_{c,g,j,\nu} \text{ by (6.25)}
\]

hence

\[
\varphi_{L,c,g,j,1}(s) - 1 = \sum_{\nu \in \mathbb{N}_0^d \setminus \{0\}} s^\nu q^e_{c,g,j,\nu} - (1 - q^e_{c,g,j,0})
\]

and rearrangement of the exponents on the right-hand side of (6.86) leads to

\[
\sum_{g \in G} \lambda_g \sum_{c=0}^C w_{c,g,j} a^j_c (\varphi_{L,c,g,j,1}(s) - 1) \\
= \sum_{\nu \in S_{j,k}} s^\nu \left(\sum_{g \in G} \lambda_g \sum_{c=0}^C w_{c,g,j} a^j_c q^e_{c,g,j,\nu} - \sum_{g \in G} \lambda_g \sum_{c=0}^C w_{c,g,j} a^j_c (1 - q^e_{c,g,j,0})\right) \\
= \lambda_{j,k} \text{ by (6.9)} \\
= \lambda_{j,k} \text{ by (6.11)} \\
= \lambda_{j,k}(\varphi_{j,k}(s) - 1)
\]

(6.88)
for every risk $k \in \{0, \ldots, K\}$ with the set $S_{j,k}$ defined in (6.10). Substituting (6.88) into (6.86), using (6.1) in the case $k \in \{1, \ldots, K\}$, taking the conditional expectation given $J$, and using the independence of $J, R_1, \ldots, R_K$, it follows that

$$E[R_1^{\gamma_1} \ldots R_K^{\gamma_K} s^L \mid J = j] = \exp(\bar{\lambda}_{j,0}(\varphi_{j,0}(s) - 1))$$

$$\times \prod_{k=1}^{K} E[R_k^{\gamma_k} \exp(\bar{\lambda}_{j,k}(\varphi_{j,k}(s) - 1)R_k) \mid J = j],$$

at least for all $s \in \mathbb{C}^d$ with $\|s\|_{\infty} \leq 1$.

To proceed further, we need to make an assumption on the distribution of the risk factors $R_1, \ldots, R_K$.

### 6.6.1 Risk Factors with a Gamma Distribution

Since $R_k \sim \text{Gamma}(\alpha_k, \beta_k)$ for every $k \in \{1, \ldots, K\}$ by Assumption 6.34, and since $R_k$ is independent of $J$, it follows from (4.40) that

$$E[R_k^{\gamma_k} \exp(\bar{\lambda}_{j,k}(\varphi_{j,k}(s) - 1)R_k) \mid J = j] = E[R_k^{\gamma_k} \left(1 - \frac{\varphi_{j,k}(s) - 1}{\beta_k}\right)^{-(\alpha_k + \gamma_k)}, (6.90)$$

Substituting (6.90) into (6.89), we obtain

$$E[R_1^{\gamma_1} \ldots R_K^{\gamma_K} s^L \mid J = j] = \exp(\bar{\lambda}_{j,0}(\varphi_{j,0}(s) - 1))$$

$$\times \prod_{k=1}^{K} E[R_k^{\gamma_k} \left(1 - \frac{\varphi_{j,k}(s) - 1}{\beta_k}\right)^{-(\alpha_k + \gamma_k)}, (6.91)$$

Transferring everything into a common exponential, we finally get for the probability-generating function under the $R_1^{\gamma_1} \ldots R_K^{\gamma_K}$-biased probability measure, defined in (6.80),

$$\varphi_{L,\gamma}(s) = \frac{1}{E[R_1^{\gamma_1} \ldots R_K^{\gamma_K}]} \sum_{j \in \mathcal{J}} E[R_1^{\gamma_1} \ldots R_K^{\gamma_K} s^L \mid J = j] P[J = j]$$

$$= \sum_{j \in \mathcal{J}} \exp(\bar{\lambda}_{j,0}(\varphi_{j,0}(s) - 1))$$

$$- \sum_{k=1}^{K} (\alpha_k + \gamma_k) \log \left(1 - \frac{\varphi_{j,k}(s) - 1}{\beta_k}\right) P[J = j],$$

at least for all $s \in \mathbb{C}^d$ with $\|s\|_{\infty} \leq 1$.

### 6.7 Algorithm for Risk Factors with a Gamma Distribution

Formula (6.92) is the probability-generating function of the accumulated $\mathbb{N}_0^d$-valued loss vector in the credit portfolio under the $R_1^{\gamma_1} \ldots R_K^{\gamma_K}$-biased probability
measure. From the definition (4.1) we know that the coefficients of the power series of \((6.92)\) provide the desired distribution on \(\mathbb{N}^d_0\). We are aiming for an algorithm that works well for small (and even zero) variances of the risk factors, so we will rewrite our main formulas in terms of the expectations \(e_k = \mathbb{E}[R_k]\) and variances \(\sigma^2_k = \text{Var}(R_k)\) for all \(k \in \{1, \ldots, K\}\) using the formulas
\[
\alpha_k = \frac{e^2_k}{\sigma^2_k} \quad \text{and} \quad \beta_k = \frac{e_k}{\sigma^2_k},
\]
(6.93) derived from (4.37) and (4.38).

Remark 6.52 (Historical remark). The computation of these coefficients, however, can lead to numerical instabilities even in the one-period case with \((\gamma_1, \ldots, \gamma_K) = 0\), cf. [24]. Therefore, this section describes an algorithm, basically due to G. Giese [24], for which Haaf, Reiß, Schoenmakers [27] proved the numerical stability. Apparently these authors didn’t notice the relation to Panjer’s recursion, see Theorem 5.8, which was pointed out in [21, Section 5.5]. The algebraic step of putting everything into a common exponential to pass from (6.91) to (6.92) reflects the fact that the negative binomial distribution is a compound Poisson distribution, where the severity distribution is a logarithmic one, see Example 4.27. Since Panjer’s recursion is numerically stable for the Poisson as well as the logarithmic distribution, see Examples 5.17 and 5.21, respectively, numerical stability is guaranteed. The idea for the multi-period extension relies on the multivariate extension of Panjer’s algorithm given by Sundt [49].

6.7.1 Expansion of the Logarithm by Panjer’s Recursion

To calculate the coefficients of the power series of \((6.92)\), we first treat the logarithmic term. For this purpose, fix a scenario \(j \in J\) and a risk factor \(k \in \{1, \ldots, K\}\). Define
\[
P_{j,k} = \frac{\lambda_{j,k}}{\beta_k + \lambda_{j,k}} = \frac{\lambda_{j,k} \sigma^2_k}{e_k + \lambda_{j,k} \sigma^2_k} \in [0, 1)
\]
(6.94)
with inverse scale parameter \(\beta_k > 0\), expectation \(e_k > 0\) and variance \(\sigma^2_k\) from Assumption 6.34 and \(\lambda_{j,k} \geq 0\) from (6.11). Note that the right-hand side of (6.94) works fine for the degenerate case \(\sigma^2_k = 0\).

We consider a random variable \(M_{j,k} \sim \text{Log}(p_{j,k})\). Let \((Y_{j,k,n})_{n \in \mathbb{N}}\) be an i.i.d. sequence of \(\mathbb{N}^d_0\)-valued random vectors, independent of \(M_{j,k}\), with probability-generating function \(\varphi_{j,k}(s)\) defined in (6.87). Then by Example 4.4 and (4.52), the probability-generating function
\[
\bar{\varphi}_{j,k}(s) = \sum_{\nu \in \mathbb{N}^d_0} b_{j,k,\nu} s^\nu, \quad s \in \mathbb{C}^d, \|s\|_{\infty} \leq 1,
\]
of the \(\mathbb{N}^d_0\)-valued random sum
\[
S_{j,k} := \sum_{n=1}^{M_{j,k}} Y_{j,k,n}
\]
125
is given by
\[ \tilde{\varphi}_{j,k}(s) = \varphi_{j,k}(s) \frac{c(p_{j,k} \varphi_{j,k}(s))}{c(p_{j,k})}, \quad s \in \mathbb{C}^d, \|s\|_\infty \leq 1, \] (6.95)
and its coefficients \((b_{j,k,\nu})_{\nu \in \mathbb{N}^d_0}\) can be computed in a numerically stable way by Panjer’s recursion for the logarithmic distribution, see Example 5.21. More explicitly, using (4.6) and (5.17), the initial value is
\[ b_{j,k,0} = q_{j,k,0} \frac{c(p_{j,k} q_{j,k,0})}{c(p_{j,k})}, \] (6.96)
and, using (5.18), the recursion formula is, for every \(\nu \in \mathbb{N}^d_0 \setminus \{0\}\),
\[ b_{j,k,\nu} = \frac{1}{1 - p_{j,k}} \left( \frac{q_{j,k,\nu}}{c(p_{j,k})} + \frac{p_{j,k}}{\nu} \sum_{n \in S_{j,k} \atop n_k < \nu_k < \nu_k} (\nu_k - n_k) q_{j,k,n} b_{j,k,\nu-n} \right), \] (6.97)
where \(i \in \{1, \ldots, d\}\) is chosen such that \(\nu_i \neq 0\), and with \(p_{j,k}\) given by (6.94), \((q_{j,k,\nu})_{\nu \in \mathbb{N}^d_0}\) given by (6.12), and \(S_{j,k}\) defined in (6.10). Note that \(\gamma_k\) does not enter into this recursion. If \(p_{j,k} = 0\), then (6.96) and (6.97) simplify dramatically to
\[ b_{j,k,\nu} = q_{j,k,\nu}\] for all \(\nu \in \mathbb{N}^d_0\). To calculate the function \(c\) from (4.5) in a numerically stable way, see the corresponding comment in Example 4.4.

Rearranging and using (6.94) shows that
\[
1 - \tilde{\lambda}_{j,k} \varphi_{j,k}(s) - 1 = \frac{\beta_k + \tilde{\lambda}_{j,k}}{\beta_k} \left( 1 - \frac{\tilde{\lambda}_{j,k}}{\beta_k + \tilde{\lambda}_{j,k}} \varphi_{j,k}(s) \right)
\]
\[= \frac{1}{1 - p_{j,k}} \left( 1 - p_{j,k} \varphi_{j,k}(s) \right),\]
hence using (4.5) and (6.95) the logarithmic term in (6.92) can be rewritten as
\[
- \log \left( 1 - \tilde{\lambda}_{j,k} \varphi_{j,k}(s) - 1 \right) = - \log (1 - p_{j,k} \varphi_{j,k}(s)) + \log (1 - p_{j,k})
\]
\[= p_{j,k} \varphi_{j,k}(s) c(p_{j,k} \varphi_{j,k}(s)) - p_{j,k} c(p_{j,k})
=p_{j,k} c(p_{j,k}) \left( \tilde{\varphi}_{j,k}(s) - 1 \right). \] (6.98)

Substituting (6.98) into (6.92) gives
\[
\varphi_{L,\gamma}(s) = \sum_{j \in J} \exp \left( \tilde{\lambda}_{j,0} (\varphi_{j,0}(s) - 1) + \sum_{k=1}^{K} (\alpha_k + \gamma_k) p_{j,k} c(p_{j,k}) (\tilde{\varphi}_{j,k}(s) - 1) \right) \mathbb{P}[J = j], \] (6.99)
at least for all \(s \in \mathbb{C}^d\) with \(\|s\|_\infty \leq 1\).
6.7.2 Expansion of the Exponential by Panjer’s Recursion

To calculate the coefficients of the power series of (6.99), we first rewrite the argument of the exponential function. Define

\[
\lambda_j = \bar{\lambda}_{j,0} + \sum_{k=1}^{K} \frac{e_k^2 + \gamma_k \sigma_k^2}{e_k + \chi_{j,k} \sigma_k^2} c(p_{j,k}), \quad j \in J,
\]

with the shape parameter \(\alpha_k > 0\) expectation \(e_k > 0\) and variance \(\sigma_k^2\) given in Assumption 6.34. Poisson intensity \(\bar{\lambda}_{j,0} \geq 0\) given in (6.11), parameter \(p_{j,k} \in [0, 1)\) of the logarithmic distribution given in (6.94), and function \(c\) defined in (4.5).

Note that only non-negative terms are added in (6.100) and that its right-hand side even works in the degenerate case \(\sigma_k^2 = 0\), both facts guarantee numerical stability. For every \(j \in J\) with \(\lambda_j > 0\), we define \(\tilde{\varphi}_j(s) = \frac{1}{\lambda_j} \left( \bar{\lambda}_{j,0} \varphi_{j,0}(s) + \sum_{k=1}^{K} \frac{e_k^2 + \gamma_k \sigma_k^2}{e_k + \chi_{j,k} \sigma_k^2} c(p_{j,k}) \tilde{\varphi}_j(s) \right)\), at least for all \(s \in \mathbb{C}^d\) with \(\|s\|_\infty \leq 1\). Note that the coefficients of the power series \(\tilde{\varphi}_j(s) = \sum_{\nu \in \mathbb{N}_0} c_{j,\nu} s^\nu\), \(s \in \mathbb{C}^d\), \(\|s\|_\infty \leq 1\), are given as convex combinations of the corresponding coefficients of \(\varphi_{j,0}\) and \(\tilde{\varphi}_{j,1}, \ldots, \tilde{\varphi}_{j,K}\), which is a numerically stable operation. More explicitly,

\[
c_{j,\nu} = \frac{1}{\lambda_j} \left( \bar{\lambda}_{j,0} q_{j,0,\nu} + \sum_{k=1}^{K} b_{j,k,\nu} \chi_{j,k} \frac{e_k^2 + \gamma_k \sigma_k^2}{e_k + \chi_{j,k} \sigma_k^2} c(p_{j,k}) \right), \quad \nu \in \mathbb{N}_0^d, \tag{6.101}
\]

with \((q_{j,0,\nu})_{\nu \in \mathbb{N}_0^d}\) given by (6.12) or (6.13) and \((b_{j,k,\nu})_{\nu \in \mathbb{N}_0^d}\) given by (6.96) and (6.97). For every \(j \in J\) with \(\lambda_j = 0\), we define \(\tilde{\varphi}_j(s) = 1\) for all \(s \in \mathbb{C}^d\) and

\[
c_{j,\nu} = \begin{cases} 1 & \text{for } \nu = 0 \in \mathbb{N}_0^d, \\
0 & \text{for } \nu \in \mathbb{N}_0^d \setminus \{0\}. \end{cases} \tag{6.102}
\]

In every case, \(\tilde{\varphi}_j\) is again a probability-generating function, and (6.99) can be written as

\[
\varphi_{L,\gamma}(s) = \sum_{j \in J} \exp(\lambda_j(\tilde{\varphi}_j(s) - 1)) \mathbb{P}[J = j]. \tag{6.103}
\]

Fix a scenario \(j \in J\), let \(M_j \sim \text{Poisson}(\lambda_j)\) and consider an independent sequence \((Y_{j,n})_{n \in \mathbb{N}}\) of i.i.d. random variables, each one with probability-generating function \(\tilde{\varphi}_j\). Then by Example 4.3 and (4.52), the probability-generating function \(\psi_j\) of the distribution of the random sum

\[
S_j := \sum_{n=1}^{M_j} Y_{j,n}
\]
is given by
\[ \psi_j(s) = \exp(\lambda_j(\varphi_j(s) - 1)), \quad s \in \mathbb{C}^d, \|s\|_\infty \leq 1, \]
and its coefficients, let’s call them \((d_{j,\nu})_{\nu \in \mathbb{N}_0^d}\), can be computed in a numerically stable way by Panjer’s recursion for the Poisson distribution, see Example 5.17. Explicitly, (5.12) implies for the initial value
\[ d_{j,0} = \exp(\lambda_j(c_{j,0} - 1)) \quad (6.104) \]
(in case of numerical underflow, see Remark 5.19 for a remedy) and the recursion formula (5.13) turns, for every \(\nu = (\nu_1, \ldots, \nu_d) \in \mathbb{N}_0^d \setminus \{0\}\), into
\[ d_{j,\nu} = \frac{\lambda_j}{\nu_i} \sum_{n \in \mathbb{N}_0^d} n_i c_{j,n} d_{j,\nu-n}, \quad (6.105) \]
where \(i \in \{1, \ldots, d\}\) is chosen such that \(\nu_i \neq 0\), with \(\lambda_j\) given by (6.100) and the coefficients \((c_{j,\nu})_{\nu \in \mathbb{N}_0^d}\) given by (6.101) and (6.102), respectively. See Remark 5.10 to omit terms in (6.105) with value zero.

The weighted probability-generating function (6.103) simplifies to
\[ \varphi_{L,\gamma}(s) = \sum_{j \in J} \psi_j(s) \mathbb{P}[J = j], \quad s \in \mathbb{C}^d, \|s\|_\infty \leq 1, \]
and the coefficients of this power series are convex combinations of the corresponding coefficients of \((\psi_j)_{j \in J}\). These operations are numerically stable. Explicitly, the coefficients in (6.80) are determined by
\[ \mathbb{P}_\gamma[L = \nu] = \sum_{j \in J} d_{j,\nu} \mathbb{P}[J = j], \quad \nu \in \mathbb{N}_0^d, \]
with \((d_{j,\nu})_{\nu \in \mathbb{N}_0} \) given by (6.104) and (6.105).

**Exercise 6.53** (Implementation of the algorithm). Assume that there are \(m \in \mathbb{N}\) obligors, where obligor \(i \in \{1, \ldots, m\}\) has default probability \(p_i = 1/(20 + i)\) within one period, and that there is the idiosyncratic cause and \(C = 3\) additional default causes. Assume that the loss given default of obligor \(i \in \{1, \ldots, m\}\) due to cause \(c \in \{0, \ldots, C\}\) has the distribution \(\text{Bin}(i + c, i/(2i + 2c))\) and that all susceptibilities are equal to \(1/(C + 1)\). Let \(\Lambda_1, \ldots, \Lambda_C\) be default cause intensities with \(\mathbb{E}[\Lambda_c] = 1\) and \(\Lambda_c \geq 1/3c\) for all \(c \in \{1, \ldots, C\}\). Assume the there are only one-element risk groups and that there are two scenarios \(J = \{0, 1\}\). Extending Example 6.38, let \(J\) be \(J\)-valued and consider the \((C + 1) \times (K + 1)\)-matrix
\[ A_J = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ * & J & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ * & 0 & 1 - J & * & 0 & 0 & 0 & 0 & 0 & 0 \\ * & 0 & 0 & * & * & * & * & * & * & * \end{pmatrix}, \quad (6.106) \]
where \(*\) denotes non-zero, deterministic entries.
(a) With the given constraints, set up a flexible model satisfying Assumptions 6.34 and 6.35 such that $\text{Cov}(\Lambda_1, \Lambda_2) < 0$ and $\text{Cov}(\Lambda_2, \Lambda_3) > 0$.

(b) Calculate the expectations, variances and covariances of the default cause intensities $\Lambda_1, \Lambda_2, \Lambda_3$ (see Remark 6.32) in your model.

(c) Calculate the expected total credit portfolio loss. Does the result depend on your specific choice of the dependence structure?

(d) Calculate the distribution of the total credit portfolio loss numerically for an $m \geq 50$ of your choice for your specific dependence structure.

6.8 Algorithm for Risk Factors with a Tempered Stable Distribution

6.9 Special Cases

In order to test the algorithm, its implementation and its numerical stability, it is helpful to consider special cases of the parameters, where the corresponding distribution of the total loss $L$ given in (6.19) can be calculated directly. In this section we assume that all group losses are multiples of some $C \in \mathbb{N}$, meaning that we have

$$L_{g,k,n} = CL'_{g,k,n}.$$ 

with an $\mathbb{N}_0$-valued $L'_{g,k,n}$ for every loss $n \in \mathbb{N}$ of risk group $g \in G$ due to risk $k \in \{0, \ldots, K\}$. We adopt the notation from (6.15), (6.17) and (6.19). In this section, we will not attribute the group loss to its individual members.

6.9.1 Pure Poisson Case

We only consider the degenerate case $\sigma_1^2 = \cdots = \sigma_K^2 = 0$, for which the algorithm described in Section 6.7 works and for which $\Lambda_k \equiv 1$ almost surely for all $k \in \{1, \ldots, K\}$. In this case the family

$$\{N_{g,k} \mid g \in G, k \in \{0, \ldots, K\}\}$$

consists of independent, Poisson distributed random variables.

Bernoulli Loss Distribution  Assume that every $L'_{g,k,n}$ is a Bernoulli random variable, i.e.,

$$p := \mathbb{P}[L'_{g,k,n} = 1] = 1 - \mathbb{P}[L'_{g,k,n} = 0]$$

with $p \in [0,1]$ for all $g \in G, k \in \{0, \ldots, K\}$ and $n \in \mathbb{N}$. Then, by (6.9), (??) and (6.11), $\lambda_{k,\nu} = 0$ for every $\nu \in \mathbb{N} \setminus \{C\}$, $\nu_k \in \{0, C\}$ and $\bar{\lambda}_k = \lambda_{k,C}$ for every risk $k \in \{0, \ldots, K\}$. By (??),

$$L'_{g,k} := \sum_{n=1}^{N_{g,k}} L'_{g,k,n} \sim \text{Poisson}(p \lambda_g w_{g,k}).$$
By the Poisson summation property (3.5), we obtain for

$$L' := \sum_{g \in G} \sum_{k=0}^{K} L'_{g,k}(6.107)$$

that $L' \sim \text{Poisson}(p\lambda)$ with

$$\lambda := \sum_{g \in G} \lambda_g \sum_{k=0}^{K} w_{g,k} \cdot \frac{\lambda_g}{\log p} := 1 \text{ by (6.2)}.$$  

Therefore, the distribution of $L = CL'$ satisfies

$$\mathbb{P}(L = l) = \begin{cases} \left(\frac{(p\lambda)^n}{n!}\right) e^{-p\lambda} & \text{if } n := l/C \in \mathbb{N}_0, \\ 0 & \text{otherwise}. \end{cases} \quad (6.109)$$

**Logarithmic Loss Distribution** Assume that every $L'_{g,k,n} \sim \text{Log}(q)$ with $q \in (0,1)$. According to Example 4.27, the compound Poisson sum $L'_{g,k}$ has the distribution NegBin($\alpha_{g,k}, p$) with parameters $p := 1 - q$ and

$$\alpha_{g,k} := -\frac{\lambda_g w_{g,k}}{\log p} \geq 0.$$  

By Lemma 4.24, the sum $L'$ defined in (6.107) has distribution NegBin($\alpha, p$) with $\alpha := -\lambda/\log p$ and $\lambda$ given by (6.108). Therefore, $L = CL'$ satisfies

$$\mathbb{P}(L = l) = \begin{cases} \left(\alpha + \frac{n-1}{n}\right) p^n q^n & \text{if } n := l/C \in \mathbb{N}_0, \\ 0 & \text{otherwise}. \end{cases} \quad (6.109)$$

**General Loss Distributions** Let $Q^s_{g,k} = (q^s_{g,k,\nu})_{\nu \in \mathbb{N}_0}$ be a general distribution for the i.i.d. group losses $(L_{g,k,n})_{n \in \mathbb{N}}$, depending on the group $g \in G$ and the risk $k \in \{0, \ldots, K\}$. Then every $L_{g,k} \sim \text{CPoisson}(\lambda_g w_{g,k}, Q^s_{g,k})$ has a compound Poisson distribution. By (4.54), its generating function is

$$\varphi_{L_{g,k}}(s) = \exp \left( \sum_{\nu \in \mathbb{N}_0} q^s_{g,k,\nu}s^{\nu} - 1 \right) = \varphi_{L_{g,k},1}(s). \quad (6.110)$$

Assume that the sum $\lambda$ of all weighted intensities, given by (6.108), is strictly positive. Define the probability distribution $Q = (q_{\nu})_{\nu \in \mathbb{N}_0}$ by

$$q_{\nu} = \frac{1}{\lambda} \sum_{g \in G} \sum_{k=0}^{K} \lambda_g w_{g,k} q^s_{g,k,\nu}, \quad \nu \in \mathbb{N}_0.$$
Due to independence, the generating function \( \varphi_L \) of the total loss \( L \) is the product of the individual functions from (6.110), hence
\[
\varphi_L(s) = \prod_{g \in G} \prod_{k=0}^{K} \varphi_{L_{g,k}}(s) = \exp \left( \lambda \left( \sum_{\nu \in \mathbb{N}_0} q_{\nu} s^{\nu} - 1 \right) \right),
\]
in particular \( L \sim \text{CPoisson}(\lambda, Q) \) has a compound Poisson distribution. Hence, the distribution of \( L \) can be calculated by the Panjer recursion formula (5.13), i.e.
\[
P[L = l] = \frac{\lambda^l}{l!} \sum_{\nu=1}^{l} \nu q_\nu \ P[L = l - \nu], \quad l \in \mathbb{N},
\]
starting from
\[
P[L = 0] = \varphi_L(0) = e^{\lambda (q_0 - 1)}.
\]

### 6.9.2 Case of Negative Binomial Distribution

Here we assume absence of idiosyncratic risk, meaning that \( \lambda_{0,\nu} = 0 \) for all \( \nu \in \mathbb{N} \) and \( \lambda_0 = 0 \), see (6.9) and (6.11).

**Bernoulli Loss Distribution** Assume that \( L'_{g,k,n} \) is a Bernoulli random variable with risk-dependent distribution, i.e.,
\[
p_k := P[L'_{g,k,n} = 1] = 1 - P[L'_{g,k,n} = 0]
\]
with \( p_k \in [0,1] \) for all \( g \in G, k \in \{1, \ldots, K\} \) and \( n \in \mathbb{N} \). Then, by (6.9) and (6.11), \( \lambda_{k,\nu} = 0 \) for every \( \nu \in \mathbb{N} \setminus \{C\} \) and \( \tilde{\lambda}_k = \lambda_{k,C} \) for every risk \( k \in \{1, \ldots, K\} \). Furthermore, assume that there exist a non-empty \( I \subset \{1, \ldots, K\} \) and \( p \in (0,1) \) such that \( \sigma_k^2 \tilde{\lambda}_k = (1 - p)/p \) for all \( k \in I \) and \( \tilde{\lambda}_k = 0 \) for all \( k \in \{1, \ldots, K\} \setminus I \). By (??) this means \( \nu_k = C \) for all \( k \in I \) and \( \nu_k = 0 \) for all \( k \in \{1, \ldots, K\} \setminus I \). Define
\[
\alpha = \sum_{k \in I} \frac{1}{\sigma_k^2}.
\]
Then (??) simplifies to
\[
\mathbb{E}[s^L] = \left( 1 + \frac{1 - p}{p} (1 - s^C) \right)^{-\alpha} = \left( \frac{p}{1 - q s^C} \right)^{\alpha}
\]
with \( q := 1 - p \), which by (4.47) means that \( L' := L/C \sim \text{NegBin}(\alpha, p) \), hence \( L \) has the distribution given by (6.109).

---

32This section has to be adapted to the new notation and the generalized setting.
General Loss Distributions  We assume that the i.i.d. losses \((L_{g,k,n})_{n \in \mathbb{N}}\) have the same distribution \(Q = (q_\nu)_{\nu \in \mathbb{N}_0}\) for every group \(g \in G\) and every risk \(k \in \{1, \ldots, K\}\). Since \(\mathcal{L}(N_{g,k} | \Lambda_k) \overset{a.s.}{=} \text{Poisson}(\lambda_g w_{g,k} \Lambda_k)\) by Assumption 6.29 and since \((N_{g,k})_{g \in G}\) are conditionally independent given \(\Lambda_k\) by Assumption (6.30), Lemma 3.2 for sums of independent Poisson random variables implies that
\[
\mathcal{L}(N_{(k)} | \Lambda_k) \overset{a.s.}{=} \text{Poisson}(\lambda_{(k)} \Lambda_k)
\]
for every \(k \in \{1, \ldots, K\}\), where
\[
N_{(k)} := \sum_{g \in G} N_{g,k} \quad \text{and} \quad \lambda_{(k)} := \sum_{g \in G} \lambda_g w_{g,k}.
\]
Here \(N_{(k)}\) is the number of defaults in the portfolio caused by risk \(k \in \{1, \ldots, K\}\). Since \(\Lambda_k \sim \Gamma(\alpha_k, \beta_k)\) with \(\alpha_k = \beta_k = 1/\sigma^2_k\) by Assumption 6.34, hence
\[
\lambda_{(k)} \Lambda_k \sim \Gamma(\alpha_k, \beta_k/\lambda_{(k)}),
\]
we get for the unconditional distribution that
\[
N_{(k)} \sim \text{NegBin}(\alpha_k, p_k) \quad \text{with} \quad p_k := \frac{\beta_k/\lambda_{(k)}}{1 + \beta_k/\lambda_{(k)}},
\]
where we use the notation from (4.43). Assuming that \(\lambda_{(k)} \sigma^2_k\) and, therefore, \(p := p_k\) are the same for every risk \(k \in \{1, \ldots, K\}\), then we get for the total number \(N := \sum_{k=1}^K N_{(k)}\) of defaults caused by all the independent risk factors that
\[
N \sim \text{NegBin}(\alpha, p) \quad \text{with} \quad \alpha := \alpha_1 + \cdots + \alpha_K,
\]
see Lemma 4.24. Therefore we have a compound negative binomial distribution for the loss \(L\) given in (6.19), meaning that
\[
L = \sum_{g \in G} \sum_{k=1}^K L_{g,k,n} \overset{d}{=} \sum_{n=1}^N X_n \sim \text{CNegBin}(\alpha, p, Q)
\]
with an i.i.d. sequence \((X_n)_{n \in \mathbb{N}}\) with \(X_n \sim Q\). Therefore, the distribution of \(L\) can be calculated by the Panjer recursion formula (5.15)
\[
\mathbb{P}[L = l] = \frac{1}{\lambda_{(k)} \Gamma(\alpha_k)} \sum_{\nu=0}^l \left( \frac{\nu}{\nu + l - \nu} \right) \mathbb{P}[L = \nu], \quad l \in \mathbb{N},
\]
starting from
\[
\mathbb{P}[L = 0] = \varphi_N(q_0) = \left( \frac{p}{1 - (1 - p)q_0} \right) ^\alpha,
\]
see (5.14).

Exercise 6.54. Consider a logarithmic distribution for the idiosyncratic losses and a Bernoulli distribution for the losses due to the risks \(k \in \{1, \ldots, K\}\), everything in multiples of \(C \in \mathbb{N}\). By combining the above results and putting appropriate conditions on the parameters, show that the portfolio loss \(L\) has a distribution given by (6.109).
7 Risk Measures and Risk Contributions

Knowing the distribution of the portfolio loss $L$ given in (6.19), we can calculate risk measures $\rho(L)$. The quantity $\rho(L)$ can be interpreted as the amount of money that has to be added to the portfolio risk $L$ to make it “acceptable.” For expected shortfall as risk measure, we will also calculate risk contributions in the context of extended CreditRisk$^+$. These contributions indicate the conditional expected loss caused by individual obligors, given a large portfolio loss occurs.

When comparing some of the following definitions with the literature, note that our losses have a positive sign.

7.1 Quantiles and Value-at-Risk

**Definition 7.1.** For a real-valued random variable $X$ and a level $\delta \in (0, 1)$, define the lower $\delta$-quantile of $X$ by

$$ q_\delta(X) = \min \{ x \in \mathbb{R} \mid \mathbb{P}[X \leq x] \geq \delta \} \tag{7.1} $$

and the upper $\delta$-quantile of $X$ by

$$ q^{\delta}(X) = \inf \{ x \in \mathbb{R} \mid \mathbb{P}[X \leq x] > \delta \}. \tag{7.2} $$

Since the distribution function $\mathbb{R} \ni x \mapsto F_X(x) = \mathbb{P}[X \leq x]$ of $X$ is right-continuous, the minimum defining the lower quantile exists. Note that the quantiles depend on $X$ only via the distribution function $F_X$. If we don’t specify lower/upper in the following, we always refer to the lower quantile. Obviously, we always have that $q_\delta(X) \leq q^{\delta}(X)$.

**Exercise 7.2.** Give an example were $q_\delta(X) < q^{\delta}(X)$.

The lower quantile is the smallest threshold such that $q_\delta(X) - X$ is non-negative with probability at least $\delta$. In financial risk management, the lower quantile $q_\delta(X)$ of a loss variable $X$ is called Value-at-Risk (VaR) at level $1 - \delta$ and used as a tool to quantify risk. Rewriting (7.1) as

$$ q_\delta(X) = \min \{ x \in \mathbb{R} \mid \mathbb{P}[X > x] \leq 1 - \delta \}, $$

we see that $q_\delta(X)$ is the smallest threshold which is exceeded by the loss $X$ with probability at most $1 - \delta$.

**Exercise 7.3.** Give an example were $0 < \delta \mapsto q_\delta(X)$ is discontinuous.

The following example shows that small variations of $X$ can lead to substantial jumps of the quantile $q_\delta(X)$, the subsequent lemma gives a condition, when this does not happen.

**Example 7.4.** Consider the unit interval $\Omega = [0, 1]$ equipped with Borel $\sigma$-algebra $\mathcal{B}([0,1])$. Let $\mathbb{P}$ denote the Lebesgue measure restricted to $\mathcal{B}([0,1])$. Given a level $\delta \in (0, 1)$ and $n \in \mathbb{N}$, define $\delta_n = \max\{0, \delta - 1/n\}$ and the Bernoulli random variables $X_n = \mathbb{1}_{[\delta_n, 1]}$ and $X = \mathbb{1}_{[\delta, 1]}$. Then $X_n \downarrow X$ pointwise as $n \to \infty$, $q_\delta(X_n) = 1$ for all $n \in \mathbb{N}$ but $q_\delta(X) = 0$. 

133
Exercise 7.5. Modify Example [7,4] such that $X_n \not	o X$ pointwise as $n \to \infty$, $q^\delta(X_n) = 0$ for all $n \in \mathbb{N}$ but $q^\delta(X) = 1$.

Lemma 7.6. Fix a level $\delta \in (0,1)$. Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of real-valued random variables converging to $X$ in probability, i.e.,
\[ \lim_{n \to \infty} P(|X - X_n| \geq \varepsilon) = 0 \quad \text{for every } \varepsilon > 0. \]

(a) The lower $\delta$-quantiles satisfy
\[ \liminf_{n \to \infty} q^\delta(X_n) \geq q^\delta(X). \]
(b) The upper $\delta$-quantiles satisfy
\[ \limsup_{n \to \infty} q^\delta(X_n) \leq q^\delta(X). \]
(c) If the distribution of $X$ satisfies $P[X \leq x] > \delta$ for all $x > q^\delta(X)$, which is equivalent to $q^\delta(X) = q^\delta(X)$, then
\[ \lim_{n \to \infty} q^\delta(X_n) = q^\delta(X) \quad \text{and} \quad \lim_{n \to \infty} q^\delta(X_n) = q^\delta(X). \]

Proof. (a) If $x < y < q^\delta(X)$, then
\[ P[X_n \leq x] \leq P[X \leq y] + P[|X - X_n| \geq y - x], \]
\[ \longrightarrow 0 \quad \text{as } n \to \infty, \]
\[ \limsup_{n \to \infty} P[X_n \leq x] \leq \gamma := P[X \leq y] < \delta. \]

by the definition of $q^\delta(X)$ in (7.1). Therefore $P[X_n \leq x] \leq (\delta + \gamma)/2 < \delta$ for all sufficiently large $n \in \mathbb{N}$, hence $q^\delta(X_n) \geq x$ for these $n$ and $\liminf_{n \to \infty} q^\delta(X_n) \geq x$. Since $x < q^\delta(X)$ was arbitrary, the lower bound in (a) follows.

(b) The proof is very similar to part (a). If $x > y > q^\delta(X)$, then
\[ P[X_n \leq x] \geq P[X \leq y] - P[|X - X_n| \geq x - y], \]
\[ \longrightarrow 0 \quad \text{as } n \to \infty, \]
\[ \liminf_{n \to \infty} P[X_n \leq x] \geq \gamma := P[X \leq y] > \delta. \]

by the definition of $q^\delta(X)$ in (7.2). Therefore $P[X_n \leq x] \geq (\delta + \gamma)/2 > \delta$ for all sufficiently large $n \in \mathbb{N}$, hence $q^\delta(X_n) \leq x$ for these $n$ and $\limsup_{n \to \infty} q^\delta(X_n) \leq x$. Since $x > q^\delta(X)$ was arbitrary, the upper bound in (b) follows.

(c) follows from (a) and (b).}

If we have an estimate for the Kolmogorov–Smirnov distance of two distributions, then we get bounds for the quantiles of these distributions.
Lemma 7.7 (Quantiles and Kolmogorov–Smirnov metric). Let $X$ and $Y$ be real-valued random variables and denote the Kolmogorov–Smirnov distance of their distributions by $d = d_{KS}(\mathcal{L}(X), \mathcal{L}(Y))$. Then the lower quantiles of $X$ and $Y$ satisfy

(a) $q_{\delta-d}(X) \leq q_{\delta}(Y)$ for every level $\delta \in (d, 1)$ and

(b) $q_{\delta}(Y) \leq q_{\delta+d}(X)$ for every level $\delta \in (0, 1-d)$.

Proof. [a] Given a level $\delta \in (d, 1)$, we use the definition (7.1) of the lower quantile and insert the term $\mathbb{P}[X \leq q_{\delta}(Y)]$, hence

$$\delta \leq \mathbb{P}[Y \leq q_{\delta}(Y)] \leq \mathbb{P}[X \leq q_{\delta}(Y)] + |\mathbb{P}[Y \leq q_{\delta}(Y)] - \mathbb{P}[X \leq q_{\delta}(Y)]|.$$ 

Due to $d_{KS}(\mathcal{L}(X), \mathcal{L}(Y)) = \sup_{x \in \mathbb{R}} |\mathbb{P}[X \leq x] - \mathbb{P}[Y \leq x]|$, this implies

$$\delta \leq \mathbb{P}[X \leq q_{\delta}(Y)] + d,$$

hence $\mathbb{P}[X \leq q_{\delta}(Y)] \geq \delta - d$, therefore $q_{\delta-d}(X) \leq q_{\delta}(Y)$ by (7.1).

[b] Note that the assumptions of the lemma are symmetric in $X$ and $Y$. Applying [a] with $X$ and $Y$ interchanged and $\delta' := \delta + d$ yields $q_{\delta'-d}(Y) \leq q_{\delta'}(X)$, which proves part (b).

Exercise 7.8. In the setting of Lemma 7.7 show the following:

(a) There is a non-trivial example (i.e. one with $\mathcal{L}(X) \neq \mathcal{L}(Y)$) such that $q_{\delta-d}(X) = q_{\delta}(Y) = q_{\delta+d}(X)$ for at least for one level $\delta$.

(b) There is an example with $q_{\delta-d}(X) < q_{\delta}(Y) < q_{\delta+d}(X)$ for at least for one level $\delta$.

(c) Formulate and prove a version of Lemma 7.7 for upper quantiles.

Contrary to its widespread use, VaR is not suitable as a risk measure for two economic reasons. First of all, it does not take into account the size of losses, which occur with probability at most $1-\delta$, meaning that it disregards risks with high effects but low probability. Secondly, VaR is not subadditive in general, i.e., it can happen that $\text{VaR}(X) + \text{VaR}(Y) < \text{VaR}(X + Y)$ for loss variables $X$ and $Y$, meaning that diversification might seem to increase risk when it is measured with VaR, see Example 7.9. Due to these deficiencies, we do not pursue the topic of Value-at-Risk in more detail.

Example 7.9 (VaR is not subadditive). Consider a loan of 100 Euro with default probability $p = 0.8\%$, which leads to a VaR at level 1% of zero. On the other hand, if we consider two independent loans of 50 Euro each with the same default probability $p = 0.8\%$, then the probability of at least one default is $2p - p^2 > 1.59\%$ and thus the VaR at level 1% equals 50 Euro. This means we would prefer the 100 Euro loan as the safer investment, which contradicts the idea of diversification.
7.1.1 Calculation and Smoothing of Lower Quantiles in Extended CreditRisk$^+$

**Remark 7.10** (Calculation of quantiles in extended CreditRisk$^+$). Given a level $\delta \in (0, 1)$, the lower quantile $q_\delta(L)$ of the credit portfolio loss $L$ as given in (6.19) can be calculated in extended CreditRisk$^+$ by adding up the probabilities $P[L = l]$ for $l = 0, 1, 2 \ldots$ until the sum reaches or exceeds $\delta$, see (7.1).

However, this means that $q_\delta(L)$ as a function of $\delta$, when multiplied by the basic loss unit $E$, will jump by this quantity $E$. Since the basic loss unit represents a compromise between precision and computation time, it might not be desirable to have it clearly visible in the output of the extended CreditRisk$^+$ model, hence some smoothing of the quantile might be desirable. If stochastic rounding (see Subsection 6.2.2 for a discussion of this discretisation procedure) has been applied to the individual losses, then somehow “reversing” this step is a legitimate wish.

**Remark 7.11** (Smoothing of lower quantiles in extended CreditRisk$^+$). Let $L$ denote the $\mathbb{N}_0$-valued loss and let $U$ be an independent real-valued random variable, bounded below by $-1$ and such that $E[U] = 0$. Define the smoothed loss $L_s$ by

$$L_s = L + 1_{\{L > 0\}}U. \quad (7.3)$$

Then $L_s$ takes values in $[0, \infty)$ and by independence


hence the smoothing doesn’t change the expectation. Let $(p_n)_{n \in \mathbb{N}_0}$ denote the probability mass function of the $\mathbb{N}_0$-valued loss $L$ and let $U$ be uniformly distributed on the interval $[-\frac{1}{2}, \frac{1}{2}]$. Then the artificially introduced smoothing error $|L - L_s|$ is bounded by $\frac{1}{2}$ and the distribution function of $L_s$ is given by

$$F_{L_s}(x) = \begin{cases} 0 & \text{for } x < 0, \\ p_0 & \text{for } x \in [0, \frac{1}{2}), \\ \sum_{k=0}^{n-1} p_k + p_n(x - n + \frac{1}{2}) & \text{for } x \in [n - \frac{1}{2}, n + \frac{1}{2}) \text{ with } n \in \mathbb{N}. \end{cases}$$

Note that $F_{L_s}$ is continuous on $[0, \infty)$ and has flat parts on $[0, \frac{1}{2})$ and whenever there is an $n \in \mathbb{N}$ with $p_n = 0$. For a level $\delta \in (0, 1)$ the smoothed lower quantile $q_\delta(L_s)$ is given by $q_\delta(L_s) = 0$ if $q_\delta(L) = 0$ and

$$q_\delta(L_s) = q_\delta(L) + \frac{1}{2} - \frac{P[L \leq q_\delta(L)] - \delta}{P[L = q_\delta(L)]} \quad (7.4)$$

if $q_\delta(L) > 0$. Note that the smoothed lower quantile jumps at $\delta = p_0$ if $p_0 > 0$, and that it jumps whenever $q_\delta(L)$ jumps by at least 2. Furthermore, besides the possible atom of size $p_0$ in zero, the distribution of $L_s$ has a piecewise constant density which can never be continuous unless we are in the degenerate case $p_0 = 1$.

**Remark 7.12** (More general smoothing). For a more general smoothing of the lower quantile, we can consider the smoothed loss in (7.3) where $U = V_1 - V_2$.
with independent $V_1, V_2 \sim \text{Beta}(\alpha, \beta)$. Of course, then the formula (7.4) for the smoothed quantile $q_\delta(L_s)$ will be more complicated, but at least in the case $\alpha = \beta = 1$, which means that $V_1, V_2$ are uniformly distributed on the unit interval, it can be done explicitly and $L_s$ has a continuous density on $(0, \infty)$.

7.2 Expected Shortfall

**Definition 7.13.** Let $X$ be a real-valued random variable. Then the expected shortfall of the loss variable $X$ at level $\delta \in (0, 1)$ is defined as

$$ES_\delta[X] = \frac{\mathbb{E}[X \mathbb{1}_{\{X > q_\delta(X)\}}] + q_\delta(X)(\mathbb{P}[X \leq q_\delta(X)] - \delta)}{1 - \delta}$$

(7.5)

with the understanding that $ES_\delta[X] := \infty$ if $\mathbb{E}[X \mathbb{1}_{\{X > q_\delta(X)\}}] = \infty$. (Note that the random variable $X \mathbb{1}_{\{X > q_\delta(X)\}}$ is bounded below by $\min\{0, q_\delta(X)\}$.)

**Remark 7.14.** If $\mathbb{P}[X \leq q_\delta(X)] = \delta$, in particular if the distribution function $R(x) \ni x \mapsto \mathbb{P}[X \leq x]$ of $X$ is also left-continuous at $x = q_\delta(X)$, then (7.5) simplifies to

$$ES_\delta[X] = \mathbb{E}[X | X > q_\delta(X)].$$  

(7.6)

When expected shortfall is taken as a risk measure, then (contrary to VaR) the sizes of large losses exceeding the threshold $q_\delta(X)$ are clearly taken into account by this conditional average. The additional term in (7.5) is necessary to prove the sub-additivity of expected shortfall in Lemma 7.20. The representation (7.6) justifies the name *conditional value-at-risk*, which is also used in the literature.

**Remark 7.15 (Alternative representation of expected shortfall).** Using the observation that

$$\mathbb{E}[X \mathbb{1}_{\{X > q_\delta(X)\}}] = \mathbb{E}[(X - q_\delta(X))^+] + q_\delta(X)\mathbb{P}[X > q_\delta(X)],$$

we obtain from (7.5) the alternative representation

$$ES_\delta[X] = q_\delta(X) + \frac{\mathbb{E}[(X - q_\delta(X))^+]}{1 - \delta}$$

(7.7)

of expected shortfall, which clearly shows that $ES_\delta[X] \geq q_\delta(X)$. See Lemma 7.20(g) for the special property of $q_\delta(X)$ in (7.7).

**Exercise 7.16.** Give an example with $\mathbb{P}[X \leq q_\delta(X)] = \delta$, where the distribution function of $X$ is discontinuous at $q_\delta(X)$.

**Exercise 7.17.** Show that expected shortfall is *law determined* (sometime called *law invariant* in the literature) by representing $ES_\delta[X]$ in terms of the distribution function $F_X$ of $X$.

**Remark 7.18 (Representation of expected shortfall with a density).** Let $X$ be a real-valued random variable. On the underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$ define $f_X: \Omega \to [0, \infty)$ by

$$f_X = \frac{\mathbb{1}_{\{X > q_\delta(X)\}} + \beta X \mathbb{1}_{\{X = q_\delta(X)\}}}{1 - \delta},$$

(7.8)
where the constant $\beta_X$ is given by

$$\beta_X = \begin{cases} \frac{\mathbb{P}(X \leq q_\delta(X)) - \delta}{\mathbb{P}[X = q_\delta(X)]} & \text{if } \mathbb{P}[X = q_\delta(X)] > 0, \\ 0 & \text{otherwise.} \end{cases} \quad (7.9)$$

It follows from the definition of the lower $\delta$-quantile of $X$ in (7.1) that $\beta_X \in [0, 1]$, hence $f_X$ is bounded by $1/(1 - \delta)$. Note that

$$\mathbb{E}[f_X] = \frac{1}{1 - \delta} \left( \mathbb{P}[X > q_\delta(X)] + \beta_X \mathbb{P}[X = q_\delta(X)] \right) = 1, \quad (7.10)$$

hence $f_X$ is a probability density. By the definition of expected shortfall in (7.5),

$$\mathbb{E}[Xf_X] = \mathbb{E}[X1_{\{X > q_\delta(X)\}}] + q_\delta(X) \beta_X \mathbb{P}[X = q_\delta(X)]$$

hence $f_X$ is the expectation of $X$ taken with respect to a probability measure $Q_X$ which has density $f_X$ relative to $\mathbb{P}$. This density raises the probability of the unfavourable event $\{X > q_\delta(X)\}$ by the factor $1/(1 - \delta)$.

### 7.2.1 Calculation of Expected Shortfall in Extended CreditRisk$^+$

**Remark 7.19.** Since the credit portfolio loss $L$, given in (6.19), is a discrete random variable, we have to apply the more complicated definition (7.5). As mentioned in Remark 7.10, the lower quantile $q_\delta(L)$ and $\mathbb{P}[L \leq q_\delta(L)]$ can be calculated using the extended CreditRisk$^+$ algorithm. Furthermore, note that

$$\mathbb{E}[L1_{\{L > q_\delta(L)\}}] = \mathbb{E}[L] - \mathbb{E}[L1_{\{L \leq q_\delta(L)\}}], \quad (7.12)$$

with $\mathbb{E}[L]$ given by (6.70) and

$$\mathbb{E}[L1_{\{L \leq q_\delta(L)\}}] = \sum_{l=1}^{q_\delta(L)} l \mathbb{P}[L = l].$$

If $\mathbb{E}[L] = \infty$, then $\text{ES}_\delta[L] = \infty$. If $\mathbb{E}[L] < \infty$, then the expected shortfall $\text{ES}_\delta[L]$ from (7.5) can be computed numerically using the first terms of the distribution of $L$. Note that the differences in (7.5) and (7.12) can lead to cancellation effects, in particular when $\mathbb{E}[L] \approx \mathbb{E}[L1_{\{L \leq q_\delta(L)\}}]$ for large quantiles.

### 7.2.2 Theoretical Properties of Expected Shortfall

The following lemma lists important properties of expected shortfall. We will need some additional notation. For a level $\delta \in (0, 1)$, let $F_\delta$ denote the set of all probability densities on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ bounded by $1/(1 - \delta)$. For a real-valued random variable $X$ define

$$\mathcal{F}_{\delta,X} := \{ f \in F_\delta \mid \mathbb{E}[X^+ f] < \infty \text{ or } \mathbb{E}[X^- f] < \infty \}, \quad (7.13)$$

138
where $X^+ := \max\{X, 0\}$ and $X^- := \max\{-X, 0\}$ so that $X = X^+ - X^-$. For a density $f \in \mathcal{F}_{\delta,X}$, the expectation $\mathbb{E}[Xf] = \mathbb{E}[X^+f] - \mathbb{E}[X^-f]$ is a well-defined value in $[-\infty, \infty]$. Note that $f_X$ given in (7.8) is in $\mathcal{F}_{\delta}$ and that $X^-f_X$ is a random variable bounded above by $|q_{\delta}(X)|/(1 - \delta)$, hence $\mathbb{E}[X^-f] < \infty$ and therefore $f_X \in \mathcal{F}_{\delta,X}$.

**Lemma 7.20.** Expected shortfall at level $\delta \in (0, 1)$ has, for all real-valued random variables $X$ and $Y$, the following properties:

(a) Positive homogeneity: If $\alpha > 0$, then $ES_\delta[\alpha X] = \alpha ES_\delta[X]$.

(b) Translation (or cash) invariance: If $a \in \mathbb{R}$, then $ES_\delta[X + a] = ES_\delta[X] + a$.

(c) Scenario representation:

(i) $ES_\delta[X] = \sup_{f \in \mathcal{F}_{\delta,X}} \mathbb{E}[Xf]$,

(ii) if $\mathbb{E}[X^+] < \infty$, then $ES_\delta[X] = \sup_{f \in \mathcal{F}_{\delta}} \mathbb{E}[Xf]$.

(d) Sub-additivity: $ES_\delta[X + Y] \leq ES_\delta[X] + ES_\delta[Y]$.

(e) Monotonicity: If $X \leq Y$, then $ES_\delta[X] \leq ES_\delta[Y]$.

(f) Convexity: If $\alpha \in (0, 1)$, then

$$ES_\delta[\alpha X + (1 - \alpha) Y] \leq \alpha ES_\delta[X] + (1 - \alpha) ES_\delta[Y].$$

(g) Minimization property:

$$ES_\delta[X] = \min_{q \in \mathbb{R}} \left( q + \frac{\mathbb{E}[(X - q)^+]}{1 - \delta} \right),$$

and the minimum is attained if and only if $q \in [q_\delta(X), q^\delta(X)]$.

(h) Bounds: For every $q \in \mathbb{R}$,

$$q_\delta(X) \leq ES_\delta[X] \leq q + \frac{\mathbb{E}[(X - q)^+]}{1 - \delta},$$

where the lower bound is an equality if and only if $\mathbb{P}[X \leq q_\delta(X)] = 1$, and the upper bound is an equality if and only if $q \in [q_\delta(X), q^\delta(X)]$.

(i) Quantile representation:

$$ES_\delta[X] = \frac{1}{1 - \delta} \int_{[\delta, 1)} q_\alpha(X) \, du.$$ 

(j) Let $(X_n)_{n \in \mathbb{N}}$ be bounded below, i.e., there exists a constant $a \in [0, \infty)$ such that $X_n \geq -a$ for all $n \in \mathbb{N}$. Then $X := \lim \inf_{n \to \infty} X_n$ satisfies

$$ES_\delta[X] \leq \lim \inf_{n \to \infty} ES_\delta[X_n].$$

(7.14)
(k) Let \((X_n)_{n \in \mathbb{N}}\) be bounded below and converging in probability to a random variable \(X\). Then (7.14) holds, too.

**Corollary 7.21.** For every real-valued random variable \(X\), the map

\[ (0,1) \ni \delta \mapsto \text{ES}_\delta[X] \in \mathbb{R} \cup \{\infty\} \]

is continuous and non-decreasing.

**Proof of Corollary 7.21.** Continuity follows from the quantile representation in part (i). For \(\delta \leq \delta'\) we have \(\mathcal{F}_{\delta,X} \subset \mathcal{F}_{\delta',X}\) which implies \(\text{ES}_\delta[X] \leq \text{ES}_{\delta'}[X]\) by the scenario representation (c). \(\square\)

**Remark 7.22.** A coherent risk measure is defined by monotonicity, positive homogeneity, translation invariance and sub-additivity, cf. Artzner, Delbaen, Eber and Heath [3]. A convex risk measure is defined by monotonicity, translation invariance and convexity, cf. Föllmer and Schied [20]. Note that risk measures are often defined for random variables representing the profit and loss, while in our notation losses have a positive sign. For more details on expected shortfall, see Acerbi and Tasche [1]. The minimization property (g) can be found in Rockafellar and Uryasev [42].

**Remark 7.23.** We excluded the cases \(\alpha = 0\) in (a) and (f), and \(\alpha = 1\) in (f) to avoid expression of the form \(0 \cdot \infty\).

**Remark 7.24.** Concerning the properties in Lemma 7.20, some comments might be useful:

(a) If all losses are scaled, then the risk and the needed capital scales in the same way.

(b) If a constant loss is added, the corresponding amount of capital is needed in addition to make the risk acceptable.

(c) If probabilities of events can be raised by at most the factor \(1/(1-\delta)\), then \(\text{ES}_\delta[X]\) is the worst expected loss possible.

(d), (f) Diversification does not increase the risk.

(e) Smaller losses need less capital.

(g) For an economic interpretation, assume that you can choose an amount \(q\) and enter into a special stop-loss insurance contract such that, whenever your loss \(X\) is above \(q\), you must to pay the fair insurance premium \(\mathbb{E}[(X-q)^+]\) multiplied with the security loading factor \(\frac{1}{1-\delta}\) and receive in return the (possibly smaller, maybe higher) amount \(X - q\) to cover your losses above \(q\). Which \(q\) is optimal for you and how much do you lose given the loss \(X\) exceeds \(q\)? If \(q\) is too high, your deductible is high when \(X > q\) happens; if \(q\) is too small, your premium is high when \(X > q\) happens, the optimal compromise is given by \(q \in [q_\delta(X), q^{\delta}(X)]\).

(i) The quantile representation implies that the expected shortfall varies continuously with the level \(\delta\), contrary to the quantile function \((0,1) \ni \delta \mapsto q_\delta(X)\), which can jump, cf. Exercise 7.3. For discrete distributions like the loss distribution in the extended CreditRisk\(^+\) model, the quantile function has to
jump unless the loss is degenerate. The quantile representation also justifies the name average value-at-risk for expected shortfall.

Remark 7.18 in particular (7.11), shows that equality holds for \( f_X \in F_{\delta,X} \).

Proof of Lemma 7.20. (a) follows from the homogeneity of the expectation and the observation that \( q_\delta(\alpha X) = \alpha q_\delta(X) \).

(b) holds because of the translation invariance of the expectation and the observation that \( q_\delta(X + a) = q_\delta(X) + a \).

(c) Remark 7.18 in particular (7.11), shows that equality holds for \( f_X \in F_{\delta,X} \). Therefore, the supremum is an upper estimate and (i) holds in the case \( ES_\delta[X] = \infty \). If \( ES_\delta[X] < \infty \), then necessarily \( \mathbb{E}[X^+] < \infty \), hence \( F_{\delta,X} = F_\delta \). Consider \( f \in F_\delta \) with \( \mathbb{E}[f - f_X] > -\infty \). We have \( \mathbb{E}[f - f_X] = 0 \), hence

\[
\mathbb{E}[Xf] - \mathbb{E}[Xf_X] = \mathbb{E}[(X - q_\delta(X))(f - f_X)]
\]

\[
= \mathbb{E}[(X - q_\delta(X))(f - f_X)1_{\{X > q_\delta(X)\}}] > 0
\]

\[
+ \mathbb{E}[(X - q_\delta(X))(f - f_X)1_{\{X < q_\delta(X)\}}] \leq 0,
\]

which means that the supremum is identical with \( \mathbb{E}[Xf_X] \).

(d) It suffices to consider the case where \( ES_\delta[X] < \infty \) and \( ES_\delta[Y] < \infty \). Then \( \mathbb{E}[X^+] \), \( \mathbb{E}[Y^+] \) and \( \mathbb{E}[(X + Y)^+] \) are finite and with the the representation from (c), part (ii), we get

\[
ES_\delta[X + Y] = \sup \mathbb{E}[(X + Y)f] = \sup_{f \in F_\delta} \mathbb{E}[Xf] + \sup_{f \in F_\delta} \mathbb{E}[Yf] = ES_\delta[X] + ES_\delta[Y].
\]

(e) Note that \( ES_\delta[X] \leq ES_\delta[X - Y] + ES_\delta[Y] \) by subadditivity (f). For \( Z := X - Y \leq 0 \), we have \( ES_\delta[Z] \leq 0 \) according to (7.5) because \( \mathbb{E}[Z1_{\{Z > q_\delta(Z)\}}] \leq 0 \) and \( q_\delta(Z) \leq 0 \) for a non-positive random variable and \( \mathbb{P}[Z \leq q_\delta(Z)] \geq \delta \) by the definition of the lower quantile.

(f) follows from (e) and (a).

(g) By the alternative representation (7.7), we have equality for \( q = q_\delta(X) \).

Note that \( X - q_\delta(X) = (q - q_\delta(X)) + (X - q) \) for every \( q \in \mathbb{R} \). Consider the case \( q > q_\delta(X) \). Then

\[
(X - q_\delta(X))^+ \leq (q - q_\delta(X))1_{\{X > q_\delta(X)\}} + (X - q)^+
\]

with strict inequality precisely on the event \( \{q_\delta(X) < X < q\} \). Adding to both sides \( q_\delta(X)(1 - \delta) \) and taking expectations, it follows that

\[
q_\delta(X)(1 - \delta) + \mathbb{E}[(X - q_\delta(X))^+] \leq q_\delta(X)(1 - \delta) + (q - q_\delta(X))\mathbb{P}[X > q_\delta(X)] + \mathbb{E}[(X - q)^+] \leq q(1 - \delta) + \mathbb{E}[(X - q)^+]
\]

141
with equality if and only if $\mathbb{P}[q_\delta(X) < X < q] = 0$ and $\mathbb{P}[X \leq q_\delta(X)] = \delta$, which by \eqref{7.1} and \eqref{7.2} is equivalent to $q_\delta(X) < q \leq q(\delta)$. Finally, consider the case $q < q_\delta(X)$. Then

$$(X - q_\delta(X))^+ \leq (q - q_\delta(X))\mathbb{1}_{\{X \geq q_\delta(X)\}} + (X - q)^+$$

with strict inequality precisely on the event $\{q < X < q_\delta(X)\}$. It follows that

$$q_\delta(X)(1 - \delta) + \mathbb{E}[(X - q_\delta(X))^+] \leq q_\delta(X)(1 - \delta) + \mathbb{E}[\mathbb{1}_{\{X \geq q_\delta(X)\}} + (X - q)^+] \leq q(1 - \delta) + \mathbb{E}[(X - q)^+]$$

with equality if and only if $\mathbb{P}[q < X < q_\delta(X)] = 0$ and $\mathbb{P}[X < q_\delta(X)] = \delta$. By the minimizing property of the lower quantile $q_\delta(X)$ defined in \eqref{7.1}, these two conditions cannot be satisfied simultaneously for a $q < q_\delta(X)$.

\[\triangleright\] The lower bound together with the discussion of equality follows directly from the alternative representation \eqref{7.7}, the upper bound follows from \eqref{7.8}.

\[\triangleright\] By extending the probability space if necessary, we may assume the existence of a random variable $U$ on $(\Omega, \mathcal{A}, \mathbb{P})$ which is uniformly distributed on $(0, 1)$, meaning that $\mathbb{P}[U \leq u] = u$ for all $u \in [0, 1]$. Let $q_U(X)$ denote the random quantile $\Omega \ni \omega \mapsto q_U(\omega)(X)$. For every $x \in \mathbb{R}$ and $u \in (0, 1)$ we have

$$q_u(X) \leq x \implies \mathbb{P}[X \leq x] \geq u \quad \text{and} \quad q_u(X) > x \implies \mathbb{P}[X \leq x] < u$$

by the definition \eqref{7.1} of the lower quantile, hence

$$\mathbb{P}[q_U(X) \leq x] = \mathbb{P}[U \leq \mathbb{P}[X \leq x]] = \mathbb{P}[X \leq x]$$

for all $x \in \mathbb{R}$, meaning that $q_U(X)$ and $X$ have the same distribution.

Define $\delta' = \mathbb{P}[X \leq q_\delta(X)]$. Note that $\delta' \geq \delta$ and $q_u(X) = q_\delta(X)$ for every $u \in [\delta, \delta']$. Using the above implications for $x = q_\delta(X)$ shows that $\{U > \delta'\} = \{q_U(X) > q_\delta(X)\}$. Therefore,

$$\int_{[\delta, 1]} q_u(X) \, du = \int_{(\delta', 1]} q_u(X) \, du + \int_{[\delta, \delta']} q_u(X) \, du = \mathbb{E}[q_U(X) \mathbb{1}_{\{U > \delta'\}}] + q_\delta(X)(\delta' - \delta) = \mathbb{E}[X \mathbb{1}_{\{X > q_\delta(X)\}}] + q_\delta(X)(\mathbb{P}[X \leq q_\delta(X)] - \delta).$$

Division by $1 - \delta$ gives the right-hand side of \eqref{7.5}, which is the result.

\[\triangleright\] By translation invariance from \eqref{7.1}, we may assume without loss of generality that every $X_n$ is non-negative. Using the density $f_X$ from \eqref{7.8}, the representation of expected shortfall with the density $f_X$ given in \eqref{7.11}, Fatou’s lemma for $(X_n f_X)_{n \in \mathbb{N}}$ and the scenario representation from \eqref{c}, we get

$$\text{ES}_\delta[X] = \mathbb{E}[X f_X] \leq \liminf_{n \to \infty} \mathbb{E}[X_n f_X] \leq \text{ES}_\delta[X_n].$$
By passing to a subsequence if necessary, we may assume that the sequence $(\varepsilon n)_{n \in \mathbb{N}}$ converges to the limit inferior in (7.14). By passing to a further subsequence if necessary, we may assume that $(X_n)_{n \in \mathbb{N}}$ converges almost surely to $X$. Now, (7.14) follows from (j).

If we have an estimate for the Wasserstein distance of two distributions, see Definition 3.14, then we get bounds for the expected shortfall of these distributions.

**Lemma 7.25** (Expected shortfall and Wasserstein distance). Let $X$ and $Y$ be real-valued, integrable random variables and denote the Wasserstein distance of their distributions by $d_W(\mathcal{L}(X), \mathcal{L}(Y))$. Then the expected shortfall of $X$ and $Y$ satisfies, for every level $\delta \in (0, 1)$,

$$\left| \text{ES}_\delta[X] - \text{ES}_\delta[Y] \right| \leq \frac{d_W(\mathcal{L}(X), \mathcal{L}(Y))}{1 - \delta}. \tag{7.15}$$

**Proof.** Let $(a_i)_{i \in I}$ and $(b_i)_{i \in I}$ be non-empty collections of real numbers, which are bounded below. Define

$$a = \inf_{i \in I} a_i, \quad b = \inf_{i \in I} b_i \quad \text{and} \quad c = \sup_{i \in I} |a_i - b_i|.$$

Then $a_i \leq b_i + c$ for every $i \in I$, hence $a \leq b + c$. Similarly $b \leq a + c$, hence $|a - b| \leq c$. Using this observation and the minimization property from Lemma 7.20(g), it follows that

$$\left| \text{ES}_\delta[X] - \text{ES}_\delta[Y] \right| \leq \frac{1}{1 - \delta} \sup_{q \in \mathbb{R}} \mathbb{E}\left[ (X - q)^+ \right] - \mathbb{E}\left[ (Y - q)^+ \right].$$

For every $q \in \mathbb{R}$, the function $\mathbb{R} \ni x \mapsto (x - q)^+$ is Lipschitz continuous with constant 1, hence (7.15) follows directly from the lower bound (3.16).

### 7.3 Contributions to Expected Shortfall

If the risk and the necessary risk capital for a portfolio loss are calculated with expected shortfall, the question about the risk contributions of individual components of the portfolio arises. Let $\mathcal{L}_0(\mathbb{P}) = \mathcal{L}_0(\Omega, \mathcal{A}, \mathbb{P})$ denote the vector space of all random variables $X$: $\Omega \to \mathbb{R}$ on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Let $\mathcal{L}_1^- (\mathbb{P})$ denote the cone of those $X \in \mathcal{L}_0(\mathbb{P})$, for which the negative part $X^- = \max\{0, -X\}$ is $\mathbb{P}$-integrable. Let $\mathcal{L}_1(\mathbb{P})$ denote the vector space of all $\mathbb{P}$-integrable $X \in \mathcal{L}_0(\mathbb{P})$.

Then, if $Z \in \mathcal{L}_0(\mathbb{P})$ denotes a portfolio loss and $X_1, \ldots, X_n \in \mathcal{L}_1^- (\mathbb{P})$ with $X_1 + \cdots + X_n = Z$ denote the losses of the $n$ subportfolios, we can ask how to allocate the risk capital $\text{ES}_\delta[Z]$ to the $n$ subportfolios in a fair and risk-adequate way.

**Definition 7.26** (Allocation of risk capital by expected shortfall). For a portfolio loss $Z \in \mathcal{L}_0(\mathbb{P})$ and a level $\delta \in (0, 1)$, consider a subportfolio loss $X \in \mathcal{L}_0(\mathbb{P})$
with $X \mathbb{1}_{\{Z \geq q_\delta(Z)\}} \in \mathcal{L}_1^-(\mathbb{P})$. Then the expected shortfall contribution of the subportfolio loss $X$ to $Z$ at level $\delta$ is defined by

$$ES_\delta[X, Z] = \frac{E[X \mathbb{1}_{\{Z > q_\delta(Z)\}}] + \beta_Z E[X \mathbb{1}_{\{Z = q_\delta(Z)\}}]}{1 - \delta} \quad (7.16)$$

with $\beta_Z$ as in (7.9), i.e.

$$\beta_Z = \begin{cases} \frac{\mathbb{P}[Z \leq q_\delta(Z)] - \delta}{\mathbb{P}[Z = q_\delta(Z)]} & \text{if } \mathbb{P}[Z = q_\delta(Z)] > 0, \\ 0 & \text{otherwise}. \end{cases} \quad (7.17)$$

**Remark 7.27.** Note that $ES_\delta[X, Z] = \infty$ is possible and that the condition $X \mathbb{1}_{\{Z \geq q_\delta(Z)\}} \in \mathcal{L}_1^-(\mathbb{P})$ is certainly satisfied for all $X \in \mathcal{L}_1^-(\mathbb{P})$.

**Remark 7.28.** If $\mathbb{P}[Z \leq q_\delta(Z)] = \delta$, then $\beta_Z = 0$ and (7.16) simplifies to

$$ES_\delta[X, Z] = E[X | Z > q_\delta(Z)],$$

cf. Remark 7.14. Therefore, $ES_\delta[X, Z]$ is the conditional expectation of the subportfolio loss $X$ given a large portfolio loss $Z$ occurs. This allocation principle was already presented in [48].

**Remark 7.29.** With the density $f_Z$ defined as in (7.8), we get the representation

$$ES_\delta[X, Z] = E[X f_Z].$$

### 7.3.1 Theoretical Properties

Allocation of risk capital by the expected shortfall principle has a number of good properties. For an axiomatic approach to risk capital allocation, see Kalkbrener [31].

**Lemma 7.30.** Expected shortfall contribution at level $\delta \in (0, 1)$ has, for all $X, Y, Z \in \mathcal{L}_1^-(\mathbb{P})$ and $Z \in \mathcal{L}_0^-(\mathbb{P})$, the following properties:


(b) Diversification: $ES_\delta[X, Z] \leq ES_\delta[X, X]$.

(c) Linearity: For all $\alpha, \beta > 0$,

$$ES_\delta[\alpha X + \beta Y, Z] = \alpha ES_\delta[X, Z] + \beta ES_\delta[Y, Z].$$

If $X, Y \in \mathcal{L}_1^-(\mathbb{P})$, the equality holds for all $\alpha, \beta \in \mathbb{R}$.

(d) Translation (or cash) invariance: If $a \in \mathbb{R}$, then

$$ES_\delta[X + a, Z] = ES_\delta[X, Z] + a.$$

(e) Monotonicity: If $X \leq Y$, then $ES_\delta[X, Z] \leq ES_\delta[Y, Z]$.

(f) Independence: If $X$ and $Z$ are independent, then $ES_\delta[X, Z] = E[X]$. 144
(g) Invariance of portfolio scale: \( \text{ES}_\delta[X, \alpha Z] = \text{ES}_\delta[X, Z] \) for all \( \alpha > 0 \).

(h) Subportfolio continuity: If \( Y \in \mathcal{L}_1(\mathbb{P}) \), then

\[
|\text{ES}_\delta[X, Z] - \text{ES}_\delta[Y, Z]| \leq \text{ES}_\delta[|X - Y|, Z] \leq \frac{\mathbb{E}[|X - Y|]}{1 - \delta}.
\]

(i) Portfolio continuity: Suppose that \( X \in \mathcal{L}_1(\mathbb{P}) \). If \( \mathbb{P}[Z \leq q_\delta(Z)] = \delta \) or if \( X \) is almost surely constant on \( \{Z = q_\delta(Z)\} \), then capital allocation for \( X \) by expected shortfall at level \( \delta \) is continuous at \( Z \), i.e., for every sequence \( (Z_n)_{n\in\mathbb{N}} \subset \mathcal{L}_0(\mathbb{P}) \) converging to \( Z \) in probability,

\[
\lim_{n \to \infty} \text{ES}_\delta[X, Z_n] = \text{ES}_\delta[X, Z]. 
\tag{7.18}
\]

(j) Representation of expected shortfall contribution by directional derivative:

If capital allocation for \( X \in \mathcal{L}_1(\mathbb{P}) \) by expected shortfall is continuous at \( Z \) as specified in part (i), then

\[
\text{ES}_\delta[X, Z] = \lim_{\varepsilon \to 0} \frac{\text{ES}_\delta[Z + \varepsilon X] - \text{ES}_\delta[Z]}{\varepsilon}. 
\tag{7.19}
\]

Remark 7.31. Property (b) shows that \( X \) considered as a subportfolio of any other portfolio \( Z \) does not need more risk capital than on its own, meaning that diversification never increases the risk capital. The proof of (i) is due to the author.

Example 7.32. To see that the continuity in part (i) and the representation as directional derivative from part (j) don’t hold for all \( Z \), consider on \( \Omega = \{0, 1\} \) with \( \mathbb{P}[\{0\}] = \delta \) the random variables given by \( X(\omega) = \omega \) and \( Z(\omega) = 0 \) for all \( \omega \in \Omega \). Define \( Z_\varepsilon = \varepsilon X \). Then \( Z_\varepsilon \to Z \) pointwise as \( \varepsilon \to 0 \). Furthermore, \( \text{ES}_\delta[X, Z] = \mathbb{E}[X] = 1 - \delta \) by independence, \( \text{ES}_\delta[X, Z_\varepsilon] = \text{ES}_\delta[X, X] = \text{ES}_\delta[X] = 1 \) for all \( \varepsilon > 0 \) by scale invariance (g), consistency (a), and Remark 7.14 using \( q_\delta(X) = 0 \). Since \( \text{ES}_\delta[Z] = 0 \) and \( \text{ES}_\delta[Z + \varepsilon X] = \varepsilon \text{ES}_\delta[X] = \varepsilon \), the directional derivative in (7.19) equals \( 1 \neq 1 - \delta = \text{ES}_\delta[X, Z] \).

Proof of Lemma 7.30. (a) By (7.11) and Remark 7.29

\[ \text{ES}_\delta[Z, Z] = \mathbb{E}[Z f_Z] = \text{ES}_\delta[Z]. \]

(b) By Remark 7.29, Lemma 7.20(c) and part (a)

\[ \text{ES}_\delta[X, Z] = \mathbb{E}[X f_Z] \leq \sup_{f \in \mathcal{F}_\delta} \mathbb{E}[X f] = \text{ES}_\delta[X] = \text{ES}_\delta[X, X]. \]

(c), (d) follow from Remark 7.29 and the linearity of the expectation.

(e) follows from Remark 7.29 and \( \text{ES}_\delta[X, Z] = \mathbb{E}[X f_Z] \leq \mathbb{E}[Y f_Z] = \text{ES}_\delta[Y, Z]. \)

(f) By Remark 7.29

\[ \text{ES}_\delta[X, Z] = \mathbb{E}[X f_Z] = \mathbb{E}[X] \mathbb{E}[f_Z] = \mathbb{E}[X]. \]
By Lemma 7.6(a), doesn’t use the additional assumptions given in (i). Note that

\[ E[|X|I_{\{|X|\leq M\}}] = E[|X|I_{\{|X|> M\}}] \leq \varepsilon. \]

By the subportfolio continuity (h), it therefore suffices to prove (7.18) for all bounded \( \varepsilon > 0 \) and \( n_{\varepsilon} \in \mathbb{N} \). By the right-continuity of the distribution function of \( |Z-q| \), there exists \( \eta > 0 \) such that

\[ \mathbb{P}[0 < |Z-q| < 2\eta] \leq \varepsilon. \] (7.20)

Define the abbreviations \( q^- = q - 2\eta \) and \( q^+ = q + 2\eta \). Since \( (Z_n)_{n \in \mathbb{N}} \) converges to \( Z \) in probability, there exists \( n_{\varepsilon} \in \mathbb{N} \) such that

\[ \mathbb{P}[|Z - Z_n| \geq \eta] \leq \varepsilon \quad \text{for all } n \geq n_{\varepsilon}. \] (7.21)

and, by Lemma 7.6(a),

\[ q_n \geq q - \eta \quad \text{for all } n \geq n_{\varepsilon}. \] (7.22)

We will show below by considering the cases \( q_n \leq q + \eta \) and \( q_n > q + \eta \) that

\[ |ES_\delta[X, Z_{\varepsilon}] - ES_\delta[X, Z]| \leq 6\varepsilon \] (7.23)

for every \( n \geq n_{\varepsilon} \). Since \( \varepsilon > 0 \) is arbitrary, (7.23) implies the desired result (7.18).

Note that \( E[|1_A - 1_B|] = \mathbb{P}[A \cap B^c] + \mathbb{P}[A^c \cap B] \) for all \( A, B \in \mathcal{A} \).

**Case I:** The proof of (7.23) for the case \( q_n > q + \eta \) is the easier one and doesn’t use the additional assumptions given in (i). Note that

\[
(1 - \beta Z_n) E[I_{\{Z_n = q_n\}}] = \delta - \mathbb{P}[Z_n < q_n]
\leq \mathbb{P}[Z \leq q] - \mathbb{P}[Z_n < q_n]
\leq \mathbb{P}[Z \leq q, Z_n \geq q_n] \leq \varepsilon
\] (7.24)

by (7.21). By partitioning \( \{Z_n \geq q_n\} \), we obtain

\[
1 - \delta \leq \mathbb{P}[Z_n \geq q_n] = \mathbb{P}[Z > q, Z_n \geq q_n] + \mathbb{P}[Z \leq q, Z_n \geq q_n].
\]

by \( \geq \) by (7.24).
hence $P[A] \geq 1 - \delta - \varepsilon$. Partitioning $\{Z > q\}$ yields

$$1 - \delta \geq P[Z > q] = P[A] + P[Z > q, Z_n < q_n],$$

thus $P[C] \leq \varepsilon$. Finally, using (7.16), $E[X 1_{\{Z=q\}}] = 0$, and $\|X\|_\infty \leq 1 - \delta$,

$$|ES_\delta[X, Z_n] - ES_\delta[X, Z]| \leq (1 - \beta Z_n) E[1_{\{Z_n=q_n\}}] + \varepsilon [1_{\{Z_n\geq q_n\}} - 1_{\{Z>q\}}] \leq 3\varepsilon,$$

which proves (7.23) for the case $q_n > q + \eta$.

**Case II:** We will now prove estimate (7.23) in the case $q_n \leq q + \eta$ for the two different assumptions given in Lemma 7.30(i). Define $E = \{Z > q, Z_n \leq q_n\}$ and $F = \{Z \leq q, Z_n > q_n\}$. Note that

$$P[E] = P[q < Z < q^+, Z_n \leq q_n] + P[Z \geq q^+, Z_n \leq q_n] \leq 2\varepsilon. \quad (7.25)$$

**Case II(a):** Let the assumption $P[Z \leq q] = \delta$ be satisfied. By partitioning $\{Z_n \leq q_n\}$, we obtain

$$\delta \leq P[Z_n \leq q_n] = P[Z \leq q, Z_n \leq q_n] + P[E],$$

hence $P[D] \geq \delta - 2\varepsilon$ by (7.25). Partitioning $\{Z \leq q\}$ yields

$$\delta = P[Z \leq q] = P[D] + P[F],$$

thus $P[D] \leq \delta$ and $P[F] \leq 2\varepsilon$. Furthermore, using (7.25)

$$\beta Z_n E[1_{\{Z_n=q_n\}}] = P[Z_n \leq q_n] - \delta = P[D] + P[F] - \delta \leq 2\varepsilon. \quad (7.26)$$

Finally, using (7.16), $E[X 1_{\{Z=q\}}] = 0$, and $\|X\|_\infty \leq 1 - \delta$,

$$|ES_\delta[X, Z_n] - ES_\delta[X, Z]| \leq \beta Z_n E[1_{\{Z_n=q_n\}}] + \varepsilon [1_{\{Z_n\geq q_n\}} - 1_{\{Z>q\}}] \leq 6\varepsilon,$$

which proves (7.23) for the Case II(a).

**Case II(b):** Let now $X$ be a.s. constant on $\{Z = q\}$. Then $E[X 1_{\{Z=q\}}] = 0$ implies $E[|X| 1_{\{Z=q, Z_n=q_n\}}] = 0$ and $E[|X| 1_{\{Z=q, Z_n>q_n\}}] = 0$. Therefore,

$$\frac{E[|X| 1_{\{Z_n=q_n\}}]}{1 - \delta} \leq P[Z \neq q, Z_n = q_n] \leq P[0 < |Z - q| < 2\eta] + P[|Z - q| \geq 2\eta, Z_n = q_n] \leq 2\varepsilon$$

$$\leq \varepsilon \text{ by (7.26) and (7.22)} \quad (7.27)$$
\[
\frac{\mathbb{E}[|X|_F]}{1 - \delta} \leq \mathbb{P}[Z < q, Z_n > q_n] + \mathbb{P}[Z \leq q^{-}, Z_n > q_n] \leq 2\varepsilon. \tag{7.28}
\]

Using (7.16), \(\beta_n Z_n \in [0, 1], \mathbb{E}[X|_{Z=q}] = 0\), and \(\|X\|_{\infty} \leq 1 - \delta\),

\[
\left| \text{ES}_\delta[X, Z_n] - \text{ES}_\delta[X, Z] \right| \leq \frac{\mathbb{E}[|X|_{\{Z_n=q_n\}}]}{1 - \delta} + \frac{\mathbb{E}[|X|_E]}{1 - \delta} + \frac{\mathbb{E}[|X|_F]}{1 - \delta} \leq 6\varepsilon,
\]

which proves (7.23) for the Case II(b).

\[\square\]

Let \(\varepsilon > 0\). By consistency (a), diversification (b) and linearity (c),

\[
\text{ES}_\delta[Z + \varepsilon X] = \text{ES}_\delta[Z + \varepsilon X, Z + \varepsilon X] \geq \text{ES}_\delta[Z + \varepsilon X, Z] = \text{ES}_\delta[Z] + \varepsilon \text{ES}_\delta[X, Z],
\]

hence

\[
\frac{\text{ES}_\delta[Z + \varepsilon X] - \text{ES}_\delta[Z]}{\varepsilon} \geq \text{ES}_\delta[X, Z].
\]

Similarly,

\[
\text{ES}_\delta[Z] = \text{ES}_\delta[Z, Z] \geq \text{ES}_\delta[Z, Z + \varepsilon X] = \text{ES}_\delta[Z + \varepsilon X] - \varepsilon \text{ES}_\delta[X, Z + \varepsilon X],
\]

hence

\[
\text{ES}_\delta[X, Z + \varepsilon X] \geq \frac{\text{ES}_\delta[Z + \varepsilon X] - \text{ES}_\delta[Z]}{\varepsilon}.
\]

Since capital allocation for \(X\) by expected shortfall is assumed to be continuous at \(Z\),

\[
\text{ES}_\delta[X, Z] = \lim_{\varepsilon \searrow 0} \frac{\text{ES}_\delta[Z + \varepsilon X] - \text{ES}_\delta[Z]}{\varepsilon}.
\]

If \(\varepsilon \nearrow 0\), apply this result for \(\varepsilon' = -\varepsilon\) and \(X' = -X\) and use \(-\text{ES}_\delta[X', Z] = \text{ES}_\delta[X, Z]\) to obtain (7.19). \(\square\)

### 7.3.2 Calculation of Risk Contributions in Extended CreditRisk+$^+$

\[\text{ES}_\delta[L_g,i,k, L] = \frac{\mathbb{E}[L_{g,i,k} \mathbb{I}_{\{L > q_g(L)\}}]}{1 - \delta} + \beta L \mathbb{E}[L_{g,i,k} \mathbb{I}_{\{L = q_g(L)\}}] \tag{7.29}
\]
as contribution attributed to obligor \(i \in \{1, \ldots, m\}\) due to group \(g \in G_i\) and risk \(k \in \{0, \ldots, K\}\) to the expected shortfall \(\text{ES}_\delta[L]\). Since \(L\) has a discrete

\[\text{ES}_\delta[L_g,i,k, L] = \frac{\mathbb{E}[L_{g,i,k} \mathbb{I}_{\{L > q_g(L)\}}]}{1 - \delta} + \beta L \mathbb{E}[L_{g,i,k} \mathbb{I}_{\{L = q_g(L)\}}] \tag{7.29}
\]
as contribution attributed to obligor \(i \in \{1, \ldots, m\}\) due to group \(g \in G_i\) and risk \(k \in \{0, \ldots, K\}\) to the expected shortfall \(\text{ES}_\delta[L]\). Since \(L\) has a discrete
and, for every risk $k \in \{1, \ldots, K\}$, 

$$
E[L_{g,i,k}\mathbb{1}_{\{L=q_{g}(L)\}}] = \lambda_g w_{g,k} \sum_{\nu=1}^{l} E[L_{g,i,k,1}\mathbb{1}_{\{L=q_{g}(L)\}}] E[\Lambda_{k}\mathbb{1}_{\{L=l-\nu\}}].
$$

(7.31)

**Remark 7.34.** The algorithm presented in Section 6.7 calculates in a numerically stable way the quantities $E[L = l - \nu]$ and $E[\Lambda_{k}\mathbb{1}_{\{L=l-\nu\}}]$ used in the above lemma. Note that the coefficients $(b_{k,l})_{l \in \mathbb{N}_0}$, which originate from the expansion of the logarithm and are given by (??), (??) and (??), are the same for both expressions. For $E[\Lambda_{k}\mathbb{1}_{\{L=l-\nu\}}]$ the coefficients $(c_{l})_{l \in \mathbb{N}_0}$ given by (??) and (??) and well as the coefficients $(d_{n})_{n \in \mathbb{N}_0}$ given by (??) and (??) have to be recalculated.

**Remark 7.35.** For every obligor $i \in \{1, \ldots, m\}$, every group $g \in G_i$, every risk $k \in \{0, \ldots, K\}$ and every group loss $\nu \in \mathbb{N}_0$, we get from Assumption 6.16 

$$
E[L_{g,i,k,1}\mathbb{1}_{\{L=q_{g}(L)\}}] = \sum_{\mu=(\mu_{j})_{j \in g} \in \mathbb{N}_0^g} \mu_j \mathbb{P}[L_{g,i,k,1} = \mu_{j} \text{ for all } j \in g] \text{ if } \|\mu\|_{1}=\nu, \text{ by (6.24)}.
$$

(7.32)

which can be calculated directly from the input data in a numerically stable way, because only non-negative numbers are multiplied and added.

(a) In the case $g = \{i\}$, which is in particular the case in the classical Credit-Risk+ model (cf. Remarks 6.6 and 6.49), the result (7.32) simplifies to

$$
E[L_{g,i,k,1}\mathbb{1}_{\{L=q_{g}(L)\}}] = \nu q_{g,k,\nu}.
$$

(7.33)
(b) If the group loss $\nu$ is attributed in a deterministic way to its members as described in Example 6.18, then
\[ E[L_{g,i,k,1} 1\{L_{g,k,1} = \nu\}] = h_{g,i,k}(\nu) q_{g,k,\nu}^i. \] (7.34)

(c) Note that by the linearity of the expectation,
\[ \nu q_{g,k,\nu}^i = E[L_{g,k,1} 1\{L_{g,k,1} = \nu\}] = \sum_{i \in g} E[L_{g,i,k,1} 1\{L_{g,k,1} = \nu\}]. \] (7.35)

If $(L_{g,i,k,1})_{i \in g}$ are exchangeable (in particular when they are i.i.d.), then all expectations on the right-hand side of (7.35) are equal and
\[ E[L_{g,i,k,1} 1\{L_{g,k,1} = \nu\}] = \nu |g| q_{g,k,\nu}^i \] for all $i \in g$. (7.36)

**Proof of Lemma 7.33.** Fix a risk $k \in \{0, \ldots, K\}$, an obligor $i \in \{1, \ldots, m\}$ and a group $g \in G_i$ which contains $i$. Recall that $L_{g,k} = \sum_{n=1}^{N_{g,k}} L_{g,k,n}$ by (6.15) and note that $L_{g,k} = 0$ if $N_{g,k} = 0$. Furthermore, if $L = l$, then no single loss can exceed $l$, in particular it suffices to consider $l \geq 1$. Define $M = L - L_{g,k}$ as the sum of all losses coming not from group $g$ due to risk $k$. For every $\mu \in \mathbb{N}$ and $n \in \{1, \ldots, \mu\}$ define
\[ M_{\mu,n} = \sum_{r=1, r \neq n}^{\mu} L_{g,k,r} \]
as the sum of the first $\mu$ losses of group $g$ due to risk $k$, omitting the $n$th loss. Then
\[ E[L_{g,i,k} 1\{L = l\}] = \sum_{\mu=1}^{\infty} E \left[ \sum_{n=1}^{\mu} L_{g,i,k,n} 1\{L = l, N_{g,k} = \mu\} \right] \]
\[ = \sum_{\mu=1}^{\infty} \sum_{n=1}^{\mu} \sum_{\nu=1}^{l} E[L_{g,i,k,n} 1\{L = l, N_{g,k} = \mu, L_{g,k,n} = \nu\}] \]
\[ = \{M + M_{\mu,n} + L_{g,k,n} = l, N_{g,k} = \mu\} \] (7.37)

It follows from Assumption 6.16 that the random vector $(L_{g,i,k,n})_{i \in g}$ together with the sum $L_{g,k,n}$ of its components given in (6.14) is independent jointly from $M$, $M_{\mu,n}$ and $N_{g,k}$, hence
\[ E[L_{g,i,k,n} 1\{M = l, M_{\mu,n} + L_{g,k,n} = \nu, N_{g,k} = \mu\}] = E[L_{g,i,k,n} 1\{L_{g,k,n} = \nu\}] \mathbb{P}[M = l - \nu, N_{g,k} = \mu]. \] (7.38)

By Assumption 6.16 the loss vectors $(L_{g,i,k,n})_{i \in g}$ and $(L_{g,i,k,1})_{i \in g}$ have the same distribution, hence we can replace $n$ by 1 in the expectation on the right-hand side of (7.38). The same assumption implies that $M_{\mu,n}$ is independent from
\((M, N_{g,k})\) and that \(M_{1,\ldots,\mu,\mu}\) are identically distributed, hence, for every \(n \in \{1, \ldots, \mu\}\),

\[
P[M + M_{\mu,n} = l - \nu, N_{g,k} = \mu] = P[M + M_{\mu,\mu} = l - \nu, N_{g,k} = \mu].
\] (7.39)

Consider now the case \(k \in \{1, \ldots, K\}\). By the conditional independence from Assumption 6.30 and the conditional Poisson distribution from Assumption 6.29,

\[
P[M + M_{\mu,\mu} = l - \nu, N_{g,k} = \mu] = \mathbb{E}[\Lambda_k \mathbb{P}[M + M_{\mu,\mu} = l - \nu, N_{g,k} = \mu - 1 | \Lambda_k]],
\]

where we used

\[
P[N_{g,k} = \mu | \Lambda_k] \xrightarrow{a.s.} \frac{(\lambda g w_{g,k} \Lambda_k)^\mu}{\mu!} \exp(-\lambda g w_{g,k} \Lambda_k)
\]

\[
= \frac{\lambda g w_{g,k} \Lambda_k}{\mu} P[N_{g,k} = \mu - 1 | \Lambda_k].
\]

Substituting (7.38), (7.39) and (7.40) into (7.37) and noting that the sum over \(n \in \{1, \ldots, \mu\}\) cancels with the denominator \(\mu\), we obtain

\[
\mathbb{E}[L_{g,i,k} 1_{\{L=l\}}] = \lambda g w_{g,k} \sum_{\mu=1}^{\infty} \sum_{\nu=1}^{l} \mathbb{E}[L_{g,i,k,1} 1_{\{L_{g,k,1}=\nu\}}] \mathbb{E}[\Lambda_k \mathbb{1}_{\{L=l-\nu, N_{g,k}=\mu-1\}}]
\]

\[
= \lambda g w_{g,k} \sum_{\nu=1}^{l} \mathbb{E}[L_{g,i,k,1} 1_{\{L_{g,k,1}=\nu\}}] \mathbb{E}[\Lambda_k \mathbb{1}_{\{L=l-\nu\}}].
\]

For the case \(k = 0\) the calculation in the last paragraph is easier and left as an exercise.

Remark 7.36. As we constructed \(N_{i,k}\) as conditionally Poisson distributed random variable, we have that \(P(N_{i,k} \geq n) > 0\) for every \(n \in \mathbb{N}\). Hence it is possible that the risk contributions become greater than the maximal exposure.

8 Application to Operational Risk

8.1 The Regulatory Framework

The quantification of operational risk of financial institutions gained importance due to the regulatory prescriptions in column 1 of the Basel II accord for capital requirements [7]. A profound introduction to the mathematical modelling of operational risk can be found in McNeil, Frey and Embrechts [38, Chap. 10].

Operational losses occur frequently with low impact, but there are also rare events with high impact such that their arrival can cause serious trouble for a
financial institution. Famous events that are subject of operational risk are the bankruptcy of the British Barings Bank in 1995 and the terror attacks on the World Trade Center in New York City on September 11th, 2001.

Another characteristic that distinguishes operational risk from credit or market risk is that there is no chance for profit. Operational risk comes along with any process of a bank’s business despite of all efforts to avoid malfunctions.

The Basel committee allows three approaches with increasing complexity to quantify a bank’s operational risk, namely

- the basic indicator approach (BIA),
- the standardized approach (SA),
- the advanced measurement approach (AMA).

The basic indicator approach and the standardized approach provide exact formulae how to calculate the regulatory capital. In the advanced measurement approach, the risk capital is determined by an internal risk measurement system that needs to fulfill various criteria. For exact definitions of these approaches and the criteria for an advanced measurement approach, consult the Basel committee’s final document [7].

In these lecture notes we will focus on the mathematical and numerical machinery to model and aggregate operational risk for an advanced measurement approach. We therefore adopt the extended CreditRisk+ methodology from Section 6 to this new kind of risk. The application of this methodology to the problem of operational risk seems even more appropriate than the application to credit risk: the modelling error caused by the approximation of a sum of Bernoulli random variables by a Poisson random variable (cf. Theorem 3.23) is not an issue for operational risk modelling, because the a priori use of Poisson distributions in the setting of operational loss occurrences is more natural.

In the standardized approach eight business lines are defined:

1. Corporate finance
2. Trading & sales
3. Retail banking
4. Commercial banking
5. Payment & settlement
6. Agency services
7. Asset management
8. Retail brokerage

These business lines are supposed to serve as categories for an advanced measurement approach as well. Furthermore, seven loss event types have to be distinguished in an advanced measurement approach [7, p. 147]:

1. Internal fraud,
2. External fraud,
3. Employment practices & workplace safety,
4. Clients, products & business practice,
(5) Damage to physical assets,
(6) Business disruption & system failures,
(7) Execution, delivery & process management.

For an exact definition and the subcategories, we refer to the Basel committee’s final document [7, Annex 9]. A bank that once has proceeded to an advanced approach will not be allowed to revert to a simpler one without supervisory approval—unless it does not fulfil the necessary criteria anymore and is therefore forced to revert to a simpler approach in at least some of its operations.

Nonetheless, the motivation for an advanced measurement approach is obvious. The formulae prescribed in the basic indicator and the standardized approach use externally given values that can in general hardly reflect the very structure of the respective financial institution. Internal models are potentially capable of detecting risk and allocating risk capital where it is really required. An advanced measurement approach can therefore lead to reduced risk capital requirements. But the regulatory capital can not be reduced arbitrarily as an initial floor of 75% of the risk capital required by the standardized approach is dictated [8, p. 6].

8.2 Characteristics of Operational Risk Data

Whereas credit loss data of various kind and market data for nearly any desirable security and rate is available for a long time horizon, there is only little data available on operational risk. The estimation of frequent losses can probably be managed using internal data, but for rare events causing high losses often external data has to be used. Another difficulty of the statistical analysis of the available data is a reporting bias coming from the increasing awareness of the importance of collecting operational risk data.

Moscadelli [40] did an in-depth statistical analysis of operational loss data and found several characteristics. In his analysis, estimated severity distributions are heavy-tailed. Light- and medium-tailed distributions as the Gumbel distribution or the lognormal distribution model the body of the severity distribution fairly well but fail to fit the tails of the loss severities. The modelling of operational risk therefore calls for the application of extreme value theory, cf. [16, 19] and [38, Chap. 7].

Moscadelli [40] even found that six business lines (among the eight mentioned before) yield estimations of distributions with infinite mean. This fact has to be considered if one wants to calculate risk measures (one would have problems explaining expected shortfall with infinite mean of severities). In this case one will have to use quantile-based risk measures such as value-at-risk. As long as the data allows us, we will use coherent risk measures such as expected shortfall in order to calculate risk contributions as a basis for the allocation of risk capital to business lines as well as to operational loss event types.
8.3 Application of the Extended CreditRisk\(^+\) Methodology

We want to keep the notation in full generality for the case that one wants to model more than the eight business lines and seven event types mentioned in the Basel committee’s final paper. For the application to operational risk, we basically have to reinterpret the notation used in Section [3]

- The number \( m \) of obligors turns into the number of business lines, \( m = 8 \) for the ones given in (8.1) is an appropriate choice.
- The basic loss unit \( E \) stays the same. The Basel committee allows the negligence of operational losses below 10 000 Euro when reporting for internal data collection [21, p. 149], which motivates the choice \( E = 10 000 \).
- The number \( K \) of non-idiosyncratic risk factors turns into the number of loss types; \( K = 7 \) for the types given above is a possible choice, but a finer subdivision is possible.
- The numbers \( \sigma_k^2 > 0 \) denote the relative variance of occurrences of losses of type \( k \in \{1, \ldots, K\} \).
- The collection \( G \) contains the subsets of all business lines which can incur a loss due to the same event.

For every group \( g \in G \) of business lines, we need

- the (one year) intensity \( \lambda_g \geq 0 \) for being hit by an operational loss event,
- the conditional probability \( w_{g,0} \in [0,1] \) for an idiosyncratic operational loss event not to belong to the types in \( \{1, \ldots, K\} \), of course \( w_{g,0} = 0 \) is a possible choice,
- the conditional probabilities \( w_{g,k} \in [0,1] \) for an operational loss event to be of type \( k \in \{1, \ldots, K\} \),
- the multivariate probability distribution \( Q_{g,k} = (q_{g,k,\mu})_{\mu \in \mathbb{N}_0^g} \) on \( \mathbb{N}_0^g \) describing the severity of the stochastic losses of the business lines \( i \in g \) in multiples of the basic loss unit \( E \) in case an operational loss event of type \( k \in \{0, \ldots, K\} \) hits the group \( g \) of business lines.

The stochastic losses (within a year) get the following interpretation:

- \( L_{g,k} \) given by (6.15) is the operational loss of the group \( g \in G \) of business lines due to common losses of type \( k \in \{0, \ldots, K\} \),
- \( L_{i,k} \) given by (6.21) is the operational loss of business line \( i \in \{1, \ldots, m\} \) due to loss type \( k \in \{0, \ldots, K\} \),
- \( L_i \) given in (6.23) is the total operational loss of business line \( i \in \{1, \ldots, m\} \), and
- \( L \) given by (6.19) is the total operational loss of the bank.

\(^{34}\)This section has to be adapted to the new notation and the generalized setting.
With the extended CreditRisk\(^+\) methodology it is therefore possible to quantify operational risk consistent with the Basel committee’s requirements for an advanced measurement approach. The probability-generating function of the total operational loss can be evaluated in a numerically stable way and in the case of finite-mean severity distributions, we can use expected shortfall and even achieve a risk capital allocation to business lines as well as operational loss event types. Our approach does not need any Monte Carlo simulations and therefore proposes a quick analysis of the bank’s operational risk situation without the stochastic simulation error.

9 Acknowledgments

The original version of these lecture notes was based on a course given at the Université Louis Pasteur (Strasbourg, France) in February 2004 and on workshops given for the Austrian Central Bank (ÖNB, www.oenb.at) in Vienna on March 19, March 26 and July 12, 2004. The lecture notes were carefully worked out and typesetted by Mag. Severin Resch, who also implemented the numerically stable algorithm for the standard version of CreditRisk\(^+\) in Java. Ms. Sabine Wimmer prepared a revision to include stochastic exposures for individual obligors, together with DI Richard Warnung she added the risk contributions. Further important additions were made by DI Warnung in cooperation with the author (proof with Stein–Chen method, coherence of expected shortfall, capital allocation, risk groups and multivariate loss distributions, applications to operational risk). Project-oriented financial support for all three co-workers through the Austrian Central Bank and financial support for DI Warnung through the project Mathematics and Credit Risk funded by the Vienna Science and Technology Fund (WWTF, www.wwtf.at) is gratefully acknowledged.

In April 2006, a version of these lecture notes including risk groups and applications to operational risk was presented by DI Warnung and the author at the Workshop on Risk Analysis and Management, preceding the First Conference on Advanced Mathematical Methods in Finance in Side, Antalya, Turkey. Travel support by the European Sciences Foundation through the AMaMeF Programme is gratefully acknowledged.

Additional research concerning a generalization of Panjer’s recursion and numerically stable risk aggregation, leading to the paper [21], was done jointly with Dr. Stefan Gerhold and DI Richard Warnung. The papers [22] and [21] are part of R. Warnung’s Ph. D. thesis [56], both have won the Best Paper Award of the Faculty of Mathematics and Geoinformation of the Vienna University of Technology. Since January 2010, there is ongoing joint research with Dipl.-Math. Cordelia Rudolph on generalizations of Panjer’s recursion for dependent claim numbers [46] as well as on approximations of Poisson mixture models via Panjer’s recursion [45] leading to her Ph. D. thesis [44]. Since Autumn 2012, DI Karin Hirhager and Jonas Hirz (MSc) work jointly with the author on conditional quantiles, conditional weighted expected shortfall and applications to capital
allocation to extend the results presented in Section 7. This joint research was financially supported by the Christian Doppler Research Association (CDG). The authors gratefully acknowledge the fruitful collaboration and support by the Bank Austria, the Oesterreichische Kontrollbank AG (OeKB) and the Austrian Federal Financing Agency (ÖBFA) through CDG and the Christian Doppler Laboratory for Portfolio Risk Management (PRisMa Lab).

During the summer term 2013, these lecture notes were expanded and used for part of the course on Credit Risk Models and Derivatives at the Vienna University of Technology. With further extensions, in particular to treat the total variation and the Wasserstein metric together with their applications to quantiles (Lemma 7.7) and expected shortfall (Lemma 7.25), and with additional exercise problems, these lecture notes are currently used again for the same course in 2014.

List of Abbreviated Distributions

- Beta(·, ·), beta distribution, see Definition 2.5
- Bin(1, ·), Bernoulli distribution
- Bin(·, ·), binomial distribution, see (2.9)
- Gamma(·, ·), gamma distribution, see Subsection 4.3
- Log(·), univariate logarithmic distribution, see Example 4.4
- MPoisson(·, ·, ·), multivariate Poisson distribution, see Definition 3.40
- Multinomial(1, ·), multivariate Bernoulli distribution, see Example 4.5
- Multinomial(·, ·), multinomial distribution, see Example 4.16
- Poisson(·), Poisson distribution, see Definition 3.1

Recent Changes

March 2021

- Remark 5.2 added for clarity.

April 2021

- Subsection 4.2 extended.
- List of abbreviated distributions started.

References


M. Moscadelli, The modelling of operational risk: experience with the analysis of the data collected by the Basel Committee, Temi di Discussione del


Index

actuarial model, 5
additive set function, 23
advanced measurement approach, 152
aggregation property, 46
multinomial distribution, 51
multivariate Bernoulli distribution, 15
multivariate binomial distribution, 67
multivariate logarithmic distribution, 64
negative multinomial distribution, 66
allocation of risk capital by expected shortfall, 143
approximation by Poisson distribution, 25
Ars conjectandi, 5
asset value model, 5
assumption default cause intensities, 105
extended CreditRisk+ conditional independence of default numbers, 106
default numbers, 106
distribution of default numbers, 105
gamma-distributed risk factors, 107
independence of risk factors and scenario, 106
normalization of default causes, 108
risk group, 92
susceptibility, 92
automobile collision insurance, 91
average value-at-risk, see expected shortfall

\text{B}(\alpha, \beta), \text{see beta function}
Banach space of signed measures, 23
basic indicator approach, 152
basic loss units, 91
Bernoulli
Jacob, 5
Nicolaus, 5
Bernoulli distribution
multivariate, 50
aggregation property, 45
and binomial distribution, 66
and negative binomial distribution, 64
and Poisson distribution, 62
covariance, 45
definition, 44
expectation, 45
generating function, 44, 47
permutation property, 45
variance, 45
univariate, 49
definition, 43
expectation, 4
generating function, 43, 47
variance, 4
Bernoulli mixture model
conditional independence, 7
construction, 8
covariance, 8
expectation, 8
joint distribution, 7
one-factor expectation, 15
homogeneous, 14
variance, 15
uniform, 9
variance, 9
Bernoulli model, 5
general mixture, 7
one-factor mixture, 14
Berry–Esseen theorem, 31
Beta(\alpha, \beta), \text{see beta distribution}
beta distribution, 136
biased probability measure, 13
definition, 11
expectation, 12
moments, 12
variance, 12
beta function, 11, 53
beta-binomial distribution, 11, 12
factorial moments, 12
initial value, 12
recursion formula, 12
BetaBin(\alpha, \beta, m), \text{see beta-binomial distribution}
beta function, 11, 53
biased probability measure beta distribution, 13
definition, 13
gamma distribution, 13
Bin(1, p), \text{see Bernoulli distribution}
Bin(m, p), \text{see binomial distribution}
binomial distribution, 10
generating function, 49
initial value, 78
numerical instability, 78
Panjer class, 78
probability mass function, 6
summation property, 51
Blackwell–Girshick equation, 70
bounds
expected shortfall, 139
business lines
operational risk, 152
calibration
of Poisson distribution, 17
cancellation of significant digits, 54, 76, 86
Cartesian product, 36
cash invariance
expected shortfall, 139
of contribution to expected shortfall, 144
categorical distribution, see Bernoulli distribution, multivariate
χ²-distribution, 55
Cholesky decomposition, 110
claims reserving, 91
CNegBin(α,p,Q), see compound negative binomial distribution
collateral, 115
combinatorial interpretation
negative binomial distribution, 56
negative multinomial distribution, 64
common Poisson shock model, 93
comonotonic losses
in risk group, 101, 103
competition group, 93
composition
of generating functions, 58
compound binomial distribution
with multivariate Bernoulli distribution, 59
compound negative binomial distribution, 69
definition, 58
generating function, 59
with multivariate Bernoulli distribution, 64
compound Poisson distribution, 69
conditional
conditional covariance, 70
conditional expectation, 70
conditional variance, 70
covariance, 70
definition, 58
expectation, 70
generating function, 58
summation property, 61
variance, 70
with logarithmic distribution, 59
with multivariate Bernoulli distribution, 62
conditional compound Poisson distribution
conditional covariance, 70
conditional expectation, 70
conditional variance, 70
conditional covariance, see covariance, conditional
conditional compound Poisson distribution, 70
of random sum, 70
conditional expectation
conditional compound Poisson distribution, 70
of random sum, 70
with independent random variables, 8
conditional independence
Bernoulli mixture model, 7
multivariate Poisson mixture model, 38
of default numbers, 105
conditional Poisson distribution
of default numbers, 105
conditional value-at-risk, see expected shortfall
conditional variance, 40
conditional compound Poisson distribution, 70
of random sum, 70
consistence
of contribution to expected shortfall, 144
of default intensities, 116
continuity
quantile, 134
contribution
to expected shortfall, 143
as conditional expectation, 144
as directional derivative, 145
cash invariance, 144
consistency, 144
diversification, 144
in extended CreditRisk⁺, 148
independence, 144
linearity, 144
monotonicity
portfolio continuity
scale invariance
subportfolio continuity
translation invariance

convexity
expected shortfall
convolution
gamma distributions
Corr(·, ·), see correlation
correlation
stochastic rounding
coupling method
Cov(·, ·), see covariance, conditional
covariance
Bernoulli distribution
multivariate
Bernoulli mixture model
compound Poisson distribution
conditional
default cause intensities
multinomial distribution
multivariate binomial distribution
multivariate logarithmic distribution
multivariate Poisson distribution
negative multinomial distribution
random sum
stochastic rounding
via generating function
covariance matrix
decomposition
with random matrix
Cox process
C\textsc{Poisson}(λ, Q), see compound Poisson distribution
credit guarantee
credit risk model
actuarial
asset value
intensity-based
reduced form
structural
Credit Suisse First Boston
Credit\textsc{Risk}⁷, see also extended Credit\textsc{Risk}⁷
historical remark
decomposition
Cholesky
covariance matrix
deductible
default cause intensities
covariance
expectation
lower bounds
negative correlation
structure
variance
default causes
examples
hierarchical order
default numbers
conditional independence
distribution
expectation
default probability
risk group
consistency
density
beta distribution
gamma distribution
dependence scenario
dependent defaults
in risk group
dilogarithm
directional derivative
do of contribution to expected shortfall
distance, see metric
distribution
beta, see beta distribution
beta-binomial, see beta-binomial distribution
binomial, see binomial distribution
Erlang
exponential
gamma, see gamma distribution
geometric
logarithmic, see logarithmic distribution
multinomial, see multinomial distribution
multivariate binomial, see multivariate binomial distribution
multivariate Poisson
Poisson distribution, multivariate
negative binomial, see negative binomial distribution
compound
negative multinomial, see
negative multinomial distribution
Panjer class, 71
Poisson, see Poisson distribution
compound, 58, 69
trunced, 71
uniform, see uniform distribution
diversification
of contribution to expected shortfall, 144
Dudley metric, see Wasserstein metric
E[·], see expectation
equation
Blackwell–Girshick, 70
Wald, 70
Erlang distribution, see also gamma distribution, 63
Euler, 74, 82
event types
operational risk, 152
exercise, 132
beta-binomial distribution, 12
factorial moments, 12
characterization of Poisson(λ), 32
comparison of bounds, 30
complete cancellation, 76
compound Poisson distribution, 69
computation of conditional expectation, 8
construction of general
Bernoulli mixture model, 8
multivariate Poisson mixture model, 132
covariance for mixture distribution, 12
extended logarithmic distribution, 89
ExtLog(2, 1), 89
historical comment, 62
implementation of extended CreditRisk+, 128
jump at lower quantile, 137
Kolmogorov–Smirnov distance
estimates for quantiles, 135
law determined, 137
logarithmic distribution, 48
lower and upper quantiles, 133
moments from factorial moments
multivariate, 49
univariate, 13
moments of beta distribution, 12
multinomial distribution, 51
multivariate beta function, 11
multivariate binomial distribution, 67
multivariate logarithmic distribution, 64
multivariate Poisson distribution, 37
negative binomial distribution, 57
negative multinomial distribution, 66
normal approximation, 31
Panjer(a, b, 0) class, 72
Poisson approximation, 30
quantile function, 133
Stein equation, 33
summation property
compound distributions, 61
summation property of multinomial distribution, 52
multivariate binomial distribution, 68
summation property of negative multinomial distribution, 66
total variation metric, 22
variational characterization, 50
total variation norm, 23
upper quantile, 134
variance of sums, 9
Wasserstein metric, 20
characterization of convergence, 23
scaling property, 23
expectation, see also conditional expectation
Bernoulli distribution, 6
multivariate, 45
Bernoulli mixture model, 8
general, 8
one-factor, 15
uniform, 10
beta distribution, 12
beta-binomial distribution, 12
compound Poisson distribution, 70
default cause intensity, 107
default numbers, 115
gamma distribution, 54
logarithmic distribution
multivariate, 64
univariate, 48
multinomial distribution, 51, 68
negative binomial distribution, 56
negative multinomial distribution, 66
Poisson distribution, 16
multivariate, 37
random sum, 70
stochastic rounding, 95
via generating function, 47
infinite value, 48
expected shortfall
alternative representation, 137
as conditional expectation, 137
as function of level, 140
bounds, 139
cash invariance, 139
contribution
as conditional expectation, 144
as directional derivative, 145
cash invariance, 144
consistency, 144
definition, 143
diversification, 144
in extended CreditRisk+, 148
independence, 144
linearity, 144
monotonicity, 144
portfolio continuity, 145
scale invariance, 145
subportfolio continuity, 145
translation invariance, 144
convexity, 139
definition, 139
estimate with Wasserstein distance, 143
Fatou property, 140
in extended CreditRisk+, 138
minimization property, 139, 143
economic interpretation, 140
monotonicity, 139
positive homogeneity, 139
quantile representation, 139
representation with density, 137
scenario representation, 139
sub-additivity, 139
theoretical properties, 138
translation invariance, 139
exponential distribution, see also
gamma distribution, 63
exponential moments
gamma distribution, 54
extended CreditRisk+
basic loss units, 91
conditional independence of default numbers, 106
cumulative Poisson intensity, 98
default cause, 91
default cause intensity, 105
default numbers
conditional independence, 105
default probability, 92
dependence scenario, 92
derived parameters, 97
distribution of default numbers, 105
106
expected shortfall, 138
gamma-distributed risk factors, 107
group loss distribution, 97
group loss vector, 101
independence of risk factors and scenario, 106
input parameters, 91
list of extensions, 90
multi-period extension, 91
negative correlation of default cause intensities, 108
number of obligors, 91
number of periods, 91
Poisson intensity for group, 97
probabilistic assumptions, 101
quantile calculation, 136
quantile smoothing, 136
risk factor, 91
risk group, 92
assumption, 92
stochastic losses, 92
susceptibility, 92
value-at-risk, 136
smoothing, 136
extended logarithmic distribution, see
logarithmic distribution, extended
extended negative binomial distribution, see
negative binomial distribution, extended
extended Panjer recursion, see
Panjer recursion
ExtLog(k,p), see
logarithmic distribution, extended
ExtNegBin(α,k,p), see negative binomial distribution, extended
factorial moment
beta-binomial distribution, 12
calculating moments
multivariate, 49
<table>
<thead>
<tr>
<th>Term</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>univariate, logarithmic distribution</td>
<td>13</td>
</tr>
<tr>
<td>multivariate</td>
<td>64</td>
</tr>
<tr>
<td>univariate</td>
<td>48</td>
</tr>
<tr>
<td>negative binomial distribution</td>
<td>57</td>
</tr>
<tr>
<td>negative multinomial distribution</td>
<td>66</td>
</tr>
<tr>
<td>Poisson distribution</td>
<td>16</td>
</tr>
<tr>
<td>via generating function</td>
<td>47</td>
</tr>
<tr>
<td>factorial moment generating function</td>
<td>43</td>
</tr>
<tr>
<td>Fatou property</td>
<td></td>
</tr>
<tr>
<td>expected shortfall</td>
<td>140</td>
</tr>
<tr>
<td>formal power series</td>
<td>74</td>
</tr>
<tr>
<td>Fortet–Mourier metric, see Wasserstein metric</td>
<td>46</td>
</tr>
<tr>
<td>function</td>
<td></td>
</tr>
<tr>
<td>hypergeometric</td>
<td>44</td>
</tr>
<tr>
<td>functional equation</td>
<td></td>
</tr>
<tr>
<td>gamma function</td>
<td>10</td>
</tr>
<tr>
<td>Gamma(α, β), see gamma distribution</td>
<td></td>
</tr>
<tr>
<td>Γ(α), see gamma function</td>
<td></td>
</tr>
<tr>
<td>gamma distribution</td>
<td></td>
</tr>
<tr>
<td>biased probability measure</td>
<td>55</td>
</tr>
<tr>
<td>convolution</td>
<td>53</td>
</tr>
<tr>
<td>definition</td>
<td>53</td>
</tr>
<tr>
<td>density</td>
<td>53</td>
</tr>
<tr>
<td>expectation</td>
<td>54</td>
</tr>
<tr>
<td>exponential moments</td>
<td>54</td>
</tr>
<tr>
<td>infinitely divisible</td>
<td>53</td>
</tr>
<tr>
<td>Laplace transform</td>
<td>54</td>
</tr>
<tr>
<td>moment</td>
<td>54</td>
</tr>
<tr>
<td>peculiar relation</td>
<td>54</td>
</tr>
<tr>
<td>summation property</td>
<td>53</td>
</tr>
<tr>
<td>variance</td>
<td>54</td>
</tr>
<tr>
<td>gamma function</td>
<td>10</td>
</tr>
<tr>
<td>functional equation</td>
<td></td>
</tr>
<tr>
<td>gamma-mixed Poisson distribution</td>
<td>55</td>
</tr>
<tr>
<td>generating function</td>
<td></td>
</tr>
<tr>
<td>Bernoulli distribution</td>
<td>43</td>
</tr>
<tr>
<td>multivariate</td>
<td>47</td>
</tr>
<tr>
<td>binomial distribution</td>
<td>49</td>
</tr>
<tr>
<td>multivariate</td>
<td>67</td>
</tr>
<tr>
<td>composition</td>
<td>58</td>
</tr>
<tr>
<td>compound negative binomial distribution</td>
<td>59</td>
</tr>
<tr>
<td>compound Poisson distribution</td>
<td>58</td>
</tr>
<tr>
<td>conditional</td>
<td></td>
</tr>
<tr>
<td>compound Poisson distribution</td>
<td>68</td>
</tr>
<tr>
<td>Poisson distribution</td>
<td></td>
</tr>
<tr>
<td>covariance</td>
<td>48</td>
</tr>
<tr>
<td>definition</td>
<td>43</td>
</tr>
<tr>
<td>expected value</td>
<td>47</td>
</tr>
<tr>
<td>infinite value</td>
<td>48</td>
</tr>
<tr>
<td>factorial moment</td>
<td>43</td>
</tr>
<tr>
<td>linear transformation</td>
<td>45</td>
</tr>
<tr>
<td>logarithmic distribution</td>
<td></td>
</tr>
<tr>
<td>multivariate</td>
<td>63</td>
</tr>
<tr>
<td>univariate</td>
<td>44</td>
</tr>
<tr>
<td>multinomial distribution</td>
<td>50</td>
</tr>
<tr>
<td>negative binomial distribution</td>
<td>56</td>
</tr>
<tr>
<td>derivatives</td>
<td>59</td>
</tr>
<tr>
<td>extended</td>
<td>61</td>
</tr>
<tr>
<td>negative multinomial distribution</td>
<td></td>
</tr>
<tr>
<td>multivariate</td>
<td>65</td>
</tr>
<tr>
<td>Poisson distribution</td>
<td>43</td>
</tr>
<tr>
<td>multivariate</td>
<td>50</td>
</tr>
<tr>
<td>random sum</td>
<td>58</td>
</tr>
<tr>
<td>variance</td>
<td>58</td>
</tr>
<tr>
<td>geometric distribution, see negative binomial distribution</td>
<td>144</td>
</tr>
<tr>
<td>group loss</td>
<td></td>
</tr>
<tr>
<td>in extended CreditRisk⁺ distribution</td>
<td>97</td>
</tr>
<tr>
<td>group loss vector</td>
<td></td>
</tr>
<tr>
<td>in extended CreditRisk⁺, 101</td>
<td></td>
</tr>
<tr>
<td>guarantor</td>
<td>104</td>
</tr>
<tr>
<td>Hadamard factorization theorem</td>
<td>17</td>
</tr>
<tr>
<td>health insurance contract</td>
<td>91</td>
</tr>
<tr>
<td>hindering defaults</td>
<td>93</td>
</tr>
<tr>
<td>historical remark</td>
<td></td>
</tr>
<tr>
<td>CreditRisk⁺</td>
<td></td>
</tr>
<tr>
<td>numerical stability</td>
<td>125</td>
</tr>
<tr>
<td>negative binomial distribution</td>
<td>59</td>
</tr>
<tr>
<td>Panjer recursion</td>
<td>52</td>
</tr>
<tr>
<td>hypergeometric function</td>
<td>44</td>
</tr>
<tr>
<td>independence</td>
<td></td>
</tr>
<tr>
<td>and contribution to expected shortfall</td>
<td>144</td>
</tr>
<tr>
<td>of risk factors and scenario</td>
<td>106</td>
</tr>
<tr>
<td>pairwise</td>
<td>58</td>
</tr>
<tr>
<td>stochastic rounding</td>
<td>55</td>
</tr>
<tr>
<td>independent losses</td>
<td>97</td>
</tr>
<tr>
<td>in risk group</td>
<td>102</td>
</tr>
<tr>
<td>infinite divisibility</td>
<td></td>
</tr>
<tr>
<td>gamma distribution</td>
<td>53</td>
</tr>
<tr>
<td>Poisson distribution</td>
<td></td>
</tr>
<tr>
<td>multivariate</td>
<td>37</td>
</tr>
<tr>
<td>univariate</td>
<td>17</td>
</tr>
<tr>
<td>initial value for Panjer’s recursion</td>
<td></td>
</tr>
<tr>
<td>binomial distribution</td>
<td>78</td>
</tr>
</tbody>
</table>
logarithmic distribution, 79
negative binomial distribution, 77
Poisson distribution, 77

intensity-based model, 5

joint distribution
Bernoulli mixture model, 7
multivariate Poisson mixture model, 38

Kantorovich $D_{1,1}$ metric, see Wasserstein metric
Kolmogorov–Smirnov metric, definition, 19
estimate for quantiles, 135

Laplace transform
gamma distribution, 54
law determined, 137
law invariance, see law determined linearity
of contribution to expected shortfall, 144
Lipschitz constant, 21
definition, 20
Log($p$), see logarithmic distribution, univariate

logarithmic distribution
extended
definition, 81
ExtLog(2, 1), 89
numerical instability, 82
Panjer class, 82
weighted convolution, 89
multivariate
aggregation property, 64
covariance, 64
definition, 63
expectation, 64
factorial moment, 64
generating function, 63
permutation property, 64
probability mass function, 63
variance, 64
normalising factor, 14
univariate, 69
and Poisson distribution, 59
definition, 14
expectation, 18
factorial moment, 14
generating function, 14
initial value, 79

numerical stability, 79
Panjer class, 79
variance, 48

loss event types
operational risk, 152
lower bound
default cause intensity, 105
lower quantile
definition, 133
semicontinuity, 134

marginal distribution
stochastic rounding, 95
marginal distribution of
multinomial, 67
one-dimensional, 51
multivariate binomial distribution, 67
negative multinomial, 66
MBin($m, p_1, \ldots, p_d$), see multivariate binomial distribution

measure
biased, 13
method
Stein–Chen, 31
metric
for probability measures, 18
Kolmogorov–Smirnov, 19, 27, 28
Wasserstein, 20
metric space
separable, 21
minimization property
of expected shortfall, 139, 143
economic interpretation, 140

mixture distribution
in extended CreditRisk$^+$, 98
MLog($p_1, \ldots, p_d$), see logarithmic distribution, multivariate

model
Bernoulli, 5
general mixture, 7
one-factor mixture, 14
Poisson, 15
general multivariate mixture, see Poisson mixture model, multivariate
one-factor mixture, 12

moment
beta distribution, 12
factorial, 17
beta-binomial distribution, 12
logarithmic distribution, 64
negative binomial distribution, 57
negative multinomial distribution, 66
Poisson distribution, 16
univariate logarithmic distribution, 68
from factorial moments
multivariate, 49
univariate, 13
gamma distribution, 54
moment-generating function
gamma distribution, 54
monotonicity
equipped shortfall, 144
of contribution to expected shortfall, 144
MPoisson\((G, \lambda, m)\), see Poisson distribution, multivariate
multi-period extension
extended CreditRisk\+, 91
Multinomial(1, ·), see Bernoulli distribution, multivariate
Multinomial\((m, \cdot)\), see multinomial distribution
multinomial coefficient, 51, 65
multinomial distribution
aggregation property, 51
covariance, 51
definition, 50
expectation, 51
generating function, 50
marginal distribution, 51, 67
Multinomial\((1, p_1, \ldots, p_d)\), 44
permutation property, 51
probability mass function, 50
summation property, 51
variance, 51
multinomial theorem, 50
multiplication theorem for generating functions, 49
multivariate beta function, 11
multivariate binomial distribution
aggregation property, 68
covariance, 68
definition, 66
expectation, 68
generating function, 67
marginal distribution, 67
permutation property, 67
summation property, 68
variance, 68
multivariate distribution
Bernoulli, see Bernoulli distribution, multivariate
binomial, see multinomial distribution
logarithmic, see logarithmic distribution, multivariate
multivariate binomial, see multivariate binomial distribution
negative multinomial, see negative multinomial distribution
Poisson, see Poisson distribution, multivariate
multivariate Poisson distribution, see Poisson distribution, multivariate
multivariate Poisson mixture model, see Poisson mixture model, multivariate
mutual singularity
total variation metric, 19
negative binomial distribution, 69
combinatorial interpretation, 69
compound, 58, 69
generating function, 59
summation property, 61
compound Poisson, 59
definition, 55
expectation, 66
extended definition, 79
ExtNegBin\((\alpha - 1, 1, 1)\), 87
generating function, 81
numerical instability, 86
numerically stable algorithm, 88
Panjer class, 81
weighted convolution, 86
factorial moment, 57
generating function, 56
initial value, 77
interpretation, 59
numerical stability, 77
Panjer class, 77
summation property, 57
variance, 66
negative correlation
of default cause intensities, 108
negative multinomial distribution
aggregation property, 66
combinatorial interpretation, 65
covariance, 66
definition, 65
expectation, 66
factorial moment, 66
generating function, 65
marginal distribution, 66
permutation property, 66
probability mass function, 65
summation property, 65
variance, 66

NegBin(\(\alpha, p\)), see negative binomial distribution

NegMult(\(\alpha, p_1, \ldots, p_d\)), see negative multinomial distribution

normal total variation, 23
normal approximation, 31
normalization of default causes, 108
sufficient conditions, 108
number of defaults notation, 99
numerical instability
binomial distribution, 78
elementary, 86
extended logarithmic distribution, 82
extended negative binomial distribution, 86
numerical stability
logarithmic distribution, 79
negative binomial distribution, 77
Poisson distribution, 77
numerical underflow, 78
numerically stable algorithm
extended negative binomial distribution, 88

obligor with guarantee, 104
one-factor Bernoulli mixture model, see Bernoulli model
operational risk, 151
advanced measurement approach, 152
basic indicator approach, 152
business lines, 152
data, 153
loss event types, 152
regulatory framework, 151
standardized approach, 152

Panjer(\(a, b, k\)), see Panjer class
Panjer class
binomial distribution, 78
characterisation, 71
definition, 71
extended logarithmic distribution
definition, 81
extended negative binomial distribution
logarithmic distribution, 79
negative binomial distribution, 77
Poisson distribution, 77
truncation, 71
uniqueness, 74
Panjer recursion, 72
choice of \(c_n\), 73
computational speed-up, 72
for truncated distribution, 85
generalization, 83
proof, 83
historical remark, 82
proof, 83
starting value, 78
technical assumption, 72
partial order on \(N^d\), 72
partition of unity, 95
partitions of a set, 13
permutation property, 46
multinomial distribution, 51
multivariate Bernoulli distribution, 45
multivariate binomial distribution, 67
multivariate logarithmic distribution, 64
negative multinomial distribution, 66
personal liability insurance, 91
\(\varphi_X\), see generating function
Poisson(\(\cdot\)), see Poisson distribution
Poisson approximation, 25 30
heuristics, 51
Poisson distribution, 15 69
calibration, 17
characterization, 52
compound, 58 69
covariance, 70
elementary, 69
expectation, 70
generating function, 58
summation property, 61
variance, 70
conditional
of default numbers, 105
expectation, 16
factorial moment, 16
gamma-mixture, 55
generating function, 43
infinite divisibility, 15
initial value, 77
multivariate
  compound Poisson distribution, 61
covariance, 37
definition, 36
expectation, 37
generating function, 50
independent components, 37
infinite divisibility, 37
summation property, 37
numerical stability, 77
Panjer class, 77
probability mass function, 15
Raikov’s theorem, 17
summation property
  univariate, 16
  variance, 16
Poisson intensity
  for risk group in extended CreditRisk+, 97
  in extended CreditRisk+, 98
Poisson mixture model
  multivariate, 38
  construction, 38
  covariance, 41
  expectation, 39
  variance, 42
  one-factor mixture, 42
  uniform, 42
Poisson model, 15
Poisson summation theorem
  multivariate, 37
  proof, 16
  univariate, 16
  proof, 49
portfolio continuity
  of contribution to expected shortfall, 145
positive homogeneity
  expected shortfall, 139
positive semi-definite matrix, 109
probability distribution, see distribution
probability mass function
  binomial distribution, 6
  logarithmic distribution
    multivariate, 63
    univariate, 44
  multinomial distribution, 50
  negative binomial distribution, 55
  negative multinomial distribution, 65
  Poisson distribution, 15
  truncated distribution, 71
probability measure
  biased, 13
  metric, 18
  pseudometric, 18
probability-generating function, see generating function
pseudo risk factor, 107
pseudometric
  for probability measures, 18
quantile, 133
  continuity, 134
  estimate with Kolmogorov–Smirnov metric, 135
  extended CreditRisk+ calculation, 136
  smoothing, 136
lower
  definition, 133
  semicontinuity, 134
upper
  definition, 133
  semicontinuity, 134
quantile representation
  expected shortfall, 139
Raikov’s theorem, 17
random default probabilities, 7
random matrix, 41
random Poisson intensities, 38, 39, 42
random sum, 57
  conditional covariance, 70
  conditional expectation, 70
  conditional variance, 70
  expectation, 70
  generating function, 58
  variance, 70
reduced form model, 5
references, 156
regulatory framework
  operational risk, 151
risk capital
  allocation by expected shortfall, 143
risk factor, 91
  gamma distribution, 107
  pseudo, 107
risk group, 92
  assumption, 92
default probability, 115, 116
smoothing, 136
not subadditive, 135
example, 135
\(\text{Var}(\cdot|\cdot)\), see variance, conditional
\(\text{Var}(\cdot)\), see variance

Bernoulli distribution, 6
multivariate, 45
Bernoulli mixture model, 9
general, 9
one-factor, 15
uniform, 10
beta distribution, 12
beta-binomial distribution, 12
compound Poisson distribution, 70
conditional, 70
conditional compound Poisson distribution, 70
default cause intensities, 107
gamma distribution, 54
multinomial distribution, 51 68
multivariate logarithmic distribution, 64
negative binomial distribution, 56
negative multinomial distribution, 66
of sum of random variables, 9
Poisson distribution, 16
multivariate, 57
random sum, 70
stochastic rounding, 97
univariate logarithmic distribution, 48
via generating function, 48

variational characterization
total variation metric, 30
Vasershtein metric, see Wasserstein metric

Wald’s equation, 70
Wasserstein metric, 22 25
and weak convergence, 21 23
bounds, 21
definition, 20
estimate for expected shortfall, 143
scaling property, 23
well-defined, 20
weak convergence
and Wasserstein metric, 21 28
weighted convolution
extended logarithmic distribution, 89
extended negative binomial distribution, 86

Zorn’s lemma, 23