

# MOD-POISSON APPROXIMATION SCHEMES: APPLICATIONS TO CREDIT RISK

PIERRE-LOÏC MÉLIOT, ASHKAN NIKEGHBALI, AND GABRIELE VISENTIN

ABSTRACT. We introduce a new numerical approximation method for functionals of factor credit portfolio models based on the theory of mod- $\phi$  convergence and mod- $\phi$  approximation schemes. The method can be understood as providing correction terms to the classic Poisson approximation, where higher order corrections lead to asymptotically better approximations as the number of obligors increases. We test the model empirically on two tasks: the estimation of risk measures (VaR and ES) and the computation of CDO tranche prices. We compare it to other commonly used methods – such as the recursive method, the large deviations approximation, the Chen–Stein method and the Monte Carlo simulation technique (with and without importance sampling) – and we show that it leads to more accurate estimates while requiring less computational time.

## CONTENTS

1. Introduction	2
2. Mod- $\phi$ convergence	2
3. Mod- $\phi$ approximation schemes	3
4. Mod-Poisson approximation	8
5. Mod-compound Poisson approximation	12
6. Application: Estimation of risk measures	14
6.1. Background	14
6.2. Estimation of risk measures	15
6.3. Computational time	19
7. Application: CDO pricing	22
7.1. Background	22
7.2. Estimation of call prices	23
7.3. Default leg, premium leg, and fair spread	24
8. Conclusions	27
Appendix A. Overview of estimation methods	28
A.1. Recursive methodology	28
A.2. Large deviations approximation	29
A.3. Stein’s method: first-order correction to the Gaussian approximation	30
A.4. Chen–Stein’s method: first-order correction to the Poisson approximation	30
A.5. Monte Carlo simulation	31
A.6. Importance sampling	32
Appendix B. Tail of the Poisson distribution	34
Appendix C. Incidence algebras and the Möbius function	35
Appendix D. Relation between the coefficients of the approximation scheme and the moments of the total loss variable	37
References	39

---

*Date:* October 24, 2022.

*Key words and phrases.* Mod- $\phi$  convergence, Mod- $\phi$  approximation schemes, Credit risk, Risk measures, CDO pricing, Poisson approximation.

## 1. INTRODUCTION

Mod- $\phi$  convergence [FMN16] is a new notion of convergence for sequences of random variables which provides a unified framework for the derivation of refinements of classical limit theorems, such as the central limit theorem, the Berry–Esseen theorem, precise large and moderate deviation results, local limit theorems and more. The results presented in this paper rely on mod- $\phi$  approximation schemes, which were first introduced in [BKN14] and [CDMN20], together with many examples of applications to probability theory, analytic number theory and combinatorics.

The paper is organized as follows. Sections 2 and 3 provide a self-contained introduction to the main concepts of mod- $\phi$  convergence and mod- $\phi$  approximation schemes respectively. Section 4 discusses the application of mod-Poisson approximation schemes to credit portfolio models. The main result is Theorem 4.1, which states that the total number of portfolio defaults  $L_n = \sum_{i=1}^n Y_i$  converges mod-Poisson conditionally on the mixing factor and can therefore be approximated using mod-Poisson approximation schemes  $(\nu_n^{(r)})_{n \in \mathbb{N}}$ , where  $r \geq 1$  is the order of approximation. Higher orders lead to better asymptotic approximations and can therefore be used to improve accuracy for finite  $n$ . The performance of this approximation is empirically tested on two benchmark applications – the estimation of risk measures in Section 6 and the pricing of synthetic CDO tranches in Section 7 – and is compared in terms of accuracy and computational time to the following commonly used estimation methods: recursive methodology [HW04, Bra04], the large deviations approximation [DDD04], Chen–Stein’s method and the zero-bias transformation method [EKJK08, EKJ09] and Monte Carlo simulation, with and without importance sampling [GL05]. These methods and the corresponding algorithms are presented in Appendix A. Section 5 presents the derivation of mod-compound Poisson approximation schemes for the case of portfolio losses  $L_n = \sum_{i=1}^n Z_i Y_i$  with conditionally i.i.d. exposures  $(Z_i)_{i=1}^n$ . Section 8 concludes, and Appendices B, C and D present the detailed proofs of certain theoretical results. In particular, the framework of mod- $\phi$  convergence relies on certain combinatorial arguments related to the theory of Möbius inversion and to the theory of symmetric functions; for this later topic, we refer to [Mac95, Chapter 1], and everything required is recalled in Section D.

2. MOD- $\phi$  CONVERGENCE

Given a  $\mathbb{Z}$ -valued random variable  $X$  with probability law  $\mu_X$ , we can define its characteristic function as the Fourier transform of its law:

$$\widehat{\mu}_X(\xi) = \mathbb{E} [e^{i\xi X}] = \sum_{k \in \mathbb{Z}} \mu_X(\{k\}) e^{ik\xi}, \quad \xi \in \mathbb{T} := \mathbb{R}/2\pi\mathbb{Z}.$$

Notice that the Fourier transform  $\widehat{\mu}_X$  in this case is well-defined on  $\mathbb{T}$ , because we are assuming  $X$  to be an integer-valued random variable.

If  $\mu$  is an infinitely divisible distribution on  $\mathbb{Z}$ , then its Fourier transform admits the following representation:

$$\widehat{\mu}(\xi) = e^{\phi(\xi)}, \quad \xi \in \mathbb{T},$$

where  $\phi$  is a periodic function of period  $2\pi$ , called the Lévy–Khintchine exponent [SVH03] of the distribution. Infinitely divisible laws will play a fundamental role in mod- $\phi$  convergence. Two very important examples are given below.

*Example 2.1. (Poisson distribution)* Let  $X$  be a Poisson random variable with parameter  $\lambda$ , i.e.  $X \sim \text{Po}(\lambda)$ . Then, its Lévy–Khintchine exponent is given by:

$$\phi(\xi) = \lambda(e^{i\xi} - 1), \quad \xi \in \mathbb{T}.$$

*Example 2.2. (Compound Poisson distribution)* A random variable  $X$  follows a compound Poisson distribution if it admits the following representation:

$$X = \sum_{i=1}^N Z_i,$$

where  $(Z_i)_{i=1}^{\infty}$  is a family of i.i.d.  $\mathbb{N}$ -valued random variables distributed like  $Z$ , and  $N$  is an independent Poisson random variable with parameter  $\lambda$ . In this case we write  $X \sim \text{CPo}(\lambda, Z)$ . If  $X \sim \text{CPo}(\lambda, Z)$ , then its Lévy–Khintchine exponent is given by:

$$\phi(\xi) := \log (\mathbb{E} [e^{i\xi X}]) = \lambda(\widehat{\mu}_Z(\xi) - 1),$$

where  $\widehat{\mu}_Z$  is the Fourier transform of the law of  $Z$ . For  $\mathbb{N}$ -valued random variables, it turns out that every infinitely divisible distribution is a compound Poisson distribution: if  $X$  is a  $\mathbb{N}$ -valued infinitely divisible distribution, then it necessarily follows that there exists a  $\lambda > 0$  and a  $\mathbb{N}$ -valued random variable  $Z$  such that  $X \sim \text{CPo}(\lambda, Z)$ ; see [SVH03, Theorem 3.2].

By Lévy’s continuity theorem [Kal02, Thm. 4.3], knowledge of  $\widehat{\mu}_X$  is equivalent to knowledge of the full probability law  $\mu_X$  and limit theorems for probability laws can be derived directly in Fourier space in terms of pointwise convergence of characteristic functions. For instance, the so-called *law of small numbers* (which is a generalization of the original Poisson convergence theorem) states that if  $Y_{i,n}, i = 1, \dots, n$  is a triangular array of independent Bernoulli random variables with success probabilities  $p_{i,n}$  such that  $\sum_{i=1}^n p_{i,n} \rightarrow \lambda \in \mathbb{R}$  and  $\max_{1 \leq i \leq n} p_{i,n} \rightarrow 0$ , then  $X_n := \sum_{i=1}^n Y_{i,n}$  converges in law to a random variable  $Y$ , with  $Y \sim \text{Po}(\lambda)$ .

This convergence can be proved from the pointwise convergence of  $\widehat{\mu}_{X_n}$  to the characteristic function of  $Y$  as  $n \rightarrow \infty$  (see, for instance, [Dur10, Theorem 3.6.1]). But this limit law can also be interpreted as a non-asymptotic approximation result, stating that the law of  $X_n$  can be approximated by a Poisson variable with parameter  $\lambda_n = \sum_{i=1}^n p_{i,n}$ , which is just a sum of  $\lambda_n$  independent Poisson random variables with parameter one. We would expect this approximation to work well for  $n \rightarrow \infty$ , or equivalently for  $\lambda_n \rightarrow \infty$ .

More generally, given a sequence  $(X_n)_{n \in \mathbb{N}}$  of  $\mathbb{Z}$ -valued random variables, we are interested in the problem of approximating the law  $\mu_{X_n}$  of  $X_n$  – which might be difficult to compute or simulate – by a sum  $Y_n$  of  $\lambda_n$  i.i.d. copies of a given  $\mathbb{Z}$ -valued infinitely divisible law with Lévy–Khintchine exponent  $\phi$ , for  $\lambda_n \rightarrow \infty$ . Let us notice that we can make sense of the law of  $Y_n$  even if  $\lambda_n$  is not an integer, by placing oneself in the convolution semigroup of probability measures generated by  $\phi$ . This approximation problem leads to the following definition:

**Definition 2.3** (Mod- $\phi$  convergence). *We say that  $(X_n)_{n \in \mathbb{N}}$  converges mod- $\phi$  with parameters  $(\lambda_n)_{n \in \mathbb{N}}$  and limiting function  $\psi$  if  $\lambda_n \rightarrow \infty$  and*

$$\widehat{\mu}_{X_n}(\xi) e^{-\lambda_n \phi(\xi)} = \psi_n(\xi), \quad \xi \in \mathbb{T} \tag{1}$$

with

$$\lim_{n \rightarrow \infty} \psi_n(\xi) = \psi(\xi).$$

The convergence  $\psi_n \rightarrow \psi$  usually occurs in a space of continuously derivable functions  $\mathcal{C}^r(\mathbb{T})$  endowed with the norm  $\|f\|_{\mathcal{C}^r} = \sup_{|\alpha| \leq r} \sup_{\xi \in \mathbb{T}} |\partial^\alpha f(\xi)|$ . When a sequence  $(X_n)_{n \in \mathbb{N}}$  converges mod- $\phi$ , we can use the functions  $\psi_n$  and  $\psi$  to extract precise information about the behavior of  $\mu_{X_n}$ , both asymptotically and non-asymptotically, and to construct good approximations by the reference infinitely divisible law, as we will see in the next section.

### 3. MOD- $\phi$ APPROXIMATION SCHEMES

The right space to study these approximations turns out to be the Wiener algebra  $\mathcal{A}(\mathbb{T})$  of absolutely convergent Fourier series on  $\mathbb{T}$ , which is a Banach algebra under the pointwise product and the norm

$$\|\widehat{\mu}\|_{\mathcal{A}(\mathbb{T})} = \sum_{n \in \mathbb{Z}} |c_n(\widehat{\mu})|,$$

where  $c_n(\widehat{\mu}) := \int_{\mathbb{T}} \widehat{\mu}(\xi) e^{-in\xi} d\xi$  is the  $n$ -th Fourier coefficient of  $\widehat{\mu}$ .

Working in the Wiener algebra  $\mathcal{A}(\mathbb{T})$  is essential for at least two reasons. First of all, the characteristic function  $\widehat{\mu}$  of a probability law  $\mu$  always lies in  $\mathcal{A}(\mathbb{T})$  and its Fourier coefficients satisfy

$c_n(\widehat{\mu}) = \mu(n)$ , so that the Wiener algebra norm of  $\widehat{\mu}$  is actually equal to the total variation norm of the law  $\mu$  itself (or twice this norm depending on the chosen convention). This means that good approximation bounds in the Wiener algebra directly translate into good approximation bounds in total variation distance.

Second, Wiener's  $1/f$  theorem guarantees that  $e^{-\lambda_n \phi(\xi)}$  lies in  $\mathcal{A}(\mathbb{T})$ . This fact, together with Eq. (1), implies that  $\psi_n \in \mathcal{A}(\mathbb{T})$  and that it can be thought of as the deconvolution residue of the law  $\mu_{X_n}$  by a sum of  $\lambda_n$  independent copies of the infinitely divisible reference law with Lévy–Khinchine exponent  $\phi$ . In other words, we can think of  $\psi_n$  as that element of  $\mathcal{A}(\mathbb{T})$  that satisfies the following equation:

$$\widehat{\mu}_{X_n}(\xi) = \psi_n(\xi) e^{\lambda_n \phi(\xi)}. \quad (2)$$

Clearly Equation (2) implies that we can always use the deconvolution residue  $\psi_n$  to reconstruct  $\widehat{\mu}_{X_n}$  perfectly by pointwise multiplication with the Fourier transform of the reference infinitely divisible, but in practice  $\psi_n$  might be as hard to compute as  $\widehat{\mu}_{X_n}$  itself. Nevertheless one can construct good approximations for  $\widehat{\mu}_{X_n}$  by substituting  $\psi_n$  in Eq. (2) with another function  $\chi_n \in \mathcal{A}(\mathbb{T})$  that approximates it sufficiently well on  $\mathbb{T}$  and that is easier to compute. This leads to the following definition of mod- $\phi$  approximation schemes, which was first given in [CDMN20]. In the sequel, we focus on  $\mathbb{N}$ -valued random variables, hence non-negative. This choice fits with the application that we have in mind, namely, the approximation of the distribution of the total loss variable of a credit portfolio. However, notice that if we were working with variables that can be positive or negative, then a straightforward extension of our methods to  $\mathbb{Z}$ -valued random variables exist and is described in [CDMN20, Example 1.11].

**Definition 3.1** (Mod- $\phi$  approximation scheme of order  $r$ ). *Let  $(X_n)_{n \in \mathbb{N}}$  be a sequence of  $\mathbb{N}$ -valued random variables that converges mod- $\phi$  with parameters  $(\lambda_n)_{n \in \mathbb{N}}$ . We suppose that  $\psi_n$  admits the following series expansion around zero:*

$$\psi_n(\xi) = 1 + \sum_{k=1}^{\infty} b_{k,n} (e^{i\xi} - 1)^k.$$

*Then the mod- $\phi$  approximation scheme of order  $r$  for  $(X_n)_{n \in \mathbb{N}}$  is a sequence of discrete signed measures  $(\nu_n^{(r)})_{n \in \mathbb{N}}$  on  $\mathbb{Z}$ , such that*

$$\widehat{\nu}_n^{(r)}(\xi) = \chi_n^{(r)}(\xi) e^{\lambda_n \phi(\xi)}, \quad (3)$$

*where  $\chi_n^{(r)}$  is the polynomial of degree  $r$  that approximates  $\psi_n$  around 0 up to order  $r$ :*

$$\chi_n^{(r)}(\xi) = 1 + \sum_{k=1}^r b_{k,n} (e^{i\xi} - 1)^k.$$

An explicit formula for the measure  $\widehat{\nu}_n^{(r)}$  in terms of  $\lambda_n$  and of the coefficients  $b_{k,n}$  is given in [CDMN20, Lemma 3.8 and Remark 3.9]. For instance, if  $\phi = \text{Po}(1)$  and  $r = 2$ , then we have:

$$\nu_n^{(2)}(k) = \frac{e^{-\lambda_n} (\lambda_n)^k}{k!} \left( 1 + b_{2,n} \left( 1 - \frac{2k}{\lambda_n} + \frac{k(k-1)}{(\lambda_n)^2} \right) \right),$$

so  $\nu_n^{(2)}$  is in this case a perturbation of  $\nu_n^{(0)} = \text{Po}(\lambda_n)$ . In general, the zero-th order approximation  $\nu_n^{(0)}$  corresponds to approximating  $\mu_{X_n}$  with a sum of  $\lambda_n$  i.i.d. copies of the reference infinitely divisible law, while higher order approximations will correspond to signed measures that approximate the law  $\mu_{X_n}$  increasingly well. Let us be a bit more precise on this claim. In a companion paper [MNV22], we prove that if  $X_n$  is a sum of  $n$  independent Bernoulli variables with parameters  $p_1, \dots, p_n$ , then the corresponding approximations  $\nu_n^{(r)}$  satisfy

$$\sum_{k \in \mathbb{N}} |\mathbb{P}(X_n = k) - \nu_n^{(r)}(k)| \leq A \left( \frac{K \sigma_n}{\sqrt{\lambda_n}} \right)^{r+1}, \quad (4)$$

with  $\lambda_n = \sum_{i=1}^n p_i$ ,  $(\sigma_n)^2 = \sum_{i=1}^n (p_i)^2$  and  $A$  and  $K$  universal constants. The mod-Poisson convergence of such sequences  $(X_n)_{n \in \mathbb{N}}$  is explained in Theorem 4.1 below. Thus, the quality of the approximation  $\nu_n^{(r)}$  of  $\mu_{X_n}$  indeed increases with the order of approximation  $r$ . The results that we shall present in Sections 6 and 7 are numerical evidences of the general theoretical result (4), and applications for the study of credit risk models.

In general, the way in which the measures  $\nu_n^{(r)}$  incorporate the information contained in  $\mu_{X_n}$  and in the residue  $\psi_n$  as  $r$  increases can also be understood in terms of factorial cumulants.

**Definition 3.2** (Factorial cumulant generating function). *If  $X$  is a  $\mathbb{N}$ -valued random variable, then its factorial cumulant generating function (when it exists, which is always the case in this paper) is defined as*

$$\log (\mathbb{E} [(z+1)^X]) = \sum_{k=1}^{\infty} \frac{1}{k!} \kappa_k(X) z^k, \quad z \in \mathbb{C},$$

and the coefficient  $\kappa_k(X)$  is called the  $k$ -th factorial cumulant of  $X$ .

*Example 3.3.* If  $X \sim \text{Po}(\lambda)$ , then

$$\kappa_k(X) = \begin{cases} \lambda & \text{if } k = 1, \\ 0 & \text{otherwise.} \end{cases}$$

We note that if  $X$  is a random variable with law  $\mu$ , then the coefficients in the expansion of its Fourier transform  $\widehat{\mu}(\xi)$  in powers of  $z := (e^{i\xi} - 1)$  are precisely its factorial cumulants, so that the coefficients of a mod- $\phi$  approximation scheme are naturally related to these quantities. In particular, the following proposition shows that the measures  $\nu_n^{(r)}$  achieve better approximations by matching exactly the factorial cumulants of  $X_n$  up to order  $r$ , while maintaining the factorial cumulants of the reference infinitely divisible law for all higher orders.

**Proposition 3.4.** *Let  $(X_n)_{n \in \mathbb{N}}$  be a sequence of  $\mathbb{N}$ -valued random variables that converges mod- $\phi$  with parameters  $(\lambda_n)_{n \in \mathbb{N}}$ , and let  $(Y_n)_{n \in \mathbb{N}}$  follow the reference infinitely divisible laws with exponents  $(\lambda_n \phi)_{n \in \mathbb{N}}$ . If  $(\nu_n^{(r)})_{n \in \mathbb{N}}$  is a mod- $\phi$  approximation scheme of order  $r$  for  $(X_n)_{n \in \mathbb{N}}$  and if we denote by  $\kappa_{k,n}^{(r)}$  the  $k$ -th factorial cumulant of  $\nu_n^{(r)}$ , then*

$$\kappa_{k,n}^{(r)} = \begin{cases} \kappa_k(X_n) & \text{for } k = 1, \dots, r, \\ \kappa_k(Y_n) & \text{for } k \geq r + 1, \end{cases}$$

where  $\kappa_k(X)$  is the  $k$ -th factorial cumulant of the random variable  $X$ .

*Proof.* From Equation (2) we compute:

$$\begin{aligned} \psi_n(\xi) &= \exp \left( \sum_{k=1}^{\infty} \frac{1}{k!} (\kappa_k(X_n) - \kappa_k(Y_n)) z^k \right) \\ &= 1 + \sum_{k=1}^{\infty} \frac{1}{k!} \sum_{\pi \in \Pi(k)} \prod_{B \in \pi} (\kappa_{|B|}(X_n) - \kappa_{|B|}(Y_n)) z^k, \end{aligned}$$

by using the first identity of Theorem C.6 in the last step. Since  $\chi_n^{(r)}$  approximates  $\psi_n$  up to order  $r$  in powers of  $z$ , we have:

$$\log (\chi_n^{(r)}(\xi)) = \log \left( 1 + \sum_{k=1}^r \frac{1}{k!} \sum_{\pi \in \Pi(k)} \prod_{B \in \pi} (\kappa_{|B|}(X_n) - \kappa_{|B|}(Y_n)) z^k \right).$$

We can now compute the coefficients of the series expansion of  $\log(\chi_n^{(r)})$  in powers of  $z$  by using the second identity of Theorem C.6:

$$\begin{aligned}
s! [z^s] \log(\chi_n^{(r)}(\xi)) &= \sum_{\sigma \in \Pi(s)} \mu(\sigma, \widehat{1}_s) \prod_{D \in \sigma} \sum_{\pi \in \Pi(|D|)} \prod_{B \in \pi} (\kappa_{|B|}(X_n) - \kappa_{|B|}(Y_n)) \mathbb{1}_{\{|B| \leq r\}} \\
&= \sum_{\sigma \in \Pi(s)} \mu(\sigma, \widehat{1}_s) \sum_{\tau \leq \sigma} \prod_{B \in \tau} (\kappa_{|B|}(X_n) - \kappa_{|B|}(Y_n)) \mathbb{1}_{\{|B| \leq r\}} \\
&= \sum_{\tau \in \Pi(s)} \sum_{\sigma \in \Pi(s)} \zeta(\tau, \sigma) \mu(\sigma, \widehat{1}_s) \prod_{B \in \tau} (\kappa_{|B|}(X_n) - \kappa_{|B|}(Y_n)) \mathbb{1}_{\{|B| \leq r\}} \\
&= \sum_{\tau \in \Pi(s)} \delta(\tau, \widehat{1}_s) \prod_{B \in \tau} (\kappa_{|B|}(X_n) - \kappa_{|B|}(Y_n)) \mathbb{1}_{\{|B| \leq r\}} \\
&= (\kappa_s(X_n) - \kappa_s(Y_n)) \mathbb{1}_{\{s \leq r\}}
\end{aligned}$$

where we have used the convolution relation  $\zeta \star \mu = \delta$  for the Möbius function of the poset  $\Pi(s)$ . We refer the reader to Appendix C for a primer on the Möbius function and its basic properties.

Finally, we can compute the factorial cumulants of the measures  $\nu_n^{(r)}$  by taking logarithms in Equation (3):

$$\begin{aligned}
\log(\widehat{\nu}_n^{(r)}(\xi)) &= \sum_{k=1}^{\infty} \frac{1}{k!} (\kappa_k(Y_n)) z^k + \log(\chi_n^{(r)}(\xi)) \\
&= \sum_{k=1}^{\infty} \frac{1}{k!} (\kappa_k(Y_n)) z^k + \sum_{k=1}^r \frac{1}{k!} (\kappa_k(X_n) - \kappa_k(Y_n)) z^k \\
&= \sum_{k=1}^r \frac{1}{k!} (\kappa_k(X_n)) z^k + \sum_{k=r+1}^{\infty} \frac{1}{k!} (\kappa_k(Y_n)) z^k.
\end{aligned}$$

□

As a particular case of Proposition 3.4, we remark that in the case of mod-Poisson approximation schemes, the signed measures  $\nu_n^{(r)}$  have all factorial cumulants equal to zero (because the Poisson distribution itself does) with the exception of the first  $r$  cumulants, which exactly match the factorial cumulants of  $X_n$ . Since factorial cumulant generating functions fully characterize probability distributions, Proposition 3.4 also implies that  $\nu_n^{(\infty)} = \mu_{X_n}$ .

Despite the availability of this interpretation in terms of factorial cumulants, the signed measures  $\nu_n^{(r)}$  are in general difficult to compute, even if we have full knowledge of the coefficients  $(b_{k,n})_{k=1}^r$ . Nevertheless, the computation of expectations of functions of these measures can be done efficiently, as the following proposition shows.

**Proposition 3.5.** *Let  $(X_n)_{n \in \mathbb{N}}$  be a sequence of  $\mathbb{N}$ -valued random variables that converges mod- $\phi$  with parameters  $(\lambda_n)_{n \in \mathbb{N}}$ , and let  $(\nu_n^{(r)})_{n \in \mathbb{N}}$  be its mod- $\phi$  approximation scheme of order  $r$ . Then, for any bounded function  $f : \mathbb{N} \rightarrow \mathbb{R}$ , the integral of  $f$  with respect to  $\nu_n^{(r)}$  is given by:*

$$\nu_n^{(r)}(f) = \sum_{j \in \mathbb{N}} f(j) \nu_n^{(r)}(\{j\}) = \mathbb{E}[f(Y_n)] + \mathbb{E}[\Delta_n(r, f)(Y_n)]$$

where  $Y_n$  follows the reference infinitely divisible law with exponent  $\lambda_n \phi$ , and where the correction term  $\Delta_n(r, f)$  is given by:

$$\Delta_n(r, f)(j) = \sum_{k=1}^r b_{k,n} (\Delta_+^k(f))(j).$$

Here,  $\Delta_+^k$  denotes the  $k$ -th power of the forward finite difference operator:

$$(\Delta_+^k(f))(j) = \sum_{l=0}^k (-1)^{k-l} \binom{k}{l} f(j+l).$$

*Proof.* A similar result with functions  $f \in \ell^2(\mathbb{N})$  is stated in [CDMN20, Proposition 1.12]; in the sequel, we give a new proof when  $f$  is only assumed to be bounded, and at the end we shall even explain how to extend the result to polynomially bounded functions. Since  $\nu_n^{(r)}$  is a finite signed measure, it is in  $\ell^1(\mathbb{N}) \subset \ell^2(\mathbb{N})$ . On the other hand, for any  $f, g \in \ell^2(\mathbb{N})$ , the Parseval formula holds:

$$\sum_{j \in \mathbb{N}} f(j) g(j) = \int_0^{2\pi} \widehat{f}(\xi) \overline{\widehat{g}(\xi)} \frac{d\xi}{2\pi}.$$

In particular, if  $f \in \ell^2(\mathbb{N})$  and  $g = \nu_n^{(r)}$ , we obtain:

$$\nu_n^{(r)}(f) = \nu_n^{(0)}(f) + \sum_{k=1}^r b_{k,n} \left( \int_0^{2\pi} \widehat{f}(\xi) \overline{(e^{-i\xi} - 1)^k} \widehat{\nu}_n^{(0)}(\xi) \frac{d\xi}{2\pi} \right),$$

where  $\widehat{\nu}_n^{(0)}(\xi) = e^{\lambda_n \phi(\xi)}$ . However, for  $f \in \ell^1(\mathbb{N})$ , we have

$$\begin{aligned} \widehat{f}(\xi) (e^{-i\xi} - 1) &= \sum_{j \in \mathbb{N}} f(j) e^{ij\xi} (e^{-i\xi} - 1) \\ &= f(0) e^{-i\xi} + \sum_{j \in \mathbb{N}} (f(j+1) - f(j)) e^{ij\xi} \\ &= f(0) e^{-i\xi} + \widehat{\Delta_+(f)}(\xi) \end{aligned}$$

so

$$\begin{aligned} \int_0^{2\pi} \widehat{f}(\xi) \overline{(e^{-i\xi} - 1)^k} \widehat{\nu}_n^{(0)}(\xi) \frac{d\xi}{2\pi} &= f(0) \int_0^{2\pi} e^{i\xi} \widehat{\nu}_n^{(0)}(\xi) \frac{d\xi}{2\pi} + \int_0^{2\pi} \widehat{\Delta_+(f)}(\xi) \overline{(e^{-i\xi} - 1)^k} \widehat{\nu}_n^{(0)}(\xi) \frac{d\xi}{2\pi} \\ &= \nu_n^{(0)}(\Delta_+(f)) \end{aligned}$$

since the first integral vanishes ( $\widehat{\nu}_n^{(0)}$  has only positive Fourier coefficients). By an immediate induction,

$$\int_0^{2\pi} \widehat{f}(\xi) \overline{(e^{-i\xi} - 1)^k} \widehat{\nu}_n^{(0)}(\xi) \frac{d\xi}{2\pi} = \nu_n^{(0)}(\Delta_+^k(f))$$

for any  $k \geq 1$ , whence the result for  $f \in \ell^1(\mathbb{N})$ . So, we have the equality of linear forms on the space of summable real functions on  $\mathbb{N}$ :

$$\nu_n^{(r)} = \nu_n^{(0)} \circ \left( \text{id} + \sum_{k=1}^r b_{k,n} \Delta_+^k \right).$$

Consider now a bounded function  $f : \mathbb{N} \rightarrow \mathbb{R}$ , and for  $L \in \mathbb{N}$ , denote  $f_L(j) = \mathbb{1}_{j \leq L} f(j)$ , which is in  $\ell^1(\mathbb{N})$ . Since  $\nu_n^{(0)}$  and  $\nu_n^{(r)}$  are in  $\ell^1(\mathbb{N})$ , for any sequence of functions  $(g_L)_{L \in \mathbb{N}}$  which are uniformly bounded by a constant  $K$  and such that  $g_L(j) \rightarrow_{L \rightarrow \infty} g(j)$  for any  $j \in \mathbb{N}$ ,  $\nu_n^{(0)}(g_L) \rightarrow_{L \rightarrow \infty} \nu_n^{(0)}(g)$  and  $\nu_n^{(r)}(g_L) \rightarrow_{L \rightarrow \infty} \nu_n^{(r)}(g)$  by the dominated convergence theorem. Here, we have obviously

$$\begin{aligned} f_L(j) &\rightarrow_{L \rightarrow \infty} f(j); \\ \left( f_L(j) + \sum_{k=1}^r b_{k,n} (\Delta_+^k f_L)(j) \right) &\rightarrow_{L \rightarrow \infty} \left( f(j) + \sum_{k=1}^r b_{k,n} (\Delta_+^k f)(j) \right), \end{aligned}$$

so

$$\nu_n^{(r)}(f) = \lim_{L \rightarrow \infty} \nu_n^{(r)}(f_L) = \lim_{L \rightarrow \infty} \nu_n^{(0)} \left( f_L + \sum_{k=1}^r b_{k,n} \Delta_+^k (f_L) \right) = \nu_n^{(0)} \left( f + \sum_{k=1}^r b_{k,n} \Delta_+^k (f) \right).$$

Let us remark that the dominated convergence argument works for a larger class of functions: a sufficient assumption is that  $(f(j))_{j \in \mathbb{N}}$  and its shifts  $(f(j+l))_{j \in \mathbb{N}}$  with  $1 \leq l \leq r$  are bounded by functions which are integrable against the reference infinitely divisible distribution  $\nu_n^{(0)}$ . In particular, if  $\phi$  is the exponent of the Poisson distribution, then the formula of the proposition holds for any  $f$  bounded by a polynomial function. More generally, if  $\phi = \text{CPo}(\lambda, Z)$  with  $\mathbb{E}[Z^r] < +\infty$ , then the formula of the proposition holds for any  $f$  bounded by a polynomial function with degree  $r$ .  $\square$

We can summarize Proposition 3.5 by saying that expectations of the form  $\mathbb{E}[f(X_n)]$  can be approximated with a mod- $\phi$  approximation scheme  $(\nu_n^{(r)})_{n \in \mathbb{N}}$  in two steps:

- (1) Replace  $X_n$  by the infinitely divisible random variable  $Y_n$ .
- (2) Correct the function  $f$  by adding the correction term  $\Delta_n(r, f)$ .

The main advantage of this approximation procedure is that computing integrals of the reference infinitely divisible law is typically easier, as they may even admit closed-form expressions in terms of well-known special functions.

*Remark 3.6.* If the reference infinitely divisible law  $\phi$  has a moment of order 3, and if the convergence of residues  $\psi_n(\xi) \rightarrow \psi(\xi)$  occurs in the space  $\mathcal{C}^{r+2}(\mathbb{T})$ , then [CDMN20, Theorem 3.11] shows that for any  $r \geq 0$  and any bounded function  $f : \mathbb{N} \rightarrow \mathbb{R}$ , the difference  $|\mathbb{E}[f(X_n)] - \mathbb{E}[f(Y_n)] - \mathbb{E}[\Delta_n(r, f)(Y_n)]|$  goes to 0 as  $n$  goes to infinity, with a speed of convergence which improves with the order of approximation  $r$ . Indeed,

$$\begin{aligned} |\mathbb{E}[f(X_n)] - \mathbb{E}[f(Y_n)] - \mathbb{E}[\Delta_n(r, f)(Y_n)]| &= |\mu_{X_n}(f) - \nu_n^{(r)}(f)| \\ &\leq \|f\|_\infty d_{\text{TV}}(\mu_{X_n}, \nu_n^{(r)}) = O\left(\frac{1}{(\lambda_n)^{\frac{r+1}{2}}}\right). \end{aligned}$$

In the case which we shall examine in Section 4 (sums of independent Bernoulli variables), the result follows also from the general estimate (4).

#### 4. MOD-POISSON APPROXIMATION

In the context of credit risk we are interested in approximating the total losses of a credit portfolio with  $n$  counterparties, which is given by

$$L_n = \sum_{i=1}^n Z_i Y_i,$$

where  $Y_i$  is the default indicator function for the  $i$ -th counterparty (i.e. a Bernoulli random variable with  $\mathbb{P}(Y_i = 1) = 1 - \mathbb{P}(Y_i = 0) = p_i$ , where  $p_i$  is the default probability) and  $Z_i$  is the monetary loss incurred by the portfolio due to that counterparty's default.

It is customary in credit risk to assume that the random variables  $(Y_i)_{i=1}^n$  are conditionally independent given some underlying (macroeconomic or purely statistical) latent factor,  $\Psi$ , so that one can write  $\mathbb{P}(Y_i = 1 | \Psi) = p_i(\Psi)$ , for some measurable function  $p_i$ , which depends on the particular credit risk model we are interested in studying. On the other hand, the exposures  $(Z_i)_{i=1}^n$  can be either constant or random. In the latter case, it is common practice to assume them to be i.i.d. random variables, independent of  $(Y_i)_{i=1}^n$ . A less common choice is to incorporate the dependence on the underlying factor  $\Psi$  by choosing  $(Z_i)_{i=1}^n$  to be conditionally i.i.d. given  $\Psi$ .

In this section we focus on the case of constant unit exposures (i.e.  $Z_i = 1$ , for all  $i = 1, \dots, n$ ), which is related to the classical Poisson approximation problem, but we anticipate that all results presented here will be extended in Section 5 to the general case of  $(Z_i)_{i=1}^n$  conditionally i.i.d. given

$\Psi$ . The basic idea is to derive mod-Poisson convergence of  $L_n$  conditionally on  $\Psi$  and use mod-Poisson approximation schemes to estimate functionals of  $L_n$  conditionally on  $\Psi$ . Unconditional estimates would then follow by integrating (numerically) on  $\Psi$ .

**Theorem 4.1 (Mod-Poisson convergence).** *Let  $(L_n)_{n \in \mathbb{N}}$  be a sequence of total portfolio losses given by  $L_n = \sum_{i=1}^n Y_i$ , where we assume the  $Y_i$ 's to be independent random variables such that  $Y_i \sim \text{Be}(p_i)$ .*

- (1) *If  $\sum_{i=1}^{\infty} p_i = +\infty$  and  $\sum_{i=1}^{\infty} (p_i)^2 < +\infty$ , then  $(L_n)_{n \in \mathbb{N}}$  converges mod- $\phi$  with parameters  $\lambda_n = \sum_{i=1}^n p_i$  and with reference infinitely divisible law the Poisson distribution  $\text{Po}(1)$  (i.e.  $\phi(\xi) = e^{i\xi} - 1$ ).*
- (2) *Furthermore,  $(L_n)_{n \in \mathbb{N}}$  admits a mod-Poisson approximation scheme of order  $r$  with the following coefficients:*

$$b_{k,n} = \frac{1}{k!} \sum_{\substack{\pi \in \Pi(k) \\ \forall B \in \pi, |B| \geq 2}} \mu(\widehat{0}_k, \pi) \left( \prod_{B \in \pi} \mathfrak{p}_{|B|,n} \right) \quad (5)$$

where  $\Pi(k)$  denotes the poset of set partitions of  $\{1, \dots, k\}$ ;  $\mu(\cdot, \cdot)$  is the Möbius function for the incidence algebra of the poset  $\Pi(k)$ ; and  $\mathfrak{p}_{k,n} := \sum_{j=1}^n (p_j)^k$  for  $k \geq 2$ .

*Proof.* We compute the deconvolution residues as in Equation (1) using the explicit form for the characteristic function of  $L_n$  and of the  $\text{Po}(1)$  distribution, obtaining:

$$\begin{aligned} \psi_n(\xi) &= \widehat{\mu}_{L_n}(\xi) e^{-\sum_{j=1}^n p_j (e^{i\xi} - 1)} \\ &= \prod_{j=1}^n (1 + p_j (e^{i\xi} - 1)) e^{-p_j (e^{i\xi} - 1)} \\ &= \exp \left( \sum_{j=1}^n \log(1 + p_j (e^{i\xi} - 1)) - p_j (e^{i\xi} - 1) \right) \\ &= \exp \left( \sum_{k=2}^{\infty} \frac{(-1)^{k-1}}{k} (e^{i\xi} - 1)^k \mathfrak{p}_{k,n} \right), \end{aligned} \quad (6)$$

where we have defined  $\mathfrak{p}_{k,n} := \sum_{j=1}^n (p_j)^k$ .

The condition  $\sum_{j=1}^{\infty} (p_j)^2 < +\infty$  guarantees that  $\psi_n(\xi)$  converges to a limit  $\psi(\xi)$  uniformly in  $\mathbb{T}$ , so the mod- $\phi$  convergence is proved. In order to compute the coefficients  $(b_{k,n})_{k=1}^r$  for the mod- $\phi$  approximation scheme of order  $r$ , we need to extract the coefficients of the series expansion of  $\psi_n(\xi)$  in powers of  $(e^{i\xi} - 1)$ . This is again an application of the first identity of Theorem C.6, since  $\psi_n(\xi)$  is the exponential of the generating series  $G(z) = \sum_{k=2}^{\infty} \frac{g_k}{k!} z^k$  with  $g_{k \geq 2} = (-1)^{k-1} (k-1)! \mathfrak{p}_{k,n}$ . Hence,

$$\begin{aligned} b_{k,n} &= [z^k] \exp(G(z)) = \frac{1}{k!} \sum_{\substack{\pi \in \Pi(k) \\ \forall B \in \pi, |B| \geq 2}} \prod_{B \in \pi} g_{|B|} \\ &= \frac{1}{k!} \sum_{\substack{\pi \in \Pi(k) \\ \forall B \in \pi, |B| \geq 2}} \mu(\widehat{0}_k, \pi) \prod_{B \in \pi} \mathfrak{p}_{|B|,n} \end{aligned}$$

by using the formula for the Möbius function of the poset  $\Pi(k)$  computed in the Appendix C. Notice that the sum runs over all set partitions with each block of size at least two.  $\square$

*Remark 4.2.* The computational time for the evaluation of the coefficients  $(b_{k,n})_{k=1}^r$  can be substantially reduced by noticing that in (5) the term in the summation depends only on the type of the

set partition  $\pi$ . This leads to the following equivalent, but computationally more advantageous expression:

$$b_{k,n} = \sum_{\substack{\lambda \in P(k) \\ \lambda = (\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_\ell \geq 2)}} \frac{(-1)^{(k-\ell(\lambda))}}{z_\lambda} \mathfrak{p}_{\lambda,n},$$

where  $P(k)$  is the set of integer partitions of  $k$ , i.e. the set of all finite non-increasing sequences of positive integers,  $\lambda = (\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_\ell)$ , such that  $\sum_{i=1}^{\ell} \lambda_i = k$ . We also denote  $\ell(\lambda)$  the number of parts of the integer partition  $\lambda$ ;  $\mathfrak{p}_{\lambda,n} = \prod_{i=1}^{\ell(\lambda)} \mathfrak{p}_{\lambda_i,n}$ ; and  $z_\lambda = \prod_{k \geq 1} k^{m_k(\lambda)} (m_k(\lambda))!$ , where  $m_k(\lambda)$  is the number of parts of  $\lambda$  of size  $k$ . In the expression of  $b_{k,n}$  in terms of the  $\mathfrak{p}_{\lambda,n}$ , the sum runs over those integer partitions  $\lambda \in P(k)$  such that all the parts of  $\lambda$  are larger than 2.

The first few coefficients of the mod- $\phi$  approximation in terms of the obligors' default probabilities are:

$$\begin{aligned} b_{1,n} &= 0 & ; & & b_{2,n} &= -\frac{1}{2} \sum_{i=1}^n (p_i)^2; \\ b_{3,n} &= \frac{1}{3} \sum_{i=1}^n (p_i)^3 & ; & & b_{4,n} &= -\frac{1}{4} \sum_{i=1}^n (p_i)^4 + \frac{1}{8} \left( \sum_{i=1}^n (p_i)^2 \right)^2. \end{aligned}$$

We remark that the first-order correction to the Poisson approximation presented in [EKJK08, EKJ09] and based on the Chen–Stein method corresponds to a mod-Poisson approximation scheme of order  $r = 2$ . On the other hand, higher order approximations have so far remained inaccessible to the Chen–Stein method.

*Remark 4.3.* Equation (5) shows that the coefficients  $b_{k,n}$  are symmetric functions in the default probabilities  $p_i$ , because they are polynomials in the Newton power sums  $\mathfrak{p}_{k,n}$ . By using the combinatorics of symmetric functions, one can rewrite the coefficients  $(b_{k,n})_{k=1}^r$  as polynomials in the moments  $(\mathbb{E}[(L_n)^k])_{k=1}^r$  of the total portfolio loss variable. Thus, if  $M_{k,n} = \mathbb{E}[(L_n)^k]$  (which can be estimated numerically), then

$$b_{k,n} = \frac{(-1)^k}{k!} (M_{1,n})^k + \sum_{1 \leq m \leq l \leq k} \frac{(-1)^{k-m}}{(k-l)! l!} \begin{bmatrix} l \\ m \end{bmatrix} M_{m,n} (M_{1,n})^{k-l}, \quad (7)$$

where  $\begin{bmatrix} l \\ m \end{bmatrix}$  is the Stirling number which counts the permutations of size  $l$  with exactly  $m$  disjoint cycles (taking into account the fixed points as cycles with length 1). In particular, knowledge of these first moments is sufficient in order to construct the  $r$ -th order approximation scheme  $\nu_n^{(r)}$ , and we do not need to know all the individuals default probabilities. The first coefficients are:

$$\begin{aligned} b_{1,n} &= 0 & ; & & b_{2,n} &= \frac{1}{2} (M_{2,n} - M_{1,n} - (M_{1,n})^2); \\ b_{3,n} &= \frac{1}{2} ((M_{1,n})^2 - M_{2,n} - M_{2,n} M_{1,n}) + \frac{1}{3} (M_{1,n} + (M_{1,n})^3) + \frac{1}{6} M_{3,n}. \end{aligned}$$

The proof of Eq. (7) is given in Appendix D.

The following proposition provides explicit closed-form expressions for the mod-Poisson approximations of the expectation of two common functions in credit risk, namely the tail function and the call function, which will be used in the numerical simulations of Sections 6 and 7.

**Proposition 4.4** (Estimation formulæ). *Let  $\gamma$  be the lower incomplete gamma function given by:*

$$\gamma(x, \lambda) = \int_0^\lambda t^{x-1} e^{-t} dt.$$

*Then:*

(1) **Tail function.** Fix a value  $x \in \mathbb{R}$ , and let  $f(j) = \mathbf{1}_{\{j > x\}} = \mathbf{1}_{\{j > \lfloor x \rfloor\}}$ . Then,

$$\nu_n^{(r)}(f) = \frac{1}{\lfloor x \rfloor!} \gamma(\lfloor x \rfloor + 1, \lambda_n) + e^{-\lambda_n} \sum_{j=\lfloor x \rfloor - r + 1}^{\lfloor x \rfloor} \frac{(\lambda_n)^j}{j!} \Delta_n(r, f)(j). \quad (8)$$

where  $\Delta_n(r, f)$  is the correction term for  $f$  as in Proposition 3.5.

(2) **Call function.** Fix a value  $K \in \mathbb{R}$ , and let  $f(j) = (j - K)^+$ , which is used in the pricing of call options and CDOs. Then,

$$\begin{aligned} \nu_n^{(r)}(f) &= \frac{\lambda_n}{(\lceil K \rceil - 2)!} \gamma(\lceil K \rceil - 1, \lambda_n) - \frac{K}{(\lceil K \rceil - 1)!} \gamma(\lceil K \rceil, \lambda_n) \\ &\quad + e^{-\lambda_n} \sum_{j=\lceil K \rceil - r + 1}^{\lfloor K \rfloor} \frac{(\lambda_n)^j}{j!} \Delta_n(r, f)(j) \end{aligned} \quad (9)$$

where  $\Delta_n(r, f)$  is the correction term for  $f$  as in Proposition 3.5.

*Proof.* We must compute in each case the following expectations:

$$\nu_n^{(r)}(f) = \mathbb{E}[f(Y_n)] + \mathbb{E}[\Delta_n(r, f)(Y_n)], \quad (10)$$

where  $Y_n \sim \text{Po}(\sum_{i=1}^n p_i)$ . Indeed, for the tail function, Proposition 3.5 applies readily because  $f$  is bounded. For the call function, we can use the remark at the end of the proof of Proposition 3.5: the formula is valid because  $f$  is bounded by a polynomial function. The term  $\mathbb{E}[f(Y_n)]$  can be computed in closed form for both functions. Indeed the tail function of a Poisson random variable is known to admit a closed-form expression in terms of the incomplete lower gamma function (see Proposition B.1 for the formula and its proof), while for the call function, one can proceed as follows:

$$\begin{aligned} \mathbb{E}[(Y_n - K)^+] &= \sum_{j=0}^{\infty} (j - K)^+ e^{-\lambda_n} \frac{\lambda_n^j}{j!} \\ &= \sum_{j=\lceil K \rceil}^{\infty} (j - K) e^{-\lambda_n} \frac{(\lambda_n)^j}{j!} \\ &= e^{-\lambda_n} \sum_{j=\lceil K \rceil}^{\infty} \frac{\lambda_n^j}{(j-1)!} - K \mathbb{P}(Y_n > \lceil K \rceil - 1) \\ &= e^{-\lambda_n} \lambda_n \sum_{j=\lceil K \rceil - 1}^{\infty} \frac{(\lambda_n)^j}{j!} - K \mathbb{P}(Y_n > \lceil K \rceil - 1) \\ &= \lambda_n \mathbb{P}(Y_n > \lceil K \rceil - 2) - K \mathbb{P}(Y_n > \lceil K \rceil - 1). \end{aligned}$$

The term  $\mathbb{E}[\Delta_n(r, f)(Y_n)]$  of Equation (10) corresponds to the summations in Equations (8) and (9), the only difference being that the integration with respect to the distribution of  $Y_n$  has been explicitly restricted to the integer-valued interval  $[\lfloor x \rfloor - r + 1, \lfloor x \rfloor]$ . This follows from the fact that the correction term  $\Delta_n(r, f)$  actually vanishes outside that interval for both the tail and the call function, as can be seen by their explicit formulæ. Indeed, for the tail function one has

$$\Delta_n(r, f)(j) = \sum_{k=1}^r b_{k,n} \sum_{l=0}^k (-1)^{k-l} \binom{k}{l} \mathbf{1}_{\{j+l > \lfloor x \rfloor\}}.$$

On the one hand if  $j \leq \lfloor x \rfloor - r$ , then  $j + l \leq \lfloor x \rfloor$  for all values of  $l$ , so  $\Delta_n(r, f)(j)$  vanishes. On the other hand, if  $j \geq \lfloor x \rfloor + 1$ , then  $j + l > \lfloor x \rfloor$  for all values of  $l$ , so that the inner summation yields  $\sum_{l=0}^k (-1)^{k-l} \binom{k}{l} = 0$  for all  $k = 1, \dots, r$  and  $\Delta_n(r, f)(j)$  is zero.

For the call function, instead, we have that

$$\Delta_n(r, f)(j) = \sum_{k=1}^r b_{k,n} \sum_{l=0}^k (-1)^{k-l} \binom{k}{l} (j+l-K)^+.$$

On the one hand if  $j \leq \lfloor K \rfloor - r$ , then  $j+l \leq \lfloor K \rfloor$  for all values of  $l$ , so  $(j+l-K)^+$  is identically zero and  $\Delta_n(r, f)(j)$  vanishes. On the other hand, if  $j \geq \lceil K \rceil$ , then  $j+l \geq \lceil K \rceil$  for all values of  $l$ , so that the inner summation becomes  $\sum_{l=0}^k (-1)^{k-l} \binom{k}{l} (j+l-K)$  which vanishes because  $k \geq 2$  ( $b_{1,n} = 0$ ) and  $\sum_{l=0}^k (-1)^{k-l} \binom{k}{l} P(l) = 0$  for all polynomials  $P$  of order strictly lower than  $k$  (see, for instance, Corollary 2 in [Rui96]).  $\square$

Formulae (8) and (9) are particularly suitable for numerical implementation. Indeed, the lower incomplete gamma function can be estimated efficiently from its power series expansion (as shown in [Tem94], on which the python implementation of the gamma function in `scipy` is based) while the correction term  $\Delta_n(r, f)$ , which requires numerical integration, is non-zero on at most  $r$  points and is therefore easy to integrate.

## 5. MOD-COMPOUND POISSON APPROXIMATION

This section deals with the extension of the results presented in Section 4 to the case of credit portfolios  $L_n = \sum_{i=1}^n Z_i Y_i$  with random i.i.d. exposures  $(Z_i)_{i=1}^n$ . In this case it is possible to prove a mod-compound Poisson convergence of the sequence  $(L_n)_{n \in \mathbb{N}}$ , as shown in the following theorem.

**Theorem 5.1 (Mod-compound Poisson convergence).** *Define a sequence  $(L_n)_{n \in \mathbb{N}}$  of total portfolio losses given by  $L_n = \sum_{i=1}^n Z_i Y_i$ , where we assume the  $Y_i$ 's and  $Z_i$ 's to be mutually independent,  $Y_i \sim \text{Be}(p_i)$ , and the  $Z_i$ 's are i.i.d. random variables with the same law as  $Z$ , for  $Z$  a given  $\mathbb{N}$ -valued random variable.*

- (1) *If  $\sum_{i=1}^{\infty} p_i = +\infty$  and  $\sum_{i=1}^{\infty} (p_i)^2 < +\infty$ , then  $(L_n)_{n \in \mathbb{N}}$  converges mod- $\phi$  with parameters  $\lambda_n = \sum_{i=1}^n p_i$  and with reference infinitely divisible law the compound Poisson distribution  $\text{CPo}(1, Z)$ .*
- (2) *Furthermore,  $(L_n)_{n \in \mathbb{N}}$  admits a mod-compound Poisson approximation scheme of order  $r$  with the following coefficients:*

$$b_{k,n} = \frac{1}{k!} \sum_{\sigma \in \Pi(k)} \sum_{\substack{\tau \in \Pi(k) \\ \tau \leq \sigma}} \mu(\tau, \sigma) \prod_{D \in \tau} \mathbb{E}[(Z)_{|D|}] \prod_{B \in \sigma} \mathfrak{p}_{n_{\tau}^{\sigma}(B), n},$$

where  $\Pi(k)$  is the poset of set partitions of  $\{1, \dots, k\}$ ;  $n_{\tau}^{\sigma}(B)$  denotes the number of blocks of  $\tau$  contained in the block  $B$  of  $\sigma$ ;  $\mu(\tau, \sigma) = (-1)^{|\tau|-|\sigma|} \prod_{B \in \sigma} (n_{\tau}^{\sigma}(B) - 1)!$  is the Möbius function for the incidence algebra of the poset  $\Pi(k)$ , and  $\mathfrak{p}_{k,n} := \sum_{j=1}^n (p_j)^k$ , with the convention  $\mathfrak{p}_{1,n} := 0$ .

*Proof.* We compute the deconvolution residue as in Equation (1), using the explicit form of the Lévy–Khintchine exponent of  $\text{CPo}(1, Z)$ . We obtain:

$$\begin{aligned}
\psi_n(\xi) &= \widehat{\mu}_{L_n}(\xi) e^{-\sum_{i=1}^n p_i(\widehat{\mu}_Z(\xi)-1)} \\
&= \prod_{j=1}^n (1 + p_j(\widehat{\mu}_Z(\xi) - 1)) e^{-p_j(\widehat{\mu}_Z(\xi)-1)} \\
&= \exp\left(\sum_{j=1}^n \log(1 + p_j(\widehat{\mu}_Z(\xi) - 1)) - p_j(\widehat{\mu}_Z(\xi) - 1)\right) \\
&= \exp\left(\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} (\widehat{\mu}_Z(\xi) - 1)^k \mathbf{p}_{k,n}\right), \tag{11}
\end{aligned}$$

where we have defined  $\mathbf{p}_{k,n} := \sum_{j=1}^n (p_j)^k$ , with the convention  $\mathbf{p}_{1,n} := 0$ .

The condition  $\sum_{j=1}^{\infty} (p_j)^2 < +\infty$  guarantees that  $\psi_n(\xi)$  converges to a limit,  $\psi(\xi)$ , uniformly in  $\mathbb{T}$ , so the mod- $\phi$  convergence is proved. Let us then extract the coefficients  $(b_{k,n})_{k=1}^r$  for the mod- $\phi$  approximation scheme of order  $r$ : thus, we need to compute the power series expansion of  $\psi_n(\xi)$  in powers of  $z = e^{i\xi} - 1$ . We start by noticing that

$$\widehat{\mu}_Z(\xi) = \sum_{j=0}^{\infty} \frac{1}{j!} \mathbb{E}[(Z)_j] (e^{i\xi} - 1)^j,$$

where  $(Z)_j := Z(Z-1)\cdots(Z-k+1)$  is the  $j$ -th falling factorial and  $\mathbb{E}[(Z)_j]$  is therefore the  $j$ -th factorial moment of  $Z$ . As a consequence, the series in the exponential (11) rewrites as:

$$\begin{aligned}
&\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} \left(\sum_{j=1}^{\infty} \frac{1}{j!} \mathbb{E}[(Z)_j] z^j\right)^k \mathbf{p}_{k,n} \\
&= \sum_{l=1}^{\infty} z^l \left(\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} \mathbf{p}_{k,n} \sum_{\substack{l=j_1+\dots+j_k \\ j_1, \dots, j_k \geq 1}} \frac{\mathbb{E}[(Z)_{j_1}] \cdots \mathbb{E}[(Z)_{j_k}]}{(j_1)! \cdots (j_k)!}\right) \\
&= \sum_{l=1}^{\infty} z^l \left(\sum_{k=1}^l (-1)^{k-1} (k-1)! \mathbf{p}_{k,n} \sum_{\substack{l=\lambda_1+\dots+\lambda_k \\ \lambda_1 \geq \dots \geq \lambda_k \geq 1}} \frac{\mathbb{E}[(Z)_{\lambda_1}] \cdots \mathbb{E}[(Z)_{\lambda_k}]}{(m_1(\lambda))! \cdots (m_l(\lambda))! (\lambda_1)! \cdots (\lambda_k)!}\right).
\end{aligned}$$

In these formulæ, the sum on the second line runs over compositions  $l = j_1 + \cdots + j_k$  of size  $l$  and length  $k$  (sequences that sum to  $l$ ), and the sum on the third line runs over integer partitions  $l = \lambda_1 + \cdots + \lambda_k$  with  $\lambda = (\lambda_1 \geq \cdots \geq \lambda_k)$  (non-increasing sequences). For  $i \leq l$ ,  $m_i(\lambda)$  is the number of parts  $\lambda_j$  equal to  $i$ . Similar combinatorial arguments are detailed in the proof of Theorem C.6; see Appendix C. Now, since the number of set partitions  $\pi \in \Pi(l)$  with sizes of blocks given by an integer partition  $\lambda = (\lambda_1 \geq \cdots \geq \lambda_k)$  is

$$\frac{l!}{(m_1(\lambda))! \cdots (m_l(\lambda))! \lambda_1! \cdots \lambda_k!},$$

and since the Möbius function  $\mu(\pi, \widehat{1}_l)$  of a set partition with  $k = |\pi|$  blocks is  $(-1)^{k-1} (k-1)!$ , we can rewrite:

$$\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} \left(\sum_{j=1}^{\infty} \frac{1}{j!} \mathbb{E}[(Z)_j] z^j\right)^k \mathbf{p}_{k,n} = \sum_{l=1}^{\infty} \frac{z^l}{l!} \left(\sum_{\pi \in \Pi(l)} \mu(\pi, \widehat{1}_l) \mathbf{p}_{|\pi|,n} \left(\prod_{B \in \pi} \mathbb{E}[(Z)_{|B|}]\right)\right).$$

Substituting this expansion in Equation (11) and using the first part of Theorem C.6 with  $g_l = \sum_{\pi \in \Pi(l)} \mu(\pi, \hat{1}_l) \mathfrak{p}_{|\pi|,n}(\prod_{B \in \pi} \mathbb{E}[(Z)_{|B|}])$ , we get:

$$b_{k,n} = \frac{1}{k!} \sum_{\sigma \in \Pi(k)} \prod_{B \in \sigma} g_{|B|} = \frac{1}{k!} \sum_{\sigma \in \Pi(k)} \prod_{B \in \sigma} \left( \sum_{\tau_B \in \Pi(|B|)} \mu(\tau_B, \hat{1}_{|B|}) \mathfrak{p}_{|\tau_B|,n} \left( \prod_{C \in \tau_B} \mathbb{E}[(Z)_{|C|}] \right) \right).$$

We notice that choosing one partition  $\tau_B \in \Pi(|B|)$  for each block  $B \in \sigma$  is equivalent to choosing a subpartition  $\tau$  of  $\sigma$ , so the coefficient can be expressed more compactly as:

$$b_{k,n} = \frac{1}{k!} \sum_{\sigma \in \Pi(k)} \sum_{\substack{\tau \in \Pi(k) \\ \tau \leq \sigma}} \mu(\tau, \sigma) \prod_{D \in \tau} \mathbb{E}[(Z)_{|D|}] \prod_{B \in \sigma} \mathfrak{p}_{n_{\tau}^{\sigma}(B),n}$$

where  $n_{\tau}^{\sigma}(B)$  denotes the number of blocks of  $\tau$  contained in the block  $B$  of  $\sigma$ ; and where we have furthermore recognized the Möbius function for the incidence algebra of the poset  $\Pi(k)$ , given by  $\mu(\tau, \sigma) = (-1)^{|\tau| - |\sigma|} \prod_{B \in \sigma} (n_{\tau}^{\sigma}(B) - 1)! = \prod_{B \in \sigma} \mu(\tau_B, \hat{1}_{|B|})$ .  $\square$

The implementation of mod-compound Poisson approximation schemes for credit risk applications is numerically more problematic than in the case of mod-Poisson schemes. In particular, integrals with respect to the reference infinitely divisible law are more difficult to evaluate and closed-form expressions, such as formulæ (8) and (9), are not available anymore, except in few very specialized cases. Nevertheless these integrals can be evaluated numerically (for instance by estimating the compound Poisson law via Panjer recursion) but, depending on the application at hand and the particular distribution of the random exposures, this might lead to estimations that are as computationally expensive as the recursive methodology.

## 6. APPLICATION: ESTIMATION OF RISK MEASURES

**6.1. Background.** A commonly used risk measure for market and credit risk applications is the *Value at Risk* (VaR), defined as

$$\text{VaR}_{\alpha}(L_n) := \inf \{t \in \mathbb{R} \mid \mathbb{P}(L_n \leq t) \geq \alpha\},$$

which quantifies the minimum capital required to cover all portfolio losses with a probability at least equal to  $\alpha$ . The parameter  $\alpha$  is known as the *confidence level* and higher values of this parameter correspond to higher and more stringent capital requirements. The definition of  $\text{VaR}_{\alpha}(L_n)$  is mathematically equivalent to the generalized inverse of the distribution function of  $L_n$ , which is also known as the  $\alpha$ -quantile of  $L_n$ .

The VaR is the most commonly used risk measure in financial practice, despite the fact that it is not a coherent risk measure [ADEH99], which means that it fails to account for diversification effects when risk is aggregated across several portfolios. A risk measure that is commonly employed to solve this problem is the *Expected Shortfall* (ES), also known as *conditional VaR*, which is defined as follows:

$$\text{ES}_{\alpha}(L_n) := \frac{1}{1 - \alpha} \int_{\alpha}^1 \text{VaR}_u(L_n) du = \mathbb{E}[L_n | L_n > \text{VaR}_{\alpha}(L_n)].$$

One can think of  $\text{ES}_{\alpha}(L_n)$  at confidence level  $\alpha$  as the expected value of portfolio losses, given that these losses already exceed  $\text{VaR}_{\alpha}(L_n)$ . One can show that  $\text{ES}_{\alpha}(L_n)$  is always greater or equal to  $\text{VaR}_{\alpha}(L_n)$  for all confidence levels  $\alpha$ , so that ES is more conservative than VaR. It is also a coherent risk measure, because it incorporates information about all potential losses, including the ones above the confidence level  $\alpha$ , and it is therefore the simplest modification of VaR that yields a theoretically acceptable risk measure. The computation of ES is more demanding than the computation of VaR, because it requires the availability of accurate estimations for the entire tail function of the loss distribution: this is the reason why simulation-based methods, such as Monte Carlo simulation, typically perform poorly in estimations of portfolio ES.

We remark that if  $L_n$  is a discrete random variable (this is typically the case if the distribution has been discretized for purposes of numerical evaluation) with  $N := \|L_n\|_\infty < \infty$ , then  $\text{VaR}_\alpha(L_n)$  is integer-valued and is given by the following left-continuous step function:

$$\text{VaR}_\alpha(L_n) = \sum_{k=0}^N k \mathbb{1}_{(\mathbb{P}(L_n \leq k-1), \mathbb{P}(L_n \leq k)]}(\alpha), \quad (12)$$

while  $\text{ES}_\alpha(L_n)$  is given by

$$\text{ES}_\alpha(L_n) = \frac{1}{1-\alpha} \left( (\mathbb{P}(L_n \leq \text{VaR}_\alpha(L_n)) - \alpha) \text{VaR}_\alpha(L_n) + \sum_{k > \text{VaR}_\alpha(L_n)}^N k \mathbb{P}(L_n = k) \right). \quad (13)$$

**6.2. Estimation of risk measures.** It is clear from Equations (12) and (13) that the estimation of the VaR and the ES depends crucially on accurate estimates for the tail function  $\mathbb{P}(L_n \geq x)$  of portfolio losses. It is therefore instructive to compare estimation methods first of all on the task of tail function estimation for a representative credit portfolio model, in our case a single-factor Gaussian copula with  $n = 250$  obligors, heterogeneous average default probabilities uniformly distributed in  $[2\%, 8\%]$  and equicorrelation parameter  $\rho = 0.3$ . The results do not vary qualitatively for different choices of the parameters.

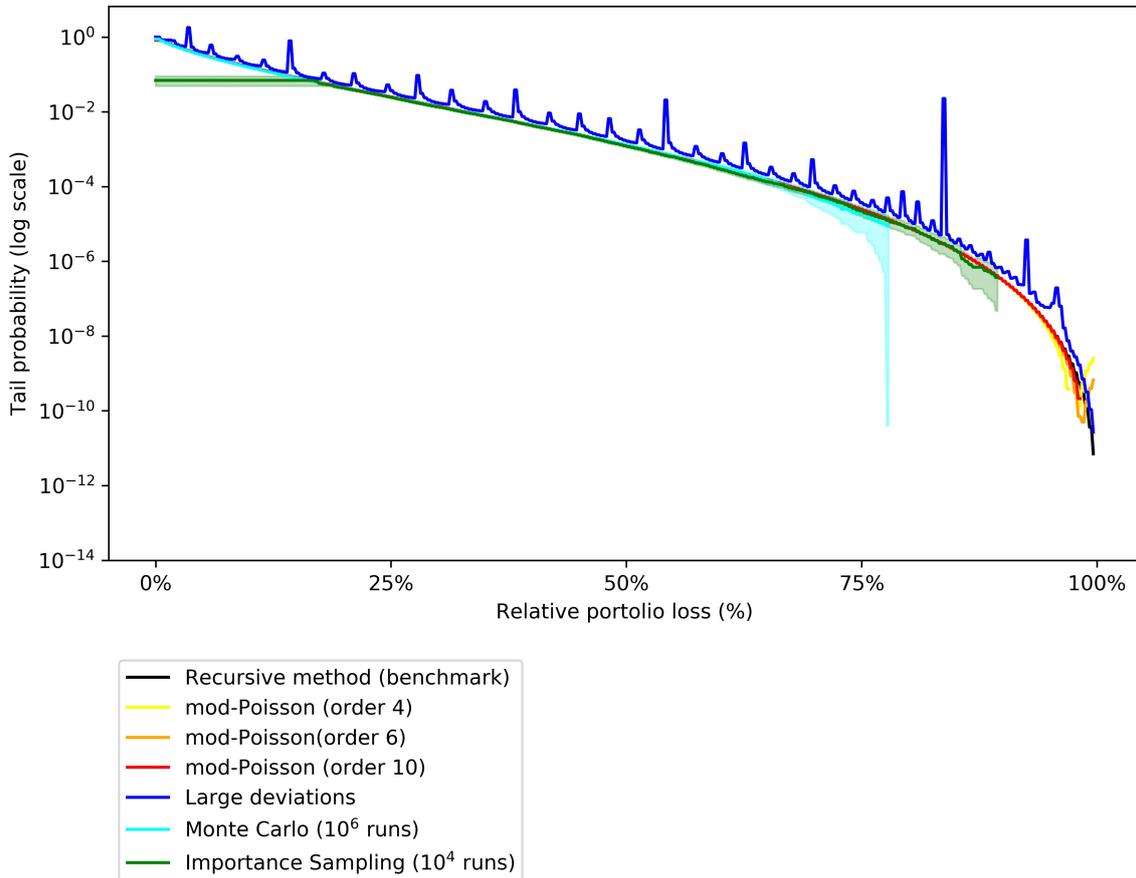


FIGURE 1. Estimated tail functions on a logarithmic scale. For simulation-based methods (i.e. Monte Carlo and Importance Sampling) mean estimates are reported within their 99% asymptotic confidence interval (shaded areas).

Figure 1 shows a comparison of the estimated tail functions for an exact method (the recursive method), two semi-analytical methods (mod-Poisson with varying order and large deviations approximation) and two simulation-based methods (Monte Carlo and importance sampling). The

reader is referred to Appendix A for a self-contained presentation of all these estimation models, together with full details of their numerical implementation.

- The recursive method can be used as a benchmark to assess the accuracy of the other methods, because it is an exact procedure for the computation of the loss distribution, up to the numerical integration error due to integration over the portfolio mixing variable and any rounding errors due to finite machine precision, both of which are in practice of order  $10^{-15}$ . A better measure of the performance of each estimation method can be obtained by looking at the signed relative errors of the estimated tail probabilities (computed with respect to the benchmark) as a function of the tail point, as plotted in Figure 2.

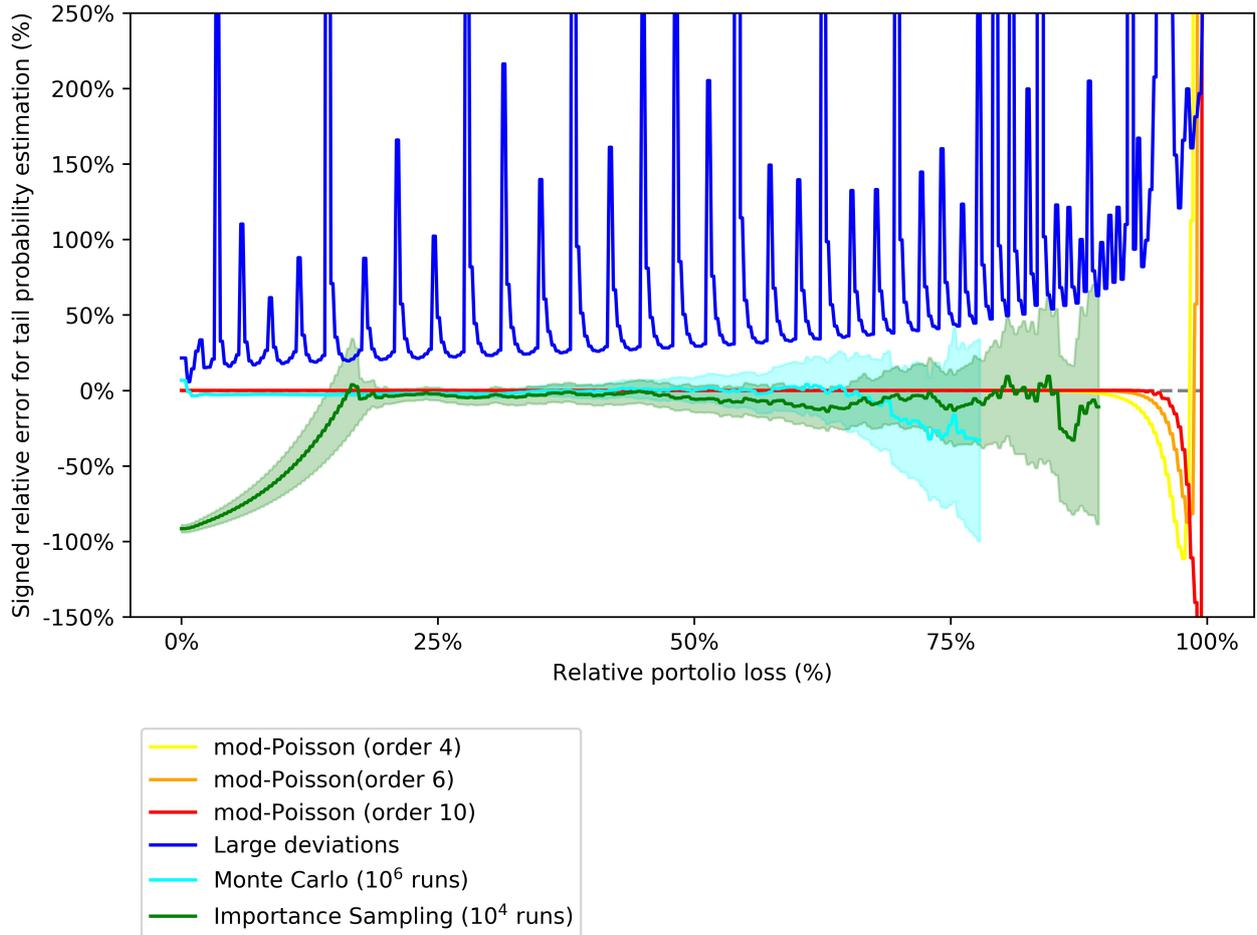


FIGURE 2. Signed relative errors of estimated tail probabilities for all levels of relative portfolio losses. Benchmark value (i.e. the assumed true value with respect to which errors are computed) computed via the recursive method. For simulation-based methods (i.e. Monte Carlo and Importance Sampling) mean estimates are reported within their 99% asymptotic confidence interval (shaded areas).

- The plain Monte Carlo method has been implemented as in Algorithm 2 with  $10^6$  simulation runs. Due to the finite number of simulations, it is able to estimate the probability of relative losses only up to the 75% level and with increasing uncertainty, as shown by the widening of the asymptotic confidence intervals around the mean estimates. The mean estimates themselves show a negative bias, due to the undersampling of rare, large losses. Higher accuracy can be obtained by suitably increasing the number of simulations at the price of higher computational times.

- The importance sampling method, by addressing the rare-event simulation problem as explained in Section A.6, is able to estimate tail probabilities at higher levels of relative losses with only  $10^4$  simulation runs. Nevertheless, this method is also characterized by widening confidence intervals, with the result that only estimates close to 85% of relative losses can be conceivably used with any statistical confidence. We further remark that higher uncertainty is visible in Figure 2 not only for high levels of relative losses, but also for low ones. This is because after performing the exponential tilting described in Algorithm 4 the mean of the loss distribution has been shifted and low losses have become rare events. In practice one could fix this issue by gluing together the estimates coming from a plain Monte Carlo method for low levels of losses and the estimates from the importance sampling method for high levels, as suggested in [GL05].
- The large deviations approximation is characterized by an erratic behavior. This is partly due to the discreteness of the distribution  $L_n$ , but also to the numerical integration over the mixing variable of the Gaussian copula, which is unstable, possibly because of the denominator in Equation (19). In general one can expect this method to perform well only in the asymptotic regime as the number of obligors becomes large, but for finite portfolios of hundreds of obligors the performance is disappointing and the estimates are biased and affected by large relative errors.
- The mod-Poisson approximation schemes show a remarkable accuracy for most of the tail function, except at the level of relative losses higher than 90%, i.e. losses of probability lower than  $10^{-8}$  for our representative portfolio, as can be deduced from Figure 1). Above that level the approximation order becomes an important tuning parameter: approximation schemes with higher order are able to maintain higher accuracy farther in the tail. Nevertheless, the performance at low approximation orders – such as  $r = 4, 6, 10$  as shown in Figure 2 – is already very satisfactory, since mod-Poisson approximation schemes show high relative errors only on probabilities of order smaller than  $10^{-8}$ , whose contribution to the estimation of risk measures is negligible, as will be shown next.

Table 1 compares the estimates of the VaR and ES (obtained using Equations (12) and (13)) of the representative portfolio for a selection of typically used confidence levels, ranging from 95% to 99.9999%, while Figure 3 shows the signed relative errors in the estimation of VaR as a function of the confidence levels on a logarithmic scale.

The best performing methods are clearly the mod-Poisson approximation schemes, which yield estimates that are almost always identical to the benchmark value for any confidence value, already for order  $r = 4$ . The other methods are characterized by diminishing accuracy at higher confidence levels and increasing uncertainty in the estimates, in the case of simulation-based methods.

VaR level	Benchmark (recursive)	Large deviations	Monte Carlo ( $10^6$ runs) Mean [99% CI]	Importance Sampling ( $10^4$ runs) Mean [99% CI]	Mod-Poisson (order=4)	Mod-Poisson (order=6)	Mod-Poisson (order=10)
95%	48	54	47	47	48	48	48
99%	82	89	82 [81, 83]	81 [80, 83]	82	82	82
99.99%	169	173	168 [164, 172]	168 [165, 170]	169	169	169
99.9999%	218	222	n.a.	215 [212, 221]	218	218	218

ES level	Benchmark (recursive)	Large deviations	Monte Carlo ( $10^6$ runs) Mean [99% CI]	Importance Sampling ( $10^4$ runs) Mean [99% CI]	Mod-Poisson (order=4)	Mod-Poisson (order=6)	Mod-Poisson (order=10)
95%	68.91	79.08	68.50 [67.99, 69.01]	68.16 [66.06, 69.28]	68.91	68.91	68.91
99%	103.06	118.92	102.81 [101.85, 103.76]	102.38 [100.86, 103.84]	103.06	103.06	103.06
99.99%	181.65	198.71	160.70 [147.39, 172.90]	179.99 [177.49, 182.09]	181.64	181.65	181.65
99.9999%	223.73	231.13	n.a.	n.a.	223.57	223.70	223.73

TABLE 1. Estimates for VaR (above) and ES (below) across estimation methods for a selection of confidence levels (95%, 97.5%, 99%, and 99.99%). For simulation-based methods, such as Monte Carlo and Importance Sampling, the mean estimate is shown together with its 95% asymptotic confidence interval in square brackets.

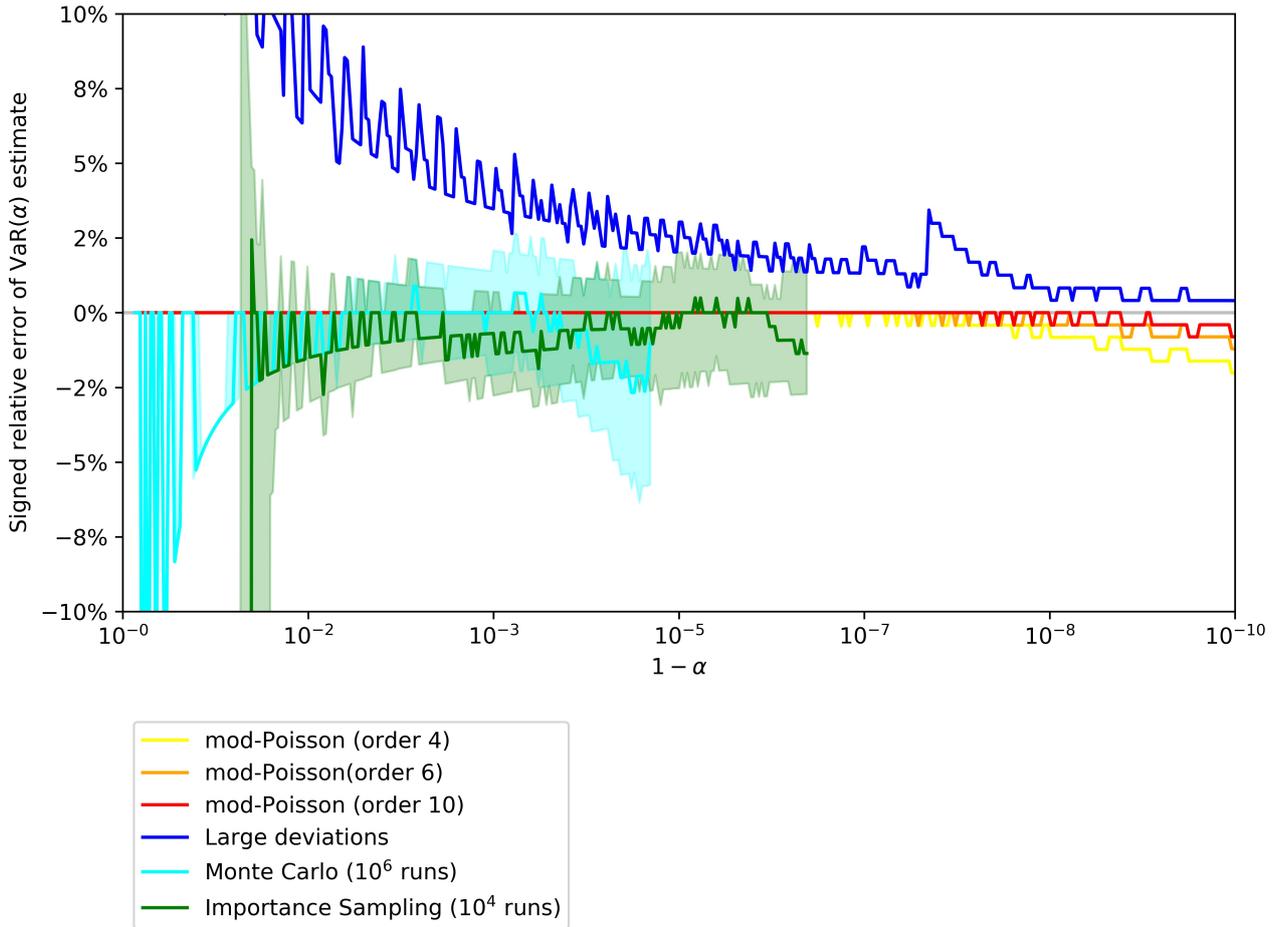


FIGURE 3. Signed relative errors of estimated VaR for confidence levels on a logarithmic scale. Benchmark value (i.e. the assumed true value with respect to which errors are computed) computed via the recursive method. For simulation-based methods (i.e. Monte Carlo and Importance Sampling) mean estimates are reported within their 99% asymptotic confidence interval (shaded areas).

**6.3. Computational time.** The performance of an estimation method in terms of computational time can be particularly important for certain financial applications. In this section we compare empirically the performance of various methods for the estimation of tail probabilities and discuss relative advantages and disadvantages. All empirical tests are performed on a representative portfolio model, specifically a single-factor Gaussian copula with average default probabilities uniformly distributed in  $[2\%, 8\%]$  and equicorrelation  $\rho = 0.3$ .

Figure 4 compares the elapsed CPU clock time for the estimation of a single tail probability of portfolio losses (i.e. an evaluation of  $\mathbb{P}(L_n > x)$  for a given value  $x \geq 0$ ) as a function of the number of obligors.

- The recursive methodology is the only known method for the exact estimation (up to numerical integration error) of the loss distribution, but its computational complexity scales quadratically in the number of obligors, as we already anticipated in the remarks following Algorithm 1 and as can be seen in Figure 4. For this reason the recursive method is considered computationally expensive in the case of large portfolios (i.e. with more than one hundred obligors) and approximate methods – either semi-analytical or simulation-based – are typically preferred.

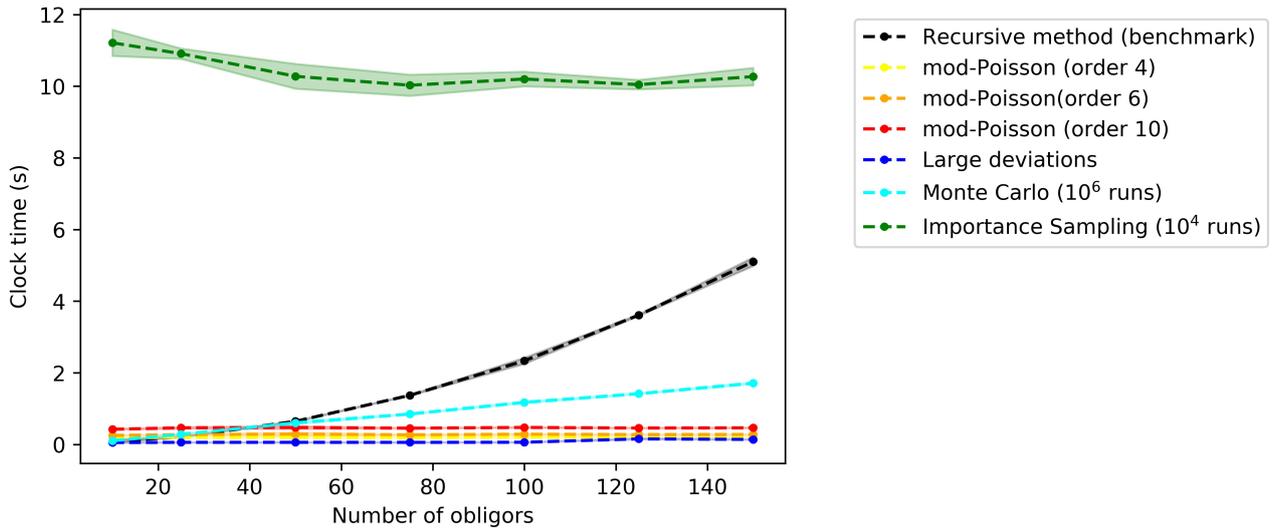


FIGURE 4. CPU clock time for estimation of a single tail probability for a selection of estimation methods as a function of the number of obligors. Shaded areas are 95% asymptotic confidence intervals from 10 runs of the estimation methods.

Furthermore, an important disadvantage of the recursive method comes from the fact that it is designed to output the full distribution of the portfolio losses, which might be wasteful in some applications. For instance, when computing risk measures it is necessary to compute the tail function of the loss distribution only at a few points – more specifically in the part of the tail corresponding to high losses – so that information of the full distribution is effectively useless. In contrast, semi-analytical methods, such as mod-Poisson approximation schemes or the large deviation approximation, yield approximations of the tail function at a single point and can thus be used to compute risk measures much more efficiently. Another application for which the recursive methodology tends to be inefficient can be found in model risk management, where sensitivity analysis requires computing risk measures and other portfolio metrics repeatedly for a given credit risk model under slight perturbations of the model parameters. In this case the recursive methodology requires the expensive computation of a large number of very similar loss distributions, thus compounding the wastefulness issue discussed above.

- As shown in Figure 4, simulation-based methods – such as Monte Carlo integration and importance sampling – scale more favorably in the number of obligors. In particular, both methods scale linearly in the number of obligors, since they both require simulating a matrix of obligors’ default indicators with number of columns equal to the number of obligors and number of rows equal to the number of simulation runs.

While for the Monte Carlo method this linear dependence is evident in Figure 4, in the case of the importance sampling method it is concealed by the computational overhead stemming from the determination of the shifted mean  $\mu$  in the first part of Algorithm 4. This preliminary optimization step turns out to be computationally expensive and makes the importance sampling algorithm unappealing for financial applications that require fast execution, such as product pricing for trading desks or the estimation of pre-trade risk for proprietary trading. Furthermore, it is worthwhile recalling that the importance sampling algorithm applies exclusively to the Gaussian factor copula model, so that its field of application is in any case already quite limited. As far as the Monte Carlo method is concerned, a correct assessment of its time complexity must take into account the dependence on the number of simulations needed. While the importance sampling procedure can produce accurate estimations far into the tail with a low number of simulations (in practice of the

order  $10^4$ ), the plain Monte Carlo approach requires a much higher number of simulation runs, as explained in Section A.5. This is particularly problematic when high accuracy is required as in the computation of the Expected Shortfall of a portfolio and other distortion measures with non-zero spectrum at high quantile levels.

- Semi-analytical methods, such as mod-Poisson approximation schemes and the large deviations approximation, boast the best performance in terms of computational time, due to the fact that they require only the evaluation of known functions and a numerical integration over the copula factor, both of which can be performed efficiently. To be more precise, these methods depend on the number of obligors only through the computation of a few coefficients. In the case of mod-Poisson approximation schemes it is necessary to compute the coefficients  $p_{k,n}$  for  $k = 2, \dots, r$ , where  $r$  is the approximation scheme order, while in the case of the large deviations approximation it is only needed to evaluate the cumulant generating function  $F$ , its second derivative  $F''$  and the optimal tilting  $\lambda_x$ . Nevertheless, these operations amount to computing specific functions of the vector of default probabilities  $(p_1, p_2, \dots, p_n)$  and are easily vectorized on any modern CPU, which results in a linear dependence on the number of obligors with a very small coefficient. Indeed this linear dependence is empirically negligible for portfolios of even thousands of obligors and is effectively invisible in Figure 4.

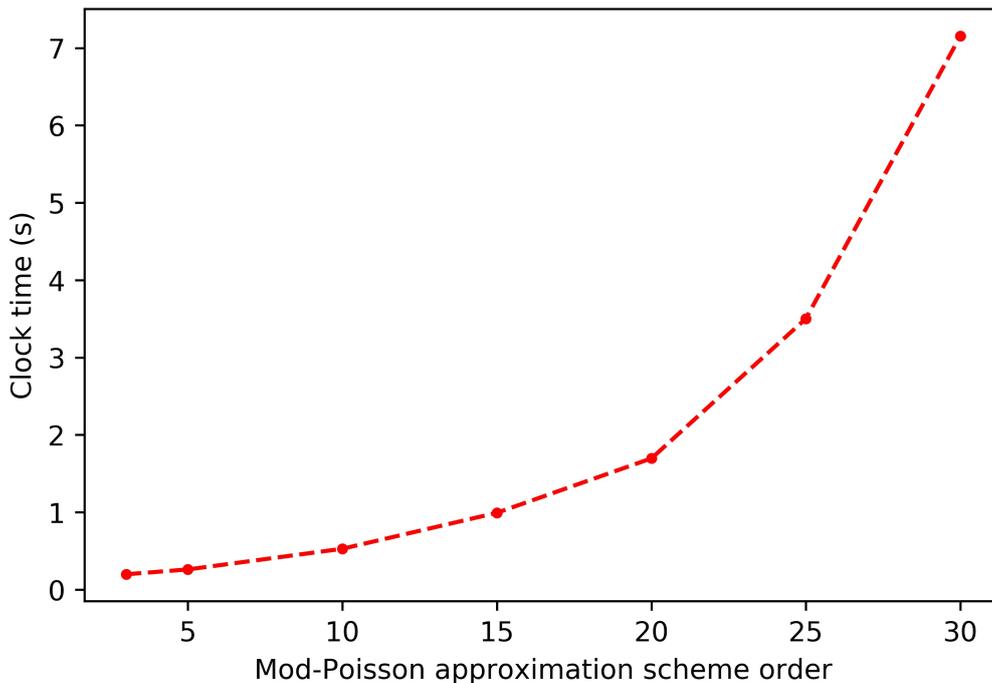


FIGURE 5. CPU clock time for estimation of a single tail probability for mod-Poisson approximation schemes as a function of the order.

The accuracy of mod-Poisson approximation schemes increases in the approximation order, as shown empirically in Figure 2. It is therefore natural to investigate how the computational time burden increases when the order is increased. From Theorem 4.1 and Remark 4.2 it is clear that the computation of the coefficient  $b_{k,n}$  depends on  $k$  as the number of integer partitions of  $k$  with minimum block size 2. In practice integer partitions can be generated efficiently by encoding them as ascending compositions – rather than descending compositions, as is conventionally done – and by exploiting some properties of this representation, as explained in [KO09]. One can then select only integer partitions with minimum block size 2 by acceptance-rejection. This implementation

yields an exponential time complexity in the square-root of  $r$ , as checked empirically for orders up to  $r = 30$  in Figure 5.

## 7. APPLICATION: CDO PRICING

**7.1. Background.** A CDO (Collateralized Debt Obligation) is a financial product for the *securitization* of credit portfolios, such as pools of residential mortgages or consumer loans. We refer the reader to [MFE15, Chapter 12] for an introduction to CDO pricing. Here we only mention that CDO pricing requires the valuation and comparison of payment cashflows, therefore the temporal evolution of the credit portfolio becomes important. Therefore we introduce the time dependence by denoting the total portfolio losses up to time  $t$  as:

$$L_{t,n} = \sum_{i=1}^n Z_i Y_{t,i}.$$

Let us denote by  $m$  the total number of tranches in the CDO, then the *notional value* of the  $j$ -th tranche at time  $t$  as a function of the underlying portfolio losses is given by:

$$N_t^{(j)}(L_{t,n}) = \begin{cases} K_j - K_{j-1} & \text{if } L_{t,n} < K_{j-1}, \\ K_j - L_{t,n} & \text{if } K_{j-1} \leq L_{t,n} \leq K_j, \\ 0 & \text{if } L_{t,n} > K_j, \end{cases}$$

where  $0 = K_0 < K_1 < \dots < K_m$ . The two values  $K_{j-1}$  and  $K_j$  are called the *attachment* and *detachment* points respectively. If the portfolio losses are below the attachment point, the tranche has a fixed value of  $K_j - K_{j-1}$ . As losses increase above that level, the tranche must absorb them and correspondingly loses value, up until the losses reach the detachment point and the tranche has become worthless.

The notional value of tranches can also be expressed more compactly as follows:

$$N_t^{(j)} = (K_j - L_{t,n})^+ - (K_{j-1} - L_{t,n})^+, \quad j = 1, \dots, m, \quad (14)$$

which shows that CDO tranches have the same payoff as put spreads on the underlying credit portfolio. Similarly it is possible to define the cumulative *tranche loss* up to time  $t$ , given by:

$$L_t^{(j)} = K_j - K_{j-1} - N_t^{(j)} = (L_{t,n} - K_{j-1})^+ - (L_{t,n} - K_j)^+, \quad j = 1, \dots, m, \quad (15)$$

which represents the losses incurred by the tranche and has the same payoff as a call spread on the underlying credit portfolio.

In this section we are actually interested in pricing *synthetic* CDOs, which involves the evaluation and comparison of the two cashflows, or *legs*, of the counterparties: the premium payments leg and the default payments leg. The CDO issuer makes premium payments at regular times, say  $0 = t_0 < t_1 < \dots < t_N = T$ . Assuming time is measured in years, we can express these payments in terms of an annualized *spread*, denoted by  $s$ . The premium payment at time  $t_n$  from the  $j$ -th tranche is then equal to  $s(t_n - t_{n-1})N_{t_n}^{(j)}$ , where  $N_{t_n}^{(j)}$  is just the notional value of the tranche at time  $t_n$ , as given in Equation (14). In actual practice premium payments also include so called *accrued payments*. More specifically, if an obligor in the reference portfolio defaults at a random time  $T \in (t_{n-1}, t_n]$ , then at time  $t_n$  the CDO issuer is also required to pay the premium accrued over the time before the default occurred, i.e.  $s(T - t_{n-1})(L_T^{(j)} - L_{T-}^{(j)})$ , where  $L_T^{(j)}$  is the tranche loss at time  $T$ , as given in Equation (15). By assuming a sufficiently thick time grid (i.e.  $N$  is sufficiently large), we can safely ignore accrued payments, as we will do in the following.

The total value at time  $t = 0$  of the premium cashflow can then be computed by taking the expectation of the discounted cashflow under an equivalent martingale measure, obtaining:

$$L_{\text{premium}}(s) = s \sum_{n=1}^N e^{-rt_n} (t_n - t_{n-1}) \mathbb{E} \left[ N_{t_n}^{(j)} \right], \quad (16)$$

where  $r$  denotes the deterministic risk-free interest rate. Notice that the uncertainty due to interest rate risk is many orders of magnitude smaller than the uncertainty due to default dependence. This is why incorporating interest rate risk in credit risk models typically leads to negligible contributions and the assumption of a deterministic interest rate is common in many credit risk applications.

The CDO buyer makes default payments every time an obligor in the reference portfolio defaults. The discounted value at time  $t = 0$  of the default cashflow is given by the following integral:

$$\int_0^T e^{-rt} dL_t^{(j)},$$

which is to be understood as a pathwise Riemann–Stieltjes integral, and which can be approximated as a stochastic Riemann sum over the premium payments time grid:

$$\sum_{n=1}^N e^{-rt_n} (L_{t_n}^{(j)} - L_{t_{n-1}}^{(j)}).$$

The value at time  $t = 0$  of the default cashflow can then be computed, analogously to the premium case, as the expectation of the discounted cashflow under an equivalent martingale measure, yielding:

$$L_{\text{default}} \approx \sum_{n=1}^N e^{-rt_n} \left( \mathbb{E} \left[ L_{t_n}^{(j)} \right] - \mathbb{E} \left[ L_{t_{n-1}}^{(j)} \right] \right). \quad (17)$$

The fair value of the CDO can be deduced by equating the premium leg  $L_{\text{premium}}(s)$  in Equation (16) with the default leg  $L_{\text{default}}$  in Equation (17). The value of the spread  $s$  for which equality holds is:

$$s = \frac{L_{\text{default}}}{L_{\text{premium}}(1)},$$

and is called the *fair spread*. This is the quantity that is quoted in CDO exchanges and used to assess the relative cost a CDO contract.

**7.2. Estimation of call prices.** From the definitions of tranche notional value and tranche loss in Equations (14) and (15) it is clear that the computations of the two payment legs – and therefore of the fair spread – can be reduced to the problem of computing call and put spreads on the reference portfolio. Moreover, since the payoff of a put option can be expressed in terms of the payoff of a call option, this task can be further reduced to the accurate estimation of call prices only. This section is devoted to an empirical comparison of several estimation models on this particular task. All estimations refer to a representative credit portfolio model, a single-factor Gaussian copula with  $n = 100$  obligors, equicorrelation parameter  $\rho = 0.1$  and heterogeneous average default probabilities sampled from a log-normal distribution with varying mean  $p$  and standard deviation  $\sigma = 0.2$ . The choice of a log-normal distribution is of course arbitrary and is done in analogy to (and to ease comparison with) the numerical experiments presented in [EKJK08, EKJ09]. The results do not vary qualitatively for different choices of the parameters.

Figure 6 shows the signed relative error for estimated call prices measured in basis points (bp) as a function of the strike. Each subfigure refers to a different choice of the mean  $p$  of the log-normal distribution from which the average single-obligor default probabilities are sampled. The value of  $p$  is the correct parameter to study the performance of these estimation methods, because as it increases the total portfolio losses move from a Poisson regime to a Gaussian one, correspondingly affecting the performance of the methods.

- For low values of  $p$  all methods based on the Poisson approximation perform best, with negligible errors, while the Stein first-order correction to the Gaussian approximation yields comparatively larger errors, further characterized by an oscillation in the strike due to the approximation to a discrete distribution.

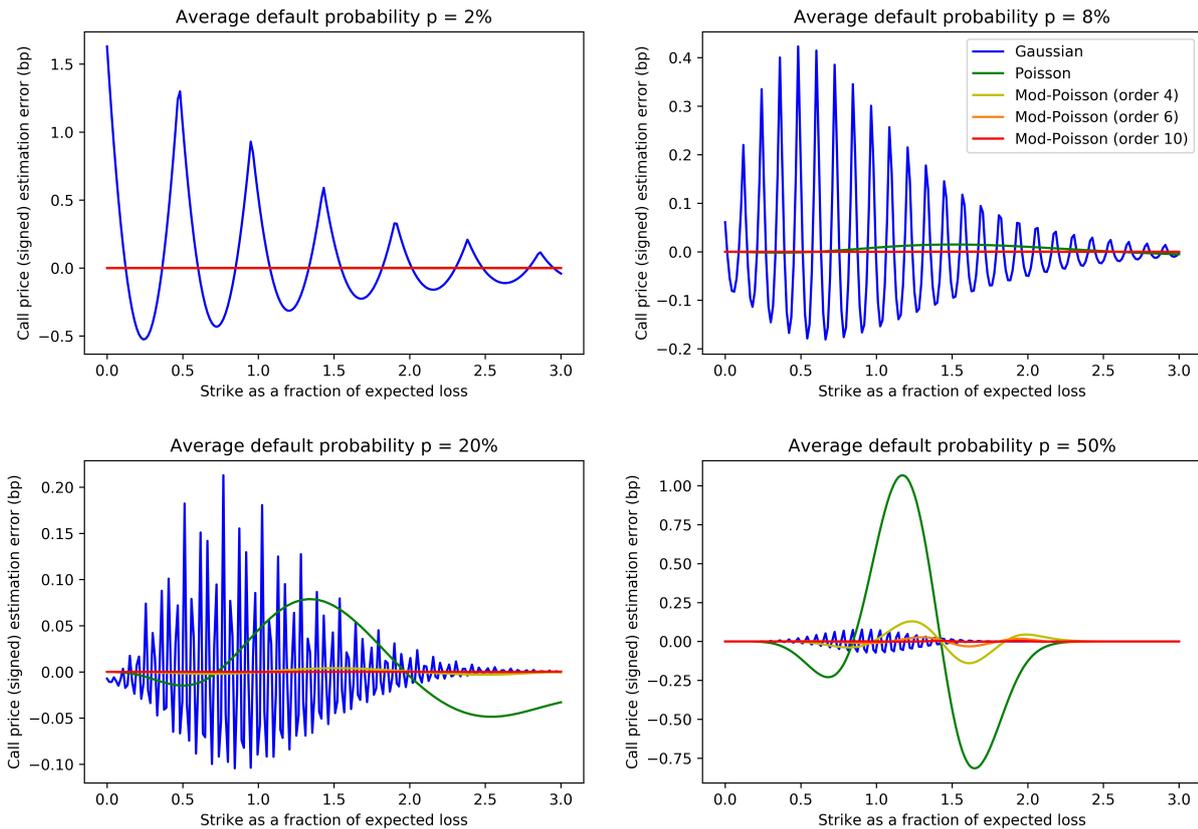


FIGURE 6. Signed relative error for call price estimation measured in basis points (bp) as a function of the strike for different estimation methods and for increasing values of the average default probability in the portfolio. Method “Poisson” denotes the Chen–Stein first order correction to the Poisson approximation (which corresponds also to the mod-Poisson approximation scheme of order 2), while the method “Gaussian” denotes the Chen–Stein first order correction to the Gaussian approximation (see [EKJK08, EKJ09] or Appendix A).

- As the default probability increases, the error associated with the Chen–Stein first-order correction to the Poisson approximation increases, while the Gaussian approximation yields better and better estimates. For this reason the authors in [EKJK08, EKJ09] (where the Chen–Stein’s method is first applied to CDO pricing) propose a hybrid estimation method, in which they suggest to use either the first-order correction to the Gaussian approximation or to the Poisson one, depending on the specific value of  $p$ .
- As shown in Figure 6, mod-Poisson approximation schemes at higher order perform very well even in a Gaussian regime. For instance, in the case of  $p = 50\%$  – which is an extremely high average default probability in any conceivable credit risk setting – the mod-Poisson approximation schemes of order 6 and 10 perform better than the first-order correction to the Gaussian approximation, so that the domain of validity of the Poisson approximation, if properly corrected, includes all credit risk applications, without the need for a hybrid method.

**7.3. Default leg, premium leg, and fair spread.** Accurate estimates of call prices can then be used in Equations (17) and (16) for the determination of the default and premium legs of the CDO and finally for the estimation of the fair spread, which is effectively used to price a CDO contract. Table 2 compares the estimation of default legs, premium legs and fair spreads for five standardized tranches of a CDO written on a representative portfolio.

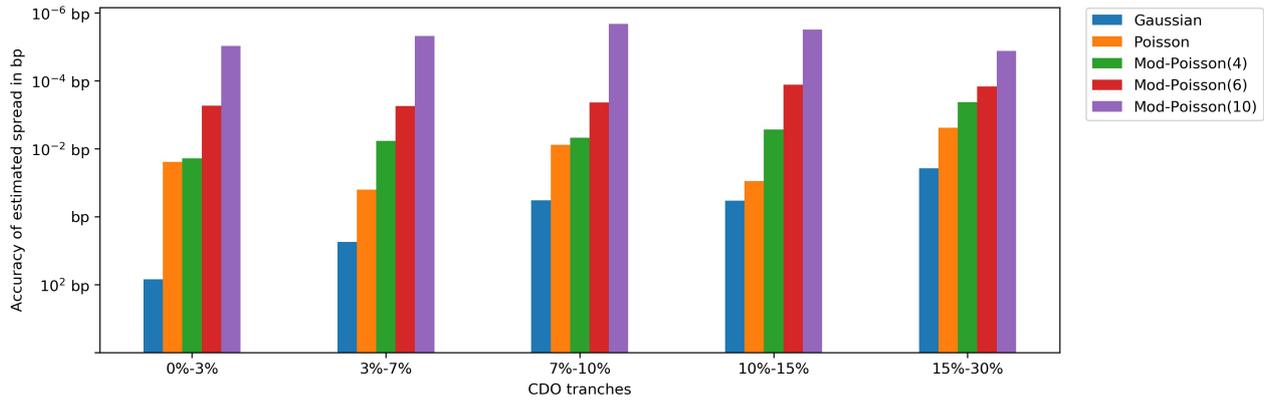


FIGURE 7. Decimal place accuracy of fair spread estimates for all estimation methods by tranche.

The accuracy of the models' estimates can be more readily compared from Figure 7, where the decimal place accuracy of the fair spread estimates in Table 2 are compared. The mod-Poisson approximation schemes yield the most accurate estimates, with exponentially improving precision as the approximation scheme order increases.

Attachment points	Benchmark (recursive)	Gaussian approximation	Poisson approximation	Mod-Poisson (order=4)	Mod-Poisson (order=6)	Mod-Poisson (order=10)
0% - 3%	Default leg	232.5975 bp	228.8759 bp	232.5996 bp	232.5979 bp	232.5975 bp
	Premium leg	452.2145 bp	451.0626 bp	452.2208 bp	452.2137 bp	452.2145 bp
	Fair spread	5143.5210 bp	5074.1488 bp	5143.4961 bp	5143.5404 bp	5143.5204 bp
3% - 7%	Default leg	200.2722 bp	200.7338 bp	200.2540 bp	200.2716 bp	200.2723 bp
	Premium leg	1364.6971 bp	1362.7014 bp	1364.7217 bp	1364.6987 bp	1364.6971 bp
	Fair spread	1467.5213 bp	1473.0575 bp	1467.3613 bp	1467.5153 bp	1467.5218 bp
7% - 10%	Default leg	62.8105 bp	62.7749 bp	62.8088 bp	62.8099 bp	62.8104 bp
	Premium leg	1248.7606 bp	1248.8878 bp	1248.7468 bp	1248.7608 bp	1248.7606 bp
	Fair spread	502.9824 bp	502.6464 bp	502.9747 bp	502.9777 bp	502.9820 bp
10% - 15%	Default leg	33.6304 bp	33.5575 bp	33.6500 bp	33.6310 bp	33.6304 bp
	Premium leg	2204.4540 bp	2204.5755 bp	2204.4246 bp	2204.4529 bp	2204.4540 bp
	Fair spread	152.5566 bp	152.2176 bp	152.6473 bp	152.5594 bp	152.5565 bp
15% - 30%	Default leg	7.2444 bp	7.2698 bp	7.2461 bp	7.2447 bp	7.2445 bp
	Premium leg	6738.6074 bp	6738.5758 bp	6738.6165 bp	6738.6076 bp	6738.6073 bp
	Fair spread	10.7506 bp	10.7883 bp	10.7531 bp	10.7511 bp	10.7508 bp

TABLE 2. Default leg, premium leg and fair spread for five tranches computed using different techniques. Benchmark values are exact and computed using the recursive methodology.

## 8. CONCLUSIONS

In this paper we have introduced mod-Poisson approximation schemes for the semi-analytical estimation of functionals of factor credit portfolio models. This technique is based on the theory of mod- $\phi$  convergence and mod- $\phi$  approximation schemes and relies on the computation of higher-order correction terms to the classic Poisson approximation. We also show how to extend the method to credit models with stochastic exposures using mod-compound Poisson approximation schemes.

The method has been compared empirically with the recursive method, the large deviations approximation, the Chen–Stein method and the Monte Carlo simulation technique (with and without importance sampling). The tests show that mod-Poisson approximation schemes lead to more accurate estimates for risk measures (such as VaR and ES) and CDO tranche prices. Furthermore, due to the semi-analytical nature of the approximations, they require substantially less computational time, especially in the large portfolio limit.

## APPENDIX A. OVERVIEW OF ESTIMATION METHODS

**A.1. Recursive methodology.** The recursive methodology was first introduced in [BH84] (but see [KZ03] for a more concise introduction) in the context of reliability theory, where the main quantity of interest is the failure probability of a system constituted by a large number of independent sub-components. The same estimation technique was rediscovered in the context of credit portfolios in [HW04, Bra04] and is nowadays well-known among financial practitioners. In its full generality, this method allows the exact computation of the law of  $L_n = \sum_{i=1}^n Y_i$ , where the random variables  $(Y_i)_{i=1}^n$  are assumed to be integer-valued and independent, but not necessarily identically distributed. Under the conditional independence assumption typical of credit risk models, this method can be used to compute the conditional distribution of  $L_n$  given  $\Psi$ , i.e.  $\mathbb{P}(L_n = k | \Psi)$  for all  $k \in \mathbb{N}$  and all  $\psi \in \text{Range}(\Psi)$ . The unconditional distribution of  $L_n$  can then be obtained via numerical integration over the mixing variable  $\Psi$ .

Before presenting the recursive methodology algorithm in its full generality, it is instructive to consider the simpler case of the Poisson binomial distribution. Let  $L_n = \sum_{i=1}^n Y_i$ , where  $(Y_i)_{i=1}^n$  are independent Bernoulli random variables, with  $\mathbb{P}(Y_i = 1) = p_i$ . The basic idea is to compute the distribution of the sum  $L_n$  recursively by adding one Bernoulli random variable at a time.

**Data:** distributions of the  $Y_i$  as  $q_{i,j} = \mathbb{P}(Y_i = j)$ , for  $j \in \text{Range}(Y_i)$ .

**Result:** distribution  $p_k = \mathbb{P}(L_n = k)$ , for  $k \in \text{Range}(L_n)$ .

```

 $p_0^{(0)} \leftarrow 1$ 
 $p_k^{(0)} \leftarrow 0$ , for  $k \in \text{Range}(L_n) \setminus \{0\}$ 
for  $i = 1$  to  $n$  do
   $p_k^{(i)} \leftarrow p_k^{(i-1)}$ , for  $k \in \text{Range}(L_n)$ 
  for  $k \in \text{Range}(L_n)$  do
    for  $j \in \text{Range}(Y_i)$  do
       $\ell \leftarrow k + j$ 
       $p_k^{(i)} - = p_k^{(i-1)} q_{i,j}$ 
       $p_\ell^{(i)} + = p_k^{(i-1)} q_{i,j}$ 
    end
  end
end

```

**return**  $p_k \leftarrow p_k^{(n)}$ , for  $k \in \text{Range}(L_n)$ .

**Algorithm 1:** Recursive methodology for loss distribution under assumption of independence.

If  $L_n$  is the sum of the first term only, i.e.  $L_n = Y_1$ , then its distribution is simply the same as that of  $Y_1$ . We can denote it as follows:

$$p_k^{(1)} = \mathbb{P}(L_n = k) = \begin{cases} 1 - p_1 & \text{if } k = 0, \\ p_1 & \text{if } k = 1 \end{cases}$$

where the superscript (1) indicates that this is the distribution of  $L_n$  as a sum of only one term. If  $L_n = L_1 + L_2$ , then clearly we have that:

$$p_k^{(2)} = \mathbb{P}(L_n = k) = \begin{cases} (1 - p_1)(1 - p_2) & \text{if } k = 0, \\ p_1(1 - p_2) + p_2(1 - p_1) & \text{if } k = 1, \\ p_1 p_2 & \text{if } k = 2, \end{cases}$$

which can also be expressed in terms of  $p_k^{(1)}$  as follows:

$$p_k^{(2)} = p_k^{(1)}(1 - p_2) + p_{k-1}^{(1)} p_2, \quad k = 0, 1, 2,$$

provided we set  $p_{-1}^{(1)} := 0$ . This shows that adding one more summand to the sum is equivalent to shifting a probability mass equal to  $p_k^{(1)} p_2$  from the point  $k$  to the point  $k + 1$ , for all possible values of  $k$ .

This recursion can be generalized, for any number of summands:

$$p_k^{(m)} = p_k^{(m-1)}(1 - p_m) + p_{k-1}^{(m-1)} p_m, \quad k = 0, 1, \dots, m.$$

Running the recursion from  $m = 1$  to  $m = n$ , one finally recovers the distribution of  $L_n$ . In the general case with  $L_n = \sum_{i=1}^n Y_i$ , where the  $Y_i$  are independent, integer-valued random variables not necessarily Bernoulli distributed, at the  $m$ -th step of the recursion we must shift a probability mass equal to  $p_k^{(m-1)} \mathbb{P}(Y_m = j)$  from the point  $k$  to the point  $\ell := k + j$ , for every possible value  $j$  of  $Y_m$  and for every possible value  $k$ . Algorithm 1 shows the recursion in this general case.

The complexity of the algorithm is

$$O\left(n \sum_{i=1}^{|\text{Range}(L_n)|} |\text{Range}(Y_i)|\right).$$

In the Poisson binomial case the complexity is  $O(n^2)$ , which makes the recursive methodology computationally expensive in the case of large portfolios, as shown in Section 6.3. We further remark that the recursive method always outputs the full loss distribution, which is computationally wasteful in the case of applications for which only a part of the distribution is needed, for instance its tail or a particular tail value.

**A.2. Large deviations approximation.** The large deviations theory provides tools for the analytical approximation of probabilities of rare events. Most results can be derived using a variety of techniques, but since the '80s a unified approach to the field has emerged, as expounded in several monographs on the subject [DS01, Ell06, DZ10, DH08]. The theory has been applied to many financial problems, including credit risk management (see [Pha07] for an overview). Here we follow the results presented in [DDD04], which can be used to estimate conditional probabilities of the form  $\mathbb{P}(L_n > nx | \Psi)$  for a given tail value  $x \in \mathbb{R}^+$  and a generic credit portfolio  $L_n = \sum_{i=1}^n Z_i Y_i$ .

The key quantity, as in all large deviations applications, is the cumulant generating function of the random variable of interest. In this case we want to approximate the conditional distribution of  $L_n$  given  $\Psi$ , so we denote its conditional cumulant generating function by  $F(\cdot, \Psi)$  and compute it as:

$$F(\lambda, \Psi) = \frac{1}{n} \sum_{i=1}^n \log(1 + p_i(\Psi)(\mathbb{E}[e^{\lambda Z_i} | \Psi] - 1)), \quad (18)$$

where  $\mathbb{P}(Y_i = 1 | \Psi) = p_i(\Psi)$  are the default probabilities as a function of the mixing variable and  $\mathbb{E}[e^{\lambda Z_i} | \Psi]$  is the conditional moment generating function of  $Z_i$ . The latter quantity depends on the particular distributional assumptions of the credit risk model and should be known explicitly or be easily computable.

We remark that the normalization term  $1/n$  in Equation (18) is necessary for adapting the proof of Bahadur–Rao's theorem (see [DZ10, Theorem 3.7.4.]) to the case of still independent, but not identically distributed summands. More specifically,  $F(\lambda, \Psi)$  can be thought of as the cumulant generating function of a mixture of the summands' distributions, each taken with weight  $1/n$ .

The second key quantity is the Legendre–Fenchel transform of  $F(\lambda, \psi)$  for a given tail value  $x$ , which is defined as:

$$\lambda_x(\Psi) := \sup_{\lambda \geq 0} \{\lambda x - F(\lambda, \Psi)\}.$$

Even in the case of the simplest credit models the value  $\lambda_x(\psi)$  cannot be computed analytically, but fortunately the corresponding optimization problem can be solved efficiently, since the objective function is convex. Finally,  $F(\cdot, \Psi)$ , its second derivative  $F''(\cdot, \Psi)$  and  $\lambda_x(\Psi)$  are used in the

computation of the following large deviations tail probability estimator:

$$\mathbb{P}(L_n > nx | \Psi) \approx \frac{1}{\sqrt{2\pi n \lambda_x(\Psi)^2 F''(\lambda_x(\Psi), \Psi)}} e^{-n(\lambda_x(\Psi)x - F(\lambda_x(\Psi), \Psi))} \quad (19)$$

The unconditional tail probability can then be obtained numerically by integrating over the mixing variable  $\Psi$ . This latter integration actually suffers from numerical instabilities due to the presence of a vanishing denominator in Equation (19) for some values of  $\Psi$ , as discussed in Section 6. We remark that the large deviations approximation is optimal only asymptotically in  $n$ , therefore we can expect it to perform well in the limit of very large portfolios.

**A.3. Stein's method: first-order correction to the Gaussian approximation.** In this section we present a method introduced by El Karoui and Jiao in [EKJ09] and applied to CDO pricing in [EKJK08]. Their results rely on Stein's method [Ste72] and on the zero-bias transformation framework [GR97] developed by Goldstein and Reinert. The results we are interested in rely on the following lemma.

**Lemma A.1.** *Let  $X_1, \dots, X_n \in L^4$  be independent mean-zero random variables and let  $W = X_1 + \dots + X_n$ , with  $\sigma_W^2 = \text{Var}(W)$ . Then for any function  $h$  such that  $\|h''\|_\infty$  is finite, the following approximation holds:*

$$\mathbb{E}[h(W)] \approx \mathbb{E}[h(Z)] + \frac{\sum_{i=1}^n \mathbb{E}[(X_i)^3]}{2\sigma_W^4} \mathbb{E}\left[\left(\frac{Z^2}{3\sigma_W^2} - 1\right) Zh(Z)\right],$$

where  $Z \sim \mathcal{N}(0, \sigma_W^2)$ .

This lemma provides a first-order correction to the classical approximation of  $W$  in terms of the Gaussian random variable  $Z$ . Its proof, together with an explicit bound on the approximation error, is beyond the scope of this work and can be found in [EKJK08, Theorem 2.1]. Unfortunately Lemma A.1 cannot be used directly for the estimation of tail probabilities or call prices, because the regularity assumption on the function  $h$  (namely, that its second derivative exists everywhere) is not satisfied neither for indicator functions,  $h(x) = \mathbb{1}_{\{x > k\}}$ , nor for call functions,  $h(x) = (x - k)^+$ . Nevertheless the authors are able to show the result still holds for these functional choices: see [EKJ09, Propositions 3.5 and 3.6].

If  $L_n = \sum_{i=1}^n Z_i Y_i$  is a generic credit portfolio with mixing variable  $\Psi$ , then we can approximate the expectation of a call function of  $L_n$  with strike  $K$  using Lemma A.1 applied to the centered random variables  $X_i := Z_i Y_i - \mathbb{E}[Z_i Y_i]$ , obtaining:

$$\mathbb{E}[(L_n - K)^+ | \Psi] \approx \mathbb{E}[(Z - K)^+ | \Psi] + \frac{1}{6\sigma_W^2} \sum_{i=1}^n \mathbb{E}[X_i^3] K \phi_{\sigma_W}(K),$$

where  $\sigma_W^2 = \sum_{i=1}^n \text{Var}(Z_i Y_i | \Psi)$ ,  $Z \sim \mathcal{N}(0, \sigma_W^2)$ ,  $\phi_\sigma$  denotes the Gaussian density with variance  $\sigma^2$ , and the first-order correction term has been computed explicitly using Bachelier's formula. The approximation for the unconditional expectation can then be obtained by numerical integration over the distribution of the mixing variable  $\Psi$ .

**A.4. Chen-Stein's method: first-order correction to the Poisson approximation.** This method was also introduced in [EKJ09, EKJK08], together with the first-order Gaussian correction seen in the previous section. In fact the two methods share the same basic techniques, just different reference laws. When the reference law is the Poisson law, Stein's method can still be applied but for a different choice of operator, as first noticed by Chen in [Che75]. The approximation we are interested in relies on the following lemma.

**Lemma A.2.** *Let  $X_1, \dots, X_n \in L^3$  be non-negative, integer-valued random variables and let  $W = X_1 + \dots + X_n$ , with  $\lambda_W = \mathbb{E}[W]$  and  $\sigma_W^2 = \text{Var}(W)$ . Then for any bounded function  $h$ , the following*

approximation holds:

$$\mathbb{E}[h(W)] \approx \mathbb{E}[h(Z)] + \frac{\sigma_W^2 - \lambda_W}{2} \mathbb{E}[\Delta_+^2(h)(Z)],$$

where  $Z \sim \text{Po}(\lambda_W)$  and  $\Delta_+^2(h)(x) = h(x+2) - 2h(x+1) + h(x)$  is the second-order forward finite difference operator.

Also in this case, it is clear that the lemma provides a first-order correction to the classical Poisson approximation of  $W$  in terms of  $Z$ . Its proof, together with an explicit bound on the approximation error, can be found in [EKJ09, Theorem 4.2].

In a credit risk setting we can apply this result to the case of a credit portfolio with unit exposures  $L_n = \sum_{i=1}^n Y_i$  and mixing variable  $\Psi$ . For instance, we can approximate the expectation of a call function of  $L_n$  with strike  $K$  using Lemma A.2 applied to the default indicators  $Y_i \sim \text{Be}(p_i(\Psi))$ , which yields:

$$\mathbb{E}[(L_n - K)^+ | \Psi] \approx \mathbb{E}[(Z - K)^+ | \Psi] - \frac{1}{2} \sum_{i=1}^n p_i(\Psi) e^{-\lambda} \frac{\lambda^{K-1}}{(K-1)!},$$

where  $\lambda = \sum_{i=1}^n p_i(\Psi)$  and  $Z \sim \text{Po}(\lambda)$ . The method can trivially be extended to the case of homogeneous deterministic recovery rates, that is portfolios of the form  $L_n = \sum_{i=1}^n Z_i Y_i$ , with  $Z_i = 1 - R$ , for some  $R \in [0, 1]$ , but the authors in [EKJK08, EKJ09] are unable to provide an extension to the case of stochastic, heterogeneous recovery rates. In Section 5 we show how higher order correction terms to the Poisson approximation can be derived in this setting using mod-compound Poisson approximation schemes. Finally, let us remark that this approximation method corresponds to the mod-Poisson approximation scheme of order  $r = 2$ .

**A.5. Monte Carlo simulation.** Simulation-based methods are very popular in financial applications and standard monographs in the field are [AG07] and [Gla04]. The simplest simulation-based method is Monte Carlo integration. Algorithm 2 presents the naive Monte Carlo estimator for the tail probabilities of a generic credit risk model  $L_n = \sum_{i=1}^n L_i$ .

The simulation is divided in two steps: first we simulate a realization of the mixing variable  $\Psi$  and then, conditionally on this value, we simulate the portfolio losses. The process is then repeated for a sufficiently high number of simulation runs.

**Data:** number of Monte Carlo simulations  $M$ .

**Result:** tail probability estimator  $\theta_x \approx \mathbb{P}(L_n > x)$ , for some  $x \in \mathbb{R}^+$ .

**for**  $m = 1$  **to**  $M$  **do**

    Sample  $\psi_m$  from the distribution of  $\Psi$

**for**  $i = 1$  **to**  $n$  **do**

        Sample  $L_i^{(m)}$  given  $\Psi = \psi_m$

**end**

**end**

$$\theta_x \leftarrow \frac{1}{M} \sum_{m=1}^M \mathbb{1}_{\{\sum_{i=1}^n L_i^{(m)} > x\}}$$

**return**  $\theta_x$ .

**Algorithm 2:** Monte Carlo estimator for probability tail function.

One key advantage of this naive Monte Carlo methodology is its generality: it's straightforward to introduce new sources of randomness and additional model parameters (provided, of course, the final losses can still be simulated efficiently), so that even very complicated models can be estimated.

Additionally, it is possible to estimate the full loss distribution with only one batch of simulations, by running Algorithm 2 for different values of  $x$  on the same simulated sample.

On the other hand, the naive Monte Carlo methodology requires at least  $10^n$  simulations for the estimation of probabilities of order  $10^{-n}$ . This is a simple consequence of the fact that rare events occur rarely also in simulations, thus leading to estimators that underestimate the probability of rare events. In practice, even more simulations are needed in order to obtain an accurate estimate within any reasonable asymptotic confidence interval: see Figure 3 for an empirical illustration. This problem is particularly severe in risk management applications, since large losses tend to occur with low probability but contribute substantially to the risk of a position.

**A.6. Importance sampling.** In order to overcome the limitations of the naive Monte Carlo methodology for the estimation of rare events, importance sampling techniques have been developed. Our presentation follows [GL05], where an importance sampling algorithm for the Gaussian copula model was first introduced, but good treatments can also be found in [Gla04, MFE15]. Additionally we show explicitly how to incorporate stochastic exposures. To illustrate the approach, let us first consider a simple Poisson binomial model of the form  $L_n = \sum_{i=1}^n Y_i$  and suppose that we are interested in estimating the rare event probability  $\mathbb{P}(L_n > x)$ , for  $x$  sufficiently large. Let  $f(y_1, \dots, y_n) = \mathbb{P}(Y_1 = y_1, \dots, Y_n = y_n)$  be the probability mass function of the random vector  $(Y_1, \dots, Y_n)$  on  $\{0, 1\}^n$  and let  $g$  denote the probability mass function of another random vector of obligors' losses,  $(\tilde{Y}_1, \dots, \tilde{Y}_n)$ , yet to be determined. Then a simple calculation shows that

$$\mathbb{P}(L_n > x) = \mathbb{E} [\mathbb{1}_{\{Y_1 + \dots + Y_n > x\}}] = \mathbb{E} \left[ \mathbb{1}_{\{\tilde{Y}_1 + \dots + \tilde{Y}_n > x\}} \frac{f(\tilde{Y}_1, \dots, \tilde{Y}_n)}{g(\tilde{Y}_1, \dots, \tilde{Y}_n)} \right], \quad (20)$$

where the right-hand side integral is computed only with respect to the distribution of  $(\tilde{Y}_1, \dots, \tilde{Y}_n)$  and the ratio  $f(\cdot)/g(\cdot)$  is known as the *likelihood ratio function* of the two distributions.

The main idea of importance sampling is that we can choose  $(\tilde{Y}_1, \dots, \tilde{Y}_n)$  in such a way that the event  $\{\tilde{Y}_1 + \dots + \tilde{Y}_n > x\}$  is much more likely than  $\{Y_1 + \dots + Y_n > x\}$  on the support of the likelihood ratio function, so we can approximate the rare event probability on the left-hand side of (20) by estimating the integral on the right-hand side using a naive Monte Carlo estimation. The large deviations theory suggests that a natural choice for the loss distribution  $(\tilde{Y}_1, \dots, \tilde{Y}_n)$  comes from exponentially tilting the total loss  $L_n$ , which leads to the following parametrized family:

$$g_\lambda(y_1, \dots, y_n) = e^{\lambda(y_1 + \dots + y_n)} \frac{f(y_1, \dots, y_n)}{\mathbb{E} [e^{\lambda(Y_1 + \dots + Y_n)}]} \quad (21)$$

$$\begin{aligned} &= e^{\lambda(y_1 + \dots + y_n)} \frac{\prod_{i=1}^n (p_i)^{y_i} (1 - p_i)^{1 - y_i}}{\prod_{i=1}^n (1 + p_i(e^\lambda - 1))} \\ &= \prod_{i=1}^n q_i(\lambda)^{y_i} (1 - q_i(\lambda))^{1 - y_i} \end{aligned} \quad (22)$$

where we defined

$$q_i(\lambda) = \frac{p_i e^\lambda}{1 + p_i(e^\lambda - 1)}, \quad \lambda \in \mathbb{R}^+.$$

Equation (22) shows that sampling from the distribution  $g_\lambda$  is equivalent to simulating losses with new default probabilities  $q_i(\lambda)$  instead of  $p_i$ .

Ideally the parameter  $\lambda$  should be chosen by minimizing the variance of the importance sampling estimator in Equation (20) (or equivalently its second moment), because the estimator with minimum variance will require the least number of simulations for any given level of confidence. Unfortunately this minimization is in general intractable, but by expressing the likelihood ratio function from Equation (21) in terms of  $F(\lambda) := \log(\mathbb{E} [e^{\lambda(Y_1 + \dots + Y_n)}])$  (i.e. the cumulant generating function of  $L_n$ ), we obtain the following bound for the second moment of the exponentially

tilted importance sampling estimator:

$$\mathbb{E} \left[ \left( \mathbb{1}_{\{\tilde{Y}_1 + \dots + \tilde{Y}_n > x\}} \frac{f(\tilde{Y}_1, \dots, \tilde{Y}_n)}{g_\lambda(\tilde{Y}_1, \dots, \tilde{Y}_n)} \right)^2 \right] \leq e^{-2\lambda x + 2F(\lambda)}. \quad (23)$$

This bound, which is equivalent to the Chernoff bound, turns out to be remarkably sharp, so that the minimizer  $\lambda_x$  of the right-hand side of (23) can be used to obtain a very efficient importance sampling estimator. The estimation procedure we just outlined applies to a Poisson binomial distribution, but it is readily extended to any Gaussian copula model of the form  $L_n = \sum_{i=1}^n Z_i Y_i$  with mixing variable  $\Psi$ , assuming that the variables  $Z_i Y_i$  are light-tailed.

We first sample several realizations of the mixing variable  $\Psi$  and compute the exponentially tilted importance sampling estimator for each realization. The presence of the stochastic exposures,  $Z_i$ , slightly modifies the formulæ already presented through their moment generating functions,  $\mathbb{E}[e^{\lambda Z_i}]$ , which must be explicitly computable, but are otherwise easy to incorporate. A full description of this procedure can be found in Algorithm 3.

**Data:** number of Monte Carlo simulations  $M$ .

**Result:** tail probability estimator  $\theta_x \approx \mathbb{P}(L_n > x)$ , for some  $x \in \mathbb{R}^+$ .

**for**  $m = 1$  **to**  $M$  **do**

    Sample  $\psi_m$  from  $\mathcal{N}(0, I)$

    Compute  $\lambda_x(\psi_m)$

**for**  $i = 1$  **to**  $n$  **do**

$$q_i \leftarrow \frac{p_i(\psi_m) \mathbb{E}[e^{\lambda_x(\psi_m) Z_i}]}{1 + p_i(\psi_m) (\mathbb{E}[e^{\lambda_x(\psi_m) Z_i}] - 1)}$$

    Sample  $\tilde{L}_i^{(m)} = Z_i \tilde{Y}_i$ , with  $\tilde{Y}_i \sim \text{Be}(q_i)$

**end**

**end**

$$\theta_x \leftarrow \frac{1}{M} \sum_{m=1}^M \mathbb{1}_{\{\sum_{i=1}^n \tilde{L}_i^{(m)} > x\}} e^{-\lambda_x(\psi_m) \sum_{i=1}^n \tilde{L}_i^{(m)} + F(\lambda_x(\psi_m), \psi_m)}$$

**return**  $\theta_x$ .

**Algorithm 3:** One-step importance sampling algorithm for probability tail function.

Nevertheless, this importance sampling procedure yields estimators that are far from being optimal, as shown in [GL05]. The problem lies in the way the mixing variable  $\Psi$  is handled. Recall that in a Gaussian copula model large losses tend to occur for large realizations of the mixing variable  $\Psi$  (up to sign conventions), but in Algorithm 3 large realizations of  $\Psi$  will be sampled only rarely. In other words, the exponential tilting must be applied to the unconditional portfolio losses and not just to the conditional ones. This leads to a two-step exponential tilting procedure, in which the Gaussian mixing variable  $\Psi$  is tilted first – which amounts to a shift of its mean from zero to a new value  $\mu$  – while portfolio losses are tilted in a second step conditionally on each realization of the shifted mixing variable, exactly as in Algorithm 3. It should be mentioned that the first step relies heavily on the specific parametric choice of a Gaussian copula and furthermore on solving an approximate optimization, as the objective function itself needs to be approximated. Nevertheless several possible approximations are feasible and the interested reader is referred to [GL05, Section 5.1] for an overview of choices.

The final procedure is described in detail in Algorithm 4 and this is also the procedure used for the empirical tests of Section 6. Finally, we remark that if the full tail function of the loss distribution needs to be estimated, then it is not necessary to re-compute the shifted mean  $\mu$  of  $\Psi$  for each value

of  $x$ , since in practice the same shift yields efficient estimators for a large neighborhood of tail points.

**Data:** number of Monte Carlo simulations  $M$ .

**Result:** tail probability estimator  $\theta_x \approx \mathbb{P}(L_n > x)$ , for some  $x \in \mathbb{R}^+$ .

$$\mu \leftarrow \sup_{z \in \mathbb{R}} \left\{ F(\lambda_x(z), z) - \lambda_x(z)x - \frac{1}{2}|z|^2 \right\}$$

**for**  $m = 1$  **to**  $M$  **do**

    Sample  $\psi_m$  from  $\mathcal{N}(\mu, I)$

    Compute  $\lambda_x(\psi_m)$

**for**  $i = 1$  **to**  $n$  **do**

$$q_i \leftarrow \frac{p_i(\psi_m) \mathbb{E}[e^{\lambda_x(\psi_m) Z_i}]}{1 + p_i(\psi_m) (\mathbb{E}[e^{\lambda_x(\psi_m) Z_i}] - 1)}$$

    Sample  $\tilde{L}_i^{(m)} = Z_i \tilde{Y}_i$ , with  $\tilde{Y}_i \sim \text{Be}(q_i)$

**end**

**end**

$$\theta_x \leftarrow \frac{1}{M} \sum_{m=1}^M \mathbb{1}_{\{\sum_{i=1}^n \tilde{L}_i^{(m)} > x\}} e^{-\lambda_x(\psi_m) \sum_{i=1}^n \tilde{L}_i^{(m)} + F(\lambda_x(\psi_m), \psi_m) + \frac{1}{2}|\mu|^2 - \mu^T \psi_m}$$

**return**  $\theta_x$ .

**Algorithm 4:** Two-step importance sampling algorithm for probability tail function.

#### APPENDIX B. TAIL OF THE POISSON DISTRIBUTION

**Proposition B.1** (Tail function of the Poisson distribution). *If  $X$  is a  $\text{Po}(\lambda)$  random variable, then:*

$$\mathbb{P}(X > k) = \frac{1}{k!} \gamma(k+1, \lambda), \quad \forall k \in \mathbb{N}$$

where  $\gamma$  is the lower incomplete gamma function given by:

$$\gamma(x, \lambda) = \int_0^\lambda t^{x-1} e^{-t} dt.$$

*Proof.* For a Poisson distribution, the derivative of the tail probability with respect to the distribution parameter  $\lambda$  is given by the probability mass function. This can be shown as follows:

$$\begin{aligned} \frac{d}{d\lambda} \mathbb{P}(X > k) &= \sum_{j=k+1}^{\infty} \frac{1}{j!} \frac{d}{d\lambda} (e^{-\lambda} \lambda^j) \\ &= \sum_{j=k+1}^{\infty} e^{-\lambda} \frac{\lambda^{j-1}}{(j-1)!} - \sum_{j=k+1}^{\infty} e^{-\lambda} \frac{\lambda^j}{(j)!} \\ &= \sum_{j=k}^{\infty} e^{-\lambda} \frac{\lambda^j}{j!} - \sum_{j=k+1}^{\infty} e^{-\lambda} \frac{\lambda^j}{(j)!} = e^{-\lambda} \frac{\lambda^k}{k!}. \end{aligned}$$

Then, by the fundamental theorem of calculus, one has:

$$\mathbb{P}(X > k) = \int_0^\lambda e^{-t} \frac{t^k}{k!} dt = \frac{1}{k!} \gamma(k+1, \lambda).$$

□

## APPENDIX C. INCIDENCE ALGEBRAS AND THE MÖBIUS FUNCTION

**Definition C.1.** A poset (or partially ordered set)  $P$  is a set together with a binary order relation, denoted  $\leq$ , satisfying the following axioms:

- (1)  $x \leq x, \quad \forall x \in P,$
- (2) if  $x \leq y$  and  $y \leq x$ , then  $x = y$ ,
- (3) if  $x \leq y$  and  $y \leq z$ , then  $x \leq z$ .

We say that  $P$  has a minimal element, denoted  $\hat{0}$ , if there exists an element  $\hat{0} \in P$  such that  $\hat{0} \leq x$  for all  $x \in P$ . Analogously, has a maximal element  $\hat{1}$ , if there exists an element  $\hat{1} \in P$  such that  $\hat{1} \geq x$  for all  $x \in P$ .

*Example C.2* (The poset of set partitions  $\Pi(n)$ ). Let  $A$  be a finite set. A set partition  $\pi = \{B_1, \dots, B_k\}$  of  $A$  is a collection of non-empty, mutually disjoint subsets of  $A$ , such that  $\cup_{i=1}^k B_i = A$ . The sets  $B_1, \dots, B_k$  are called the blocks of  $\pi$  and the number of blocks of  $\pi$  is denoted by  $|\pi|$ .

For  $n \in \mathbb{N}$ , define  $\Pi(n)$  as the set of all set partitions of  $\{1, 2, \dots, n\}$ . Given two set partitions  $\pi$  and  $\sigma$ , we denote  $\pi \leq \sigma$  if every block of  $\pi$  is contained in a block of  $\sigma$ . Then, the poset  $(\Pi(n), \leq)$  admits a minimal element

$$\hat{0}_n = \{\{k\}, k = 1, \dots, n\} \quad (\text{the partition with } n \text{ blocks})$$

and a maximal element

$$\hat{1}_n = \{\{1, \dots, n\}\} \quad (\text{the partition with only one block}).$$

An interval of a poset  $P$ , denoted  $[x, y]$  for some  $x, y \in P$  with  $x \leq y$ , is a subset of  $P$  defined as  $[x, y] = \{z \in P \mid x \leq z \leq y\}$ . We denote by  $\text{Int}(P)$  the set of all intervals of  $P$  and we say that  $P$  is locally finite if every interval of  $P$  is finite. It is quite natural to define functions on intervals, for instance if we want to count the number of elements of an interval, and more generally functions  $f : \text{Int}(P) \rightarrow \mathbb{K}$  for some field  $\mathbb{K}$ . The space of all such functions can be turned into an associative algebra, as the following definition shows.

**Definition C.3** (Incidence algebra  $I(P, \mathbb{K})$ ). The incidence algebra  $I(P, \mathbb{K})$  of  $P$  over  $\mathbb{K}$  is the  $\mathbb{K}$ -algebra of all functions  $f : \text{Int}(P) \rightarrow \mathbb{K}$  with operation, called convolution, given by:

$$(f \star g)(x, y) := \sum_{x \leq z \leq y} f(x, z) g(z, y).$$

The algebra  $I(P, \mathbb{K})$  is associative, and its multiplicative identity is

$$\delta(x, y) = \begin{cases} 1 & \text{if } x = y, \\ 0 & \text{otherwise.} \end{cases}$$

Another important element of the incidence algebra is the zeta function of the poset, defined as

$$\zeta(x, y) = \begin{cases} 1 & \text{if } x \leq y, \\ 0 & \text{otherwise.} \end{cases}$$

It can be shown that the function  $\zeta$  of a poset is invertible and its inverse is called the Möbius function of the poset and is denoted by  $\mu$ . In particular one has that  $\mu$  satisfies:

$$\mu \star \zeta = \zeta \star \mu = \delta.$$

Furthermore, the following important result holds.

**Theorem C.4** (Möbius inversion formula). *Let  $P$  be a locally finite poset and let  $g, h : P \rightarrow \mathbb{K}$ . Then*

$$h(x) = \sum_{y \leq x} g(y), \quad \forall x \in P$$

is equivalent to

$$g(x) = \sum_{y \leq x} h(y) \mu(y, x), \quad \forall x \in P.$$

*Example C.5.* The Möbius function of the poset of set partitions  $\Pi(n)$  admits the following representation:

$$\mu(\pi, \sigma) = (-1)^{|\pi| - |\sigma|} \prod_{B \in \sigma} (n_\pi^\sigma(B) - 1)!, \quad \forall \pi \leq \sigma$$

where  $n_\pi^\sigma(B)$  is the number of blocks of  $\pi$  contained in the block  $B$  of  $\sigma$ .

We refer to [Rot64] for details on these constructions. For the manipulation of Fourier and Laplace transforms of probability distributions, the formalism of posets and Möbius functions enables one to go from a generating series to its exponential or logarithm.

**Theorem C.6** (Exponential and logarithm of generating series). *Let  $G(z) = \sum_{n=1}^{\infty} \frac{g_n}{n!} z^n$  be the exponential generating series of a sequence of coefficients  $(g_n)_{n \geq 1}$ . If  $H(z) = \exp(G(z))$ , then*

$$H(z) = 1 + \sum_{n=1}^{\infty} \frac{1}{n!} \left( \sum_{\pi \in \Pi(n)} \prod_{B \in \pi} g_{|B|} \right) z^n.$$

Conversely, if  $H(z) = 1 + \sum_{n=1}^{\infty} \frac{h_n}{n!} z^n$  and  $G(z) = \log(H(z))$ , then

$$G(z) = \sum_{n=1}^{\infty} \frac{1}{n!} \left( \sum_{\pi \in \Pi(n)} \mu(\pi, \hat{1}_n) \prod_{B \in \pi} h_{|B|} \right) z^n,$$

with  $\mu(\pi, \hat{1}_n) = (-1)^{|\pi| - 1} (|\pi| - 1)!$ .

*Proof.* We expand the exponential of  $G(z) = \sum_{n=1}^{\infty} \frac{g_n}{n!} z^n$ , and we collect the coefficient of  $z^n$ . This is

$$[z^n] H(z) = \sum_{l=1}^n \sum_{\substack{c_1 + \dots + c_l = n \\ c_1 \geq 1, \dots, c_l \geq 1}} \frac{1}{l!} \frac{g_{c_1} \cdots g_{c_l}}{c_1! \cdots c_l!}.$$

The sum above runs over compositions of  $n$ , that is to say sequences  $(c_1, \dots, c_l)$  of positive integers with sum equal to  $n$ . By replacing these *compositions* of size  $n$  by their non-increasing reorderings called *integer partitions* of size  $n$ , we obtain a sum over a smaller set:

$$[z^n] H(z) = \sum_{l=1}^n \sum_{\substack{\lambda_1 + \dots + \lambda_l = n \\ \lambda_1 \geq \dots \geq \lambda_l \geq 1}} \frac{1}{m_1(\lambda)! \cdots m_n(\lambda)!} \frac{g_{\lambda_1} \cdots g_{\lambda_l}}{\lambda_1! \cdots \lambda_l!},$$

where  $m_i(\lambda)$  denotes the number of parts  $\lambda_{1 \leq j \leq l}$  of  $\lambda$  equal to  $i$ . Indeed, given an integer partition  $\lambda = (\lambda_1 \geq \dots \geq \lambda_l)$  with sum  $n$ , the number of compositions whose non-increasing reordering is  $\lambda$  is the multinomial coefficient  $\frac{l!}{m_1(\lambda)! \cdots m_n(\lambda)!}$ . Now, for any integer partition  $\lambda$  with size  $n$ ,

$$\frac{n!}{m_1(\lambda)! \cdots m_n(\lambda)! \lambda_1! \cdots \lambda_l!}$$

is the number of *set partitions*  $\pi$  with size  $n$  and type  $\lambda$ , that is to say that the sizes of the blocks of  $\pi$  are given by the integer partition  $\lambda$ . Therefore,

$$n! [z^n] H(z) = \sum_{\pi \in \Pi(n)} \prod_{B \in \pi} g_{|B|}.$$

This proves the first formula. In order to get the second formula, let us define two functions  $g$  and  $h$  on the poset  $\Pi(n)$  :

$$\begin{aligned} g(\pi) &= \prod_{B \in \pi} (|B|! [z^{|B|}] G(z)) = \prod_{B \in \pi} g_{|B|}; \\ h(\pi) &= \prod_{B \in \pi} (|B|! [z^{|B|}] H(z)) = \prod_{B \in \pi} h_{|B|}. \end{aligned}$$

If  $H(z) = \exp(G(z))$ , then we have shown that  $h(\widehat{1}_n) = \sum_{\pi \in \Pi(n)} g(\pi)$ , from which we deduce that  $h(\sigma) = \sum_{\pi \leq \sigma} g(\pi)$ . By Möbius inversion,  $g(\sigma) = \sum_{\pi \leq \sigma} \mu(\pi, \sigma) g(\pi)$ , so in particular,

$$g(\widehat{1}_n) = g_n = n! [z^n] G(z) = \sum_{\pi \in \Pi(n)} \mu(\pi, \widehat{1}_n) \prod_{B \in \pi} h_{|B|}.$$

Note that an alternative way to get these inversion formulæ is by means of the Faà-di-Bruno formula

$$\frac{d^n(a \circ b)}{dz^n}(z) = \sum_{\pi \in \Pi(n)} \frac{d^{|\pi|} a}{dz^{|\pi|}}(b(z)) \prod_{B \in \pi} \frac{d^{|B|} b}{dz^{|B|}}(z),$$

with  $b$  taken equal to  $G(z)$  or  $H(z)$ , and  $a = \exp$  or  $a = \log$ .  $\square$

#### APPENDIX D. RELATION BETWEEN THE COEFFICIENTS OF THE APPROXIMATION SCHEME AND THE MOMENTS OF THE TOTAL LOSS VARIABLE

In this appendix, we prove Formula (7). It is convenient to introduce the elementary symmetric functions

$$\mathbf{e}_{k,n} = \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} p_{i_1} p_{i_2} \dots p_{i_k}.$$

By using the well known relation between the coefficients of a polynomial and its roots, we get:

$$\begin{aligned} 1 + \sum_{k=1}^{\infty} \mathbf{e}_{k,n} z^k &= \prod_{i=1}^n (1 + p_i z) = \exp \left( \sum_{i=1}^n \log(1 + p_i z) \right) \\ &= \exp \left( \sum_{i=1}^n \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} (p_i z)^k \right) = \exp \left( \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} \mathbf{p}_{k,n} z^k \right). \end{aligned}$$

By using Theorem C.6, we then obtain the relation between the coefficients  $\mathbf{e}_{k,n}$  and the coefficients  $\mathbf{p}_{k,n}$ :

$$\mathbf{e}_{k,n} = \frac{1}{k!} \sum_{\pi \in \Pi(k)} \mu(\widehat{0}_k, \pi) \left( \prod_{B \in \pi} \mathbf{p}_{|B|,n} \right).$$

This is the same relation as Equation (5), except that the sum runs over all set partitions, and not only those with blocks of size larger than 2. As a consequence, the coefficients  $\mathbf{e}_{k,n}$  and  $b_{k,n}$  are related by the following inclusion-exclusion formula:

$$b_{k,n} = \sum_{l=0}^k \frac{(-1)^l}{l!} (\mathbf{e}_{1,n})^l \mathbf{e}_{k-l,n}. \quad (24)$$

Indeed, let us replace on the right-hand side each  $\mathbf{e}_{k-l,n}$  by its expansion over set partitions. We get:

$$\text{RHS} = \sum_{l=0}^k \sum_{\pi \in \Pi(k-l)} \frac{(-1)^l \mu(\widehat{0}_{k-l}, \pi)}{l! (k-l)!} (\mathbf{p}_{1,n})^l \left( \prod_{B \in \pi} \mathbf{p}_{|B|,n} \right).$$

Since there are  $\frac{k!}{l!(k-l)!}$  subsets of the integer interval  $\llbracket 1, k \rrbracket = \{1, 2, \dots, k\}$  with size  $l$ , we can rewrite the formula above as a sum over pairs  $(\sigma, S)$ , where  $\sigma \in \Pi(k)$ , and  $S$  is a subset of  $S(\sigma) = \bigsqcup_{B \in \sigma, |B|=1} B$ , which is the union of the blocks of  $\sigma$  with size 1. Thus,

$$\text{RHS} = \frac{1}{k!} \sum_{\sigma \in \Pi(k)} \sum_{S \subset S(\sigma)} (-1)^{|S|} \mu(\widehat{0}_k, \sigma) \left( \prod_{B \in \sigma} \mathfrak{p}_{|B|, n} \right).$$

Given a set partition  $\sigma$ , the alternate sum  $\sum_{S \subset S(\sigma)} (-1)^{|S|}$  vanishes unless  $S(\sigma) = \emptyset$ , so we conclude that

$$\text{RHS} = \frac{1}{k!} \sum_{\substack{\sigma \in \Pi(k) \\ \forall B \in \sigma, |B| \geq 2}} \mu(\widehat{0}^k, \sigma) \left( \prod_{B \in \sigma} \mathfrak{p}_{|B|, n} \right) = b_{k, n}.$$

Now, the coefficients  $\mathfrak{e}_{k, n}$  also appear in the computations of the moments of  $L_n = \sum_{i=1}^n Y_i$  with  $Y_i \sim \text{Be}(p_i)$ :

$$\begin{aligned} \mathbb{E}[(L_n)^r] &= \sum_{1 \leq i_1, i_2, \dots, i_r \leq n} \mathbb{E}[Y_{i_1} Y_{i_2} \cdots Y_{i_r}] = \sum_{f: \llbracket 1, r \rrbracket \rightarrow \llbracket 1, n \rrbracket} \left( \prod_{i \in f(\llbracket 1, r \rrbracket)} p_i \right) \\ &= \sum_{s=1}^r \sum_{1 \leq i_1 < i_2 < \dots < i_s \leq n} s! \left\{ \begin{matrix} r \\ s \end{matrix} \right\} p_{i_1} p_{i_2} \cdots p_{i_s} = \sum_{s=1}^r s! \left\{ \begin{matrix} r \\ s \end{matrix} \right\} \mathfrak{e}_{s, n}, \end{aligned}$$

where  $\left\{ \begin{matrix} r \\ s \end{matrix} \right\}$  is the Stirling number of the second kind, which counts set partitions of  $\llbracket 1, r \rrbracket$  in  $s$  parts. Indeed, to go from the first line to the second line, we gather the functions  $f: \llbracket 1, r \rrbracket \rightarrow \llbracket 1, n \rrbracket$  according to their range  $f(\llbracket 1, r \rrbracket) = \{i_1 < i_2 < \dots < i_s\}$ ; if this range is fixed, then there are  $s! \left\{ \begin{matrix} r \\ s \end{matrix} \right\}$  functions with this range. Set  $M_{r, n} = \mathbb{E}[(L_n)^r]$ . The relation above can be inverted by introducing the Stirling number of the first kind  $\left[ \begin{matrix} r \\ s \end{matrix} \right]$ , which counts permutations of size  $r$  with  $s$  disjoint cycles. Hence,

$$\mathfrak{e}_{r, n} = \frac{1}{r!} \sum_{s=1}^r (-1)^{r-s} \left[ \begin{matrix} r \\ s \end{matrix} \right] M_{s, n}; \quad (25)$$

see [Sta97, Sections 1.3 and 1.4] for the combinatorial properties of the two kinds of Stirling numbers. Injecting Equation (25) into Formula (24), we get Equation (7).

## REFERENCES

- [ADEH99] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath, *Coherent measures of risk*, *Mathematical Finance* **9** (1999), no. 3, 203–228.
- [AG07] Søren Asmussen and Peter W. Glynn, *Stochastic Simulation: Algorithms and Analysis*, vol. 57, Springer, 2007.
- [BH84] Richard E. Barlow and Klaus D. Heidtmann, *Computing  $k$ -out-of- $n$  system reliability*, *IEEE Transactions on Reliability* **33** (1984), no. 4, 322–323.
- [BKN14] Andrew D. Barbour, Emmanuel Kowalski, and Ashkan Nikeghbali, *Mod-discrete expansions*, *Probability Theory and Related Fields* **158** (2014), no. 3, 859–893.
- [Bra04] Hans-Juergen Brasch, *A note on efficient pricing and risk calculation of credit basket products*, Tech. report, Working paper, TD Securities. [http://www.defaultrisk.com/pp\\_crdrv\\_54.htm](http://www.defaultrisk.com/pp_crdrv_54.htm), 2004.
- [CDMN20] Reda Chhaibi, Freddy Delbaen, Pierre-Loïc Méliot, and Ashkan Nikeghbali, *Mod- $\phi$  convergence: Approximation of discrete measures and harmonic analysis on the torus*, *Annales de l’Institut Fourier* **70** (2020), no. 3, 1115–1197.
- [Che75] Louis H. Y. Chen, *Poisson approximation for dependent trials*, *The Annals of Probability* **3** (1975), no. 3, 534–545.
- [DDD04] Amir Dembo, Jean-Dominique Deuschel, and Darrell Duffie, *Large portfolio losses*, *Finance and Stochastics* **8** (2004), no. 1, 3–16.
- [DH08] Frank Den Hollander, *Large Deviations*, vol. 14, American Mathematical Soc., 2008.
- [DS01] Jean-Dominique Deuschel and Daniel W. Stroock, *Large Deviations*, vol. 342, American Mathematical Soc., 2001.
- [Dur10] Rick Durrett, *Probability: Theory and Examples*, Cambridge University Press, 2010.
- [DZ10] Amir Dembo and Ofer Zeitouni, *Large Deviations Techniques and Applications*, Springer, 2010.
- [EKJ09] Nicole El Karoui and Ying Jiao, *Stein’s method and zero bias transformation for cdo tranche pricing*, *Finance and Stochastics* **13** (2009), no. 2, 151–180.
- [EKJK08] Nicole El Karoui, Ying Jiao, and David Kurtz, *Gauss and Poisson approximation: applications to CDO tranches pricing*, *Journal of Computational Finance* **12** (2008), no. 2, 31–58.
- [Ell06] Richard S. Ellis, *Entropy, Large Deviations, and Statistical Mechanics*, Springer, 2006.
- [FMN16] Valentin Féray, Pierre-Loïc Méliot, and Ashkan Nikeghbali, *Mod- $\varphi$  convergence: Normality zones and precise deviations*, Springer, 2016.
- [GL05] Paul Glasserman and Jingyi Li, *Importance sampling for portfolio credit risk*, *Management Science* **51** (2005), no. 11, 1643–1656.
- [Gla04] Paul Glasserman, *Monte Carlo Methods in Financial Engineering*, vol. 53, Springer, 2004.
- [GR97] Larry Goldstein and Gesine Reinert, *Stein’s method and the zero bias transformation with application to simple random sampling*, *The Annals of Applied Probability* **7** (1997), no. 4, 935–952.
- [HW04] John C. Hull and Alan D. White, *Valuation of a CDO and an  $n$ -th to default CDS without Monte Carlo simulation*, *The Journal of Derivatives* **12** (2004), no. 2, 8–23.
- [Kal02] Olav Kallenberg, *Foundations of Modern Probability*, 2nd ed., Springer, 2002.
- [KO09] Jerome Kelleher and Barry O’Sullivan, *Generating all partitions: a comparison of two encodings*, <https://arxiv.org/abs/0909.2331>, 2009.
- [KZ03] Way Kuo and Ming J. Zuo, *Optimal Reliability Modeling: Principles and Applications*, John Wiley & Sons, 2003.
- [Mac95] I. G. Macdonald, *Symmetric Functions and Hall Polynomials*, Oxford Mathematical Monographs, Clarendon Press, 1995.
- [MFE15] Alexander J. McNeil, Rüdiger Frey, and Paul Embrechts, *Quantitative Risk Management: Concepts, Techniques and Tools*, revised ed., Princeton University Press, 2015.
- [MNV22] Pierre-Loïc Méliot, Ashkan Nikeghbali, and Gabriele Visentin, *Mod-poisson approximation schemes and higher-order Chen–Stein inequalities*, In preparation, 2022.
- [Pha07] Huyên Pham, *Some applications and methods of large deviations in finance and insurance*, Paris-Princeton Lectures on Mathematical Finance 2004, Springer, 2007, pp. 191–244.
- [Rot64] Gian-Carlo Rota, *On the foundations of combinatorial theory I: Theory of Möbius functions*, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **2** (1964), no. 4, 340–368.
- [Rui96] Sebastián Martín Ruiz, *An algebraic identity leading to Wilson’s theorem*, *The Mathematical Gazette* **80** (1996), no. 489, 579–582.
- [Sta97] R. P. Stanley, *Enumerative Combinatorics. Volume 1*, Cambridge Studies in Advanced Mathematics, vol. 49, Cambridge University Press, 1997.
- [Ste72] Charles Stein, *A bound for the error in the normal approximation to the distribution of a sum of dependent random variables*, Proceedings of the sixth Berkeley symposium on mathematical statistics and probability, volume 2: Probability theory, vol. 6, University of California Press, 1972, pp. 583–603.

- [SVH03] Fred W. Steutel and Klaas Van Harn, *Infinite Divisibility of Probability Distributions on the Real Line*, CRC Press, 2003.
- [Tem94] N. M. Temme, *A set of algorithms for the incomplete gamma functions*, Probability in the Engineering and Informational Sciences **8** (1994), no. 2, 291–307.

INSTITUT DE MATHÉMATIQUES D'ORSAY, UNIVERSITÉ PARIS-SACLAY, FRANCE  
*Email address:* pierre-loic.meliot@universite-paris-saclay.fr

INSTITUTE OF MATHEMATICS, UNIVERSITÄT ZÜRICH, SWITZERLAND  
*Email address:* ashkan.nikeghbali@math.uzh.ch

DEPARTMENT OF MATHEMATICS, RISKLAB, ETH ZÜRICH, SWITZERLAND  
*Email address:* gabriele.visentin@math.ethz.ch