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 Statistics and Mathematical Methods in Economics
Vienna University of Technology (TU Wien)
Wiedner Hauptstraße 8–10/E105-1
A-1040 Vienna, Austria

2004–2025

schmock@fam.tuwien.ac.at

Version of Notes: June 19, 2025, 9:00 UTC
(Incomplete revision)

Updates on fam.tuwien.ac.at/∼schmock/notes/ExtensionsCreditRiskPlus.pdf

Contents

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 $(N_1, \dots, N_m)$ be a random vector of Bernoulli[1] 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

$$p_i := \mathbb{P}[N_i = 1] \in [0, 1] \tag{2.1}$$

denote the probability of default[2] of obligor $i \in \{1, \dots, m\}$ within a specified time frame, typically one year, and

$$N := \sum_{i=1}^{m} N_i \tag{2.2}$$

[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 $p_1, \dots, p_m$ in practice can be a challenging task.

be the random variable representing the total number of defaults. Obviously

$$\mathbb{E}[N_i] = p_i \tag{2.3}$$

and, using $N_i^2 = N_i$,

$$\operatorname{Var}(N_i) = \mathbb{E}\big[N_i^2\big] - (\mathbb{E}[N_i])^2 \overset{(2.3)}{=} p_i(1-p_i). \tag{2.4}$$

The expected number of defaults (within one period) is given by

$$\mathbb{E}[N] = \sum_{i=1}^{m} \mathbb{E}[N_i] \overset{(2.3)}{=} \sum_{i=1}^{m} p_i. \tag{2.5}$$

If $N_1, \dots, N_m$ are uncorrelated, meaning that

$$\operatorname{Cov}(N_i, N_j) = \mathbb{E}\big[(N_i - \mathbb{E}[N_i])(N_j - \mathbb{E}[N_j])\big] = 0$$

for all $i, j \in \{1, \dots, m\}$ with $i \neq j$, then the variance of N is

$$\operatorname{Var}(N) = \sum_{i=1}^{m} \operatorname{Var}(N_i) \overset{(2.4)}{=} \sum_{i=1}^{m} p_i(1-p_i); \tag{2.6}$$

see (2.19) and Exercise 2.4 for a more general formula.

The probability of exactly $n \in \{0, \dots, m\}$ defaults is the sum over the probabilities of all possible subsets of n obligors defaulting during the period, i.e.

$$\mathbb{P}[N = n] = \sum_{\substack{I \subseteq \{1,\dots,m\} \\ |I| = n}} \mathbb{P}[N_i = 1 \text{ for } i \in I,\, N_i = 0 \text{ for } i \in \{1, \dots, m\} \setminus I]. \tag{2.7}$$

Moreover, if the $N_1, \dots, N_m$ are independent (which is a strong assumption), then

$$\mathbb{P}[N = n] = \sum_{\substack{I \subseteq \{1,\dots,m\} \\ |I| = n}} \biggl(\prod_{i \in I} p_i\biggr) \prod_{i \in \{1,\dots,m\} \setminus I} (1 - p_i). \tag{2.8}$$

For $n = 100$ defaults in a portfolio of $m = 1000$ obligors, assuming pairwise different $p_1, \dots, p_m$, 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, suitable approximations,[3] and more sophisticated algorithms.[4]

In the special case of equal default probabilities for all obligors, i.e.

$$p_1 = \dots = p_m =: p,$$

[3] See e.g. Theorem 3.23 below.

[4] See e.g. Exercise 5.4, Theorem 5.16.

the distribution in (2.8) simplifies to

$$\mathbb{P}[N=n]=\binom{m}{n}p^n(1-p)^{m-n},\quad n\in\{0,\dots,m\},\tag{2.9}$$

which is the binomial distribution[5] $\mathrm{Bin}(m,p)$ 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, $N_1,\dots,N_m$ are typically dependent on each other!

Exercise 2.1 (Factorial moments of the binomial distributions)**.** Show for $N\sim\mathrm{Bin}(m,p)$ with $m\in\mathbb{N}$ and $p\in[0,1]$ (and the convention $0^0:=1$) that

$$\mathbb{E}\biggl[\prod_{k=0}^{l-1}(N-k)\biggr]=p^l\prod_{k=0}^{l-1}(m-k),\qquad l\in\mathbb{N}_0.\tag{2.10}$$

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 precisely known today. The uncertainty is expressed by introducing a distribution for them as follows.

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

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 $i\in\{1,\dots,m\}$, i.e.

$$\mathbb{P}[N_i=1\,|\,P_1,\dots,P_m]\overset{\text{a.s.}}{=}\mathbb{P}[N_i=1\,|\,P_i]\overset{\text{a.s.}}{=}P_i.\tag{2.11}$$

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

$$\begin{aligned}\mathbb{P}[N_1=n_1,\dots,N_m=n_m\,|\,P_1,\dots,P_m]&\overset{\text{a.s.}}{=}\prod_{i=1}^m\mathbb{P}[N_i=n_i\,|\,P_1,\dots,P_m]\\&\overset{\text{a.s.}}{=}\prod_{i=1}^m P_i^{n_i}(1-P_i)^{1-n_i},\end{aligned}\tag{2.12}$$

[5] The name refers to the binomial theorem, which can be used to show that the terms in (2.9) add up to one.

where we used (2.11), 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 (2.12). Note that, for every $i \in \{1,\dots,m\}$,

$$\sum_{n_i \in \{0,1\}} P_i^{n_i}(1-P_i)^{1-n_i} = 1. \tag{2.13}$$

In the unconditional case, the joint distribution is obtained by integration of (2.12) over all possible values of $(P_1,\dots,P_m)$ with respect to the distribution function F, or formally

$$\begin{aligned} \mathbb{P}[N_1 = n_1,\dots,N_m = n_m] &= \mathbb{E}\biggl[\prod_{i=1}^m P_i^{n_i}(1-P_i)^{1-n_i}\biggr] \\ &= \int_{[0,1]^m} \prod_{i=1}^m p_i^{n_i}(1-p_i)^{1-n_i}\, F(\mathrm{d}p_1,\dots,\mathrm{d}p_m). \end{aligned} \tag{2.14}$$

If $I \subseteq \{1,\dots,m\}$ is any subset of obligors, then iterative summation over all $n_i \in \{0,1\}$ with $i \in \{1,\dots,m\} \setminus I$ using (2.13) implies that

$$\mathbb{P}[N_i = n_i \text{ for all } i \in I\,] = \mathbb{E}\biggl[\prod_{i\in I} P_i^{n_i}(1-P_i)^{1-n_i}\biggr]. \tag{2.15}$$

Exercise 2.2 (Conditional expectation involving independent random variables)**.** Let $(\Omega,\mathcal{A},\mathbb{P})$ be a probability space, $\mathcal{B} \subseteq \mathcal{A}$ a sub-σ-algebra, $(S_1,\mathcal{S}_1)$ and $(S_2,\mathcal{S}_2)$ measurable spaces, $X\colon \Omega \to S_1$ and $Y\colon \Omega \to S_2$ random variables, and $F\colon S_1 \times S_2 \to \mathbb{R}$ an $\mathcal{S}_1 \otimes \mathcal{S}_2$-measurable function, which is bounded or non-negative. Suppose that X is $\mathcal{B}$-measurable and Y is independent of $\mathcal{B}$. Prove that

$$\mathbb{E}[F(X,Y)\,|\,\mathcal{B}] \overset{\text{a.s.}}{=} H(X), \tag{2.16}$$

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

Hint: Show that the set

$$\mathcal{F} := \{F\colon S_1 \times S_2 \to \mathbb{R} \mid F \text{ is bounded and } \mathcal{S}_1 \otimes \mathcal{S}_2\text{-measurable satisfying (2.16)}\,\}$$

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

Remark: In the case $\mathcal{B} = \sigma(X)$, this exercise can be used to illustrate the Doob–Dynkin lemma, because H is explicitly determined here.

Exercise 2.3 (Explicit construction of the general Bernoulli mixture model)**.** Consider a $[0,1]^m$-valued random vector $(P_1,\dots,P_m)$ and let $U_1,\dots,U_m$ be independent random variables, uniformly distributed on $[0,1]$, and independent of $(P_1,\dots,P_m)$. Define, for every obligor $i\in\{1,\dots,m\}$,

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

Use Exercise 2.2 to show that $N_1,\dots,N_m$ satisfy (2.11) and (2.12).

Hint: For (2.12) with $n \coloneqq (n_1,\dots,n_m)$ apply Exercise 2.2 with

$$H_n(p_1,\dots,p_m) = \mathbb{P}\bigl[\mathbb{1}_{[0,p_i]}(U_i) = n_i \text{ for } i\in\{1,\dots,m\}\bigr].$$

2.2.2 Number of Default Events, Expected Value and Variance

With the assumptions (2.11) and (2.12) 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 $i\in\{1,\dots,m\}$,

$$\mathbb{E}[N_i] = \mathbb{P}[N_i=1] = \mathbb{E}\bigl[\mathbb{P}[N_i=1\,|\,P_1,\dots,P_m]\bigr] \overset{(2.11)}{=} \mathbb{E}[P_i]\,, \tag{2.17}$$

where we also used a defining property of conditional expectation, or more directly by (2.15) with $I=\{i\}$ and $n_i=1$. Using (2.2), we obtain for the expected number of defaults, cf. (2.5),

$$\mathbb{E}[N] = \sum_{i=1}^m \mathbb{E}[N_i] \overset{(2.17)}{=} \sum_{i=1}^m \mathbb{E}[P_i]. \tag{2.18}$$

For the variance, first note that by the general formula for sums of square-integrable random variables (see Exercise 2.4 below),

$$\operatorname{Var}(N) = \sum_{i=1}^m \operatorname{Var}(N_i) + \sum_{\substack{i,j=1\\ i\ne j}}^m \operatorname{Cov}(N_i,N_j)\,. \tag{2.19}$$

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

$$\operatorname{Var}(N_i) = \mathbb{E}\bigl[N_i^2\bigr] - (\mathbb{E}[N_i])^2 = \mathbb{E}[N_i] - (\mathbb{E}[N_i])^2 \overset{(2.17)}{=} \mathbb{E}[P_i]\,(1-\mathbb{E}[P_i]) \tag{2.20}$$

for every $i\in\{1,\dots,m\}$. Next we compute the covariance. From (2.15) we get for $i\ne j$ in $\{1,\dots,m\}$

$$\mathbb{E}[N_iN_j] = \mathbb{P}[N_i=1,\,N_j=1] = \mathbb{E}[P_iP_j], \tag{2.21}$$

hence with (2.17)

$$\begin{aligned} \operatorname{Cov}(N_i,N_j) &= \mathbb{E}[N_iN_j] - \mathbb{E}[N_i]\,\mathbb{E}[N_j]\\ &= \mathbb{E}[P_iP_j] - \mathbb{E}[P_i]\,\mathbb{E}[P_j]\\ &= \operatorname{Cov}(P_i,P_j)\,. \end{aligned} \tag{2.22}$$

Equations (2.19), (2.20) and (2.22) together yield the variance

$$\operatorname{Var}(N) = \sum_{i=1}^{m} \mathbb{E}[P_i](1 - \mathbb{E}[P_i]) + \sum_{\substack{i,j=1 \\ i\neq j}}^{m} \operatorname{Cov}(P_i, P_j)\,. \tag{2.23}$$

Exercise 2.4. Prove (2.19) for real-valued square-integrable random variables $N_1, \dots, N_m$. You may use the first equality in (2.22) as definition.

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 $n_1, \dots, n_m \in \{0,1\}$ and $n \coloneqq n_1 + \cdots + n_m$ denoting the total number of defaults, it follows from (2.14) that

$$\mathbb{P}[N_1 = n_1, \dots, N_m = n_m] = \int_0^1 p^n (1-p)^{m-n}\, F(\mathrm{d}p). \tag{2.24}$$

Given that the identity of the defaulting obligors is unknown (as in the case above), the probability for $n \in \{0, \dots, m\}$ defaults is given by

$$\begin{aligned} \mathbb{P}[N = n] &= \mathbb{E}\big[\mathbb{P}[N = n \,|\, P]\big] \\ &= \mathbb{E}\Bigg[\underbrace{\binom{m}{n} P^n (1-P)^{m-n}}_{\text{binomial distribution}}\Bigg] \\ &= \binom{m}{n} \int_0^1 p^n (1-p)^{m-n}\, F(\mathrm{d}p), \end{aligned} \tag{2.25}$$

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

In the case of such a uniform portfolio, the expectation in (2.18) reduces to

$$\mathbb{E}[N] = m\, \mathbb{E}[P] \tag{2.26}$$

and the variance of the total number of defaults can be computed using (2.23). For $i \neq j$ in $\{1, \dots, m\}$ we have $\operatorname{Cov}(P_i, P_j) = \operatorname{Var}(P) \geq 0$ and therefore

$$\operatorname{Var}(N) = m\, \mathbb{E}[P](1 - \mathbb{E}[P]) + m(m-1)\operatorname{Var}(P). \tag{2.27}$$

Hence, the variance of N is composed of the binomial component $m\,\mathbb{E}[P](1-\mathbb{E}[P])$ with success probability $\mathbb{E}[P]$ and a non-negative additional variance term arising from the uncertainty of P. In essence, using the uniform Bernoulli mixture model can only increase the variance of the total number of defaults.

More generally, the factorial moments can be calculated using (2.10) from Exercise 2.1, because for every $l \in \mathbb{N}_0$ by conditioning

$$\mathbb{E}\biggl[\prod_{k=0}^{l-1}(N-k)\biggr] = \mathbb{E}\biggl[\mathbb{E}\biggl[\prod_{k=0}^{l-1}(N-k)\biggm| P\biggr]\biggr] \overset{(2.10)}{=} \mathbb{E}[P^l]\prod_{k=0}^{l-1}(m-k). \tag{2.28}$$

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 is defined[6] by

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}\,\mathrm{e}^{-x}\,\mathrm{d}x, \qquad \alpha > 0. \tag{2.29}$$

By partial integration,

$$\alpha\,\Gamma(\alpha) = \Gamma(\alpha+1), \qquad \alpha > 0, \tag{2.30}$$

which is the functional equation of the gamma function. Iterated application of (2.30) yields

$$\Gamma(\alpha+n) = \Gamma(\alpha)\prod_{i=0}^{n-1}(\alpha+i), \qquad \alpha > 0,\, n \in \mathbb{N}_0. \tag{2.31}$$

Since $\Gamma(1) = 1$ by (2.29), the case $\alpha = 1$ shows that $\Gamma(n+1) = n!$ for all $n \in \mathbb{N}_0$.

Exercise 2.5 (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

$$\Delta_{d-1} = \{(x_1, \ldots, x_{d-1}) \in (0,1)^{d-1} \mid x_1 + \cdots + x_{d-1} < 1\}. \tag{2.32}$$

Show by direct calculation for the multivariate beta function[7] that

$$\begin{aligned} B(\alpha_1, \ldots, \alpha_d) &:= \int_{\Delta_{d-1}} \biggl(\prod_{i=1}^{d-1} x_i^{\alpha_i - 1}\biggr)(1 - x_1 - \cdots - x_{d-1})^{\alpha_d - 1}\,\mathrm{d}(x_1, \ldots, x_{d-1}) \\ &= \frac{\prod_{i=1}^d \Gamma(\alpha_i)}{\Gamma(\alpha_1 + \cdots + \alpha_d)}, \qquad \alpha_1, \ldots, \alpha_d > 0, \end{aligned} \tag{2.33}$$

[6] The gamma function is actually a 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.

[7] The proof of Lemma 4.35 below contains a probabilistic argument for the case $d = 2$.

which in the case $d = 2$ simplifies to

$$B(\alpha,\beta) := \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,\mathrm{d}x = \frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}, \qquad \alpha,\beta > 0. \tag{2.34}$$

Using a particular choice of $\alpha_1,\dots,\alpha_d$, conclude that the $(d-1)$-dimensional volume of Δ_{d-1} is $1/(d-1)!$.

Hint: Write down $\prod_{i=1}^d \Gamma(\alpha_i)$ and use a d-dimensional integral substitution with $(x_1,\dots,x_{d-1},1-x_1-\cdots-x_{d-1})z$ where $(x_1,\dots,x_{d-1}) \in \Delta_{d-1}$ and $z \in (0,\infty)$.

Definition 2.6 (Beta distribution[8])**.** A density of the beta distribution with real shape parameters $\alpha,\beta > 0$ is given by

$$f_{\alpha,\beta}(p) = \begin{cases} \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)} & \text{for } p \in (0,1), \\ 0 & \text{for } p \in \mathbb{R}\setminus(0,1), \end{cases} \tag{2.35}$$

where B denotes the beta function, see (2.34). For a random variable P with a beta distribution, we use the notation $P \sim \mathrm{Beta}(\alpha,\beta)$.

When the mixing random variable P in the uniform Bernoulli mixture model (as presented in Section 2.3) follows a beta distribution, we can derive a more explicit distribution for the number of defaults. From (2.25) we get that

$$\begin{aligned} \mathbb{P}[N=n] &= \binom{m}{n}\int_0^1 p^n(1-p)^{m-n}\frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}\,\mathrm{d}p \\ &= \binom{m}{n}\frac{1}{B(\alpha,\beta)}\underbrace{\int_0^1 p^{\alpha+n-1}(1-p)^{\beta+m-n-1}\,\mathrm{d}p}_{=B(\alpha+n,\beta+m-n)\text{ by (2.34)}} \\ &= \binom{m}{n}\frac{B(\alpha+n,\beta+m-n)}{B(\alpha,\beta)}, \qquad n \in \{0,1,\dots,m\}, \end{aligned} \tag{2.36}$$

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 $\mathrm{BetaBin}(\alpha,\beta,m)$.

Exercise 2.7 (Moments of the beta distribution)**.** Let $P \sim \mathrm{Beta}(\alpha,\beta)$ with $\alpha,\beta > 0$. Show that

$$\mathbb{E}\big[P^\gamma(1-P)^\delta\big] = \frac{B(\alpha+\gamma,\beta+\delta)}{B(\alpha,\beta)}, \qquad \gamma > -\alpha,\ \delta > -\beta, \tag{2.37}$$

and, using the relation (2.34) for the beta function and the functional equation (2.31) of the gamma function, conclude that

$$\mathbb{E}[P^l] = \prod_{k=0}^{l-1}\frac{\alpha+k}{\alpha+\beta+k}, \qquad l \in \mathbb{N}_0, \tag{2.38}$$

in particular

$$\mathbb{E}[P] = \frac{\alpha}{\alpha+\beta} \qquad\text{and}\qquad \mathrm{Var}(P) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}. \tag{2.39}$$

[8] For the multivariate generalization, see Definition 4.26 below.

Exercise 2.8 (Computation of the beta-binomial distribution)**.** Using the relation (2.34) for the beta function and the functional equation (2.30) of the gamma function, show that the beta-binomial distribution (2.36) can be computed in an elementary way by

$$\mathbb{P}[N=n]=\biggl(\prod_{i=0}^{n-1}\frac{\alpha+i}{i+1}\biggr)\biggl(\prod_{i=0}^{m-n-1}\frac{\beta+i}{i+1}\biggr)\prod_{i=0}^{m-1}\frac{i+1}{\alpha+\beta+i}$$

for every $n\in\{0,\dots,m\}$, and conclude that it can also be calculated recursively from the initial value

$$\mathbb{P}[N=0]=\prod_{i=0}^{m-1}\frac{\beta+i}{\alpha+\beta+i}$$

and the recursion formula

$$\mathbb{P}[N=n]=\frac{(\alpha+n-1)(m-n+1)}{n(\beta+m-n)}\,\mathbb{P}[N=n-1]\,,\qquad n\in\{1,\dots,m\},$$

in a numerically stable way, because only differences of integers are calculated.

Exercise 2.9 (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 $l\in\mathbb{N}_0$, the l-th factorial moment is given by

$$\mathbb{E}\biggl[\prod_{k=0}^{l-1}(N-k)\biggr]=\prod_{k=0}^{l-1}\frac{(\alpha+k)(m-k)}{\alpha+\beta+k},\tag{2.40}$$

and conclude from (2.40) using $N^2=N+N(N-1)$ that

$$\mathbb{E}[N]=\frac{\alpha m}{\alpha+\beta}\qquad\text{and}\qquad\operatorname{Var}(N)=\frac{\alpha\beta m(\alpha+\beta+m)}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Hint: Combine (2.28) and (2.38).

Exercise 2.10 (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),

$$x^n=\sum_{l=0}^{n}\left\{{n\atop l}\right\}\prod_{k=0}^{l-1}(x-k),\qquad n\in\mathbb{N}_0,\tag{2.41}$$

where $\left\{{n\atop l}\right\}$ denotes the Stirling number of the second kind,[9] defined recursively by

$$\left\{{n+1\atop l}\right\}=\left\{{n\atop l-1}\right\}+l\left\{{n\atop l}\right\},\qquad l\in\mathbb{N}\text{ and }n\in\mathbb{N}_0,\tag{2.42}$$

[9] The Stirling number of the second kind $\left\{{n\atop l}\right\}$ gives the number of ways to partition a set of $n\in\mathbb{N}$ elements into $l\in\{1,\dots,n\}$ non-empty subsets: Obviously $\left\{{1\atop 1}\right\}=1$. To explain the recursion formula (2.42) by induction, you can add $\{n+1\}$ as a new subset to the partition of $\{1,\dots,n\}$ into $l-1$ subsets, or you can put $n+1$ into one of the l existing sets of the partition.

with initial conditions $\left\{ {0 \atop 0} \right\} := 1$, $\left\{ {n \atop 0} \right\} := 0$ and $\left\{ {0 \atop l} \right\} := 0$ 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

$$\mathbb{E}[N^n] = \sum_{l=0}^{n} \left\{ {n \atop l} \right\} \mathbb{E}\biggl[\prod_{k=0}^{l-1}(N-k)\biggr], \qquad n \in \mathbb{N}_0. \tag{2.43}$$

Show that (2.43) is also true for $\mathbb{C}$-valued random variables, provided the absolute factorial moments for the right-hand side of (2.43) are finite or the absolute nth moment for the left-hand side is finite. Explain how (2.43) can be applied to random $\mathbb{C}^{d\times d}$-matrices and see Exercise 4.14 for the multivariate extension.

Hint: Show for all $l, n \in \mathbb{N}$ that $\left\{ {n \atop l} \right\} = 0$ if $l > n$ and $\left\{ {n \atop l} \right\} = 1$ if $l = n$. Use $x = (x - l) + l$ to prove (2.41).

2.3.2 Biased Measure and the Beta Distribution

Definition 2.11 (Biased probability measure)**.** Let Λ be a $[0,\infty)$-valued random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $0 < \mathbb{E}[\Lambda] < \infty$. Then the Λ-biased probability measure $\mathbb{P}_\Lambda$ on $(\Omega, \mathcal{F})$ is defined by

$$\mathbb{P}_\Lambda[A] = \frac{\mathbb{E}[\Lambda \mathbb{1}_A]}{\mathbb{E}[\Lambda]}, \qquad A \in \mathcal{F}. \tag{2.44}$$

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

Lemma 2.12 (Biased beta distribution[10])**.** *Assume that $P \sim \mathrm{Beta}(\alpha, \beta)$ with parameters $\alpha, \beta > 0$ and that $\gamma \in (-\alpha, \infty)$ and $\delta \in (-\beta, \infty)$. Then $\mathbb{P}_{P^\gamma(1-P)^\delta} P^{-1} = \mathrm{Beta}(\alpha+\gamma, \beta+\delta)$, that means the distribution of P under the $P^\gamma(1-P)^\delta$-biased probability measure $\mathbb{P}_{P^\gamma(1-P)^\delta}$ given by Definition 2.11 is the $\mathrm{Beta}(\alpha+\gamma, \beta+\delta)$ distribution.*

Proof. By (2.37) and (2.44), a density of the $P^\gamma(1-P)^\delta$-biased probability measure $\mathbb{P}_{P^\gamma(1-P)^\delta}$ w.r.t. $\mathbb{P}$ is given by

$$\frac{\mathrm{d}\mathbb{P}_{P^\gamma(1-P)^\delta}}{\mathrm{d}\mathbb{P}} = \frac{B(\alpha,\beta)}{B(\alpha+\gamma,\beta+\delta)} P^\gamma (1-P)^\delta.$$

Let μ denote the Lebesgue–Borel measure on $\mathbb{R}$. Using the density $f_{\alpha,\beta}$ from (2.35) shows that, for μ-almost all $p \in (0,1)$,

$$\begin{aligned}
\frac{\mathrm{d}(\mathbb{P}_{P^\gamma(1-P)^\delta} P^{-1})}{\mathrm{d}\mu}(p) &= \frac{\mathrm{d}(\mathbb{P}_{P^\gamma(1-P)^\delta} P^{-1})}{\mathrm{d}(\mathbb{P}P^{-1})}(p) \cdot \frac{\mathrm{d}(\mathbb{P}P^{-1})}{\mathrm{d}\mu}(p) \\
&= \frac{B(\alpha,\beta)}{B(\alpha+\gamma,\beta+\delta)} p^\gamma (1-p)^\delta \cdot f_{\alpha,\beta}(p) \\
&= \frac{p^{\alpha+\gamma-1}(1-p)^{\beta+\delta-1}}{B(\alpha+\gamma,\beta+\delta)},
\end{aligned}$$

which by (2.35) gives a density of the $\mathrm{Beta}(\alpha+\gamma, \beta+\delta)$ distribution. □

[10] See Lemma 4.29 below for the generalization to the Dirichlet 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 $p_1, \ldots, p_m$ w.r.t. the macroeconomic random variable, which don't need to be equal.

Definition 2.13 (One-factor Bernoulli mixture model)**.** Consider Bernoulli random variables $N_1, \ldots, N_m$. Let Λ be a $[0, \infty)$-valued random variable such that $0 < \mathbb{E}[\Lambda] < \infty$. If there exist $p_1, \ldots, p_m \in [0, \infty)$ such that

$$\mathbb{P}[N_i = 1 \,|\, \Lambda] \overset{\text{a.s.}}{=} p_i \Lambda, \qquad i \in \{1, \ldots, m\}, \tag{2.45}$$

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

$$\mathbb{P}[N_1 = n_1, \ldots, N_m = n_m \,|\, \Lambda] \overset{\text{a.s.}}{=} \prod_{i=1}^{m} \mathbb{P}[N_i = n_i \,|\, \Lambda] \tag{2.46}$$

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 susceptibilities $p_1, \ldots, p_m$. If $p_1 = \cdots = p_m$, then we call the model homogeneous.

Condition (2.45) implies that $\max\{p_1, \ldots, p_m\}\Lambda \le 1$ $\mathbb{P}$-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 susceptibilities $p_1, \ldots, p_m$ are the individual default probabilities within the next period as introduced in (2.1).

Remark 2.14 (Discussion of expectation and variance)**.** Let $(N_1, \ldots, N_m, \Lambda)$ be a one-factor Bernoulli mixture model with susceptibilities $p_1, \ldots, p_m$, let $N = N_1 + \cdots + N_m$ denote the number of defaults, and define $\lambda = p_1 + \cdots + p_m$. Then (2.45) implies that

$$\mathbb{E}[N \,|\, \Lambda] \overset{\text{a.s.}}{=} (p_1 + \cdots + p_m)\Lambda = \lambda\Lambda,$$

hence $\mathbb{E}[N] = \lambda \, \mathbb{E}[\Lambda]$. For the variance we see from (2.23) that

$$\operatorname{Var}(N) = \sum_{i=1}^{m} p_i \, \mathbb{E}[\Lambda] \, (1 - p_i \, \mathbb{E}[\Lambda]) + \sum_{\substack{i,j=1 \\ i \ne j}}^{m} \underbrace{\operatorname{Cov}(p_i\Lambda, p_j\Lambda)}_{= p_i p_j \operatorname{Var}(\Lambda)}. \tag{2.47}$$

Using the abbreviation $\lambda_2 := p_1^2 + \cdots + p_m^2$ and noting that the double sum over $p_i p_j$ in (2.47) has all terms of λ^2 except $p_1^2, \ldots, p_m^2$, it follows that

$$\operatorname{Var}(N) = \lambda \, \mathbb{E}[\Lambda] - \lambda_2 (\mathbb{E}[\Lambda])^2 + (\lambda^2 - \lambda_2) \operatorname{Var}(\Lambda)\,, \tag{2.48}$$

which can be smaller or larger than $\mathbb{E}[N] = \lambda \, \mathbb{E}[\Lambda]$ depending on $(\lambda^2 - \lambda_2) \operatorname{Var}(\Lambda)$. 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 $p_\mathrm{b} \coloneqq (\lambda^2 - \lambda_2)/\lambda^2$ defines a strictly positive probability. If, for a given mean $\mu > 0$, the susceptibilities satisfy $p_i \le p_\mathrm{b}/\mu$ for every $i \in \{1, \dots, m\}$, then there exists a random variable Λ with $\mathbb{E}[\Lambda] = \mu$ and $p_i\Lambda \le 1$ for all $i \in \{1, \dots, m\}$ satisfying

$$\operatorname{Var}(\Lambda) = \frac{\lambda_2}{\lambda^2 - \lambda_2}(\mathbb{E}[\Lambda])^2; \tag{2.49}$$

a simple (but extreme) example is a random variable Λ with $\mathbb{P}[\Lambda = 0] = 1 - p_\mathrm{b}$ and $\mathbb{P}[\Lambda = \mu/p_\mathrm{b}] = p_\mathrm{b}$, because $\mathbb{E}[\Lambda] = \mu$ and $\mathbb{E}[\Lambda^2] = \mu^2/p_\mathrm{b}$, hence

$$\operatorname{Var}(\Lambda) = \mathbb{E}[\Lambda^2] - (\mathbb{E}[\Lambda])^2 = \Big(\frac{1}{p_\mathrm{b}} - 1\Big)\mu^2 = \frac{\lambda_2}{\lambda^2 - \lambda_2}\mu^2.$$

In the case (2.49), the expectation and the variance of N agree,[11] see (2.48).

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[12] with parameter $\lambda \ge 0$ if

$$\mathbb{P}[N = n] = \frac{\lambda^n}{n!}\,\mathrm{e}^{-\lambda}, \qquad n \in \mathbb{N}_0, \tag{3.1}$$

where we use the convention $0^0 \coloneqq 1$. We will use the notation $N \sim \operatorname{Poisson}(\lambda)$.

In a credit risk context, if N describes the number of defaults of an obligor within one period, then mainly the events $N = 0$ and $N = 1$ 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 within one period.

First we consider moments. Suppose $N \sim \operatorname{Poisson}(\lambda)$ 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

$$\mathbb{E}\Bigg[\prod_{k=0}^{l-1}(N-k)\Bigg] \overset{(3.1)}{=} \sum_{n=l}^{\infty} \Bigg(\underbrace{\prod_{k=0}^{l-1}(n-k)}_{=0 \text{ for } n \in \{0,\dots,l-1\}}\Bigg)\frac{\lambda^n}{n!}\,\mathrm{e}^{-\lambda} = \lambda^l\,\mathrm{e}^{-\lambda}\underbrace{\sum_{n=l}^{\infty}\frac{\lambda^{n-l}}{(n-l)!}}_{=\mathrm{e}^{\lambda}} = \lambda^l. \tag{3.2}$$

[11] The property $\mathbb{E}[N] = \operatorname{Var}(N)$ is shared with the Poisson distribution, see Definition 3.1 as well as (3.3) and (3.4) below.

[12] Named after the French mathematician Siméon Denis Poisson (1781–1840).

For $l = 1$ this gives the expected value

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

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

$$\begin{aligned}\operatorname{Var}(N) &= \mathbb{E}[N^2] - (\mathbb{E}[N])^2 \\ &= \mathbb{E}[N] + \mathbb{E}[N(N-1)] - (\mathbb{E}[N])^2 = \lambda + \lambda^2 - \lambda^2 = \lambda.\end{aligned} \tag{3.4}$$

To calculate higher moments of N, use (2.43) from Exercise 2.10.

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 $N_1, \dots, N_k$ are independent with $N_i \sim \text{Poisson}(\lambda_i)$ for all $i \in \{1, \dots, k\}$, then*

$$N := \sum_{i=1}^{k} N_i \sim \text{Poisson}(\lambda_1 + \dots + \lambda_k)\,. \tag{3.5}$$

We give a direct proof below; for a short one using probability-generating functions, see (4.32). For the multivariate generalization, see Lemma 3.43.

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,

$$\begin{aligned}\mathbb{P}[X+Y=n] &= \sum_{l=0}^{n} \underbrace{\mathbb{P}[X=n-l, Y=l]}_{=\,\mathbb{P}[X=n-l]\,\mathbb{P}[Y=l] \text{ by independence}} \\ &= \sum_{l=0}^{n} \mathrm{e}^{-\lambda}\,\frac{\lambda^{n-l}}{(n-l)!}\,\mathrm{e}^{-\mu}\,\frac{\mu^l}{l!} = \mathrm{e}^{-(\lambda+\mu)}\,\frac{1}{n!}\underbrace{\sum_{l=0}^{n}\binom{n}{l}\lambda^{n-l}\mu^l}_{=\,(\lambda+\mu)^n},\end{aligned} \tag{3.6}$$

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 $\lambda \ge 0$, the Poisson distribution Poisson(λ) is infinitely divisible, because for every $k \in \mathbb{N}$ the distribution of $N_1 + \dots + N_k$ is Poisson(λ), when $N_1, \dots, N_k$ are independent with $N_i \sim \text{Poisson}(\lambda/k)$ for every $i \in \{1, \dots, k\}$.

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 five calibration options available. The Poisson parameter $\lambda \geq 0$ of a random variable $N \sim \text{Poisson}(\lambda)$ can be determined as follows when a Bernoulli distribution with success probability $p \in [0, 1]$ is provided:

(a) Given $p \in [0, 1)$, choose $\lambda \in [0, \infty)$ so that the probability of no default coincides with the one in the Bernoulli model, i.e.

$$\mathrm{e}^{-\lambda} = \mathbb{P}[N = 0] = 1 - p, \tag{3.7}$$

or equivalently, using the Taylor expansion,[13]

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

However, when the probability p of one default is sufficiently close to 1, it might be better to approximate p by $\mathbb{P}[N = 1] = \lambda\,\mathrm{e}^{-\lambda}$ as best as possible. By considering the derivative of $[0, \infty) \ni \lambda \mapsto \lambda\,\mathrm{e}^{-\lambda}$, which is $\lambda \mapsto (1 - \lambda)\,\mathrm{e}^{-\lambda}$, it follows that $\mathbb{P}[N = 1]$ attains its maximum $1/\mathrm{e}$ for $\lambda = 1$. Hence at least for $p \geq 1 - 1/\mathrm{e}$, the calibration $\lambda = 1$ instead of the larger value arising from (3.8) should be chosen, see also (d) and (e) below.

(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.

$$\lambda = \mathbb{E}[N] = p, \tag{3.9}$$

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.

$$\lambda = \operatorname{Var}(N) = p(1 - p), \tag{3.10}$$

where (2.4) for the variance of a Bernoulli random variable and (3.4) for the variance of N are used.

[13] Note that the terms on the right-hand side of (3.8) are (up to normalization) those of the logarithmic distribution $\text{Log}(p)$, see Definition 4.4 below.

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.

There are two additional calibration methods discussed in Subsection 3.4.1 below and mentioned here for completeness, both are variants of (3.8):

(d) For $p \in [0, 1]$ take $\lambda = -\log(1 - \min\{p, 1/2\}) \in [0, \log 2]$, which minimizes the Wasserstein distance between Poisson(λ) and the Bernoulli distribution Bin$(1, p)$, see Definition 3.14 and Remark 3.28 below.

(e) For $p \in [0, 1]$ take $\lambda = -\log(1 - \min\{p, 1 - 1/\mathrm{e}\}) \in [0, 1]$, which minimizes the total variation distance between Poisson(λ) and Bin$(1, p)$, see Definition 3.7 and Exercise 3.33 below.

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[14], $\mathcal{M}_1(S, \mathcal{S})$ 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 σ-algebra $\mathcal{S}$ in the notation. Define the set

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

of all probability measures μ such that $\mathcal{F} \subseteq L^1(\mu)$. Then

$$d_{\mathcal{F}}(\mu, \nu) = \sup_{f \in \mathcal{F}} \Big| \int_S f \,\mathrm{d}\mu - \int_S f \,\mathrm{d}\nu \Big|, \qquad \mu, \nu \in \mathcal{M}_1^{\mathcal{F}}(S), \tag{3.12}$$

defines an $\overline{\mathbb{R}}_+$-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 $\int_S f \,\mathrm{d}\mu \neq \int_S f \,\mathrm{d}\nu$.

Remark 3.5. Note that the supremum in (3.12) can result in $d_{\mathcal{F}}(\mu, \nu) = \infty$, which is normally not an allowed value for a metric or a pseudometric. This

[14] 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 σ-algebra $\mathcal{B}_{\mathbb{R}}$ on $\mathbb{R}$, respectively.

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 $d'_{\mathcal{F}}(\mu, \nu) \coloneqq \min\{r, d_{\mathcal{F}}(\mu, \nu)\}$. 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.

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

$$\int_S (c_f - f)\,\mathrm{d}\mu - \int_S (c_f - f)\,\mathrm{d}\nu = \int_S f\,\mathrm{d}\nu - \int_S f\,\mathrm{d}\mu, \qquad f \in \mathcal{F},$$

because μ and ν are probability measures, hence we can omit the absolute value in the definition (3.12) of $d_{\mathcal{F}}$.

We will consider three different choices for $\mathcal{F}$, giving rise to three different metrics.[15] The first one arises from the set $\mathcal{F}_{\mathrm{TV}} \coloneqq \{\, \mathbb{1}_A \mid A \in \mathcal{S}\,\}$ 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_{TV} on the set $\mathcal{M}_1(S)$ of all probability measures on the measurable space $(S, \mathcal{S})$ is defined by

$$d_{\mathrm{TV}}(\mu, \nu) = \sup_{A \in \mathcal{S}} (\mu(A) - \nu(A)), \qquad \mu, \nu \in \mathcal{M}_1(S).$$

Remark 3.8. Note that $d_{\mathrm{TV}}(\mu, \nu) \le 1$ for all $\mu, \nu \in \mathcal{M}_1(S)$. If μ and ν are mutually singular, then $d_{\mathrm{TV}}(\mu, \nu) = 1$. The reverse direction is also true and follows from Exercise 3.19(c) 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 σ-algebra $\mathcal{B}_{\mathbb{R}^d}$, we consider the collection

$$\mathcal{F}_{\mathrm{KS}} \coloneqq \{\, \mathbb{1}_{(-\infty, a_1] \times \cdots \times (-\infty, a_d]} \mid (a_1, \ldots, a_d) \in \mathbb{R}^d \,\}.$$

Since the distribution function F_μ of a probability measure μ on $\mathbb{R}^d$, defined by $F_\mu(a_1, \ldots, a_d) = \mu\big((-\infty, a_1] \times \cdots \times (-\infty, a_d]\big)$ for all $(a_1, \ldots, a_d) \in \mathbb{R}^d$, uniquely determines[16] μ, the collection $\mathcal{F}_{\mathrm{KS}}$ separates the probability measures on $\mathbb{R}^d$.

[15] There are other notions of "distances" for probability measures like the Hellinger metric, the pth Wasserstein metric for $p > 1$, the Levý–Prokhorov metric metricizing the so-called weak topology, the Kullback–Leibler divergence (which is not a metric), and so on, cf. [24]. For connections to optimal transport, see the textbooks by C. Villani [55, 56].

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

Definition 3.9 (Kolmogorov–Smirnov metric)**.** The Kolmogorov–Smirnov metric[17] d_{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

$$d_{\mathrm{KS}}(\mu,\nu) = \sup_{a\in\mathbb{R}^d} |F_\mu(a) - F_\nu(a)| = \|F_\mu - F_\nu\|_\infty, \qquad \mu,\nu \in \mathcal{M}_1(\mathbb{R}^d), \tag{3.13}$$

where F_μ and F_ν denote the distribution functions of μ and ν, respectively.

Remark 3.10. For probability measures μ and ν on $\mathbb{R}^d$, it follows from $\mathcal{F}_{\mathrm{KS}} \subseteq \mathcal{F}_{\mathrm{TV}}$ that

$$d_{\mathrm{KS}}(\mu,\nu) \le d_{\mathrm{TV}}(\mu,\nu). \tag{3.14}$$

The Kolmogorov–Smirnov metric is useful to obtain estimates for quantiles and value-at-risk, see Lemma 8.7 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 μ denote the uniform distribution on $[0,1]$ and define $\mu_n = (1/n)\sum_{i=1}^n \delta_{i/n}$. Then $\mu(\{1/n,\dots,n/n\}) = 0$ and $\mu_n(\{1/n,\dots,n/n\}) = 1$, hence $d_{\mathrm{TV}}(\mu,\mu_n) = 1$ by Remark 3.8, while $d_{\mathrm{KS}}(\mu,\mu_n) = 1/n$ 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 $d_{\mathrm{KS}}(\mu,\mu_n) = 1$ for every $n\in\mathbb{N}$. On the other hand, $\int_{\mathbb{R}} f\,\mathrm{d}\mu_n = f(1/n) \to f(0) = \int_S f\,\mathrm{d}\mu$ as $n\to\infty$ for every bounded and continuous function $f\colon \mathbb{R}\to\mathbb{R}$, which means weak convergence of $(\mu_n)_{n\in\mathbb{N}}$ to μ.

For the last one of the three metrics, consider a metric space (S,d) with Borel σ-algebra $\mathcal{S}$ and let $\mathcal{F}_{\mathrm{W}}$ denote the set of all functions $f\colon S\to\mathbb{R}$, which are Lipschitz continuous with constant at most 1, i.e.,

$$|f(x) - f(y)| \le d(x,y), \qquad x,y\in S.$$

Note that $\mathcal{F}_{\mathrm{W}}$ has the property discussed in Remark 3.6 with $c_f = 0$. Define $\mathcal{M}_1^{\mathcal{F}_{\mathrm{W}}}(S)$ according to (3.11).

Exercise 3.13 (Separating functions for $\mathcal{M}_1(S)$)**.** Let (S,d) be a metric space. Show that already the bounded functions in $\mathcal{F}_{\mathrm{W}}$ separate the probability measures in $\mathcal{M}_1(S)$.

Hint: Consider $f_{A,n}(x) = (1 - n\,\mathrm{dist}(A,x))^+$ for closed $A\subseteq S$ and $n\in\mathbb{N}$. Use a corollary of Dynkin's lemma, see e.g. [49, Corollary 15.69].

[17] Named after Andrey Kolmogorov (1903–1987) and Nikolai Smirnov (1900–1966), because the metric appears in the test statistic in their Kolmogorov–Smirnov test.

Definition 3.14 (Wasserstein metric)**.** Let (S,d) be metric space with Borel σ-algebra $\mathcal{S}$. The Wasserstein metric[18] d_{W} induced by d is defined by

$$d_{\mathrm{W}}(\mu,\nu) = \sup_{f\in\mathcal{F}_{\mathrm{W}}} \biggl(\int_S f\,\mathrm{d}\mu - \int_S f\,\mathrm{d}\nu\biggr), \qquad \mu,\nu\in\mathcal{M}_1^{\mathcal{F}_{\mathrm{W}}}(S). \tag{3.15}$$

Remark 3.15 (The Wasserstein metric is well defined on $\mathcal{M}_1^{\mathcal{F}_{\mathrm{W}}}(S)$)**.** Consider a point $x_0\in S$ and two probability measures $\mu,\nu\in\mathcal{M}_1^{\mathcal{F}_{\mathrm{W}}}(S)$. Then, for every function $f\colon S\to\mathbb{R}$ having Lipschitz constant

$$\operatorname{Lip}(f) := \sup_{\substack{x,y\in S\\ x\neq y}} \frac{|f(x)-f(y)|}{d(x,y)} < \infty, \tag{3.16}$$

the expectations $\int_S f\,\mathrm{d}\mu$ and $\int_S f\,\mathrm{d}\nu$ are well defined, because $|f(x)|\le|f(x_0)|+\operatorname{Lip}(f)\,d(x,x_0)$ for all $x\in S$, and the function $S\ni x\mapsto d(x,x_0)\in\mathbb{R}$ is in $\mathcal{F}_{\mathrm{W}}$ by the reverse triangle inequality: $|d(x,x_0)-d(y,x_0)|\le d(x,y)$ for all $x,x_0,y\in S$. Furthermore,

$$\begin{aligned}\biggl|\int_S f\,\mathrm{d}\mu - \int_S f\,\mathrm{d}\nu\biggr| &= \biggl|\int_S (f(x)-f(x_0))\,\mu(\mathrm{d}x) - \int_S (f(x)-f(x_0))\,\nu(\mathrm{d}x)\biggr| \\ &\le \operatorname{Lip}(f)\biggl(\int_S d(x,x_0)\,\mu(\mathrm{d}x) + \int_S d(x,x_0)\,\nu(\mathrm{d}x)\biggr),\end{aligned}$$

which in particular implies that $d_{\mathrm{W}}(\mu,\nu)$ in (3.15) is finite, cf. Remark 3.5.

Remark 3.16 (Bounds for the Wasserstein metric)**.** Consider two probability measures $\mu,\nu\in\mathcal{M}_1^{\mathcal{F}_{\mathrm{W}}}(S)$. 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 $f\colon S\to\mathbb{R}$ 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

$$\bigl|\mathbb{E}[f(X)]-\mathbb{E}[f(Y)]\bigr| \le \operatorname{Lip}(f)\,d_{\mathrm{W}}(\mu,\nu), \tag{3.17}$$

which will be used in Lemma 8.25 below to estimate differences of expected shortfalls. If the metric $d\colon S\times S\to[0,\infty)$ is $\mathcal{S}\otimes\mathcal{S}$-measurable (which is certainly the case when the metric space (S,d) is separable and equipped with the Borel σ-algebra $\mathcal{S}$ as before, see [49, Corollary 15.55]), then $d(X,Y)$ is a random variable. Then by (3.16), for every function $f\colon S\to\mathbb{R}$ with Lipschitz constant $\operatorname{Lip}(f)<\infty$,

$$\bigl|\mathbb{E}[f(X)]-\mathbb{E}[f(Y)]\bigr| \le \mathbb{E}[|f(X)-f(Y)|] \le \operatorname{Lip}(f)\,\mathbb{E}[d(X,Y)]\,, \tag{3.18}$$

and taking the supremum in (3.18) over all functions f with $\operatorname{Lip}(f)\le 1$,

$$d_{\mathrm{W}}(\mu,\nu) \overset{(3.15)}{=} \sup_{f\in\mathcal{F}_{\mathrm{W}}} \bigl(\mathbb{E}[f(X)]-\mathbb{E}[f(Y)]\bigr) \le \mathbb{E}[d(X,Y)]\,. \tag{3.19}$$

To obtain a good upper bound, we can optimize the right-hand side of (3.19) with respect to the dependence of X and Y.

[18] Named after the Russian-American mathematician Leonid Nisonovich Vaserstein, 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 next example shows that weak convergence in general does not imply convergence in the Wasserstein metric, because there are unbounded functions in $\mathcal{F}_{\mathrm{W}}$. 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 $d_{\mathrm{W}}(\mu, \mu_n) \geq 1$ for all $n \in \mathbb{N}$. On the other hand, $\bigl|\int_S f\,\mathrm{d}\mu - \int_S f\,\mathrm{d}\mu_n\bigr| = |f(0) - f(n)|/n \leq 2\|f\|_\infty/n \to 0$ as $n \to \infty$, for every bounded and continuous function $f\colon \mathbb{R} \to \mathbb{R}$, which verifies weak convergence.

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

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

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

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

$$\sum_{n\in S} |n|\mu(\{n\}) < \infty \qquad \text{and} \qquad \sum_{n\in S} |n|\nu(\{n\}) < \infty, \tag{3.20}$$

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

For $S \subseteq \mathbb{Z}$ the Wasserstein distance $d_{\mathrm{W}}(\mu, \nu)$ between the probability measures μ and ν takes into account not only the amounts by which their individual probabilities differ, as in the total variation distance $d_{\mathrm{TV}}(\mu, \nu)$, 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 $A \subseteq S$,

$$\frac{1}{2}\sum_{n\in S} |e_n| \geq \frac{1}{2}\sum_{n\in A} e_n - \frac{1}{2}\sum_{n\in S\setminus A} e_n = \sum_{n\in A} e_n - \frac{1}{2}\underbrace{\sum_{n\in S} e_n}_{=0} = \mu(A) - \nu(A),$$

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

Exercise 3.19 (Representation of the total variation metric with densities)**.** Let $(S, \mathcal{S})$ be a measurable space and consider $\mu, \nu \in \mathcal{M}_1(S, \mathcal{S})$. Let λ be a non-negative σ-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 $f = \mathrm{d}\mu/\mathrm{d}\lambda$ and $g = \mathrm{d}\nu/\mathrm{d}\lambda$.

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

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

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

Hint: (c) Verify and use that $\frac{1}{2}|a - b| = \frac{1}{2}(a + b) - \min\{a, b\}$ for all $a, b \in \mathbb{R}$.

Exercise 3.20 (Total variation norm for signed and $\mathbb{C}$-valued measures)**.** Let $(S, \mathcal{S})$ be a measurable space and consider the set $\mathcal{M}(S, \mathcal{S})$ 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 $\mu \in \mathcal{M}(S, \mathcal{S})$ is the zero measure, i.e. $\mu(A) = 0$ for all $A \in \mathcal{S}$. Prove:

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

Hint: Measures on σ-algebras with values in $\mathbb{R}$ or $\mathbb{C}$ are always bounded, see e.g. [44, Theorem 6.4].

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

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

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

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

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

Hint: For $n \in \mathbb{N}$ consider the discrete uniform probability distribution μ_n on $\{1, \ldots, n\}$.

(e) Explain where the proof of σ-additivity for a limiting candidate μ in item (b) goes wrong when the sequence $(A_k)_{k \in \mathbb{N}}$ with $A_k = \{k\}$ in the setting of (d) is considered.[19]

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 $c \in \mathbb{K} \setminus \{0\}$,

$$d_{\mathrm{W}}(\mathcal{L}(cX), \mathcal{L}(cY)) = |c|\, d_{\mathrm{W}}(\mathcal{L}(X), \mathcal{L}(Y)).$$

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

[19] To 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 σ-additive, see [35, Chapter V, Section 10, Problems 34–41].

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 (i) implies (ii):

   (i) $\int_S \|x\|\,\mu(\mathrm{d}x) < \infty$ and $d_{\mathrm{W}}(\mathcal{L}(X_n), \mu) \to 0$ as $n \to \infty$.

   (ii) The set $\{X_n\}_{n\in\mathbb{N}}$ is uniformly integrable, i.e.

$$\lim_{c\to\infty} \sup_{n\in\mathbb{N}} \mathbb{E}\big[\|X_n\| \mathbb{1}_{\{\|X_n\|>c\}}\big] = 0,$$

   and converges weakly to μ.

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

(b) For $C \subseteq S$ with $C \neq \varnothing$ let $f\colon C \to \mathbb{R}$ denote a function with $\mathrm{Lip}(f) \le 1$. Show that $g(x) := \inf_{z\in C}(f(z) + \|z-x\|)$ for all $x \in S$ is in $\mathcal{F}_{\mathrm{W}}$ and extends f.

Always assume (ii) for the following steps.

(c) Prove that $\int_S \|x\|\,\mu(\mathrm{d}x) < \infty$.

(d) Prove for each $f \in \mathcal{F}_{\mathrm{W}}$ that $\lim_{n\to\infty} \mathbb{E}[f(X_n)] = \int_S f\,\mathrm{d}\mu$.

It remains to show that the convergence in (d) is uniform in $f \in \mathcal{F}_{\mathrm{W}}$.

(e) For each $b > 0$ define $\mathcal{F}_{\mathrm{W},b} := \{ f \in \mathcal{F}_{\mathrm{W}} \mid \|f\|_\infty \le b\}$ and assume that

$$d_{\mathrm{W},b}(\mathcal{L}(X_n), \mu) := \sup_{f\in\mathcal{F}_{\mathrm{W},b}} \Big(\mathbb{E}[f(X_n)] - \int_S f\,\mathrm{d}\mu\Big) \to 0 \quad \text{as } n \to \infty.$$

   Prove that $d_{\mathrm{W}}(\mathcal{L}(X_n), \mu) \to 0$ as $n \to \infty$.

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

Hints: (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 $c > 0$ the Lipschitz continuous function h_c, defined by $h_c(x) = \max\{0, \|x\| - \max\{0, c(c - \|x\|)\}\}$ satisfies $\|x\|\mathbb{1}_{\{\|x\|>c\}} \le h_c(x) \le \|x\|$ for all $x \in S$ and $h_c \searrow 0$ as $c \to \infty$. (c) Uniform integrability implies boundedness in $L^1(\mathbb{P})$, 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. (d) Restrict to $f \in \mathcal{F}^0_{\mathrm{W}} := \{f \in \mathcal{F}_{\mathrm{W}} \mid f(0) = 0\}$, then use $|f(x)| \le \|x\|$ for $x \in S$, part (c) and uniform integrability of $\{X_n\}_{n\in\mathbb{N}}$. (e) Similar to (d). (f) For a proof by contradiction, assume that there are $b, \varepsilon > 0$ such that, after passing to a subsequence if necessary, $d_{\mathrm{W},b}(\mathcal{L}(X_n), \mu) \ge 8\varepsilon$ for every $n \in \mathbb{N}$. Then there exists a sequence $(f_n)_{n\in\mathbb{N}}$ in $\mathcal{F}_{\mathrm{W},b}$ such that $\mathbb{E}[f_n(X_n)] - \int_S f_n \,\mathrm{d}\mu \ge 7\varepsilon$ 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 $\mu(C) \ge 1-\varepsilon/b$ and $\mathbb{P}[X_n \in C] \ge 1-\varepsilon/b$ for every $n \in \mathbb{N}$. By the Arzelà–Ascoli theorem, there exists a subsequence $(f_{n_k})_{k\in\mathbb{N}}$ converging uniformly to a function $f\colon C \to \mathbb{R}$. Show that $\|f\|_\infty \le b$ and $\mathrm{Lip}(f) \le 1$. Apply (b), 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 $h \in \mathcal{F}_{\mathrm{W},b}$ as well as $\|f_n - h\|_\infty \le 2b$. Take k_ε so large that $|f_{n_k}(x) - h(x)| \le \varepsilon$ for all $x \in C$ and $k \ge k_\varepsilon$. Then

$$\mathbb{E}[f_{n_k}(X_{n_k})] - \int_S f_{n_k} \,\mathrm{d}\mu \le 6\varepsilon + \mathbb{E}[h(X_{n_k})] - \int_S h \,\mathrm{d}\mu, \quad k \ge k_\varepsilon,$$

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 well 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 (Unbiased Poisson approximation)**.** *Let $X_1, \dots, X_m$ 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*

$$d_{\mathrm{TV}}\big(\mathrm{Poisson}(\lambda)\,, \mathcal{L}(W)\big) \le \frac{1-\mathrm{e}^{-\lambda}}{\lambda} \sum_{i=1}^m p_i^2, \tag{3.21}$$

see 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 $\lambda \searrow 0$). In addition,

$$d_{\mathrm{W}}\big(\mathrm{Poisson}(\lambda)\,, \mathcal{L}(W)\big) \le \min\biggl\{1, \frac{4}{3}\sqrt{\frac{2}{\mathrm{e}\,\lambda}}\biggr\} \sum_{i=1}^m p_i^2. \tag{3.22}$$

Remark 3.24. Since $\mathrm{e}^{-\lambda} > 0$ and $1 - \mathrm{e}^{-\lambda} \le \lambda$, we have the upper bound

$$\frac{1-\mathrm{e}^{-\lambda}}{\lambda} \le \min\Bigl\{1, \frac{1}{\lambda}\Bigr\}, \qquad \lambda > 0, \tag{3.23}$$

which is illustrated in Figure 3.1.

Remark 3.25. In the Theorem 3.23, the Poisson parameter λ is chosen such that the expectations of W and $N \sim \mathrm{Poisson}(\lambda)$ agree, which corresponds to the calibration method (3.9). If $p_1, \dots, p_m$ are small, then the estimate (3.21) can be improved by using the calibration method of (3.7) to obtain the bound (3.37) from Exercise 3.33, see also Remark 3.34 and Table 3.2.

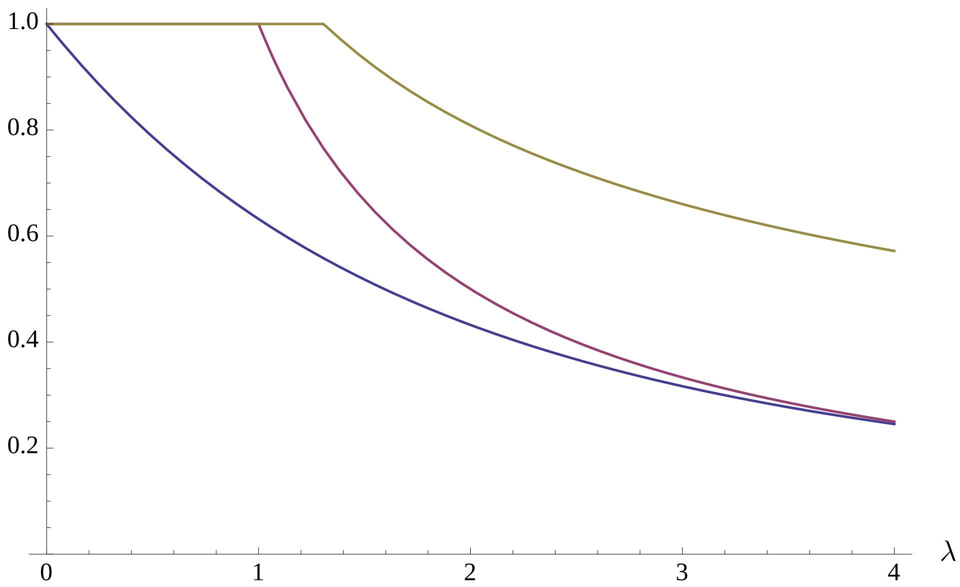

Figure 3.1: The factor $[0,\infty) \ni \lambda \mapsto (1-\mathrm{e}^{-\lambda})/\lambda$ *in* (3.21) *and its upper bound* $\lambda \mapsto \min\{1, 1/\lambda\}$ *from* (3.23)*. The upper line is the factor from* (3.22) *with a kink at* $\lambda \approx 1.144$*.*

3.4.1 Results Using an Elementary Coupling Method

In this subsection we want to prove a weaker version of the unbiased approximation bound (3.22), namely the estimate

$$d_{\mathrm{W}}\big(\mathrm{Poisson}(\lambda)\,,\mathcal{L}(W)\big) \le \sum_{i=1}^{m} p_i^2, \tag{3.24}$$

which actually agrees with (3.22) for $\lambda \le \frac{32}{9\,\mathrm{e}} \approx 1.3080$, and which by Lemma 3.18(c) also implies

$$d_{\mathrm{TV}}\big(\mathrm{Poisson}(\lambda)\,,\mathcal{L}(W)\big) \le \sum_{i=1}^{m} p_i^2, \tag{3.25}$$

which is (3.21) without the factor $(1-\mathrm{e}^{-\lambda})/\lambda$, see Le Cam [36]. This can be done using the so-called coupling method (see Lindvall [38] for a textbook presentation). The proof below will be slightly more general, so that we can also treat some biased Poisson approximations.

Example 3.26 (Comparison of Poisson approximation bounds)**.** To see that the difference between the estimates (3.21) and (3.25) can be substantial, consider the case $p_1 = \cdots = p_m = 1/\sqrt{m}$. Then the right-hand side of (3.25) is 1 and therefore useless (see Remark 3.8), while the right-hand side of (3.21) 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.

m	K.–S. dist.	total var.	Bound (3.21)	Percentage	Bound (3.37)
1	0.367879	0.632121	0.632121	100.00 %	0.63212
2	0.169948	0.327278	0.535197	61.15 %	0.67845
3	0.101422	0.199464	0.475205	41.97 %	0.64008
4	0.093506	0.173882	0.432332	40.22 %	0.61371
5	0.085456	0.144796	0.399416	36.25 %	0.59765
6	0.077148	0.134432	0.373001	36.04 %	0.58665
8	0.062349	0.108886	0.332656	32.73 %	0.57225
10	0.050520	0.091307	0.302842	30.15 %	0.56305
20	0.034107	0.060421	0.221053	27.33 %	0.54209
50	0.020212	0.036644	0.141301	25.93 %	0.52539
100	0.014170	0.025829	0.099996	25.83 %	0.51755
200	0.009583	0.017777	0.070711	25.14 %	0.51222
500	0.005915	0.011105	0.044721	24.83 %	0.50762
1000	0.004131	0.007805	0.031623	24.68 %	0.50536
2000	0.002880	0.005481	0.022361	24.51 %	0.50377
5000	0.001792	0.003450	0.014142	24.40 %	0.50237
10^4	0.001258	0.002435	0.010000	24.35 %	0.50168

Table 3.1: Quality of Poisson approximation. For various $m \in \mathbb{N}$ the second column gives the Kolmogorov–Smirnov distance, see Definition 3.9, of the binomial distribution $\mathrm{Bin}(m, 1/\sqrt{m})$ *and the Poisson distribution* $\mathrm{Poisson}(\sqrt{m})$*, while the third column gives the total variation distance. The fourth column gives the upper bound (3.21) 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 (3.25) always gives 1 in this example and is not shown; instead the last column shows the slightly improved bound from (3.37) when* $\mathrm{Poisson}(-m\log(1 - \min\{1/\sqrt{m}, 1 - 1/\mathrm{e}\}))$ *is used for the approximation (where the minimum is $1/\sqrt{m}$ for $m \ge 3$). It converges to $1/2$.*

Proof of (3.24) *using the coupling method.* Since the estimate (3.24) 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

$$\mathbb{P}_i(\{n\}) = \begin{cases} 1 - p_i & \text{for } n = 0, \\ \lambda_i^n \,\mathrm{e}^{-\lambda_i}/n! & \text{for } n \in \mathbb{N}, \\ \mathrm{e}^{-\lambda_i} - (1 - p_i) & \text{for } n = -1, \end{cases} \tag{3.26}$$

where $\lambda_i \in [0, \infty)$ satisfies

$$\lambda_i \le -\log(1 - p_i) \quad \text{if } p_i < 1, \tag{3.27}$$

see (3.8), so that $\mathbb{P}_i(\{-1\}) \ge 0$. 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

m	K.–S. dist.	total var.	Bound (3.21)	Percentage	Bound (3.37)
1	0.367879	0.632121	0.632121	100.00 %	0.632121
2	0.117879	0.198181	0.316060	62.70 %	0.306853
3	0.071583	0.114848	0.210707	54.51 %	0.189070
4	0.051473	0.080993	0.158030	51.25 %	0.136954
5	0.040199	0.062581	0.126424	49.50 %	0.107426
6	0.032982	0.050997	0.105353	48.41 %	0.088392
7	0.027963	0.043035	0.090303	47.66 %	0.075096
8	0.024271	0.037225	0.079015	47.11 %	0.065280
9	0.021440	0.032797	0.070236	46.70 %	0.057736
10	0.019201	0.029312	0.063212	46.37 %	0.051755
20	0.009394	0.014211	0.031606	44.96 %	0.025427
50	0.003710	0.005583	0.012642	44.16 %	0.010067
100	0.001847	0.002775	0.006321	43.90 %	0.005017
200	0.000922	0.001384	0.003161	43.78 %	0.002504
500	0.000368	0.000552	0.001264	43.70 %	0.001001
1000	0.000184	0.000276	0.000632	43.67 %	0.000500
2000	0.000092	0.000138	0.000316	43.66 %	0.000250
5000	0.000037	0.000055	0.000126	43.65 %	0.000100
10^4	0.000018	0.000028	0.000063	43.65 %	0.000050

Table 3.2: Quality of Poisson approximation as in Table 3.1, but here the binomial distribution $\mathrm{Bin}(m, 1/m)$ *is approximated by the Poisson distribution* $\mathrm{Poisson}(1)$*. In this example, the elementary coupling bound* (3.25) *always gives* $1/m$ *and is greater than* (3.21) *by the factor* $1/(1-\mathrm{e}^{-1}) \approx 1.58198$*; it is not shown here. The last column shows the improved bound from* (3.37)*, when* $\mathrm{Poisson}(-m\log(1-1/m))$ *for* $m \ge 2$ *is used for the approximation. It gives a better upper bound for the corresponding approximation than* (3.21)*, but the expectations of the two distributions do not agree.*

$i \in \{1, \dots, m\}$ and $\omega = (\omega_1, \dots, \omega_m) \in \Omega$, define

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

and

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

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

$$N := N_1 + \dots + N_m \sim \mathrm{Poisson}(\lambda) \quad \text{with} \quad \lambda := \lambda_1 + \dots + \lambda_m.$$

All together we now have the means to derive the upper estimate (3.24). Using the upper bound (3.19) and the triangle inequality,

$$d_{\mathrm{W}}\big(\mathcal{L}(N), \mathcal{L}(W)\big) \leq \mathbb{E}[|N - W|] \leq \sum_{i=1}^{m} \mathbb{E}[|N_i - X_i|]\,. \tag{3.28}$$

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

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

Since $\mathbb{E}[N_i] \overset{(3.3)}{=} \lambda_i$ and $\mathbb{E}[X_i] \overset{(2.3)}{=} p_i$ as well as $\mathbb{P}[N_i = 0,\, X_i = 1] = \mathbb{P}_i(\{-1\}) \overset{(3.26)}{=} \mathrm{e}^{-\lambda_i} + p_i - 1$, it follows that

$$\mathbb{E}[|N_i - X_i|] = \lambda_i - p_i + 2(\mathrm{e}^{-\lambda_i} + p_i - 1), \qquad i \in \{1, \dots, m\}. \tag{3.29}$$

Note that by (3.8), the condition (3.27) allows the choice $\lambda_i := p_i$ for all $i \in \{1, \dots, m\}$, which corresponds to the unbiased calibration from (3.9). Since the function $f\colon [0, \infty) \to \mathbb{R}$ with $f(x) := 2(\mathrm{e}^{-x} + x - 1)$ satisfies $f(0) = f'(0) = 0$, hence by applying the fundamental theorem of calculus twice,

$$f(x) = \int_0^x f'(y)\,\mathrm{d}y = \int_0^x \int_0^y \underbrace{f''(z)}_{\leq 2}\,\mathrm{d}z\,\mathrm{d}y \leq \int_0^x 2y\,\mathrm{d}y = x^2 \tag{3.30}$$

for all $x \in [0, \infty)$. Combining (3.28), (3.29) and applying (3.30) gives (3.24). $\square$

Remark 3.27. By omitting the application of (3.30) in the above proof, we obtain a slightly better estimate for the unbiased Poisson approximation with $\lambda = \mathbb{E}[W]$, namely

$$d_{\mathrm{W}}\big(\mathrm{Poisson}(\lambda)\,, \mathcal{L}(W)\big) \leq 2 \sum_{i=1}^{m} (\mathrm{e}^{-p_i} + p_i - 1), \tag{3.31}$$

see Figure 3.2. By Lemma 3.18(c), the result (3.31) implies the same upper bound for $d_{\mathrm{TV}}\big(\mathrm{Poisson}(\lambda)\,, \mathcal{L}(W)\big)$.

Remark 3.28 (Biased Poisson approximation in the Wasserstein metric)**.** When we willing to accept a biased Poisson approximation, i.e. $\lambda \neq \mathbb{E}[W]$, then we can improve the upper bound in (3.31) by optimizing in (3.29). The partial derivative of the right-hand side of (3.29) w.r.t. λ_i is given by $[0, \infty) \ni \lambda_i \mapsto 1 - 2\,\mathrm{e}^{-\lambda_i}$, which changes sign at $\lambda_i = \log 2$. Hence the right-hand side of (3.29) is decreasing for $\lambda_i \in [0, \log 2]$ and increasing afterwards. Taking the constraint (3.27) into account, we minimize the right-hand side of (3.29) by taking

$$\lambda_i := -\log(1 - \min\{p_i, \tfrac{1}{2}\}) \qquad i \in \{1, \dots, m\}. \tag{3.32}$$

Substitution into (3.28) leads to an improved version of (3.31) for the adjusted $\lambda := \lambda_1 + \dots + \lambda_m$, namely

$$d_{\mathrm{W}}\big(\mathrm{Poisson}(\lambda)\,, \mathcal{L}(W)\big) \leq 2 \sum_{i=1}^{m} \Big(\mathrm{e}^{-\lambda_i} + \frac{\lambda_i + p_i}{2} - 1\Big) \tag{3.33}$$

see Figure 3.2.

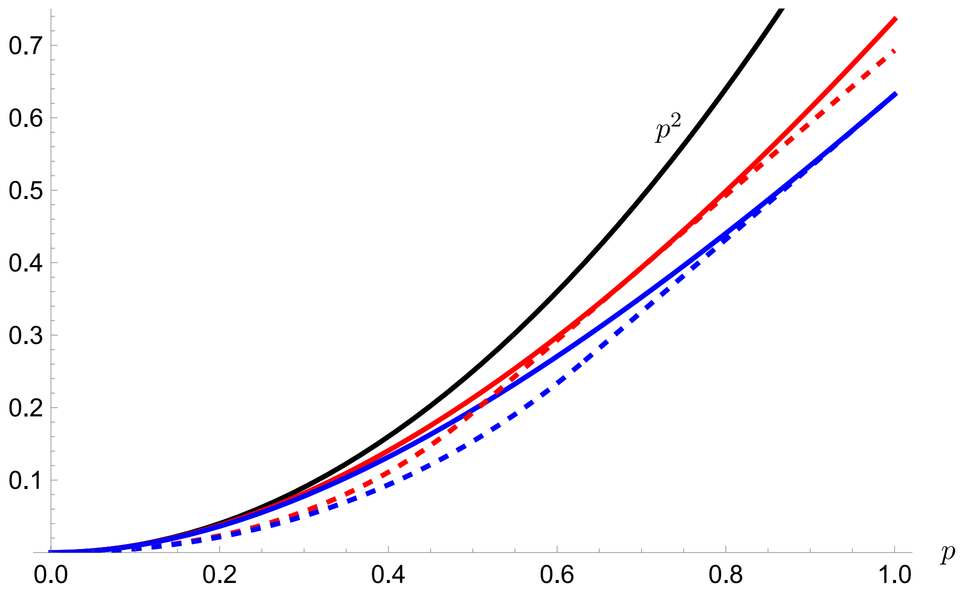

Figure 3.2: Comparison of the individual terms of the Poisson approximation bounds as functions of the Bernoulli success probability $p \in [0,1]$. Namely the general bound $p \mapsto p^2$ from (3.24) and (3.25) (upper black curve) as well as the improved unbiased Poisson approximations with $p \mapsto 2(\mathrm{e}^{-p} + p - 1)$ from (3.31) (Wasserstein metric, red curve) and $p \mapsto p(1-\mathrm{e}^{-p})$ from (3.34) (total variation metric, blue curve). The estimates for the quality of the biased Poisson approximations are illustrated by the dashed curves of corresponding colour, namely $p \mapsto 2\,\mathrm{e}^{-\lambda} + \lambda + p - 2$ with $\lambda := -\log(1-\min\{p, 1/2\})$ from (3.33) for the Wasserstein metric (dashed red curve, linear for $p \ge 1/2$) and $p \mapsto p - \lambda\,\mathrm{e}^{-\lambda}$ with $\lambda := -\log(1-\min\{p, 1-1/\mathrm{e}\})$ from (3.37) for the total variation metric (dashed blue curve, linear for $p \ge 1 - 1/\mathrm{e} \approx 0.6321$). See also Figure 3.3.

Remark 3.29 (Unbiased Poisson approximation in the total variation metric)**.** An additional slight improvement, see Figure 3.2, in the unbiased case $\lambda = \mathbb{E}[W]$, namely

$$d_{\mathrm{TV}}\big(\mathrm{Poisson}(\lambda)\,,\mathcal{L}(W)\big) \le \sum_{i=1}^{m} p_i(1-\mathrm{e}^{-p_i}), \tag{3.34}$$

is derived below by estimating the total variation distance directly. Note that for $m = 1$, estimate (3.34) agrees with (3.21).

To derive (3.34), define $A = \{n \in \mathbb{N}_0 \mid \mathbb{P}[N = n] > \mathbb{P}[W = n]\}$. By Lemma 3.18(a),

$$\begin{aligned}
&d_{\mathrm{TV}}\big(\mathcal{L}(N), \mathcal{L}(W)\big)\\
&\qquad = \mathbb{P}[N \in A] - \mathbb{P}[W \in A]\\
&\qquad = \mathbb{P}[\underbrace{N \in A}_{\text{omit}}, N \ne W] + \mathbb{P}[\underbrace{N \in A, N = W}_{\subseteq \{W \in A\}}] - \mathbb{P}[W \in A]\\
&\qquad \le \mathbb{P}[N \ne W] \le \sum_{i=1}^{m} \mathbb{P}[N_i \ne X_i]\,,
\end{aligned} \tag{3.35}$$

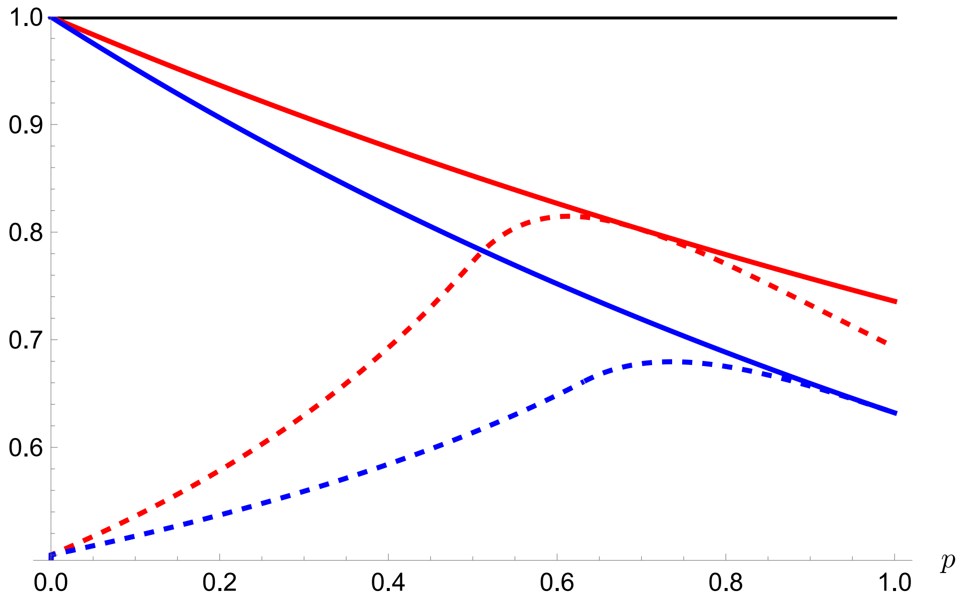

Figure 3.3: Comparison of the bounds from Figure 3.2 relative to the upper bound $[0,1] \ni p \mapsto p^2$,

where we used in the last estimate that $N_1 + \cdots + N_m \neq X_1 + \cdots + X_m$ is only possible if $N_i \neq X_i$ for at least one $i \in \{1, \ldots, m\}$. Furthermore, using (3.26),

$$\begin{aligned} \mathbb{P}[N_i \neq X_i] &= 1 - \mathbb{P}[N_i = X_i] = 1 - \mathbb{P}_i(\{0,1\}) \\ &= 1 - (1 - p_i + \lambda_i \,\mathrm{e}^{-\lambda_i}) = p_i - \lambda_i \,\mathrm{e}^{-\lambda_i} . \end{aligned} \tag{3.36}$$

In the unbiased case $\lambda_i := p_i$ for all $i \in \{1, \ldots, m\}$, see (3.9), the combination of (3.35) and (3.36) yields the estimate (3.34).

Exercise 3.30 (Comparison of Poisson approximation bounds)**.** Prove directly that the right-hand side of (3.34) is indeed smaller than the right-hand side of (3.31). Hint: Use the method from (3.30).

Exercise 3.31 (Upper bound for the total variation metric)**.** Let (S, d) be a metric space with $(\mathcal{S} \otimes \mathcal{S})$-measurable metric d, and let X and Y be two S-valued random variables, defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Prove that $d_{\mathrm{TV}}(\mathcal{L}(X), \mathcal{L}(Y)) \le \mathbb{P}[X \neq Y]$. Hint: See Exercise 3.19(a) and (3.35). Remark: Without the measurability of d, the outer $\mathbb{P}$-measure of $\{X \neq Y\}$ can be used.

Remark 3.32 (Cancellation of individual Poisson approximation errors)**.** 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, see (3.26), there is a trade-off for the large values $N_i \ge 2$, 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 (3.28) and (3.35) do not take this cancellation effect of individual approximation errors into account, hence there is room for improvement. The Stein–Chen method introduced below does this in an ingenious way, see Example 3.26 for a comparison.

Exercise 3.33 (Biased Poisson approximation for the total variation metric)**.** Let $X_1, \dots, X_m$ be independent Bernoulli random variables with $p_i := \mathbb{P}[X_i = 1]$ for all $i \in \{1, \dots, m\}$. Define $W = X_1 + \cdots + X_m$ and $\lambda = \lambda_1 + \cdots + \lambda_m$, where $\lambda_i := -\log(1 - \tilde{p}_i)$ with $\tilde{p}_i := \min\{p_i, 1 - 1/\mathrm{e}\}$. Use the coupling method to prove (see Figure 3.2)

$$d_{\mathrm{TV}}\bigl(\mathrm{Poisson}(\lambda), \mathcal{L}(W)\bigr) \le \sum_{i=1}^{m} \bigl(p_i - \lambda_i\, \mathrm{e}^{-\lambda_i}\bigr) \tag{3.37}$$

Hint: Optimize in (3.36) taking (3.27) into account.

Remark 3.34 (Comparison of biased and unbiased Poisson approximation)**.** If $p_1, \dots, p_m$ and their sum $p_1 + \cdots + p_m$ are small, then the approximation used in Exercise 3.33 gives the (without much work obtainable) upper bound (3.37) for the approximation error, which can be as small as about half the size of the bound (3.21) in Theorem 3.23 relying on the Stein–Chen method. To be specific, consider the example $p_i := 1 - \mathrm{e}^{-1/m^2}$ for all $i \in \{1, \dots, m\}$, hence p_i and $\tilde{p}_i$ in Exercise 3.33 agree and $\lambda_i = 1/m^2$. As preparation, note that

$$\mathrm{e}^x - 1 - x = \sum_{n=2}^{\infty} \frac{x^n}{n!} = \frac{x^2}{2} \sum_{n=2}^{\infty} \underbrace{\frac{2}{n(n-1)}}_{\le 1} \frac{x^{n-2}}{(n-2)!} \le \frac{x^2}{2}\, \mathrm{e}^x, \qquad x \ge 0.$$

Then the right-hand side of (3.37) simplifies to

$$m\Bigl(1 - \mathrm{e}^{-1/m^2} - \frac{1}{m^2}\, \mathrm{e}^{-1/m^2}\Bigr) = m\, \mathrm{e}^{-1/m^2} \underbrace{\Bigl(\mathrm{e}^{1/m^2} - 1 - \frac{1}{m^2}\Bigr)}_{\le \frac{1}{2m^4}\, \mathrm{e}^{1/m^2}} \le \frac{1}{2m^3}.$$

On the other hand, with $\lambda := p_1 + \cdots + p_m = m(1 - \mathrm{e}^{-1/m^2})$, the right-hand side of (3.21) yields, for large $m \in \mathbb{N}$,

$$\frac{1 - \mathrm{e}^{-\lambda}}{\lambda} m (1 - \mathrm{e}^{-1/m^2})^2 = (1 - \mathrm{e}^{-\lambda})(1 - \mathrm{e}^{-1/m^2}) = \frac{1}{m^3} - \frac{1}{2m^4} + \mathcal{O}\Bigl(\frac{1}{m^5}\Bigr)$$

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

Exercise 3.35 (Unbiased normal approximation)**.** Using a computer and suitable software of your choice, compute similarly to Table 3.1 the Kolmogorov–Smirnov distance, see Definition 3.9, between the binomial distribution $\mathrm{Bin}(m, 1/\sqrt{m})$ and the normal distribution $\mathcal{N}(\sqrt{m}, \sqrt{m} - 1)$ with expecation $\sqrt{m}$ 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),

$$\lambda\,\mathbb{P}[N = n-1] = \frac{\lambda^n}{(n-1)!}\,\mathrm{e}^{-\lambda} = n\,\mathbb{P}[N = n]\,, \qquad n \in \mathbb{N}, \tag{3.38}$$

and this recursion relation[20] uniquely determines the Poisson distribution with parameter λ: If N is $\mathbb{N}_0$-valued, then (3.38) implies by induction that

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

and $\mathbb{P}[N = 0] = \mathrm{e}^{-\lambda}$ gives the correct starting probability to obtain a probability distribution. The recursion (3.38) implies that, for every function $g\colon \mathbb{N}_0 \to \mathbb{R}$ which is bounded below,

$$\begin{aligned}\lambda\,\mathbb{E}[g(N+1)] &= \sum_{n=1}^{\infty} \lambda\, g(n)\,\mathbb{P}[N = n-1] \\ &= \sum_{n=1}^{\infty} n\, g(n)\,\mathbb{P}[N = n] = \mathbb{E}[N g(N)]\,.\end{aligned} \tag{3.39}$$

Relation (3.39) applied to the functions $g_n = \mathbb{1}_{\{n\}}$ for $n \in \mathbb{N}$ reduces to (3.38), hence (3.39) for all indicator functions $\mathbb{1}_{\{n\}}\colon \mathbb{N}_0 \to \{0,1\}$ 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 (3.39) is violated for at least one bounded $g\colon \mathbb{N}_0 \to \mathbb{R}$.

Exercise 3.36 (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 $\mathbb{1}_{(n,n+1)}$ for $n \in \mathbb{N}_0$.

The idea of the Stein–Chen method[21] 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

$$\lambda\,\mathbb{E}[g(W+1)] - \mathbb{E}[Wg(W)] \tag{3.40}$$

of inequality in (3.39), for a specific function g or a suitable collection of them.

If $\lambda = 0$, then $p_1 = \cdots = p_m = 0$, and $N = W \equiv 0$ almost surely, hence (3.21) and (3.22) 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

$$d_{\text{TV}}(\mathcal{L}(W), \mathcal{L}(N)) = \mathbb{P}[W \in A] - \mathbb{P}[N \in A]\,. \tag{3.41}$$

[20] The recursion relation (3.38) also shows that the Poisson distribution with parameter $\lambda \geq 0$ agrees with the Panjer$(0,\lambda,0)$ distribution, see Example 5.21 below.

[21] Named after Charles M. Stein (1920–2016) and his former Ph.D. student Louis H. Y. Chen, Emeritus Professor at the National University of Singapore.

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

$$f = \mathbb{1}_A - \mathbb{P}[N \in A]\,. \tag{3.42}$$

Note that

$$\mathbb{E}[f(W)] = \mathbb{P}[W \in A] - \mathbb{P}[N \in A] \tag{3.43}$$

is the right-hand side of (3.41), 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 (3.43) by (3.40). We do this more general, not just for the function f from (3.42), because we also want to use the result for the Wasserstein metric in Subsection 3.4.3 below.

Lemma 3.37 (Solution of the Stein equation)**.** *Let $f\colon \mathbb{N}_0 \to \mathbb{R}$ be a function and $\lambda > 0$. Then the function $g\colon \mathbb{N}_0 \to \mathbb{R}$ given by $g(0) = 0$ (or any other value) and*[22]

$$g(l+1) = \frac{l!}{\lambda^{l+1}} \sum_{n=0}^{l} \frac{\lambda^n}{n!} f(n), \qquad l \in \mathbb{N}_0, \tag{3.44}$$

solves the so-called Stein equation *for the Poisson distribution with parameter λ, i.e.*

$$f(l) = \lambda g(l+1) - l g(l), \qquad l \in \mathbb{N}_0. \tag{3.45}$$

Proof. By direct inspection of (3.44) for $l = 0$, we get that $\lambda g(1) = f(0)$, which is (3.45) for $l = 0$. For every $l \in \mathbb{N}$,

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

□

Exercise 3.38. In the setting of Lemma 3.37, let $N \sim \text{Poisson}(\lambda)$ and show that

$$g(l+1) = \frac{\mathbb{E}[f(N)\mathbb{1}_{\{N \le l\}}]}{\lambda\, \mathbb{P}[N = l]}, \qquad l \in \mathbb{N}_0. \tag{3.46}$$

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 $\{0, \dots, m\}$, the expectations $\mathbb{E}[g(W+1)]$, $\mathbb{E}[Wg(W)]$ and $\mathbb{E}[f(W)]$ are well defined and the Stein equation (3.45) implies that

$$\mathbb{E}[f(W)] = \lambda\, \mathbb{E}[g(W+1)] - \mathbb{E}[Wg(W)]\,. \tag{3.47}$$

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 (3.42).

[22] The representation (3.44) can be derived from the Stein equation (3.45) by recursion, hence there is no need to memorize it.

Lemma 3.39. *For every function* $g\colon \mathbb{N}_0 \to \mathbb{R}$,

$$\lambda\,\mathbb{E}[g(W+1)] - \mathbb{E}[W g(W)] \le \max_{l\in\{1,\dots,m\}} \Delta g(l) \sum_{i=1}^m p_i^2 \tag{3.48}$$

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

Proof. Using that $\lambda = p_1 + \dots + p_m$ and $W = X_1 + \dots + X_m$, we obtain for the left-hand side of (3.48) that

$$\lambda\,\mathbb{E}[g(W+1)] - \mathbb{E}[W g(W)] = \sum_{i=1}^m \bigl(p_i\,\mathbb{E}[g(W+1)] - \mathbb{E}[X_i\, g(W)]\bigr).$$

Define $W_i = W - X_i$ for every $i \in \{0,\dots,m\}$. 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

$$\mathbb{E}[X_i\, g(W)] = \sum_{j\in\{0,1\}} \mathbb{E}\bigl[X_i\, g(W_i + X_i)\mathbb{1}_{\{X_i = j\}}\bigr] = \mathbb{E}[g(W_i+1)]\, p_i.$$

Repeating this reasoning and noting that W_i takes values in $\{0,\dots,m-1\}$,

$$\begin{aligned}\lambda\,\mathbb{E}[g(W+1)] - \mathbb{E}[W g(W)] &= \sum_{i=1}^m p_i \underbrace{\mathbb{E}\bigl[g(W_i + X_i + 1) - g(W_i+1)\bigr]}_{\substack{=\mathbb{E}[(g(W_i+X_i+1)-g(W_i+1))\mathbb{1}_{\{X_i=1\}}] \\ =\mathbb{E}[g(W_i+2)-g(W_i+1)]\,p_i \text{ by indep. of } W_i \text{ and } X_i \\ =\mathbb{E}[\Delta g(W_i+1)]\,p_i}} \\ &\le \max_{l\in\{1,\dots,m\}} \Delta g(l) \sum_{i=1}^m p_i^2.\end{aligned}$$

□

Combining (3.41), (3.43), (3.47) and (3.48), we just need the result of the next lemma to obtain (3.21).

Lemma 3.40. *For the function* $f = \mathbb{1}_A - \mathbb{P}[N \in A]$ *defined in* (3.42), *the solution* g *of the Stein equation* (3.45) *given by Lemma* 3.37 *satisfies* $\Delta g(l) \le (1 - \mathrm{e}^{-\lambda})/\lambda$ *for all* $l \in \mathbb{N}$ *(with equality for* $A = \{1\}$ *and* $l = 1$*).*

Proof. For every $n \in \mathbb{N}_0$ define the function

$$f_n(l) = \mathbb{1}_{\{n\}}(l) - \mathbb{P}[N = n], \qquad l \in \mathbb{N}_0. \tag{3.49}$$

Due to Lemma 3.37 and (3.46), a corresponding solution $g_n\colon \mathbb{N}_0 \to \mathbb{R}$ of the Stein equation (3.45) is given by $g_n(0) = 0$ and, for every $l \in \mathbb{N}$,

$$g_n(l+1) = \frac{\mathbb{E}[f_n(N)\mathbb{1}_{\{N\le l\}}]}{\lambda\,\mathbb{P}[N = l]} = \frac{\mathbb{1}_{\{n,n+1,\dots\}}(l) - \mathbb{P}[N \le l]}{\lambda\,\mathbb{P}[N = l]}\,\mathbb{P}[N = n], \tag{3.50}$$

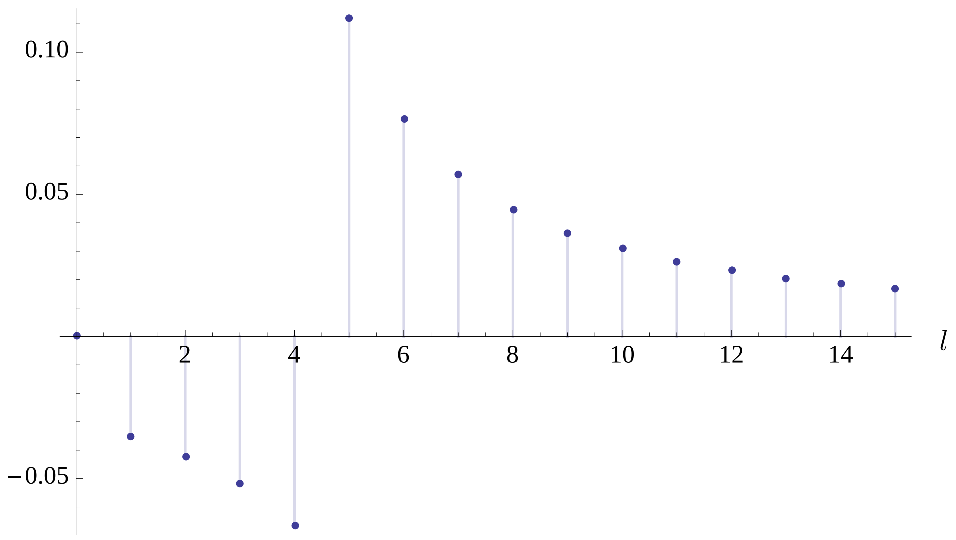

Figure 3.4: The function $\mathbb{N}_0 \ni l \mapsto g_n(l)$ from (3.50) for $\lambda = 5$ and $n = 4$. The increments of this Stein solution are estimated by (3.51).

because $\mathbb{E}[\mathbb{1}_{\{n\}}(N)\mathbb{1}_{\{N\le l\}}] = \mathbb{1}_{\{n,n+1,\dots\}}(l)\,\mathbb{P}[N=n]$. Since $A \subseteq \{0,\dots,m\}$ is finite, $f = \sum_{n\in A} f_n$. Since the Stein equation (3.45) 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 $l \in \mathbb{N}_0$. Hence it suffices to show that, see Figure 3.4,

$$\Delta g_n(l) \le \begin{cases} (1-\mathrm{e}^{-\lambda})/\lambda & \text{for } l = n \in \mathbb{N}, \\ 0 & \text{for } l \in \mathbb{N} \text{ and } n \in \mathbb{N}_0 \text{ with } l \ne n. \end{cases} \tag{3.51}$$

Using (3.50) and the recursion formula

$$\lambda\,\mathbb{P}[N = l-1] = \frac{\lambda^l}{(l-1)!}\,\mathrm{e}^{-\lambda} = l\,\mathbb{P}[N=l]\,, \qquad l \in \mathbb{N}, \tag{3.52}$$

see (3.38), we see that for $l = n \in \mathbb{N}$,

$$\begin{aligned} g_n(n+1) - g_n(n) &= \frac{1-\mathbb{P}[N\le n]}{\lambda} + \frac{\mathbb{P}[N\le n-1]}{\lambda\,\mathbb{P}[N=n-1]}\,\mathbb{P}[N=n] \quad \text{by (3.50)} \\ &= \frac{\mathbb{P}[N\ge n+1]}{\lambda} + \frac{\mathbb{P}[N\le n-1]}{n} \\ &= \frac{\mathbb{P}[N\ge n+1]}{\lambda} + \frac{1}{\lambda}\sum_{k=1}^{n} \underbrace{\frac{1}{n}}_{\le 1/k}\, \underbrace{\lambda\,\mathbb{P}[N=k-1]}_{=k\,\mathbb{P}[N=k] \text{ by (3.52)}} \\ &\le \frac{\mathbb{P}[N\ge 1]}{\lambda} = \frac{1-\mathrm{e}^{-\lambda}}{\lambda} \end{aligned}$$

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

$$g_n(l+1) - g_n(l) \overset{(3.50)}{=} \left(-\frac{\mathbb{P}[N\le l]}{\mathbb{P}[N=l]} + \frac{\mathbb{P}[N\le l-1]}{\mathbb{P}[N=l-1]}\right)\frac{\mathbb{P}[N=n]}{\lambda}.$$

The term in parentheses is negative, because by the recursion formula (3.52)

$$\frac{\mathbb{P}[N \le l-1]}{\mathbb{P}[N = l-1]} = \sum_{k=1}^{l} \frac{\lambda\,\mathbb{P}[N = k-1]}{\lambda\,\mathbb{P}[N = l-1]} = \sum_{k=1}^{l} \underbrace{\frac{k}{l}}_{\le 1} \frac{\mathbb{P}[N = k]}{\mathbb{P}[N = l]} < \frac{\mathbb{P}[N \le l]}{\mathbb{P}[N = l]}.$$

For $l \in \mathbb{N}$ and $n \in \mathbb{N}_0$ with $l > n$ we get from (3.50)

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

Again, the term in parentheses is negative, because, using (3.52),

$$\frac{\mathbb{P}[N \ge l]}{\mathbb{P}[N = l-1]} = \sum_{k=l+1}^{\infty} \frac{\lambda\,\mathbb{P}[N = k-1]}{\lambda\,\mathbb{P}[N = l-1]} = \sum_{k=l+1}^{\infty} \underbrace{\frac{k}{l}}_{>1} \frac{\mathbb{P}[N = k]}{\mathbb{P}[N = l]} > \frac{\mathbb{P}[N \ge l+1]}{\mathbb{P}[N = l]}.$$

Therefore, the estimate (3.51) for Δg_n is proved. □

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. (3.22), we can follow the strategy used in the previous subsection. Let $N \sim \text{Poisson}(\lambda)$ and let $f\colon \mathbb{N}_0 \to \mathbb{R}$ 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 3.37, the corresponding solution g of the Stein equation is given by (3.44), and Lemma 3.39 applies to g. In view of the definition of the Wasserstein metric in (3.15), all we need for (3.22) is the following lemma.

Lemma 3.41. *Let $f\colon \mathbb{N}_0 \to \mathbb{R}$ 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*

$$\Delta g \le \min\left\{1, \frac{4}{3}\sqrt{\frac{2}{\mathrm{e}\,\lambda}}\right\}.$$

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 3.38, the solution g 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 [37]; with different notation it is also given in [53, Chapter 20.1]. It will easily allow us to model joint defaults of obligors.

Definition 3.42 (Multivariate Poisson distribution)**.** Let $m \in \mathbb{N}$, consider a collection $G \subseteq \mathcal{P}(\{1,\dots,m\})$ of subsets of $\{1,\dots,m\}$ with $\varnothing \notin G$, and Poisson parameters[23] $\lambda = (\lambda_g)_{g\in G} \in [0,\infty)^G$. 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

$$N = \sum_{g\in G} c_g N_g, \tag{3.53}$$

where the vector[24] $c_g = (c_{g,1},\dots,c_{g,m})^\top \in \{0,1\}^m$ is given by

$$c_{g,i} = \mathbb{1}_g(i) = \begin{cases} 1 & \text{if } i \in g, \\ 0 & \text{if } i \notin g, \end{cases} \tag{3.54}$$

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

In the credit risk interpretation, the obligors in the group $g \subseteq \{1,\dots,m\}$ default together with Poisson intensity λ_g, independent of the other groups in G. An empty group of obligors cannot cause any default, for this reason we excluded $\varnothing$ from G. For practical applications we should assume that $\{1,\dots,m\} \subseteq \bigcup_{g\in G} g$, because otherwise there would exist obligors who can never default. If $G = \varnothing$, then (3.53) 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 $G = \mathcal{P}(\{1,\dots,m\}) \setminus \{\varnothing\}$ for greatest generality, but then there are $2^m - 1$ Poisson parameters $(\lambda_g)_{g\in G}$, which already for $m = 1\,000$ obligors are far too many to yield a practically useful model.

The next result is the multivariate generalization of Lemma 3.2.

Lemma 3.43 (Summation property of the multivariate Poisson distribution)**.** *If $N_1,\dots,N_k$ are independent with $N_i \sim \text{MPoisson}(G_i,\lambda_i,m)$ for all $i \in \{1,\dots,k\}$ with $\lambda_i = (\lambda_{i,g})_{g\in G_i}$ according to Definition 3.42, 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

$$\lambda_g = \sum_{\substack{i=1\\ G_i \ni g}}^k \lambda_{i,g}, \qquad g \in G.$$

[23] We consider $[0,\infty)^G$ as the set of all functions $\lambda\colon G \to [0,\infty)$, where the image of $g \in G$ is denoted by λ_g, hence λ 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 $\mathbb{R}^{\{1,\dots,d\}}$ and $\mathbb{N}_0^{\{1,\dots,d\}}$.

[24] The vector c_g points to a corner of the m-dimensional hypercube.

Exercise 3.44 (Proof of Lemma 3.43 under extra independence)**.** Write $N_i = \sum_{g\in G_i} c_g N_{i,g}$ with $N_{i,g} \sim \text{Poisson}(\lambda_{i,g})$ for each $i \in \{1,\dots,k\}$ and $g \in G_i$ according to Definition 3.42, where $(N_{i,g})_{g\in G_i}$ are independent for each $i \in \{1,\dots,k\}$. Under the additional assumption that the collection $(N_{i,g})_{i\in\{1,\dots,k\},g\in G_i}$ is independent, use Lemma 3.2 to prove Lemma 3.43. (Without the extra independence, the proof is given as Exercise 4.33 below.)

Remark 3.45 (Infinite divisibility of the multivariate Poisson distribution)**.** Lemma 3.43 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 $N_1,\dots,N_k$ are independent with $N_i \sim \text{MPoisson}\big(G,\lambda^{(k)},m\big)$ for every $i \in \{1,\dots,k\}$, where $\lambda^{(k)} = (\lambda_g/k)_{g\in G}$.

Lemma 3.46 (Moments of the multivariate Poisson distribution)**.** *Assume that $N = (N_1,\dots,N_m)^\top \sim \text{MPoisson}(G,\lambda,m)$. Then, with the notation from Definition 3.42, for all $i,j \in \{1,\dots,m\}$,*

$$\mathbb{E}[N_i] = \sum_{\substack{g\in G\\ g\ni i}} \lambda_g \tag{3.55}$$

and for the components of the covariance matrix of N,

$$\text{Cov}(N_i,N_j) = \sum_{\substack{g\in G\\ g\ni i,j}} \lambda_g. \tag{3.56}$$

Proof. Equation (3.55) follows from (3.53), (3.54) and (3.3). Similarly, using the bi-linearity of the covariance and the independence of $(N_g)_{g\in G}$,

$$\text{Cov}(N_i,N_j) = \sum_{\substack{g\in G\\ g\ni i}} \sum_{\substack{g'\in G\\ g'\ni j}} \underbrace{\text{Cov}\big(N_g,N_{g'}\big)}_{=0 \text{ if } g\neq g'} = \sum_{\substack{g\in G\\ g\ni i,j}} \text{Var}(N_g)\,.$$

Using (3.4), the result (3.56) follows. ☐

Remark 3.47. Note that by (3.56) the components of a multivariate Poisson distribution can only have a non-negative covariance.

Lemma 3.48 (Multivariate Poisson distribution with independent components)**.** *Consider $N = (N_1,\dots,N_m)^\top \sim \text{MPoisson}(G,\lambda,m)$ and $m \ge 2$. Then, with the notation from Definition 3.42, the following properties are equivalent:*

(a) The components $N_1,\dots,N_m$ are independent.

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

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

Proof. Note that (a) implies (b), which in turn implies (c) via (3.56). If (c) holds, then $N_g \stackrel{\text{a.s.}}{=} 0$ for all $g \in G$ with $|g| \geq 2$, hence

$$\begin{pmatrix} N_1 \\ \vdots \\ N_m \end{pmatrix} \stackrel{\text{a.s.}}{=} \sum_{\substack{i=1 \\ \{i\}\in G}}^{m} c_{\{i\}} N_{\{i\}}$$

by (3.53), hence $N_i \stackrel{\text{a.s.}}{=} N_{\{i\}}$ if $\{i\} \in G$, and $N_i \stackrel{\text{a.s.}}{=} 0$ otherwise. Since $(N_g)_{g\in G,\,|g|=1}$ are independent by Definition 3.42, 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

$$\mathbb{P}[N_g = n_g \,|\, (\Lambda_h)_{h\in G}] \stackrel{\text{a.s.}}{=} \mathbb{P}[N_g = n_g \,|\, \Lambda_g] \stackrel{\text{a.s.}}{=} \mathrm{e}^{-\Lambda_g} \frac{\Lambda_g^{n_g}}{n_g!} \tag{3.57}$$

for every $g \in G$ and $n_g \in \mathbb{N}_0$, cf. (2.11), 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 $g \in G$,

$$\begin{aligned} \mathbb{P}[N_g = n_g \text{ for all } g \in G \,|\, (\Lambda_h)_{h\in G}] &\stackrel{\text{a.s.}}{=} \prod_{g\in G} \mathbb{P}[N_g = n_g \,|\, (\Lambda_h)_{h\in G}] \\ &\stackrel{\text{a.s.}}{=} \prod_{g\in G} \frac{\Lambda_g^{n_g}}{n_g!}\,\mathrm{e}^{-\Lambda_g} \quad \text{by (3.57)}, \end{aligned} \tag{3.58}$$

cf. (2.12). The unconditional joint distribution of $(N_g)_{g\in G}$ can be obtained by integrating over the random intensities, i.e.

$$\mathbb{P}[N_g = n_g \text{ for all } g \in G] = \mathbb{E}\biggl[\prod_{g\in G} \frac{\Lambda_g^{n_g}}{n_g!}\,\mathrm{e}^{-\Lambda_g}\biggr]. \tag{3.59}$$

Exercise 3.49 (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 $K\colon [0,\infty)^G \times \mathcal{P}(\mathbb{N}_0^G) \to [0,1]$ with

$$K(\lambda, B) := \sum_{(n_g)_{g\in G}\in B} \prod_{g\in G} \frac{\lambda_g^{n_g}}{n_g!}\,\mathrm{e}^{-\lambda_g} \tag{3.60}$$

for all $\lambda = (\lambda_g)_{g\in G} \in [0,\infty)^G$ and $B \subseteq \mathbb{N}_0^G$ is a well-defined stochastic transition kernel. Hint: (3.60) can be expressed as $K(\lambda,\cdot) = \bigotimes_{g\in G} \mathrm{Poisson}(\lambda_g)$.

(b) Show that a well-defined probability measure $\mathbb{P}$ on the product space $(\Omega, \mathcal{A})$ is uniquely determined by

$$\mathbb{P}[A \times B] = \mathbb{E}_{\mathbb{P}'}[\mathbb{1}_A K(\Lambda', B)], \qquad A \in \mathcal{A}',\, B \subseteq \mathbb{N}_0^G. \tag{3.61}$$

Hint: Consider $\mathbb{P}' \otimes \nu$ on $(\Omega, \mathcal{A})$, where ν is the counting measure on $\mathbb{N}_0^G$, and consider the product in (3.60) as probability density. Alternatively, apply [33, 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 (3.57) and (3.58) are satisfied. Hint: Use (3.61) 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 (3.57) we obtain that $\mathbb{E}[N_g\,|\,\Lambda_g] \overset{\text{a.s.}}{=} \Lambda_g$ and $\operatorname{Var}(N_g\,|\,\Lambda_g) \overset{\text{a.s.}}{=} \Lambda_g$ for every $g \in G$. For the numbers $N_1, \dots, N_m$ of default events of the individual obligors $1, \dots, m$, we have the representation

$$\begin{pmatrix} N_1 \\ \vdots \\ N_m \end{pmatrix} = \sum_{g\in G} c_g N_g \tag{3.62}$$

from (3.53), hence

$$\begin{pmatrix} \mathbb{E}[N_1] \\ \vdots \\ \mathbb{E}[N_m] \end{pmatrix} = \sum_{g\in G} c_g\, \mathbb{E}\bigl[\underbrace{\mathbb{E}[N_g\,|\,\Lambda_g]}_{\overset{\text{a.s.}}{=}\,\Lambda_g}\bigr] = \sum_{g\in G} c_g\, \mathbb{E}[\Lambda_g]\,,$$

or, written out componentwise,

$$\mathbb{E}[N_i] = \sum_{\substack{g\in G\\ g\ni i}} \mathbb{E}[\Lambda_g]\,, \qquad i \in \{1, \dots, m\}, \tag{3.63}$$

cf. (3.55). 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 (3.62),

$$N := N_1 + \dots + N_m = \sum_{i=1}^m \sum_{g\in G} c_{g,i} N_g = \sum_{g\in G} |g| N_g \tag{3.64}$$

is the random variable representing the overall number of defaults in the credit portfolio. Similarly, using (3.63),

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

To calculate the variances and covariances of $N_1, \dots, N_m$, we start with a general formula, which is helpful in particular for mixture models. We will apply (3.66) below with $X \coloneqq N_g$ and the sub-σ-algebra $\mathcal{B} \coloneqq \sigma(\Lambda_g)$ containing all the information about Λ_g.

Lemma 3.50 (Law of total covariance)**.** *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 $\mathcal{B} \subseteq \mathcal{A}$ a sub-σ-algebra. Then the covariance matrix of size $c \times d$ satisfies*

$$\operatorname{Cov}(X,Y) = \mathbb{E}\big[\operatorname{Cov}(X,Y\,|\,\mathcal{B})\big] + \operatorname{Cov}\big(\mathbb{E}[X\,|\,\mathcal{B}], \mathbb{E}[Y\,|\,\mathcal{B}]\big), \tag{3.65}$$

where expectations are taken componentwise. If $c = d = 1$ and $X = Y$, then (3.65) *reduces to the law of total variance*

$$\operatorname{Var}(X) = \mathbb{E}\big[\operatorname{Var}(X\,|\,\mathcal{B})\big] + \operatorname{Var}\big(\mathbb{E}[X\,|\,\mathcal{B}]\big). \tag{3.66}$$

Remark 3.51 (Vanishing conditional covariance)**.** In the setting of Lemma 3.50, when X and Y are conditionally independent given $\mathcal{B}$, then $\operatorname{Cov}(X,Y\,|\,\mathcal{B}) \overset{\text{a.s.}}{=} 0$ by [49, Exercise 9.11(b)]. This observation can be useful when applying (3.65).

Proof of Lemma 3.50. The formula for the variance follows from the one for the covariance matrix. It therefore suffices to prove (3.65). We view X and Y as column vectors. Using the definition of the covariance matrix, subtracting and adding conditional expectations, we get that

$$\begin{aligned}
\operatorname{Cov}(X,Y) &= \mathbb{E}\big[(X - \mathbb{E}[X])\,(Y - \mathbb{E}[Y])^{\mathsf{T}}\big] \\
&= \mathbb{E}\big[\big((X - \mathbb{E}[X\,|\,\mathcal{B}]) + (\mathbb{E}[X\,|\,\mathcal{B}] - \mathbb{E}[X])\big) \\
&\qquad\quad \times \big((Y - \mathbb{E}[Y\,|\,\mathcal{B}]) + (\mathbb{E}[Y\,|\,\mathcal{B}] - \mathbb{E}[Y])\big)^{\mathsf{T}}\big].
\end{aligned}$$

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

$$\begin{aligned}
\operatorname{Cov}(X,Y) = \mathbb{E}\big[&\underbrace{\mathbb{E}[(X - \mathbb{E}[X\,|\,\mathcal{B}])(Y - \mathbb{E}[Y\,|\,\mathcal{B}])^{\mathsf{T}}\,|\,\mathcal{B}]}_{=:\operatorname{Cov}(X,Y|\mathcal{B})}\big] \\
&+ \mathbb{E}\big[\underbrace{\big(\mathbb{E}[X\,|\,\mathcal{B}] - \mathbb{E}[X]\big)}_{\mathcal{B}\text{-measurable}}\,\underbrace{\mathbb{E}\big[Y - \mathbb{E}[Y\,|\,\mathcal{B}]\,\big|\,\mathcal{B}\big]^{\mathsf{T}}}_{\overset{\text{a.s.}}{=}\,0}\big] \\
&+ \mathbb{E}\big[\underbrace{\mathbb{E}\big[X - \mathbb{E}[X\,|\,\mathcal{B}]\,\big|\,\mathcal{B}\big]}_{\overset{\text{a.s.}}{=}\,0}\,\underbrace{\big(\mathbb{E}[Y\,|\,\mathcal{B}] - \mathbb{E}[Y]\big)^{\mathsf{T}}}_{\mathcal{B}\text{-measurable}}\big] \\
&+ \underbrace{\mathbb{E}\big[\big(\mathbb{E}[X\,|\,\mathcal{B}] - \mathbb{E}[X]\big)\big(\mathbb{E}[Y\,|\,\mathcal{B}] - \mathbb{E}[Y]\big)^{\mathsf{T}}\big]}_{=\operatorname{Cov}(\mathbb{E}[X|\mathcal{B}],\mathbb{E}[Y|\mathcal{B}])}.
\end{aligned}$$

□

Corollary 3.52. *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*

$$\operatorname{Cov}(AX, BY) = \mathbb{E}\big[A\operatorname{Cov}(X,Y)\,B^{\mathsf{T}}\big] + \operatorname{Cov}\big(A\,\mathbb{E}[X]\,, B\,\mathbb{E}[Y]\big).$$

Proof. We apply (3.65) from Lemma 3.50 with $\mathcal{B} = \sigma(A, B)$. Since A and B are $\mathcal{B}$-measurable, $\mathbb{E}[AX \mid \mathcal{B}] \overset{\text{a.s.}}{=} A\,\mathbb{E}[X \mid \mathcal{B}]$ and $\mathbb{E}[BY \mid \mathcal{B}] \overset{\text{a.s.}}{=} B\,\mathbb{E}[Y \mid \mathcal{B}]$ as well as

$$\begin{aligned}\operatorname{Cov}(AX, BY \mid \mathcal{B}) &\overset{\text{a.s.}}{=} \mathbb{E}\big[(AX - \mathbb{E}[AX \mid \mathcal{B}])(BY - \mathbb{E}[BY \mid \mathcal{B}])^{\mathsf{T}} \,\big|\, \mathcal{B}\big] \\ &\overset{\text{a.s.}}{=} A\,\mathbb{E}\big[(X - \mathbb{E}[X \mid \mathcal{B}])(Y - \mathbb{E}[Y \mid \mathcal{B}])^{\mathsf{T}} \,\big|\, \mathcal{B}\big] B^{\mathsf{T}} \\ &\overset{\text{a.s.}}{=} A \operatorname{Cov}(X, Y \mid \mathcal{B})\, B^{\mathsf{T}}.\end{aligned}$$

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

Now we are in a position to calculate the variances and covariances of $N_1, \dots, N_m$ given by (3.62) as well as the variance of $N = N_1 + \dots + N_m$, provided that these default numbers have a finite expectation. Using (3.66) from Lemma 3.50 and (3.57),

$$\operatorname{Var}(N_g) = \mathbb{E}\big[\underbrace{\operatorname{Var}(N_g \mid \Lambda_g)}_{\overset{\text{a.s.}}{=} \Lambda_g \text{ by (3.4)}}\big] + \operatorname{Var}\big(\underbrace{\mathbb{E}[N_g \mid \Lambda_g]}_{\overset{\text{a.s.}}{=} \Lambda_g \text{ by (3.3)}}\big) = \mathbb{E}[\Lambda_g] + \operatorname{Var}(\Lambda_g) \tag{3.67}$$

for every $g \in G$. By the conditional independence of N_g and N_h, see (3.58), and (3.57),

$$\mathbb{E}[N_g N_h] = \mathbb{E}\big[\mathbb{E}[N_g N_h \mid (\Lambda_{g'})_{g' \in G}]\big] = \mathbb{E}\big[\underbrace{\mathbb{E}[N_g \mid \Lambda_g]}_{\overset{\text{a.s.}}{=} \Lambda_g} \underbrace{\mathbb{E}[N_h \mid \Lambda_h]}_{\overset{\text{a.s.}}{=} \Lambda_h}\big] = \mathbb{E}[\Lambda_g \Lambda_h]$$

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

$$\begin{aligned}\operatorname{Cov}(N_g, N_h) &= \mathbb{E}[N_g N_h] - \mathbb{E}[N_g]\,\mathbb{E}[N_h] \\ &= \mathbb{E}[\Lambda_g \Lambda_h] - \mathbb{E}[\Lambda_g]\,\mathbb{E}[\Lambda_h] \\ &= \operatorname{Cov}(\Lambda_g, \Lambda_h)\,,\end{aligned} \tag{3.68}$$

cf. (2.22). Using the representation (3.62), in particular $N_i = \sum_{g \in G,\, g \ni i} N_g$ and $N_j = \sum_{h \in G,\, h \ni j} N_h$, it follows with the linearity of covariance in both arguments that

$$\begin{aligned}\operatorname{Cov}(N_i, N_j) &= \sum_{\substack{g,h \in G \\ g \ni i,\, h \ni j}} \operatorname{Cov}(N_g, N_h) \\ &= \sum_{\substack{g \in G \\ g \ni i,j}} \operatorname{Var}(N_g) + \sum_{\substack{g,h \in G,\, g \neq h \\ g \ni i,\, h \ni j}} \operatorname{Cov}(N_g, N_h)\end{aligned}$$

for all obligors $i, j \in \{1, \dots, m\}$. By inserting (3.67) and (3.68),

$$\operatorname{Cov}(N_i, N_j) = \sum_{\substack{g \in G \\ g \ni i,j}} \big(\mathbb{E}[\Lambda_g] + \operatorname{Var}(\Lambda_g)\big) + \sum_{\substack{g,h \in G,\, g \neq h \\ g \ni i,\, h \ni j}} \operatorname{Cov}(\Lambda_g, \Lambda_h)\,.$$

For the case $i = j$, we obtain that

$$\operatorname{Var}(N_i) = \sum_{\substack{g\in G\\ g\ni i}} \biggl(\mathbb{E}[\Lambda_g] + \operatorname{Var}(\Lambda_g) + \sum_{\substack{h\in G\setminus\{g\}\\ h\ni i}} \operatorname{Cov}(\Lambda_g,\Lambda_h)\biggr), \qquad i\in\{1,\dots,m\}.$$

Using the representation (3.64) and formula (2.19), it follows for the total number of defaults in the portfolio that

$$\operatorname{Var}(N) = \sum_{g\in G} |g|^2 \operatorname{Var}(N_g) + \sum_{\substack{g,h\in G\\ g\neq h}} |g|\,|h| \operatorname{Cov}(N_g,N_h)\,;$$

rearranging and inserting (3.67) and (3.68), it follows that

$$\operatorname{Var}(N) = \sum_{g\in G} |g| \biggl(|g| \bigl(\mathbb{E}[\Lambda_g] + \operatorname{Var}(\Lambda_g)\bigr) + \sum_{h\in G\setminus\{g\}} |h| \operatorname{Cov}(\Lambda_g,\Lambda_h)\biggr).$$

Exercise 3.53. Rederive (2.22) using (3.65) and the conditional independence formulated in (2.12).

3.6.2 One-Factor Poisson Mixture Model

As a special case of the general multivariate Poisson mixture model, assume that $G = \{\{1\},\dots,\{m\}\}$, that there exists a single $[0,\infty)$-valued random variable Λ, let F denote its distribution function, and assume that there are parameters $\mu_1,\dots,\mu_m \ge 0$ such that $\Lambda_{\{i\}} = \mu_i\Lambda$ for all $i\in\{1,\dots,m\}$. Then $N_i = N_{\{i\}}$ by (3.62) for all $i\in\{1,\dots,m\}$ and (3.59) simplifies, i.e., for all $n_1,\dots,n_m\in\mathbb{N}_0$,

$$\begin{aligned}\mathbb{P}[N_1=n_1,\dots,N_m=n_m] &= \biggl(\prod_{i=1}^m \frac{\mu_i^{n_i}}{n_i!}\biggr)\, \mathbb{E}\bigl[\Lambda^{n_1+\dots+n_m}\,\mathrm{e}^{-\mu\Lambda}\bigr]\\ &= \biggl(\prod_{i=1}^m \frac{\mu_i^{n_i}}{n_i!}\biggr) \int_0^\infty \lambda^{n_1+\dots+n_m}\,\mathrm{e}^{-\mu\lambda}\; F(\mathrm{d}\lambda)\end{aligned} \tag{3.69}$$

with $\mu := \mu_1+\dots+\mu_m$.

Since $N_1,\dots,N_m$ are conditionally independent given Λ, the summation property of the Poisson distribution, see Lemma 3.2, implies that the conditional distribution of the sum $N = N_1+\dots+N_m$ given Λ is almost surely Poisson$(\mu\Lambda)$. Hence, for all $n\in\mathbb{N}_0$,

$$\mathbb{P}[N=n] = \int_0^\infty \mathbb{P}[N=n\,|\,\Lambda=\lambda]\,F(\mathrm{d}\lambda) = \int_0^\infty \frac{(\mu\lambda)^n}{n!}\,\mathrm{e}^{-\mu\lambda}\; F(\mathrm{d}\lambda)\,. \tag{3.70}$$

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 $\mu_1 = \dots = \mu_m = 1$, hence $\mu = m$. Then (3.69) simplifies, i.e., for all $n_1,\dots,n_m\in\mathbb{N}_0$,

$$\mathbb{P}[N_1=n_1,\dots,N_m=n_m] = \int_0^\infty \frac{\lambda^{n_1+\dots+n_m}}{n_1!\dots n_m!}\,\mathrm{e}^{-m\lambda}\; F(\mathrm{d}\lambda),$$

and (3.70) 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 $\mathbb{N}_0$-valued or, more generally, $\mathbb{N}_0^d$-valued random variables. Especially, as will be shown, a probability-generating function uniquely determines a probability distribution on $\mathbb{N}_0^d$ 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 (4.31) below. In the following we will use some multi-index notation, which we will introduce when convenient.

4.1.1 Definition and Basic Examples

Definition 4.1. For a random vector $X = (X_1, \dots, X_d)\colon \Omega \to \mathbb{N}_0^d$ define the *probability-generating function*[25] of its distribution by[26]

$$\varphi_X(s) := \mathbb{E}\Bigl[\underbrace{\prod_{i=1}^{d} s_i^{X_i}}_{=:\, s^X}\Bigr] = \sum_{n=(n_1,\dots,n_d)\in\mathbb{N}_0^d} \Bigl(\underbrace{\prod_{i=1}^{d} s_i^{n_i}}_{=:\, s^n}\Bigr)\, \mathbb{P}[X = n]\,, \tag{4.1}$$

where the series is absolutely convergent at least for all $s = (s_1, \dots, s_d) \in \mathbb{C}^d$ with $\|s\|_\infty := \max\{|s_1|, \dots, |s_d|\} \le 1$, so the generating function is defined at least on the closed and centred unit polydisk of $\mathbb{C}^d$, meaning 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 4.2 (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 $B \sim \mathrm{Bin}(1, p)$. Its probability-generating function is given by

$$\varphi_B(s) \overset{(4.1)}{=} \mathbb{P}[B = 0] + \mathbb{P}[B = 1]\, s = (1 - p) + ps = 1 + p(s - 1), \qquad s \in \mathbb{C}. \tag{4.2}$$

[25] The factorial moment generating function $s \mapsto \mathbb{E}[s^X]$, defined at least for all $s = (s_1, \dots, s_d) \in \mathbb{C}^d$ with $|s_i| = 1$ for all $i \in \{1, \dots, d\}$, extends the notion of the probability-generating function to $\mathbb{R}^d$-valued random vectors, but we will not need this extension. However, we will use the moment-generating property of the probability-generating function, see (4.25) below.

[26] Recall that $z^0 := 1$ for all $z \in \mathbb{C}$, because it is a special case of the empty product.

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

$$\varphi_N(s) \overset{(4.1)}{=} \mathbb{E}\big[s^N\big] \overset{(3.1)}{=} \sum_{n\in\mathbb{N}_0} s^n \frac{\lambda^n}{n!}\,\mathrm{e}^{-\lambda} = \mathrm{e}^{\lambda s}\,\mathrm{e}^{-\lambda} = \mathrm{e}^{\lambda(s-1)}, \qquad s \in \mathbb{C}. \tag{4.3}$$

Example 4.4 (Univariate logarithmic distribution)**.** Consider an $\mathbb{N}$-valued random variable N with univariate logarithmic distribution $\text{Log}(p)$ with parameter $p \in [0, 1)$, i.e.,

$$\mathbb{P}[N = n] = \frac{p^{n-1}}{c(p)\,n}, \qquad n \in \mathbb{N}, \tag{4.4}$$

with normalising factor[27]

$$c(p) := \sum_{n\in\mathbb{N}} \frac{p^{n-1}}{n} = \begin{cases} -\frac{\log(1-p)}{p} & \text{if } p \in (0,1), \\ 1 & \text{if } p = 0, \end{cases} \tag{4.5}$$

see the Taylor series (3.8). Using this Taylor series again, we see that

$$\varphi_N(s) \overset{(4.1)}{=} \frac{s}{c(p)} \sum_{n\in\mathbb{N}} \frac{(ps)^{n-1}}{n} = s\,\frac{c(ps)}{c(p)} = \begin{cases} \frac{\log(1-ps)}{\log(1-p)} & \text{if } p \in (0,1), \\ s & \text{if } p = 0, \end{cases} \tag{4.6}$$

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 $p \le 0.1$ 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 [31, Chapter 7], for the multivariate version see Definition 4.49 below.

Example 4.5 (Multivariate Bernoulli distribution)**.** For $d \in \mathbb{N}$ consider a random vector $B = (B_1, \dots, B_d)^\top$ with a multivariate Bernoulli distribution with parameter vector $p = (p_1, \dots, p_d)^\top \in [0, 1]^d$ satisfying $p_1 + \cdots + p_d = 1$, i.e.

$$\mathbb{P}[B = e_i] = p_i, \qquad i \in \{1, \dots, d\}, \tag{4.7}$$

where $e_i = (0, \dots, 0, 1, 0, \dots, 0)^\top \in \{0, 1\}^d$ denotes the ith unit vector with the digit 1 at position i. It is also called categorical distribution on the finite set $\{e_1, \dots, e_d\}$. In analogy to Example 4.2, we will consider this distribution as a special case of the multinomial distribution, see Example 4.19 below, and write $B \sim \text{Multinomial}(1, p_1, \dots, p_d)$. It can be used, for example, to model the time of default in a model with d periods, see Remark 7.1 below. Its probability-generating function is given by

$$\varphi_B(s) \overset{(4.1)}{=} \sum_{i=1}^{d} s_i\,\mathbb{P}[B = e_i] \overset{(4.7)}{=} \sum_{i=1}^{d} p_i s_i, \qquad s = (s_1, \dots, s_d) \in \mathbb{C}^d. \tag{4.8}$$

[27] The function c is the hypergeometric function ${}_2F_1(1, 1; 2; \cdot)$ and also the derivative of the dilogarithm Li_2.

Note that B_i is $\{0,1\}$-valued and $\mathbb{P}[B_i = 1] = p_i$, hence $B_i \sim \mathrm{Bin}(1, p_i)$ for every component $i \in \{1, \dots, d\}$ of $B = (B_1, \dots, B_d)^\mathsf{T}$, in particular $\mathbb{E}[B] = p$ and $\mathrm{Var}(B_i) = p_i(1 - p_i)$. Since $\|e_i\|_1 = 1$ for every $i \in \{1, \dots, d\}$, it follows that

$$\|B\|_1 = B_1 + \cdots + B_d \equiv 1. \tag{4.9}$$

The multivariate Bernoulli distribution has the aggregation property

$$(B_1, \dots, B_i, B_{i+1} + \cdots + B_d)^\mathsf{T} \sim \mathrm{Multinomial}(1, p_1, \dots, p_i, p_{i+1} + \cdots + p_d) \tag{4.10}$$

for every $i \in \{1, \dots, d-1\}$, and the permutation property

$$(B_{\sigma(1)}, \dots, B_{\sigma(d)})^\mathsf{T} \sim \mathrm{Multinomial}(1, p_{\sigma(1)}, \dots, p_{\sigma(d)}) \tag{4.11}$$

for every permutation σ of $\{1, \dots, d\}$. Properties (4.9), (4.10) and (4.11) will imply corresponding properties for compound distributions involving the multivariate Bernoulli distribution, see Exercises 4.20, 4.50 and 4.55 below. If $d \geq 2$, then exactly one of the components of B attains the value 1, all others are zero, hence for all $i, j \in \{1, \dots, d\}$ with $i \neq j$,

$$\mathrm{Cov}(B_i, B_j) = \mathbb{E}[\underbrace{B_i B_j}_{=0}] - \mathbb{E}[B_i]\,\mathbb{E}[B_j] = -p_i p_j, \tag{4.12}$$

which implies dependence unless $p_i = 0$ or $p_j = 0$. Together with the variance of the components of B calculated above,

$$\mathrm{Cov}(B) = \mathrm{diag}(p) - pp^\mathsf{T}, \tag{4.13}$$

where $\mathrm{diag}(p)$ denotes the diagonal matrix with the entries of p on the diagonal. For the generalizations of the properties (4.9), (4.10), (4.11) and (4.13) to the general multinomial distribution, see Exercise 4.20 below. For the general multivariate Bernoulli mixture model, see Example 4.24 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 $X = (X_1, \dots, X_d)$ following directly from (4.1) are

$$\varphi_X(0, \dots, 0) = \mathbb{P}[X = 0]\,, \tag{4.14}$$

$$\varphi_X(1, \dots, 1) = \sum_{n \in \mathbb{N}_0^d} \mathbb{P}[X = n] = 1 \tag{4.15}$$

and

$$\varphi_X^{(n)}(0, \dots, 0) = n_1! \dots n_d!\,\mathbb{P}[X = n]\,, \qquad n = (n_1, \dots, n_d) \in \mathbb{N}_0^d, \tag{4.16}$$

where $\varphi_X^{(n)} = \varphi_X^{(n_1, \dots, n_d)}$ means n_i partial derivatives with respect to the ith variable iteratively for all[28] $i \in \{1, \dots, d\}$. Because of (4.1) and (4.16), φ_X uniquely

[28] We only differentiate probability generating functions in the interior of the set where the defining power series (4.1) converges absolutely. There, by Schwarz's theorem, the order in which the partial derivatives are computed, does not matter.

determines the distribution of X and vice versa. It follows from the power series representation in (4.1) that $[0,1]^d \ni (s_1,\dots,s_d) \mapsto \varphi_X(s_1,\dots,s_d)$ is monotonically increasing, meaning that for all $s,t \in [0,1]^d$ with $s \le t$ componentwise,[29] it follows that $\varphi_X(s) \le \varphi_X(t)$. In particular, φ_X is monotonically increasing separately in every argument.

The probability-generating function φ_X contains the information about all distributions arising from X by an affine transformation with coefficients in $\mathbb{N}_0$.

Lemma 4.6. *Let $X = (X_1,\dots,X_d)^\mathsf{T}$ be an $\mathbb{N}_0^d$-valued random vector with probability-generating function φ_X, let $A = (a_{i,j})_{i\in\{1,\dots,c\},j\in\{1,\dots,d\}} \in \mathbb{N}_0^{c\times d}$ be a matrix and $b = (b_1,\dots,b_c)^\mathsf{T} \in \mathbb{N}_0^c$. Then the probability-generating function of the distribution of the $\mathbb{N}_0^c$-valued random vector $AX+b$ satisfies*

$$\varphi_{AX+b}(s) = s^b \varphi_X(t_1,\dots,t_d) \quad \textit{with} \quad t_j := \prod_{i=1}^{c} s_i^{a_{i,j}}, \quad j \in \{1,\dots,d\}, \tag{4.17}$$

at least for every $s = (s_1,\dots,s_c) \in \mathbb{C}^c$ with $\|s\|_\infty \le 1$.

Proof. Using the definitions in (4.1) and (4.17),

$$\begin{aligned}\varphi_{AX+b}(s_1,\dots,s_c) = \mathbb{E}\Biggl[\prod_{i=1}^{c} s_i^{\sum_{j=1}^d a_{i,j}X_j + b_i}\Biggr] &= s^b\, \mathbb{E}\Biggl[\prod_{j=1}^{d}\biggl(\prod_{i=1}^{c} s_i^{a_{i,j}}\biggr)^{X_j}\Biggr] \\ &= s^b \varphi_X(t_1,\dots,t_d).\end{aligned}$$

□

Example 4.7. Let us rewrite (4.17) for several special cases with $b = 0$.

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

$$\varphi_{(X_1,\dots,X_c)}(s_1,\dots,s_c) = \varphi_X(s_1,\dots,s_c,1,\dots,1), \tag{4.18}$$

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

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

This result generalizes to all other c-dimensional marginal distributions by writing the ones in (4.18) as well as the zero column vectors of A at the positions not used for the marginal distribution.

(b) For the sum of all the d components of X,

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

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

[29] Hence $([0,1]^d, \le)$ is a partially ordered set, which is directed upwards as well as downwards (take the componentwise maximum and minimum, respectively).

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

$$\varphi_{(X_1,\dots,X_{c-1},X_c+\dots+X_d)}(s_1,\dots,s_c) = \varphi_X(s_1,\dots,s_{c-1},s_c,\dots,s_c), \tag{4.19}$$

because

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

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

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

$$\varphi_{(X_{\sigma(1)},\dots,X_{\sigma(d)})}(s_1,\dots,s_d) = \varphi_X(s_{\sigma^{-1}(1)},\dots,s_{\sigma^{-1}(d)}), \tag{4.20}$$

because $a_{i,j} = \delta_{\sigma(i),j} = \delta_{i,\sigma^{-1}(j)}$ for all $i,j \in \{1,\dots,d\}$.

Example 4.8 (Multivariate Bernoulli distribution revisited)**.** Assume that the random vector $B = (B_1,\dots,B_d)$ with $d \geq 2$ has a multivariate Bernoulli distribution, i.e. $B \sim \mathrm{Multinomial}(1,p_1,\dots,p_d)$ as in Example 4.5. Using the probability-generating function from (4.2) and (4.8)

$$\varphi_{B_i}(s_i) = p_i s_i + (1-p_i) = p_i s_i + \sum_{\substack{j=1 \\ j\neq i}}^{d} p_j = \varphi_B(1,\dots,1,s_i,1,\dots,1), \qquad s_i \in \mathbb{C},$$

for every $i \in \{1,\dots,d\}$, which illustrates (4.18). See Remark 4.56 below for higher-dimensional marginal distributions of $\mathrm{Multinomial}(1,p_1,\dots,p_d)$.

Example 4.9 (Finite convex combination of probability measures)**.** Fix $k \in \mathbb{N}$. For each $i \in \{1,\dots,k\}$ let $Q_i = (q_{i,n})_{n\in\mathbb{N}_0^d}$ denote a probability distribution on $\mathbb{N}_0^d$ with probability-generating function φ_{Q_i}. Furthermore, let $p = (p_1,\dots,p_k) \in [0,1]^k$ with $p_1+\dots+p_k = 1$ be a probability vector. Then the probability distribution $Q = (q_n)_{n\in\mathbb{N}_0^d}$, defined as convex combination of $Q_1,\dots,Q_k$ with weights $p_1,\dots,p_k$, is given by

$$q_n = \sum_{i=1}^{k} p_i q_{i,n}, \qquad n \in \mathbb{N}_0^d. \tag{4.21}$$

We use the notation $\mathrm{Convex}((p_i,Q_i)_{i\in\{1,\dots,k\}})$. The probability-generating function of Q is given by

$$\varphi_Q(s) \overset{(4.1)}{=} \sum_{n\in\mathbb{N}_0^d} q_n s^n \overset{(4.21)}{=} \sum_{i=1}^{k} p_i \sum_{n\in\mathbb{N}_0^d} q_{i,n} s^n \overset{(4.1)}{=} \sum_{i=1}^{k} p_i \varphi_{Q_i}(s), \tag{4.22}$$

for all $s \in \mathbb{C}^d$ for which the power series defining $\varphi_{Q_1}(s), \dots, \varphi_{Q_k}(s)$ converge, hence at least for all $s \in \mathbb{C}^d$ with $\|s\|_\infty \le 1$. For these s the equality in the centre of (4.22) is justified. For a stochastic representation of Q, consider $B = (B_1, \dots, B_k) \sim \text{Multinomial}(1, p_1, \dots, p_k)$ as in Example 4.5. For each $i \in \{1, \dots, k\}$ let X_i be an $\mathbb{N}_0^d$-valued random vector with $X_i \sim Q_i$ which is independent of B_i. Then $Y := \sum_{i=1}^k B_i X_i$ satisfies

$$\mathbb{P}[Y = n] = \sum_{i=1}^k \mathbb{P}[\underbrace{B = e_i,\, Y = n}_{= \{B_i = 1,\, X_i = n\}}] = \sum_{i=1}^k \mathbb{P}[B_i = 1]\, \mathbb{P}[X_i = n] = \sum_{i=1}^k p_i q_{i,n} \overset{(4.21)}{=} q_n$$

for each $n \in \mathbb{N}_0^d$, hence $Y \sim Q$. Note that no independence assumption between $X_1, \dots, X_k$ is necessary.

Remark 4.10 (Calculation of multivariate factorial moments)**.** 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 $i \in \{1, \dots, d\}$. At least for all $s = (s_1, \dots, s_d) \in \mathbb{C}^d$ with $\|s\|_\infty \le 1$ and $|s_i| < 1$,

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

and

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

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

$$\varphi_X^{(n)}(s) = \mathbb{E}\biggl[\prod_{i=1}^d \biggl(s_i^{X_i - n_i} \prod_{l_i = 0}^{n_i - 1} (X_i - l_i)\biggr)\biggr], \qquad n = (n_1, \dots, n_d) \in \mathbb{N}_0^d.$$

It follows from (4.1) that φ_X and its derivatives are monotonically increasing on $[0, 1)^d$. By monotone convergence for the left-sided limit at the ith position, for every $i \in \{1, \dots, d\}$,

$$\frac{\partial}{\partial s_i} \varphi_X(1, \dots, 1, s_i, 1, \dots, 1)\Big|_{s_i = 1-} = \mathbb{E}[X_i] \tag{4.23}$$

and

$$\frac{\partial^2}{\partial s_i^2} \varphi_X(1, \dots, 1, s_i, 1, \dots, 1)\Big|_{s_i = 1-} = \mathbb{E}[X_i (X_i - 1)], \tag{4.24}$$

and generally for the multivariate factorial moments,

$$\varphi_X^{(n)}(1-, \dots, 1-) = \mathbb{E}\biggl[\prod_{i=1}^d \prod_{l_i = 0}^{n_i - 1} (X_i - l_i)\biggr], \qquad n = (n_1, \dots, n_d) \in \mathbb{N}_0^d, \tag{4.25}$$

where the precaution with the left-sided limit is unnecessary for those $i \in \{1, \dots, d\}$ which satisfy $n_i = 0$. The precaution is also unnecessary when there exists a radius $r > 1$ 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 [59, Section A5.1] or an iterated application of the monotone convergence theorem that $\varphi_X^{(n)}(1-,\dots,1-)$ does not depend on the order in which the left-sided limits are taken. As the next example shows, these left-sided limits can be infinite, which is also the reason for calculating partial derivatives of φ_X only in the interior of the domain of definition.

Example 4.11 (A distribution on $\mathbb{N}$ with infinite expectation)**.** Consider an $\mathbb{N}$-valued random variable X with

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

Since $\mathbb{P}[X \in \{1,\dots,k\}] = 1 - \frac{1}{k+1} \nearrow 1$ as $k \to \infty$, this is indeed a probability distribution. Its probability-generating function φ_X satisfies

$$\varphi_X'(s) = \biggl(\sum_{n=1}^\infty \frac{s^n}{n(n+1)}\biggr)' = \sum_{n=1}^\infty \frac{s^{n-1}}{n+1}, \qquad |s| < 1.$$

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

Remark 4.12 (Variances and Covariances)**.** Consider an $\mathbb{N}_0^d$-valued random vector X. For every component $i \in \{1,\dots,d\}$ with $\mathbb{E}[X_i] < \infty$, we can use

$$\operatorname{Var}(X_i) = \mathbb{E}[X_i^2] - (\mathbb{E}[X_i])^2 = \mathbb{E}[X_i(X_i-1)] - \mathbb{E}[X_i]\bigl(\mathbb{E}[X_i] - 1\bigr) \tag{4.26}$$

as well as (4.23) and (4.24) to calculate the variance. For $i, j \in \{1,\dots,d\}$ with $i \neq j$, a special case of (4.25) is

$$\frac{\partial^2 \varphi_X}{\partial s_i\,\partial s_j}(1,\dots,1,1-,1,\dots,1,1-,1,\dots,1) = \mathbb{E}[X_i X_j], \tag{4.27}$$

where the left-sided limits are considered for the ith and jth argument. Therefore, if $\mathbb{E}[X_i] < \infty$ and $\mathbb{E}[X_j] < \infty$, then we can use

$$\operatorname{Cov}(X_i, X_j) = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\,\mathbb{E}[X_j] \tag{4.28}$$

together with (4.23) and (4.27) to calculate the covariance of X_i and X_j, which is allowed to be infinite here.

Exercise 4.13 (Factorial moments and variance of the univariate logarithmic distribution)**.** Suppose that $N \sim \operatorname{Log}(p)$ with $p \in [0,1)$, see Example 4.4. Show that

$$\mathbb{E}\biggl[\prod_{l=0}^{n-1}(N-l)\biggr] = \frac{(n-1)!\,p^{n-1}}{c(p)\,(1-p)^n}, \qquad n \in \mathbb{N}, \tag{4.29}$$

and

$$\operatorname{Var}(N) = \frac{c(p)-1}{c^2(p)\,(1-p)^2} \tag{4.30}$$

with $c(p)$ given by (4.5). For the multivariate case, see Exercise 4.50.

Exercise 4.14 (Calculating mixed moments from multivariate factorial moments)**.** Extending Exercise 2.10 to the multivariate case, show that in the polynomial ring $R[x_1,\dots,x_d]$ of d variables over a commutative ring R (with 1),

$$x^n=\sum_{\substack{l\in\mathbb{N}_0^d\\ l\le n}}\prod_{i=1}^{d}\left\{{n_i \atop l_i}\right\}\prod_{k_i=0}^{l_i-1}(x_i-k_i),\qquad n=(n_1,\dots,n_d)\in\mathbb{N}_0^d,$$

where $x=(x_1,\dots,x_d)$ and the inequality $l\le n$ is understood componentwise. Conclude that, for every $\mathbb{N}_0^d$-valued random vector $N=(N_1,\dots,N_d)$, the mixed moments can be calculated from the multivariate factorial moments given in (4.25) by the formula

$$\mathbb{E}[N^n]=\sum_{\substack{l\in\mathbb{N}_0^d\\ l\le n}}\biggl(\prod_{i=1}^{d}\left\{{n_i \atop l_i}\right\}\biggr)\,\mathbb{E}\biggl[\prod_{i=1}^{d}\prod_{k_i=0}^{l_i-1}(N_i-k_i)\biggr],\qquad n=(n_1,\dots,n_d)\in\mathbb{N}_0^d,$$

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.

Lemma 4.15 (Characterization of independence using probability-generating functions)**.** *Let $X\colon\Omega\to\mathbb{N}_0^c$ and $Y\colon\Omega\to\mathbb{N}_0^d$ be two random vectors. Then X and Y are independent if and only if $\varphi_{(X,Y)}(s,t)=\varphi_X(s)\varphi_Y(t)$ at least for all $s\in\mathbb{C}^c$ and $t\in\mathbb{C}^d$ satisfying $\|s\|_\infty\le 1$ and $\|t\|_\infty\le 1$.*

Proof. Note that $(s,t)^{(X,Y)}=\bigl(\prod_{i=1}^{c}s_i^{X_i}\bigr)\prod_{j=1}^{d}t_j^{Y_j}=s^Xt^Y$ in multi-index notation. If X and Y are independent, then

$$\varphi_{(X,Y)}(s,t)=\mathbb{E}\bigl[s^Xt^Y\bigr]=\mathbb{E}\bigl[s^X\bigr]\,\mathbb{E}\bigl[t^Y\bigr]=\varphi_X(s)\varphi_Y(t)$$

for all s and t mentioned in the lemma. The reverse direction follows because $\varphi_{(X,Y)}$ uniquely determines the distribution of (X,Y), see (4.16). $\square$

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

$$\varphi_{X+Y}(s)=\mathbb{E}\bigl[s^{X+Y}\bigr]=\mathbb{E}\bigl[s^X\bigr]\,\mathbb{E}\bigl[s^Y\bigr]=\varphi_X(s)\varphi_Y(s)\tag{4.31}$$

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

Note that Theorem 4.16 also follows from Lemma 4.15 with $c=d$ and an application of Lemma 4.6 to the $\mathbb{N}_0^{2d}$-valued random variable (X,Y) and the matrix $A:=(I_d,I_d)$, cf. Example 4.7(b) for $\mathbb{N}_0$-valued X and Y.

An application of formula (4.31) 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. Then

$$\varphi_{X+Y}(s) \overset{(4.31)}{=} \varphi_X(s)\varphi_Y(s) \overset{(4.3)}{=} \mathrm{e}^{\lambda(s-1)}\,\mathrm{e}^{\mu(s-1)} = \mathrm{e}^{(\lambda+\mu)(s-1)}, \quad s \in \mathbb{C}. \tag{4.32}$$

Hence $X + Y \sim \text{Poisson}(\lambda + \mu)$, because φ_{X+Y} uniquely determines the distribution of $X + Y$, see (4.16). $\square$

Example 4.17 (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 $i \in \{1, \ldots, m\}$,

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

hence the multiplication theorem of probability-generating functions, see (4.31), implies that

$$\varphi_N(s) = \prod_{i=1}^{m} \varphi_{B_i}(s) = (1 + p(s-1))^m, \qquad s \in \mathbb{C}. \tag{4.33}$$

Remark 4.18 (Motivation of the Poisson approximation)**.** The following observation uses generating functions to make the Poisson approximation of Theorem 3.23 plausible. Let φ_{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 φ_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 $1 + x \approx \mathrm{e}^x$ for $|x|$ small, we get

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

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 Multinomial Mixture Model

4.2.1 Multinomial Distribution

We start by introducing and discussing the multinomial distribution.

Example 4.19 (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. $B_i \sim \text{Multinomial}(1,p)$ for each $i \in \{1,\dots,m\}$. We can interpret B_i as describing the result of the ith trial, which can have d different outcomes. Then the jth component N_j of $N := B_1 + \cdots + B_m$ describes the number of outcomes of type j in a sequence of m independent trials, for every $j \in \{1,\dots,d\}$. By definition, N has a multinomial distribution,[30] which we denote by $\text{Multinomial}(m,p_1,\dots,p_d)$ or $\text{Multinomial}(m,p)$ for short. We can add the case $m=0$ with the understanding that $N \equiv 0 \in \mathbb{N}_0^d$ (empty sum convention). By (4.8), the probability-generating function of B_i is given by

$$\varphi_{B_i}(s) = \sum_{j=1}^d p_j s_j, \qquad s=(s_1,\dots,s_d)\in\mathbb{C}^d,$$

for every $i \in \{1,\dots,m\}$, hence by the multiplication theorem of probability-generating functions, see (4.31),

$$\varphi_N(s) = \prod_{i=1}^m \varphi_{B_i}(s) = \biggl(\sum_{j=1}^d p_j s_j\biggr)^m, \qquad s=(s_1,\dots,s_d)\in\mathbb{C}^d. \tag{4.34}$$

Either by using (4.16) to derive the probability mass function from φ_N, or by using the multinomial theorem to expand $\varphi_N(s) = (p_1s_1+\cdots+p_ds_d)^m$, it follows that

$$\mathbb{P}[N=n] = m!\prod_{i=1}^d \frac{p_i^{n_i}}{n_i!} = \binom{m}{n} p^n \tag{4.35}$$

in multi-index notation for all $n=(n_1,\dots,n_d)\in\mathbb{N}_0^d$ with $n_1+\cdots+n_d=m$. Note that

$$\binom{m}{n} := \binom{m}{n_1,\dots,n_d} := \frac{m!}{n_1!\dots n_d!}$$

is the multinomial coefficient, which can be defined more generally for $z\in\mathbb{C}$ by

$$\binom{z}{n} := \binom{z}{n_1,\dots,n_d} := \frac{1}{n_1!\dots n_d!}\prod_{i=0}^{n_1+\cdots+n_d-1}(z-i). \tag{4.36}$$

Exercise 4.20 (Some properties of the multinomial distribution)**.** Let $N=(N_1,\dots,N_d)\sim\text{Multinomial}(m,p_1,\dots,p_d)$ with $m\in\mathbb{N}$ trials and probability vector $p=(p_1,\dots,p_d)\in[0,1]^d$ 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 $\mathbb{E}[N]=mp$ and $\text{Var}(N_i)=mp_i(1-p_i)$ for every $i\in\{1,\dots,d\}$. (See Remark 4.56 below for higher-dimensional marginal distributions.)

[30] For the generalization to the case $p_1+\cdots+p_d\in[0,1]$, see the multivariate binomial distribution in Subsection 4.7.3 below.

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

$$(N_1, \dots, N_i, N_{i+1} + \dots + N_d) \sim \text{Multinomial}(m, p_1, \dots, p_i, p_{i+1} + \dots + p_d).$$

(d) Permutation property: For every permutation σ of $\{1, \dots, d\}$,

$$(N_{\sigma(1)}, \dots, N_{\sigma(d)}) \sim \text{Multinomial}(m, p_{\sigma(1)}, \dots, p_{\sigma(d)}).$$

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

Lemma 4.21 (Summation property of the multinomial distribution)**.** *Let $d, k \in \mathbb{N}$, $m_1, \dots, m_k \in \mathbb{N}_0$ and $p_1, \dots, p_d \in [0,1]$ with $p_1 + \dots + p_d = 1$. If $N_1, \dots, N_k$ are independent with $N_i \sim \text{Multinomial}(m_i, p_1, \dots, p_d)$ for every $i \in \{1, \dots, k\}$, then*

$$N := \sum_{i=1}^{k} N_i \sim \text{Multinomial}(m_1 + \dots + m_k, p_1, \dots, p_d).$$

Exercise 4.22. Prove Lemma 4.21 using (4.34).

Remark 4.23 (Summation property of the binomial distribution)**.** Using Lemma 4.21 for $d = 2$ and looking at the one-dimensional marginal distribution (see Exercise 4.20(b)), we obtain the summation property of the binomial distribution. Of course, this also follows directly using (4.33).

4.2.2 General Multinomial Mixture Model

As usual for mixture models, we now want to replace the probability vector $p = (p_1, \dots, p_d) \in [0,1]^d$ satisfying $p_1 + \dots + p_d = 1$ by a random probability vector. Therefore, for the remaining part of this subsection, let $P = (P_1, \dots, P_d)^\top$ denote a random vector with values in $[0,1]^d$ satisfying $P_1 + \dots + P_d = 1$. Note that $P_1, \dots, P_d$ in general are stochastically dependent, because for each $j \in \{1, \dots, d\}$ the component P_j can be expressed using all the other components.

Example 4.24 (General multivariate Bernoulli mixture model)**.** For $d \in \mathbb{N}$ consider a random vector $B = (B_1, \dots, B_d)^\top$, taking values in the set $\{e_1, \dots, e_d\}$ of unit vectors of $\mathbb{R}^d$. We generalize (4.7) from Example 4.5 by requiring

$$\mathbb{P}[B = e_i \,|\, P] \overset{\text{a.s.}}{=} P_i, \qquad i \in \{1, \dots, d\}, \tag{4.37}$$

which is equivalent to the vector equation

$$\mathbb{E}[B \,|\, P] \overset{\text{a.s.}}{=} P. \tag{4.38}$$

Furthermore $\|B_1\|_1 = B_1 + \dots + B_d \equiv 1$ corresponding to (4.9). Since $\mathcal{L}(B|P) \overset{\text{a.s.}}{=}$ Multinomial$(1, P)$, the calculation (4.8) implies for the probability-generating function of the conditional distribution $\mathcal{L}(B|P)$ that

$$\mathbb{E}[s^B \,|\, P] \overset{\text{a.s.}}{=} \varphi_{B|P}(s) \overset{\text{a.s.}}{=} \sum_{i=1}^{d} P_i s_i, \qquad s = (s_1, \dots, s_d) \in \mathbb{C}^d.$$

Hence by taking expectations,

$$\varphi_B(s) = \mathbb{E}\big[\mathbb{E}[s^B\,|\,P]\big] = \sum_{i=1}^{d} \mathbb{E}[P_i]\, s_i, \qquad s = (s_1, \dots, s_d) \in \mathbb{C}^d.$$

Since the probability-generating function determines the distribution uniquely, we see by comparison with (4.8) that $B \sim \text{Multinomial}(1, \mathbb{E}[P])$. It follows from (4.38) that

$$\mathbb{E}[B] = \mathbb{E}[P] \tag{4.39}$$

and from (4.13) combined with $\mathcal{L}(B|P) \overset{\text{a.s.}}{=} \text{Multinomial}(1, P)$ that

$$\text{Cov}(B\,|\,P) \overset{\text{a.s.}}{=} \text{diag}(P) - PP^{\mathsf{T}}, \tag{4.40}$$

and combined with $B \sim \text{Multinomial}(1, \mathbb{E}[P])$ that

$$\text{Cov}(B) = \text{diag}(\mathbb{E}[P]) - \mathbb{E}[P]\,\mathbb{E}[P]^{\mathsf{T}}.$$

Combining (3.65) from Lemma 3.50 with (4.40) and (4.38) gives the same result. The aggregation property (4.10) and the permutation property (4.11) are transferred accordingly.

Example 4.25 (General multinomial mixture model)**.** By combining Examples 4.19 and 4.24, given $m \in \mathbb{N}$, we consider general multivariate Bernoulli random vectors $B_1, \dots, B_m$ with $\mathcal{L}(B_i|P) \overset{\text{a.s.}}{=} \text{Multinomial}(1, P)$ for each $i \in \{1, \dots, m\}$, which are conditionally independent given $P = (P_1, \dots, P_d)^{\mathsf{T}}$, meaning that for all $x_1, \dots, x_m \in \{e_1, \dots, e_d\}$,

$$\mathbb{P}[B_1 = x_1, \dots, B_m = x_m\,|\,P] \overset{\text{a.s.}}{=} \prod_{i=1}^{m} \mathbb{P}[B_i = x_i\,|\,P]\,. \tag{4.41}$$

Writing $n = (n_1, \dots, n_d) \coloneqq x_1 + \dots + x_m$, which satisfies $n_1 + \dots + n_d = m$, the equations (4.37) and (4.41) imply that

$$\mathbb{P}[B_1 = x_1, \dots, B_m = x_m\,|\,P] \overset{\text{a.s.}}{=} P_1^{n_1} \cdots P_d^{n_d} = P^n. \tag{4.42}$$

Define $N = B_1 + \dots + B_m$ and take expectations in (4.42). Since for each $n = (n_1, \dots, n_d) \in \mathbb{N}_0^d$ with $n_1 + \dots + n_d = m$ there are $\binom{m}{n}$ possibilities to choose $x_1, \dots, x_m \in \{e_1, \dots, e_d\}$ with $x_1 + \dots + x_m = n$, it follows that

$$\mathbb{P}[N = n] = \binom{m}{n}\,\mathbb{E}[P^n] \tag{4.43}$$

in multi-index notation, hence (4.35) is generalized. By linearity and (4.38),

$$\mathbb{E}[N\,|\,P] \overset{\text{a.s.}}{=} \sum_{i=1}^{m} \mathbb{E}[B_i\,|\,P] \overset{\text{a.s.}}{=} mP, \quad \text{hence} \quad \mathbb{E}[N] = m\,\mathbb{E}[P]\,. \tag{4.44}$$

Using (3.65) from Lemma 3.50 and (4.44) afterwards,

$$\operatorname{Cov}(N) = \mathbb{E}\big[\operatorname{Cov}(N\,|\,P)\big] + \operatorname{Cov}(\mathbb{E}[N\,|\,P]) = \sum_{i,j=1}^{m} \mathbb{E}\big[\operatorname{Cov}(B_i, B_j\,|\,P)\big] + \operatorname{Cov}(mP)\,.$$

Due to the conditional independence of $B_1, \dots, B_m$, see (4.41), and Remark 3.51, $\operatorname{Cov}(B_i, B_j\,|\,P) \overset{\text{a.s.}}{=} 0$ for all $i, j \in \{1, \dots, m\}$ with $i \neq j$. Using (4.40) for the remaining terms,

$$\begin{aligned} \operatorname{Cov}(N) &= m\big(m \operatorname{Cov}(P) + \operatorname{diag}(\mathbb{E}[P]) - \mathbb{E}[PP^{\mathsf{T}}]\big) \\ &= m\big((m-1)\operatorname{Cov}(P) + \operatorname{diag}(\mathbb{E}[P]) - \mathbb{E}[P]\,\mathbb{E}[P]^{\mathsf{T}}\big), \end{aligned} \tag{4.45}$$

because $\operatorname{Cov}(P) = \mathbb{E}[PP^{\mathsf{T}}] - \mathbb{E}[P]\,\mathbb{E}[P]^{\mathsf{T}}$. Since $\mathcal{L}(N|P) \overset{\text{a.s.}}{=} \operatorname{Multinomial}(m, P)$ by Example 4.19, (4.34) implies that

$$\varphi_{N|P}(s) \overset{\text{a.s.}}{=} \bigg(\sum_{j=1}^{d} P_j s_j\bigg)^{m} = \sum_{\substack{n \in \mathbb{N}_0^d \\ \|n\|_1 = m}} \binom{m}{n} P^n s^n, \qquad s = (s_1, \dots, s_d) \in \mathbb{C}^d,$$

where we used the multinomial theorem to expand the power and multi-index notation as in (4.43). Taking expectations shows that

$$\varphi_N(s) = \sum_{\substack{n \in \mathbb{N}_0^d \\ \|n\|_1 = m}} \binom{m}{n} \mathbb{E}[P^n]\, s^n, \qquad s = (s_1, \dots, s_d) \in \mathbb{C}^d,$$

which also follows directly from (4.1) and (4.43).

4.2.3 Dirichlet and Dirichlet-Multinomial Distribution

To proceed, we now introduce a popular distribution for the random probability vector $P = (P_1, \dots, P_d)$, which is the multi-dimensional analogue of the beta distribution, see Remark 4.27 below. We can think of $(P_1, \dots, P_d)$ as the random lengths of intervals created by cutting the unit interval $[0, 1]$ at the points $P_1 + \dots + P_i$ for $i \in \{1, \dots, d-1\}$.

Definition 4.26 (Dirichlet distribution[31])**.** Let $d \geq 2$ be an integer. A density of the Dirichlet distribution with shape parameter vector $\alpha = (\alpha_1, \dots, \alpha_d) \in (0, \infty)^d$ on the open standard orthogonal $(d-1)$-dimensional simplex Δ_{d-1} defined in (2.32) is given by

$$f_\alpha(x) = \begin{cases} \frac{1}{B(\alpha)} \prod_{i=1}^{d} x_i^{\alpha_i - 1} & \text{for } x = (x_1, \dots, x_{d-1}) \in \Delta_{d-1}, \\ 0 & \text{for } x \in \mathbb{R}^{d-1} \setminus \Delta_{d-1}, \end{cases} \tag{4.46}$$

where $x_d \coloneqq 1 - (x_1 + \dots + x_{d-1}) > 0$ for notational reasons, and B denotes the multivariate beta function, see (2.33). We denote an $\mathbb{R}^d$-valued random probability vector $P = (P_1, \dots, P_d)$ with a Dirichlet distribution by $P \sim \operatorname{Dirichlet}(\alpha)$.

[31] Named after the German mathematician Peter Gustav Lejeune Dirichlet (1805–1859).

Remark 4.27 (Beta distribution)**.** Suppose that $d = 2$ and $P = (P_1, P_2) \sim \text{Dirichlet}(\alpha, \beta)$ with real shape parameters $\alpha, \beta > 0$. Then (4.46) simplifies to (2.35), hence P_1 has the beta distribution $\text{Beta}(\alpha, \beta)$, see Definition 2.6 (compared to the Dirichlet distribution, the component $P_2 = 1 - P_1$ is usually omitted).

Exercise 4.28 (Mixed moments and covariance matrix of the Dirichlet distribution)**.** Let $P = (P_1, \ldots, P_d)^\top \sim \text{Dirichlet}(\alpha)$ with $\alpha = (\alpha_1, \ldots, \alpha_d)^\top \in (0, \infty)^d$. Show for all $\gamma = (\gamma_1, \ldots, \gamma_d)^\top \in \mathbb{R}^d$ with $\gamma > -\alpha$ componentwise that

$$\mathbb{E}\big[P^\gamma\big] := \mathbb{E}\big[P_1^{\gamma_1} \cdots P_d^{\gamma_d}\big] = \frac{B(\alpha + \gamma)}{B(\alpha)}, \tag{4.47}$$

define the probability vector $\tilde{\alpha} = \alpha / \|\alpha\|_1$ and, using the relation (2.33) for the multivariate beta function and the functional equation (2.30) of the gamma function, conclude that

$$\mathbb{E}[P] = \tilde{\alpha} \qquad \text{and} \qquad \operatorname{Cov}(P) = \frac{\operatorname{diag}(\tilde{\alpha}) - \tilde{\alpha}\tilde{\alpha}^\top}{1 + \|\alpha\|_1}, \tag{4.48}$$

where $\operatorname{diag}(\tilde{\alpha})$ denotes the diagonal matrix with the entries of $\tilde{\alpha}$ on the diagonal.

Remark: Note that $\tilde{\alpha}$ marks the intersection in $\mathbb{R}^d$ of the one-dimensional span of α with the $(d-1)$-dimensional simplex $\{(x_1, \ldots, x_d) \in (0,1)^d \mid x_1 + \cdots + x_d = 1\}$, in which $P = (P_1, \ldots, P_d)$ takes its values, and which is also the graph of $\Delta_{d-1} \ni (x_1, \ldots, x_{d-1}) \mapsto 1 - (x_1 + \cdots + x_{d-1})$ with Δ_{d-1} given by (2.32). The structure of $\operatorname{Cov}(P)$ is determined by $\tilde{\alpha}$, its scale by $\|\alpha\|_1$. Higher values of $\|\alpha\|_1$ lead to a stronger concentration of $\mathcal{L}(P)$ around $\tilde{\alpha}$. Also note that P_i and P_j have negative covariance for $i \neq j$ in $\{1, \ldots, d\}$, as expected by the interpretation of random lengths of subintervals created by cutting the unit interval $[0, 1]$ at the points $P_1 + \cdots + P_i$ for $i \in \{1, \ldots, d-1\}$.

Here is the generalization of Lemma 2.12:

Lemma 4.29 (Biased Dirichlet distribution)**.** *Consider $d \in \mathbb{N}$ with $d \geq 2$. Assume that $P \sim \text{Dirichlet}(\alpha)$ with shape parameter vector $\alpha \in (0, \infty)^d$ and take $\gamma \in \mathbb{R}^d$ with $\gamma > -\alpha$ componentwise. Then $\mathbb{P}_{P^\gamma} P^{-1} = \text{Dirichlet}(\alpha + \gamma)$, that means the distribution of P under the P^γ-biased probability measure $\mathbb{P}_{P^\gamma}$ given by Definition 2.11 is the $\text{Dirichlet}(\alpha + \gamma)$ distribution.*

Proof. By (2.44) and (4.47), a density of the P^γ-biased probability measure $\mathbb{P}_{P^\gamma}$ is given by

$$\frac{\mathrm{d}\mathbb{P}_{P^\gamma}}{\mathrm{d}\mathbb{P}} = \frac{B(\alpha)}{B(\alpha + \gamma)} P^\gamma.$$

Let μ denote the Lebesgue–Borel measure on $\mathbb{R}^{d-1}$. Using the density f_α from (4.46) shows that, for μ-almost all $p = (p_1, \ldots, p_{d-1})$ in the open simplex Δ_{d-1} defined in (2.32), writing $\tilde{p} = (p_1, \ldots, p_{d-1}, 1 - (p_1 + \cdots + p_{d-1}))$,

$$\begin{aligned}
\frac{\mathrm{d}(\mathbb{P}_{P^\gamma} P^{-1})}{\mathrm{d}\mu}(p) &= \frac{\mathrm{d}(\mathbb{P}_{P^\gamma} P^{-1})}{\mathrm{d}(\mathbb{P} P^{-1})}(p) \cdot \frac{\mathrm{d}(\mathbb{P} P^{-1})}{\mathrm{d}\mu}(p) \\
&= \frac{B(\alpha)}{B(\alpha + \gamma)} \tilde{p}^\gamma \cdot f_\alpha(p) \overset{(4.46)}{=} \frac{\tilde{p}^{\alpha + \gamma - \mathbb{1}}}{B(\alpha + \gamma)}
\end{aligned}$$

with $\mathbb{1} = (1, \dots, 1) \in \mathbb{N}^d$, which by (4.46) is a density of the Dirichlet$(\alpha + \gamma)$ distribution. $\square$

Definition 4.30 (Dirichlet-multinomial distribution[32]). Consider $d, m \in \mathbb{N}$ with $d \geq 2$ and $\alpha \in (0, \infty)^d$. Let $P = (P_1, \dots, P_d) \sim \text{Dirichlet}(\alpha)$ as in Definition 4.26. Combining (4.43) and (4.47), the probability mass function of $N \sim \text{DirichletMultinomial}(\alpha, m)$ is given by

$$\mathbb{P}[N = n] = \binom{m}{n} \frac{B(\alpha + n)}{B(\alpha)} \tag{4.49}$$

for all $n = (n_1, \dots, n_d) \in \mathbb{N}_0^d$ with $n_1 + \dots + n_d = m$; it generalizes (2.36).

Exercise 4.31 (Expectation and covariance matrix of the Dirichlet-multinomial distribution). Consider $d, m \in \mathbb{N}$ with $d \geq 2$ and $\alpha \in (0, \infty)^d$. Let $N \sim \text{DirichletMultinomial}(\alpha, m)$. Define the probability column vector $\tilde{\alpha} = \alpha / \|\alpha\|_1$ and show that $\mathbb{E}[N] = m\tilde{\alpha}$ and

$$\operatorname{Cov}(N) = m \frac{m + \|\alpha\|_1}{1 + \|\alpha\|_1} \big(\operatorname{diag}(\tilde{\alpha}) - \tilde{\alpha}\tilde{\alpha}^\top\big).$$

Hint: Use (4.44), (4.45) and (4.48).

4.3 Application to the General Multivariate Poisson Mixture Model

We start by using the multiplication theorem (Theorem 4.16) to calculate the probability-generating function of the multivariate Poisson distribution given by Definition 3.42

Example 4.32 (Multivariate Poisson distribution). Assume that N has the multivariate Poisson distribution $\text{MPoisson}(G, (\lambda_g)_{g \in G}, m)$. By the representation (3.53), using multi-index notation, the probability-generating function is given by

$$\varphi_N(s) \overset{(4.1)}{=} \mathbb{E}\big[s^N\big] \overset{(3.53)}{=} \mathbb{E}\bigg[\prod_{g \in G} (s^{c_g})^{N_g}\bigg], \qquad s \in \mathbb{C}^m, \tag{4.50}$$

where $s^{c_g} = \prod_{i \in g} s_i$ by (3.54). Using the independence of $(N_g)_{g \in G}$ and the multiplication theorem (4.31) of probability-generating functions,

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

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

$$\varphi_N(s) = \prod_{g \in G} \exp\big(\lambda_g\big(s^{c_g} - 1\big)\big) = \exp\bigg(\sum_{g \in G} \lambda_g\big(s^{c_g} - 1\big)\bigg), \qquad s \in \mathbb{C}^m. \tag{4.51}$$

[32] Also called multivariate Pólya distribution, because it appears in Pólya's urn model, see [49].

Exercise 4.33 (Proof of the summation property of the multivariate Poisson distribution)**.** Use (4.51) and Theorem 4.16 to prove Lemma 3.43 without the extra assumption in Exercise 3.44.

Let us reconsider the general multivariate Poisson mixture model introduced in Subsection 3.6.

Example 4.34 (General multivariate Poisson mixture model from Section 3.6)**.** We start by considering the random vector $(N_1, \dots, N_m)^\mathsf{T}$ of defaults of the individual m obligors and use the representation (3.62) and the calculation (4.50) to see that the corresponding probability-generating function is given by

$$\varphi_{(N_1,\dots,N_m)}(s) = \mathbb{E}\biggl[\prod_{g\in G}(s^{c_g})^{N_g}\biggr] \qquad \text{with} \qquad s^{c_g} = \prod_{i\in g} s_i$$

at least for all $s = (s_1, \dots, s_m) \in \mathbb{C}^m$ with $\|s\|_\infty \le 1$. By conditioning on the random intensities $(\Lambda_h)_{h\in G}$ and using conditional independence, see (3.58), as well as (3.57), it follows that

$$\begin{aligned}
\varphi_{(N_1,\dots,N_m)}(s) &= \mathbb{E}\biggl[\mathbb{E}\biggl[\prod_{g\in G}(s^{c_g})^{N_g}\,\bigg|\,(\Lambda_h)_{h\in G}\biggr]\biggr] \\
&= \mathbb{E}\biggl[\prod_{g\in G}\underbrace{\mathbb{E}\bigl[(s^{c_g})^{N_g}\,\big|\,\Lambda_g\bigr]}_{\overset{\text{a.s.}}{=}\exp(\Lambda_g(s^{c_g}-1))\text{ by (4.3)}}\biggr] = \mathbb{E}\biggl[\exp\biggl(\sum_{g\in G}\Lambda_g(s^{c_g}-1)\biggr)\biggr].
\end{aligned}$$

For the number of defaults $N := N_1 + \cdots + N_m$ in the portfolio as considered in (3.64), Example 4.7(b) yields

$$\varphi_N(s) = \varphi_{(N_1,\dots,N_m)}(\underbrace{s,\dots,s}_{d\text{ entries}}) = \mathbb{E}\biggl[\exp\biggl(\sum_{g\in G}\Lambda_g(s^{|g|}-1)\biggr)\biggr],$$

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

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

and, provided $\Lambda_1, \dots, \Lambda_m$ are independent, to

$$\varphi_N(s) = \prod_{i=1}^m \mathbb{E}\bigl[\mathrm{e}^{\Lambda_i(s-1)}\bigr].$$

4.4 Properties of the Gamma Distribution

In Subsection 4.3, no assumption was made about the distribution of any Λ_i. In this subsection we will consider only one factor Λ. An arbitrary, but well-accepted choice for mathematical convenience, is the gamma distribution. Therefore, suppose Λ to be gamma-distributed (notation $\Lambda \sim \mathrm{Gamma}(\alpha,\beta)$) with shape parameter $\alpha > 0$ and rate (or inverse scale) parameter $\beta > 0$, i.e., Λ has a density

$$f(\lambda) := \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \, \mathrm{e}^{-\beta\lambda} & \text{for } \lambda > 0, \\ 0 & \text{for } \lambda \le 0, \end{cases} \tag{4.52}$$

where Γ denotes the gamma function given in (2.29). The integral substitution $x := \beta\lambda$ shows that f is indeed a probability density.

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

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

Lemma 4.35 (Summation property of the gamma distribution)**.** *Let $k \in \mathbb{N}$ and $\alpha_1, \dots, \alpha_k, \beta > 0$. If $\Lambda_1, \dots, \Lambda_k$ are independent random variables with $\Lambda_i \sim \mathrm{Gamma}(\alpha_i, \beta)$ for every $i \in \{1, \dots, k\}$, then*

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

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 (4.52) for $\Lambda_1 \sim \mathrm{Gamma}(\alpha_1, \beta)$ and $\Lambda_2 \sim \mathrm{Gamma}(\alpha_2, \beta)$, respectively. Due to the independence of Λ_1 and Λ_2, a density f for $\Lambda := \Lambda_1 + \Lambda_2$ is given by the convolution, i.e., for all $\lambda > 0$,

$$\begin{aligned} f(\lambda) &= \int_0^\lambda f_1(\mu) \, f_2(\lambda - \mu) \, \mathrm{d}\mu \\ &= \int_0^\lambda \frac{\beta^{\alpha_1}}{\Gamma(\alpha_1)} \mu^{\alpha_1 - 1} \, \mathrm{e}^{-\beta\mu} \, \frac{\beta^{\alpha_2}}{\Gamma(\alpha_2)} (\lambda - \mu)^{\alpha_2 - 1} \, \mathrm{e}^{-\beta(\lambda - \mu)} \, \mathrm{d}\mu. \end{aligned}$$

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

$$f(\lambda) = \underbrace{\frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \, \mathrm{e}^{-\beta\lambda}}_{\mathrm{Gamma}(\alpha,\beta)\text{-density}} \cdot \frac{\Gamma(\alpha)}{\Gamma(\alpha_1) \, \Gamma(\alpha_2)} \int_0^1 x^{\alpha_1 - 1} (1 - x)^{\alpha_2 - 1} \, \mathrm{d}x, \quad \lambda > 0,$$

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 2.5 and (2.34). $\square$

4.4.1 Moments of the Gamma Distribution

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

$$\begin{aligned}\mathbb{E}[\Lambda^\gamma \,\mathrm{e}^{\Lambda z}] &= \int_0^\infty \lambda^\gamma \,\mathrm{e}^{\lambda z}\, \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1}\,\mathrm{e}^{-\beta\lambda}\,\mathrm{d}\lambda \qquad \text{by (4.52)} \\ &= \frac{\Gamma(\alpha+\gamma)}{\Gamma(\alpha)} \frac{\beta^\alpha}{(\beta-z)^{\alpha+\gamma}} \int_0^\infty \underbrace{\frac{(\beta-z)^{\alpha+\gamma}}{\Gamma(\alpha+\gamma)} \lambda^{\alpha+\gamma-1}\,\mathrm{e}^{-(\beta-z)\lambda}}_{\text{Gamma}(\alpha+\gamma,\beta-z)\text{-density}}\,\mathrm{d}\lambda \\ &= \frac{\Gamma(\alpha+\gamma)}{\beta^\gamma\,\Gamma(\alpha)} (1-z/\beta)^{-(\alpha+\gamma)}.\end{aligned} \tag{4.53}$$

For $z = 0$, the calculation (4.53) gives all the moments

$$\mathbb{E}[\Lambda^\gamma] = \frac{\Gamma(\alpha+\gamma)}{\beta^\gamma\,\Gamma(\alpha)}, \qquad \gamma \in (-\alpha, \infty), \tag{4.54}$$

in particular, using the functional equation (2.30) for the gamma function,

$$\mathbb{E}[\Lambda] = \frac{\Gamma(\alpha+1)}{\beta\,\Gamma(\alpha)} = \frac{\alpha}{\beta}, \tag{4.55}$$

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

and

$$\mathrm{Var}(\Lambda) = \mathbb{E}[\Lambda^2] - (\mathbb{E}[\Lambda])^2 = \frac{\alpha}{\beta^2}. \tag{4.56}$$

For $\gamma = 0$, the calculation (4.53) gives the exponential moments and the moment-generating function

$$\mathbb{E}\big[\mathrm{e}^{\Lambda z}\big] = (1-z/\beta)^{-\alpha}, \qquad z \in (-\infty, \beta), \tag{4.57}$$

and the Laplace transform

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

Given $\gamma \in (-\alpha, \infty)$, let $\Lambda' \sim \text{Gamma}(\alpha+\gamma, \beta)$, where the shape parameter is shifted by γ. Then (4.53), (4.54) and (4.57) imply the peculiar relation

$$\mathbb{E}[\Lambda^\gamma \,\mathrm{e}^{\Lambda z}] = \mathbb{E}[\Lambda^\gamma](1-z/\beta)^{-(\alpha+\gamma)} = \mathbb{E}[\Lambda^\gamma]\,\mathbb{E}\big[\mathrm{e}^{\Lambda' z}\big], \qquad z \in (-\infty, \beta), \tag{4.58}$$

which we will use to derive (7.77) below.

4.4.2 Biased Gamma Distribution

The following lemma makes clear that the peculiar relation (4.58) is the consequence of a more general observation, which is very similar to Lemma 2.12 for the beta distribution and Lemma 4.29 for the Dirichlet distribution.

Lemma 4.36. *Assume that* $\Lambda \sim \mathrm{Gamma}(\alpha,\beta)$ *with parameters* $\alpha,\beta>0$ *and that* $\gamma\in(-\alpha,\infty)$ *and* $\delta\in(-\beta,\infty)$*. Then* $\mathbb{P}_{\Lambda^\gamma\,\mathrm{e}^{-\delta\Lambda}}\Lambda^{-1} = \mathrm{Gamma}(\alpha+\gamma,\beta+\delta)$*, that means the distribution of* Λ *under the* $\Lambda^\gamma\,\mathrm{e}^{-\delta\Lambda}$*-biased probability measure* $\mathbb{P}_{\Lambda^\gamma\,\mathrm{e}^{-\delta\Lambda}}$ *given by Definition 2.11 is the* $\mathrm{Gamma}(\alpha+\gamma,\beta+\delta)$ *distribution.*

Proof. By (2.44) and (4.53), a density of the $\Lambda^\gamma\,\mathrm{e}^{-\delta\Lambda}$-biased probability measure $\mathbb{P}_{\Lambda^\gamma\,\mathrm{e}^{-\delta\Lambda}}$ is given by

$$\frac{\mathrm{d}\mathbb{P}_{\Lambda^\gamma\,\mathrm{e}^{-\delta\Lambda}}}{\mathrm{d}\mathbb{P}} = \frac{\beta^\gamma\,\Gamma(\alpha)}{\Gamma(\alpha+\gamma)}\Big(1+\frac{\delta}{\beta}\Big)^{\alpha+\gamma}\Lambda^\gamma\,\mathrm{e}^{-\delta\Lambda} = \frac{(\beta+\delta)^{\alpha+\gamma}\,\Gamma(\alpha)}{\beta^\alpha\,\Gamma(\alpha+\gamma)}\Lambda^\gamma\,\mathrm{e}^{-\delta\Lambda}\,.$$

Let μ denote the Lebesgue–Borel measure on $\mathbb{R}$. Using the density f from (4.52) shows that, for μ-almost all $\lambda>0$,

$$\begin{aligned}\frac{\mathrm{d}(\mathbb{P}_{\Lambda^\gamma\,\mathrm{e}^{-\delta\Lambda}}\Lambda^{-1})}{\mathrm{d}\mu}(\lambda) &= \frac{\mathrm{d}(\mathbb{P}_{\Lambda^\gamma\,\mathrm{e}^{-\delta\Lambda}}\Lambda^{-1})}{\mathrm{d}(\mathbb{P}\Lambda^{-1})}(\lambda)\cdot\frac{\mathrm{d}(\mathbb{P}\Lambda^{-1})}{\mathrm{d}\mu}(\lambda)\\ &= \frac{(\beta+\delta)^{\alpha+\gamma}\,\Gamma(\alpha)}{\beta^\alpha\,\Gamma(\alpha+\gamma)}\lambda^\gamma\,\mathrm{e}^{-\delta\lambda}\cdot f(\lambda)\\ &= \frac{(\beta+\delta)^{\alpha+\gamma}}{\Gamma(\alpha+\gamma)}\lambda^{\alpha+\gamma-1}\,\mathrm{e}^{-(\beta+\delta)\lambda},\end{aligned}$$

which by (4.52) gives a density of the $\mathrm{Gamma}(\alpha+\gamma,\beta+\delta)$ distribution. □

4.5 Gamma-Mixed Poisson Distribution

To continue the investigation of Poisson mixture models, assume that $\Lambda\sim\mathrm{Gamma}(\alpha,\beta)$ with $\alpha,\beta>0$ and that the conditional distribution of N given Λ is $\mathrm{Poisson}(\Lambda)$, notation $\mathcal{L}(N|\Lambda)\overset{\text{a.s.}}{=}\mathrm{Poisson}(\Lambda)$, meaning that

$$\mathbb{P}[N=n\,|\,\Lambda] \overset{\text{a.s.}}{=} \frac{\Lambda^n}{n!}\,\mathrm{e}^{-\Lambda},\qquad n\in\mathbb{N}_0. \tag{4.59}$$

Combining (4.59) and (4.53) with $z=-1$, the unconditional distribution of N is

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

for all $n\in\mathbb{N}_0$. Using the abbreviations

$$p=\frac{1}{1+\beta}\in(0,1)\quad\text{and}\quad q=1-p=\frac{\beta}{1+\beta}, \tag{4.60}$$

and then the functional equation (2.31) of the gamma function, we get

$$\mathbb{P}[N=n] = \frac{\Gamma(\alpha+n)}{n!\,\Gamma(\alpha)}p^n\,q^\alpha = \binom{\alpha+n-1}{n}p^n\,q^\alpha,\qquad n\in\mathbb{N}_0, \tag{4.61}$$

which is called the negative binomial distribution.[33] We will use the notation $N\sim\mathrm{NegBin}(\alpha,p)$. We will interpret $\mathrm{NegBin}(0,p)$ with $p\in[0,1)$ and

[33] The term $\binom{\alpha+n-1}{n}p^n$ in (4.61) shows up when considering the binomial series for $(1-p)^{-\alpha}$, see (5.29) and (5.30), which is a negative power. This might motivate the name.

NegBin$(\alpha, 0)$ with $\alpha \geq 0$ as the degenerate distribution concentrated in 0. Note that the right-hand sides of (4.62), (4.63), (4.64) and (4.65) below, hence also (4.66) and (4.67) are correct for these cases.

If $\alpha \in \mathbb{N}$, then (4.61) gives the probability of exactly $n \in \mathbb{N}_0$ successes before the α-th failure in a sequence of independent Bernoulli trials with success probability p. For $\alpha = 1$, the negative binomial distribution (4.61) 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 $\mathcal{L}(N|\Lambda) \overset{\text{a.s.}}{=} \text{Poisson}(\Lambda)$ by assumption, we have

$$\mathbb{E}[N] = \mathbb{E}\big[\underbrace{\mathbb{E}[N|\Lambda]}_{\overset{\text{a.s.}}{=}\Lambda \text{ by (3.3)}}\big] = \mathbb{E}[\Lambda] \overset{(4.55)}{=} \frac{\alpha}{\beta} = \frac{\alpha p}{1-p} \tag{4.62}$$

by the substitution $\beta = \frac{1-p}{p}$ arising from (4.60). Using the law of total variance, i.e. (3.66) from Lemma 3.50, as well as (4.55) for the mean and (4.56) for the variance of Λ, we obtain

$$\begin{aligned}\operatorname{Var}(N) &= \mathbb{E}[\underbrace{\operatorname{Var}(N\,|\,\Lambda)}_{\overset{\text{a.s.}}{=}\Lambda \text{ by (3.4)}}] + \operatorname{Var}(\underbrace{\mathbb{E}[N\,|\,\Lambda]}_{\overset{\text{a.s.}}{=}\Lambda \text{ by (3.3)}}) \\ &= \mathbb{E}[\Lambda] + \operatorname{Var}(\Lambda) = \frac{\alpha}{\beta} + \frac{\alpha}{\beta^2} = \alpha\frac{\beta+1}{\beta^2} = \frac{\alpha p}{(1-p)^2},\end{aligned} \tag{4.63}$$

where we used (4.60) and $\beta = \frac{1-p}{p}$ for the last equation. It remains to calculate the corresponding probability-generating function. We present two different approaches. Using (4.61) and extending the fraction by $(1-ps)^\alpha$, it follows that

$$\begin{aligned}\varphi_N(s) &\overset{(4.1)}{=} \mathbb{E}\big[s^N\big] = \sum_{n=0}^{\infty} s^n\, \mathbb{P}[N=n] \\ &= \frac{q^\alpha}{(1-ps)^\alpha} \sum_{n=0}^{\infty} \underbrace{\binom{\alpha+n-1}{n}(ps)^n(1-ps)^\alpha}_{\text{NegBin}(\alpha,ps)\text{-distribution}} = \Big(\frac{q}{1-ps}\Big)^\alpha\end{aligned} \tag{4.64}$$

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

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

as well as the exponential moments (4.57) of $\Lambda \sim \text{Gamma}(\alpha, \beta)$,

$$\begin{aligned}\varphi_N(s) = \mathbb{E}\big[\mathbb{E}\big[s^N\,|\,\Lambda\big]\big] = \mathbb{E}\big[\mathrm{e}^{\Lambda(s-1)}\big] &= \Big(1 - \frac{s-1}{\beta}\Big)^{-\alpha} \\ &= \Big(\frac{\beta}{1+\beta-s}\Big)^\alpha \overset{(4.60)}{=} \Big(\frac{q}{1-ps}\Big)^\alpha\end{aligned} \tag{4.65}$$

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

$$\varphi_N^{(n)}(s) = \frac{p^n q^\alpha}{(1-ps)^{\alpha+n}} \prod_{l=0}^{n-1} (\alpha + l), \qquad n \in \mathbb{N}, \tag{4.66}$$

it follows via (4.25) for the factorial moments of the negative binomial distribution that

$$\mathbb{E}\Big[\prod_{l=0}^{n-1}(N-l)\Big] = \frac{p^n}{q^n} \prod_{l=0}^{n-1} (\alpha + l), \qquad n \in \mathbb{N}. \tag{4.67}$$

Here is the analogue of the Poisson and gamma summation properties given in Lemma 3.2 and Lemma 4.35, respectively, transferred to independent random variables with a negative binomial distribution (see Lemma 4.53 below for a multi-dimensional generalization):

Lemma 4.37 (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 $N_i \sim \mathrm{NegBin}(\alpha_i, p)$ for every $i \in \{1, \ldots, k\}$, then*

$$N \coloneqq \sum_{i=1}^{k} N_i \sim \mathrm{NegBin}(\alpha_1 + \cdots + \alpha_k, p)\,.$$

Proof. By independence, see (4.31), and generating function from (4.65),

$$\varphi_N(s) = \prod_{i=1}^{k} \varphi_{N_i}(s) = \prod_{i=1}^{k} \Big(\frac{q}{1-ps}\Big)^{\alpha_i} = \Big(\frac{q}{1-ps}\Big)^{\alpha_1+\cdots+\alpha_k} \tag{4.68}$$

for all $s \in \mathbb{C}$ satisfying $p|s| < 1$. Therefore, $N \sim \mathrm{NegBin}(\alpha, p)$ with $\alpha = \alpha_1 + \cdots + \alpha_k$, because the probability-generating function uniquely determines the distribution, see (4.16). □

Exercise 4.38. Give a more probabilistic derivation of Lemma 4.37 by considering the negative binomial distribution as a gamma-mixed Poisson distribution and using Lemma 3.2, Lemma 4.35, and the setup of the general multivariate Poisson mixture model, see (3.57) and (3.58).

4.6 Generating Function of Compound Distributions

To study random sums, let N be an $\mathbb{N}_0$-valued random variable and $(X_n)_{n\in\mathbb{N}}$ a sequence of $\mathbb{N}_0^d$-valued, independent, identically distributed random vectors, which is independent of N. In actuarial science, N describes the number of insurance claims during a given period and $(X_n)_{n\in\mathbb{N}}$ denote the claim sizes (only the first N are observed during the period) arising from a homogeneous portfolio of insurance contracts. The total claim amount is given by the $\mathbb{N}_0^d$-valued random sum

$$S \coloneqq \sum_{n=1}^{N} X_n. \tag{4.69}$$

This is called *collective risk model* for the total claim amount and used in ruin theory. Of course, this model can also be applied to credit risks. Using $Q \coloneqq \mathcal{L}(X_1)$, we introduce the notation

$$\mathrm{Compound}(\mathcal{L}(N), Q) \coloneqq \mathcal{L}(S). \tag{4.70}$$

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 $\mathbb{E}[\|X_1\|] < \infty$ using the elementary definition $\mathbb{E}[S \,|\, N = k] = \mathbb{E}[S\mathbb{1}_{\{N=k\}}]/\,\mathbb{P}[N = k]$ of the conditional expectation that

$$\mathbb{E}[S \,|\, N = k] = \mathbb{E}[X_1 + \cdots + X_k \,|\, N = k] = k\,\mathbb{E}[X_1]\,,$$

and in the case $\mathbb{E}\big[\|X_1\|^2\big] < \infty$ that

$$\begin{aligned}\mathrm{Cov}(S \,|\, N = k) &= \mathrm{Cov}(X_1 + \cdots + X_k \,|\, N = k)\\ &= \mathrm{Cov}(X_1 + \cdots + X_k)\\ &= k\,\mathrm{Cov}(X_1)\end{aligned}$$

by independence. These two results can be rewritten as

$$\mathbb{E}[S \,|\, N] \overset{\text{a.s.}}{=} N\,\mathbb{E}[X_1] \tag{4.71}$$

and

$$\mathrm{Cov}(S \,|\, N) \overset{\text{a.s.}}{=} N\,\mathrm{Cov}(X_1)\,, \tag{4.72}$$

where the last equation gives the conditional variances on the diagonal. Therefore, if N and X_1 are integrable, we get a special case of Wald's equation

$$\mathbb{E}[S] = \mathbb{E}[\mathbb{E}[S \,|\, N]] = \mathbb{E}[N]\,\mathbb{E}[X_1] \tag{4.73}$$

and, if they are square integrable, using Lemma 3.50,

$$\begin{aligned}\mathrm{Cov}(S) &= \mathbb{E}[\mathrm{Cov}(S \,|\, N)] + \mathrm{Cov}(\mathbb{E}[S \,|\, N])\\ &= \mathbb{E}[N]\,\mathrm{Cov}(X_1) + \mathrm{Var}(N)\,\mathbb{E}[X_1]\,\mathbb{E}[X_1]^{\mathsf{T}}\,,\end{aligned} \tag{4.74}$$

which is a special case of the Blackwell–Girshick equation.

We compute the probability-generating function φ_S. Using the multi-index notation as in Definition 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 $(X_n)_{n\in\mathbb{N}}$,

$$\begin{aligned}\varphi_S(s) \overset{(4.1)}{=} \mathbb{E}\big[s^{X_1+\cdots+X_N}\big] &= \sum_{n=0}^{\infty} \mathbb{E}\big[s^{X_1+\cdots+X_n}\mathbb{1}_{\{N=n\}}\big]\\ &= \sum_{n=0}^{\infty} \underbrace{\mathbb{E}\big[s^{X_1+\cdots+X_n}\big]}_{=\,(\mathbb{E}[s^{X_1}])^n\,=\,(\varphi_{X_1}(s))^n}\mathbb{P}[N = n]\\ &= \varphi_N(\varphi_{X_1}(s)),\end{aligned} \tag{4.75}$$

where is calculation is valid for all $s \in \mathbb{C}^d$ such that the power series defining $\varphi_{X_1}(s)$ is absolutely convergent and such that the power series defining φ_N converges at $|\varphi_{X_1}(s)|$. This is the case at least for all $s \in \mathbb{C}^d$ with $\|s\|_\infty \le 1$; note that $|\varphi_{X_1}(s)| \le 1$ for these s.

Example 4.39 (Pairwise independence is not enough for (4.75))**.** 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[34] for (4.75). 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 $\varphi_{X_1}(s) = \frac{1}{2}(1+s)$ 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, for all $i \in \{1, 2\}$ and $j \in \{0, 1\}$,

$$\mathbb{P}[N = 1, X_i = j] = \mathbb{P}[X_i = j, X_{3-i} = 1 - j] = \tfrac{1}{4}$$

as well as

$$\mathbb{P}[N = 2, X_i = j] = \mathbb{P}[X_1 = j, X_2 = j] = \tfrac{1}{4},$$

hence N and X_i are independent for every $i \in \{1, 2\}$. Note that $\varphi_N(s) = \frac{1}{2}s + \frac{1}{2}s^2$ and

$$\varphi_N(\varphi_{X_1}(s)) = \tfrac{1}{4}(1 + s) + \tfrac{1}{8}(1 + s)^2 = \tfrac{3}{8} + \tfrac{1}{2}s + \tfrac{1}{8}s^2, \qquad s \in \mathbb{C}. \tag{4.76}$$

However, for the compound sum S given by (4.69), 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

$$\varphi_S(s) = \tfrac{1}{2} + \tfrac{1}{4}s + \tfrac{1}{4}s^2, \qquad s \in \mathbb{C},$$

which differs from (4.76), hence (4.75) does not hold in this case.

Let $Q = (q_\nu)_{\nu\in\mathbb{N}_0^d}$ with $q_\nu := \mathbb{P}[X_1 = \nu]$ denote the distribution of X_1. If $N \sim \text{Poisson}(\lambda)$ with $\lambda \ge 0$, then the random sum S in (4.69) has a so-called compound Poisson distribution and we use the notation $S \sim \text{CPoisson}(\lambda, Q)$. Since $\varphi_N(s) = \mathrm{e}^{\lambda(s-1)}$ for all $s \in \mathbb{C}$ by (4.3), the calculation in (4.75) implies that

$$\varphi_S(s) = \exp\bigl(\lambda(\varphi_{X_1}(s) - 1)\bigr) \tag{4.77}$$

for all $s \in \mathbb{C}^d$ for which the power series defining $\varphi_{X_1}(s)$ converges, which is the case at least when $\|s\|_\infty \le 1$.

Similarly, if $N \sim \text{NegBin}(\alpha, p)$ with $\alpha \ge 0$ and $p \in [0, 1)$, then S from (4.69) 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 (4.65), the calculation in (4.75) implies that

$$\varphi_S(s) = \Bigl(\frac{q}{1 - p\varphi_{X_1}(s)}\Bigr)^\alpha \tag{4.78}$$

for all $s \in \mathbb{C}^d$ for which the power series defining $\varphi_{X_1}(s)$ is absolutely convergent and for which $p|\varphi_{X_1}(s)| < 1$, which is the case at least when $\|s\|_\infty \le 1$.

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

[34] A careful study of (4.75) shows that the independence of $X_1 + \cdots + X_n$ from $\{N = n\}$ for each $n \in \mathbb{N}$ is sufficient. Therefore, modifying Example 4.39 by defining $N = (X_1 + X_2) \mod 2$ makes N dependent on $X_1 + X_2$ but has no influence on the distribution of the random sum S.

Example 4.40 (Negative binomial distribution as compound Poisson distribution)**.** Let $(X_n)_{n\in\mathbb{N}}$ denote i.i.d. random variables, where $X_1 \sim \mathrm{Log}(p)$ has a univariate logarithmic distribution with parameter $p \in (0,1)$, see Example 4.4. Recall (4.6) to see that

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

According to (4.77), the compound Poisson sum S has the generating function

$$\varphi_S(s) = \exp\biggl(\underbrace{\lambda\Bigl(\frac{\log(1-ps)}{\log(1-p)} - 1\Bigr)}_{=\frac{\lambda}{\log(1-p)}\log\frac{1-ps}{1-p}}\biggr) = \Bigl(\frac{1-p}{1-ps}\Bigr)^{\alpha}, \qquad |s| < 1/p,$$

with

$$\alpha := -\frac{\lambda}{\log(1-p)} \geq 0, \tag{4.79}$$

which according to (4.65) is the probability-generating function of a negative binomial distribution, hence

$$\mathrm{CPoisson}(\lambda, \mathrm{Log}(p)) = \mathrm{NegBin}(\alpha, p)\,. \tag{4.80}$$

Remark 4.41 (Historical remark)**.** Note that the result of Example 4.40 can be traced back at least to H. Ammeter[35] [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 α and p to specify $\mathrm{NegBin}(\alpha,p)$, hence P is the expectation of the distribution, see (4.62). At the bottom of the page, he obtains the logarithmic distribution with parameter $\frac{\chi}{1+\chi}$ 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 (4.79).

Remark 4.42 (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 $N \sim \mathrm{Poisson}(\lambda)$ with $\lambda > 0$ and visualize N as the number of events of a homogeneous Poisson process of intensity λ (see [40, Section 2.1]) during $[0,1]$. To reflect the imprecise knowledge of the rate parameter λ, we can model it by a random factor $\Lambda \sim \mathrm{Gamma}(\alpha,\beta)$ with $\alpha,\beta > 0$ such that $\mathbb{E}[\Lambda] = 1$ and express the uncertainty by $\mathrm{Var}(\Lambda) = \sigma^2 > 0$. We assume that $\mathcal{L}(N|\Lambda) \overset{\text{a.s.}}{=} \mathrm{Poisson}(\lambda\Lambda)$, which implies that $\mathbb{E}[N] = \mathbb{E}[\mathbb{E}[N|\Lambda]] = \mathbb{E}[\lambda\Lambda] = \lambda$. Since $\mathbb{E}[\Lambda] = \alpha/\beta$ and $\mathrm{Var}(\Lambda) = \alpha/\beta^2$ by (4.55) and (4.56), this means $\alpha = \beta = 1/\sigma^2$. Then $\lambda\Lambda \sim \mathrm{Gamma}(\alpha, \beta/\lambda) = \mathrm{Gamma}(1/\sigma^2, 1/(\lambda\sigma^2))$, hence $N \sim \mathrm{NegBin}\bigl(1/\sigma^2, p\bigr)$ with

$$p \overset{(4.60)}{:=} \frac{1}{1+1/(\lambda\sigma^2)} = \frac{\lambda\sigma^2}{1+\lambda\sigma^2} \tag{4.81}$$

[35] Prof. Dr. h.c. Hans A. Ammeter (1912–1986), president of the *Schweizerische Lebensversicherungs- und Rentenanstalt* (now *Swiss Life*) from 1973 to 1978.

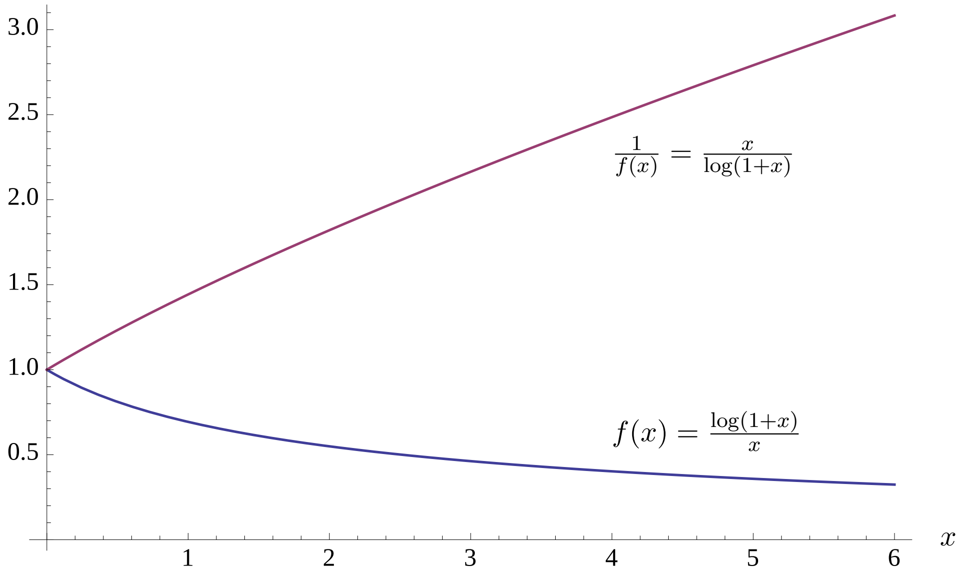

Figure 4.1: Illustration of the factor $f(x)$ from (4.83) reducing the Poisson intensity in (4.82) with increasing variance, and increasing the expectation of the number of defaults happening together, see (4.84).

as shown in Section 4.5, and we can visualize N as the number of events of a mixed Poisson process of random intensity $\lambda\Lambda$ during $[0, 1]$ (see [40, Section 2.3], it is also a special version of a Cox process).

Example 4.40 offers another interpretation of the distribution of N: We can consider a compound Poisson process with reduced intensity

$$\lambda' \overset{(4.79)}{:=} -\alpha \log(1-p) \overset{(4.81)}{=} -\frac{1}{\sigma^2} \log \frac{1}{1+\lambda\sigma^2} = \lambda f(\lambda\sigma^2), \tag{4.82}$$

where

$$f(x) := \frac{1}{x} \log(1+x), \qquad x > 0, \tag{4.83}$$

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

$$\mathbb{E}[X_i] = -\frac{p}{(1-p)\log(1-p)} \overset{(4.81)}{=} \frac{\lambda\sigma^2}{\log(1+\lambda\sigma^2)} \overset{(4.83)}{=} \frac{1}{f(\lambda\sigma^2)}, \qquad i \in \mathbb{N}, \tag{4.84}$$

see (4.29) with $n = 1$ and (4.5). By (4.80), this leads to the same distribution of the number of defaults during $[0, 1]$, namely $N \sim \mathrm{CPoisson}(\lambda', \mathrm{Log}(p))$.

As a corollary to the summation property of the Poisson distribution (Lemma 3.2) and the negative binomial distribution (Lemma 4.37), we get the corresponding property for the compound distributions.

Corollary 4.43 (Summation property for some compound distributions)**.** *Fix $k \in \mathbb{N}$. Let $Q, Q_1, \ldots, Q_k$ denote probability distributions on $\mathbb{N}_0^d$ and let $S_1, \ldots, S_k$ be independent $\mathbb{N}_0^d$-valued random vectors.*

(a) *Let* $\lambda_1, \dots, \lambda_k \ge 0$. *If* $S_i \sim \mathrm{CPoisson}(\lambda_i, Q_i)$ *for every* $i \in \{1, \dots, k\}$, *then*

$$S_1 + \cdots + S_k \sim \mathrm{CPoisson}(\lambda_1 + \cdots + \lambda_k, Q)\,,$$

if Q *satisfies*[36] $(\lambda_1 + \cdots + \lambda_k)Q = \lambda_1 Q_1 + \cdots + \lambda_k Q_k$.

(b) *Let* $\alpha_1, \dots, \alpha_k \ge 0$ *and* $p \in [0, 1)$. *If* $S_i \sim \mathrm{CNegBin}(\alpha_i, p, Q)$ *for every* $i \in \{1, \dots, k\}$, *then*[37]

$$S_1 + \cdots + S_k \sim \mathrm{CNegBin}(\alpha_1 + \cdots + \alpha_k, p, Q)\,.$$

Exercise 4.44. Prove Corollary 4.43. Hint: Use probability-generating functions, (4.32), (4.68), (4.75), (4.77) and (4.78).

Remark 4.45. The definitions and Corollary 4.43 can be extended to probability distributions $Q, Q_1, \dots, Q_k$ on $\mathbb{R}^d$. In this case the proof can be done using characteristic functions.

Lemma 4.46 (Representation of the multivariate Poisson distribution as compound Poisson distribution)**.** *Given* $\mathrm{MPoisson}(G, (\lambda_g)_{g\in G}, m)$ *as in Definition 3.42, define the total intensity by* $\lambda = \sum_{g\in G} \lambda_g$ *and let* μ *be a probability measure on* $\{0,1\}^m$ *satisfying* $\lambda\mu = \sum_{g\in G} \lambda_g \delta_{c_g}$, *where* δ_{c_g} *denotes the Dirac measure concentrated in* $c_g \in \{0,1\}^m$ *given by* (3.54). *Then* $\mathrm{MPoisson}(G, (\lambda_g)_{g\in G}, m) = \mathrm{CPoisson}(\lambda, \mu)$.

Proof. The probability-generating function of the Dirac measure δ_{c_g} is given in multi-index notation by $\varphi_{\delta_{c_g}}(s) = s^{c_g}$ for all $s \in \mathbb{C}^m$. When $\lambda > 0$, then $\mu = \mathrm{Convex}((\lambda_g/\lambda, \delta_{c_g})_{g\in G})$, hence by (4.22) in Example 4.9,

$$\lambda\varphi_\mu(s) = \sum_{g\in G} \lambda_g s^{c_g}, \qquad s \in \mathbb{C}^m.$$

Therefore, using (4.77), the probability-generating function φ of $\mathrm{CPoisson}(\lambda, \mu)$ is given by

$$\varphi(s) = \exp\bigl(\lambda(\varphi_\mu(s) - 1)\bigr) = \exp\biggl(\sum_{g\in G} \lambda_g (s^{c_g} - 1)\biggr), \qquad s \in \mathbb{C}^m,$$

which agrees with the probability-generating function (4.51) of the multivariate Poisson distribution $\mathrm{MPoisson}(G, (\lambda_g)_{g\in G}, m)$. □

[36] If $\lambda := \lambda_1 + \cdots + \lambda_k > 0$, then the degenerate case in excluded and Q is uniquely determined as the convex combination $\mathrm{Convex}((\lambda_i/\lambda, Q_i)_{i\in\{1,\dots,k\}})$ of $Q_1, \dots, Q_k$, see Example 4.9.

[37] For a generalization, see Corollary 4.61 below.

4.7 Some Compound Distributions Arising from the Multivariate Bernoulli Distribution

The purpose of this subsection is to introduce several multivariate discrete distributions on $\mathbb{N}_0^d$ and to discuss their characteristics. Throughout this subsection, let $(B_m)_{m\in\mathbb{N}}$ denote i.i.d. multivariate Bernoulli random vectors with $B_1 \sim \text{Multinomial}(1,\tilde{p}_1,\dots,\tilde{p}_d)$, where $\tilde{p}_1,\dots,\tilde{p}_d \in [0,1]$ with $\tilde{p}_1+\dots+\tilde{p}_d = 1$, see Example 4.5. Then $\varphi_{B_1}(s) = \sum_{i=1}^d \tilde{p}_i s_i$ for all $s=(s_1,\dots,s_d)\in\mathbb{C}^d$ by (4.8). Furthermore, let M be an $\mathbb{N}_0$-valued random variable, independent of $(B_m)_{m\in\mathbb{N}}$, and consider the random sum

$$N = (N_1,\dots,N_d) = \sum_{m=1}^{M} B_m. \tag{4.85}$$

Remark 4.47 (Covariance of components)**.** Suppose that $\mathrm{Var}(M) < \infty$. Using the representation from (4.85), the law of total covariance (Lemma 3.50) applied with $\mathcal{B} = \sigma(M)$, as well as (4.71), (4.72), (4.7) and (4.12),

$$\begin{aligned}
\mathrm{Cov}(N_i,N_j) &= \mathrm{Cov}\big(\underbrace{\mathbb{E}[N_i\,|\,M]}_{=\tilde{p}_iM},\underbrace{\mathbb{E}[N_j\,|\,M]}_{=\tilde{p}_jM}\big) + \mathbb{E}\big[\underbrace{\mathrm{Cov}(N_i,N_j\,|\,M)}_{=-\tilde{p}_i\tilde{p}_jM \text{ if } i\neq j}\big]\\
&= \tilde{p}_i\tilde{p}_j\big(\mathrm{Var}(M)-\mathbb{E}[M]\big), \qquad i,j\in\{1,\dots,d\} \text{ with } i\neq j.
\end{aligned}$$

Hence the sign of the covariance of two different 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 4.48 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 4.50(b) below.

Example 4.48. (Compound Poisson) Let $M\sim\text{Poisson}(\lambda)$ with $\lambda\ge 0$. Then (4.8) substituted into (4.77) implies for the random sum (4.85) that

$$\varphi_N(s) = \exp\Bigg(\lambda\underbrace{\Bigg(\sum_{i=1}^d \tilde{p}_i s_i - 1\Bigg)}_{=\sum_{i=1}^d \tilde{p}_i(s_i-1)}\Bigg) = \prod_{i=1}^d \exp\big(\lambda\tilde{p}_i(s_i-1)\big) \tag{4.86}$$

for all $s=(s_1,\dots,s_d)\in\mathbb{C}^d$, hence the components of N are independent and satisfy $N_i\sim\text{Poisson}(\lambda\tilde{p}_i)$ for every $i\in\{1,\dots,d\}$. This independence may come as a surprise, because different 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 4.7.1) or the negative binomial distribution (see Subsection 4.7.2) is considered for M.

If $\mathbb{P}[M=m]=1$ for an $m\in\mathbb{N}$, then $N\sim\text{Multinomial}(m,\tilde{p}_1,\dots,\tilde{p}_d)$ for the random variable in (4.85), see Example 4.19. More generally, given $(n_1,\dots,n_d)\in$

$\mathbb{N}_0^d$, 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

$$\begin{aligned}\mathbb{P}[N = (n_1, \ldots, n_d)] &= \mathbb{P}[M = m]\, \mathbb{P}[B_1 + \cdots + B_m = (n_1, \ldots, n_d) \,|\, M = m] \\ &= \mathbb{P}[M = m]\, \mathbb{P}[B_1 + \cdots + B_m = (n_1, \ldots, n_d)]\,.\end{aligned}$$

Since $B_1 + \cdots + B_m \sim \text{Multinomial}(m, \tilde{p}_1, \ldots, \tilde{p}_d)$, it follows from (4.35) that

$$\mathbb{P}[N = (n_1, \ldots, n_d)] = \mathbb{P}[M = m] \cdot m! \prod_{i=1}^{d} \frac{\tilde{p}_i^{n_i}}{n_i!} \tag{4.87}$$

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 5.26), the extended logarithmic distribution (see Example 5.27) and truncations of these distribution (see Definition 5.11).

4.7.1 Multivariate Logarithmic Distribution

Consider $M \sim \text{Log}(p)$ with $p \in (0, 1)$, see Example 4.4. It follows from (4.4) and (4.87) that, for every $(n_1, \ldots, n_d) \in \mathbb{N}_0^d \setminus \{(0, \ldots, 0)\}$,

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

with $m \coloneqq n_1 + \cdots + n_d$ and $p_i \coloneqq p\tilde{p}_i$ for $i \in \{1, \ldots, d\}$. This motivates the following definition:

Definition 4.49 (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 $\text{MLog}(p_1, \ldots, p_d)$ with parameters $p_1, \ldots, p_d \in [0, 1)$ satisfying $0 < p \coloneqq p_1 + \cdots + p_d < 1$, if

$$\mathbb{P}[N = (n_1, \ldots, n_d)] = \frac{(n_1 + \cdots + n_d - 1)!}{c(p)\, p} \prod_{i=1}^{d} \frac{p_i^{n_i}}{n_i!} \tag{4.88}$$

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

$$c(p) \coloneqq -\frac{\log(1-p)}{p}.$$

For $d = 1$, Definition 4.49 reduces to the univariate logarithmic distribution given in Example 4.4, which is well defined also for $p = 0$.

With

$$\varphi_M(s) = \frac{\log(1-ps)}{\log(1-p)}, \qquad |s| < 1/p,$$

given by (4.6) and $\varphi_{B_1}(s) = \sum_{i=1}^d \tilde{p}_i s_i$ for $s = (s_1, \dots, s_d) \in \mathbb{C}^d$ given by (4.8), it follows from (4.75) for the probability-generating function of N that

$$\varphi_N(s) = \frac{\log\bigl(1 - \sum_{i=1}^d p_i s_i\bigr)}{\log(1-p)} \tag{4.89}$$

for all $s = (s_1, \dots, s_d) \in \mathbb{C}^d$ with $\bigl|\sum_{i=1}^d p_i s_i\bigr| < 1$, which is certainly the case if $\|s\|_\infty < 1/p$.

Exercise 4.50 (Properties of the multivariate logarithmic distribution)**.** Assume that $N = (N_1, \dots, N_d)^\mathsf{T} \sim \mathrm{MLog}(p_1, \dots, p_d)$ with $p_1, \dots, p_d \in [0, 1)$ satisfying $0 < p \coloneqq p_1 + \dots + p_d < 1$, see Definition 4.49. Show:

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

$$\mathbb{E}\biggl[\prod_{i=1}^d \prod_{l_i=0}^{n_i-1} (N_i - l_i)\biggr] = -\frac{(n_1 + \dots + n_d - 1)!}{\log(1-p)} \prod_{i=1}^d \Bigl(\frac{p_i}{1-p}\Bigr)^{n_i},$$

which generalizes (4.29), hence for the components,

$$e_i \coloneqq \mathbb{E}[N_i] = -\frac{p_i}{(1-p)\log(1-p)}, \qquad i \in \{1, \dots, d\}.$$

(b) Covariance matrix: With expectation vector $e \coloneqq (e_1, \dots, e_d)^\mathsf{T}$,

$$\mathrm{Cov}(N) = \mathrm{diag}(e) - (1 + \log(1-p))\, e e^\mathsf{T},$$

which generalizes (4.30). When $p_1, \dots, p_d > 0$, conclude for all $i, j \in \{1, \dots, d\}$ with $i \ne j$ that $\mathrm{Cov}(N_i, N_j) \ge 0$ and $\mathrm{Var}(N_i) \ge \mathbb{E}[N_i]$ for all $p \ge 1 - \frac{1}{\mathrm{e}} \approx 0.6321$ and reversed inequalities otherwise.

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

$$(N_{\sigma(1)}, \dots, N_{\sigma(d)}) \sim \mathrm{MLog}(p_{\sigma(1)}, \dots, p_{\sigma(d)}).$$

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

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

(e) $N_1 + \dots + N_d \sim \mathrm{Log}(p)$.

Remark 4.51. Parts (a) and (b) of Exercise 4.50 can be solved using probability-generating functions, see (4.23), (4.24), (4.26), (4.27), (4.28) and (4.89), or they can be solved using the representation (4.85) together with the law of total covariance (Lemma 3.50) and results for the multinomial distribution and the univariate logarithmic distribution, see Exercises 4.20 and 4.13, respectively.

4.7.2 Negative Multinomial Distribution

Let $M \sim \mathrm{NegBin}(\alpha, p)$ with $\alpha > 0$ and $p \in [0,1)$, see (4.61). It follows from (4.61) and (4.87) that, for every $(n_1, \ldots, n_d) \in \mathbb{N}_0^d$,

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

with $m := n_1 + \cdots + n_d$ and $q := 1 - p$ as well as $p_i := p\tilde{p}_i$ for $i \in \{1, \ldots, d\}$. This motivates the following definition:

Definition 4.52 (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 $\mathrm{NegMult}(\alpha, p_1, \ldots, p_d)$ 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

$$\mathbb{P}[N = (n_1, \ldots, n_d)] = \frac{\Gamma(\alpha + n_1 + \cdots + n_d)}{\Gamma(\alpha)} q^\alpha \prod_{i=1}^d \frac{p_i^{n_i}}{n_i!} \tag{4.90}$$

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

For $d = 1$, Definition 4.52 reduces to the negative binomial distribution given by (4.61).

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 (2.31) of the gamma function, (4.90) can be rewritten with a multinomial coefficient, see (4.36), as

$$\mathbb{P}[N = (n_1, \ldots, n_d)] = \binom{\alpha - 1 + n_1 + \cdots + n_d}{n_1, \ldots, n_d} q^\alpha \prod_{i=1}^d p_i^{n_i} \tag{4.91}$$

for $(n_1, \ldots, n_d) \in \mathbb{N}_0^d$, and the product in (4.91) could be written using multi-index notation. In a sequence of independent trials, (4.91) gives the probability of $n_1, \ldots, n_d \in \mathbb{N}_0$ successes of types $1, \ldots, d$ before the αth failure happens.

With

$$\varphi_M(s) = \Big(\frac{q}{1 - ps}\Big)^\alpha, \qquad s \in \mathbb{C} \text{ with } p|s| < 1,$$

given by (4.64) and $\varphi_{B_1}(s) = \sum_{i=1}^d \tilde{p}_i s_i$ for $s = (s_1, \ldots, s_d) \in \mathbb{C}^d$ given by (4.8), it follows from (4.75) for the probability-generating function of N that

$$\varphi_N(s) = \biggl(\frac{q}{1 - \sum_{i=1}^d p_i s_i}\biggr)^\alpha \tag{4.92}$$

for all $s = (s_1, \ldots, s_d) \in \mathbb{C}^d$ with $\bigl|\sum_{i=1}^d p_i s_i\bigr| < 1$, which is certainly the case if $(p_1 + \cdots + p_d)\|s\|_\infty < 1$. Note that the calculation leading to (4.92) is correct for $p_1 = \cdots = p_d = 0$, and the result (4.92) is also correct for $\alpha = 0$.

Here is the multi-dimensional generalization of Lemma 4.37, which also implies that the negative multinomial distribution is infinitely divisible:

Lemma 4.53 (Summation property of the negative multinomial distribution)**.** *Let $k \in \mathbb{N}$, $\alpha_1, \dots, \alpha_k \ge 0$ and $p_1, \dots, p_d \in [0, 1)$ with $p_1 + \cdots + p_d < 1$. If $N_1, \dots, N_k$ are independent with $N_i \sim \operatorname{NegMult}(\alpha_i, p_1, \dots, p_d)$ for every $i \in \{1, \dots, k\}$, then*

$$N := \sum_{i=1}^{k} N_i \sim \operatorname{NegMult}(\alpha_1 + \cdots + \alpha_k, p_1, \dots, p_d). \tag{4.93}$$

Exercise 4.54. Prove Lemma 4.53.

Exercise 4.55 (Properties of the negative multinomial distribution)**.** Assume that $N = (N_1, \dots, N_d) \sim \operatorname{NegMult}(\alpha, p_1, \dots, p_d)$ with $\alpha \ge 0$ and $p_1, \dots, p_d \in [0, 1)$ satisfying $q := 1 - (p_1 + \cdots + p_d) \in (0, 1]$, see Definition 4.52. Show:

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

$$\mathbb{E}\Biggl[\prod_{i=1}^{d} \prod_{l_i=0}^{n_i-1} (N_i - l_i)\Biggr] = \Biggl(\prod_{l=0}^{n_1+\cdots+n_d-1} (\alpha + l)\Biggr) \prod_{i=1}^{d} \Bigl(\frac{p_i}{q}\Bigr)^{n_i},$$

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

$$\operatorname{Var}(N_i) = \frac{\alpha p_i (p_i + q)}{q^2}.$$

In the case $d = 1$, these results coincide with (4.67) and (4.63), respectively.

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

$$\operatorname{Cov}(N_i, N_j) = \alpha \frac{p_i p_j}{q^2}.$$

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

$$(N_{\sigma(1)}, \dots, N_{\sigma(d)}) \sim \operatorname{NegMult}(\alpha, p_{\sigma(1)}, \dots, p_{\sigma(d)}).$$

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

$$(N_1, \dots, N_i, N_{i+1} + \cdots + N_d) \sim \operatorname{NegMult}(\alpha, p_1, \dots, p_i, p_{i+1} + \cdots + p_d).$$

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

(f) Marginal distributions: For every $i \in \{1, \dots, d\}$,

$$(N_1, \dots, N_i) \sim \operatorname{NegMult}\Bigl(\alpha, \frac{p_1}{1 - p_{i+1} - \cdots - p_d}, \dots, \frac{p_i}{1 - p_{i+1} - \cdots - p_d}\Bigr),$$

in particular $N_i \sim \operatorname{NegBin}\bigl(\alpha, \frac{p_i}{p_i + q}\bigr)$.

4.7.3 Multivariate Binomial Distribution

Let $M \sim \mathrm{Bin}(m, p)$ with $m \in \mathbb{N}_0$ and $p \in [0, 1]$. It follows from (2.9) and (4.87) that, for every $(n_1, \ldots, n_d) \in \mathbb{N}_0^d$ with $l := n_1 + \cdots + n_d \le m$,

$$\begin{aligned} \mathbb{P}[N = (n_1, \ldots, n_d)] &= \binom{m}{l} p^l (1-p)^{m-l} \cdot l! \prod_{i=1}^{d} \frac{\tilde{p}_i^{n_i}}{n_i!} \\ &= \frac{m!}{(m-l)!} (1-p)^{m-l} \prod_{i=1}^{d} \frac{p_i^{n_i}}{n_i!} \end{aligned} \tag{4.94}$$

with $p_i := p\tilde{p}_i$ for $i \in \{1, \ldots, d\}$. This can be called multivariate binomial distribution $\mathrm{MBin}(m, p_1, \ldots, p_d)$ with $m \in \mathbb{N}_0$ independent trials and success probabilities $p_1, \ldots, p_d \in [0, 1]$ satisfying $p_1 + \cdots + p_d \le 1$. For $d = 1$, this coincides with the binomial distribution, compare (2.9) with (4.94). If $p = 1$, hence $p_1 + \cdots + p_d = 1$, then $\mathrm{MBin}(m, p_1, \ldots, p_d) = \mathrm{Multinomial}(m, p_1, \ldots, p_d)$.

With $\varphi_M(s) = (1 + p(s-1))^m$ for $s \in \mathbb{C}$ as in (4.33) and $\varphi_{B_1}(s) = \sum_{i=1}^{d} \tilde{p}_i s_i$ for $s = (s_1, \ldots, s_d) \in \mathbb{C}^d$ given by (4.8), it follows from (4.75) for the probability-generating function of N that

$$\varphi_N(s) = \left(1 + \sum_{i=1}^{d} p_i(s_i - 1)\right)^m, \qquad s = (s_1, \ldots, s_d) \in \mathbb{C}^d, \tag{4.95}$$

which is a generalization of (4.34).

Remark 4.56 (Relation to multinomial distribution)**.** While the multivariate binomial distribution generalizes the multinomial distribution (see Example 4.19), it is not a new distribution but already contained in the multinomial distribution of higher dimension by looking at marginals: More precisely, if $(N_1, \ldots, N_d) \sim \mathrm{MBin}(m, p_1, \ldots, p_d)$ with $m \in \mathbb{N}_0$ and $p_1, \ldots, p_d \in [0, 1]$ satisfying $p_1 + \cdots + p_d \le 1$, then it follows from (4.35) and (4.94) that

$$\begin{aligned} &\big(N_1, \ldots, N_d, m - (N_1 + \cdots + N_d)\big) \\ &\qquad \sim \mathrm{Multinomial}\big(m, p_1, \ldots, p_d, 1 - (p_1 + \cdots + p_d)\big). \end{aligned} \tag{4.96}$$

The other way round, if $(N_1, \ldots, N_d) \sim \mathrm{Multinomial}(m, p_1, \ldots, p_d)$ 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 4.20(c) and (4.96),

$$(N_1, \ldots, N_i) \sim \mathrm{MBin}(m, p_1, \ldots, p_i) \tag{4.97}$$

for every $i \in \{1, \ldots, d\}$. Of course, (4.96) and (4.97) can also be proved by applying (4.18) to the probability generating functions (4.34) and (4.92).

Due to Remark 4.56, the multivariate binomial distribution inherits many properties of the multinomial distribution given in Exercise 4.20.

Exercise 4.57 (Properties of the multivariate binomial distribution)**.** Let $N = (N_1, \dots, N_d) \sim \mathrm{MBin}(m, p_1, \dots, p_d)$ with parameters $m \in \mathbb{N}_0$ and $p_1, \dots, p_d \in [0,1]$ satisfying $p_1 + \cdots + p_d \le 1$. Show the following:

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

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

$$(N_1, \dots, N_i, N_{i+1} + \cdots + N_d) \sim \mathrm{MBin}(m, p_1, \dots, p_i, p_{i+1} + \cdots + p_d).$$

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

$$(N_1, \dots, N_i) \sim \mathrm{MBin}(m, p_1, \dots, p_i).$$

(d) Permutation property: For every permutation σ of $\{1, \dots, d\}$,

$$(N_{\sigma(1)}, \dots, N_{\sigma(d)}) \sim \mathrm{MBin}(m, p_{\sigma(1)}, \dots, p_{\sigma(d)}).$$

(e) Expectations and variances: $\mathbb{E}[N_i] = mp_i$ and $\mathrm{Var}(N_i) = mp_i(1 - p_i)$ for every $i \in \{1, \dots, d\}$.

(f) Covariances: $\mathrm{Cov}(N_i, N_j) = -mp_ip_j$ for all $i, j \in \{1, \dots, d\}$ with $i \ne j$.

Lemma 4.58 (Summation property of the multivariate binomial distribution)**.** *Let $k \in \mathbb{N}$, $m_1, \dots, m_k \in \mathbb{N}_0$ and $p_1, \dots, p_d \in [0,1]$ with $p_1 + \cdots + p_d \le 1$. If $N_1, \dots, N_k$ are independent with $N_i \sim \mathrm{MBin}(m_i, p_1, \dots, p_d)$ for every $i \in \{1, \dots, k\}$, then*

$$N := \sum_{i=1}^{k} N_i \sim \mathrm{MBin}(m_1 + \cdots + m_k, p_1, \dots, p_d). \tag{4.98}$$

Exercise 4.59. Prove Lemma 4.58 (using (4.92) or Lemma 4.21 and (4.96)).

4.8 Conditional Compound Distributions

In the next step we look at the case, where N is conditionally Poisson-distributed, namely $\mathcal{L}(N|\Lambda) \overset{\text{a.s.}}{=} \mathrm{Poisson}(\Lambda)$ for a non-negative random variable Λ. To compute the generating function of the random sum S given in (4.69), conditioned on Λ, first note that

$$\varphi_{N|\Lambda}(s) := \mathbb{E}\big[s^N \,\big|\, \Lambda\big] \overset{\text{a.s.}}{=} \exp(\Lambda(s-1)), \qquad s \in \mathbb{C}, \tag{4.99}$$

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 (Λ, N). Then, for every $n \in \mathbb{N}_0$, using multi-index notation and the multiplication theorem for probability-generating functions,

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

hence by summation over $n \in \mathbb{N}_0$,

$$\mathbb{E}\big[s^{X_1+\cdots+X_N}\,\big|\,\Lambda, N\big] \overset{\text{a.s.}}{=} \big(\varphi_{X_1}(s)\big)^N \tag{4.100}$$

for all $s \in \mathbb{C}^d$ for which the power series defining $\varphi_{X_1}(s)$ converges,[38] which is the case at least for all $s \in \mathbb{C}^d$ with $\|s\|_\infty \le 1$. Hence for these $s \in \mathbb{C}^d$, by using the tower property of conditional expectation,

$$\begin{aligned}\varphi_{S|\Lambda}(s) &:= \mathbb{E}\big[s^{X_1+\cdots+X_N}\,\big|\,\Lambda\big]\\ &\overset{\text{a.s.}}{=} \mathbb{E}\big[\mathbb{E}\big[s^{X_1+\cdots+X_N}\,|\,\Lambda, N\big]\,\big|\,\Lambda\big]\\ &\overset{\text{a.s.}}{=} \mathbb{E}\big[\big(\varphi_{X_1}(s)\big)^N\,\big|\,\Lambda\big] \qquad \text{by (4.100)}\\ &\overset{\text{a.s.}}{=} \varphi_{N|\Lambda}(\varphi_{X_1}(s))\\ &\overset{\text{a.s.}}{=} \exp\big(\Lambda(\varphi_{X_1}(s)-1)\big) \qquad \text{by (4.99)},\end{aligned}$$

and therefore

$$\varphi_S(s) = \mathbb{E}[\varphi_{S|\Lambda}(s)] = \mathbb{E}\big[\exp\big(\Lambda(\varphi_{X_1}(s)-1)\big)\big] \tag{4.101}$$

at least for all $s \in \mathbb{C}^d$ with $\|s\|_\infty \le 1$, which generalizes (4.77).

If $\Lambda \sim \text{Gamma}(\alpha,\beta)$ with $\alpha,\beta > 0$, then $N \sim \text{NegBin}(\alpha,p)$ with $p = \frac{1}{1+\beta}$ by (4.60) and (4.61), hence $S \sim \text{CNegBin}(\alpha,p,Q)$, where Q denotes the distribution of $\mathbb{N}_0^d$-valued X_1, and the probability-generating function of S is given by (4.78). Evaluating the right-hand side of (4.101) using the exponential moment of Λ given by (4.57) and $\beta = \frac{1-p}{p}$ leads to

$$\varphi_S(s) = \left(1 - \frac{\varphi_{X_1}(s)-1}{\beta}\right)^{-\alpha} = \left(\frac{1-p}{1-p\varphi_{X_1}(s)}\right)^{\alpha} \tag{4.102}$$

at least for all $s \in \mathbb{C}^d$ with $\|s\|_\infty \le 1$, which agrees with (4.78).

Exercise 4.60 (Generalization of (4.80) to compound distributions)**.** Let the array $(X_{m,n})_{m,n\in\mathbb{N}}$ consist of $\mathbb{N}_0^d$-valued i.i.d. random vectors and define $Q = \mathcal{L}(X_{1,1})$. Given $\alpha,\beta,\lambda > 0$, let $\Lambda \sim \text{Gamma}(\alpha,\beta)$ and $\mathcal{L}(N|\Lambda) \overset{\text{a.s.}}{=} \text{Poisson}(\lambda\Lambda)$. Assume that $(X_{m,n})_{m,n\in\mathbb{N}}$ and (Λ, N) are independent. Define $p = \lambda/(\beta+\lambda) \in (0,1)$ and $\mu = -\alpha\log(1-p) > 0$. 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 that 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

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

[38] For these s the random variable $s^{X_1+\cdots+X_N}$ is σ-integrable, and the corresponding generalization of conditional expectations should be used.

that they have the same distribution, which is then a compound negative binomial as well as a compound Poisson distribution, cf. (4.80), i.e.

$$\mathrm{CNegBin}(\alpha, p, Q) = \mathrm{CPoisson}(\mu, \mathrm{CLog}(p, Q)) . \tag{4.103}$$

(c) Assume that $\mathbb{E}[\Lambda] = 1$ and $\mathrm{Var}(\Lambda) = \sigma^2 > 0$. Show that $\alpha = \beta = 1/\sigma^2$ and conclude that

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

and

$$\mu = -\frac{1}{\sigma^2} \log\Bigl(1 - \frac{\lambda\sigma^2}{1 + \lambda\sigma^2}\Bigr) = \frac{\lambda}{1 + \lambda\sigma^2} \biggl(1 + \sum_{n=2}^{\infty} \frac{1}{n} \Bigl(\frac{\lambda\sigma^2}{1 + \lambda\sigma^2}\Bigr)^{n-1}\biggr),$$

where the right-hand side is numerically stable for small σ^2. Show that $p \searrow 0$ and $\mu \to \lambda$ as $\sigma^2 \searrow 0$ and, using Definition 4.4, that the right-hand side of (4.103) equals $\mathrm{CPoisson}(\lambda, Q)$ for the limiting values.

Corollary 4.61 (General summation property for compound negative binomial distributions, generalization of Corollary 4.43(b))**.** *With $k \in \mathbb{N}$, consider for each $i \in \{1, \dots, k\}$ a parameter $\alpha_i > 0$, a probability $p_i \in (0, 1)$, a probability distribution Q_i on $\mathbb{N}_0^d$, and a random vector $S_i \sim \mathrm{CNegBin}(\alpha_i, p_i, Q_i)$. Define $\mu_i = -\alpha_i \log(1 - p_i)$ for every $i \in \{1, \dots, k\}$ and $\mu = \mu_1 + \cdots + \mu_k$. If $S_1, \dots, S_k$ are independent, then*

$$S_1 + \cdots + S_k \sim \mathrm{CPoisson}(\mu, Q) .$$

where

$$Q \coloneqq \mathrm{Convex}((\mu_i/\mu, \mathrm{CLog}(p_i, Q_i))_{i \in \{1,\dots,k\}}).$$

Exercise 4.62. Prove Corollary 4.61. Hint: Combine (4.103) and Corollary 4.43(a).

4.8.1 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 $X_n = (X_{n,1}, \dots, X_{n,d})$, which is independent of N. We want to calculate the expectations, variances and covariances of the components $(S_1, \dots, S_d)$ of the random sum $S \coloneqq X_1 + \cdots + X_N$ considered in (4.69).

We now specialize to the case where N is conditionally Poisson-distributed, namely $\mathcal{L}(N|\Lambda) \overset{\text{a.s.}}{=} \mathrm{Poisson}(\Lambda)$ for a non-negative random variable Λ. Then the random sum S is conditionally compound Poisson given Λ, hence, if $\mathbb{E}[\|X_1\|] < \infty$, then (4.71) is turned into

$$\mathbb{E}[S\,|\,\Lambda, N] \overset{\text{a.s.}}{=} N\,\mathbb{E}[X_1] ,$$

and taking conditional expectations using the tower property replaces (4.73) by

$$\mathbb{E}[S\,|\,\Lambda] \overset{\text{a.s.}}{=} \underbrace{\mathbb{E}[N\,|\,\Lambda]}_{\overset{\text{a.s.}}{=}\Lambda \text{ by } (3.3)} \mathbb{E}[X_1] \overset{\text{a.s.}}{=} \Lambda\,\mathbb{E}[X_1] \tag{4.104}$$

and, if $\mathbb{E}\big[\|X_1\|^2\big] < \infty$, then (4.72) is turned into

$$\operatorname{Cov}(S\,|\,\Lambda, N) \overset{\text{a.s.}}{=} N \operatorname{Cov}(X_1)\,,$$

and (4.74) turns into

$$\begin{aligned}\operatorname{Cov}(S\,|\,\Lambda) &\overset{\text{a.s.}}{=} \underbrace{\mathbb{E}[N\,|\,\Lambda]}_{\overset{\text{a.s.}}{=}\Lambda \text{ by } (3.3)} \operatorname{Cov}(X_1) + \underbrace{\operatorname{Var}(N\,|\,\Lambda)}_{\overset{\text{a.s.}}{=}\Lambda \text{ by } (3.4)} \mathbb{E}[X_1]\,\mathbb{E}[X_1]^{\mathsf{T}} \\ &\overset{\text{a.s.}}{=} \Lambda\,\mathbb{E}\big[X_1 X_1^{\mathsf{T}}\big].\end{aligned} \tag{4.105}$$

If N is unconditionally Poisson distributed, i.e., $\mathcal{L}(N) = \text{Poisson}(\lambda)$, then (4.73) and (4.74) simplify to

$$\mathbb{E}[S] = \lambda\,\mathbb{E}[X_1] \tag{4.106}$$

and

$$\operatorname{Cov}(S) = \lambda\,\mathbb{E}\big[X_1 X_1^{\mathsf{T}}\big]. \tag{4.107}$$

5 Recursive Algorithms and Weighted Convolutions

For $j, n \in \mathbb{N}_0^d$ we write $j \le n$ if this is true for all d components, and we write $j < n$ if $j \le n$ and $j \ne n$, meaning that there is strict inequality for at least one component. Note that $\le$ is then a partial order on $\mathbb{N}_0^d$. We write $\langle \cdot, \cdot \rangle$ for the standard inner product in $\mathbb{R}^d$.

5.1 Convolutions

Remark 5.1 (Convolution)**.** Let X and Y be two independent $\mathbb{N}_0^d$-valued random vectors, let $P = (p_n)_{n\in\mathbb{N}_0^d}$ and $Q = (q_n)_{n\in\mathbb{N}_0^d}$ denote their distributions, and φ_X as well as φ_Y their probability-generating functions given via (4.1), respectively. Then $P * Q := \mathcal{L}(X+Y)$ denotes the distribution of their sum and can be computed in a numerically stable way by

$$\mathbb{P}[X+Y=n] = \sum_{j\in\mathbb{N}_0^d} \underbrace{\mathbb{P}[X=n-j,\,Y=j]}_{\substack{=\mathbb{P}[X=n-j]\,\mathbb{P}[Y=j]\\ \text{by independence}}} = \sum_{\substack{j\in\mathbb{N}_0^d\\ j\le n}} p_{n-j} q_j, \quad n \in \mathbb{N}_0^d, \tag{5.1}$$

where $j \le n$ is understood componentwise. Comparison with (4.1) shows that (5.1) is a way to calculate the coefficients of the power series φ_{X+Y}.

Recall that $\varphi_{X+Y}(s) = \varphi_X(s)\varphi_Y(s)$ at least for all $s \in \mathbb{C}^d$ with $\|s\|_\infty \le 1$ by Theorem 4.16, hence an application of the general Leibniz rule for every one of the d dimensions gives in multi-index notation

$$\varphi_{X+Y}^{(n)}(s) = \sum_{\substack{j\in\mathbb{N}_0^d\\ j\le n}} \varphi_X^{(n-j)}(s)\,\varphi_Y^{(j)}(s)\prod_{i=1}^d \binom{n_i}{j_i}, \quad n = (n_1,\dots,n_d) \in \mathbb{N}_0^d,$$

which via (4.16) also leads to (5.1).

If X and Y have the same distribution Q, then by commutativity of multiplication, the number of terms on the right-hand side of (5.1) can be cut in half (only approximately if the scaled vector $\frac{1}{2}n$ is in $\mathbb{N}_0^d$) and we can replace (5.1) by

$$\mathbb{P}[X+Y=n] = \mathbb{1}_{\mathbb{N}_0^d}(n/2)q_{n/2}^2 + 2\sum_{\substack{j\in\mathbb{N}_0^d, j\le n\\ j<_{\mathrm{t}} n-j}} q_{n-j}q_j, \quad n \in \mathbb{N}_0^d, \tag{5.2}$$

where $j \le n$ is understood componentwise and $<_{\mathrm{t}}$ denotes a total order of $\mathbb{N}_0^d$, for example the lexicographic order.

The binary operation $*$ for distributions is called convolution. Since addition of $\mathbb{N}_0^d$-valued random vectors is commutative and associative, the same is true for the convolution operation. Note that the Dirac measure δ_0 concentrated in the origin of $\mathbb{N}_0^d$ is the neutral element w.r.t. convolution. Given $\mathbb{N}_0^d$-valued i.i.d. random vectors $X_1,\dots,X_k$ with distribution Q, the distribution of their sum $S_k \coloneqq X_1+\cdots+X_k$ is denoted by the convolution power Q^{*k} for each $k \in \mathbb{N}_0$. Note that $Q^{*1} = Q$ and $Q^{*0} = \delta_0$ (by the convention for the empty sum).

Algorithm 5.2 (Calculation of convolution powers)**.** Fix $k \in \mathbb{N}$ with $k \ge 2$ and $\mathbb{N}_0^d$-valued i.i.d. random vectors $X_1,\dots,X_k$ with distribution Q. The convolution power $Q^{*k} \coloneqq \mathcal{L}(S_k)$ for $S_k \coloneqq X_1+\cdots+X_k$ can be calculated in two ways:

(a) The recursive and numerically stable calculation of

$$Q^{*(i+1)} = Q^{*i} * Q, \quad i \in \{1,\dots,k-1\}, \tag{5.3}$$

via (5.2) if $i=1$ and (5.1) otherwise, requiring $k-1$ convolutions.

(b) Starting with $k \ge 4$, there is a method to calculate Q^{*k} with a fewer number of convolutions, similar to the exponentiation by squaring or the Russian peasant multiplication. Define $l = \lfloor \log_2 k \rfloor$ and represent k in the binary form $k = \sum_{i=0}^l b_i 2^i$ with $b_l = 1$ and $b_0,\dots,b_{l-1} \in \{0,1\}$. Calculate iteratively via (5.2) the convolution powers

$$Q^{*2^{i+1}} = Q^{*2^i} * Q^{*2^i}, \quad i \in \{0,\dots,l-1\}, \tag{5.4}$$

which requires l convolutions. If $k = 2^l$, then we are done, otherwise Q^{*k} is obtained by using (5.1) to calculate the convolution

$$Q^{*k} = Q^{2^l b_l} * \cdots * Q^{*2^2 b_2} * Q^{*2^1 b_1} * Q^{*2^0 b_0}, \tag{5.5}$$

where $Q^{*2^i b_i} = Q^{*0} = \delta_0$ is the convolution-neutral element when $b_i = 0$ for $i \in \{0, \dots, l-1\}$. This requires $b_0 + \dots + b_{l-1}$ additional convolutions, so there are $l + b_0 + \dots + b_{l-1} \le 2l$ altogether. This algorithm is numerically slightly more precise than the $k-1$ recursive convolutions for (5.3), because a smaller number of operations for the calculation of Q^{*k} and, therefore, a smaller number of rounding errors to machine precision are needed. Furthermore, it can be substantially faster for large k (see however Example 5.3 below), and work on (5.5) can start in parallel as soon as the first results needed from (5.4) are available.

Example 5.3 (Comparison of the two algorithms for computing Q^{*k})**.** Due to an effect discussed also in Remark 5.19 below, the speed-up by Algorithm 5.2(b) compared to version (a) might not be as large as the logarithmically smaller number of convolutions suggests. On the contrary, the seemingly faster algorithm can be much slower when counting the total number of multiplications yielding a non-zero product!

In general, it doesn't seem to be easy to predict and count the total number of non-zero terms appearing on the right-hand side of (5.1) or (5.2), respectively, when calculating the convolutions in (5.3), (5.4) and (5.5). Therefore, we will focus on a special case, for which some notation is helpful. Fix $d \in \mathbb{N}$. Define the standard discrete d-dimensional simplex in $\mathbb{N}_0^d$ of size $\nu \in \mathbb{N}_0$ by

$$\Delta_{d,\nu} = \{ n \in \mathbb{N}_0^d \mid \|n\|_1 \le \nu \} \tag{5.6}$$

Note that each $n = (n_1, \dots, n_d) \in \Delta_{d,\nu}$ is via $\tilde{n} := (n_1, \dots, n_d, \nu - (n_1 + \dots + n_d))$ in a one-to-one correspondence to an element of $\{\tilde{n} \in \mathbb{N}_0^{d+1} \mid \|\tilde{n}\|_1 = \nu\}$, hence[39]

$$|\Delta_{d,\nu}| = \binom{d+\nu}{d}, \qquad \nu \in \mathbb{N}_0. \tag{5.7}$$

To illustrate the above claim concerning Algorithms 5.2(a) and (b), fix $k, \nu \in \mathbb{N}$ with $k \ge 2$ and suppose that X_1 takes only values in $\Delta_{d,\nu}$, each one with strictly positive probability. Then $\mathbb{P}[S_i = n] > 0$ if and only if $n \in \Delta_{d,i\nu}$ for each $i \in \{1, \dots, k\}$. Hence when calculating the distribution $Q^{*(i+1)}$ given by (5.3) for $i \in \{1, \dots, k-1\}$, i.e. $\mathbb{P}[S_{i+1} = n]$ for all $n \in \Delta_{d,(i+1)\nu}$, then there are in total exactly $|\Delta_{d,i\nu}| \cdot |\Delta_{d,\nu}|$ pairs appearing on the right-hand side of (5.1) giving a non-zero product. In the case $i = 1$, (5.2) reduces the number of non-zero products to $\frac{1}{2}|\Delta_{d,\nu}|(|\Delta_{d,\nu}| + 1)$. Hence when calculating Q^{*k} using Algorithm 5.2(a), there are in total exactly

$$N_{\mathrm{a}}(d, k, \nu) := |\Delta_{d,\nu}| \biggl(\frac{|\Delta_{d,\nu}| + 1}{2} + \sum_{i=2}^{k-1} |\Delta_{d,i\nu}| \biggr) \tag{5.8}$$

non-zero summands (and products).

[39] Line up $d + \nu$ equal objects and choose d of them to mark the boundaries between $d + 1$ boxes containing the other ν objects. The number of choices is given by the binomial coefficient.

	$\lvert\Delta_{d,\nu}\rvert$	$k=3$	$k=7$	$k=15$	$k=31$	$k=63$
$N_{\mathrm{a}}(1,k,2)$	3	21	141	669	2 877	11 901
$N_{\mathrm{b}}(1,k,2)$	3	21	99	399	1 575	6 231
$N_{\mathrm{r}}(1,k,2)$	3	12	28	60	124	252
$N_{\mathrm{a}}(2,k,1)$	3	24	246	2 034	16 362	131 034
$N_{\mathrm{b}}(2,k,1)$	3	24	195	1 935	23 778	331 767
$N_{\mathrm{r}}(2,k,1)$	3	13	57	241	993	4 033
$N_{\mathrm{a}}(3,k,1)$	4	50	830	12 230	185 494	2 882 870
$N_{\mathrm{b}}(3,k,1)$	4	50	805	21 235	825 634	40 460 879
$N_{\mathrm{r}}(3,k,1)$	4	31	253	2 041	16 369	131 041
upper est.		≤ 58	≤ 358	$\le 2\,446$	$\le 17\,950$	$\le 137\,278$
$N_{\mathrm{a}}(3,k,2)$	10	405	10 305	191 545	3 278 505	54 183 305
$N_{\mathrm{b}}(3,k,2)$	10	405	14 895	687 510	36 866 995	2 150 779 180
$N_{\mathrm{r}}(3,k,2)$	10	≤ 748	$\le 6\,112$	$\le 49\,096$	$\le 393\,112$	$\le 3\,145\,528$

Table 5.1: Comparison of the total number of non-zero terms to be calculated for the right-hand side of (5.1) or (5.2), respectively, when calculating the convolution power Q^{*k}, provided the support of Q agrees with the discrete d-dimensional simplex $\Delta_{d,\nu}$ of size ν in $\mathbb{N}_0^d$ given by (5.6). We concentrate on $k = 2^{l+1} - 1$ for $l \in \{1,\dots,5\}$ such that $b_0 = \cdots = b_l = 1$ in the binary representation of k, and compare $N_{\mathrm{a}}(d,k,\nu)$ given by (5.8) with $N_{\mathrm{b}}(d,k,\nu)$ given by (5.9) corresponding to Algorithm 5.2(a) and (b), respectively. For comparison, also the number $N_{\mathrm{r}}(d,k,\nu)$ of terms (or the upper estimate from (5.16), or both) required by the recursive algorithm from Theorem 5.6 are given, see Remark 5.7(c). The size $\lvert\Delta_{d,\nu}\rvert$ of the support of Q is determined via (5.7).

In Algorithm 5.2(b), by the same reasoning and referring to (5.2), there are $\frac{1}{2}\sum_{i=0}^{l-1}\lvert\Delta_{d,2^i\nu}\rvert(\lvert\Delta_{d,2^i\nu}\rvert+1)$ non-zero summands (and products) in total to calculate the l convolution powers in (5.4). Let $i_{\min} \coloneqq \min\{i \in \{0,\dots,l\} \mid b_i = 1\}$ indicate the first non-trivial convolution factor in (5.5). Then there are in total $\lvert\Delta_{d,2^i\nu}\rvert \cdot \lvert\Delta_{d,(2^0b_0+\cdots+2^{i-1}b_{i-1})\nu}\rvert$ non-zero terms (and products) for each $i \in \{1,\dots,l\}$ satisfying $i > i_{\min}$ and $b_i = 1$ when calculating the convolution of Q^{*2^i} with $Q^{2^{i-1}b_{i-1}} * \cdots * Q^{*2^0 b_0}$ in (5.5). In total we get

$$N_{\mathrm{b}}(d,k,\nu) \coloneqq \frac{1}{2}\sum_{i=0}^{l-1}\lvert\Delta_{d,2^i\nu}\rvert\bigl(\lvert\Delta_{d,2^i\nu}\rvert+1\bigr) + \sum_{\substack{i=i_{\min}+1\\ b_i=1}}^{l}\lvert\Delta_{d,2^i\nu}\rvert\cdot\lvert\Delta_{d,(2^{i-1}b_{i-1}+\cdots+2^0b_0)\nu}\rvert \tag{5.9}$$

terms. Depending on the choice of the dimension $d \in \mathbb{N}$, the size parameter $\nu \in \mathbb{N}$ for the support of Q, and the power $k \in \mathbb{N}$ for Q^{*k}, the resulting number $N_{\mathrm{b}}(d,k,\nu)$ in (5.9) can be smaller or, particularly in higher dimensions, substantially larger than $N_{\mathrm{a}}(d,k,\nu)$ in (5.8), see Table 5.1.

Exercise 5.4 (Revisit of the basic Bernoulli model)**.** Consider independent Bernoulli random variables $N_1,\dots,N_m$, allowing for pairwise different success probabilities $p_1,\dots,p_m \in [0,1]$. Show that the distribution of $N \coloneqq N_1+\cdots+N_m$, see (2.2), can be calculated using convolutions by $(m+1)m-2$ multiplications of probabilities and compare this result with the approach via (2.8).

Hint: Study Algorithm 5.2(a) and Example 5.3.

Instead of using iterated convolutions as explained in Algorithm 5.2, it is possible to use a direct recursion based on the following observation (which is well known for powers of formal power series, cf. [57], and goes back at least to Euler [18, Chapter 4, Section 76], see also Remark 5.28 below).

Lemma 5.5 (Relation for convolution powers)**.** *Let $(X_i)_{i\in\mathbb{N}}$ be a sequence of $\mathbb{N}_0^d$-valued, independent, identically distributed random vectors. For $k\in\mathbb{N}_0$ define $S_k=\sum_{i=1}^k X_i$, where the empty sum is the zero vector in $\mathbb{N}_0^d$. Then, for every $k\in\mathbb{N}_0$ and $n\in\mathbb{N}_0^d$,*

$$\sum_{\substack{j\in\mathbb{N}_0^d\\ j\le n}} \bigl((k+1)j-n\bigr)\,\mathbb{P}[S_k=n-j]\,\mathbb{P}[X_1=j]=0. \tag{5.10}$$

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 (5.10) 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 (5.1) in the form

$$n\,\mathbb{P}[S_{k+1}=n]=n\sum_{\substack{j\in\mathbb{N}_0^d\\ j\le n}}\mathbb{P}[S_k=n-j]\,\underbrace{\mathbb{P}[X_{k+1}=j]}_{=\,\mathbb{P}[X_1=j]}. \tag{5.11}$$

Furthermore,

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

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

$$\begin{aligned} n\,\mathbb{P}[S_{k+1}=n]&=(k+1)\,\mathbb{E}\bigl[X_{k+1}\mathbb{1}_{\{S_{k+1}=n\}}\bigr]\\ &=(k+1)\sum_{\substack{j\in\mathbb{N}_0^d\\ j\le n}} j\,\underbrace{\mathbb{P}[S_k=n-j,\,X_{k+1}=j]}_{=\,\mathbb{P}[S_k=n-j]\,\mathbb{P}[X_1=j]}. \end{aligned} \tag{5.12}$$

Subtracting (5.11) from (5.12) yields (5.10). □

Theorem 5.6 (Recursion for convolution powers)**.**

(a) *There exists a vector $c\in[0,\infty)^d\setminus\{0\}$ such that there is 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 $\mathbb{P}[X_1=j]>0$.*

(b) *For c and m satisfying (a), and for every natural number $k \geq 2$, the distribution of $S_k = X_1 + \cdots + X_k$ can be calculated by*

$$\mathbb{P}[S_k = n] = \begin{cases} 0 & \text{if } n \in \mathbb{N}_0^d \setminus \{km\} \text{ with } \langle c,n\rangle \leq k\langle c,m\rangle, \\ \bigl(\mathbb{P}[X_1 = m]\bigr)^k & \text{if } n = km, \end{cases} \tag{5.13}$$

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

$$\begin{aligned} \mathbb{P}[S_k = n] = {} & \frac{1}{\langle c, n-km\rangle\,\mathbb{P}[X_1 = m]} \\ & \times \sum_{j\in N_{m,n}} \langle c,(k+1)j-m-n\rangle\,\mathbb{P}[S_k = m+n-j]\,\mathbb{P}[X_1 = j] \end{aligned} \tag{5.14}$$

with summation over

$$\begin{aligned} N_{m,n} \coloneqq \bigl\{ j \in \mathbb{N}_0^d \bigm| {} & \mathbb{P}[X_1 = j] > 0 \text{ and } j \leq m+n \\ & \text{and } \langle c,m\rangle < \langle c,j\rangle \leq \langle c, n-(k-1)m\rangle \bigr\}. \end{aligned} \tag{5.15}$$

Proof. (a) 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 (5.14). 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:

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 $c \coloneqq (1,\ldots,1) \in \mathbb{N}^d$ have the properties required in (a).

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 $\|j\|_1 \geq \tilde{m}$.

For the general case, consider

$$M \coloneqq \bigl\{ m \in \mathbb{N}_0^d \bigm| \|m\|_1 = \tilde{m} \text{ and } \mathbb{P}[X_1 = m] > 0 \bigr\}$$

and let

$$\begin{aligned} M_0 \coloneqq \bigl\{ (m_{d:d}, m_{d-1:d}, \ldots, m_{1:d}) \bigm| {} & (m_1,\ldots,m_d) \in M, \\ & m_{1:d} \leq m_{2:d} \leq \cdots \leq m_{d:d} \bigr\} \end{aligned}$$

be the set of tuples arising from M such that their components are ordered is a decreasing way. To optimize componentwise, iteratively for each $i \in \{1,\ldots,d\}$, let

$$M_i \coloneqq \bigl\{ (m_1,\ldots,m_d) \in M_{i-1} \bigm| m_i = \max\{n_i \mid (n_1,\ldots,n_d) \in M_{i-1}\} \bigr\}.$$

Then M_d contains a single element. Now fix an $m = (m_1,\ldots,m_d) \in M$ such that there exists a permutation π of $\{1,\ldots,d\}$ satisfying $(m_{\pi(1)},\ldots,m_{\pi(d)}) \in M_d$.

Define $c = (c_1, \dots, c_d) \in \mathbb{N}^d$ by $c_{\pi(i)} = \tilde{c} + i$ for all $i \in \{1, \dots, d\}$, where $\tilde{c} := \sum_{i=1}^{d}(i-1)m_{\pi(i)}$.

To verify that (a) holds for m and c, first consider $j = (j_1, \dots, j_d) \in \mathbb{N}_0^d$ such that $\|j\|_1 \geq \tilde{m} + 1$. Then

$$\langle c, j \rangle = \sum_{i=1}^{d}(\tilde{c} + i)j_{\pi(i)} \geq (\tilde{c}+1)\|j\|_1 \geq (\tilde{c}+1)(\tilde{m}+1),$$

hence

$$\langle c, m \rangle = \sum_{i=1}^{d}(\tilde{c} + i)m_{\pi(i)} = (\tilde{c}+1)\tilde{m} + \underbrace{\sum_{i=1}^{d}(i-1)m_{\pi(i)}}_{=\tilde{c}} < \langle c, j \rangle.$$

It remains to consider $j = (j_1, \dots, j_d) \in M \backslash \{m\}$. Since $c_{\pi(1)} < c_{\pi(2)} < \dots < c_{\pi(d)}$, we may assume by the rearrangement inequality that $j_{\pi(1)} \geq j_{\pi(2)} \geq \dots \geq j_{\pi(d)}$, because $\langle c, m \rangle < \langle c, j \rangle$ otherwise. Hence $(j_{\pi(1)}, \dots, j_{\pi(d)}) \in M_0$. By the above construction of m we must have $(j_{\pi(1)}, \dots, j_{\pi(d)}) = (m_{\pi(1)}, \dots, m_{\pi(d)})$ unless there exists an $i_0 \in \{1, \dots, d\}$ with $j_{\pi(i_0)} < m_{\pi(i_0)}$. In this case take the smallest i_0. Since $\|j\|_1 = \|m\|_1$, there exists a largest $i_1 \in \{i_0 + 1, \dots, d\}$ with $j_{\pi(i_1)} > m_{\pi(i_1)}$. Define $j' = (j'_1, \dots, j'_d)$ by

$$j'_{\pi(i)} = \begin{cases} j_{\pi(i_0)} + 1 & \text{for } i = i_0, \\ j_{\pi(i_1)} - 1 & \text{for } i = i_1, \\ j_{\pi(i)} & \text{for } i \in \{1, \dots, d\} \setminus \{i_0, i_1\}. \end{cases}$$

Then $\langle c, j \rangle = \langle c, j' \rangle + i_1 - i_0 > \langle c, j' \rangle$ and still $j'_{\pi(1)} \geq j'_{\pi(2)} \geq \dots \geq j'_{\pi(d)}$, because if $i_0 \geq 2$ then $j'_{\pi(i_0-1)} = j_{\pi(i_0-1)} = m_{\pi(i_0-1)} \geq m_{\pi(i_0)} \geq j'_{\pi(i_0)}$ by the minimality of i_0, and if $i_1 \leq d-1$, then $j'_{\pi(i_1)} \geq m_{\pi(i_1)} \geq m_{\pi(i_1+1)} \geq j_{\pi(i_1+1)} = j'_{\pi(i_1+1)}$ by the maximality of i_1. Iterating this procedure, we arrive at $(m_{\pi(1)}, \dots, m_{\pi(d)})$, which proves that $\langle c, j \rangle > \langle c, m \rangle$, verifying the assumption in Theorem 5.6.

(b) 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 $\mathbb{R}_+$-valued random variables. Fix the natural number $k \geq 2$. Define $S'_k = X'_1 + \dots + X'_k$. Then $0 \leq S'_k = \langle c, S_k - km \rangle$ and $\{S'_k = 0\} = \{X'_1 = 0, \dots, X'_k = 0\} = \{X_1 = m, \dots, X_k = m\}$. Using the i.i.d. assumption, (5.13) follows.

To prove (5.14) for a given $n \in \mathbb{N}_0^d$ with $\langle c, n \rangle > k\langle c, m \rangle$, rewrite (5.10) 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 (5.14) 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 (5.14) with $\langle c, m+n-j \rangle < k\langle c, m \rangle$. These are the ones with $\langle c, j \rangle > \langle c, m+n \rangle - k\langle c, m \rangle = \langle c, n-(k-1)m \rangle$. Since $j \leq m+n$ by the rewritten version of (5.10), this justifies to sum only over $j \in N_{m,n}$ given by (5.15). □

Remark 5.7 (Applicability of Theorem 5.6)**.**

(a) Equation (5.14) 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, since for $l := m + n - j$ this follows from the last strict inequality in definition (5.15) of $N_{m,n}$.

(b) Contrary to the approach given in Algorithm 5.2, *the recursion* (5.14) *can be numerically unstable,* because already in one dimension (hence $c := 1$ works), for a distribution with $m = 0$, and $n > k + 1$ the term $(k + 1)j - n$ changes sign as j runs from 1 to n. For an example, see Exercise 5.8 below.

(c) As in Example 5.3, assume that X_1 attains every vector in the d-dimensional simplex $\Delta_{d,\nu}$ given by (5.6) with strictly positive probability. Then $\mathbb{P}[S_k = n] > 0$ if and only if $n \in \Delta_{d,k\nu}$, and for every $n \in \Delta_{d,k\nu} \setminus \{0\}$ there are at most $|\Delta_{d,\nu}| - 1$ non-zero terms in the sum in (5.14), hence with the term for $n = 0$ given by (5.13) the total number $N_{\mathrm{r}}(d, k, \nu)$ of terms in the recursive method satisfies

$$N_{\mathrm{r}}(d, k, \nu) \le 1 + (|\Delta_{d,k\nu}| - 1)(|\Delta_{d,\nu}| - 1), \tag{5.16}$$

see (5.7) for further evaluation and Table 5.1 for comparison. For $d = 1$ or $\nu = 1$ the value of $N_{\mathrm{r}}(d, k, \nu)$ can be determined explicitly.

Exercise 5.8 (Complete cancellation in recursion (5.14))**.** Fix $l \in \mathbb{N}$ with $l \ge 3$ 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 (5.14) for $k = l - 1$ and $n = 2l - 1$ contains exactly two non-zero terms of opposite sign, hence complete cancellation occurs and $\mathbb{P}[S_{l-1} = 2l - 1] = 0$.

5.2 Panjer Distributions and Extended Panjer Recursion

As in Subsection 4.6, 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, which is independent of N. We want to calculate the distribution

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

of the random sum $S = X_1 + \cdots + X_N$ defined in (4.69). If the distribution

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

of N satisfies the recursion formula given in Definition 5.9 below, then Theorem 5.16 shows that there is an efficient way to do this.

Definition 5.9 (Panjer distribution)**.** 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 $q_0 = q_1 = \cdots = q_{k-1} = 0$ and

$$q_n = \Big(a + \frac{b}{n}\Big) q_{n-1} \quad \text{for all } n \in \mathbb{N} \text{ with } n \ge k + 1. \tag{5.17}$$

Remark 5.10 (Uniqueness)**.** Given $a, b \in \mathbb{R}$ and $k \in \mathbb{N}_0$, the linearity of (5.17) implies that there exists at most one probability distribution $(q_n)_{n\in\mathbb{N}_0}$ satisfying Definition 5.9, because there can be at most one $q_k \in (0, 1]$ such that $\sum_{n=k}^{\infty} q_n = 1$.

Definition 5.11 (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 $(\tilde{q}_n)_{n\in\mathbb{N}_0}$ of $(q_n)_{n\in\mathbb{N}_0}$ is defined by $\tilde{q}_0 = \cdots = \tilde{q}_{l-1} = 0$ and

$$\tilde{q}_n = \frac{q_n}{1 - \sum_{j=0}^{l-1} q_j}, \quad n \geq l. \tag{5.18}$$

For $N \sim (q_n)_{n\in\mathbb{N}}$ the l-truncated probability distribution $(\tilde{q}_n)_{n\in\mathbb{N}_0}$ is the conditional distribution satisfying $\tilde{q}_n = \mathbb{P}[N = n \,|\, N \geq l]$ for all $n \in \mathbb{N}_0$.

Lemma 5.12 (Truncations of Panjer distributions)**.** *Suppose* $(q_n)_{n\in\mathbb{N}_0}$ *is the* $\mathrm{Panjer}(a, b, k)$ *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* $\mathrm{Panjer}(a, b, l)$ *distribution.*

Exercise 5.13. Prove Lemma 5.12 using the linearity of the recursion equation (5.17)

Remark 5.14 (List of Panjer distributions)**.** All probability distributions satisfying Definition 5.9 were identified by Sundt and Jewell [52] for the case $k = 0$, Willmot [60] for the case $k = 1$, and finally Hess, Liewald and Schmidt [29] for general $k \in \mathbb{N}_0$. The Panjer distributions are the following:

(a) Poisson distribution (see Example 5.21),

(b) Negative binomial distribution (see Example 5.22),

(c) Binomial distribution (see Example 5.24),

(d) Logarithmic distribution (see Example 5.25),

(e) Extended negative binomial distribution (see Example 5.26),

(f) Extended logarithmic distribution (see Example 5.27),

(g) All truncations of these distributions (see Definition 5.11 and Lemma 5.12).

Exercise 5.15 (Sundt and Jewell [52])**.** Prove that the only non-degenerate probability distributions in the class $\{\mathrm{Panjer}(a, b, 0) \mid a, b \in \mathbb{R}\}$ are the Poisson, the binomial, and the negative binomial distributions.

Hint: Discuss the cases $a < 0$, $a = 0$ and $a > 0$ separately.

The following theorem combines results of Panjer [42] and Hess, Liewald and Schmidt [29] with the multivariate extension of Sundt [51].

Theorem 5.16 (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. The distribution $(p_n)_{n\in\mathbb{N}_0^d}$ of the random sum S defined in (4.69) is denoted by* CPanjer(a,b,k,Q) *with $Q := \mathcal{L}(X_1)$. If $a\,\mathbb{P}[X_1=0]\neq 1$, then $\mathcal{L}(S)$ can be calculated by*

$$p_0 = \varphi_N(\mathbb{P}[X_1=0]) = \begin{cases} q_0 & \text{if } \mathbb{P}[X_1=0]=0,\\ \mathbb{E}\big[(\mathbb{P}[X_1=0])^N\big] & \text{otherwise,}\end{cases} \tag{5.19}$$

where φ_N is the probability-generating function of N, and the recursion formula

$$p_n = \frac{1}{1-a\,\mathbb{P}[X_1=0]}\Biggl(\mathbb{P}[S_k=n]\,q_k + \sum_{\substack{j\in\mathbb{N}_0^d\\ 0\neq j\leq n}}\biggl(a+\frac{b\langle c_n,j\rangle}{\langle c_n,n\rangle}\biggr)\,\mathbb{P}[X_1=j]\,p_{n-j}\Biggr) \tag{5.20}$$

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 $c_n := (1,\dots,1)$ works in every case.

Proof. Theorem 5.16 is a corollary of Theorem 5.30(a) below, hence its proof is given just after the statement of Theorem 5.30. □

Remark 5.17 (Recursional character and parallel computing). Observe that (5.20) is indeed a recursion: For the computation of p_n only values p_{n-j} with $j\neq 0$ are used, these satisfy $\|n-j\|_1 < \|n\|_1$. In dimension $d\geq 2$, for the same reason, the values p_n for vectors n sharing the same $\|\cdot\|_1$-norm can be computed in parallel.

Remark 5.18 (Technical assumption). Of the Panjer distributions given in Remark 5.14, only the uninteresting case $\mathbb{P}[X_1=0]=1$ with $N\sim\mathrm{ExtLog}(k,1)$, see Example 5.27 below, or one of its truncations, see Lemma 5.12, violates the technical assumption $a\,\mathbb{P}[X_1=0]\neq 1$. Obviously, $p_0=1$ and $p_n=0$ for all $n\in\mathbb{N}_0^d\setminus\{0\}$ in these cases, and we define $\mathrm{CPanjer}(1,-k,k,\delta_0)=\delta_0$ for each $k\in\mathbb{N}$ with $k\geq 2$.

Remark 5.19 (Computational speed-up for small support of $\mathcal{L}(X_1)$). For $n=(n_1,\dots,n_d)\in\mathbb{N}_0^d\setminus\{0\}$, the number of terms in (5.20) 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 write $\mathcal{S}_X=\{n\in\mathbb{N}_0^d\setminus\{0\}\mid \mathbb{P}[X_1=n]>0\}$ for this support without the origin of $\mathbb{N}_0^d$. Then the sum in (5.20) runs over all $j\in\mathcal{S}_X$ satisfying $j\leq n$, i.e.

$$j\in\mathcal{S}_n(X) := \mathcal{S}_X\cap\bigtimes_{i=1}^d\{0,\dots,n_i\},$$

and their cardinalities satisfy $|\mathcal{S}_n(X)|\leq\min\{|\mathcal{S}_X|,(n_1+1)\cdots(n_d+1)-1\}$. If $|\mathcal{S}_X|<\infty$, then $|\mathcal{S}_X|$ is an upper bound for the number of terms which doesn't grow with n. Remark 5.20 below simplifies the computation of the individual terms.

Remark 5.20 (Choice of c_n)**.** While $c_n = (1, \dots, 1)$ works in (5.20) for every $n \in \mathbb{N}_0^d \setminus \{0\}$, there is a computational advantage in choosing c_n dependent on n. To illustrate this, let us take the notation of Remark 5.19 and define $\mathcal{S}_{i,n}(X) = \{j_i \mid (j_1, \dots, j_d) \in \mathcal{S}_n(X)\}$ for every $i \in \{1, \dots, d\}$. Since every $n = (n_1, \dots, n_d) \in \mathcal{S}_X$ has at least one non-zero component, let's say the ith one n_i, we can then choose the unit vector $c_n = (0, \dots, 0, 1, 0, \dots, 0)$ with the 1 at the ith 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 (5.20), i.e.,

$$\begin{aligned}\sum_{j\in\mathcal{S}_n(X)} \biggl(a + \frac{b\langle c_n, j\rangle}{\langle c_n, n\rangle}\biggr) \mathbb{P}[X_1 = j]\, p_{n-j}& \\ = \sum_{l\in\mathcal{S}_{i,n}(X)} \biggl(a + \frac{bl}{n_i}\biggr) &\sum_{\substack{j=(j_1,\dots,j_d)\in\mathcal{S}_n(X)\\ j_i = l}} \mathbb{P}[X_1 = j]\, p_{n-j}.\end{aligned}$$

Before we derive Theorem 5.16 from Theorem 5.30(a) below, let us look at several examples and keep the question of numerical stability for the recursion formula (5.20) in mind.

Example 5.21 (Poisson distribution)**.** If $(q_n)_{n\in\mathbb{N}_0}$ is Poisson(λ) with $\lambda \ge 0$, then $q_0 = \mathrm{e}^{-\lambda}$ and

$$q_n \overset{(3.1)}{=} \frac{\lambda^n}{n!}\,\mathrm{e}^{-\lambda} = \frac{\lambda}{n} q_{n-1}, \quad n \in \mathbb{N},$$

hence by comparison[40] with (5.17), Poisson(λ) is the Panjer($0, \lambda, 0$) distribution. Using (4.3), the initial value (5.19) turns into

$$p_0 = \mathrm{e}^{\lambda(\mathbb{P}[X_1=0]-1)}\,. \tag{5.21}$$

The recursion formula (5.20) can be simplified to

$$p_n = \frac{\lambda}{n_i} \sum_{\substack{j\in\mathbb{N}_0^d\\ 0\ne j\le n}} j_i\, \mathbb{P}[X_1 = j]\, p_{n-j}, \tag{5.22}$$

for every $n = (n_1, \dots, n_d) \in \mathbb{N}_0^d \setminus \{0\}$, where $i \in \{1, \dots, d\}$ is chosen such that $n_i \ne 0$. See Remark 5.19 to omit terms in (5.22) with value zero. The recursion (5.22) is numerically stable because only non-negative numbers are multiplied and added.

[40] Here and in the following examples we use that in the vector space of all functions $\mathbb{N} \to \mathbb{R}$ the constant function $f \equiv 1$ and the function $g(n) := 1/n$ for all $n \in \mathbb{N}$ are linearly independent, hence the coefficients for functions in the linear span of them are uniquely determined.

Example 5.22 (Negative binomial distribution)**.** If $(q_n)_{n\in\mathbb{N}_0}$ is $\mathrm{NegBin}(\alpha,p)$ with parameters $\alpha>0$ and $p\in[0,1)$ as specified in (4.61), 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 $\mathrm{NegBin}(\alpha,p)$ is the $\mathrm{Panjer}(p,(\alpha-1)p,0)$ distribution by comparison with (5.17). Using (4.65), the initial value (5.19) turns into

$$p_0=\Bigl(\frac{q}{1-p\,\mathbb{P}[X_1=0]}\Bigr)^{\alpha}. \tag{5.23}$$

The recursion formula (5.20) can be simplified to

$$p_n=\frac{p}{n_i(1-p\,\mathbb{P}[X_1=0])}\sum_{\substack{j\in\mathbb{N}_0^d\\0\neq j\leq n}}(\alpha j_i+n_i-j_i)\,\mathbb{P}[X_1=j]\,p_{n-j} \tag{5.24}$$

for every $n=(n_1,\dots,n_d)\in\mathbb{N}_0^d\setminus\{0\}$, where $i\in\{1,\dots,d\}$ is chosen such that $n_i\neq0$. See Remark 5.19 for the possibility to omit terms in (5.24) with value zero. The recursion (5.24) is numerically stable because $n_i-j_i\in\mathbb{N}_0$ (this requires proper programming, αj_i has to be added afterwards) and otherwise only non-negative numbers are multiplied and added to calculate the sum.

Remark 5.23 (Algorithm for a very small initial value)**.** To apply the d-dimensional extended Panjer recursion (5.20), the probability p_0 of a loss of zero is needed as starting value, see (5.19). If $N\sim\mathrm{Poisson}(\lambda)$ with $\lambda\geq0$, then p_0 is given by (5.21). If $N\sim\mathrm{NegBin}(\alpha,p)$ with $\alpha>0$ and $p\in[0,1)$, then p_0 is given by (5.23). When modeling large portfolios with the collective risk model (4.69) using one of these two claim number distributions, it can happen for large λ or α, respectively, that p_0 is so small that it can only be represented as zero on a computer (arithmetic underflow). The recursion (5.20) then produces $p_n=0$ for all $n\in\mathbb{N}_0^d\setminus\{0\}$, which is clearly wrong. The standard solution, cf. [34, Section 6.6.2], is to perform Panjer's recursion with the reduced parameter $\lambda':=\lambda/2^l$ (resp. $\alpha':=\alpha/2^l$) instead, where $l\in\mathbb{N}$ is chosen such that the new starting value p_0 is properly representable on the computer. Afterwards, l iterative and numerically stable convolutions are needed to calculate the original probability distribution, see (5.4) in Subsection 5.1. This approach works because for independent $N_1,\dots,N_{2^l}\sim\mathrm{Poisson}(\lambda/2^l)$, we have that $N=N_1+\dots+N_{2^l}\sim\mathrm{Poisson}(\lambda)$ by the summation property, see Lemma 3.2, similarly for the negative binomial distribution, see Lemma 4.37. For the corresponding property of the compound distributions, see Corollary 4.43. In general, this approach works for claim number distributions closed under convolutions.

Example 5.24 (Binomial distribution)**.** Let $(q_n)_{n\in\mathbb{N}_0}$ denote the binomial distribution $\mathrm{Bin}(m,p)$ with success probability $p\in[0,1)$ and number of trials $m\in\mathbb{N}$.

Let $q \coloneqq 1-p$ denote the failure probability. Then, for every $n \in \mathbb{N}$,

$$\begin{aligned} q_n = \binom{m}{n} p^n q^{m-n} &= \frac{m-n+1}{n}\frac{p}{q} \cdot \binom{m}{n-1} p^{n-1} q^{m-(n-1)} \\ &= \biggl(\underbrace{-\frac{p}{q}}_{=:\,a} + \underbrace{\frac{(m+1)p}{q}}_{=:\,b}\frac{1}{n}\biggr) q_{n-1} \end{aligned}$$

by comparison with (5.17), hence $\mathrm{Bin}(m,p)$ is the $\mathrm{Panjer}(-p/q, (m+1)p/q, 0)$ distribution. The recursion factor $a + b/n$ is zero for $n = m+1$, giving $q_n = 0$ for $n \ge m+1$ as expected. Using (4.33), the initial value (5.19) turns into

$$p_0 = \bigl(1 + p(\mathbb{P}[X_1 = 0] - 1)\bigr)^m. \tag{5.25}$$

Consider Panjer's recursion formula (5.20) for $n = (n_1, \dots, n_d) \in \mathbb{N}_0^d \setminus \{0\}$ with $n_1 \ge m+2$ and $n_2 = \cdots = n_d = 0$. Without loss of generality we can take $c_n = (1, 0, \dots, 0)$. Then the term

$$a + \frac{b\langle c_n, j\rangle}{\langle c_n, n\rangle} = -\frac{p}{q}\Bigl(1 - \frac{m+1}{n_1} j_1\Bigr)$$

changes sign as $j = (j_1, 0, \dots, 0)$ varies between $(1, 0, \dots, 0)$ and $(n_1, 0, \dots, 0)$. Therefore, the recursion can be numerically unstable because cancellations can occur. The problem with numerical underflow during the calculation of the initial value p_0 given in (5.25) can also occur for large m, see Remark 5.23. Since

$$\varphi_S(s) = \varphi_N\bigl(\varphi_{X_1}(s)\bigr) = \bigl(q + p\varphi_{X_1}(s)\bigr)^m = \prod_{\substack{k=0\\ b_k=1}}^{l} \bigl(q + p\varphi_{X_1}(s)\bigr)^{2^k}$$

at least for all $s \in \mathbb{C}^d$ with $\|s\|_\infty \le 1$, where $m = \sum_{k=0}^{l} b_k 2^k$ with $b_1, \dots, b_{l-1} \in \{0,1\}$, $b_l = 1$ and $l = \lfloor \log_2 m \rfloor$ denotes the binary representation of m, we see that the distribution $(p_n)_{n \in \mathbb{N}_0^d}$ of S can be computed in a numerically stable way with $b_0 + \cdots + b_{l-1} + l \le 2l$ convolutions, see Algorithm 5.2.

Example 5.25 (Logarithmic distribution)**.** If $(q_n)_{n \in \mathbb{N}_0}$ is $\mathrm{Log}(p)$ with $p \in [0,1)$, see Example 4.4, then $q_0 = 0$, $q_1 = 1/c(p)$ with $c(p)$ defined by (4.5), and

$$q_n = \frac{p^{n-1}}{c(p)\, n} = p\frac{n-1}{n} q_{n-1} \quad \text{for } n \in \mathbb{N},\ n \ge 2,$$

hence by comparison with (5.17), $\mathrm{Log}(p)$ is the $\mathrm{Panjer}(p, -p, 1)$ distribution. Using (4.6), the initial value (5.19) turns into

$$p_0 = \mathbb{P}[X_1 = 0]\, \frac{c(p\,\mathbb{P}[X_1 = 0])}{c(p)}. \tag{5.26}$$

The recursion formula (5.20) simplifies to

$$p_n = \frac{1}{1 - p\,\mathbb{P}[X_1 = 0]}\biggl(\frac{\mathbb{P}[X_1 = n]}{c(p)} + \frac{p}{n_i} \sum_{\substack{j \in \mathbb{N}_0^d\\ 0 \ne j < n\\ j_i < n_i}} (n_i - j_i)\, \mathbb{P}[X_1 = j]\, p_{n-j}\biggr) \tag{5.27}$$

for every $n = (n_1, \dots, n_d) \in \mathbb{N}_0^d \setminus \{0\}$, where $i \in \{1, \dots, d\}$ is chosen such that $n_i \neq 0$. See Remark 5.19 about the possibility to omit further terms in (5.27) with value zero. The recursion (5.27) 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 $p = 0$, hence $c(p) = 1$ by (4.5), the recursion (5.27) simplifies dramatically to $p_n = \mathbb{P}[X_1 = n]$ for all $n \in \mathbb{N}_0^d \setminus \{0\}$.

Example 5.26 (Extended negative binomial distribution)**.** For parameters $k \in \mathbb{N}$, $\alpha \in (-k, 0) \setminus \{-1, -2, \dots, -(k-1)\}$ and $p \in (0, 1]$, define $q_0 = \dots = q_{k-1} = 0$ and, using the abbreviation $q := 1 - p$,

$$q_n = \frac{\binom{\alpha+n-1}{n} p^n}{q^{-\alpha} - \sum_{j=0}^{k-1} \binom{\alpha+j-1}{j} p^j} \qquad \text{for } n \geq k. \tag{5.28}$$

We will verify below that (5.28) is a well-defined probability distribution $(q_n)_{n\in\mathbb{N}_0}$, called extended negative binomial distribution, notation $\mathrm{ExtNegBin}(\alpha, k, p)$.

First note that the k-truncation, see (5.18) in Definition 5.11, of the negative binomial distribution defined in (4.61) gives the same formula, however with $\alpha > 0$ and valid for $p \in [0, 1)$. A short calculation shows that the k-truncation of an $\mathrm{ExtNegBin}(\alpha, l, p)$ distribution with $l \in \{1, \dots, k-1\}$, $\alpha \in (-l, -l+1)$ and $p \in (0, 1]$ is also given by (5.28). Hence, for every $k \in \mathbb{N}$, the formula (5.28) defines a probability distribution for all $\alpha \in (-k, \infty) \setminus \{0, -1, -2, \dots\}$ and $p \in (0, 1)$. If $-\alpha \in \mathbb{N}_0$, then $\binom{\alpha+n-1}{n} = 0$ 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 $n \in \{1, \dots, -\alpha\}$, hence (5.28) cannot define an interesting probability distribution is this case.

To verify that (5.28) defines a probability distribution, note that, for every $n \in \mathbb{N}_0$,

$$\begin{aligned}\binom{\alpha+n-1}{n} &= \frac{1}{n!}\prod_{j=1}^{n}(\alpha+n-j) \\ &= \frac{(-1)^n}{n!}\prod_{l=0}^{n-1}(-\alpha-l) = (-1)^n\binom{-\alpha}{n},\end{aligned} \tag{5.29}$$

and, for all integers $n \geq k$,

$$\binom{\alpha+n-1}{n} = \left(\prod_{j=1}^{k}\frac{\alpha+j-1}{j}\right)\prod_{j=k+1}^{n}\underbrace{\Bigl(1+\frac{\overbrace{\alpha-1}^{>-k-1}}{j}\Bigr)}_{>0}$$

has the same sign. Using $\log(1+x) \leq x$ for $x > -1$ and noting that $\alpha - 1 < 0$,

$$\log\prod_{j=k+1}^{n}\Bigl(1+\frac{\alpha-1}{j}\Bigr) \leq \sum_{j=k+1}^{n}\frac{\alpha-1}{j} \leq (\alpha-1)\int_{k+1}^{n+1}\frac{\mathrm{d}x}{x} = \log\Bigl(\frac{n+1}{k+1}\Bigr)^{\alpha-1}$$

for all integers $n \geq k$. Therefore,

$$\sum_{n=k}^{\infty} \left| \binom{\alpha+n-1}{n} \right| \leq \left| \prod_{j=1}^{k} \frac{\alpha+j-1}{j} \right| \sum_{n=k}^{\infty} \Bigl(\frac{k+1}{n+1}\Bigr)^{1-\alpha} < \infty$$

by the integral test for convergence, because $1-\alpha > 1$ and

$$\sum_{n=k}^{\infty} \frac{1}{(n+1)^{1-\alpha}} \leq \int_k^{\infty} \frac{\mathrm{d}x}{x^{1-\alpha}} = -\frac{1}{\alpha k^{-\alpha}} < \infty.$$

Using (5.29), we see that the binomial series

$$(1+x)^{-\alpha} = \sum_{n\in\mathbb{N}_0} \binom{-\alpha}{n} x^n \tag{5.30}$$

converges absolutely for all $x \in \mathbb{C}$ with $|x| \leq 1$; for $x = -p$ we see that

$$\sum_{n\in\mathbb{N}_0} \binom{\alpha+n-1}{n} p^n = \sum_{n\in\mathbb{N}_0} \binom{-\alpha}{n} (-p)^n = (1-p)^{-\alpha} = q^{-\alpha}. \tag{5.31}$$

We conclude that the nominators in (5.28) are all of the same sign and, by (5.31), 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 (5.29) and an index shift, we see that, for every $n \geq k+1$,

$$\begin{aligned}
\binom{\alpha+n-1}{n} p^n &= \frac{\alpha+n-1}{n} p \cdot \frac{p^{n-1}}{(n-1)!} \prod_{j=1}^{n-1} (\alpha+n-1-j) \\
&= \Bigl(1 + \frac{\alpha-1}{n}\Bigr) p \cdot \binom{\alpha+n-2}{n-1} p^{n-1},
\end{aligned}$$

hence by comparison with (5.17), $\mathrm{ExtNegBin}(\alpha, k, p)$ is the $\mathrm{Panjer}(p, (\alpha-1)p, k)$ distribution. Consider Panjer's recursion formula (5.20) 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

$$a + \frac{b\langle c_n, j\rangle}{\langle c_n, n\rangle} = \Bigl(1 + \frac{\alpha-1}{n_1} j_1\Bigr) p$$

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 5.34 below.

To calculate the probability-generating function of a random variable $N \sim \mathrm{ExtNegBin}(\alpha, k, p)$, note that by (5.31) applied to ps in place of p,

$$\sum_{n\in\mathbb{N}_0} \binom{\alpha+n-1}{n} p^n s^n = (1-ps)^{-\alpha}, \qquad |s| \leq \frac{1}{p},$$

therefore

$$\varphi_N(s) = \sum_{n\in\mathbb{N}_0} q_n s^n \overset{(5.28)}{=} \frac{(1-ps)^{-\alpha} - \sum_{j=0}^{k-1}\binom{\alpha+j-1}{j}(ps)^j}{q^{-\alpha} - \sum_{j=0}^{k-1}\binom{\alpha+j-1}{j}p^j}, \qquad |s| \le \frac{1}{p}. \tag{5.32}$$

For $k = 1$, hence $\alpha \in (-1, 0)$, this simplifies to

$$\varphi_N(s) = \frac{1-(1-ps)^{-\alpha}}{1-q^{-\alpha}}, \qquad |s| \le \frac{1}{p}. \tag{5.33}$$

Example 5.27 (Extended logarithmic distribution)**.** Assume that $(q_n)_{n\in\mathbb{N}_0}$ is an extended logarithmic distribution, notation ExtLog(k,p), with parameters $k \in \mathbb{N}$, $k \ge 2$, and $p \in (0,1]$, which means that $q_0 = \cdots = q_{k-1} = 0$ and

$$q_n = \frac{\binom{n}{k}^{-1} p^n}{\sum_{l=k}^{\infty}\binom{l}{k}^{-1} p^l}, \qquad n \ge k. \tag{5.34}$$

Since, for every $m \in \mathbb{N}$ with $m \ge k$,

$$\sum_{l=k}^{m} \frac{1}{\binom{l}{k}} \le \sum_{l=k}^{m} \frac{k!}{l(l-1)} = k! \sum_{l=k}^{m}\Big(\frac{1}{l-1} - \frac{1}{l}\Big) = k!\Big(\frac{1}{k-1} - \frac{1}{m}\Big) \le \frac{k!}{k-1},$$

the extended logarithmic distribution is well defined for every $p \in (0,1]$. For $n \ge k+1$ we have

$$\binom{n}{k} = \frac{n}{n-k}\binom{n-1}{k},$$

which via (5.34) yields

$$q_n = \frac{n-k}{n} p \cdot q_{n-1} = \Big(p - \frac{kp}{n}\Big) q_{n-1},$$

hence by comparison with (5.17), ExtLog(k,p) is the Panjer$(p, -kp, k)$ distribution. Consider Panjer's recursion formula (5.20) for $n = (n_1, \dots, n_d) \in \mathbb{N}_0^d \setminus \{0\}$ with $n_1 \ge k+1$ and $n_2 = \cdots = n_d = 0$. Without loss of generality we can take $c_n = (1, 0, \dots, 0)$. Then the term

$$a + \frac{b\langle c_n, j\rangle}{\langle c_n, n\rangle} = p\Big(1 - \frac{kj_1}{n_1}\Big)$$

changes sign as $j = (j_1, 0, \dots, 0)$ varies between $(1, 0, \dots, 0)$ and $(n_1, 0, \dots, 0)$. Therefore, the recursion can be numerically unstable because cancellations might occur; see Subsection 5.5 and [22, Section 5.2] for a solution of this problem. We remark that a closed-form expression for the denominator in (5.34) is given by [22, Lemma 2.1], which makes it possible to express the probability-generating function also in closed form, cf. [22, (2.7)]

Remark 5.28 (Historical remark)**.** We mention that the one-dimensional Panjer recursion for claim number distributions given by the binomial, the negative binomial, and the extended negative binomial distribution with $k = 1$ is contained in a much older result: For $\alpha \in \mathbb{R}$ and a power series $f(s) = \sum_{j=0}^{\infty} a_j s^j$ with $a_0 \neq 0$, the coefficients $(b_n)_{n\in\mathbb{N}_0}$ of the power series $f^{-\alpha}(s)$ satisfy the recursion

$$b_n = \frac{1}{na_0} \sum_{j=1}^{n} ((1-\alpha)j - n) a_j b_{n-j}, \qquad n \in \mathbb{N}, \tag{5.35}$$

starting with $b_0 = 1/a_0^{\alpha}$. Gould [26] 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 (5.35) 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 5.29. Use (4.33), (4.65) and (5.33) to verify the last statement in Remark 5.28.

5.3 A Generalisation of the Multivariate Panjer Recursion

The multivariate extended Panjer recursion in Theorem 5.16 is a special case of part (a) of the following theorem, which combines [22, Theorem 4.5] with the multivariate idea in [51, Theorem 1] and is of independent interest for questions of numerical stability, see Subsections 5.4 and 5.5 below.

Theorem 5.30. *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 $i \in \{1,\dots,l\}$, where $(N, \tilde{N}_1,\dots,\tilde{N}_l)$ is independent of the $\mathbb{N}_0^d$-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 $S = X_1 + \dots + X_N$ and $\tilde{S}^{(i)} = X_1 + \dots + X_{\tilde{N}_i}$ for $i \in \{1,\dots,l\}$, respectively.*

(a) Assume[41] that there exist $k \in \mathbb{N}_0$ and $a_1,\dots,a_l, b_1,\dots,b_l \in \mathbb{R}$ such that

$$q_n = \sum_{i=1}^{l} \Big(a_i + \frac{b_i}{n}\Big) \tilde{q}_{i,n-i}, \quad n \in \mathbb{N} \text{ with } n \geq k + l, \tag{5.36}$$

and all probabilities not used on the right-hand side of (5.36) are zero, i.e.

$$\tilde{q}_{i,0} = \dots = \tilde{q}_{i,k+l-i-1} = 0, \quad i \in \{1,\dots,\min\{l, k+l-1\}\}. \tag{5.37}$$

Then, for every $n \in \mathbb{N}_0^d \setminus \{0\}$ and $c_n \in \mathbb{R}^d$ with $\langle c_n, n\rangle \neq 0$,

$$p_n = \sum_{j=1}^{k+l-1} \mathbb{P}[S_j = n]\, q_j + \sum_{i=1}^{l} \sum_{\substack{j\in\mathbb{N}_0^d \\ j \leq n}} \bigg(a_i + \frac{b_i \langle c_n, j\rangle}{i\langle c_n, n\rangle}\bigg) \mathbb{P}[S_i = j]\, \tilde{p}_{i,n-j} \tag{5.38}$$

with $S_j \coloneqq X_1 + \dots + X_j$, and p_0 is given by (5.19).

[41] In these lecture notes, we only apply this case with $l = 1$.

(b) *Assume that there exist* $\nu_1, \dots, \nu_l \in [0,1]$ *with* $\nu_1 + \dots + \nu_l \leq 1$ *such that* $q_n = \sum_{i=1}^{l} \nu_i \tilde{q}_{i,n}$ *for all* $n \in \mathbb{N}$. *Then* $p_n = \sum_{i=1}^{l} \nu_i \tilde{p}_{i,n}$ *for all* $n \in \mathbb{N}_0^d \setminus \{0\}$.

Remark 5.31 (Reformulation and proof of Theorem 5.30(b))**.** Let $Q := \mathcal{L}(X_1)$ denote the claim size distribution, define $\tilde{N}_0 \equiv 0$ and $\nu_0 = 1 - (\nu_1 + \dots + \nu_l) \in [0,1]$. Using the notation for finite convex combinations from Example 4.9, the assumption of Theorem 5.30(b) can be reformulated as

$$\mathcal{L}(N) = \operatorname{Convex}\big((\nu_i, \mathcal{L}(\tilde{N}_i))_{i\in\{0,\dots,l\}}\big)$$

hence $\varphi_N = \sum_{i=0}^{l} \nu_i \varphi_{\tilde{N}_i}$ by (4.22). Using (4.70), the result in (b) is

$$\operatorname{Compound}\big(\mathcal{L}(N), Q\big) = \operatorname{Convex}\big(\big(\nu_i, \operatorname{Compound}\big(\mathcal{L}(\tilde{N}_i), Q\big)\big)_{i\in\{0,\dots,l\}}\big),$$

which follows via (4.75) in Example 4.9, because

$$\varphi_S = \varphi_N \circ \varphi_Q = \left(\sum_{i=0}^{l} \nu_i \varphi_{\tilde{N}_i}\right) \circ \varphi_Q = \sum_{i=0}^{l} \nu_i (\varphi_{\tilde{N}_i} \circ \varphi_Q) = \nu_0 + \sum_{i=1}^{l} \nu_i \varphi_{\tilde{S}^{(i)}}.$$

Proof of Theorem 5.16. If $(q_n)_{n\in\mathbb{N}_0}$ is the Panjer(a,b,k) distribution, then Theorem 5.30(a) is applicable by choosing $l=1$ and $\tilde{q}_{1,n} = q_n$ for all $n \in \mathbb{N}_0$, which implies $p_n = \tilde{p}_{1,n}$ for all $n \in \mathbb{N}_0^d$. Using $q_0 = \dots = q_{k-1} = 0$, which implies (5.37), and solving (5.38) for p_n yields (5.20). □

Proof of Theorem 5.30. (a) We extend a standard proof (cf. [40, Theorem 3.3.9] for the case $k=0$ and $l=1$) with the idea from [51] for the d-dimensional setting.

To prove the representation for the initial value given in (5.19), note that

$$p_0 \overset{(4.14)}{=} \varphi_S(0) \overset{(4.75)}{=} \varphi_N(\varphi_{X_1}(0)) \overset{(4.14)}{=} \varphi_N(\mathbb{P}[X_1 = 0]).$$

We now prove (5.38) 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 $i \in \{1, \dots, l\}$. For every $m \in \mathbb{N}$ with $m \geq i$, we use the representation $S_m = X_1 + \dots + X_m = S_{m-i} + S_{i,m}$ with $S_{i,m} := X_{m-i+1} + \dots + X_m$ and independent and identically distributed $X_1, \dots, X_m$. If $\mathbb{P}[S_m = n] > 0$, then

$$\begin{aligned}\langle c, n\rangle = \mathbb{E}\big[\langle c, S_m\rangle \,\big|\, S_m = n\big] &= \sum_{j=1}^{m} \mathbb{E}\big[\langle c, X_j\rangle \,\big|\, S_m = n\big] \\ &= m\,\mathbb{E}\big[\langle c, X_m\rangle \,\big|\, S_m = n\big] = \frac{m}{i}\,\mathbb{E}\big[\langle c, S_{i,m}\rangle \,\big|\, S_m = n\big],\end{aligned}$$

hence by solving for $\frac{1}{m}$ and calculating the conditional expectation,

$$\begin{aligned}a_i + \frac{b_i}{m} &= \mathbb{E}\left[a_i + \frac{b_i \langle c, S_{i,m}\rangle}{i\langle c, n\rangle} \,\middle|\, S_m = n\right] \\ &= \sum_{\substack{j\in\mathbb{N}_0^d \\ j\leq n}} \left(a_i + \frac{b_i\langle c, j\rangle}{i\langle c, n\rangle}\right) \mathbb{P}[S_{i,m} = j \,|\, S_m = n]\,.\end{aligned} \tag{5.39}$$

For every $m \ge i$ the sums S_{m-i} and $S_{i,m}$ are independent, hence

$$\begin{aligned}\mathbb{P}[S_{i,m} = j,\, S_m = n] &= \mathbb{P}[S_{i,m} = j,\, S_{m-i} = n-j] \\ &= \underbrace{\mathbb{P}[S_{i,m} = j]}_{=\,\mathbb{P}[S_i=j]}\,\mathbb{P}[S_{m-i} = n-j]\,. \end{aligned} \tag{5.40}$$

We now rewrite $p_n = \mathbb{P}[S = n]$ using (5.36) as follows

$$\begin{aligned} p_n &= \sum_{\substack{m=1\\ q_m>0}}^{\infty} \underbrace{\mathbb{P}[S_m = n \,|\, N = m]}_{=\,\mathbb{P}[S_m=n] \text{ by indep.}}\underbrace{\mathbb{P}[N = m]}_{=\,q_m} \\ &= \sum_{m=1}^{k+l-1} \mathbb{P}[S_m = n]\,q_m + \underbrace{\sum_{m=k+l}^{\infty}\sum_{i=1}^{l}\Bigl(a_i + \frac{b_i}{m}\Bigr)\,\mathbb{P}[S_m = n]\,\tilde q_{i,m-i}\,.}_{=:\,(*)} \end{aligned} \tag{5.41}$$

Inserting (5.39) and (5.40) yields for the series

$$\begin{aligned} (*) &= \sum_{m=k+l}^{\infty}\sum_{i=1}^{l}\sum_{\substack{j\in\mathbb{N}_0^d\\ j\le n}}\biggl(a_i + \frac{b_i\langle c,j\rangle}{i\langle c,n\rangle}\biggr)\,\mathbb{P}[S_i = j]\,\mathbb{P}[S_{m-i} = n-j]\,\tilde q_{i,m-i} \\ &= \sum_{i=1}^{l}\sum_{\substack{j\in\mathbb{N}_0^d\\ j\le n}}\biggl(a_i + \frac{b_i\langle c,j\rangle}{i\langle c,n\rangle}\biggr)\,\mathbb{P}[S_i = j]\,\underbrace{\sum_{m=k+l}^{\infty}\mathbb{P}[S_{m-i} = n-j]\,\tilde q_{i,m-i}\,,}_{=:\,(**)} \end{aligned}$$

where the rearrangement from the first to the second line is admissible, because the series in the second line converge for every $i \in \{1,\dots,l\}$ and $j \in \mathbb{N}_0^d$ satisfying $j \le n$, as we show next. Using (5.37), the index shift $m - i \rightsquigarrow m$, and similar arguments as for (5.41), we get for these series

$$\begin{aligned} (**) &= \sum_{m=i}^{\infty}\mathbb{P}[S_{m-i} = n-j]\,\tilde q_{i,m-i} \\ &= \sum_{m=0}^{\infty}\mathbb{P}\bigl[S_m = n-j,\, \tilde N_i = m\bigr] = \mathbb{P}\bigl[\tilde S^{(i)} = n-j\bigr] = \tilde p_{i,n-j}. \end{aligned}$$

Substituting $(**)$ into $(*)$ and this result into (5.41) gives (5.38).

(b) Modifying the calculation in (5.41) 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

$$p_n = \sum_{m=1}^{\infty}\underbrace{\mathbb{P}[S_m = n,\, N = m]}_{=\,\mathbb{P}[S_m=n]\,\mathbb{P}[N=m]} = \sum_{i=1}^{l}\nu_i\underbrace{\sum_{m=1}^{\infty}\mathbb{P}[S_m = n]\,\mathbb{P}[\tilde N_i = m]}_{=\,\tilde p_{i,n}}$$

for every $n \in \mathbb{N}_0^d \setminus \{0\}$. □

The next corollary of Theorem 5.30(b) is useful, when only a k-truncation of a probability distribution is a Panjer(a, b, k) distribution but the first terms don't satisfy the recursion (5.17). It is the multivariate extension of [22, Corollary 4.7].

Corollary 5.32. *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 5.11. 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 (5.19) and*

$$p_n = \sum_{i=1}^{k-1} \mathbb{P}[S_i = n]\, q_i + \biggl(1 - \sum_{j=0}^{k-1} q_j\biggr)\tilde{p}_n, \quad n \in \mathbb{N}_0^d \setminus \{0\}.$$

Proof. Use Theorem 5.30(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 $n \ge k$, and all other $\tilde{q}_{i,n} = 0$. □

Exercise 5.33 (Combining Panjer's algorithm with convolutions and convex combinations)**.** Let Q denote a probability distribution of $\mathbb{N}_0^d$, and N an $\mathbb{N}_0$-valued random variable. Fix $k < l$ in $\mathbb{N}_0$ such that $\mathbb{P}[N \ge l] > 0$ and $\mathbb{P}[k \le N < l] > 0$. Assume that $\mathcal{L}(N|N \ge l) = \text{Panjer}(a, b, l)$ and that $\mathcal{L}(N|k \le N < l)$ is a k-truncated binomial distribution with $l-1$ trials and success probability $p \in [0, 1]$. Devise at least one algorithm to calculate Compound$(\mathcal{L}(N), Q)$ given by (4.70) and discuss numerical stability.

Hints: See Subsections 5.1 and 5.2, Theorem 5.30(b) and Remark 5.31 as well as Corollary 5.32.

5.4 Numerically Stable Algorithm for ExtNegBin

Remark 5.34. As noticed in Example 5.26, 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 (5.28). Choose $l \in \mathbb{N}$ with $l \ge 3$ and $\mathbb{P}[X_1 = 1] = \mathbb{P}[X_1 = l] = 1/2$ as one-dimensional loss distribution. Note that

$$p_k = \mathbb{P}[N = k,\, X_1 = \cdots = X_k = 1] = \frac{q_k}{2^k},$$

because $\mathbb{P}[N < k] = 0$, and correspondingly

$$\begin{aligned} p_{k+l-1} &= \sum_{j=1}^{k} \mathbb{P}[N = k,\, X_j = l,\, X_i = 1 \text{ for all } i \in \{1, \ldots, k\} \setminus \{j\}] \\ &\quad + \mathbb{P}[N = k + l - 1,\, X_1 = \cdots = X_{k+l-1} = 1] \\ &= \frac{kq_k}{2^k} + \frac{q_{k+l-1}}{2^{k+l-1}}. \end{aligned}$$

Recall from Example 5.26 that the frequency distribution ExtNegBin(α, k, p) is the Panjer$(p, (\alpha - 1)p, k)$ distribution. Note that S_k takes values in the set

$\{k+j(l-1) \mid j=0,\dots,k\}$, which does not contain $k+l$, hence the Panjer recursion formula (5.20) for p_{k+l} with $c_{k+l} := 1$ reduces to

$$p_{k+l} = \sum_{j=1}^{k+l} p\Big(1+\frac{\alpha-1}{k+l}j\Big)\,\mathbb{P}[X_1=j]\,p_{k+l-j}.$$

Since $\mathbb{P}[X_1=j]\neq 0$ only for $j\in\{1,l\}$, this simplifies to two summands, i.e.,

$$\begin{aligned} p_{k+l} &= p\Big(1+\frac{\alpha-1}{k+l}\Big)\frac{p_{k+l-1}}{2} + p\Big(1+\frac{\alpha-1}{k+l}l\Big)\frac{p_k}{2} \\ &= p\frac{k(l-1)+\varepsilon k}{k+l}\Big(\frac{q_k}{2^{k+1}}+\frac{q_{k+l-1}}{k2^{k+l}}\Big) - p\frac{k(l-1)-\varepsilon l}{k+l}\frac{q_k}{2^{k+1}}, \end{aligned}$$

hence severe cancellation occurs for p_{k+l} when ε 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 [22, Section 5.1], we now develop a numerically stable algorithm to compute the distribution of $(p_n)_{n\in\mathbb{N}_0^d}$ of $S=X_1+\cdots+X_N$, when N has an extended negative binomial distribution. The main ingredient is the following corollary of Theorem 5.30(a) for the case $l=1$ (we will omit the index 1 for simplicity).

Corollary 5.35. *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 $(q_n)_{n\in\mathbb{N}_0} := \mathrm{ExtNegBin}(\alpha-1,k+1,p)$ and*

$$(\tilde q_n)_{n\in\mathbb{N}_0} := \begin{cases} \mathrm{ExtNegBin}(\alpha,k,p) & \text{if } k\in\mathbb{N},\\ \mathrm{NegBin}(\alpha,p) & \text{if } k=0.\end{cases}$$

Then (5.36) 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

$$b = (\alpha-1)p\frac{q^{-\alpha}-\sum_{j=0}^{k-1}\binom{\alpha+j-1}{j}p^j}{q^{1-\alpha}-\sum_{j=0}^{k}\binom{\alpha+j-2}{j}p^j}, \tag{5.42}$$

hence (5.38) simplifies to the numerically stable weighted convolution

$$p_n = \frac{b}{n_i}\sum_{\substack{j\in\mathbb{N}_0^d\\ j\leq n,\, j_i>0}} j_i\,\mathbb{P}[X_1=j]\,\tilde p_{n-j}, \tag{5.43}$$

for every $n=(n_1,\dots,n_d)\in\mathbb{N}_0^d\setminus\{0\}$, where $i\in\{1,\dots,d\}$ is chosen such that $n_i\neq 0$. The initial value p_0 is given by (5.19) with probability-generationg function from (5.32) with parameters α and k replaced by $\alpha-1$ and $k+1$, respectively.

Proof. Using (5.28), we see that, for every $n \geq k+1$,

$$\binom{(\alpha-1)+n-1}{n} p^n = \frac{(\alpha-1)p}{n} \binom{\alpha+(n-1)-1}{n-1} p^{n-1},$$

hence $q_n = b\tilde{q}_{n-1}/n$ and Theorem 5.30(a) is applicable. $\square$

The case $k=0$, $p=1$ is excluded in the preceding corollary. We cannot reduce the calculation for a claim number $N \sim \mathrm{ExtNegBin}(\alpha-1, k+1, p)$ to the one for $N \sim \mathrm{ExtNegBin}(\alpha, k, p)$ in this case, because the negative binomial distribution is not defined for $p=1$. However, a suitable limit $p \nearrow 1$ gives the following numerically stable procedure.

Lemma 5.36 (Stable recursion for ExtNegBin$(\alpha-1,1,1)$)**.** *For $\alpha \in (0,1)$ consider a claim number $N \sim \mathrm{ExtNegBin}(\alpha-1,1,1)$. 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 $p_0 = 1 - (\mathbb{P}[X_1 \neq 0])^{1-\alpha}$ and*

$$p_n = \begin{cases} \frac{1-\alpha}{n_i} \sum_{j\in\mathbb{N}_0^d,\, 0\neq j\leq n} j_i\, \mathbb{P}[X_1 = j]\, r_{n-j} & \text{if } \mathbb{P}[X_1 \neq 0] > 0, \\ 0 & \text{if } \mathbb{P}[X_1 = 0] = 1, \end{cases}$$

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 $\mathbb{P}[X_1 \neq 0] > 0$ the non-negative sequence $(r_n)_{n\in\mathbb{N}_0^d}$ is defined by $r_0 = (\mathbb{P}[X_1 \neq 0])^{-\alpha}$ and recursively in a numerically stable way by

$$r_n = \frac{1}{n_i\, \mathbb{P}[X_1 \neq 0]} \sum_{\substack{j\in\mathbb{N}_0^d \\ 0\neq j\leq n}} (\alpha j_i + n_i - j_i)\, \mathbb{P}[X_1 = j]\, r_{n-j}$$

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 $\mathbb{P}[X_1 \neq 0] > 0$. We start with $p \in (0,1)$ and let $(\tilde{p}_n(p))_{n\in\mathbb{N}_0^d}$ denote the distribution of $\tilde{S} = X_1 + \cdots + X_{\tilde{N}}$, where $\tilde{N} \sim \mathrm{NegBin}(\alpha, p)$, and $(p_n(p))_{n\in\mathbb{N}_0^d}$ the distribution of $S = X_1 + \cdots + X_N$, where $N \sim \mathrm{ExtNegBin}(\alpha-1, 1, p)$. Since NegBin$(\alpha, p)$ is the Panjer$(p, (\alpha-1)p, 0)$ distribution, a recursion for the auxiliary sequence

$$r_n(p) := (1-p)^{-\alpha} \tilde{p}_n(p), \qquad n \in \mathbb{N}_0^d, \tag{5.44}$$

follows from Panjer's recursion (5.24) for $(\tilde{p}_n(p))_{n\in\mathbb{N}_0^d}$, namely

$$r_n(p) = \frac{p}{n_i(1 - p\,\mathbb{P}[X_1 = 0])} \sum_{\substack{j\in\mathbb{N}_0^d \\ 0\neq j\leq n}} (\alpha j_i + n_i - j_i)\, \mathbb{P}[X_1 = j]\, r_{n-j}(p) \tag{5.45}$$

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

$$r_0(p) = (1 - p\,\mathbb{P}[X_1 = 0])^{-\alpha} \tag{5.46}$$

given by (5.19) with probability-generating function from (5.23). The weighted convolution (5.43) becomes

$$p_n(p) = \frac{(1-p)^\alpha b(p)}{n_i} \sum_{\substack{j\in\mathbb{N}_0^d \\ j\le n,\, j_i>0}} j_i\,\mathbb{P}[X_1 = j]\,r_{n-j}(p) \tag{5.47}$$

for every $n = (n_1,\dots,n_d) \in \mathbb{N}_0^d \setminus \{0\}$ and $i \in \{1,\dots,d\}$ satisfying $n_i \neq 0$ and with $b(p) := (1-\alpha)p(1-p)^{-\alpha}/(1-(1-p)^{1-\alpha})$ from (5.42) and starting value

$$p_0(p) = \frac{1-(1-p\,\mathbb{P}[X_1=0])^{1-\alpha}}{1-(1-p)^{1-\alpha}} \tag{5.48}$$

given by (5.19) with probability-generating function from (5.33). The normalization in (5.44) is chosen so that we can take the limit $p \nearrow 1$ in (5.45)–(5.48), in particular $(1-p)^\alpha b(p)$ tends to $1-\alpha$. With $r_n := \lim_{p\nearrow 1} r_n(p)$ and $p_n := \lim_{p\nearrow 1} p_n(p)$, the lemma follows. □

Algorithm 5.37. Corollary 5.35 and Lemma 5.36 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 $N \sim \mathrm{ExtNegBin}(\alpha, k, p)$ 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 5.16 for $N \sim \mathrm{NegBin}(\alpha + k, p)$, followed by a stable weighted convolution according to Corollary 5.35 to pass to $N \sim \mathrm{ExtNegBin}(\alpha + k - 1, 1, p)$.
- If $p = 1$, use Lemma 5.36 to calculate the distribution of the compound sum S for $N \sim \mathrm{ExtNegBin}(\alpha + k - 1, 1, p)$.

Calculate $k - 1$ weighted convolutions according to (5.43) to pass iteratively to $N \sim \mathrm{ExtNegBin}(\alpha + k - 2, 2, p)$, …, and finally to $N \sim \mathrm{ExtNegBin}(\alpha, k, p)$.

Remark 5.38. Of course, compared to the ordinary (but possibly unstable) Panjer recursion of Theorem 5.16, Algorithm 5.37 increases the numerical effort by a factor of $k + 1$. Note that the weighted convolution in (5.43) is not a recurrence, hence unavoidable rounding errors do not propagate as in a recursive calculation.

5.5 Numerically Stable Algorithm for ExtLog

Similar results as in the previous subsection can be obtained for the extended logarithmic distribution.[42]

Corollary 5.39 ([22, Corollary 5.4])**.** *For the parameters $k \in \mathbb{N}$ and $p \in (0, 1]$ with $p < 1$ in case $k = 1$, let $(q_n)_{n\in\mathbb{N}_0} := \mathrm{ExtLog}(k + 1, p)$ and*

$$(\tilde{q}_n)_{n\in\mathbb{N}_0} := \begin{cases} \mathrm{ExtLog}(k,p) & \text{if } k \ge 2, \\ \mathrm{Log}(p) & \text{if } k = 1. \end{cases}$$

[42] The results of this subsection will not be used in the remaining part of the lecture notes.

Then (5.36) *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*

$$b = (k+1)p \frac{\sum_{l=k}^{\infty} \binom{l}{k}^{-1} p^l}{\sum_{l=k+1}^{\infty} \binom{l}{k+1}^{-1} p^l}$$

hence (5.38) *simplifies to the numerically stable weighted convolution* (5.43) *and* p_0 *is given by* (5.19)*.*

Exercise 5.40. Use Theorem 5.30(a) to prove Corollary 5.39.

In the excluded case $(k,p) = (1,1)$, we cannot reduce the calculation for $N \sim \mathrm{ExtLog}(2,p)$ to that for $N \sim \mathrm{Log}(p)$, 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 5.41 (Multi-dimensional version of [22, Lemma 5.5], stable recursion for $\mathrm{ExtLog}(2,1)$)**.** *Assume that* $N \sim \mathrm{ExtLog}(2,1)$*. 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*

$$p_0 = \mathbb{P}[X_1 = 0] + \mathbb{P}[X_1 \neq 0] \log \mathbb{P}[X_1 \neq 0]$$

with the convention $0 \log 0 = 0$*, and*

$$p_n = \begin{cases} \frac{1}{n_i} \sum_{j\in\mathbb{N}_0^d,\, 0<j\leq n} j_i \, \mathbb{P}[X_1 = j]\, r_{n-j} & \text{if } \mathbb{P}[X_1 \neq 0] > 0, \\ 0 & \text{if } \mathbb{P}[X_1 = 0] = 1, \end{cases}$$

for every $n = (n_1, \ldots, n_d) \in \mathbb{N}_0^d \setminus \{0\}$ *and* $i \in \{1, \ldots, d\}$ *satisfying* $n_i \neq 0$*, where for the case* $\mathbb{P}[X_1 \neq 0] > 0$ *the non-negative sequence* $(r_n)_{n\in\mathbb{N}_0^d}$ *is defined by* $r_0 = -\log \mathbb{P}[X_1 \neq 0]$ *and recursively in a numerically stable way by*

$$r_n = \frac{1}{\mathbb{P}[X_1 \neq 0]} \left(\mathbb{P}[X_1 = n] + \frac{1}{n_i} \sum_{\substack{j\in\mathbb{N}_0^d\setminus\{0\} \\ j<n,\, j_i<n_i}} (n_i - j_i)\, \mathbb{P}[X_1 = j]\, r_{n-j} \right)$$

for every $n = (n_1, \ldots, n_d) \in \mathbb{N}_0^d \setminus \{0\}$ *and* $i \in \{1, \ldots, d\}$ *satisfying* $n_i \neq 0$*.*

Exercise 5.42. Prove Lemma 5.41. Hints: For $p \in (0,1)$ consider $\tilde{N} \sim \mathrm{Log}(p)$, let $(\tilde{p}_n(p))_{n\in\mathbb{N}_0^d}$ denote the distribution of $\tilde{S} = X_1 + \cdots + X_{\tilde{N}}$, and let $(p_n(p))_{n\in\mathbb{N}_0^d}$ denote the distribution of $S = X_1 + \cdots + X_N$, where $N \sim \mathrm{ExtLog}(2,p)$. Define the auxiliary sequence

$$r_n(p) := -\tilde{p}_n(p) \log(1-p), \qquad n \in \mathbb{N}_0^d.$$

and proceed in a similar way as in the proof of Lemma 5.36. Consider the limit $p \nearrow 1$ at the end.

6 Stochastic Rounding and Copula-Based Aggregation

6.1 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 $E_1, \dots, E_d$ 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 6.1 (Stochastic rounding)**.** *Let $X = (X_1, \dots, X_d)$ be an $\mathbb{R}^d$-valued random vector. Define*

$$p_n = \mathbb{E}\Bigl[\prod_{i=1}^{d}(1 - |X_i - n_i|)^+\Bigr], \qquad n = (n_1, \dots, n_d) \in \mathbb{Z}^d, \tag{6.1}$$

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 $Y = (Y_1, \dots, Y_d)$ be a $\mathbb{Z}^d$-valued random vector with distribution $(p_n)_{n\in\mathbb{Z}^d}$ given by (6.1) and let I be a non-empty subset of $\{1, \dots, d\}$.

(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 $g_i\colon \mathbb{R} \to \mathbb{R}$ be a function which changes sign only at integers and which is piecewise affine between adjacent integers, i.e.

$$\lambda g_i(k) + (1-\lambda) g_i(k+1) = g_i\bigl(\lambda k + (1-\lambda)(k+1)\bigr) \tag{6.2}$$

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

$$\mathbb{E}\Bigl[\prod_{i\in I} g_i(X_i)\Bigr] = \mathbb{E}\Bigl[\prod_{i\in I} g_i(Y_i)\Bigr]. \tag{6.3}$$

Remark 6.2. Part (e) applied to $I = \{i\}$ with $i \in \{1, \dots, d\}$ 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 $I = \{i, j\} \subseteq \{1, \dots, d\}$ with $i \neq j$ and g_j also the identity function, we see that $\mathbb{E}[X_i X_j] = \mathbb{E}[Y_i Y_j]$, hence $\operatorname{Cov}(X_i, X_j) = \operatorname{Cov}(Y_i, Y_j)$, provided X_i, X_j and their product $X_i X_j$ are integrable.

Proof of Lemma 6.1. For each integer $k \in \mathbb{Z}$ define $f_k \colon \mathbb{R} \to [0, 1]$ by $f_k(x) = (1 - |x - k|)^+$ for all $x \in \mathbb{R}$. Note that $\prod_{i=1}^d f_{n_i}(X_i)$ coincides with the product in (6.1) for each $(n_1, \dots, n_d) \in \mathbb{Z}^d$. Let $g \colon \mathbb{R} \to \mathbb{R}$ be a function which is piecewise affine between adjacent integers, see (6.2). 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 (6.2) for the third equality,

$$\begin{aligned}\sum_{k \in \mathbb{Z}} f_k(x) g(k) &= f_{k_x}(x) g(k_x) + f_{k_x+1}(x) g(k_x + 1) \\ &= \underbrace{(1 - (x - k_x))}_{=:\,\lambda \,\in\, [0,1]} g(k_x) + \underbrace{(1 - (k_x + 1 - x))}_{=\,1-\lambda\,=\,x-k_x} g(k_x + 1) \\ &= g\big((1 - (x - k_x))k_x + (x - k_x)(k_x + 1)\big) = g(x).\end{aligned} \tag{6.4}$$

Note that no convergence problems arise on the left-hand side of (6.4), since at most two terms are different from zero. Using (6.4) for $g \equiv 1$, we see that $\{f_k\}_{k \in \mathbb{Z}}$ is a partition of unity, meaning in particular that

$$\sum_{k \in \mathbb{Z}} f_k(x) = 1, \qquad x \in \mathbb{R}. \tag{6.5}$$

(a) Using (6.5) for every dimension and expanding (keeping in mind that at most 2^d terms can be different from zero) leads to

$$\sum_{(n_1, \dots, n_d) \in \mathbb{Z}^d} \prod_{i=1}^d f_{n_i}(x_i) = \prod_{i=1}^d \sum_{n_i \in \mathbb{Z}} f_{n_i}(x_i) = 1, \qquad (x_1, \dots, x_d) \in \mathbb{R}^d.$$

Hence by monotone convergence,

$$\sum_{n \in \mathbb{Z}^d} p_n = \mathbb{E}\bigg[\sum_{n \in \mathbb{Z}^d} \prod_{i=1}^d f_{n_i}(X_i)\bigg] = 1.$$

(b) For every $n = (n_1, \dots, n_d) \in \mathbb{Z}^d \setminus \mathbb{N}_0^d$ there exists $i \in \{1, \dots, d\}$ 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 $J \coloneqq \{1, \dots, d\} \setminus I$. Using σ-additivity, monotone convergence and factoring,

$$\begin{aligned}\mathbb{P}[Y_i = n_i \text{ for all } i \in I] &= \sum_{(n_j)_{j \in J} \in \mathbb{Z}^J} \underbrace{\mathbb{P}[(Y_1, \dots, Y_d) = (n_1, \dots, n_d)]}_{=\,\mathbb{E}[\prod_{i=1}^d f_{n_i}(X_i)] \text{ by } (6.1)} \\ &= \mathbb{E}\bigg[\bigg(\prod_{i \in I} f_{n_i}(X_i)\bigg) \prod_{j \in J} \underbrace{\sum_{n_j \in \mathbb{Z}} f_{n_j}(X_j)}_{=\,1 \text{ by } (6.5)}\bigg].\end{aligned}$$

(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),

$$\mathbb{P}[Y_i = n_i \text{ for all } i \in I] = \mathbb{E}\Biggl[\prod_{i\in I} f_{n_i}(X_i)\Biggr] = \prod_{i\in I} \mathbb{E}[\,f_{n_i}(X_i)] = \prod_{i\in I} \mathbb{P}[Y_i = n_i]\,.$$

(e) Note that, if the functions g_i change sign only at integers, then the functions $\mathbb{R} \ni x \mapsto |g_i(x)|$ are also piecewise affine between adjacent integers, see (6.2), and (6.4) applies to them. Since all terms are non-negative, using the monotone convergence theorem,

$$\begin{aligned}\mathbb{E}\Biggl[\prod_{i\in I} |g_i(Y_i)|\Biggr] &= \sum_{(n_i)_{i\in I}\in\mathbb{Z}^I} \Biggl(\prod_{i\in I} |g_i(n_i)|\Biggr) \underbrace{\mathbb{P}[Y_i = n_i \text{ for all } i \in I]}_{=\mathbb{E}[\prod_{i\in I} f_{n_i}(X_i)] \text{ by part (c)}}\\ &= \sum_{(n_i)_{i\in I}\in\mathbb{Z}^I} \mathbb{E}\Biggl[\prod_{i\in I} f_{n_i}(X_i)\,|g_i(n_i)|\Biggr] = \mathbb{E}\Biggl[\prod_{i\in I} \underbrace{\sum_{n_i\in\mathbb{Z}} f_{n_i}(X_i)\,|g_i(n_i)|}_{=|g_i(X_i)| \text{ by (6.4)}}\Biggr],\end{aligned}$$

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 (6.3). □

Example 6.3 (Stochastic rounding can change the variance)**.** Consider a degenerate random variable X with $\mathbb{P}\bigl[X = \frac{1}{2}\bigr] = 1$, which has zero variance. Stochastic rounding produces the Bernolli distribution $\mathrm{Bin}(1, \frac{1}{2})$, which has variance $\frac{1}{4}$.

Example 6.4 (Stochastic rounding can change the correlation)**.** While Lemma 6.1(e) 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 $Z \sim \mathrm{Bin}(2, \frac{1}{2})$. Then $\mathrm{Var}(Z) = \frac{1}{2}$, hence $\mathrm{Cov}(X_1, X_2) = \frac{1}{4}\,\mathrm{Var}(Z) = \frac{1}{8}$. Since X_1 and X_2 are comonotone, or by noting that $\mathrm{Var}(X_1) = \mathrm{Var}(X_2) = \frac{1}{4}\,\mathrm{Var}(Z) = \frac{1}{8}$, it follows that $\mathrm{Corr}(X_1, X_2) = 1$. Stochastic rounding produces 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 $\mathrm{Cov}(Y_1, Y_2) = \frac{1}{8}$ by explicit calculation or an application of Lemma 6.1(e). Since $Y_1, Y_2 \sim \mathrm{Bin}(1, \frac{1}{2})$, it follows that $\mathrm{Var}(Y_1) = \mathrm{Var}(Y_2) = \frac{1}{4}$, hence $\mathrm{Corr}(Y_1, Y_2) = \frac{1}{2} \neq 1$.

Example 6.5 (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 $\mathrm{Cov}(X_1, X_2)$ is well defined, then Lemma 6.1(e) shows that $\mathrm{Cov}(X_1, X_2) = 0$ is a necessary condition for this phenomenon to occur. As an example, consider a random vector (X_1, X_2) taking with probability $\frac{1}{4}$ the four values $(1, 0)$, $(1, 1)$, $(\frac{1}{2}, \frac{1}{2})$ and $(\frac{3}{2}, \frac{1}{2})$, respectively, which are located on a square. The components X_1 and X_2 are clearly dependent, because

$$\mathbb{P}\bigl[X_1 = 1,\, X_2 = \tfrac{1}{2}\bigr] = 0 \neq \tfrac{1}{4} = \mathbb{P}[X_1 = 1]\,\mathbb{P}\bigl[X_2 = \tfrac{1}{2}\bigr]\,.$$

Stochastic rounding moves one quarter of the probability of $(\frac{1}{2}, \frac{1}{2})$ equally to each of its is four neighbouring lattice points in $\mathbb{Z}^2$, the same happens to the probability of $(\frac{3}{2}, \frac{1}{2})$. Hence $p_{(0,0)} = p_{(0,1)} = p_{(2,0)} = p_{(2,1)} = \frac{1}{16}$ 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 Introduction to Copulas

6.3 Aggregation of Integer-Valued Risks with Copula-Induced Dependency Structure

This subsection is based on joint work with Martin Schmidt contained in [48], see this reference for further details and illustrations.

Let $d \in \mathbb{N}$ with $d \geq 2$ and let $X_1, \dots, X_d$ denote $\mathbb{N}_0$-valued random variables such that $X_i \sim F_i$ for each $i \in \{1, \dots, d\}$ with given univariate distribution functions $F_1, \dots, F_d$. The random variables $X_1, \dots, X_d$ can represent claim sizes in an insurer's portfolio or credit losses in banking. The main objective of this subsection is to examine the distribution of the aggregated components, i.e., the sum $S \coloneqq X_1 + \dots + X_d$.

Recall that $\Delta_{d,\nu} = \{n \in \mathbb{N}_0^d \mid \|n\|_1 \leq \nu\}$ denotes the standard discrete d-dimensional simplex in $\mathbb{N}_0^d$ of size $\nu \in \mathbb{N}_0$ as given in (5.6). Let $\partial\Delta_{d,\nu} \coloneqq \{n \in \mathbb{N}_0^d \mid \|n\|_1 = \nu\}$ denote the discrete inner boundary of $\Delta_{d,\nu}$ in $\mathbb{N}_0^d$, i.e., the set of all $n \in \Delta_{d,\nu}$ such that there exists $x \in \mathbb{N}_0^d \setminus \Delta_{d,\nu}$ with $|n - x| = 1$.

Let us first consider the case where the random vector $X = (X_1, \dots, X_d)$ has independent components. Let $Q_i = (q_{i,n})_{n\in\mathbb{N}_0}$ denote the distribution of X_i, i.e. $q_{i,n} \coloneqq \mathbb{P}[X_i = n]$ for each $n \in \mathbb{N}_0$ and $i \in \{1, \dots, d\}$. Then the probability mass function of the sum S can be written down explicitly as

$$\mathbb{P}[S = \nu] = \sum_{n\in\partial\Delta_{d,\nu}} \underbrace{\mathbb{P}[X_1 = n_1, \dots, X_d = n_d]}_{=\prod_{i=1}^d q_{i,n_i}}, \quad \nu \in \mathbb{N}_0. \tag{6.6}$$

Starting with $d \geq 3$, it can be computationally more efficient to calculate the probability mass function of S using iterative convolutions $Q_1 * \dots * Q_d$, see Subsection 5.1 and Exercise 5.4. If $Q_1 = \dots = Q_d$, then Algorithm 5.2 and the recursion from Theorem 5.6 are available.

Suppose from now on that the dependence structure of the components of the random vector $X = (X_1, \dots, X_d)$ is given by a d-dimensional copula C, see Subsection 6.2. Let $\mathcal{I}_d \coloneqq \{0, 1\}^d$ denote the set of vertices of the d-dimensional unit hypercube and define $\operatorname{sign}(i) = (-1)^{i_1+\dots+i_d}$ for every $i = (i_1, \dots, i_d) \in \mathcal{I}_d$.

By the defining property of the copula C,

$$\mathbb{P}[A_n] = C\big(F_1(n_1), \dots, F_d(n_d)\big), \quad n = (n_1, \dots, n_d) \in \mathbb{N}_0^d,$$

with $A_n \coloneqq \{X_1 \leq n_1, \dots, X_d \leq n_d\}$. Defining $A_{j,n} = \{X_j \leq n_j - 1\} \cap A_n$ for each $j \in \{1, \dots, d\}$, it follows that

$$\{X = n\} = \{X_1 = n_1, \dots, X_d = n_d\} = A_n \setminus (A_{1,n} \cup \dots \cup A_{d,n}) \tag{6.7}$$

and by the inclusion–exclusion principle

$$\mathbb{P}[A_{1,n} \cup \cdots \cup A_{d,n}] = \sum_{(i_1,\dots,i_d)\in\mathcal{I}_d\setminus\{0\}} (-1)^{i_1+\cdots+i_d+1}\,\mathbb{P}\biggl[\bigcap_{\substack{j\in\{1,\dots,d\}\\ i_j=1}} A_{j,n}\biggr], \tag{6.8}$$

hence by (6.7) and (6.8), for every $n = (n_1,\dots,n_d) \in \mathbb{N}_0^d$,

$$\begin{aligned}\mathbb{P}[X = n] &= \mathbb{P}[A_n] - \mathbb{P}[A_{1,n} \cup \cdots \cup A_{d,n}] \\ &= \sum_{i\in\mathcal{I}_d} \operatorname{sign}(i)\,\underbrace{\mathbb{P}[X_1 \le n_1 - i_1,\dots,X_d \le n_d - i_d]}_{=C(F_1(n_1-i_1),\dots,F_d(n_d-i_d))}\end{aligned} \tag{6.9}$$

Since X takes values in $\mathbb{N}_0^d$, the defining property of the copula C implies that

$$C\bigl(F_1(n_1),\dots,F_d(n_d)\bigr) = \mathbb{P}[X \le n] = 0, \quad n = (n_1,\dots,n_d) \in \mathbb{Z}^d \setminus \mathbb{N}_0^d,$$

hence we can restrict the summation in (6.9) to all $i \in \mathcal{I}_d$ with $i \le n$. As a generalization of (6.6), the probability mass function of S is given by

$$\begin{aligned}\mathbb{P}[S = \nu] &= \sum_{n\in\partial\Delta_{d,\nu}} \mathbb{P}[X = n] \\ &= \sum_{n\in\partial\Delta_{d,\nu}} \sum_{\substack{i\in\mathcal{I}_d\\ i\le n}} \operatorname{sign}(i)\,C\bigl(F_1(n_1-i_1),\dots,F_d(n_d-i_d)\bigr), \quad \nu \in \mathbb{N}_0.\end{aligned} \tag{6.10}$$

Using that the discrete simplex satisfies $\Delta_{d,\nu} = \bigcup_{l=0}^{\nu} \partial\Delta_{d,l}$, where the union is disjoint, it follows from (6.10) that the distribution function of S is given by

$$\begin{aligned}\mathbb{P}[S \le \nu] &= \sum_{l=0}^{\nu} \mathbb{P}[S = l] \\ &= \sum_{n\in\Delta_{d,\nu}} \sum_{\substack{i\in\mathcal{I}_d\\ i\le n}} \operatorname{sign}(i)\,C\bigl(F_1(n_1-i_1),\dots,F_d(n_d-i_d)\bigr), \quad \nu \in \mathbb{N}_0.\end{aligned} \tag{6.11}$$

Using (5.7), it follows that the right-hand side of (6.11) without the restriction $i \le n$ has $2^d\binom{d+\nu}{d}$ terms to be summed up. Due to the inner summation over the elements of $\mathcal{I}_d$, most terms can appear up to 2^d times with positive or negative sign. This calls for a more efficient way to add the terms, which is given in the next lemma.

Lemma 6.6 (Distribution function of $S = X_1 + \cdots + X_d$)**.** *Define*

$$c_j = \sum_{n\in\partial\Delta_{d,j}} C\bigl(F_1(n_1),\dots,F_d(n_d)\bigr), \quad j \in \mathbb{N}_0. \tag{6.12}$$

Then

$$\mathbb{P}[S \le \nu] = \sum_{l=0}^{(d-1)\wedge\nu} (-1)^l \binom{d-1}{l} c_{\nu-l}, \quad \nu \in \mathbb{N}_0. \tag{6.13}$$

Remark 6.7. Compared to (6.10) and (6.11), each evaluation of the copula is only needed once for the calculation of the corresponding c_j via (6.12). For the calculation of the distribution function of S on $\{0,\dots,\nu\}$, the upper bound of $2^d\binom{d+\nu}{d}$ terms to be summed up via (6.11) is substantially reduced to $d\nu+\binom{d+\nu}{d}$ by the method from Lemma 6.6 (only the case $d\geq 2$ is of interest here). Note that due to the alternating sign in (6.13), cancellations can occur, but they don't propagate via a recursion.

The terms $C\big(F_1(n_1),\dots,F_d(n_d)\big)$ in (6.12) are computed without much effort when the distribution function C is given explicitly, examples are the Clayton copula, the Gumbel copula, and the Frank copula.

Proof of Lemma 6.6. Fix $\nu\in\mathbb{N}_0$. Note that $\mathcal{I}_d=\bigcup_{k=0}^d\mathcal{I}_{d,k}$ with $I_{d,k}:=\{i\in\mathcal{I}_d:\|i\|_1=k\}$ and $\Delta_{d,\nu}=\bigcup_{l=0}^{\nu}\partial\Delta_{d,l}$, where the unions are disjoint. Rewriting (6.11),

$$\begin{aligned}\mathbb{P}[S\leq\nu]&=\sum_{i\in\mathcal{I}_d}\operatorname{sign}(i)\sum_{\substack{n\in\Delta_{d,\nu}\\ i\leq n}}C\big(F_1(n_1-i_1),\dots,F_d(n_d-i_d)\big)\\&=\sum_{k=0}^{d\wedge\nu}(-1)^k\sum_{l=k}^{\nu}\sum_{i\in\mathcal{I}_{d,k}}\sum_{\substack{n\in\partial\Delta_{d,l}\\ i\leq n}}C\big(F_1(n_1-i_1),\dots,F_d(n_d-i_d)\big)\end{aligned}\tag{6.14}$$

Given $k\in\{0,\dots,d\wedge\nu\}$ and $l\in\{k,\dots,\nu\}$, note that for every pair $(i,n)\in\mathcal{I}_{d,k}\times\partial\Delta_{d,l}$ with $i\leq n$ there exists $\tilde{n}:=n-i$ such that $(i,\tilde{n})\in\mathcal{I}_{d,k}\times\partial\Delta_{d,l-k}$. For the reverse direction, given $(i,\tilde{n})\in\mathcal{I}_{d,k}\times\partial\Delta_{d,l-k}$, there exists $n:=\tilde{n}+i$ such that $(i,n)\in\mathcal{I}_{d,k}\times\partial\Delta_{d,l}$ and $i\leq n$. Due to this bijection, the last sum in (6.14) equals c_{l-k} given by (6.12). Since $|\mathcal{I}_{d,k}|=\binom{d}{k}$, (6.14) can be rewritten as

$$\mathbb{P}[S\leq\nu]=\sum_{k=0}^{d\wedge\nu}(-1)^k\binom{d}{k}\sum_{l=k}^{\nu}c_{l-k}=\sum_{k=0}^{d\wedge\nu}(-1)^k\binom{d}{k}\sum_{l=k}^{\nu}c_{\nu-l}=\sum_{l=0}^{\nu}c_{\nu-l}J_{d,l}\tag{6.15}$$

with

$$J_{d,l}:=\sum_{k=0}^{d\wedge l}(-1)^k\binom{d}{k},\qquad l\in\mathbb{N}_0,$$

where we used the index substitution $l\rightsquigarrow\nu-l+k$ for the second equality and interchanged the sums for the last one. We claim that

$$J_{d,l}=\begin{cases}(-1)^l\binom{d-1}{l} & \text{for } l\in\{0,\dots,d-1\},\\ 0 & \text{otherwise},\end{cases}\tag{6.16}$$

which follows inductively: By direct evaluation, $J_{d,0}=1$. For $l\in\{1,\dots,d\}$,

$$(-1)^lJ_{d,l}=\binom{d}{l}+(-1)^lJ_{d,l-1}=\binom{d}{l}-\binom{d-1}{l-1}=\frac{d-l}{l}\binom{d-1}{l-1}=\binom{d-1}{l}.$$

Hence $J_{d,d}=0$ and the second case in (6.16) follows. Substitution of (6.16) into (6.15) proves (6.13). □

The probability mass function of the sum S can be calculated via $\mathbb{P}[S = \nu] = \mathbb{P}[S \le \nu] - \mathbb{P}[S \le \nu - 1]$ for $\nu \in \mathbb{N}$ and (6.13) from Lemma 6.6. Using this idea, there is also a direct formula available, which looks very similar to (6.13):

Lemma 6.8 (Probability mass function of $S = X_1 + \cdots + X_d$)**.** *With* $(c_j)_{j\in\mathbb{N}_0}$ *given by* (6.12),

$$\mathbb{P}[S = \nu] = \sum_{l=0}^{d\wedge\nu} (-1)^l \binom{d}{l} c_{\nu-l}, \quad \nu \in \mathbb{N}_0. \tag{6.17}$$

Proof. For $\nu = 0$, (6.17) reduces to $\mathbb{P}[S = 0] = c_0$, which is correct. Fix $\nu \in \mathbb{N}$. We use (6.13) from Lemma 6.6 to see that

$$\mathbb{P}[S \le \nu] = c_\nu + \sum_{l=1}^{(d-1)\wedge\nu} (-1)^l \binom{d-1}{l} c_{\nu-l}$$

and, performing an index shift,

$$\mathbb{P}[S \le \nu - 1] = \sum_{l=0}^{(d-1)\wedge(\nu-1)} (-1)^l \binom{d-1}{l} c_{\nu-1-l} = -\sum_{l=1}^{d\wedge\nu} (-1)^l \binom{d-1}{l-1} c_{\nu-l}.$$

Using these two results and the recursive formula for binomial coefficients, which is the basis of Pascal's triangle,

$$\begin{aligned}\mathbb{P}[S = \nu] &= \mathbb{P}[S \le \nu] - \mathbb{P}[S \le \nu - 1] \\ &= c_\nu + \sum_{l=1}^{(d-1)\wedge\nu} (-1)^l \Biggl\{ \underbrace{\binom{d-1}{l} + \binom{d-1}{l-1}}_{=\binom{d}{l}} \Biggr\} c_{\nu-l} + (-1)^d c_{\nu-d} \mathbb{1}_{d\le\nu},\end{aligned}$$

where the last term is only present when $d \le \nu$ and can be included to the sum by changing the upper bound of the summation index to $d \wedge \nu$. Including also c_ν by starting the summation with $l = 0$, the claim (6.17) follows. □

7 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.

7.1 Introduction

With the tools developed in the previous chapters 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, see Remark 3.32.
- 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 7.1 (Multi-period extension)**.** The extension to several periods can be used in various ways and is also applicable in actuarial mathematics.

(a) *Several periods:* 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) *Immediate payments and actuarial reserves:* 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) *Profits and losses:* To aggregate profits as well as losses, that means $\mathbb{Z}$-valued random variables, we consider $\mathbb{N}_0^2$-valued random vectors whose components are the positive and negative parts, i.e. for each $\mathbb{Z}$-valued X we consider (X^+, X^-) and aim to determine the two-dimensional distribution of all profits and losses. Of course, netting can be done afterwards. In general, $\mathbb{Z}^d$-valued random vectors are converted into $\mathbb{N}_0^{2d}$-valued ones.

(d) *Different types of claim payments:* 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.

(e) *Stochastic claims reserving:* In the context of stochastic claims reserving (see e.g. [61] 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.

7.2 Description of the Model

We now assemble the necessary input parameters and the notation of the extended CreditRisk$^+$ methodology.

7.2.1 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 $R_1, \dots, R_K$,
- 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 $A_j = (a^j_{c,k})_{c\in\{0,\dots,C\},k\in\{0,\dots,K\}}$ of size $(C+1) \times (K+1)$ with non-negative entries, where

$$a^j_{0,k} = 0 \quad \text{for all } j \in \mathcal{J} \text{ and } k \in \{1, \dots, K\}, \tag{7.1}$$

- the collection G of nonempty subsets of all obligors $\{1, \dots, m\}$, 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 $c \in \{1, \dots, C\}$,
- 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 $E_1, \dots, E_d$ in case the risk group g defaults due to cause $c \in \{0, \dots, C\}$.

Assumption 7.2. *Every obligor $i \in \{1, \dots, m\}$ belongs to at least one group $g \in G$. Let $G_i \coloneqq \{g \in G \mid i \in g\}$ denote the set of all groups to which obligor $i \in \{1, \dots, m\}$ belongs, by assumption $G_i \neq \varnothing$.*

Remark 7.3. While Assumption 7.2 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 7.4. *For each group $g \in G$ and each scenario $j \in \mathcal{J}$, the susceptibilities (also called weights) exhaustively describe the default causes:*

$$\sum_{c=0}^{C} w_{c,g,j} = 1, \quad g \in G, j \in \mathcal{J}. \tag{7.2}$$

Remark 7.5. Assumption 7.4 is useful for the interpretation of the default probability p_g and the default intensity λ_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 7.30 below.

The idea of risk groups modelling joint defaults is motivated by the common Poisson shock models discussed by Lindskog and McNeil [37]. The idea to have different scenarios comes from [47], it originates from the desire to make negatively correlated default causes possible, see Example 7.33 below.

Remark 7.6 (Classical CreditRisk$^+$ model)**.** The classical CreditRisk$^+$ model is contained in the above set-up by choosing $G = \{\{1\}, \{2\}, \dots, \{m\}\}$, 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 $i \in \{1, \dots, m\}$ due to cause $c \in \{0, \dots, C\}$ in scenario $j \in \mathcal{J}$. Note also that in the classical CreditRisk$^+$ model there is just one scenario, i.e. $|\mathcal{J}| = 1$, one period, i.e. $d = 1$, and default causes and risk factors are identified, which corresponds to A^j 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 7.7 (Directly dependent defaults)**.** Suppose obligor $i \in \{1, \dots, m\}$ is a large factory and the different obligors $i_1, \dots, i_l \in \{1, \dots, m\}$ are suppliers of i, being economically heavily dependent on the factory. If the factory i defaults and is subsequently closed, the suppliers $i_1, \dots, i_l$ have a high probability to default, too. Therefore, $\{i, i_1, \dots, i_l\}$ 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 $\{i_1\}, \dots, \{i_l\} \in G$ makes sense, because every supplier can individually default due to poor management and subsequently be replaced by a competing supplier. Note that different loss distributions $Q_{c,g,j}$ of the $(\mathbb{N}_0^d)^g$-valued loss vectors given default due to cause $c \in \{0, \dots, C\}$ in scenario $j \in \mathcal{J}$ can be specified for the big risk group $g = \{i, i_1, \dots, i_l\}$ and for the individual obligors represented by $g = \{i\}$ and $g = \{i_1\}, \dots, \{i_l\}$.

Remark 7.8 (Hindering defaults, competition groups)**.** Suppose that the different obligors $i_1, \dots, i_l \in \{1, \dots, m\}$ 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 $g = \{i_1, \dots, i_l\}$ with a default probability p_g and choose the multivariate loss distribution $Q_{c,g,j} = (q_{c,g,j,\mu})_{\mu \in (\mathbb{N}_0^d)^g}$ in such a way that $q_{c,g,j,\mu} = 0$ for every integer vector $\mu = (\mu_{i_1}, \dots, \mu_{i_l})$ where two or more of the components $\mu_{i_1}, \dots, \mu_{i_l} \in \mathbb{N}_0^d$ representing the losses during the d periods are different from $0 \in \mathbb{N}_0^d$. This means in case of a default of risk group g due to cause $c \in \{0, \dots, C\}$ in scenario $j \in \mathcal{J}$, 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 7.9 (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 cycles, 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 $j \in \mathcal{J}$.

Remark 7.10 (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 a pandemic, 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.

7.2.2 Derived Parameters

The following quantities are derived from the input parameters:

- The Poisson intensity λ_g for defaults of group $g \in G$ during the d periods. As explained in Subsection 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 $(\mathbb{N}_0^d)^g$ of the loss during the d periods due to a default of group $g \in G$ caused by $c \in \{0, \dots, C\}$ in scenario $j \in \mathcal{J}$, the d-dimensional distribution $Q^{\mathrm{s}}_{c,g,j} = (q^{\mathrm{s}}_{c,g,j,\nu})_{\nu \in \mathbb{N}_0^d}$ 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

$$q^{\mathrm{s}}_{c,g,j,\nu} = \sum_{\substack{\mu = (\mu_i)_{i \in g} \in (\mathbb{N}_0^d)^g \\ \sum_{i \in g} \mu_i = \nu}} q_{c,g,j,\mu}\,, \qquad \nu \in \mathbb{N}_0^d, \tag{7.3}$$

see Remark 7.12 below. In the case $d = 1$, when $Q_{c,g,j}$ is specified by the loss distribution of every obligor $i \in g$ and a copula, see Subsection 6.3 for the computation of $Q^{\mathrm{s}}_{c,g,j}$.

- The cumulative Poisson intensity

$$\lambda_{j,k,\nu} \coloneqq \sum_{g \in G} \lambda_g \sum_{c=0}^{C} w_{c,g,j}\, a^j_{c,k}\, q^{\mathrm{s}}_{c,g,j,\nu} \geq 0 \tag{7.4}$$

in scenario $j \in \mathcal{J}$ for losses of size $\nu \in \mathbb{N}_0^d \setminus \{0\}$ in the portfolio due to idiosyncratic risk $k = 0$ or risk factor $k \in \{1, \dots, K\}$. In the last case, due to (7.1), the term for $c = 0$ can be omitted in (7.4).

- The set

$$\mathcal{S}_{j,k} \coloneqq \{\, \nu \in \mathbb{N}_0^d \setminus \{0\} \mid \lambda_{j,k,\nu} > 0\} \tag{7.5}$$

of all non-zero d-period exposure vectors with strictly positive intensity in scenario $j \in \mathcal{J}$ due to risk factor $k \in \{1, \dots, K\}$ in terms of the basic loss units $E_1, \dots, E_d$. This set is used in (7.74) and (7.83) below.

- The cumulative Poisson intensity for non-zero d-period loss vectors in the portfolio in scenario $j \in \mathcal{J}$ due to risk $k \in \{0, 1, \dots, K\}$, given by

$$\begin{aligned} \bar{\lambda}_{j,k} &\coloneqq \sum_{g \in G} \lambda_g \sum_{c=0}^{C} w_{c,g,j}\, a^j_{c,k} (1 - q^{\mathrm{s}}_{c,g,j,0}) \\ &= \sum_{\nu \in \mathbb{N}_0^d \setminus \{0\}} \sum_{g \in G} \lambda_g \sum_{c=0}^{C} w_{c,g,j}\, a^j_{c,k}\, q^{\mathrm{s}}_{c,g,j,\nu} = \sum_{\nu \in \mathcal{S}_{j,k}} \lambda_{j,k,\nu} \geq 0, \end{aligned} \tag{7.6}$$

where we used (7.4) and (7.5) for the last equality. Due to (7.5), $\bar{\lambda}_{j,k} = 0$ if and only if $\mathcal{S}_{j,k} = \varnothing$.

- If $\bar{\lambda}_{j,k} > 0$ for scenario $j \in \mathcal{J}$ and risk $k \in \{0, \dots, K\}$, then we can define the d-dimensional distribution $Q_{j,k} = (q_{j,k,\nu})_{\nu \in \mathbb{N}_0^d}$ by

$$q_{j,k,\nu} = \begin{cases} \lambda_{j,k,\nu} / \bar{\lambda}_{j,k} & \text{for all } \nu \in \mathbb{N}_0^d \setminus \{0\}, \\ 0 & \text{for } \nu = 0 \in \mathbb{N}_0^d. \end{cases} \tag{7.7}$$

It is a probability distribution due to (7.6). By (7.4) and (7.6), the distribution $Q_{j,k}$ is a mixture distribution of the family $\{Q^{\mathrm{s}}_{c,g,j} \mid c \in \{0, \dots, C\}, g \in G\}$, conditioned to be non-zero. If $\bar{\lambda}_{j,k} = 0$ for a scenario $j \in \mathcal{J}$ and a risk $k \in \{0, \dots, K\}$, then no non-zero d-period loss vector in this scenario due to this risk factor is possible and we define

$$q_{j,k,\nu} = \begin{cases} 0 & \text{for all } \nu \in \mathbb{N}_0^d \setminus \{0\}, \\ 1 & \text{for } \nu = 0 \in \mathbb{N}_0^d, \end{cases} \tag{7.8}$$

to avoid notational complications.

Note that the algorithm in Section 7.7 uses the intensities from (7.4) and (7.6), but not the default intensities of individual groups, not the individual susceptibilities, not the matrices A_j with $j \in \mathcal{J}$, 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 (see Lemma 8.33 below), the individual quantities are important.

7.2.3 Notation for the Number of Default Events

For every risk group $g \in G$ and every scenario $j \in \mathcal{J}$ 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 $c \in \{1, \dots, C\}$,
- $N_{g,j} \coloneqq \sum_{c=0}^{C} N_{c,g,j}$ for the total number of defaults.

For every obligor $i \in \{1, \dots, m\}$ and every scenario $j \in \mathcal{J}$ we write analogously

- $N_{0,i,j} \coloneqq \sum_{g \in G_i} N_{0,g,j}$ for the number of idiosyncratic defaults,
- $N_{c,i,j} \coloneqq \sum_{g \in G_i} N_{c,g,j}$ for the number of defaults caused by $c \in \{1, \dots, C\}$,
- $N_{i,j} \coloneqq \sum_{c=0}^{C} N_{c,i,j} = \sum_{g \in G_i} N_{g,j}$ 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 $\mathcal{J}$. Then

- $N_{c,g} \coloneqq N_{c,g,J} = \sum_{j \in \mathcal{J}} N_{c,g,j} \mathbb{1}_{\{J=j\}}$ is the number of defaults of group $g \in G$ due to cause $c \in \{0, \dots, C\}$,
- $N_g \coloneqq N_{g,J} \coloneqq \sum_{c=0}^{C} N_{c,g}$ describes the total number of defaults of risk group $g \in G$, and
- $N_i \coloneqq N_{i,J} = \sum_{j \in \mathcal{J}} N_{i,j} \mathbb{1}_{\{J=j\}}$ describes the total number of defaults of the individual obligor $i \in \{1, \dots, m\}$.

7.2.4 Notation for Stochastic Losses

Losses are $\mathbb{N}_0^d$-multiples of the basic loss units $E_1, \dots, E_d$. As in Subsection 7.2.3, let J be a random variable selecting the scenario from $\mathcal{J}$.

- 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 $j \in \mathcal{J}$ due to cause $c \in \{1, \dots, C\}$ 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 $j \in \mathcal{J}$ caused by $c \in \{1, \dots, C\}$ or due to idiosyncratic cause $c = 0$ is defined by
$$L_{c,g,j,n} = \sum_{i \in g} L_{c,g,i,j,n}. \tag{7.9}$$

- The $\mathbb{N}_0^d$-valued loss vector in scenario $j \in \mathcal{J}$ due to risk group $g \in G$ and cause $c \in \{1, \dots, C\}$ or idiosyncratic cause $c = 0$ is defined by

$$L_{c,g,j} = \sum_{n=1}^{N_{c,g,j}} L_{c,g,j,n}. \tag{7.10}$$

- The $\mathbb{N}_0^d$-valued loss vector due to risk group $g \in G$ and cause $c \in \{0, \dots, C\}$ is defined by

$$L_{c,g} = L_{c,g,J} = \sum_{j \in \mathcal{J}} L_{c,g,j} \mathbb{1}_{\{J=j\}}. \tag{7.11}$$

- The total $\mathbb{N}_0^d$-valued loss vector in scenario $j \in \mathcal{J}$ due to group $g \in G$ is given by

$$L_{g,j} \coloneqq \sum_{c=0}^{C} L_{c,g,j}. \tag{7.12}$$

- The total $\mathbb{N}_0^d$-valued loss vector in the portfolio in scenario $j \in \mathcal{J}$ is given by

$$L_j \coloneqq \sum_{g \in G} L_{g,j}. \tag{7.13}$$

- The total $\mathbb{N}_0^d$-valued loss vector in the portfolio is given by

$$L \coloneqq L_J = \sum_{j \in \mathcal{J}} L_j \mathbb{1}_{\{J=j\}}. \tag{7.14}$$

For the interpretation of the model and the calculation of risk contributions in Subsection 8.3 below, we will also need the following definitions of $\mathbb{N}_0^d$-valued loss vectors attributed to obligor $i \in \{1, \dots, m\}$:

- 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 $c \in \{0, \dots, C\}$ is given by

$$L_{c,g,i,j} \coloneqq \sum_{n=1}^{N_{c,g,j}} L_{c,g,i,j,n}. \tag{7.15}$$

- The attributed $\mathbb{N}_0^d$-valued loss vector in scenario $j \in \mathcal{J}$ due to cause $c \in \{0, \dots, C\}$ is given by the sum over all risk groups to which obligor i belongs, i.e.,

$$L_{c,i,j} \coloneqq \sum_{g \in G_i} L_{c,g,i,j}. \tag{7.16}$$

- 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.,

$$L_{i,j} \coloneqq \sum_{c=0}^{C} L_{c,i,j}. \tag{7.17}$$

- The total attributed $\mathbb{N}_0^d$-valued loss vector is given by the loss in the randomly selected scenario, i.e.,

$$L_i := L_{i,J} = \sum_{j\in\mathcal{J}} L_{i,j} \mathbb{1}_{\{J=j\}}. \tag{7.18}$$

7.3 Probabilistic Assumptions

The following assumptions are made:

Assumption 7.11 (Independence and distribution of group losses)**.** *For every group $g \in G$, every default cause $c \in \{0,\dots,C\}$ and every dependence scenario $j \in \mathcal{J}$, 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,*[43] *with distribution*

$$\mathbb{P}[L_{c,g,i,j,1} = \mu_i \text{ for all } i \in g] = q_{c,g,j,\mu}, \quad \mu = (\mu_i)_{i\in g} \in (\mathbb{N}_0^d)^g. \tag{7.19}$$

Remark 7.12. From Assumption 7.11 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 $j \in \mathcal{J}$ due to cause $c \in \{0,\dots,C\}$ defined in (7.9) is also i.i.d. with distribution $Q^{\mathrm{s}}_{c,g,j}$ given in (7.3). More explicitly, for all $n \in \mathbb{N}$ and $\nu \in \mathbb{N}_0^d$,

$$\begin{aligned} \mathbb{P}[L_{c,g,j,n} = \nu] &\overset{(7.9)}{=} \mathbb{P}\biggl[\sum_{i\in g} L_{c,g,i,j,n} = \nu\biggr] \\ &= \sum_{\substack{\mu=(\mu_i)_{i\in g}\in(\mathbb{N}_0^d)^g \\ \sum_{i\in g}\mu_i=\nu}} \underbrace{\mathbb{P}[L_{c,g,i,j,n} = \mu_i \text{ for all } i \in g]}_{=q_{c,g,j,\mu}} \\ &\overset{(7.3)}{=} q^{\mathrm{s}}_{c,g,j,\nu}. \end{aligned} \tag{7.20}$$

In some cases the distribution of the sum of the components is available in closed form. Examples are the multivariate Bernoulli distribution, the multinomial distribution, the multivariate logarithmic distribution, and the negative multinomial distribution, see (4.9), Exercise 4.20(a), Exercise 4.50(e), and Exercise 4.55(e), respectively.

Example 7.13 (Deterministic subdivision of a loss within a risk group)**.** Given a risk group $g \in G$ with at least two obligors, a scenario $j \in \mathcal{J}$ and a default cause $c \in \{0,\dots,C\}$, 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 $h_{c,g,i,j}\colon \mathbb{N}_0^d \to \mathbb{N}_0^d$ such that

$$\sum_{i\in g} h_{c,g,i,j}(\nu) = \nu, \qquad \text{for all } \nu \in \mathbb{N}_0^d. \tag{7.21}$$

[43] This means all other sequences of loss vectors, the scenario J, the idiosyncratic default numbers $(N_{0,g})_{g\in G}$ in Assumption 7.20, the non-idiosyncratic default numbers $(N_{c,g})_{c\in\{1,\dots,C\},g\in G}$ in Assumption 7.25 and the risk factors $R_1,\dots,R_K$ in Assumption 7.26 below.

We can then divide up the n^{th} group loss $L_{c,g,j,n} \sim Q^{\mathrm{s}}_{c,g,j}$ 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 (7.21), 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

$$q_{c,g,j,\mu} = \mathbb{P}[L_{c,g,i,j,n} = \mu_i \text{ for all } i \in g] = \begin{cases} q^{\mathrm{s}}_{c,g,j,\nu}, & \text{if } \mu = (h_{c,g,i,j}(\nu))_{i\in g}, \\ 0, & \text{otherwise,} \end{cases}$$

in particular the right-hand side of (7.3) 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 $g = \{i_1, \dots, i_l\}$ as uniform as possible over its members in a comonotone way, then

$$h_{c,g,i_k,j}(\nu) = \lfloor (\nu + k - 1)/l \rfloor, \qquad \text{for all } k \in \{1, \dots, l\} \text{ and } \nu \in \mathbb{N}_0, \tag{7.22}$$

is a possible choice.

Remark 7.14. Suppose that a risk group g has at least two members and that, for a specific default cause $c \in \{0, \dots, C\}$ 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 $Q^{\mathrm{s}}_{c,g,j}$ from (7.3) and (7.20) 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 7.15. For $d = 1$, we treat the case of comonotonic losses in Example 7.16, and the mixture of independent and comonotonic losses in Example 7.17. For copula-induced dependence, see Subsection 6.3. 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 7.15 (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 $c \in \{0, \dots, C\}$, 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 $Q_{c,g,j} = (q_{c,g,j,\mu})_{\mu\in(\mathbb{N}_0^d)^g}$ is given by

$$q_{c,g,j,\mu} = \mathbb{P}[L_{c,g,i,j,1} = \mu_i \text{ for all } i \in g] = \prod_{i\in g} \underbrace{\mathbb{P}[L_{c,g,i,j,1} = \mu_i]}_{= q_{c,g,i,j,\mu_i}} \tag{7.23}$$

for every $\mu = (\mu_i)_{i\in g} \in (\mathbb{N}_0^d)^g$. The distribution $Q^{\mathrm{s}}_{c,g,j} = (q^{\mathrm{s}}_{c,g,j,\nu})_{\nu\in\mathbb{N}_0^d}$ from (7.20) for the group loss is then the convolution of the $Q_{c,g,i,j}$ with $i \in g$, explicitly

$$q^{\mathrm{s}}_{c,g,j,\nu} = \sum_{\substack{\mu=(\mu_i)_{i\in g}\in(\mathbb{N}_0^d)^g \\ \sum_{i\in g}\mu_i=\nu}} \prod_{i\in g} q_{c,g,i,j,\mu_i}, \qquad \nu \in \mathbb{N}_0^d, \tag{7.24}$$

see also Subsection 5.1.

Example 7.16 (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 $c \in \{0, \dots, C\}$, 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 $L_{c,g,i,j,n} \sim Q_{c,g,i,j} = (q_{c,g,i,j,\nu})_{\nu \in \mathbb{N}_0}$ for every $i \in g$ and $n \in \mathbb{N}$. Let

$$F_{c,g,i,j}(\mu_i) = \sum_{\nu=0}^{\mu_i} q_{c,g,i,j,\nu}, \qquad \mu_i \in \mathbb{N}_0,$$

denote the discrete distribution function of $Q_{c,g,i,j}$ for $i \in g$. In this case the distribution $Q^{\mathrm{c}}_{c,g,j} = (q^{\mathrm{c}}_{c,g,j,\mu})_{\mu \in \mathbb{N}_0^g}$, where the superscript reminds of comonotonicity, with discrete distribution function

$$F_{c,g,j}(\mu) = \sum_{\substack{\nu \in \mathbb{N}_0^g \\ \nu \le \mu}} q^{\mathrm{c}}_{c,g,j,\nu}, \qquad \mu \in \mathbb{Z}^g,$$

of the group loss vector is given recursively by

$$q^{\mathrm{c}}_{c,g,j,\mu} = \min_{i \in g} F_{c,g,i,j}(\mu_i) - \max_{i \in g} F_{c,g,j}(\mu - e_i), \qquad \mu = (\mu_i)_{i \in g} \in \mathbb{N}_0^g, \tag{7.25}$$

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 $q^{\mathrm{c}}_{c,g,j,\mu_\nu} > 0$. Hence the distribution $Q^{\mathrm{s,c}}_{c,g,j} = (q^{\mathrm{s,c}}_{c,g,j,\nu})_{\nu \in \mathbb{N}_0}$, determined via (7.3), is in the comonotonic case given by

$$q^{\mathrm{s,c}}_{c,g,j,\nu} = \begin{cases} q^{\mathrm{c}}_{c,g,j,\mu_\nu} & \text{if } \mu_\nu \text{ exists,} \\ 0 & \text{otherwise,} \end{cases} \qquad \nu \in \mathbb{N}_0. \tag{7.26}$$

The discrete distribution function $F^{\mathrm{s,c}}_{c,g,j}$ corresponding to $Q^{\mathrm{s,c}}_{c,g,j}$ 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

$$F^{\mathrm{s,c}}_{c,g,j}(\nu) = \begin{cases} 0 & \text{for } \nu \in \{0, \dots, \nu_0 - 1\}, \\ \min_{i \in g} F_{c,g,i,j}(\nu_{i,0}) & \text{for } \nu = \nu_0. \end{cases}$$

For the recursion, assume that $(\nu_{i,n})_{i \in g} \in \mathbb{N}_0^g$ and $\nu_n = \sum_{i \in g} \nu_{i,n}$ as well as $F^{\mathrm{s,c}}_{c,g,j}$ on $\{0, \dots, \nu_n\}$ are given. If $F^{\mathrm{s,c}}_{c,g,j}(\nu_n) = 1$, then we can set $F^{\mathrm{s,c}}_{c,g,j}(\nu) = 1$ for all $\nu \in \mathbb{N}$ with $\nu > \nu_n$ and we are done. Otherwise, proceed as follows: Define for every $i \in g$

$$\nu_{i,n+1} = \begin{cases} \nu_{i,n} & \text{if } F_{c,g,i,j}(\nu_{i,n}) > F^{\mathrm{s,c}}_{c,g,j}(\nu_n), \\ \min\{\nu \in \mathbb{N}_0 \mid \nu > \nu_{i,n},\, q_{c,g,i,j,\nu} > 0\} & \text{otherwise,} \end{cases}$$

$\nu_{n+1} = \sum_{i \in g} \nu_{i,n+1}$, and correspondingly

$$F^{\mathrm{s,c}}_{c,g,j}(\nu) = \begin{cases} F^{\mathrm{s,c}}_{c,g,j}(\nu_n) & \text{for } \nu \in \{\nu_n + 1, \dots, \nu_{n+1} - 1\}, \\ \min_{i \in g} F_{c,g,i,j}(\nu_{i,n+1}) & \text{for } \nu = \nu_{n+1}. \end{cases}$$

Example 7.17 (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 $c \in \{0, \dots, C\}$, we can consider a mixture distribution of independent and 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. Specifically, choose an $\alpha_{c,g,j} \in [0,1]$ and define the mixed group loss distribution $Q^{\mathrm{m}}_{c,g,j} = (q^{\mathrm{m}}_{c,g,j,\mu})_{\mu \in \mathbb{N}_0^g}$ by

$$q^{\mathrm{m}}_{c,g,j,\mu} = \alpha_{c,g,j}\, q_{c,g,j,\mu} + (1 - \alpha_{c,g,j}) q^{\mathrm{c}}_{c,g,j,\mu}, \qquad \mu \in \mathbb{N}_0^g,$$

with $q_{c,g,j,\mu}$ given by (7.23) and $q^{\mathrm{c}}_{c,g,j,\mu}$ given by (7.25). The distribution of the sum of all the losses in the group is then

$$q^{\mathrm{s,m}}_{c,g,j,\nu} = \alpha_{c,g,j}\, q^{\mathrm{s}}_{c,g,j,\nu} + (1 - \alpha_{c,g,j}) q^{\mathrm{s,c}}_{c,g,j,\nu}, \qquad \nu \in \mathbb{N}_0,$$

with $q^{\mathrm{s}}_{c,g,j,\nu}$ given by (7.24) with $d = 1$ and $q^{\mathrm{s,c}}_{c,g,j,\nu}$ given by (7.26).

Remark 7.18 (Obligors with a credit guarantee[44])**.** Suppose a bank, a regional authority or a country, let's call it obligor $a \in \{1, \dots, m\}$, gives a credit guarantee to all obligors of a group $g \subseteq \{1, \dots, m\} \setminus \{a\}$ 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 $c \in \{0, \dots, C\}$ 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 $(\mathbb{N}_0^d)^{g(a)}$ describing the stochastic loss of all the obligors in $g(a)$ for scenario $j \in \mathcal{J}$ and default cause $c \in \{0, \dots, C\}$ is needed. The following practical problems come to mind:

[44] This remark needs an update based on the Master's thesis [19] of Lukas Fabrykowski.

- If g is large, think of $|g| \geq 100$, then $Q_{c,g(a),j}$ and the corresponding sum $Q^{\mathrm{s}}_{c,g(a),j}$ from (7.3) are computationally hard to calculate. A solution might be to make additional assumptions and apply the extended CreditRisk$^+$ methodology to calculate an approximation of $Q^{\mathrm{s}}_{c,g(a),j}$.
- 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 7.19 (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 λ_g, the susceptibility $w_{0,g,J}$ and the matrix entry $a^j_{0,0}$, i.e.,*

$$\mathcal{L}(N_{0,g}|J) = \mathrm{Poisson}\big(\lambda_g\, w_{0,g,J}\, a^J_{0,0}\big) \quad \textit{for every } g \in G. \tag{7.27}$$

Assumption 7.20 (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,*[45] *in particular*

$$\begin{aligned}\mathbb{P}[N_{0,g} = n_{0,g} \textit{ for all } g \in G\,|\,J] &= \prod_{g\in G} \mathbb{P}[N_{0,g} = n_{0,g}\,|\,J] \\ &= \prod_{g\in G} \mathrm{e}^{-\lambda_g w_{0,g,J} a^J_{0,0}} \frac{(\lambda_g\, w_{0,g,J}\, a^J_{0,0})^{n_{0,g}}}{n_{0,g}!}\end{aligned}$$

for all $n_{0,g} \in \mathbb{N}_0$, where we used (7.27) for the second equality.

Assumption 7.21 (Structure of default cause intensities)**.** *The default cause intensities $\Lambda_1, \ldots, \Lambda_C$ are expressed in terms of the random matrix $A_J = \sum_{j\in\mathcal{J}} A_j \mathbb{1}_{\{J=j\}}$ of size $(C+1)\times(K+1)$ and the non-negative risk factors $R_1, \ldots, R_K$ by*

$$\Lambda_c = a^J_{c,0} + \sum_{k=1}^{K} a^J_{c,k} R_k, \quad c \in \{1, \ldots, C\}. \tag{7.28}$$

Remark 7.22 (Lower bound for default cause intensity)**.** The scenario-dependent but otherwise constant term $a^J_{c,0} \geq 0$ in (7.28) is added so that a strictly positive lower bound for the default cause intensity Λ_c can be put into the model in addition to mathematically convenient distributions (like gamma distributions) for the risk factors $R_1, \ldots, R_K$.

Remark 7.23 (Constant risk factor R_0)**.** For notational convenience, we will sometimes use a constant 'risk factor' $R_0 \equiv 1$ and a scenario-dependent default

[45] This means the random loss vectors in Assumption 7.11, the non-idiosyncratic default numbers $(N_{c,g})_{c\in\{1,\ldots,C\},g\in G}$ in Assumption 7.25 and the risk factors $R_1, \ldots, R_K$ in Assumption 7.26 below.

cause intensity $\Lambda_0 = a^J_{0,0}$ for idiosyncratic risk, see (7.1), to write (7.28) in a more compact form or in matrix notation as

$$\Lambda = A_J R \tag{7.29}$$

with column random vectors $\Lambda = (\Lambda_0, \dots, \Lambda_C)^\mathsf{T}$ and $R = (R_0, \dots, R_K)^\mathsf{T}$.

Assumption 7.24 (Conditional distribution of non-idiosyncratic default numbers)**.** *For every default cause $c \in \{1, \dots, C\}$ and every group $g \in G$, the non-idiosyncratic default number $N_{c,g}$ is, conditioned on $J, R_1, \dots, R_K$, Poisson distributed with parameter given as product of the group default intensity λ_g, the susceptibility $w_{c,g,J}$, and the default cause intensity Λ_c, this means*

$$\begin{aligned}\mathbb{P}[N_{c,g} = n \,|\, J, R_1, \dots, R_K] &\stackrel{\text{a.s.}}{=} \mathbb{P}[N_{c,g} = n \,|\, J, \Lambda_c] \\ &\stackrel{\text{a.s.}}{=} \mathrm{e}^{-\lambda_g w_{c,g,J} \Lambda_c} \frac{(\lambda_g w_{c,g,J} \Lambda_c)^n}{n!}\end{aligned} \tag{7.30}$$

for all $n \in \mathbb{N}_0$, i.e.,

$$\mathcal{L}(N_{c,g} | J, R_1, \dots, R_K) \stackrel{\text{a.s.}}{=} \mathcal{L}(N_{c,g} | J, \Lambda_c) \stackrel{\text{a.s.}}{=} \text{Poisson}(\lambda_g w_{c,g,J} \Lambda_c)\,. \tag{7.31}$$

Assumption 7.25 (Conditional independence of non-idiosyncratic default numbers)**.** *Conditionally on $J, R_1, \dots, R_K$, the family*

$$\{N_{c,g} \,|\, c \in \{1, \dots, C\}, g \in G\}$$

of default numbers is independent, hence

$$\begin{aligned}&\mathbb{P}[N_{c,g} = n_{c,g} \textit{ for } c \in \{1, \dots, C\} \textit{ and } g \in G \,|\, J, R_1, \dots, R_K] \\ &\quad\stackrel{\text{a.s.}}{=} \prod_{c=1}^{C} \prod_{g \in G} \mathbb{P}[N_{c,g} = n_{c,g} \,|\, J, R_1, \dots, R_K] \\ &\quad\stackrel{\text{a.s.}}{=} \prod_{c=1}^{C} \prod_{g \in G} \mathrm{e}^{-\lambda_g w_{c,g,J} \Lambda_c} \frac{(\lambda_g w_{c,g,J} \Lambda_c)^{n_{c,g}}}{n_{c,g}!} \qquad \textit{by (7.30)}\end{aligned}$$

for all $n_{c,g} \in \mathbb{N}_0$.

Assumption 7.26 (Independence of risk factors and scenario)**.** *The non-negative risk factors $R_1, \dots, R_K$ and the scenario variable J are stochastically independent random variables.*

The independence of J and the risk factors $R_1, \dots, R_K$ is used for the algorithm in (7.76) below. It is also useful for calculating the moments and the covariances of the default cause intensities, as the following remark shows.

Remark 7.27 (Expectation, variance and covariance of default cause intensities)**.** If $R_1, \dots, R_K \in L^1(\mathbb{P})$ and Assumptions 7.21 and 7.26 hold, then

$$\mathbb{E}[\Lambda_c \,|\, J] \stackrel{(7.28)}{=} a^J_{c,0} + \sum_{k=1}^{K} a^J_{c,k}\, \mathbb{E}[R_k] \tag{7.32}$$

hence

$$\mathbb{E}[\Lambda_c] = \mathbb{E}\big[a^J_{c,0}\big] + \sum_{k=1}^{K} \mathbb{E}\big[a^J_{c,k}\big]\,\mathbb{E}[R_k] \tag{7.33}$$

for every $c \in \{1, \dots, C\}$. If, in addition, $R_1, \dots, R_K \in L^2(\mathbb{P})$, then, for all $c, d \in \{1, \dots, C\}$,

$$\operatorname{Cov}(\Lambda_c, \Lambda_d \,|\, J) \overset{(7.28)}{=} \sum_{k,l=1}^{K} a^J_{c,k} a^J_{d,l} \underbrace{\operatorname{Cov}(R_k, R_l)}_{= \delta_{k,l} \operatorname{Var}(R_k)} = \sum_{k=1}^{K} a^J_{c,k} a^J_{d,k} \operatorname{Var}(R_k)\,, \tag{7.34}$$

hence, by (3.65) from Lemma 3.50, it follows from (7.34) and (7.32) that

$$\begin{aligned} \operatorname{Cov}(\Lambda_c, \Lambda_d) &= \mathbb{E}\big[\operatorname{Cov}(\Lambda_c, \Lambda_d \,|\, J)\big] + \operatorname{Cov}\big(\mathbb{E}[\Lambda_c \,|\, J]\,, \mathbb{E}[\Lambda_d \,|\, J]\big) \\ &= \sum_{k=1}^{K} \mathbb{E}\big[a^J_{c,k} a^J_{d,k}\big] \operatorname{Var}(R_k) + \sum_{k,l=0}^{K} \operatorname{Cov}\big(a^J_{c,k}, a^J_{d,l}\big)\, e_k e_l \end{aligned} \tag{7.35}$$

with expectations $e_0 := 1$ and $e_k := \mathbb{E}[R_k]$ for $k \in \{1, \dots, K\}$.

Remark 7.28 (Pseudo risk factors)**.** Due to the independence of the risk factors $R_1, \dots, R_K$, see Assumption 7.26, it is not always possible to give them an economic interpretation. On the other hand, the distribution of the group losses, see Assumption 7.11, 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 $\Lambda_1, \dots, \Lambda_C$ as in (7.28) by economic considerations. A solution is the introduction of a random vector $P = (P_0, \dots, P_{K'})^\mathsf{T}$ of pseudo risk factors with an economic interpretation. Then a random matrix $A'_J = \sum_{j \in \mathcal{J}} A'_j \mathbb{1}_{\{J=j\}}$ 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 $\Lambda = (\Lambda_0, \dots, \Lambda_C)^\mathsf{T}$. The dependence of $P_0, \dots, P_{K'}$ can be specified by a random matrix $\tilde{A}_J = \sum_{j \in \mathcal{J}} \tilde{A}_j \mathbb{1}_{\{J=j\}}$ of size $(K'+1) \times (K+1)$ with non-negative entries such that $P = \tilde{A}_J R$, where $R = (R_0, \dots, R_K)^\mathsf{T}$ is the column vector of the independent risk factors. Then (7.29) is satisfied for the matrix product

$$A_J = A'_J \tilde{A}_J = \sum_{j \in \mathcal{J}} A'_j \tilde{A}_j \mathbb{1}_{\{J=j\}}. \tag{7.36}$$

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 (7.1); this is certainly the case if the corresponding entries of A'_j and $\tilde{A}_j$ satisfy (7.1).

Assumption 7.29 (Gamma-distributed risk factors)**.** *The risk factors $R_1, \dots,$ R_K are gamma distributed random variables with expectation $e_k := \mathbb{E}[R_k] > 0$ and variance $\sigma_k^2 := \operatorname{Var}(R_k) > 0$, 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 $k \in \{1, \dots, K\}$ by (4.55) and (4.56).*

Assumption 7.30 (Normalization of default causes)**.** *We assume that*

$$\mathbb{E}\biggl[w_{0,g,J}\,a_{0,0}^J+\sum_{c=1}^{C}w_{c,g,J}\,\Lambda_c\biggr]=1 \tag{7.37}$$

for every group $g\in G$.

Remark 7.31. Similar to Assumption 7.4, the preceding Assumption 7.30 is useful for the interpretation of the default probability p_g and the default intensity λ_g for every risk group $g\in G$, but the assumption is not necessary for the algorithm itself.

Remark 7.32 (Sufficient conditions for Assumption 7.30)**.** If $\mathbb{E}[R_k]=1$ for every risk $k\in\{1,\dots,K\}$ and $\mathbb{E}[A_J]$ is a stochastic matrix, then $\mathbb{E}[\Lambda_c]=1$ by (7.33) for every default cause $c\in\{1,\dots,C\}$. If the weights are deterministic, meaning that they do not depend on the scenario, then due to (7.1), which implies $\mathbb{E}[a_{0,0}^J]=1$ for the stochastic matrix $\mathbb{E}[A_J]$, and due to Assumption 7.4, the condition (7.37) is satisfied for every group $g\in G$.

7.4 Covariance Structure of Default Cause Intensities

The following example, which is based on [45, 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 7.33 (Negative correlation of default cause intensities)**.** Let J attain the values in $\mathcal{J}=\{0,1\}$ with strictly positive probability. Let R_1 and R_2 be two independent and gamma distributed random variables, independent of J, with $\mathbb{E}[R_1]=\mathbb{E}[R_2]=1$. Then Assumptions 7.26 and 7.29 are satisfied. Define

$$A_J=\begin{pmatrix}1&0&0\\0&\frac{J}{\mathbb{E}[J]}&0\\0&0&\frac{1-J}{1-\mathbb{E}[J]}\end{pmatrix}.$$

Then $\Lambda_1=JR_1/\mathbb{E}[J]$ and $\Lambda_2=(1-J)R_2/\mathbb{E}[1-J]$ by (7.28). Since $\mathbb{E}[A_J]=I_3$ is a stochastic matrix, $\mathbb{E}[\Lambda_1]=\mathbb{E}[\Lambda_2]=1$. If the weights do not depend on the scenario $j\in\{0,1\}$ and satisfy Assumption 7.4, then Assumption 7.30 is satisfied, see Remark 7.32. Since the product $\Lambda_1\Lambda_2$ contains the factor $J(1-J)\equiv 0$, we get $\Lambda_1\Lambda_2\equiv 0$ and

$$\operatorname{Cov}(\Lambda_1,\Lambda_2)=-\,\mathbb{E}[\Lambda_1]\,\mathbb{E}[\Lambda_2]=-1.$$

By direct computation using $\mathbb{E}[R_k^2]=\operatorname{Var}(R_k)+1$ for $k\in\{1,2\}$ or by (7.35),

$$\operatorname{Var}(\Lambda_1)=\frac{\operatorname{Var}(R_1)+1}{\mathbb{E}[J]}-1\qquad\text{and}\qquad\operatorname{Var}(\Lambda_2)=\frac{\operatorname{Var}(R_2)+1}{1-\mathbb{E}[J]}-1.$$

The correlation is therefore given by

$$\operatorname{Corr}(\Lambda_1,\Lambda_2)=\frac{\operatorname{Cov}(\Lambda_1,\Lambda_2)}{\sqrt{\operatorname{Var}(\Lambda_1)\operatorname{Var}(\Lambda_2)}}=-\frac{\sqrt{\mathbb{E}[J]\,\mathbb{E}[1-J]}}{\sqrt{\operatorname{Var}(R_1)+1-\mathbb{E}[J]}\sqrt{\operatorname{Var}(R_2)+\mathbb{E}[J]}},$$

which attains every value in $[-1, 0)$ if suitable values for $\operatorname{Var}(R_1)$ and $\operatorname{Var}(R_2)$ in $[0, \infty)$ are chosen. For the symmetric case $\mathbb{E}[J] = 1/2$ and $\operatorname{Var}(R_1) = \operatorname{Var}(R_2)$, this simplifies to

$$\operatorname{Corr}(\Lambda_1, \Lambda_2) = -\frac{1}{1 + 2\operatorname{Var}(R_1)}.$$

Example 7.33 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 7.34. A quadratic matrix Σ of size d with real entries is called *positive semidefinite*, if Σ is symmetric and $v^\mathsf{T}\Sigma v \geq 0$ for all $v \in \mathbb{R}^d$.

Remark 7.35. If a symmetric matrix Σ with real entries is not positive semidefinite, the R-command nearPD can be used to calculate a corresponding approximation.

Lemma 7.36. *(a) Let X be a square-integrable $\mathbb{R}^d$-valued random vector. Then its covariance matrix $\operatorname{Cov}(X, X)$ is positive semidefinite.*

(b) Let Σ be a positive semidefinite $d \times d$ matrix with real entries. Then there exists a square-integrable $\mathbb{R}^d$-valued random vector with $\operatorname{Cov}(X, X) = \Sigma$.

(c) Let $X = (X_1, \dots, X_d)^\mathsf{T}$ be a square-integrable $[0, \infty)^d$-valued random vector. Then $\operatorname{Cov}(X_i, X_j) \geq -\mathbb{E}[X_i]\,\mathbb{E}[X_j]$ for all $i, j \in \{1, \dots, d\}$ with $i \neq j$.

Let $\Sigma = (\Sigma_{i,j})_{i,j\in\{1,\dots,d\}}$ be a positive semidefinite matrix with real entries.

(d) For all $i, j \in \{1, \dots, d\}$,

$$\Sigma_{i,i} \geq 0 \qquad \text{and} \qquad |\Sigma_{i,j}| \leq \sqrt{\Sigma_{i,i}\Sigma_{j,j}}\,.$$

(e) Let A be a matrix of size $d \times k$ with real entries. Then $\Sigma' := A^\mathsf{T}\Sigma A$ is positive semidefinite.

(f) Assume that Σ satisfies $\Sigma = A\Sigma' A^\mathsf{T}$ with a matrix A of size $d \times k$ and a quadratic matrix Σ' of size k, both with real entries. If $A^\mathsf{T}A$ is invertible, then Σ' is positive semidefinite.

Remark 7.37. To see that the invertibility of $A^\mathsf{T}A$ in Lemma 7.36(f) is necessary, let all entries of A be zero. Then $\Sigma = A\Sigma' A^\mathsf{T}$ is the zero matrix and gives no information about the entries of Σ', in particular $\Sigma' = -I_k$ is possible.

Proof of Lemma 7.36. (a) Note that $\operatorname{Cov}(X, X)$ is symmetric and of size d with real entries. Consider X and $v \in \mathbb{R}^d$ as column vectors. Then

$$\begin{aligned} v^\mathsf{T}\operatorname{Cov}(X, X)\,v &= v^\mathsf{T}\,\mathbb{E}\big[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^\mathsf{T}\big]v \\ &= \mathbb{E}\big[v^\mathsf{T}(X - \mathbb{E}[X])\,\underbrace{(X - \mathbb{E}[X])^\mathsf{T}v}_{= v^\mathsf{T}(X - \mathbb{E}[X])}\big] \geq 0. \end{aligned}$$

(b) Let $\Sigma = LL^\top$ be the Cholesky decomposition of Σ, where L is a lower triangular matrix of size d with real entries. Let $Y = (Y_1, \dots, Y_d)^\top$ be any square-integrable random vector with independent components satisfying $\mathrm{Var}(Y_i) = 1$ for all $i \in \{1, \dots, d\}$ (like Y having a d-dimensional standard normal distribution). Then $\mathrm{Cov}(Y, Y) = I_d$ is the identity matrix of size d and $X := LY$ satisfies

$$\begin{aligned}\mathrm{Cov}(X, X) &= \mathbb{E}\big[(LY - \mathbb{E}[LY])(LY - \mathbb{E}[LY])^\top\big] \\ &= L\,\mathbb{E}\big[(Y - \mathbb{E}[Y])(Y - \mathbb{E}[Y])^\top\big]L^\top = L\,\mathrm{Cov}(Y, Y)\,L^\top = \Sigma.\end{aligned}$$

(c) $\mathrm{Cov}(X_i, X_j) = \mathbb{E}[X_iX_j] - \mathbb{E}[X_i]\,\mathbb{E}[X_j] \ge -\,\mathbb{E}[X_i]\,\mathbb{E}[X_j]$ because $X_iX_j \ge 0$.

(d) Let $X = (X_1, \dots, X_d)$ be a random vector according to (b). Then $\Sigma_{i,i} = \mathrm{Var}(X_i) \ge 0$ and, by the Cauchy–Schwarz inequality,

$$\begin{aligned}|\Sigma_{i,j}| = \big|\mathrm{Cov}(X_i, X_j)\big| &= \big|\mathbb{E}\big[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])\big]\big| \\ &\le \sqrt{\mathbb{E}[(X_i - \mathbb{E}[X_i])^2]}\sqrt{\mathbb{E}[(X_j - \mathbb{E}[X_j])^2]} = \sqrt{\Sigma_{i,i}\,\Sigma_{j,j}}\,.\end{aligned}$$

(e) Since $\Sigma^\top = \Sigma$, the matrix Σ' is symmetric, too. Furthermore, $v^\top\Sigma' v = (Av)^\top\Sigma(Av) \ge 0$ for every $v \in \mathbb{R}^k$. Hence Σ' is positive semidefinite.

(f) Note that $A\Sigma'A^\top = \Sigma$ implies $A^\top A\Sigma'A^\top A = A^\top\Sigma A$. Since $A^\top A$ is invertible with symmetric inverse, this implies $\Sigma' = B^\top\Sigma B$ with $B := A(A^\top A)^{-1}$. Hence (f) follows from part (e). □

Remark 7.38. While the Cholesky decomposition used in the proof of Lemma 7.36(b) always gives a lower triangular matrix L with non-negative diagonal entries, the example

$$\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$$

shows that L can have negative off-diagonal entries. Hence, if Y has independent gamma distributed components, the $X = LY$ as is the proof of Lemma 7.36(b) cannot always be used to model default cause intensities, because the components of X might attain negative values. Therefore, we need a more sophisticated approach.

Theorem 7.39. [46] *Let $\Sigma = (\Sigma_{i,j})_{i,j\in\{1,\dots,d\}}$ be a positive semidefinite matrix. Then there exist an integer $k \in \{1, \dots, d\}$ and independent random variables $J_2, \dots, J_d, X_{1,1}, \dots, X_{1,k}$, where $J_2, \dots, J_d$ take values in $\{0, 1\}$ and $X_{1,1}, \dots, X_{1,k}$ are non-negative and square-integrable, and random matrices $A_{J_2}, \dots, A_{J_d}$ with non-negative entries, where A_{J_i} is $\sigma(J_i)$-measurable for every $i \in \{2, \dots, d\}$, such that their sizes are non-decreasing and compatible such that the product $X_d := A_{J_d} \dots A_{J_2}X_1$ with $X_1 := (X_{1,1}, \dots, X_{1,k})^\top$ is well defined and satisfies $\mathrm{Cov}(X_d, X_d) = \Sigma$. In addition, $\mathbb{E}[A_{J_2}], \dots, \mathbb{E}[A_{J_d}]$ are sub-stochastic matrices (meaning that the entries in every row sum to at most 1).*

[46] This theorem is true, but the proof given here is probably incomplete. Time-permitting, the full proof will be copied into these lecture notes.

Remark 7.40 (Non-uniqueness of the representation)**.** Without further conditions, the representation in Theorem 7.39 is not unique. Already for $\Sigma = I_d$, where I_d denotes the identity matrix of size $d \geq 2$, there exist several solutions: Take $k = d$ and deterministic $A_{J_l} = P_{i_l, j_l}$ with $i_l, j_l \in \{1, \dots, d\}$ for $l \in \{2, \dots, d\}$, where $P_{i,j}$ denotes the matrix permuting rows i and j, with $P_{i,j} = I_d$ if $i = j$.

Proof of Theorem 7.39. We give a constructive, inductive proof of Theorem 7.39, 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 $X_{1,1}$ with $\operatorname{Var}(X_{1,1}) = \Sigma$.

Case 2: If $d \geq 2$ and Σ is a diagonal matrix with all diagonal elements different from zero, take $k = d$ and independent and non-negative $X_{1,1}, \dots, X_{d,d}$ with $\operatorname{Var}(X_{i,i}) = \Sigma_{i,i}$ for all $i \in \{1, \dots, d\}$. Furthermore, take degenerate random variables $J_2 = \dots = J_d \equiv 0$ and deterministic $A_{J_2} = \dots = A_{J_d} = I_d$.

Case 3: Suppose there exist different $i, j \in \{1, \dots, d\}$ with $\Sigma_{i,i} \geq \Sigma_{j,j}$ and $|\Sigma_{i,j}| = \sqrt{\Sigma_{i,i}\Sigma_{j,j}}$ (according the Lemma 7.36(d) this certainly happens if Σ has a diagonal entry which is zero). Define the permutation matrix

$$P = \begin{cases} P_{d-1,i} P_{d,j} & \text{if } i \neq d \text{ and } j \neq d-1, \\ P_{d-1,d} P_{d,i} P_{d-1,j} & \text{if } i = d \text{ or } j = d-1, \end{cases}$$

which moves row i to row $d-1$ and row j to row d, taking care of special cases. Then $P^{-1} = P^{\mathsf{T}}$, and $\Sigma' := P \Sigma P^{\mathsf{T}}$ satisfies $\Sigma = P^{\mathsf{T}} \Sigma' P$ as well as $\Sigma'_{d-1,d-1} \geq \Sigma'_{d,d}$ and $\Sigma'_{d-1,d} = f\,\Sigma'_{d-1,d-1}$ with factor

$$f := \begin{cases} 0 & \text{if } \Sigma'_{d-1,d-1} = 0, \\ \sqrt{\Sigma'_{d,d}/\Sigma'_{d-1,d-1}} & \text{if } \Sigma'_{d-1,d-1} > 0 \text{ and } \Sigma'_{d-1,d} \geq 0, \\ -\sqrt{\Sigma'_{d,d}/\Sigma'_{d-1,d-1}} & \text{if } \Sigma'_{d-1,d-1} > 0 \text{ and } \Sigma'_{d-1,d} < 0. \end{cases}$$

Note that $f \in [-1, 1]$ and $\Sigma'_{d,d} = f^2 \Sigma'_{d-1,d-1}$. We can partition Σ' as

$$\Sigma' = \begin{pmatrix} \Sigma'' & v \\ v^{\mathsf{T}} & \Sigma'_{d,d} \end{pmatrix}$$

with column vector $v = (v_1, \dots, v_{d-2}, \Sigma'_{d-1,d})^{\mathsf{T}}$. Let $u = (u_1, \dots, u_{d-2}, \Sigma'_{d-1,d-1})^{\mathsf{T}}$ denote the last column vector of Σ''. If $d = 2$, then $v = fu$. To prove by contradiction that $v = fu$ also for $d \geq 3$, assume that there exists an $i \in \{1, \dots, d-2\}$ with $v_i \neq f u_i$. Define $x = -(\Sigma'_{i,i} + 1)/(2fu_i - 2v_i)$ and $z = (0, \dots, 0, 1, 0, \dots, 0, fx, -x)^{\mathsf{T}} \in \mathbb{R}^d$ with the 1 in position i. Then

$$(\Sigma' z)_j = \begin{cases} \Sigma'_{i,j} + (fu_j - v_j)x & \text{for } j \in \{1, \dots, d-2\}, \\ u_i & \text{for } j = d-1, \\ v_i & \text{for } j = d, \end{cases}$$

and $z^{\mathsf{T}} \Sigma' z = \Sigma'_{i,i} + 2(fu_i - v_i)x = -1$, which is impossible for the positive semidefinite matrix Σ'. Due to $v = fu$ and $\Sigma'_{d,d} = f^2 \Sigma'_{d-1,d-1}$, it follows that

$$\Sigma' = \begin{pmatrix} I_{d-1} \\ w^{\mathsf{T}} \end{pmatrix} \Sigma'' \begin{pmatrix} I_{d-1} & w \end{pmatrix},$$

where $w = (0, \dots, 0, f)^\mathsf{T} \in \mathbb{R}^{d-1}$.

Case 3(a): Suppose that $f \geq 0$. Define $J_d \equiv 0$ and note that

$$\Sigma = A_{J_d} \Sigma'' A_{J_d}^\mathsf{T} \qquad \text{with} \qquad A_{J_d} := P^\mathsf{T} \begin{pmatrix} I_{d-1} \\ w^\mathsf{T} \end{pmatrix}.$$

Furthermore, note that A_{J_d} is a deterministic sub-stochastic matrix of size $d \times (d-1)$, which is stochastic if and only if $f = 1$. To verify that Σ'' is positive semidefinite, note that Σ'' is symmetric and that

$$A_{J_d}^\mathsf{T} A_{J_d} = \begin{pmatrix} I_{d-1} & w \end{pmatrix} \underbrace{P P^\mathsf{T}}_{= I_d} \begin{pmatrix} I_{d-1} \\ w^\mathsf{T} \end{pmatrix} = \begin{pmatrix} I_{d-2} & 0 \\ 0 & 1 + f^2 \end{pmatrix},$$

hence $A_{J_d}^\mathsf{T} A_{J_d}$ is an invertible diagonal matrix. Hence, Σ'' is positive semidefinite by Lemma 7.36(f) and the problem is reduced by one dimension and one risk factor.

Case 3(b): Suppose that $f < 0$.

Case 4: Take an $i \in \{1, \dots, d\}$ in the following order of priorities:

(a) All off–diagonal entries of Σ in row i are zero.

(b) All entries of Σ in row i are non-negative and the diagonal entry of every column $j \in \{1, \dots, d\} \setminus \{i\}$ with $\Sigma_{i,j} > 0$ satisfies $\Sigma_{j,j} \leq \Sigma_{i,i}$.

(c) For every $j \in \{1, \dots, d\} \setminus \{i\}$ the diagonal entry satisfies $\Sigma_{j,j} \leq \Sigma_{i,i}$.

By symmetry of Σ, the same is true for column i. We use the permutation matrix $P = P_{d,i}$ to exchange rows d and i (hence $P^{-1} = P^\mathsf{T} = P$) and represent $\Sigma = P \Sigma' P$ with $\Sigma' := P \Sigma P$. Note that P is a stochastic matrix and that Σ' is positive semidefinite by Lemma 7.36(f). Now the last row and the last column of Σ' have the property (a), (b) or (c), respectively. We write

$$\Sigma' = \begin{pmatrix} B & w \\ w^\mathsf{T} & c \end{pmatrix} = \begin{pmatrix} B & cu - cv \\ cu^\mathsf{T} - cv^\mathsf{T} & c \end{pmatrix}$$

with real square matrix B of size $d-1$, constant $c \in (0, \infty)$, and column vector $w = (w_1, \dots, w_{d-1})^\mathsf{T} \in \mathbb{R}^{d-1}$ decomposed componentwise into $u = \max\{w/c, 0\}$ and $v = \max\{0, -w/c\}$ in $[0, \infty)^{d-1}$. The matrix B is positive semidefinite by Lemma 7.36(e).

Case 4(a) Here $w = 0$ and Σ' 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 $k' \in \{1, \dots, d-1\}$, independent random variables $J_2, \dots, J_{d-1}$ and $X_{1,1}, \dots, X_{1,k'}$, and matrices $A'_{J_2}, \dots, A'_{J_{d-1}}$. 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 $\mathrm{Var}(X_{1,k}) = c$. Furthermore, define

$$A_{J_l} = \begin{pmatrix} A'_{J_l} & 0 \\ 0 & 1 \end{pmatrix}, \qquad l \in \{2, \dots, d-1\}. \tag{7.38}$$

Cases 4(b) and 4(c): Define the diagonal matrix $\tilde{D} = \operatorname{diag}(\tilde{D}_{1,1}, \dots, \tilde{D}_{d-1,d-1})$ with $\tilde{D}_{j,j} = 1 - u_j$ for every $j \in \{1, \dots, d-1\}$. Note that $|w_j| \le \sqrt{c\Sigma_{j,j}} \le c$ by Lemma 7.36(d) 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 $j \in \{1, \dots, d-1\}$, hence $\tilde{D}$ is invertible. Define

$$A = \begin{pmatrix} \tilde{D} & u \\ 0 & 1 \end{pmatrix} \qquad \text{and note that} \qquad A^{-1} = \begin{pmatrix} \tilde{D}^{-1} & -\tilde{D}^{-1}u \\ 0 & 1 \end{pmatrix}$$

and that A is a stochastic matrix. Hence we have the representation $\Sigma' = A\Sigma'' A^\mathsf{T}$ with

$$\begin{aligned} \Sigma'' &= A^{-1}\Sigma'(A^{-1})^\mathsf{T} \\ &= \begin{pmatrix} \tilde{D}^{-1} & -\tilde{D}^{-1}u \\ 0 & 1 \end{pmatrix} \begin{pmatrix} B & cu - cv \\ cu^\mathsf{T} - cv^\mathsf{T} & c \end{pmatrix} \begin{pmatrix} \tilde{D}^{-1} & 0 \\ -u^\mathsf{T}\tilde{D}^{-1} & 1 \end{pmatrix} \\ &= \begin{pmatrix} \tilde{D}^{-1}(B - cuu^\mathsf{T} + cuv^\mathsf{T}) & -c\tilde{D}^{-1}v \\ cu^\mathsf{T} - cv^\mathsf{T} & c \end{pmatrix} \begin{pmatrix} \tilde{D}^{-1} & 0 \\ -u^\mathsf{T}\tilde{D}^{-1} & 1 \end{pmatrix}. \end{aligned}$$

Defining $\tilde{B} := \tilde{D}^{-1}(B - cuu^\mathsf{T} + cuv^\mathsf{T} + cvu^\mathsf{T})\tilde{D}^{-1}$ and using that $\tilde{D}v = v$, hence $\tilde{D}^{-1}v = v$, it follows that

$$\Sigma'' = \begin{pmatrix} \tilde{B} & -cv \\ -cv^\mathsf{T} & c \end{pmatrix}.$$

By Lemma 7.36(f), the matrix Σ'' is positive semidefinite. By Lemma 7.36(e), the matrix $\tilde{B}$ is positive semidefinite, too.

Case 4(b): Here $v = 0$ and Σ'' 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 $k' \in \{1, \dots, d-1\}$, independent random variables $J_2, \dots, J_{d-1}$ and $X_{1,1}, \dots, X_{1,k'}$, and matrices $A'_{J_2}, \dots, A'_{J_{d-1}}$. Define $J_d \equiv 0$ and $k = k' + 1$ as well as the deterministic $A_{J_d} = PA$, and take any independent, non-negative random variable $X_{1,k}$ with $\operatorname{Var}(X_{1,k}) = c$. Furthermore, define by $A_{J_2}, \dots, A_{J_{d-1}}$ by (7.38).

Case 4(c): It remains to treat case (c) by introducing scenarios. Let $Y = (Y_1, \dots, Y_d)$ be a square-integrable random vector and define $e_d = \mathbb{E}[Y_d]$. Let J be $\{0, 1\}$-valued with $p := \mathbb{P}[J = 1] = c/(c + e_d^2) \in (0, 1)$. Consider

$$A_J = \begin{pmatrix} \tilde{C} & fJv \\ 0 & f(1-J) \end{pmatrix},$$

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

$$\Sigma''' := \begin{pmatrix} \tilde{C}^{-1}(\tilde{B} - cvv^\mathsf{T})(\tilde{C}^\mathsf{T})^{-1} & 0 \\ 0 & 0 \end{pmatrix},$$

which is symmetric because $(C^\intercal)^{-1} = (C^{-1})^\intercal$, it follows that

$$
\begin{aligned}
A_J \Sigma''' A_J^\intercal &= \begin{pmatrix} \tilde{C} & fJv \\ 0 & f(1-J) \end{pmatrix} \begin{pmatrix} \tilde{C}^{-1}(\tilde{B} - cvv^\intercal)(\tilde{C}^\intercal)^{-1} & 0 \\ 0 & 0 \end{pmatrix} A_J^\intercal \\
&= \begin{pmatrix} (\tilde{B} - cvv^\intercal)(\tilde{C}^\intercal)^{-1} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \tilde{C}^\intercal & 0 \\ fJv^\intercal & f(1-J) \end{pmatrix} = \begin{pmatrix} \tilde{B} - cvv^\intercal & 0 \\ 0 & 0 \end{pmatrix}.
\end{aligned}
$$

Then

$$
\begin{aligned}
\operatorname{Cov}(A_J\,\mathbb{E}[Y]\,, A_J\,\mathbb{E}[Y]) &= \operatorname{Cov}\biggl(\begin{pmatrix} v \\ -1 \end{pmatrix} J, \begin{pmatrix} v \\ -1 \end{pmatrix} J\biggr)\, e_d^2 f^2 \\
&= \begin{pmatrix} v \\ -1 \end{pmatrix} \begin{pmatrix} v^\intercal & -1 \end{pmatrix} e_d^2 f^2 \operatorname{Var}(J) = \begin{pmatrix} cvv^\intercal & -cv \\ -cv^\intercal & c \end{pmatrix},
\end{aligned}
$$

because $\operatorname{Var}(J) = p(1-p)$ and $e_d^2 f^2 \operatorname{Var}(J) = e_d^2 fp = c$. Therefore,

$$
\mathbb{E}\bigl[A_J \Sigma''' A_J^\intercal\bigr] + \operatorname{Cov}\bigl(A_J\,\mathbb{E}[Y]\,, A_J\,\mathbb{E}[Y]\bigr) = \Sigma''.
$$

Note that

$$
\mathbb{E}[A_J] = \begin{pmatrix} \tilde{C} & fpv \\ 0 & f(1-p) \end{pmatrix} = \begin{pmatrix} \tilde{C} & cv/e_d^2 \\ 0 & 1 \end{pmatrix},
$$

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 Σ'' is positive semidefinite and

$$
\begin{pmatrix} I_{d-1} & v \end{pmatrix} \underbrace{\begin{pmatrix} \tilde{B} & -cv \\ -cv^\intercal & c \end{pmatrix}}_{=\,\Sigma''} \begin{pmatrix} I_{d-1} \\ v^\intercal \end{pmatrix} = \begin{pmatrix} \tilde{B} - cvv^\intercal & 0 \end{pmatrix} \begin{pmatrix} I_{d-1} \\ v^\intercal \end{pmatrix} = \tilde{B} - cvv^\intercal,
$$

it follows from Lemma 7.36(e) and (f), that the matrices $\tilde{B} - cvv^\intercal$ and $\tilde{\Sigma} := \tilde{C}^{-1}(\tilde{B} - cvv^\intercal)(\tilde{C}^\intercal)^{-1}$ of size $d-1$ are also positive semidefinite, hence we can reduce the problem by one dimension. Applying the theorem to $\tilde{\Sigma}$ yields $k' \in \{1, \dots, d-1\}$, independent random variables $J_2, \dots, J_{d-1}$ and $X_{1,1}, \dots, X_{1,k'}$, and matrices $A'_{J_2}, \dots, A'_{J_{d-1}}$. Define $J_d = J$ and $k = k' + 1$ as well as the random matrix $A_{J_d} = PAA_J$, and take any independent, non-negative random variable $X_{1,k}$ with $\mathbb{E}[X_{1,k}] = e_d$ and $\operatorname{Var}(X_{1,k}) = c$. Furthermore, define by $A_{J_2}, \dots, A_{J_{d-1}}$ by (7.38). □

7.5 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 7.5.1, 7.5.2 and 7.5.3 are actually special cases of the results of Subsection 7.5.4, see Remark 7.46.

7.5.1 Expectation of Default Numbers

Let us start with the number of defaults

$$N_i = \sum_{g\in G_i} N_g = \sum_{g\in G_i}\sum_{c=0}^{C} N_{c,g} \tag{7.39}$$

of obligor $i \in \{1,\dots,m\}$. First note that by the conditional independence specified in Assumptions 7.20 and 7.25, as well as by the conditional Poisson distribution, see (7.27) and (7.31) in Assumptions 7.19 and 7.24, respectively, it follows with the Poisson summation property (3.5) that

$$\begin{aligned}\mathcal{L}(N_i|J,R_1,\dots,R_K) &\overset{\text{a.s.}}{=} \mathcal{L}\biggl(\sum_{g\in G_i}\biggl(N_{0,g}+\sum_{c=1}^{C}N_{c,g}\biggr)\biggm| J,R_1,\dots,R_K\biggr)\\ &\overset{\text{a.s.}}{=} \text{Poisson}(\underline{\Lambda}_i)\,.\end{aligned} \tag{7.40}$$

where

$$\underline{\Lambda}_i \coloneqq \sum_{g\in G_i}\lambda_g\biggl(w_{0,g,J}\,a^J_{0,0}+\sum_{c=1}^{C}w_{c,g,J}\,\Lambda_c\biggr) \tag{7.41}$$

is the conditional default intensity of obligor i, hence

$$\mathbb{E}[N_i\,|\,J,R_1,\dots,R_K] = \underline{\Lambda}_i \tag{7.42}$$

by (3.3). By inserting a conditional expectation given $J,R_1,\dots,R_K$, using (7.42) and the normalization (7.37) given in Assumption 7.30,

$$\mathbb{E}[N_i] = \mathbb{E}[\underline{\Lambda}_i] = \sum_{g\in G_i}\lambda_g\,\mathbb{E}\biggl[w_{0,g,J}\,a^J_{0,0}+\sum_{c=1}^{C}w_{c,g,J}\,\Lambda_c\biggr] = \sum_{g\in G_i}\lambda_g. \tag{7.43}$$

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 7.41 (Defaults with zero loss)**.** Note that (7.43) gives the expected number of defaults of obligor $i\in\{1,\dots,m\}$, but not every default has to lead to a credit loss, due to a sufficiently high collateral or deductable (in case of credit insurance). A corresponding remark applies to the results of Subsections 7.5.2 and 7.5.3 below.

Example 7.42. 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 (7.43) is satisfied, which shows that default intensities of risk groups with several obligors can in general not be derived from individual default intensities.

Remark 7.43. 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 λ_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

$$\lambda_{\{i\}} = \lambda_i - \sum_{\substack{g \in G_i \\ g \neq \{i\}}} \lambda_g, \qquad i \in \{1, \ldots, m\},$$

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.

7.5.2 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 $\operatorname{Var}(N_i \,|\, J, R_1, \ldots, R_K) \stackrel{\text{a.s.}}{=} \underline{\Lambda}_i$ by (7.40), (3.3) and (3.4). Using (3.66) from Lemma 3.50 and (7.42), we obtain

$$\operatorname{Var}(N_i) = \mathbb{E}\bigl[\underbrace{\operatorname{Var}(N_i \,|\, J, R_1, \ldots, R_K)}_{\stackrel{\text{a.s.}}{=} \underline{\Lambda}_i}\bigr] + \operatorname{Var}\bigl(\underbrace{\mathbb{E}[N_i \,|\, J, R_1, \ldots, R_K]}_{\stackrel{\text{a.s.}}{=} \underline{\Lambda}_i}\bigr), \tag{7.44}$$

which corresponds to (3.67). Using (7.43) and again (3.66) from Lemma 3.50, equation (7.44) turns into

$$\operatorname{Var}(N_i) = \mathbb{E}[N_i] + \mathbb{E}\bigl[\operatorname{Var}(\underline{\Lambda}_i \,|\, J)\bigr] + \operatorname{Var}\bigl(\mathbb{E}[\underline{\Lambda}_i \,|\, J]\bigr). \tag{7.45}$$

Note that $\operatorname{Var}(N_i) \geq \mathbb{E}[N_i]$, because variances are non-negative. Using Assumption 7.21 about the structure of the default cause intensities, it follows from (7.41) that

$$\underline{\Lambda}_i = \sum_{g \in G_i} \lambda_g \biggl(\sum_{c=0}^{C} w_{c,g,J}\, a_{c,0}^{J} + \sum_{k=1}^{K} R_k \sum_{c=1}^{C} w_{c,g,J}\, a_{c,k}^{J} \biggr).$$

Using Assumption 7.26 about the independence of $J, R_1, \ldots, R_K$,

$$\mathbb{E}[\underline{\Lambda}_i \,|\, J] = \sum_{g \in G_i} \lambda_g \biggl(\sum_{c=0}^{C} w_{c,g,J}\, a_{c,0}^{J} + \sum_{k=1}^{K} \mathbb{E}[R_k] \sum_{c=1}^{C} w_{c,g,J}\, a_{c,k}^{J} \biggr) \tag{7.46}$$

and

$$\operatorname{Var}(\underline{\Lambda}_i \,|\, J) = \sum_{k=1}^{K} \operatorname{Var}(R_k) \biggl(\sum_{g \in G_i} \lambda_g \sum_{c=1}^{C} w_{c,g,J}\, a_{c,k}^{J} \biggr)^2, \tag{7.47}$$

where $\mathbb{E}[R_k]$ and $\operatorname{Var}(R_k)$ are specified by Assumption 7.29.

If there is just one scenario, then J and therefore $\mathbb{E}[\underline{\Lambda}_i \,|\, J]$ are constant, hence the last term $\operatorname{Var}(\mathbb{E}[\underline{\Lambda}_i \,|\, J])$ in (7.45) is zero and $\operatorname{Var}(\underline{\Lambda}_i \,|\, J)$ from (7.47) coincides with the term $\mathbb{E}[\operatorname{Var}(\underline{\Lambda}_i \,|\, J)]$ in (7.45).

For the general case, note that $\operatorname{Var}(\mathbb{E}[\underline{\Lambda}_i\,|\,J]) = \mathbb{E}\big[\big(\mathbb{E}[\underline{\Lambda}_i\,|\,J]\big)^2\big] - (\mathbb{E}[\underline{\Lambda}_i])^2$ with $\mathbb{E}[\underline{\Lambda}_i]$ given by (7.43) and

$$\begin{aligned}
&\mathbb{E}\big[\big(\mathbb{E}[\underline{\Lambda}_i\,|\,J]\big)^2\big] \\
&\qquad = \sum_{j\in\mathcal{J}}\biggl(\sum_{g\in G_i}\lambda_g\biggl(\sum_{c=0}^{C} w_{c,g,j}\, a_{c,0}^{j} + \sum_{k=1}^{K}\mathbb{E}[R_k]\sum_{c=1}^{C} w_{c,g,j}\, a_{c,k}^{j}\biggr)\biggr)^2\,\mathbb{P}[J=j]\,.
\end{aligned}$$

Taking the expectation of (7.47) shows that

$$\mathbb{E}\big[\operatorname{Var}(\underline{\Lambda}_i\,|\,J)\big] = \sum_{k=1}^{K}\operatorname{Var}(R_k)\sum_{j\in\mathcal{J}}\biggl(\sum_{g\in G_i}\lambda_g\sum_{c=1}^{C} w_{c,g,j}\, a_{c,k}^{j}\biggr)^2\,\mathbb{P}[J=j]\,.$$

7.5.3 Covariances of Default Numbers

For obligors $i, i' \in \{1,\dots,m\}$ with $i \neq i'$ we can calculate the covariance of N_i and $N_{i'}$. By (3.65) from Lemma 3.50,

$$\operatorname{Cov}(N_i, N_{i'}) = \operatorname{Cov}\big(\mathbb{E}[N_i\,|\,J]\,, \mathbb{E}[N_{i'}\,|\,J]\big) + \mathbb{E}\big[\operatorname{Cov}(N_i, N_{i'}\,|\,J)\big] \tag{7.48}$$

Using (7.39), the linearity of conditional covariance in both arguments, Assumption 7.20 and (3.65) from Lemma 3.50, we obtain

$$\begin{aligned}
\operatorname{Cov}(N_i, N_{i'}\,|\,J) = {} & \sum_{g\in G_i\cap G_l}\underbrace{\operatorname{Var}(N_{0,g}\,|\,J)}_{=\lambda_g w_{0,g,J} a_{0,0}^{J} \text{ by Assumption 7.19 and (3.4)}} \\
& + \sum_{g\in G_i}\sum_{h\in G_{i'}}\sum_{c,d=1}^{C}\Bigl(\mathbb{E}\bigl[\underbrace{\operatorname{Cov}(N_{c,g}, N_{d,h}\,|\,J, R_1,\dots,R_K)}_{\overset{\text{a.s.}}{=}\operatorname{Var}(N_{c,g}|J,\Lambda_c)\overset{\text{a.s.}}{=}\lambda_g w_{c,g,J}\Lambda_c \text{ for } (c,g)=(d,h)}\bigm|J\bigr] \\
& \qquad + \operatorname{Cov}\bigl(\underbrace{\mathbb{E}[N_{c,g}\,|\,J, R_1,\dots,R_K]}_{\overset{\text{a.s.}}{=}\lambda_g w_{c,g,J}\Lambda_c}, \underbrace{\mathbb{E}[N_{d,h}\,|\,J, R_1,\dots,R_K]}_{\overset{\text{a.s.}}{=}\lambda_h w_{d,h,J}\Lambda_d}\bigm|J\bigr)\Bigr),
\end{aligned}$$

where we used Assumption 7.24, (3.3) and (3.4) to calculate the conditional expectations and the conditional variance. The conditional covariance of $N_{c,g}$ and $N_{d,h}$ given $J, R_1,\dots,R_K$ vanishes if $(g,k) \neq (h,l)$ due to conditional independence formulated in Assumption 7.25. Therefore,

$$\begin{aligned}
\operatorname{Cov}(N_i, N_{i'}\,|\,J) = {} & \sum_{g\in G_i\cap G_{i'}}\lambda_g\underbrace{\biggl(w_{0,g,J}\, a_{0,0}^{J} + \sum_{c=1}^{C} w_{c,g,J}\,\mathbb{E}[\Lambda_c\,|\,J]\biggr)}_{\mathbb{E}[\,\cdot\,]=1 \text{ by (7.37)}} \\
& + \sum_{g\in G_i}\lambda_g\sum_{h\in G_{i'}}\lambda_h\sum_{c,d=1}^{C} w_{c,g,J}\, w_{d,h,J}\operatorname{Cov}(\Lambda_c, \Lambda_d\,|\,J)\,,
\end{aligned} \tag{7.49}$$

where the remaining covariance is given by (7.34). Substituting (7.34) into (7.49), and the result into (7.48) yields

$$\begin{aligned}\operatorname{Cov}(N_i, N_{i'}) &= \operatorname{Cov}(\mathbb{E}[N_i\,|\,J]\,,\mathbb{E}[N_{i'}\,|\,J]) + \sum_{g\in G_i\cap G_{i'}} \lambda_g \\ &\quad + \sum_{g\in G_i} \lambda_g \sum_{h\in G_{i'}} \lambda_h \sum_{k=1}^{K} \operatorname{Var}(R_k) \sum_{c,d=1}^{C} \mathbb{E}\big[w_{c,g,J}\, w_{d,h,J}\, a_{c,k}^J\, a_{d,k}^J\big]\,,\end{aligned} \tag{7.50}$$

and it follows from (7.42) and (7.46) that

$$\mathbb{E}[N_i\,|\,J] = \mathbb{E}[\underline{\Lambda}_i\,|\,J] = \sum_{g\in G_i} \lambda_g \biggl(\sum_{c=0}^{C} w_{c,g,J}\, a_{c,0}^J + \sum_{k=1}^{K} \mathbb{E}[R_k] \sum_{c=1}^{C} w_{c,g,J}\, a_{c,k}^J\biggr)$$

and similarly for $\mathbb{E}[N_{i'}\,|\,J]$.

If there is just one scenario, then $\mathbb{E}[N_i\,|\,J]$ and $\mathbb{E}[N_{i'}\,|\,J]$ are deterministic and the covariance in (7.50) vanishes. Furthermore, there is no need to take the expectation on the right hand side of (7.50) and (omitting the J) we obtain

$$\begin{aligned}\operatorname{Cov}(N_i, N_{i'}) &= \sum_{g\in G_i\cap G_{i'}} \lambda_g \\ &\quad + \sum_{k=1}^{K} \operatorname{Var}(R_k) \biggl(\sum_{g\in G_i} \lambda_g \sum_{c=1}^{C} w_{c,g}\, a_{c,k}\biggr)\biggl(\sum_{h\in G_{i'}} \lambda_h \sum_{d=1}^{C} w_{d,h}\, a_{d,k}\biggr).\end{aligned} \tag{7.51}$$

Remark 7.44. In the classical CreditRisk$^+$ model (cf. Remark 7.6) with only one-element risk groups, the expectation in (7.43), the variance from Subsection 7.5.2, and the covariance given in (7.51) simplify to $\mathbb{E}[N_i] = \lambda_i$,

$$\operatorname{Var}(N_i) = \lambda_i + \lambda_i^2 \sum_{k=1}^{K} w_{k,i}^2 \operatorname{Var}(R_k)$$

and

$$\operatorname{Cov}(N_i, N_{i'}) = \lambda_i \lambda_j \sum_{k=1}^{K} w_{k,i} w_{k,i'} \operatorname{Var}(R_k)$$

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 (7.50), contributions to the covariance can also come from the risk groups in $G_i \cap G_{i'}$ and from the scenarios

7.5.4 Default Losses

[47] 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 $j \in \mathcal{J}$ due to default cause $c \in \{0, \dots, C\}$, as introduced in Subsection 7.2.4, 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 $i \in \{1, \dots, m\}$ given the scenario J and the risk factors $R_1, \dots, R_K$.

$$L_i = \sum_{c=0}^{C} \sum_{g \in G_i} L_{c,g,i,J}$$

By (7.16) and (7.18),

$$\mathbb{E}[L_i \,|\, J, R_1, \dots, R_K] \overset{\text{a.s.}}{=} \sum_{g \in G_i} \biggl(\mathbb{E}[L_{0,g,i,J} \,|\, J] + \sum_{c=1}^{C} \mathbb{E}[L_{c,g,i,J} \,|\, J, R_1, \dots, R_K] \biggr), \tag{7.52}$$

where we used Assumptions 7.11, 7.20, and (7.30) to simplify the conditional expectations. By Assumptions 7.11 and 7.19, the loss $L_{0,g,i,J}$ defined in (7.15) has a compound Poisson distribution and (4.106) implies that

$$\mathbb{E}[L_{0,g,i,J} \,|\, J] = \mathbb{E}[N_{0,g,J} \,|\, J]\, \mathbb{E}[L_{0,g,i,J,1} \,|\, J]\,. \tag{7.53}$$

By Assumptions 7.11 and 7.24, the loss $L_{g,i,k}$ due to risk factor $k \in \{1, \dots, K\}$ has a conditional compound Poisson distribution given Λ_k, hence by (4.104)

$$\mathbb{E}[L_{g,i,k} \,|\, \Lambda_k] \overset{\text{a.s.}}{=} \lambda_g w_{g,k} \Lambda_k\, \mathbb{E}[L_{g,i,k,1}]\,. \tag{7.54}$$

Substitution of (7.53) and (7.54) into (7.52) yields

$$\mathbb{E}[L_i \,|\, \Lambda_1, \dots, \Lambda_K] \overset{\text{a.s.}}{=} \sum_{g \in G_i} \lambda_g \biggl(w_{g,0}\, \mathbb{E}[L_{g,i,0,1}] + \sum_{k=1}^{K} w_{g,k} \Lambda_k\, \mathbb{E}[L_{g,i,k,1}] \biggr). \tag{7.55}$$

Since $\mathbb{E}[\Lambda_k] = 1$ by Assumption 7.29, we obtain

$$\mathbb{E}[L_i] = \sum_{g \in G_i} \lambda_g \sum_{k=0}^{K} w_{g,k}\, \mathbb{E}[L_{g,i,k,1}]\,. \tag{7.56}$$

Using (7.9) and (7.20), we get for the expected credit loss in the entire portfolio

$$\mathbb{E}[L] = \sum_{i=1}^{m} \mathbb{E}[L_i] = \sum_{g \in G} \lambda_g \sum_{k=0}^{K} w_{g,k} \underbrace{\mathbb{E}[L_{g,k,1}]}_{= \sum_{\nu \in \mathbb{N}} \nu q^{\mathrm{s}}_{g,k,\nu}}. \tag{7.57}$$

[47] This section has to be adapted to the new notation and the generalized setting.

Due to (7.2), the sums over the risks $k \in \{0, \dots, K\}$ in (7.56) and (7.57) are actually convex combinations.

The next step is to calculate the conditional covariance of the losses due to obligors $i, j \in \{1, \dots, m\}$ given the risk factors $\Lambda_1, \dots, \Lambda_K$. Considering $i = j$, this calculation will give the conditional variance. We first rewrite L_i and L_j using (7.16) and (7.18). We then note that, conditioned on the risk factors $\Lambda_1, \dots, \Lambda_K$, the family of random vectors $\bigl\{(L_{g,i',k})_{i' \in g} \mid g \in G, k \in \{0, \dots, K\}\bigr\}$ is independent by Assumptions 7.11, 7.20, and 7.25, hence

$$\begin{aligned}
&\mathrm{Cov}(L_i, L_j \,|\, \Lambda_1, \dots, \Lambda_K) \\
&\quad \overset{\text{a.s.}}{=} \sum_{g \in G_i} \sum_{h \in G_j} \sum_{k,l=0}^{K} \mathrm{Cov}(L_{g,i,k}, L_{h,j,l} \,|\, \Lambda_1, \dots, \Lambda_K) \\
&\quad \overset{\text{a.s.}}{=} \sum_{g \in G_i \cap G_j} \biggl(\mathrm{Cov}(L_{g,i,0}, L_{g,j,0}) + \sum_{k=1}^{K} \mathrm{Cov}(L_{g,i,k}, L_{g,j,k} \,|\, \Lambda_k) \biggr),
\end{aligned} \tag{7.58}$$

where we used Assumptions 7.11, 7.20 and (7.30) to simplify the conditional covariances. By Assumptions 7.11 and 7.19, the loss vector $(L_{g,i,0}, L_{g,j,0})$ with components defined in (7.15) has a compound Poisson distribution and (4.107) implies that

$$\mathrm{Cov}(L_{g,i,0}, L_{g,j,0}) = \lambda_g w_{g,0}\, \mathbb{E}[L_{g,i,0,1} L_{g,j,0,1}]\,. \tag{7.59}$$

By Assumptions 7.11 and 7.24, the loss vector $(L_{g,i,k}, L_{g,j,k})$ due to risk factor $k \in \{1, \dots, K\}$ has a conditional compound Poisson distribution given Λ_k, hence by (4.105)

$$\mathrm{Cov}(L_{g,i,k}, L_{g,j,k} \,|\, \Lambda_k) \overset{\text{a.s.}}{=} \lambda_g w_{g,k} \Lambda_k\, \mathbb{E}[L_{g,i,k,1} L_{g,j,k,1}]\,. \tag{7.60}$$

Substitution of (7.59) and (7.60) into (7.58) yields

$$\begin{aligned}
&\mathrm{Cov}(L_i, L_j \,|\, \Lambda_1, \dots, \Lambda_K) \\
&\quad \overset{\text{a.s.}}{=} \sum_{g \in G_i \cap G_j} \lambda_g \biggl(w_{g,0}\, \mathbb{E}[L_{g,i,0,1} L_{g,j,0,1}] + \sum_{k=1}^{K} w_{g,k} \Lambda_k\, \mathbb{E}[L_{g,i,k,1} L_{g,j,k,1}] \biggr).
\end{aligned} \tag{7.61}$$

To calculate the covariance of the credit losses due to obligors $i, j \in \{1, \dots, m\}$, we use (3.65), substitute (7.61) and (7.55), and use Assumption 7.29 to obtain

$$\begin{aligned}
\mathrm{Cov}(L_i, L_j) &= \mathbb{E}[\mathrm{Cov}(L_i, L_j \,|\, \Lambda_1, \dots, \Lambda_K)] \\
&\qquad + \mathrm{Cov}(\mathbb{E}[L_i \,|\, \Lambda_1, \dots, \Lambda_K], \mathbb{E}[L_i \,|\, \Lambda_1, \dots, \Lambda_K]) \\
&= \sum_{g \in G_i \cap G_j} \lambda_g \sum_{k=0}^{K} w_{g,k}\, \mathbb{E}[L_{g,i,k,1} L_{g,j,k,1}] \\
&\qquad + \sum_{k=1}^{K} \biggl(\sum_{g \in G_i} \lambda_g w_{g,k}\, \mathbb{E}[L_{g,i,k,1}] \biggr) \biggl(\sum_{g \in G_j} \lambda_g w_{g,k}\, \mathbb{E}[L_{g,j,k,1}] \biggr) \overbrace{\mathrm{Var}(\Lambda_k)}^{=\sigma_k^2}.
\end{aligned} \tag{7.62}$$

For $i = j$ this result simplifies to

$$\operatorname{Var}(L_i) = \sum_{g\in G_i} \lambda_g \sum_{k=0}^{K} w_{g,k}\,\mathbb{E}\big[L_{g,i,k,1}^2\big] + \sum_{k=1}^{K}\biggl(\sum_{g\in G_i} \lambda_g w_{g,k}\,\mathbb{E}[L_{g,i,k,1}]\biggr)^2 \sigma_k^2. \quad (7.63)$$

Remark 7.45. In the classical CreditRisk$^+$ model (cf. Remarks 7.6 and 7.44) with only one-element risk groups, the results (7.56), (7.63) and (7.62) simplify to

$$\mathbb{E}[L_i] = \lambda_i \sum_{k=0}^{K} w_{i,k}\,\mathbb{E}[L_{i,k,1}]\,, \quad (7.64)$$

$$\operatorname{Var}(L_i) = \lambda_i \sum_{k=0}^{K} w_{i,k}\,\mathbb{E}\big[L_{i,k,1}^2\big] + \lambda_i^2 \sum_{k=1}^{K} \sigma_k^2 w_{i,k}^2 \big(\mathbb{E}[L_{i,k,1}]\big)^2 \quad (7.65)$$

and

$$\operatorname{Cov}(L_i, L_j) = \lambda_i\lambda_j \sum_{k=1}^{K} \sigma_k^2 w_{i,k} w_{j,k}\,\mathbb{E}[L_{i,k,1}]\,\mathbb{E}[L_{j,k,1}]\,. \quad (7.66)$$

for all $i, j \in \{1, \dots, m\}$ with $i \neq j$, where we used the abbreviations $\lambda_i \coloneqq \lambda_{\{i\}}$ and $w_{i,k} \coloneqq w_{\{i\},k}$ as well as $L_{i,k,1} \coloneqq L_{\{i\},i,k,1}$ and corresponding ones for the index j.

Remark 7.46. To see that the results of Subsections 7.5.1, 7.5.2 and 7.5.3 are actually special cases of the results of Subsection 7.5.4, define $L_{g,i,k,n} = 1$ for all risk groups $g \in G$, risks $k \in \{0, 1, \dots, K\}$, obligors $i \in g$, and defaults $n \in \mathbb{N}$. Then (7.39) and (7.15)–(7.18) imply $N_i = L_i$ for all obligors $i \in \{1, \dots, m\}$. Comparison shows that the expectation in (7.56) simplifies to (7.43), the variance in (7.63) simplifies to (**??**), and the covariance in (7.62) simplifies to (7.50).

7.5.5 Default Numbers with Non-Zero Loss

[48] The default numbers considered in Subsections 7.5.1, 7.5.2 and 7.5.3 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 $L'_{g,i,k,n} \coloneqq \mathbb{1}_{\mathbb{N}}(L_{g,i,k,n})$ instead $L_{g,i,k,n}$.

Define for every obligor $i \in \{1, \dots, m\}$ the number L'_i of defaults with non-zero loss via (7.15), (7.16), and (7.18) using the just introduced $L'_{g,i,k,n}$. The results (7.56), (7.63) and (7.62) applied to L'_i and L'_j can easily be rewritten using

$$\mathbb{E}\big[(L'_{g,i,k,1})^2\big] = \mathbb{E}\big[L'_{g,i,k,1}\big] = \mathbb{P}[L_{g,i,k,1} > 0]$$

and

$$\mathbb{E}\big[L'_{g,i,k,1} L'_{g,j,k,1}\big] = \mathbb{P}[L_{g,i,k,1} > 0,\, L_{g,j,k,1} > 0]$$

[48] This section has to be adapted to the new notation and the generalized setting.

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.

7.6 Probability-Generating Function of the Biased Loss Vector

Fix a $\gamma = (\gamma_1, \ldots, \gamma_K) \in [0, \infty)^K$ such that $0 < \mathbb{E}\bigl[R_1^{\gamma_1} \ldots R_K^{\gamma_K}\bigr] < \infty$. In this section, using multi-index notation, we calculate the coefficients of the probability-generating function of the portfolio loss vector L under the $R_1^{\gamma_1} \ldots R_K^{\gamma_K}$-biased probability measure, given according to Definition 2.11, which we denote by $\mathbb{P}_\gamma$ for short. The corresponding expectation operator is denoted by $\mathbb{E}_\gamma$. Hence we want to calculate, for all $s \in \mathbb{C}^d$ with $\|s\|_\infty \le 1$,

$$\varphi_{L,\gamma}(s) := \sum_{\nu \in \mathbb{N}_0^d} \mathbb{P}_\gamma[L = \nu]\, s^\nu = \mathbb{E}_\gamma[s^L] = \frac{\mathbb{E}\bigl[\mathbb{E}[R_1^{\gamma_1} \ldots R_K^{\gamma_K} s^L \,|\, J]\bigr]}{\mathbb{E}\bigl[R_1^{\gamma_1} \ldots R_K^{\gamma_K}\bigr]} \tag{7.67}$$

of the $\mathbb{N}_0^d$-valued total loss vector L given by (7.14). For $\gamma = (0, \ldots, 0)$, we will obtain the usual probability-generating function φ_L of L. Let

$$L' = \sum_{c=1}^{C} \sum_{g \in G} L_{c,g} \tag{7.68}$$

denote the non-ideosycratic $\mathbb{N}_0^d$-valued portfolio loss vector. By Assumptions 7.11 and 7.20, the random vectors $(L_{0,g})_{g \in G}$ given by (7.11) 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

$$\mathbb{E}\bigl[R_1^{\gamma_1} \ldots R_K^{\gamma_K} s^L \,\big|\, J\bigr] = \mathbb{E}\bigl[R_1^{\gamma_1} \ldots R_K^{\gamma_K} s^{L'} \,\big|\, J\bigr] \prod_{g \in G} \mathbb{E}\bigl[s^{L_{0,g}} \,\big|\, J\bigr]. \tag{7.69}$$

By Assumptions 7.11, 7.19 and (4.77), it follows for the compound Poisson sum $L_{0,g,j}$, defined in (7.10), of idiosyncratic loss vectors of group $g \in G$ in scenario $j \in \mathcal{J}$, that

$$\mathbb{E}\bigl[s^{L_{0,g}} \,\big|\, J = j\bigr] = \exp\bigl(\lambda_g\, w_{0,g,j}\, a_{0,0}^j (\varphi_{L_{0,g,j,1}}(s) - 1)\bigr). \tag{7.70}$$

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 7.25, hence the random sums $(L_{c,g})_{c \in \{1,\ldots,C\}, g \in G}$ in (7.68), given by (7.10) and (7.11), are also conditionally independent due to Assumption 7.11. Therefore, we obtain

$$\begin{aligned} &\mathbb{E}\bigl[R_1^{\gamma_1} \ldots R_K^{\gamma_K} s^{L'} \,\big|\, J, R_1, \ldots, R_K\bigr] \\ &\quad\overset{\text{a.s.}}{=} R_1^{\gamma_1} \ldots R_K^{\gamma_K} \prod_{c=1}^{C} \prod_{g \in G} \mathbb{E}\bigl[s^{L_{c,g}} \,\big|\, J, R_1, \ldots, R_K\bigr] \end{aligned} \tag{7.71}$$

Due to Assumption 7.11, (7.31) in Assumption 7.24, the result (**??**) and Assumption 7.21, it follows that, for every default cause $c \in \{1, \dots, C\}$ and every group $g \in G$,

$$\begin{aligned}
&\mathbb{E}\big[s^{L_{c,g}} \,\big|\, J = j, R_1, \dots, R_K\big] \\
&\quad \overset{\text{a.s.}}{=} \mathbb{E}\big[s^{L_{c,g}} \,\big|\, J = j, \Lambda_c\big] \\
&\quad \overset{\text{a.s.}}{=} \exp\big(\lambda_g w_{c,g,j} \Lambda_c (\varphi_{L_{c,g,j,1}}(s) - 1)\big) \\
&\quad = \exp\Bigg(\lambda_g w_{c,g,j} \Bigg(a^j_{c,0} + \sum_{k=1}^{K} a^j_{c,k} R_k\Bigg)(\varphi_{L_{c,g,j,1}}(s) - 1)\Bigg).
\end{aligned} \tag{7.72}$$

Substitution of (7.70), (7.71) and (7.72) into (7.69) and rearrangement leads to

$$\begin{aligned}
&\mathbb{E}\big[R_1^{\gamma_1} \dots R_K^{\gamma_K} s^L \,|\, J = j, R_1, \dots, R_K\big] \\
&\quad \overset{\text{a.s.}}{=} \exp\Bigg(\sum_{g\in G} \lambda_g \sum_{c=0}^{C} w_{c,g,j}\, a^j_{c,0} (\varphi_{L_{c,g,j,1}}(s) - 1)\Bigg) \\
&\qquad \times \prod_{k=1}^{K} R_k^{\gamma_k} \exp\Bigg(R_k \sum_{g\in G} \lambda_g \sum_{c=1}^{C} w_{c,g,j}\, a^j_{c,k} (\varphi_{L_{c,g,j,1}}(s) - 1)\Bigg).
\end{aligned} \tag{7.73}$$

For every scenario $j \in \mathcal{J}$ and risk $k \in \{0, \dots, K\}$ let

$$\varphi_{j,k}(s) = \sum_{\nu \in \mathcal{S}_{j,k} \cup \{0\}} q_{j,k,\nu}\, s^\nu = \begin{cases} \bar\lambda_{j,k}^{-1} \sum_{\nu \in \mathcal{S}_{j,k}} \lambda_{j,k,\nu}\, s^\nu & \text{if } \bar\lambda_{j,k} > 0, \\ 1 & \text{if } \bar\lambda_{j,k} = 0, \end{cases} \tag{7.74}$$

at least for all $s \in \mathbb{C}^d$ with $\|s\|_\infty \le 1$, denote the probability-generating function of the distribution $Q_{j,k} = (q_{j,k,\nu})_{\nu \in \mathbb{N}_0^d}$ defined in (7.7) and (7.8), respectively, with the set $\mathcal{S}_{j,k}$ defined in (7.5). Recall that, for all default causes $c \in \{0, \dots, C\}$, groups $g \in G$ and scenarios $j \in \mathcal{J}$,

$$\varphi_{L_{c,g,j,1}}(s) = \sum_{\nu \in \mathbb{N}_0^d} s^\nu \underbrace{\mathbb{P}[L_{c,g,j,1} = \nu]}_{= q^{\mathrm{s}}_{c,g,j,\nu} \text{ by } (7.20)},$$

hence

$$\varphi_{L_{c,g,j,1}}(s) - 1 = \sum_{\nu \in \mathbb{N}_0^d \setminus \{0\}} s^\nu q^{\mathrm{s}}_{c,g,j,\nu} - (1 - q^{\mathrm{s}}_{c,g,j,0})$$

and rearrangement of the exponents on the right-hand side of (7.73) leads to

$$\begin{aligned}
&\sum_{g\in G} \lambda_g \sum_{c=0}^{C} w_{c,g,j}\, a^j_{c,k} (\varphi_{L_{c,g,j,1}}(s) - 1) \\
&\quad = \sum_{\nu \in \mathcal{S}_{j,k}} s^\nu \underbrace{\sum_{g\in G} \lambda_g \sum_{c=0}^{C} w_{c,g,j}\, a^j_{c,k}\, q^{\mathrm{s}}_{c,g,j,\nu}}_{= \lambda_{j,k,\nu} \text{ by } (7.4)} - \underbrace{\sum_{g\in G} \lambda_g \sum_{c=0}^{C} w_{c,g,j}\, a^j_{c,k} (1 - q^{\mathrm{s}}_{c,g,j,0})}_{= \bar\lambda_{j,k} \text{ by } (7.6)} \\
&\quad = \bar\lambda_{j,k} (\varphi_{j,k}(s) - 1) \quad \text{by } (7.74)
\end{aligned} \tag{7.75}$$

for every risk $k \in \{0, \dots, K\}$ with the set $\mathcal{S}_{j,k}$ defined in (7.5). Substituting (7.75) into of (7.73), using (7.1) in the case $k \in \{1, \dots, K\}$, taking the conditional expectation given J, and using the independence of $J, R_1, \dots, R_K$, it follows that

$$\begin{aligned}\mathbb{E}\bigl[R_1^{\gamma_1} \dots R_K^{\gamma_K} s^L \,\big|\, J = j\bigr] &= \exp\bigl(\bar{\lambda}_{j,0}(\varphi_{j,0}(s) - 1)\bigr) \\ &\quad \times \prod_{k=1}^{K} \mathbb{E}\bigl[R_k^{\gamma_k} \exp\bigl(\bar{\lambda}_{j,k}(\varphi_{j,k}(s) - 1)R_k\bigr) \,\big|\, J = j\bigr],\end{aligned} \tag{7.76}$$

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

To proceed further, we need to make an assumption on the distribution of the risk factors $R_1, \dots, R_K$.

7.6.1 Risk Factors with a Gamma Distribution

Since $R_k \sim \mathrm{Gamma}(\alpha_k, \beta_k)$ for every $k \in \{1, \dots, K\}$ by Assumption 7.29, and since R_k is independent of J, it follows from (4.58) that

$$\begin{aligned}&\mathbb{E}\bigl[R_k^{\gamma_k} \exp\bigl(\bar{\lambda}_{J,k}(\varphi_{J,k}(s) - 1)R_k\bigr) \,\big|\, J = j\bigr] \\ &\qquad = \mathbb{E}\bigl[R_k^{\gamma_k}\bigr] \left(1 - \bar{\lambda}_{j,k} \frac{\varphi_{j,k}(s) - 1}{\beta_k}\right)^{-(\alpha_k + \gamma_k)}.\end{aligned} \tag{7.77}$$

Substituting (7.77) into (7.76), we obtain

$$\begin{aligned}\mathbb{E}\bigl[R_1^{\gamma_1} \dots R_K^{\gamma_K} s^L \,|\, J = j\bigr] &= \exp\bigl(\bar{\lambda}_{j,0}(\varphi_{j,0}(s) - 1)\bigr) \\ &\quad \times \prod_{k=1}^{K} \mathbb{E}\bigl[R_k^{\gamma_k}\bigr] \left(1 - \bar{\lambda}_{j,k} \frac{\varphi_{j,k}(s) - 1}{\beta_k}\right)^{-(\alpha_k + \gamma_k)}.\end{aligned} \tag{7.78}$$

Transferring everything into a common exponential, we finally get for the probability-generating function under the $R_1^{\gamma_1} \dots R_K^{\gamma_K}$-biased probability measure, defined in (7.67),

$$\begin{aligned}\varphi_{L,\gamma}(s) &= \frac{1}{\mathbb{E}\bigl[R_1^{\gamma_1} \dots R_K^{\gamma_K}\bigr]} \sum_{j \in \mathcal{J}} \mathbb{E}\bigl[R_1^{\gamma_1} \dots R_K^{\gamma_K} s^L \,|\, J = j\bigr]\, \mathbb{P}[J = j] \\ &= \sum_{j \in \mathcal{J}} \exp\Biggl(\bar{\lambda}_{j,0}(\varphi_{j,0}(s) - 1) \\ &\qquad - \sum_{k=1}^{K} (\alpha_k + \gamma_k) \log\left(1 - \bar{\lambda}_{j,k} \frac{\varphi_{j,k}(s) - 1}{\beta_k}\right)\Biggr)\, \mathbb{P}[J = j],\end{aligned} \tag{7.79}$$

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

7.7 Algorithm for Risk Factors with a Gamma Distribution

Formula (7.79) is the probability-generating function of the accumulated $\mathbb{N}_0^d$-valued loss vector in the credit portfolio under the $R_1^{\gamma_1} \dots R_K^{\gamma_K}$-biased probability measure. From the definition (4.1) we know that the coefficients of the power series of (7.79) provide the desired distribution on $\mathbb{N}_0^d$. 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 $\sigma_k^2 = \operatorname{Var}(R_k)$ for all $k \in \{1, \dots, K\}$ using the formulas

$$\alpha_k = \frac{e_k^2}{\sigma_k^2} \qquad \text{and} \qquad \beta_k = \frac{e_k}{\sigma_k^2},$$

derived from (4.55) and (4.56).

Remark 7.47 (Historical remark)**.** The computation of these coefficients, however, can lead to numerical instabilities even in the one-period case with $(\gamma_1, \dots, \gamma_K) = 0$, cf. [25]. Therefore, this section describes an algorithm, basically due to G. Giese [25], for which Haaf, Reiß, Schoenmakers [28] proved the numerical stability. Apparently these authors didn't notice the relation to Panjer's recursion, see Theorem 5.16, which was pointed out in [22, Section 5.5]. The algebraic step of putting everything into a common exponential to pass from (7.78) to (7.79) reflects the fact that the negative binomial distribution is a compound Poisson distribution, where the severity distribution is a logarithmic one, see Example 4.40. Since Panjer's recursion is numerically stable for the Poisson as well as the logarithmic distribution, see Examples 5.21 and 5.25, respectively, numerical stability is guaranteed. The idea for the multi-period extension relies on the multivariate extension of Panjer's algorithm given by Sundt [51].

7.7.1 Expansion of the Logarithm by Panjer's Recursion

To calculate the coefficients of the power series of (7.79), we first treat the logarithmic term. For this purpose, fix a scenario $j \in \mathcal{J}$ and a risk factor $k \in \{1, \dots, K\}$. Define

$$p_{j,k} = \frac{\bar{\lambda}_{j,k}}{\beta_k + \bar{\lambda}_{j,k}} = \frac{\bar{\lambda}_{j,k}\sigma_k^2}{e_k + \bar{\lambda}_{j,k}\sigma_k^2} \in [0, 1) \tag{7.80}$$

with rate parameter $\beta_k > 0$, expectation $e_k > 0$ and variance σ_k^2 from Assumption 7.29 and $\bar{\lambda}_{j,k} \ge 0$ from (7.6). Note that the right-hand side of (7.80) works fine for the degenerate case $\sigma_k^2 = 0$.

We consider a random variable $M_{j,k} \sim \operatorname{Log}(p_{j,k})$. Let $(Y_{j,k,n})_{n \in \mathbb{N}}$ be an i.i.d. sequence of $\mathbb{N}_0^d$-valued random vectors, independent of $M_{j,k}$, with probability-generating function $\varphi_{j,k}$ defined in (7.74). Then by Example 4.4, in particular (4.6), and (4.75), the probability-generating function

$$\tilde{\varphi}_{j,k}(s) = \sum_{\nu \in \mathbb{N}_0^d} b_{j,k,\nu}\, s^\nu, \qquad s \in \mathbb{C}^d,\ \|s\|_\infty \le 1,$$

of the $\mathbb{N}_0^d$-valued random sum

$$S_{j,k} := \sum_{n=1}^{M_{j,k}} Y_{j,k,n}$$

is given by

$$\tilde{\varphi}_{j,k}(s) = \varphi_{j,k}(s)\frac{c(p_{j,k}\varphi_{j,k}(s))}{c(p_{j,k})}, \qquad s \in \mathbb{C}^d,\ \|s\|_\infty \le 1, \tag{7.81}$$

and its coefficients $(b_{j,k,\nu})_{\nu\in\mathbb{N}_0^d}$ can be computed in a numerically stable way by Panjer's recursion for the logarithmic distribution, see Example 5.25. More explicitly, using (4.6) and (5.26), the initial value is

$$b_{j,k,0} = q_{j,k,0}\frac{c(p_{j,k}q_{j,k,0})}{c(p_{j,k})}, \tag{7.82}$$

and, using (5.27), the recursion formula is, for every $\nu \in \mathbb{N}_0^d \setminus \{0\}$,

$$b_{j,k,\nu} = \frac{1}{1 - p_{j,k}q_{j,k,0}}\biggl(\frac{q_{j,k,\nu}}{c(p_{j,k})} + \frac{p_{j,k}}{\nu_i}\sum_{\substack{n\in\mathcal{S}_{j,k}\\ n<\nu,\, n_i<\nu_i}}(\nu_i - n_i)q_{j,k,n}b_{j,k,\nu-n}\biggr), \tag{7.83}$$

where $i \in \{1,\dots,d\}$ is chosen such that $\nu_i \neq 0$, and with $p_{j,k}$ given by (7.80), $(q_{j,k,\nu})_{\nu\in\mathbb{N}_0^d}$ given by (7.7), and $\mathcal{S}_{j,k}$ defined in (7.5). Note that γ_k does not enter into this recursion. If $p_{j,k} = 0$, then (7.82) and (7.83) simplify dramatically to $b_{j,k,\nu} = q_{j,k,\nu}$ for all $\nu \in \mathbb{N}_0^d$. To calculate the function c from (4.5) in a numerically stable way, see the corresponding comment in Example 4.4.

Rearranging and using (7.80) shows that

$$\begin{aligned} 1 - \bar{\lambda}_{j,k}\frac{\varphi_{j,k}(s) - 1}{\beta_k} &= \frac{\beta_k + \bar{\lambda}_{j,k}}{\beta_k}\biggl(1 - \frac{\bar{\lambda}_{j,k}}{\beta_k + \bar{\lambda}_{j,k}}\varphi_{j,k}(s)\biggr)\\ &= \frac{1}{1 - p_{j,k}}\bigl(1 - p_{j,k}\varphi_{j,k}(s)\bigr), \end{aligned}$$

hence using (4.5) and (7.81) the logarithmic term in (7.79) can be rewritten as

$$\begin{aligned} -\log\biggl(1 - \bar{\lambda}_{j,k}\frac{\varphi_{j,k}(s) - 1}{\beta_k}\biggr) &= -\log\bigl(1 - p_{j,k}\varphi_{j,k}(s)\bigr) + \log(1 - p_{j,k})\\ &= p_{j,k}\varphi_{j,k}(s)c(p_{j,k}\varphi_{j,k}(s)) - p_{j,k}c(p_{j,k})\\ &= p_{j,k}c(p_{j,k})\bigl(\tilde{\varphi}_{j,k}(s) - 1\bigr). \end{aligned} \tag{7.84}$$

Substituting (7.84) into (7.79) gives

$$\begin{aligned} \varphi_{L,\gamma}(s) = \sum_{j\in\mathcal{J}}\exp\biggl(&\bar{\lambda}_{j,0}(\varphi_{j,0}(s) - 1)\\ &+ \sum_{k=1}^{K}(\alpha_k + \gamma_k)p_{j,k}c(p_{j,k})\bigl(\tilde{\varphi}_{j,k}(s) - 1\bigr)\biggr)\,\mathbb{P}[J = j]\,, \end{aligned} \tag{7.85}$$

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

7.7.2 Expansion of the Exponential by Panjer's Recursion

To calculate the coefficients of the power series of (7.85), we first rewrite the argument of the exponential function. Define

$$\lambda_j = \bar{\lambda}_{j,0} + \sum_{k=1}^{K} \underbrace{\bar{\lambda}_{j,k} \frac{e_k^2 + \gamma_k \sigma_k^2}{e_k + \bar{\lambda}_{j,k} \sigma_k^2}}_{= (\alpha_k + \gamma_k) p_{j,k}} c(p_{j,k}), \qquad j \in \mathcal{J}, \tag{7.86}$$

with the shape parameter $\alpha_k > 0$ expectation $e_k > 0$ and variance σ_k^2 given in Assumption 7.29, Poisson intensity $\bar{\lambda}_{j,0} \geq 0$ given in (7.6), parameter $p_{j,k} \in [0, 1)$ of the logarithmic distribution given in (7.80), and function c defined in (4.5). Note that only non-negative terms are added in (7.86) and that its right-hand side even works in the degenerate case $\sigma_k^2 = 0$, both facts guarantee numerical stability. For every $j \in \mathcal{J}$ with $\lambda_j > 0$, we define

$$\tilde{\varphi}_j(s) = \frac{1}{\lambda_j} \biggl(\bar{\lambda}_{j,0} \varphi_{j,0}(s) + \sum_{k=1}^{K} \bar{\lambda}_{j,k} \frac{e_k^2 + \gamma_k \sigma_k^2}{e_k + \bar{\lambda}_{j,k} \sigma_k^2} c(p_{j,k}) \tilde{\varphi}_{j,k}(s) \biggr),$$

at least for all $s \in \mathbb{C}^d$ with $\|s\|_\infty \leq 1$. Note that the coefficients of the power series

$$\tilde{\varphi}_j(s) = \sum_{\nu \in \mathbb{N}_0} c_{j,\nu} s^\nu, \qquad s \in \mathbb{C}^d, \ \|s\|_\infty \leq 1,$$

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,

$$c_{j,\nu} = \frac{1}{\lambda_j} \biggl(\bar{\lambda}_{j,0} q_{j,0,\nu} + \sum_{k=1}^{K} b_{j,k,\nu} \bar{\lambda}_{j,k} \frac{e_k^2 + \gamma_k \sigma_k^2}{e_k + \bar{\lambda}_{j,k} \sigma_k^2} c(p_{j,k}) \biggr), \quad \nu \in \mathbb{N}_0^d, \tag{7.87}$$

with $(q_{j,0,\nu})_{\nu \in \mathbb{N}_0^d}$ given by (7.7) or (7.8) and $(b_{j,k,\nu})_{\nu \in \mathbb{N}_0^d}$ given by (7.82) and (7.83). For every $j \in \mathcal{J}$ with $\lambda_j = 0$, we define $\tilde{\varphi}_j(s) = 1$ for all $s \in \mathbb{C}^d$ and

$$c_{j,\nu} = \begin{cases} 1 & \text{for } \nu = 0 \in \mathbb{N}_0^d, \\ 0 & \text{for } \nu \in \mathbb{N}_0^d \setminus \{0\}. \end{cases} \tag{7.88}$$

In every case, $\tilde{\varphi}_j$ is again a probability-generating function, and (7.85) can be written as

$$\varphi_{L,\gamma}(s) = \sum_{j \in \mathcal{J}} \exp\bigl(\lambda_j (\tilde{\varphi}_j(s) - 1)\bigr) \, \mathbb{P}[J = j] \,. \tag{7.89}$$

Fix a scenario $j \in \mathcal{J}$, 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 (4.75), the probability-generating function ψ_j of the distribution of the random sum

$$S_j := \sum_{n=1}^{M_j} Y_{j,n}$$

is given by

$$\psi_j(s) = \exp\bigl(\lambda_j(\tilde\varphi_j(s)-1)\bigr), \qquad s\in\mathbb{C}^d,\ \|s\|_\infty\le 1,$$

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 5.21. Explicitly, (5.21) implies for the initial value

$$d_{j,0} = \exp\bigl(\lambda_j(c_{j,0}-1)\bigr) \tag{7.90}$$

(in case of numerical underflow, see Remark 5.23 for a remedy) and the recursion formula (5.22) turns, for every $\nu=(\nu_1,\dots,\nu_d)\in\mathbb{N}_0^d\setminus\{0\}$, into

$$d_{j,\nu} = \frac{\lambda_j}{\nu_i} \sum_{\substack{n\in\mathbb{N}_0^d\\ 0<n\le\nu}} n_i\, c_{j,n}\, d_{j,\nu-n}, \tag{7.91}$$

where $i\in\{1,\dots,d\}$ is chosen such that $\nu_i\ne 0$, with λ_j given by (7.86) and the coefficients $(c_{j,\nu})_{\nu\in\mathbb{N}_0^d}$ given by (7.87) and (7.88), respectively. See Remark 5.19 to omit terms in (7.91) with value zero.

The weighted probability-generating function (7.89) simplifies to

$$\varphi_{L,\gamma}(s) = \sum_{j\in\mathcal{J}} \psi_j(s)\,\mathbb{P}[J=j]\,, \qquad s\in\mathbb{C}^d,\ \|s\|_\infty\le 1,$$

and the coefficients of this power series are convex combinations of the corresponding coefficients of $(\psi_j)_{j\in\mathcal{J}}$. These operations are numerically stable. Explicitly, the coefficients in (7.67) are determined by

$$\mathbb{P}_\gamma[L=\nu] = \sum_{j\in\mathcal{J}} d_{j,\nu}\,\mathbb{P}[J=j]\,, \qquad \nu\in\mathbb{N}_0^d,$$

with $(d_{j,\nu})_{\nu\in\mathbb{N}_0}$ given by (7.90) and (7.91).

Exercise 7.48 (Implementation of the algorithm)**.** Assume that there are $m\in\mathbb{N}$ obligors, where obligor $i\in\{1,\dots,m\}$ 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 $i\in\{1,\dots,m\}$ due to cause $c\in\{0,\dots,C\}$ has the distribution $\mathrm{Bin}(i+c, i/(2i+2c))$ and that all susceptibilities are equal to $1/(C+1)$. Let $\Lambda_1,\dots,\Lambda_C$ be default cause intensities with $\mathbb{E}[\Lambda_c]=1$ and $\Lambda_c\ge 1/3c$ for all $c\in\{1,\dots,C\}$. Assume the there are only one-element risk groups and that there are two scenarios $\mathcal{J}=\{0,1\}$. Extending Example 7.33, let J be $\mathcal{J}$-valued and consider the $(C+1)\times(K+1)$-matrix

$$A_J = \begin{pmatrix} 1 & 0 & 0 & 0 & 0\\ * & J & 0 & 0 & 0\\ * & 0 & 1-J & * & 0\\ * & 0 & 0 & * & * \end{pmatrix},$$

where $*$ denotes non-zero, deterministic entries.

(a) With the given constraints, set up a flexible model satisfying Assumptions 7.29 and 7.30 such that $\mathrm{Cov}(\Lambda_1, \Lambda_2) < 0$ and $\mathrm{Cov}(\Lambda_2, \Lambda_3) > 0$.

(b) Calculate the expectations, variances and covariances of the default cause intensities $\Lambda_1, \Lambda_2, \Lambda_3$ (see Remark 7.27) 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.

7.8 Algorithm for Risk Factors with a Tempered Stable Distribution

7.9 Special Cases

[49] 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 (7.14) 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

$$L_{g,k,n} = C\, L'_{g,k,n}.$$

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 (7.10), (7.12) and (7.14). In this section, we will not attribute the group loss to its individual members.

7.9.1 Pure Poisson Case

[50] We only consider the degenerate case $\sigma_1^2 = \cdots = \sigma_K^2 = 0$, for which the algorithm described in Section 7.7 works and for which $\Lambda_k \equiv 1$ almost surely for all $k \in \{1, \ldots, K\}$. In this case the family

$$\big\{ N_{g,k} \,\big|\, g \in G,\, k \in \{0, \ldots, K\} \big\}$$

consists of independent, Poisson distributed random variables.

Bernoulli Loss Distribution Assume that every $L'_{g,k,n}$ is a Bernoulli random variable, i.e.,

$$p := \mathbb{P}\big[L'_{g,k,n} = 1\big] = 1 - \mathbb{P}\big[L'_{g,k,n} = 0\big]$$

with $p \in [0,1]$ for all $g \in G$, $k \in \{0, \ldots, K\}$ and $n \in \mathbb{N}$. Then, by (7.4), (**??**) and (7.6), $\lambda_{k,\nu} = 0$ 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 (**??**),

$$L'_{g,k} := \sum_{n=1}^{N_{g,k}} L'_{g,k,n} \sim \mathrm{Poisson}(p \lambda_g w_{g,k})\,.$$

[49] This section has to be adapted to the new notation and the generalized setting.
[50] This section has to be adapted to the new notation and the generalized setting.

By the Poisson summation property (3.5), we obtain for

$$L' := \sum_{g\in G}\sum_{k=0}^{K} L'_{g,k} \tag{7.92}$$

that $L' \sim \text{Poisson}(p\lambda)$ with

$$\lambda := \sum_{g\in G}\lambda_g \underbrace{\sum_{k=0}^{K} w_{g,k}}_{=1\text{ by (7.2)}} . \tag{7.93}$$

Therefore, the distribution of $L = C\,L'$ satisfies

$$\mathbb{P}(L=l) = \begin{cases} \frac{(p\lambda)^n}{n!}\,\mathrm{e}^{-p\lambda} & \text{if } n := l/C \in \mathbb{N}_0,\\ 0 & \text{otherwise.}\end{cases}$$

Logarithmic Loss Distribution Assume that every $L'_{g,k,n} \sim \text{Log}(q)$ with $q \in (0,1)$. According to Example 4.40, the compound Poisson sum $L'_{g,k}$ has the distribution $\text{NegBin}(\alpha_{g,k}, p)$ with parameters $p := 1-q$ and

$$\alpha_{g,k} := -\frac{\lambda_g w_{g,k}}{\log p} \geq 0.$$

By Lemma 4.37, the sum L' defined in (7.92) has distribution $\text{NegBin}(\alpha, p)$ with $\alpha := -\lambda/\log p$ and λ given by (7.93). Therefore, $L = C\,L'$ satisfies

$$\mathbb{P}(L=l) = \begin{cases} \binom{\alpha+n-1}{n} p^\alpha q^n & \text{if } n := l/C \in \mathbb{N}_0,\\ 0 & \text{otherwise.}\end{cases} \tag{7.94}$$

General Loss Distributions Let $Q^{\mathrm{s}}_{g,k} = (q^{\mathrm{s}}_{g,k,\nu})_{\nu\in\mathbb{N}_0}$ 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 $k \in \{0,\dots,K\}$. Then every $L_{g,k} \sim \text{CPoisson}\big(\lambda_g w_{g,k}, Q^{\mathrm{s}}_{g,k}\big)$ has a compound Poisson distribution. By (4.77), its generating function is

$$\varphi_{L_{g,k}}(s) = \exp\biggl(\lambda_g w_{g,k}\biggl(\underbrace{\sum_{\nu\in\mathbb{N}_0} q^{\mathrm{s}}_{g,k,\nu} s^\nu}_{=\varphi_{L_{g,k,1}}(s)} -1\biggr)\biggr). \tag{7.95}$$

Assume that the sum λ of all weighted intensities, given by (7.93), is strictly positive. Define the probability distribution $Q = (q_\nu)_{\nu\in\mathbb{N}_0}$ by

$$q_\nu = \frac{1}{\lambda}\sum_{g\in G}\sum_{k=0}^{K} \lambda_g w_{g,k} q^{\mathrm{s}}_{g,k,\nu}, \qquad \nu \in \mathbb{N}_0.$$

Due to independence, the generating function φ_L of the total loss L is the product of the individual functions from (7.95), hence

$$\varphi_L(s) = \prod_{g\in G}\prod_{k=0}^{K} \varphi_{L_{g,k}}(s) = \exp\biggl(\lambda\biggl(\sum_{\nu\in\mathbb{N}_0} q_\nu s^\nu - 1\biggr)\biggr),$$

in particular $L \sim \mathrm{CPoisson}(\lambda, Q)$ has a compound Poisson distribution. Hence, the distribution of L can be calculated by the Panjer recursion formula (5.22), i.e.

$$\mathbb{P}[L = l] = \frac{\lambda}{l}\sum_{\nu=1}^{l} \nu q_\nu\, \mathbb{P}[L = l-\nu], \qquad l \in \mathbb{N},$$

starting from

$$\mathbb{P}[L = 0] = \varphi_L(0) = \mathrm{e}^{\lambda(q_0-1)}\,.$$

7.9.2 Case of Negative Binomial Distribution

[51] Here we assume absence of idiosyncratic risk, meaning that $\lambda_{0,\nu} = 0$ for all $\nu \in \mathbb{N}$ and $\bar\lambda_0 = 0$, see (7.4) and (7.6).

Bernoulli Loss Distribution Assume that $L'_{g,k,n}$ is a Bernoulli random variable with risk-dependent distribution, i.e.,

$$p_k \coloneqq \mathbb{P}\bigl[L'_{g,k,n} = 1\bigr] = 1 - \mathbb{P}\bigl[L'_{g,k,n} = 0\bigr]$$

with $p_k \in [0,1]$ for all $g \in G$, $k \in \{1,\dots,K\}$ and $n \in \mathbb{N}$. Then, by (7.4) and (7.6), $\lambda_{k,\nu} = 0$ for every $\nu \in \mathbb{N}\setminus\{C\}$ and $\bar\lambda_k = \lambda_{k,C}$ for every risk $k \in \{1,\dots,K\}$. Furthermore, assume that there exist a non-empty $I \subseteq \{1,\dots,K\}$ and $p \in (0,1)$ such that $\sigma_k^2\bar\lambda_k = (1-p)/p$ for all $k \in I$ and $\bar\lambda_k = 0$ for all $k \in \{1,\dots,K\}\setminus I$. By (**??**) this means $\nu_k = C$ for all $k \in I$ and $\nu_k = 0$ for all $k \in \{1,\dots,K\}\setminus I$. Define

$$\alpha = \sum_{k\in I}\frac{1}{\sigma_k^2}.$$

Then (**??**) simplifies to

$$\mathbb{E}\bigl[s^L\bigr] = \Bigl(1 + \frac{1-p}{p}(1-s^C)\Bigr)^{-\alpha} = \Bigl(\frac{p}{1-qs^C}\Bigr)^{\alpha}$$

with $q \coloneqq 1-p$, which by (4.65) means that $L' \coloneqq L/C \sim \mathrm{NegBin}(\alpha, p)$, hence L has the distribution given by (7.94).

[51] This 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 $k \in \{1,\dots,K\}$. Since $\mathcal{L}(N_{g,k}|\Lambda_k) \overset{\text{a.s.}}{=} \text{Poisson}(\lambda_g w_{g,k}\Lambda_k)$ by Assumption 7.24, and since $(N_{g,k})_{g\in G}$ are conditionally independent given Λ_k by Assumption (7.25), Lemma 3.2 for sums of independent Poisson random variables implies that

$$\mathcal{L}(N_{(k)}|\Lambda_k) \overset{\text{a.s.}}{=} \text{Poisson}\bigl(\lambda_{(k)}\Lambda_k\bigr)$$

for every $k \in \{1,\dots,K\}$, where

$$N_{(k)} := \sum_{g\in G} N_{g,k} \qquad \text{and} \qquad \lambda_{(k)} := \sum_{g\in G} \lambda_g w_{g,k}\,.$$

Here $N_{(k)}$ is the number of defaults in the portfolio caused by risk $k \in \{1,\dots,K\}$. Since $\Lambda_k \sim \text{Gamma}(\alpha_k,\beta_k)$ with $\alpha_k = \beta_k = 1/\sigma_k^2$ by Assumption 7.29, hence

$$\lambda_{(k)}\Lambda_k \sim \text{Gamma}(\alpha_k, \beta_k/\lambda_{(k)}),$$

we get for the unconditional distribution that

$$N_{(k)} \sim \text{NegBin}(\alpha_k, p_k) \qquad \text{with} \qquad p_k := \frac{\beta_k/\lambda_{(k)}}{1+\beta_k/\lambda_{(k)}} = \frac{1}{1+\lambda_{(k)}\sigma_k^2}\,,$$

where we use the notation from (4.61). Assuming that $\lambda_{(k)}\sigma_k^2$ and, therefore, $p := p_k$ are the same for every risk $k \in \{1,\dots,K\}$, then we get for the total number $N := \sum_{k=1}^K N_{(k)}$ of defaults caused by all the independent risk factors that

$$N \sim \text{NegBin}(\alpha, p) \qquad \text{with} \qquad \alpha := \alpha_1 + \cdots + \alpha_K,$$

see Lemma 4.37. Therefore we have a compound negative binomial distribution for the loss L given in (7.14), meaning that

$$L = \sum_{g\in G}\sum_{k=1}^{K}\sum_{n=1}^{N_{g,k}} L_{g,k,n} \overset{\text{d}}{=} \sum_{n=1}^{N} X_n \sim \text{CNegBin}(\alpha, p, Q)$$

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 (5.24)

$$\mathbb{P}[L=l] = \frac{1}{1-(1-p)q_0}\,\frac{1-p}{l}\sum_{\nu=1}^{l}(\alpha\nu + l - \nu)\,q_\nu\,\mathbb{P}[L=l-\nu], \qquad l \in \mathbb{N},$$

starting from

$$\mathbb{P}[L=0] = \varphi_N(q_0) = \Bigl(\frac{p}{1-(1-p)q_0}\Bigr)^{\alpha},$$

see (5.23).

Exercise 7.49. Consider a logarithmic distribution for the idiosyncratic losses and a Bernoulli distribution for the losses due to the risks $k \in \{1,\dots,K\}$, 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 (7.94).

8 Risk Measures and Risk Contributions

Knowing the distribution of the portfolio loss L given in (7.14), 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.

8.1 Quantiles and Value-at-Risk

Definition 8.1 (Lower and upper quantile)**.** For a real-valued random variable X and a level $\delta \in (0,1)$, define the lower δ-quantile of X by

$$q_\delta(X) = \min\{x \in \mathbb{R} \mid \mathbb{P}[X \le x] \ge \delta\} \tag{8.1}$$

and the upper δ-quantile of X by

$$q^\delta(X) = \inf\{x \in \mathbb{R} \mid \mathbb{P}[X \le x] > \delta\}. \tag{8.2}$$

Since the distribution function $\mathbb{R} \ni x \mapsto F_X(x) := \mathbb{P}[X \le x]$ 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 $q_\delta(X) \le q^\delta(X)$.

Exercise 8.2. 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 δ. 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 (8.1) as

$$q_\delta(X) = \min\{x \in \mathbb{R} \mid \mathbb{P}[X > x] \le 1-\delta\},$$

we see that $q_\delta(X)$ is the smallest threshold which is exceeded by the loss X with probability at most $1-\delta$.

Exercise 8.3. Give an example where $(0,1) \ni \delta \mapsto q_\delta(X)$ 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 8.4 (Downward jump of lower quantile)**.** Consider the unit interval $\Omega = [0,1]$ equipped with Borel σ-algebra $\mathcal{B}([0,1])$. Let $\mathbb{P}$ denote the Lebesgue–Borel 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 $X_n \searrow X$ pointwise as $n \to \infty$, $q_\delta(X_n) = 1$ for all $n \in \mathbb{N}$ but $q_\delta(X) = 0$ by (8.1).

Exercise 8.5 (Upward jump of upper quantile)**.** Modify Example 8.4 such that $X_n \nearrow X$ pointwise as $n \to \infty$, $q^\delta(X_n) = 0$ for all $n \in \mathbb{N}$ but $q^\delta(X) = 1$.

Lemma 8.6 (Convergence properties of quantiles)**.** *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.,*

$$\lim_{n\to\infty} \mathbb{P}[|X - X_n| \geq \varepsilon] = 0 \quad \textit{for every } \varepsilon > 0. \tag{8.3}$$

(a) The lower δ-quantiles satisfy

$$\liminf_{n\to\infty} q_\delta(X_n) \geq q_\delta(X).$$

(b) The upper δ-quantiles satisfy

$$\limsup_{n\to\infty} q^\delta(X_n) \leq q^\delta(X).$$

(c) If the distribution of X satisfies $\mathbb{P}[X \leq x] > \delta$ for all $x > q_\delta(X)$, which is equivalent to $q_\delta(X) = q^\delta(X)$, then

$$\lim_{n\to\infty} q_\delta(X_n) = q_\delta(X) \quad \textit{and} \quad \lim_{n\to\infty} q^\delta(X_n) = q^\delta(X).$$

Proof. (a) If $x < y < q_\delta(X)$, then $\{X_n \leq x\} \subseteq \{X \leq y\} \cup \{X - X_n \geq y - x\}$, hence

$$\mathbb{P}[X_n \leq x] \leq \mathbb{P}[X \leq y] + \underbrace{\mathbb{P}[|X - X_n| \geq y - x]}_{\to 0 \text{ as } n \to \infty \text{ by } (8.3)},$$

and therefore

$$\limsup_{n\to\infty} \mathbb{P}[X_n \leq x] \leq \gamma := \mathbb{P}[X \leq y] < \delta$$

by the definition of $q_\delta(X)$ in (8.1). Therefore $\mathbb{P}[X_n \leq x] \leq (\delta + \gamma)/2 < \delta$ 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 the one for part (a). If $x > y > q^\delta(X)$, then

$$\mathbb{P}[X_n \leq x] \geq \mathbb{P}[X \leq y] - \underbrace{\mathbb{P}[|X - X_n| \geq x - y]}_{\to 0 \text{ as } n \to \infty \text{ by } (8.3)},$$

hence

$$\liminf_{n\to\infty} \mathbb{P}[X_n \leq x] \geq \gamma := \mathbb{P}[X \leq y] > \delta$$

by the definition of $q^\delta(X)$ in (8.2). Therefore $\mathbb{P}[X_n \leq x] \geq (\delta + \gamma)/2 > \delta$ 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) because $q_\delta(X_n) \leq q^\delta(X_n)$ for all $n \in \mathbb{N}$. □

If we have an estimate for the Kolmogorov–Smirnov distance of two distributions, then we get bounds for the quantiles of these distributions.

Lemma 8.7 (Quantiles and Kolmogorov–Smirnov metric)**.** *Let X and Y be real-valued random variables and abbreviate the Kolmogorov–Smirnov distance of their distributions by $d := d_{\mathrm{KS}}(\mathcal{L}(X), \mathcal{L}(Y))$. Then the lower quantiles of X and Y satisfy*

(a) $q_{\delta-d}(X) \le q_\delta(Y)$ for every level $\delta \in (d, 1)$ and

(b) $q_\delta(Y) \le q_{\delta+d}(X)$ for every level $\delta \in (0, 1-d)$.

Proof. (a) Given a level $\delta \in (d, 1)$, we use the definition (8.1) of the lower quantile and insert the term $\mathbb{P}[X \le q_\delta(Y)]$, hence

$$\begin{aligned}\delta &\le \mathbb{P}[Y \le q_\delta(Y)] \\ &\le \mathbb{P}[X \le q_\delta(Y)] + \big|\mathbb{P}[Y \le q_\delta(Y)] - \mathbb{P}[X \le q_\delta(Y)]\big|.\end{aligned}$$

Due to $d_{\mathrm{KS}}(\mathcal{L}(X), \mathcal{L}(Y)) = \sup_{x\in\mathbb{R}}\big|\mathbb{P}[X \le x] - \mathbb{P}[Y \le x]\big|$ by (3.13), this implies

$$\delta \le \mathbb{P}[X \le q_\delta(Y)] + d,$$

hence $\mathbb{P}[X \le q_\delta(Y)] \ge \delta - d$, therefore $q_{\delta-d}(X) \le q_\delta(Y)$ by (8.1).

(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 $q_{\delta'-d}(Y) \le q_{\delta'}(X)$, which proves part (b). □

Exercise 8.8. In the setting of Lemma 8.7 show the following:

(a) There is a non-trivial example (i.e. one with $\mathcal{L}(X) \ne \mathcal{L}(Y)$) such that $q_{\delta-d}(X) = q_\delta(Y) = q_{\delta+d}(X)$ for at least for one level δ.

(b) There is an example with $q_{\delta-d}(X) < q_\delta(Y) < q_{\delta+d}(X)$ for at least for one level δ.

(c) Formulate and prove a version of Lemma 8.7 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 $\mathrm{VaR}(X) + \mathrm{VaR}(Y) < \mathrm{VaR}(X+Y)$ for loss variables X and Y, meaning that diversification might seem to increase risk[52] when it is measured with VaR, see Example 8.9. Due to these deficiencies, we do not pursue the topic of Value-at-Risk in more detail.

[52] When it comes to catastrophe risks in connection with limited liability, then diversification can actually increase the risk: Think of two companies with independent risks X and Y, respectively, which cause insolvency. Then the probability of both companies going bankrupt is considerable smaller than one company having risk $X+Y$; see Example 8.9 to work out a numerical example.

Example 8.9 (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.

8.1.1 Calculation and Smoothing of Lower Quantiles in Extended CreditRisk$^+$

Remark 8.10 (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 (7.14) can be calculated in extended CreditRisk$^+$ by adding up the probabilities $\mathbb{P}[L = l]$ for $l = 0, 1, 2\ldots$ until the sum reaches or exceeds δ, see (8.1).

However, this means that $q_\delta(L)$ as a function of δ, 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 6.1 for a discussion of this discretisation procedure) has been applied to the individual losses, then somehow "reversing" this step is a legitimate wish.

Remark 8.11 (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 $\mathbb{E}[U] = 0$. Define the smoothed loss L_s by

$$L_\mathrm{s} = L + \mathbb{1}_{\{L>0\}} U. \tag{8.4}$$

Then L_s takes values in $[0,\infty)$ and by independence

$$\mathbb{E}[L_\mathrm{s}] = \mathbb{E}[L] + \mathbb{P}[L > 0]\,\mathbb{E}[U] = \mathbb{E}[L]\,,$$

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 $|L - L_\mathrm{s}|$ is bounded by $\frac{1}{2}$ and the distribution function of L_s is given by

$$F_{L_\mathrm{s}}(x) = \begin{cases} 0 & \text{for } x < 0, \\ p_0 & \text{for } x \in [0, \frac{1}{2}), \\ \sum_{k=0}^{n-1} p_k + p_n(x - n + \frac{1}{2}) & \text{for } x \in [n - \frac{1}{2}, n + \frac{1}{2}) \text{ with } n \in \mathbb{N}. \end{cases}$$

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 $q_\delta(L_\mathrm{s})$ is given by $q_\delta(L_\mathrm{s}) = 0$ if $q_\delta(L) = 0$ and

$$q_\delta(L_\mathrm{s}) = q_\delta(L) + \frac{1}{2} - \frac{\mathbb{P}[L \le q_\delta(L)] - \delta}{\mathbb{P}[L = q_\delta(L)]} \tag{8.5}$$

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 8.12 (More general smoothing)**.** For a more general smoothing of the lower quantile, we can consider the smoothed loss in (8.4), where $U = V_1 - V_2$ with independent $V_1, V_2 \sim \mathrm{Beta}(\alpha, \beta)$. Of course, then the formula (8.5) for the smoothed quantile $q_\delta(L_\mathrm{s})$ 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 $(0, \infty)$.

8.2 Expected Shortfall

8.2.1 Definition and Representations of Expected Shortfall

Definition 8.13 (Expected shortfall)**.** 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

$$\mathrm{ES}_\delta[X] = \frac{\mathbb{E}\big[X\mathbb{1}_{\{X > q_\delta(X)\}}\big] + q_\delta(X)(\mathbb{P}[X \le q_\delta(X)] - \delta)}{1-\delta} \tag{8.6}$$

with the understanding that $\mathrm{ES}_\delta[X] \coloneqq \infty$ 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 8.14 (Simple case of expected shortfall)**.** If $\mathbb{P}[X \le q_\delta(X)] = \delta$, in particular if the distribution function $\mathbb{R} \ni x \mapsto \mathbb{P}[X \le x]$ of X is also left-continuous at $x = q_\delta(X)$, then (8.6) simplifies to

$$\mathrm{ES}_\delta[X] = \mathbb{E}[X \,|\, X > q_\delta(X)]\,. \tag{8.7}$$

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 (8.6) is necessary to prove the sub-additivity of expected shortfall in Theorem 8.20 below. The representation (8.7) justifies the name *conditional value-at-risk*, which is also used in the literature.

Remark 8.15 (Alternative representation of expected shortfall)**.** Using the identity $X = (X - q_\delta(X)) + q_\delta(X)$, it follows that

$$\mathbb{E}\big[X\mathbb{1}_{\{X > q_\delta(X)\}}\big] = \mathbb{E}\big[(X - q_\delta(X))^+\big] + q_\delta(X)\,\mathbb{P}[X > q_\delta(X)]\,,$$

hence we obtain from (8.6) the alternative representation

$$\mathrm{ES}_\delta[X] = q_\delta(X) + \frac{\mathbb{E}[(X - q_\delta(X))^+]}{1-\delta} \tag{8.8}$$

of expected shortfall, which clearly shows that $\mathrm{ES}_\delta[X] \ge q_\delta(X)$. See Theorem 8.20(g) below for the special property of $q_\delta(X)$ in (8.8).

Exercise 8.16. Give an example with $\mathbb{P}[X \le q_\delta(X)] = \delta$, where the distribution function of X is discontinuous at $q_\delta(X)$.

Exercise 8.17. Show that expected shortfall is *law determined* (sometime called *law invariant* in the literature) by representing $\mathrm{ES}_\delta[X]$ in terms of the distribution function F_X of X.

Remark 8.18 (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 $f_X \colon \Omega \to [0, \infty)$ by

$$f_X = \frac{\mathbb{1}_{\{X > q_\delta(X)\}} + \beta_X \mathbb{1}_{\{X = q_\delta(X)\}}}{1-\delta}, \tag{8.9}$$

where the constant β_X is given by

$$\beta_X = \begin{cases} \frac{\mathbb{P}[X \le q_\delta(X)] - \delta}{\mathbb{P}[X = q_\delta(X)]} & \text{if } \mathbb{P}[X = q_\delta(X)] > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{8.10}$$

It follows from the definition of the lower δ-quantile of X in (8.1) that $\beta_X \in [0, 1]$, hence f_X is bounded by $1/(1-\delta)$. Note that

$$\mathbb{E}[f_X] = \frac{1}{1-\delta}\bigl(\mathbb{P}[X > q_\delta(X)] + \underbrace{\beta_X \, \mathbb{P}[X = q_\delta(X)]}_{= \mathbb{P}[X \le q_\delta(X)] - \delta}\bigr) = 1, \tag{8.11}$$

hence f_X is a probability density. By the definition of expected shortfall in (8.6),

$$\mathbb{E}[X f_X] = \frac{\mathbb{E}[X \mathbb{1}_{\{X > q_\delta(X)\}}] + q_\delta(X)\beta_X \, \mathbb{P}[X = q_\delta(X)]}{1-\delta} = \mathrm{ES}_\delta[X]. \tag{8.12}$$

Therefore, expected shortfall at level δ can be see as the expectation of X taken with respect to a probability measure $\mathbb{Q}_X$ 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 $1/(1-\delta)$.

8.2.2 Calculation of Expected Shortfall in Extended CreditRisk$^+$

Remark 8.19. Since the credit portfolio loss L, given in (7.14), is a discrete random variable, we have to apply the more complicated definition (8.6). As mentioned in Remark 8.10, the lower quantile $q_\delta(L)$ and $\mathbb{P}[L \le q_\delta(L)]$ can be calculated using the extended CreditRisk$^+$ algorithm. Furthermore, note that

$$\mathbb{E}\bigl[L \mathbb{1}_{\{L > q_\delta(L)\}}\bigr] = \mathbb{E}[L] - \mathbb{E}\bigl[L \mathbb{1}_{\{L \le q_\delta(L)\}}\bigr] \tag{8.13}$$

with $\mathbb{E}[L]$ given by (7.57) and

$$\mathbb{E}\bigl[L \mathbb{1}_{\{L \le q_\delta(L)\}}\bigr] = \sum_{l=1}^{q_\delta(L)} l \, \mathbb{P}[L = l] \, .$$

If $\mathbb{E}[L] = \infty$, then $\mathrm{ES}_\delta[L] = \infty$. If $\mathbb{E}[L] < \infty$, then the expected shortfall $\mathrm{ES}_\delta[L]$ from (8.6) can be computed numerically using the first terms of the distribution of L. Note that the differences in (8.6) and (8.13) can lead to cancellation effects, in particular when $\mathbb{E}[L] \approx \mathbb{E}[L \mathbb{1}_{\{L \le q_\delta(L)\}}]$ for large quantiles.

8.2.3 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 $\mathcal{F}_\delta$ 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, since we do not impose a general integrability condition, we also define the restricted set

$$\mathcal{F}_{\delta,X} := \{\, f \in \mathcal{F}_\delta \mid \mathbb{E}[X^+ f] < \infty \text{ or } \mathbb{E}[X^- f] < \infty \,\}, \tag{8.14}$$

where $X^\pm := \max\{\pm X, 0\}$ 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 (8.9) is in $\mathcal{F}_\delta$ and that $X^- f_X$ is a random variable bounded above by $|q_\delta(X)|/(1-\delta)$, hence $\mathbb{E}[X^- f_X] < \infty$ and therefore $f_X \in \mathcal{F}_{\delta,X}$.

Theorem 8.20. *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 $\mathrm{ES}_\delta[\alpha X] = \alpha\, \mathrm{ES}_\delta[X]$.

(b) Translation (or cash) invariance: If $a \in \mathbb{R}$, then $\mathrm{ES}_\delta[X + a] = \mathrm{ES}_\delta[X] + a$.

(c) Scenario representation:

  (i) $\mathrm{ES}_\delta[X] = \sup_{f \in \mathcal{F}_{\delta,X}} \mathbb{E}[Xf]$,

  (ii) if[53] $\mathbb{E}[X^+] < \infty$, then $\mathrm{ES}_\delta[X] = \sup_{f \in \mathcal{F}_\delta} \mathbb{E}[Xf]$.

(d) Sub-additivity: $\mathrm{ES}_\delta[X + Y] \le \mathrm{ES}_\delta[X] + \mathrm{ES}_\delta[Y]$.

(e) Monotonicity: If $X \le Y$, then $\mathrm{ES}_\delta[X] \le \mathrm{ES}_\delta[Y]$.

(f) Convexity: If $\alpha \in (0,1)$, then

$$\mathrm{ES}_\delta[\alpha X + (1-\alpha)Y] \le \alpha\, \mathrm{ES}_\delta[X] + (1-\alpha)\, \mathrm{ES}_\delta[Y].$$

(g) Minimization property:

$$\mathrm{ES}_\delta[X] = \min_{q \in \mathbb{R}} \Big(q + \frac{\mathbb{E}[(X-q)^+]}{1-\delta} \Big),$$

and the minimum is attained if and only if $q \in [q_\delta(X), q^\delta(X)]$.

(h) Bounds: For every $q \in \mathbb{R}$,

$$q_\delta(X) \le \mathrm{ES}_\delta[X] \le q + \frac{\mathbb{E}[(X-q)^+]}{1-\delta},$$

where the lower bound is an equality if and only if $\mathbb{P}[X \le q_\delta(X)] = 1$, and the upper bound is an equality if and only if $q \in [q_\delta(X), q^\delta(X)]$.

[53] If you offer the St. Petersburg lottery, then $\mathbb{E}[X^+] = \infty$.

(i) *Quantile representation:*

$$\mathrm{ES}_\delta[X] = \frac{1}{1-\delta} \int_{[\delta,1)} q_u(X)\,\mathrm{d}u. \tag{8.15}$$

(j) *Fatou property: 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 $X := \liminf_{n\to\infty} X_n$ satisfies*

$$\mathrm{ES}_\delta[X] \leq \liminf_{n\to\infty} \mathrm{ES}_\delta[X_n]. \tag{8.16}$$

(k) *Let $(X_n)_{n\in\mathbb{N}}$ be bounded below and converging in probability to a random variable X. Then (8.16) holds, too.*

Corollary 8.21. *For every real-valued random variable X, the map*

$$(0,1) \ni \delta \mapsto \mathrm{ES}_\delta[X] \in \mathbb{R} \cup \{\infty\}$$

is continuous and non-decreasing.

Proof of Corollary 8.21. Continuity follows from the quantile representation in part (i). For $\delta \leq \delta'$ we have $\mathcal{F}_{\delta,X} \subseteq \mathcal{F}_{\delta',X}$ which implies $\mathrm{ES}_\delta[X] \leq \mathrm{ES}_{\delta'}[X]$ by the scenario representation (c). □

Remark 8.22. A coherent risk measure is defined by monotonicity (e), positive homogeneity (a), translation invariance (b) and sub-additivity (d), see Artzner, Delbaen, Eber and Heath [3]. A convex risk measure is defined by monotonicity (e), translation invariance (b) and convexity (f), see Föllmer and Schied [21]. 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 [43].

Remark 8.23. 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 8.24. Concerning the properties in Theorem 8.20, some comments might be useful:

(a) If all losses are scaled (by converting them to a different currency, for example), 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 $\mathrm{ES}_\delta[X]$ is the worst expected loss possible.

(d), (f) Diversification does not increase the risk, regardless of any dependence between X and Y.

(e) Smaller losses need less capital.

(g) For an economic interpretation in the case $\mathbb{P}[X \leq q_\delta(X)] = \delta$, assume that you can choose an amount q and enter into a special stop-loss insurance contract:

Whenever your loss X is above $q_\delta(X)$, you must pay q plus 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 to cover your losses. Which q is optimal for you and how much do you lose given the loss X exceeds $q_\delta(X)$? If q is too high, your pay too much in the case $q_\delta(X) < X < q$; if q is too small, your premium part $\mathbb{E}[(X-q)^+]/(1-\delta)$ is too high. 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 δ, contrary to the quantile function $(0,1) \ni \delta \mapsto q_\delta(X)$, which can jump, see Exercise 8.3. 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.

(k) implies the Fatou property discussed in Delbaen [14].

Proof of Theorem 8.20. (a) follows from the homogeneity of the expectation in (8.6) and the observation that $q_\delta(\alpha X) = \alpha q_\delta(X)$, see (8.1).

(b) holds because of the translation invariance of the expectation in (8.6) and the observation that $q_\delta(X+a) = q_\delta(X) + a$, see (8.1).

(c) Remark 8.18, in particular (8.12), shows that equality holds for $f_X \in \mathcal{F}_{\delta,X}$. Therefore, the supremum is an upper estimate and (i) holds in the case $\mathrm{ES}_\delta[X] = \infty$. If $\mathrm{ES}_\delta[X] < \infty$, then necessarily $\mathbb{E}[X^+] < \infty$, hence $\mathcal{F}_{\delta,X} = \mathcal{F}_\delta$ by (8.14). Consider $f \in \mathcal{F}_\delta$ with $\mathbb{E}[Xf] > -\infty$. We have $\mathbb{E}[f - f_X] = 0$, hence

$$\begin{aligned}
\mathbb{E}[Xf] - \mathbb{E}[Xf_X] &= \mathbb{E}[(X - q_\delta(X))(f - f_X)] \\
&= \mathbb{E}[\underbrace{(X - q_\delta(X))}_{>0}\underbrace{(f - f_X)}_{\le 0}\mathbb{1}_{\{X > q_\delta(X)\}}] \\
&\quad + \mathbb{E}[\underbrace{(X - q_\delta(X))}_{<0}\underbrace{(f - f_X)}_{\ge 0}\mathbb{1}_{\{X < q_\delta(X)\}}] \le 0,
\end{aligned}$$

which means that the supremum is identical with $\mathbb{E}[Xf_X]$.

(d) It suffices to consider the case where $\mathrm{ES}_\delta[X] < \infty$ and $\mathrm{ES}_\delta[Y] < \infty$. Then $\mathbb{E}[X^+]$, $\mathbb{E}[Y^+]$ and $\mathbb{E}[(X+Y)^+]$ are finite and with the representation from (c), part (ii), we get that

$$\begin{aligned}
\mathrm{ES}_\delta[X+Y] &= \sup_{f \in \mathcal{F}_\delta} \mathbb{E}[(X+Y)f] \\
&\le \sup_{f \in \mathcal{F}_\delta} \mathbb{E}[Xf] + \sup_{f \in \mathcal{F}_\delta} \mathbb{E}[Yf] = \mathrm{ES}_\delta[X] + \mathrm{ES}_\delta[Y].
\end{aligned}$$

(e) Note that $\mathrm{ES}_\delta[X] \le \mathrm{ES}_\delta[X-Y] + \mathrm{ES}_\delta[Y]$ by subadditivity (d). For $Z := X - Y \le 0$, we have $\mathrm{ES}_\delta[Z] \le 0$ according to (8.6) because $\mathbb{E}\big[Z\mathbb{1}_{\{Z > q_\delta(Z)\}}\big] \le 0$ and $q_\delta(Z) \le 0$ for a non-positive random variable as well as $\mathbb{P}[Z \le q_\delta(Z)] \ge \delta$ by the definition of the lower quantile in (8.1).

(f) follows from (d) and (a).

(g) By the alternative representation (8.8), we have equality for $q = q_\delta(X)$. Note that, for every $q \in \mathbb{R}$,

$$X - q_\delta(X) = (q - q_\delta(X)) + (X - q). \tag{8.17}$$

Consider the case $q > q_\delta(X)$. Then the inequality

$$(X - q_\delta(X))^+ \le (q - q_\delta(X))\mathbb{1}_{\{X > q_\delta(X)\}} + (X - q)^+ \tag{8.18}$$

is an equality on the event $\{X \le q_\delta(X)\}$, because all terms in (8.18) are zero. It is also an equality on $\{X \ge q\}$ since it reduces to (8.17), because the positive parts $(\,\cdot\,)^+$ are superfluous and the indicator function attains 1. On the remaining event $\{q_\delta(X) < X < q\}$ there is strict inequality in (8.18) because it arises from (8.17) using $X - q < 0 = (X - q)^+$. Adding $q_\delta(X)(1-\delta)$ to both sides of (8.18) and taking expectations, it follows that

$$\begin{aligned}
\mathrm{ES}_\delta[X](1-\delta) &= q_\delta(X)(1-\delta) + \mathbb{E}\big[(X - q_\delta(X))^+\big] \quad \text{by (8.8)}\\
&\le q_\delta(X)(1-\delta) + \underbrace{(q - q_\delta(X))}_{>0}\underbrace{\mathbb{P}[X > q_\delta(X)]}_{\le 1-\delta \text{ by (8.1)}} + \mathbb{E}\big[(X-q)^+\big]\\
&\le q(1-\delta) + \mathbb{E}\big[(X-q)^+\big]
\end{aligned}$$

with equality if and only if $\mathbb{P}[q_\delta(X) < X < q] = 0$ and $\mathbb{P}[X \le q_\delta(X)] = \delta$, which by (8.1) and (8.2) is equivalent to $q_\delta(X) < q \le q^\delta(X)$.

Finally, consider the case $q < q_\delta(X)$. Then the inequality[54]

$$(X - q_\delta(X))^+ \le (q - q_\delta(X))\mathbb{1}_{\{X \ge q_\delta(X)\}} + (X - q)^+ \tag{8.19}$$

is an equality on $\{X \le q\}$, because all terms in (8.19) are zero, and also an equality on $\{X \ge q_\delta(X)\}$, because it agrees with (8.17). On the remaining event $\{q < X < q_\delta(X)\}$ there is strict inequality in (8.19), because $X - q > 0$ and the other two terms are zero. In a similar way as above, using the expectation of (8.19) for the first inequality, it follows that

$$\begin{aligned}
\mathrm{ES}_\delta[X](1-\delta) &= q_\delta(X)(1-\delta) + \mathbb{E}\big[(X - q_\delta(X))^+\big] \quad \text{by (8.8)}\\
&\le q_\delta(X)(1-\delta) + \underbrace{(q - q_\delta(X))}_{<0}\underbrace{\mathbb{P}[X \ge q_\delta(X)]}_{\ge 1-\delta \text{ by (8.1)}} + \mathbb{E}\big[(X-q)^+\big]\\
&\le q(1-\delta) + \mathbb{E}\big[(X-q)^+\big]
\end{aligned}$$

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 (8.1), these two conditions cannot be satisfied simultaneously for $q < q_\delta(X)$, hence equality is impossible.

(h) The lower bound together with the discussion of equality follows directly from the alternative representation (8.8), the upper bound follows from (g).

[54] Note that (8.18) differs from (8.19) in an important point for estimates afterwards.

(i) 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 $\mathbb{P}[U \le u] = u$ 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

$$q_u(X) \le x \implies \mathbb{P}[X \le x] \ge u \qquad \text{and} \qquad q_u(X) > x \implies \mathbb{P}[X \le x] < u$$

by the definition of the lower quantile in (8.1), hence both implication are in fact equivalences and $\{q_U(X) \le x\} = \{U \le \mathbb{P}[X \le x]\}$. Therefore,

$$\mathbb{P}[q_U(X) \le x] = \mathbb{P}[U \le \mathbb{P}[X \le x]] = \mathbb{P}[X \le x]$$

for all $x \in \mathbb{R}$, meaning that $q_U(X)$ and X have the same distribution.

Define $\delta' = \mathbb{P}[X \le q_\delta(X)]$. Note that $\delta' \ge \delta$ and $q_u(X) = q_\delta(X)$ for every $u \in [\delta, \delta']$. Using the second of the above equivalences for $x = q_\delta(X)$ shows that $\{q_U(X) > q_\delta(X)\} = \{U > \delta'\}$. Therefore,

$$\begin{aligned}\int_{[\delta,1)} q_u(X)\,\mathrm{d}u &= \int_{(\delta',1)} q_u(X)\,\mathrm{d}u + \int_{[\delta,\delta']} q_u(X)\,\mathrm{d}u \\ &= \mathbb{E}\big[q_U(X)\mathbb{1}_{\{U>\delta'\}}\big] + q_\delta(X)(\delta' - \delta) \\ &= \mathbb{E}\big[X\mathbb{1}_{\{X>q_\delta(X)\}}\big] + q_\delta(X)(\mathbb{P}[X \le q_\delta(X)] - \delta).\end{aligned}$$

Division by $1-\delta$ gives the right-hand side of (8.6), which is the result (8.15).

(j) By translation invariance from (b), we may assume without loss of generality that every X_n is non-negative. Using the density f_X from (8.9), the representation of expected shortfall with the density f_X given in (8.12), Fatou's lemma for $(X_n f_X)_{n\in\mathbb{N}}$ and the scenario representation from (c), we get

$$\mathrm{ES}_\delta[X] = \mathbb{E}[X f_X] \le \liminf_{n\to\infty} \underbrace{\mathbb{E}[X_n f_X]}_{\le \mathrm{ES}_\delta[X_n]}.$$

(k) By passing to a subsequence if necessary, we may assume that the sequence $(\mathrm{ES}_\delta[X_n])_{n\in\mathbb{N}}$ converges to the limit inferior in (8.16). By passing to a further subsequence if necessary, we may assume that $(X_n)_{n\in\mathbb{N}}$ converges almost surely to X. Now, (8.16) 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 8.25 (Expected shortfall and Wasserstein distance)**.** *Let X and Y be real-valued, integrable random variables and denote the Wasserstein distance of their distributions by $d_{\mathrm{W}}(\mathcal{L}(X), \mathcal{L}(Y))$. Then the expected shortfall of X and Y satisfies, for every level $\delta \in (0,1)$,*

$$\big|\mathrm{ES}_\delta[X] - \mathrm{ES}_\delta[Y]\big| \le \frac{d_{\mathrm{W}}(\mathcal{L}(X), \mathcal{L}(Y))}{1-\delta}. \tag{8.20}$$

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

$$a = \inf_{i\in I} a_i, \qquad b = \inf_{i\in I} b_i \qquad \text{and} \qquad c = \sup_{i\in I} |a_i - b_i|.$$

Then $a_i \le b_i + c$ for every $i \in I$, hence $a \le b + c$. Similarly $b \le a + c$, hence $|a - b| \le c$. Using this observation, the integrability of X and Y, and the minimization property from Theorem 8.20(g), it follows that

$$\bigl|\mathrm{ES}_\delta[X] - \mathrm{ES}_\delta[Y]\bigr| \le \frac{1}{1-\delta} \sup_{q\in\mathbb{R}} \bigl|\mathbb{E}\bigl[(X-q)^+\bigr] - \mathbb{E}\bigl[(Y-q)^+\bigr]\bigr|.$$

For every $q \in \mathbb{R}$, the function $\mathbb{R} \ni x \mapsto f_q(x) := (x-q)^+$ is Lipschitz continuous with constant 1, hence (8.20) follows directly from the lower bound (3.17). □

8.3 Contributions to Expected Shortfall

8.3.1 Definition and Representation with a Density

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 $X\colon \Omega \to \mathbb{R}$ on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Let $\mathcal{L}_1^-(\mathbb{P})$ denote the set of all those $X \in \mathcal{L}_0(\mathbb{P})$, for which the negative part $X^- := \max\{0, -X\}$ is $\mathbb{P}$-integrable. Since $(\alpha X)^- = \alpha X^-$ and $(X+Y)^- \le X^- + Y^-$ for all $\alpha \in [0,\infty)$ and $X, Y \in \mathcal{L}_0(\mathbb{P})$, it follows that $\mathcal{L}_1^-(\mathbb{P})$ is a convex cone. 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 $\mathrm{ES}_\delta[Z]$ to the n subportfolios in a fair and risk-adequate way.

Definition 8.26 (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[55] $X\mathbb{1}_{\{Z \ge q_\delta(Z)\}} \in \mathcal{L}_1^-(\mathbb{P})$. Then the expected shortfall contribution of the subportfolio loss X to Z at level δ is defined by

$$\mathrm{ES}_\delta[X, Z] = \frac{\mathbb{E}[X\mathbb{1}_{\{Z > q_\delta(Z)\}}] + \beta_Z\, \mathbb{E}[X\mathbb{1}_{\{Z = q_\delta(Z)\}}]}{1-\delta} \tag{8.21}$$

with β_Z as in (8.10), i. e.

$$\beta_Z := \begin{cases} \frac{\mathbb{P}[Z \le q_\delta(Z)] - \delta}{\mathbb{P}[Z = q_\delta(Z)]} & \text{if } \mathbb{P}[Z = q_\delta(Z)] > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{8.22}$$

[55] We are quite general here to state the consistency property in Theorem 8.30(a) below without any integrability assumptions on Z.

Remark 8.27. Note that $\mathrm{ES}_\delta[X,Z]=\infty$ is possible and that the condition $X\mathbb{1}_{\{Z\ge q_\delta(Z)\}}\in\mathcal{L}_1^-(\mathbb{P})$ is certainly satisfied for all $X\in\mathcal{L}_1^-(\mathbb{P})$.

Remark 8.28 (Simple case of expected shortfall contribution)**.** If $\mathbb{P}[Z\le q_\delta(Z)]=\delta$, then $\beta_Z=0$ by (8.22), and (8.21) simplifies to

$$\mathrm{ES}_\delta[X,Z]=\mathbb{E}[X\,|\,Z>q_\delta(Z)]$$

simultaneously for all $X\in\mathcal{L}_0(\mathbb{P})$ with $X\mathbb{1}_{\{Z\ge q_\delta(Z)\}}\in\mathcal{L}_1^-(\mathbb{P})$, cf. Remark 8.14. Therefore, $\mathrm{ES}_\delta[X,Z]$ is the conditional expectation of the subportfolio loss X given a large portfolio loss Z occurs. This allocation principle was already presented in [50].

Remark 8.29 (Representation of expected shortfall contributions with a Z-adjusted probability measure)**.** With the density f_Z defined as in (8.9), we get in the setting of Definition 8.26 the representation $\mathrm{ES}_\delta[X,Z]=\mathbb{E}[Xf_Z]$.

8.3.2 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 [32].

Theorem 8.30 (Properties of expected shortfall contributions)**.** *For each level $\delta\in(0,1)$, the expected shortfall contributions have, for all $X,Y\in\mathcal{L}_1^-(\mathbb{P})$ and $Z\in\mathcal{L}_0(\mathbb{P})$, the following properties:*

(a) Consistency with expected shortfall: $\mathrm{ES}_\delta[Z,Z]=\mathrm{ES}_\delta[Z]$.

(b) Diversification: $\mathrm{ES}_\delta[X,Z]\le\mathrm{ES}_\delta[X,X]$.

(c) Linearity: For all $\alpha,\beta>0$,

$$\mathrm{ES}_\delta[\alpha X+\beta Y,Z]=\alpha\,\mathrm{ES}_\delta[X,Z]+\beta\,\mathrm{ES}_\delta[Y,Z].$$

If $X,Y\in\mathcal{L}_1(\mathbb{P})$, then the equality holds for all $\alpha,\beta\in\mathbb{R}$.

(d) Translation (or cash) invariance: If $a\in\mathbb{R}$, then

$$\mathrm{ES}_\delta[X+a,Z]=\mathrm{ES}_\delta[X,Z]+a.$$

(e) Monotonicity: If $X\le Y$, then $\mathrm{ES}_\delta[X,Z]\le\mathrm{ES}_\delta[Y,Z]$.

(f) Independence: If X and Z are independent, then $\mathrm{ES}_\delta[X,Z]=\mathbb{E}[X]$.

(g) Invariance of portfolio scale: $\mathrm{ES}_\delta[X,\alpha Z]=\mathrm{ES}_\delta[X,Z]$ *for all $\alpha>0$.*

(h) Subportfolio continuity: If $Y\in\mathcal{L}_1(\mathbb{P})$, then

$$\big|\mathrm{ES}_\delta[X,Z]-\mathrm{ES}_\delta[Y,Z]\big|\le\mathrm{ES}_\delta[|X-Y|,Z]\le\frac{\mathbb{E}[|X-Y|]}{1-\delta}.$$

(i) Portfolio continuity: Suppose that $X \in \mathcal{L}_1(\mathbb{P})$. If $\mathbb{P}[Z \le q_\delta(Z)] = \delta$ or if X is almost surely constant on $\{Z = q_\delta(Z)\}$, then capital allocation for X by expected shortfall at level δ is continuous at Z, i.e., for every sequence $(Z_n)_{n\in\mathbb{N}}$ in $\mathcal{L}_0(\mathbb{P})$ converging to Z in probability,

$$\lim_{n\to\infty} \mathrm{ES}_\delta[X, Z_n] = \mathrm{ES}_\delta[X, Z]. \tag{8.23}$$

(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 $Z \in \mathcal{L}_1(\mathbb{P})$ as specified in part (i), then

$$\mathrm{ES}_\delta[X, Z] = \lim_{\varepsilon\to 0} \frac{\mathrm{ES}_\delta[Z + \varepsilon X] - \mathrm{ES}_\delta[Z]}{\varepsilon}. \tag{8.24}$$

Remark 8.31 (Discussion of the allocation properties)**.** Let us expand on some of the given item titles in Theorem 8.30:

- 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 independence in (f) can be satisfied, when X and also $-X$ are contained in Z. Think of an financial option's payoff X together with its hedge delivering $-X$, or of an insurance risk X, which is transfered to a reinsurance company.
- For item (i) think of a reinsurance company operating with one-year contracts. During the renewal phase, the company plans to have the portfolio risk Z, and determines the individual contribution of a reinsurance contract with risk X accordingly with $\mathrm{ES}_\delta[X, Z]$. At the end of the renewal phase, the company ends up with a portfolio risk Z_n, which is close but not identical to the planned Z. In this case, the limit relation (8.23) gives a connection between the true contribution $\mathrm{ES}_\delta[X, Z_n]$ and the planned $\mathrm{ES}_\delta[X, Z]$. The proof of (i) is due to the author.

Example 8.32 (A counterexample to (8.23) and (8.24))**.** 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, $\mathrm{ES}_\delta[X, Z] = \mathbb{E}[X] = 1 - \delta$ by independence, see (f), $\mathrm{ES}_\delta[X, Z_\varepsilon] = \mathrm{ES}_\delta[X, X] = \mathrm{ES}_\delta[X] = 1$ for all $\varepsilon > 0$ by scale invariance (g), consistency (a), and Remark 8.14 using $q_\delta(X) = 0$. Therefore, (8.23) is violated. Since $\mathrm{ES}_\delta[Z] = 0$ and $\mathrm{ES}_\delta[Z + \varepsilon X] = \varepsilon\, \mathrm{ES}_\delta[X] = \varepsilon$, the directional derivative in (8.24) equals $1 \ne 1 - \delta = \mathrm{ES}_\delta[X, Z]$, hence (8.24) is violated.

Proof of Theorem 8.30. (a) By Remark 8.29 and (8.12),

$$\mathrm{ES}_\delta[Z, Z] = \mathbb{E}[Z f_Z] = \mathrm{ES}_\delta[Z].$$

(b) By Remark 8.29, the scenario representation from Theorem 8.20(c), and finally item (a),

$$\mathrm{ES}_\delta[X,Z] = \mathbb{E}[Xf_Z] \le \sup_{f\in\mathcal{F}_\delta} \mathbb{E}[Xf] = \mathrm{ES}_\delta[X] = \mathrm{ES}_\delta[X,X].$$

(c), (d) follow from Remark 8.29 and the linearity of the expectation.

(e) follows from Remark 8.29 and $\mathrm{ES}_\delta[X,Z] = \mathbb{E}[Xf_Z] \le \mathbb{E}[Yf_Z] = \mathrm{ES}_\delta[Y,Z]$.

(f) By Remark 8.29, $\mathrm{ES}_\delta[X,Z] = \mathbb{E}[Xf_Z] = \mathbb{E}[X]\,\mathbb{E}[f_Z] = \mathbb{E}[X]$.

(g) Since $q_\delta(\alpha Z) = \alpha q_\delta(Z)$, the definition (8.9) implies $f_{\alpha Z} = f_Z$. Hence, by Remark 8.29,

$$\mathrm{ES}_\delta[X,\alpha Z] = \mathbb{E}[Xf_{\alpha Z}] = \mathbb{E}[Xf_Z] = \mathrm{ES}_\delta[X,Z].$$

(h) For the first inequality use linearity (c) and monotonicity (e), for the second one use Remark 8.29 and the upper bound $1/(1-\delta)$ for the density f_Z.

(i) Since the proof is longer, let us first reduce the problem. Given $X \in \mathcal{L}_1(\mathbb{P})$ and $\varepsilon > 0$, there exists by the dominated convergence theorem a constant M such that the bounded random variable $X_\varepsilon := X\mathbb{1}_{\{|X|\le M\}}$ satisfies $\mathbb{E}[|X - X_\varepsilon|] = \mathbb{E}\big[|X|\mathbb{1}_{\{|X|>M\}}\big] \le \varepsilon$. By the subportfolio continuity (h), it therefore suffices to prove (8.23) for all bounded $X \in \mathcal{L}_1(\mathbb{P})$.

To simplify the notation for the quantiles, define $q = q_\delta(Z)$ and $q_n = q_\delta(Z_n)$. Without loss of generality we may assume that $\mathbb{E}[X\mathbb{1}_{\{Z=q\}}] = 0$, because in the case $\mathbb{P}[Z = q] > 0$ we could, using cash invariance (d), switch to $X' := X - a$ with $a := \mathbb{E}[X\,|\,Z = q]$. This simplifies (8.21). By linearity (c), we may restrict our attention to those $X \in \mathcal{L}_1(\mathbb{P})$ which are bounded by $1 - \delta$.

For $\varepsilon > 0$, we now set up $\eta > 0$ and $n_\varepsilon \in \mathbb{N}$. By the right-continuity of the distribution function of $|Z - q|$, there exists $\eta > 0$ such that

$$\mathbb{P}[0 < |Z - q| < 2\eta] \le \varepsilon. \tag{8.25}$$

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

$$\mathbb{P}[\,|Z - Z_n| \ge \eta] \le \varepsilon \qquad \text{for all } n \ge n_\varepsilon \tag{8.26}$$

and, by Lemma 8.6(a),

$$q_n \ge q - \eta \qquad \text{for all } n \ge n_\varepsilon. \tag{8.27}$$

We will show below by considering the cases $q_n \le q + \eta$ and $q_n > q + \eta$ that

$$\big|\mathrm{ES}_\delta[X, Z_n] - \mathrm{ES}_\delta[X,Z]\big| \le 6\varepsilon \tag{8.28}$$

for every $n \ge n_\varepsilon$. Since $\varepsilon > 0$ is arbitrary, (8.28) implies the desired result (8.23). Note that $\mathbb{E}[|\mathbb{1}_A - \mathbb{1}_B|] = \mathbb{P}[A \cap B^{\mathrm{c}}] + \mathbb{P}[A^{\mathrm{c}} \cap B]$ for all $A, B \in \mathcal{A}$.

Case I: The proof of (8.28) for the case $q_n > q + \eta$ is the easier one and doesn't use the additional assumptions given in (i). Note that

$$\begin{aligned}(1-\beta_{Z_n})\,\mathbb{E}\big[\mathbb{1}_{\{Z_n=q_n\}}\big] &= \delta - \mathbb{P}[Z_n < q_n]\\ &\le \mathbb{P}[Z \le q] - \mathbb{P}[Z_n < q_n]\\ &\le \mathbb{P}[Z \le q,\, Z_n \ge q_n] \le \varepsilon\end{aligned} \tag{8.29}$$

by (8.22) and (8.26). By partitioning $\{Z_n \ge q_n\}$, we obtain

$$1-\delta \le \mathbb{P}[Z_n \ge q_n] = \mathbb{P}[\underbrace{Z > q,\, Z_n \ge q_n}_{=:A}] + \underbrace{\mathbb{P}[\overbrace{Z \le q,\, Z_n \ge q_n}^{=:B}]}_{\le \varepsilon \text{ by } (8.26)},$$

hence $\mathbb{P}[A] \ge 1-\delta-\varepsilon$. Partitioning $\{Z > q\}$ yields

$$1-\delta \ge \mathbb{P}[Z > q] = \mathbb{P}[A] + \mathbb{P}[\underbrace{Z > q,\, Z_n < q_n}_{=:C}],$$

thus $\mathbb{P}[C] \le \varepsilon$. Finally, using (8.21), $\mathbb{E}[X\mathbb{1}_{\{Z=q\}}] = 0$, and $\|X\|_\infty \le 1-\delta$,

$$\begin{aligned}&\big|\mathrm{ES}_\delta[X, Z_n] - \mathrm{ES}_\delta[X, Z]\big|\\ &\qquad \le \underbrace{(1-\beta_{Z_n})\,\mathbb{E}\big[\mathbb{1}_{\{Z_n=q_n\}}\big]}_{\le \varepsilon \text{ by } (8.29)} + \underbrace{\mathbb{E}\big[|\mathbb{1}_{\{Z_n \ge q_n\}} - \mathbb{1}_{\{Z>q\}}|\big]}_{=\mathbb{P}[B]+\mathbb{P}[C]} \le 3\varepsilon,\end{aligned}$$

which proves (8.28) for the case $q_n > q + \eta$.

Case II: We will now prove estimate (8.28) in the case $q_n \le q + \eta$ for the two different assumptions given in Theorem 8.30(i). Define $E = \{Z > q,\, Z_n \le q_n\}$ and $F = \{Z \le q,\, Z_n > q_n\}$. Note that

$$\mathbb{P}[E] = \underbrace{\mathbb{P}[q < Z < q^+,\, Z_n \le q_n]}_{\le \varepsilon \text{ by } (8.25)} + \underbrace{\mathbb{P}[Z \ge q^+,\, Z_n \le q_n]}_{\le \varepsilon \text{ by } (8.26)} \le 2\varepsilon. \tag{8.30}$$

Case II(a): Let the assumption $\mathbb{P}[Z \le q] = \delta$ be satisfied. By partitioning $\{Z_n \le q_n\}$, we obtain

$$\delta \le \mathbb{P}[Z_n \le q_n] = \mathbb{P}[\underbrace{Z \le q,\, Z_n \le q_n}_{=:D}] + \mathbb{P}[E]\,,$$

hence $\mathbb{P}[D] \ge \delta - 2\varepsilon$ by (8.30). Partitioning $\{Z \le q\}$ yields

$$\delta = \mathbb{P}[Z \le q] = \mathbb{P}[D] + \mathbb{P}[F],$$

thus $\mathbb{P}[D] \le \delta$ and $\mathbb{P}[F] \le 2\varepsilon$. Furthermore, using (8.30)

$$\beta_{Z_n}\,\mathbb{E}\big[\mathbb{1}_{\{Z_n=q_n\}}\big] = \mathbb{P}[Z_n \le q_n] - \delta = \mathbb{P}[D] + \mathbb{P}[E] - \delta \le 2\varepsilon. \tag{8.31}$$

Finally, using (8.21), $\mathbb{E}[X\mathbb{1}_{\{Z=q\}}]=0$, and $\|X\|_\infty \le 1-\delta$,

$$\bigl|\mathrm{ES}_\delta[X,Z_n]-\mathrm{ES}_\delta[X,Z]\bigr| \le \underbrace{\beta_{Z_n}\,\mathbb{E}\bigl[\mathbb{1}_{\{Z_n=q_n\}}\bigr]}_{\le 2\varepsilon \text{ by } (8.31)} + \underbrace{\mathbb{E}\bigl[|\mathbb{1}_{\{Z_n>q_n\}}-\mathbb{1}_{\{Z>q\}}|\bigr]}_{=\mathbb{P}[E]+\mathbb{P}[F]\le 4\varepsilon \text{ using } (8.30)} \le 6\varepsilon,$$

which proves (8.28) for the Case II(a).

Case II(b): Let now X be a. s. constant on $\{Z=q\}$. Then $\mathbb{E}[X\mathbb{1}_{\{Z=q\}}]=0$ implies $\mathbb{E}[|X|\mathbb{1}_{\{Z=q,\,Z_n=q_n\}}]=0$ and $\mathbb{E}[|X|\mathbb{1}_{\{Z=q,\,Z_n>q_n\}}]=0$. Therefore,

$$\begin{aligned}\frac{\mathbb{E}\bigl[|X|\mathbb{1}_{\{Z_n=q_n\}}\bigr]}{1-\delta} &= \frac{\mathbb{E}\bigl[|X|\mathbb{1}_{\{Z\ne q,\,Z_n=q_n\}}\bigr]}{1-\delta}\\ &\le \mathbb{P}[\,Z\ne q,\,Z_n=q_n]\\ &\le \underbrace{\mathbb{P}[0<|Z-q|<2\eta]}_{\le\varepsilon \text{ by } (8.25)} + \underbrace{\mathbb{P}[\,|Z-q|\ge 2\eta,\,Z_n=q_n]}_{\le\varepsilon \text{ by } (8.26) \text{ and } (8.27)} \le 2\varepsilon\end{aligned} \tag{8.32}$$

and

$$\begin{aligned}\frac{\mathbb{E}[|X|\mathbb{1}_F]}{1-\delta} &\le \mathbb{P}[Z<q,\,Z_n>q_n]\\ &= \underbrace{\mathbb{P}\bigl[q^-<Z<q,\,Z_n>q_n\bigr]}_{\le\varepsilon \text{ by } (8.25)} + \underbrace{\mathbb{P}\bigl[Z\le q^-,\,Z_n>q_n\bigr]}_{\le\varepsilon \text{ by } (8.26) \text{ and } (8.27)} \le 2\varepsilon.\end{aligned} \tag{8.33}$$

Using (8.21), $\beta_{Z_n}\in[0,1]$, $\mathbb{E}[X\mathbb{1}_{\{Z=q\}}]=0$, and $\|X\|_\infty\le 1-\delta$,

$$\bigl|\mathrm{ES}_\delta[X,Z_n]-\mathrm{ES}_\delta[X,Z]\bigr| \le \underbrace{\frac{\mathbb{E}\bigl[|X|\mathbb{1}_{\{Z_n=q_n\}}\bigr]}{1-\delta}}_{\le 2\varepsilon \text{ by } (8.32)} + \underbrace{\frac{\mathbb{E}\bigl[|X|\mathbb{1}_E\bigr]}{1-\delta}}_{\le 2\varepsilon \text{ by } (8.30)} + \underbrace{\frac{\mathbb{E}\bigl[|X|\mathbb{1}_F\bigr]}{1-\delta}}_{\le 2\varepsilon \text{ by } (8.33)} \le 6\varepsilon,$$

which proves (8.28) for the Case II(b).

(j) Let $\varepsilon>0$. By consistency (a), diversification (b) and linearity (c),

$$\mathrm{ES}_\delta[Z+\varepsilon X]=\mathrm{ES}_\delta[Z+\varepsilon X,Z+\varepsilon X]\ge \mathrm{ES}_\delta[Z+\varepsilon X,Z]=\mathrm{ES}_\delta[Z]+\varepsilon\,\mathrm{ES}_\delta[X,Z],$$

hence

$$\frac{\mathrm{ES}_\delta[Z+\varepsilon X]-\mathrm{ES}_\delta[Z]}{\varepsilon}\ge \mathrm{ES}_\delta[X,Z].$$

Similarly,

$$\mathrm{ES}_\delta[Z]=\mathrm{ES}_\delta[Z,Z]\ge \mathrm{ES}_\delta[Z,Z+\varepsilon X]=\mathrm{ES}_\delta[Z+\varepsilon X]-\varepsilon\,\mathrm{ES}_\delta[X,Z+\varepsilon X],$$

hence

$$\mathrm{ES}_\delta[X,Z+\varepsilon X]\ge \frac{\mathrm{ES}_\delta[Z+\varepsilon X]-\mathrm{ES}_\delta[Z]}{\varepsilon}.$$

Since capital allocation for X by expected shortfall is assumed to be continuous at Z,

$$\mathrm{ES}_\delta[X,Z]=\lim_{\varepsilon\searrow 0}\frac{\mathrm{ES}_\delta[Z+\varepsilon X]-\mathrm{ES}_\delta[Z]}{\varepsilon}.$$

If $\varepsilon\nearrow 0$, apply this result for $\varepsilon'=-\varepsilon$ and $X'=-X$ and use $-\mathrm{ES}_\delta[X',Z]=\mathrm{ES}_\delta[X,Z]$ to obtain (8.24). $\square$

8.3.3 Calculation of Contributions in Extended CreditRisk$^+$

[56] Let us now apply the idea of risk capital allocation by expected shortfall to the credit portfolio loss L given by (7.14). We also want to calculate this allocation within the extended CreditRisk$^+$ model. If $\mathbb{E}[L] < \infty$, then the definition (8.21) gives

$$\mathrm{ES}_\delta[L_{g,i,k}, L] = \frac{\mathbb{E}\big[L_{g,i,k}\mathbb{1}_{\{L>q_\delta(L)\}}\big] + \beta_L\,\mathbb{E}\big[L_{g,i,k}\mathbb{1}_{\{L=q_\delta(L)\}}\big]}{1-\delta} \tag{8.34}$$

as contribution attributed to obligor $i \in \{1,\dots,m\}$ due to group $g \in G_i$ and risk $k \in \{0,\dots,K\}$ to the expected shortfall $\mathrm{ES}_\delta[L]$. Since L has a discrete distribution, $\mathbb{P}[L = q_\delta(L)] = 0$ is impossible due to the definition of $q_\delta(L)$ in (8.1). Note that, by consistency and linearity of the allocation given in Theorem 8.30(a) and (c),

$$\mathrm{ES}_\delta[L] = \mathrm{ES}_\delta[L, L] = \sum_{i=1}^{m}\sum_{g\in G_i}\sum_{k=0}^{K}\mathrm{ES}_\delta[L_{g,i,k}, L].$$

Since

$$\mathbb{E}\big[L_{g,i,k}\mathbb{1}_{\{L>q_\delta(L)\}}\big] = \underbrace{\mathbb{E}[L_{g,i,k}]}_{=\lambda_g w_{g,k}\,\mathbb{E}[L_{g,i,k,1}]} - \mathbb{E}\big[L_{g,i,k}\mathbb{1}_{\{L\le q_\delta(L)\}}\big]\,,$$

we need to compute $\mathbb{E}[L_{g,i,k}\mathbb{1}_{\{L=l\}}]$ for $l \in \{0,1,\dots,q_\delta(L)\}$. This can be done adapting a lemma by Tasche [54, Section 3.4], which is in turn a generalization of a formula given in [50, Slide 9].

Lemma 8.33. *For every obligor $i \in \{1,\dots,m\}$, every group $g \in G_i$ and total loss $l \in \mathbb{N}_0$,*

$$\mathbb{E}[L_{g,i,0}\mathbb{1}_{\{L=l\}}] = \lambda_g w_{g,0}\sum_{\nu=1}^{l}\mathbb{E}\big[L_{g,i,0,1}\mathbb{1}_{\{L_{g,0,1}=\nu\}}\big]\,\mathbb{P}[L = l-\nu] \tag{8.35}$$

and, for every risk $k \in \{1,\dots,K\}$,

$$\mathbb{E}[L_{g,i,k}\mathbb{1}_{\{L=l\}}] = \lambda_g w_{g,k}\sum_{\nu=1}^{l}\mathbb{E}[L_{g,i,k,1}\mathbb{1}_{\{L_{g,k,1}=\nu\}}]\,\mathbb{E}[\Lambda_k\mathbb{1}_{\{L=l-\nu\}}]. \tag{8.36}$$

Remark 8.34. The algorithm presented in Section 7.7 calculates in a numerically stable way the quantities $\mathbb{P}[L = l-\nu]$ and $\mathbb{E}[\Lambda_k\mathbb{1}_{\{L=l-\nu\}}]$ 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 $\mathbb{E}[\Lambda_k\mathbb{1}_{\{L=l-\nu\}}]$ the coefficients $(c_l)_{\in\mathbb{N}_0}$ given by (**??**) and (**??**) and well as the coefficients $(d_n)_{n\in\mathbb{N}_0}$ given by (**??**) and (**??**) have to be recalculated.

[56] This section has to be adapted to the new notation and the generalized setting.

Remark 8.35. For every obligor $i \in \{1, \dots, m\}$, every group $g \in G_i$, every risk $k \in \{0, \dots, K\}$ and every group loss $\nu \in \mathbb{N}_0$, we get from Assumption 7.11

$$\mathbb{E}[L_{g,i,k,1} \mathbb{1}_{\{L_{g,k,1}=\nu\}}] = \sum_{\substack{\mu=(\mu_j)_{j\in g}\in\mathbb{N}_0^g \\ \|\mu\|_1=\nu}} \mu_i \underbrace{\mathbb{P}[L_{g,j,k,1} = \mu_j \text{ for all } j \in g]}_{= q_{g,k,\mu} \text{ by (7.19)}}, \tag{8.37}$$

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 CreditRisk$^+$ model (cf. Remarks 7.6 and 7.44), the result (8.37) simplifies to

$$\mathbb{E}[L_{g,i,k,1} \mathbb{1}_{\{L_{g,k,1}=\nu\}}] = \nu\, q_{g,k,\nu}. \tag{8.38}$$

(b) If the group loss ν is attributed in a deterministic way to its members as described in Example 7.13, then

$$\mathbb{E}[L_{g,i,k,1} \mathbb{1}_{\{L_{g,k,1}=\nu\}}] = h_{g,i,k}(\nu)\, q^{\mathrm{s}}_{g,k,\nu}. \tag{8.39}$$

(c) Note that by the linearity of the expectation,

$$\nu q^{\mathrm{s}}_{g,k,\nu} = \mathbb{E}[L_{g,k,1} \mathbb{1}_{\{L_{g,k,1}=\nu\}}] = \sum_{i\in g} \mathbb{E}[L_{g,i,k,1} \mathbb{1}_{\{L_{g,k,1}=\nu\}}]. \tag{8.40}$$

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 (8.40) are equal and

$$\mathbb{E}[L_{g,i,k,1} \mathbb{1}_{\{L_{g,k,1}=\nu\}}] = \frac{\nu}{|g|}\, q^{\mathrm{s}}_{g,k,\nu} \qquad \text{for all } i \in g. \tag{8.41}$$

Proof of Lemma 8.33. Fix a risk $k \in \{0, \dots, K\}$, an obligor $i \in \{1, \dots, m\}$ 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 (7.10) 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 $l \ge 1$. 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 $n \in \{1, \dots, \mu\}$ define

$$M_{\mu,n} = \sum_{\substack{r=1 \\ r\neq n}}^{\mu} L_{g,k,r}$$

as the sum of the first μ losses of group g due to risk k, omitting the nth loss. Then

$$\begin{aligned} \mathbb{E}\bigl[L_{g,i,k} \mathbb{1}_{\{L=l\}}\bigr] &= \sum_{\mu=1}^{\infty} \mathbb{E}\biggl[\sum_{n=1}^{\mu} L_{g,i,k,n} \mathbb{1}_{\{L=l,\, N_{g,k}=\mu\}}\biggr] \\ &= \sum_{\mu=1}^{\infty} \sum_{n=1}^{\mu} \sum_{\nu=1}^{l} \mathbb{E}\bigl[L_{g,i,k,n} \mathbb{1}_{\{\underbrace{L=l,\, N_{g,k}=\mu}_{=\{M+M_{\mu,n}+L_{g,k,n}=l,\, N_{g,k}=\mu\}},\, L_{g,k,n}=\nu\}}\bigr]. \end{aligned} \tag{8.42}$$

It follows from Assumption 7.11 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 (7.9) is independent jointly from M, $M_{\mu,n}$ and $N_{g,k}$, hence

$$\begin{aligned}\mathbb{E}\big[L_{g,i,k,n}\mathbb{1}_{\{M+M_{\mu,n}+L_{g,k,n}=l,\,N_{g,k}=\mu,\,L_{g,k,n}=\nu\}}\big]\\ = \mathbb{E}\big[L_{g,i,k,n}\mathbb{1}_{\{L_{g,k,n}=\nu\}}\big]\,\mathbb{P}[M+M_{\mu,n}=l-\nu,\,N_{g,k}=\mu]\,.\end{aligned}\tag{8.43}$$

By Assumption 7.11, 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 (8.43). The same assumption implies that $M_{\mu,n}$ is independent from $(M,N_{g,k})$ and that $M_{\mu,1},\dots,M_{\mu,\mu}$ are identically distributed, hence, for every $n\in\{1,\dots,\mu\}$,

$$\mathbb{P}[M+M_{\mu,n}=l-\nu,\,N_{g,k}=\mu]=\mathbb{P}[M+M_{\mu,\mu}=l-\nu,\,N_{g,k}=\mu]\,.\tag{8.44}$$

Consider now the case $k\in\{1,\dots,K\}$. By the conditional independence from Assumption 7.25 and the conditional Poisson distribution from Assumption 7.24,

$$\begin{aligned}&\mathbb{P}[M+M_{\mu,\mu}=l-\nu,\,N_{g,k}=\mu]\\ &\quad=\mathbb{E}\big[\mathbb{P}[M+M_{\mu,\mu}=l-\nu\,|\,\Lambda_1,\dots,\Lambda_m]\,\mathbb{P}[N_{g,k}=\mu\,|\,\Lambda_k]\big]\\ &\quad=\frac{\lambda_g w_{g,k}}{\mu}\,\mathbb{E}\big[\Lambda_k\,\mathbb{P}[\underbrace{M+M_{\mu,\mu}=l-\nu,\,N_{g,k}=\mu-1}_{=\{L=l-\nu,\,N_{g,k}=\mu-1\}}\,|\,\Lambda_k]\big],\end{aligned}\tag{8.45}$$

where we used

$$\begin{aligned}\mathbb{P}[N_{g,k}=\mu\,|\,\Lambda_k]&\overset{\text{a.s.}}{=}\frac{(\lambda_g w_{g,k}\Lambda_k)^\mu}{\mu!}\exp(-\lambda_g w_{g,k}\Lambda_k)\\ &\overset{\text{a.s.}}{=}\frac{\lambda_g w_{g,k}\Lambda_k}{\mu}\,\mathbb{P}[N_{g,k}=\mu-1\,|\,\Lambda_k]\,.\end{aligned}$$

Substituting (8.43), (8.44) and (8.45) into (8.42) and noting that the sum over $n\in\{1,\dots,\mu\}$ cancels with the denominator μ, we obtain

$$\begin{aligned}\mathbb{E}\big[L_{g,i,k}\mathbb{1}_{\{L=l\}}\big]&=\lambda_g w_{g,k}\sum_{\mu=1}^{\infty}\sum_{\nu=1}^{l}\mathbb{E}\big[L_{g,i,k,1}\mathbb{1}_{\{L_{g,k,1}=\nu\}}\big]\,\mathbb{E}\big[\Lambda_k\mathbb{1}_{\{L=l-\nu,\,N_{g,k}=\mu-1\}}\big]\\ &=\lambda_g w_{g,k}\sum_{\nu=1}^{l}\mathbb{E}\big[L_{g,i,k,1}\mathbb{1}_{\{L_{g,k,1}=\nu\}}\big]\,\mathbb{E}\big[\Lambda_k\mathbb{1}_{\{L=l-\nu\}}\big].\end{aligned}$$

For the case $k=0$ the calculation in the last paragraph is easier and left as an exercise. □

Remark 8.36. As we constructed $N_{i,k}$ as conditionally Poisson distributed random variable, we have that $\mathbb{P}(N_{i,k}\ge n)>0$ for every $n\in\mathbb{N}$. Hence it is possible that the risk contributions become greater than the maximal exposure.

9 Application to Operational Risk

9.1 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 [39, 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 11$^{\text{th}}$, 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 7 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	(5) Payment & settlement	
(2) Trading & sales	(6) Agency services	(9.1)
(3) Retail banking	(7) Asset management	
(4) Commercial banking	(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].

9.2 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 [41] 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, 20] and [39, Chap. 7].

Moscadelli [41] 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.

9.3 Application of the Extended CreditRisk$^+$ Methodology

[57] 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 7:

- The number m of obligors turns into the number of business lines, $m = 8$ for the ones given in (9.1) 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 [7, 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.

[57] This section has to be adapted to the new notation and the generalized setting.

The stochastic losses (within a year) get the following interpretation:

- $L_{g,k}$ given by (7.10) is the operational loss of the group $g \in G$ of business lines due to common losses of type $k \in \{0, \dots, K\}$,
- $L_{i,k}$ given by (7.16) is the operational loss of business line $i \in \{1, \dots, m\}$ due to loss type $k \in \{0, \dots, K\}$,
- L_i given in (7.18) is the total operational loss of business line $i \in \{1, \dots, m\}$, and
- L given by (7.14) is the total operational loss of the bank.

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.

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 [22], was done jointly with Dr. Stefan Gerhold and DI Richard Warnung. The papers [23] and [22] are part of R. Warnung's Ph. D. thesis [58], 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 [47] as well as on approximations of Poisson mixture models via Panjer's recursion [46] leading to her Ph. D. thesis [45]. 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 [30] to extend the results presented in Section 8. 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 8.7) and expected shortfall (Lemma 8.25), and with additional exercise problems, the lecture notes were used again for the same course in 2014.

List of Abbreviated Distributions and Operations

- $\mathrm{Beta}(\alpha, \beta)$, beta distribution, see Definition 2.6
- $\mathrm{BetaBin}(\alpha, \beta, m)$, beta-binomial distribution, see (2.36)
- $\mathrm{Bin}(1, p)$, Bernoulli distribution
- $\mathrm{Bin}(m, p)$, binomial distribution, see (2.9)
- $\mathrm{CLog}(p, Q) \coloneqq \mathrm{Compound}(\mathrm{Log}(p), Q)$, compound logarithmic distribution
- $\mathrm{CNegBin}(\alpha, p, Q)$, compound negative binomial distribution, see page 68
- $\mathrm{Compound}(\mathcal{L}(N), Q)$, general compound distribution, see (4.70)
- $\mathrm{Convex}((p_i, Q_i)_{i\in\{1,\dots,k\}})$, convex combination of distributions, see Example 4.9
- $*$, convolution, see Remark 5.1
- $\mathrm{CPanjer}(a, b, k, Q)$, compound Panjer distribution, see Theorem 5.16
- $\mathrm{CPoisson}(\lambda, Q)$, compound Poisson distribution, see page 68
- $\mathrm{Dirichlet}(\alpha_1, \dots, \alpha_d)$, see Definition 4.26
- $\mathrm{DirichletMultinomial}(\alpha_1, \dots, \alpha_d, m)$, see Definition 4.30

- Gamma(α, β), gamma distribution, see Subsection 4.4
- Log(p), univariate logarithmic distribution, see Example 4.4
- MBin($m, p_1, \ldots, p_d$), multivariate binomial distribution, see (4.94)
- MLog($p_1, \ldots, p_d$), multivariate logarithmic distribution, see Definition 4.49
- MPoisson($G, (\lambda_g)_{g \in G}, m$), multivariate Poisson distribution, see Definition 3.42
- Multinomial($1, p_1, \ldots, p_d$), multivariate Bernoulli distribution, see Example 4.5
- Multinomial($m, p_1, \ldots, p_d$), multinomial distribution, see Example 4.19
- NegBin(α, p), negative binomial distribution, see (4.61)
- NegMult($\alpha, p_1, \ldots, p_d$), negative multinomial distribution, see Definition 4.52
- Panjer(a, b, k), Panjer distribution, see Definition 5.9
- Poisson(λ), Poisson distribution, see Definition 3.1

Reading Assignments, Summer 2023

1. Week (March 2): Until Lemma 2.12
2. Week (March 9): Until Remark 3.15
3. Week (March 16): Until Exercise 3.33
4. Week (March 23): Until the end of Subsection 3.4.2
5. Week (March 30): Only exercise presentations
6. Week (April 6): Until the end of Subsection 3.6.3
7. Week (April 27): Subsection 8.1 until Example 8.9, Subsection 8.2 until Theorem 8.20(g) with proofs; Remark 8.19 was omitted.
8. Week (May 4): Remaining part of Subsection 8.2, Section 4 until Example 4.5
9. Week (May 11): Continuation of Section 4 until Example 4.25
10. Week (May 18): Until the end of Subsection 4.5
11. Week (May 25): Until Definition 4.52
12. Week (June 1): Until Exercise 4.62
13. Week (June 8): Until Remark 5.19
14. Week (June 15): Until Theorem 5.30 (without proof)
15. Week (June 22): Proof of Theorem 5.30, Subsection 6.1

Reading Assignments, Summer 2025

1. Week (March 5): Until Lemma 2.12
2. Week (March 12): Lecture online via Zoom, until Remark 3.15
3. Week (March 19): Until Exercise 3.22, discussion of Exercises 2.2 and 2.3
4. Week (March 26): Until Exercise 3.35, discussion of Exercises 2.4 and 2.5
5. Week (April 2): Until Lemma 3.41, discussion of Exercises 2.7, 2.8, 2.9
6. Week (April 9): No lecture
7. Week (April 30): Subsections 3.5 and 3.6
8. Week (May 7): Start of Section 8 until Example 8.9, discussion of Exercise 2.10 (and discussion about internships and job applications)
9. Week (May 14): Subsections 8.2.1 and 8.2.3, discussion of Exercise 3.13
10. Week (May 21): Subsection 8.3 until the proof of Theorem 8.30(h), Section 4 until Remark 4.10
11. Week (May 28): Example 4.11 until Lemma 4.29, discussion of Exercises 3.19 and 3.20(a)
12. Week (June 4): Definition 4.30 until Example 4.39, discussion of the remaining items of Exercise 3.20
13. Week (June 11): Footnote of Example 4.39 until the end of Section 4 (omitting Subsubsection 4.7.3), discussion of Exercise 3.21
14. Week (June 18): Section 5 until the statement of Theorem 5.30(a), discussion of Exercises 3.30 and 3.31
15. Week (June 25):

Recent Changes

March 2021

- Remark 5.10 added for clarity.

April 2021

- Exercise 3.44, which proves Lemma 3.43, is given under additional assumptions. The full proof is added as Exercise 4.33.
- Remark 3.51 is added.
- Lemma 4.15 is added.

- Subsection 4.2 is introduced, Examples 4.24 and 4.25 as well as Definition 4.26, Remark 4.27, Exercise 4.28, Lemma 4.29, Definition 4.30 and Exercise 4.31 are added.
- Subsection 4.3 extended, Example 4.32 shifted to it.
- List of abbreviated distributions started.

May 2021

- Example 4.9 is added.
- Subsection 5.1 is new, consisting mainly of material contained previously in Subsection 5.2. Remark 5.1, Algorithm 5.2 and Example 5.3 are revised, Exercise 5.4 is added.
- Remark 5.31 and Exercise 5.33 are added.
- Section 6 is added, starting with stochastic rounding which was previously contained in Subsection 7.2.

May 2022

- Subsection 4.2 is subdivided.
- Exercise 4.50(b) with more explicit covariance matrix.
- Corollary 4.61 and Exercise 4.62 are added.

May 2023

- Theorem 5.6(a) with proof is added.
- Remark 5.7(c) and corresponding entries in Table 5.1 are added.
- Remark 5.17 is added.

March 2025

- Subsection 3.2 is extended by two calibration methods.
- Subsubsection 3.4.1 is revised for a better coverage of bounds for biased Poisson approximations.

April 2025

- Exercise 2.1 and its application (2.28) are added.

References

[1] C. Acerbi and D. Tasche, *On the coherence of expected shortfall*, J. of Banking and Finance **26** (2002), 1487–1503. 159

[2] H. Ammeter, *A generalization of the collective theory of risk in regard to fluctuating basic-probabilities*, Skand. Aktuarietidskr. **31** (1948), 171–198. 69

[3] P. Artzner, F. Delbaen, J. M. Eber, and D. Heath, *Coherent measures of risk*, Math. Finance **9** (1999), no. 3, 203–228. 159

[4] A. D. Barbour and P. Hall, *On the rate of Poisson convergence*, Math. Proc. Cambridge Philos. Soc. **95** (1984), no. 3, 473–480. 26

[5] A. D. Barbour, L. Holst, and S. Janson, *Poisson Approximation*, Oxford Studies in Probability, 1992. 38

[6] A. D. Barbour and A. Xia, *On Stein's factors for Poisson approximation in Wasserstein distance*, Bernoulli **12** (2006), no. 6, 943–954. 38

[7] Basel Committee of Banking Supervision, *Basel II: International convergence of capital measurement and capital standards: a revised framework*, Available at www.bis.org/publ/bcbs118.htm, Nov. 2005. 172, 173, 174

[8] ———, *Working paper on the regulatory treatment of operational risk*, Available at www.bis.org/publ/bcbs_wp8.pdf, Sept. 2001. 173

[9] C. Bluhm, L. Overbeck, and C. Wagner, *An Introduction to Credit Risk Modeling*, Chapman & Hall, 2003. 5

[10] L. H. Y. Chen, L. Goldstein, and Q.-M. Shao, *Normal Approximation by Stein's Method*, Springer-Verlag, Heidelberg, 2011. 38

[11] Credit Suisse First Boston, *CreditRisk$^+$: a credit risk management framework*, Available at www.csfb.com/institutional/research/assets/creditrisk.pdf, 1997. 111

[12] E. De Giorgi and J. Burkhard, *An intensity based non-parametric default model for residential mortgage portfolios*, Journal of Risk **8** (2006), no. 4, 57–95. 116

[13] E. De Giorgi, V. Komaric, and J. Burkhard, *Default risk for residential mortgage portfolios*, Wilmott Magazine (2004), 78–86. 116

[14] F. Delbaen, *Coherent risk measures on general probability spaces*, Advances in Finance and Stochastics. Essays in Honour of Dieter Sondermann (K. Sandmann and Ph. J. Schönbucher, eds.), Springer-Verlag, Berlin, 2002, pp. 1–37. 160

[15] P. Eichelsbacher, *Die Steinsche Methode*, Skripten zur Mathematischen Statistik, vol. 38, Gesellschaft zur Förderung der Mathematischen Statistik Münster, Einsteinstr. 62, D-48149 Münster, 2003 (German). 38

[16] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modelling Extremal Events for Insurance and Finance*, Applications of Mathematics (New York), vol. 33, Springer-Verlag, Berlin, 1997. 173

[17] S. N. Ethier and T. G. Kurtz, *Markov Processes, Characterization and Convergence*, John Wiley & Sons, New York, 1986. 25, 26

[18] L. Euler, *Introductio in Analysin Infinitorum*, vol. 1, Lausanne, 1748, Available at math.dartmouth.edu/∼euler/pages/E101.html. 85, 97

[19] L. Fabrykowski, *Extended CreditRisk$^+$ with Guarantees*, Master's thesis, TU Wien, 2017, Advisor: U. Schmock. 123

[20] M. Falk, J. Hüsler, and R.-D. Reiss, *Laws of Small Numbers: Extremes and Rare Events*, extended ed., Birkhäuser Verlag, Basel, 2004. 173

[21] H. Föllmer and A. Schied, *Convex measures of risk and trading constraints*, Finance Stoch. **6** (2002), 429–447. 159

[22] S. Gerhold, U. Schmock, and R. Warnung, *A generalization of Panjer's recursion and numerically stable risk aggregation*, Finance Stoch. **14** (2010), no. 1, 81–128. 96, 97, 100, 101, 103, 104, 144, 176

[23] S. Gerhold and R. Warnung, *Finding efficient recursions for risk aggregation by computer algebra*, J. Comput. Appl. Math. **223** (2009), no. 1, 499–507. 176

[24] A. L. Gibbs and F. E. Su, *On choosing and bounding probability metrics*, Int. Stat. Rev. **70** (2002), no. 3, 419–435. 20

[25] G. Giese, *Enhancing CreditRisk$^+$*, Risk Magazine **16** (2003), no. 4, 73–77. 144

[26] H. W. Gould, *Coefficient identities for powers of Taylor and Dirichlet series*, Amer. Math. Monthly **81** (1974), 3–14. 97

[27] M. Gundlach and F. Lehrbass (eds.), *CreditRisk$^+$ in the Banking Industry*, Springer-Verlag, Berlin, Heidelberg, 2004. 181, 182

[28] H. Haaf, O. Reiß, and J. Schoenmakers, *Numerically stable computation of CreditRisk$^+$*, in Gundlach and Lehrbass [27], pp. 69–77. 144

[29] K. Th. Hess, A. Liewald, and K. D. Schmidt, *An extension of Panjer's recursion*, Astin Bull. **32** (2002), no. 2, 283–297. 89, 90

[30] K. Hirhager, J. Hirz, and U. Schmock, *Conditional quantiles, conditional weighted expected shortfall and applications to capital allocation*, Work in progress, 2015. 176

[31] N. L. Johnson, S. Kotz, and A. W. Kemp, *Univariate Discrete Distributions*, 3rd ed., John Wiley & Sons, Hoboken, NJ, 2005. 47

[32] M. Kalkbrener, *An axiomatic approach to capital allocation*, Math. Finance **15** (2005), no. 3, 425–437. 164

[33] A. Klenke, *Probability Theory: A Comprehensive Course*, 2nd ed., Springer-Verlag, London, 2014. 42

[34] S. A. Klugman, H. H. Panjer, and G. E. Willmot, *Loss Models: From Data to Decisions*, 2nd ed., John Wiley & Sons, 2004. 92

[35] A. W. Knapp, *Basic Real Analysis*, Birkhäuser Boston, Inc., Boston, MA, 2005. 24

[36] L. M. Le Cam, *An approximation theorem for the Poisson binomial distribution*, Pacific J. Math. **10** (1960), 1181–1197. 27

[37] F. Lindskog and A. McNeil, *Common Poisson shock models: Applications to insurance and credit risk modelling*, Astin Bull. **33** (2003), no. 2, 209–238. 38, 115

[38] T. Lindvall, *Lectures on the Coupling Method*, Dover Publications Inc., Mineola, NY, 2002, Corrected reprint of the 1992 original. 27

[39] A. McNeil, R. Frey, and P. Embrechts, *Quantative Risk Management: Concepts, Techniques, and Tools*, Princeton University Press, 2005. 172, 173
[40] T. Mikosch, *Non-Life Insurance Mathematics, An Introduction with Stochastic Processes*, 2nd ed., Springer-Verlag, Berlin, Heidelberg, 2009. 69, 70, 98
[41] M. Moscadelli, *The modelling of operational risk: experience with the analysis of the data collected by the Basel Comittee*, Temi di Discussione del Servizio Studi **517** (2004), Available on the homepage of Banca d'Italia www.bancaditalia.it/ricerca/consultazioni/temidi. 173, 174
[42] H. H. Panjer, *Recursive evaluation of a family of compound distributions*, Astin Bull. **12** (1981), no. 1, 22–26. 90
[43] R. T. Rockafellar and S. Uryasev, *Optimization of conditional value-at-risk*, Journal of Risk **2** (2000), no. 3, 21–41. 159
[44] W. Rudin, *Real and Complex Analysis*, 3rd ed., McGraw-Hill, New York, 1987. 24
[45] C. Rudolph, *A Generalization of Panjer's Recursion for Dependent Claim Numbers and an Approximation of Poisson Mixture Models*, Ph.D. thesis, Vienna University of Technology, Austria, 2014, Advisor: U. Schmock. 127, 176
[46] C. Rudolph and U. Schmock, *An approximation via Panjer's recursion for Poisson-mixture models*, Work in progress, 2015. 176
[47] ———, *Multivariate collective risk model: Dependent claim numbers and Panjer's recursion*, Risks **8** (2020), no. 2. 115, 176
[48] M. Schmidt, *Aggregation of Integer-Valued Risks with Copula-Induced Dependency Structure*, Master's thesis, TU Wien, 2019, Advisor: U. Schmock. 108
[49] U. Schmock, *Stochastic Analysis for Financial and Actuarial Mathematics*, Lecture notes, TU Wien, March 13, 2023. 21, 22, 43, 60
[50] U. Schmock and D. Straumann, *Allocation of risk capital and performance measurement*, Talk at the Conference on Quantitative Methods in Finance, Sydney, Australia. Slides available at http://www.fam.tuwien.ac.at/~schmock/slides/AllocationSlidesOct1999.pdf, July 1999. 164, 169
[51] B. Sundt, *On multivariate Panjer recursions*, Astin Bull. **29** (1999), no. 1, 29–45. 90, 97, 98, 144
[52] B. Sundt and W. S. Jewell, *Further results on recursive evaluation of compound distributions*, Astin Bull. **12** (1981), no. 1, 27–39. 89
[53] B. Sundt and R. Vernic, *Recursions for Convolutions and Compound Distributions with Insurance Applications*, EAA Lecture Notes, Springer-Verlag, Berlin, 2009. 38
[54] D. Tasche, *Capital allocation with CreditRisk$^+$*, in Gundlach and Lehrbass [27], pp. 25–43. 169
[55] C. Villani, *Topics in Optimal Transportation*, American Mathematical Society, Providence, RI, 2003. 20
[56] ———, *Optimal Transport, Old and New*, Springer-Verlag, Berlin, 2009. 20

[57] R. E. von Holdt, *Rational powers of power series*, Amer. Math. Monthly **72** (1965), no. 7, 740–743. 85
[58] R. Warnung, *The Construction of an Integrand and Improved Recursions for Risk Aggregation*, Ph.D. thesis, Vienna University of Technology, Austria, 2008. 176
[59] D. Williams, *Probability with Martingales*, Cambridge University Press, Cambridge, 1991. 52
[60] G. E. Willmot, *Sundt and Jewell's family of discrete distributions*, Astin Bull. **18** (1988), no. 1, 17–29. 89
[61] M. V. Wüthrich and M. Merz, *Stochastic Claims Reserving Methods in Insurance*, John Wiley & Sons, Chichester, 2008. 113

Index

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

M

N

Q

R