Affine Point Processes: Refinements to Large-Time Asymptotics

Xuefeng Gao1, Lingjiong Zhu2

March 18, 2019

Abstract

Affine point processes are a class of simple point processes with self- and mutually-exciting properties, and they have found applications in many areas. In this paper, we obtain large-time asymptotic expansions in large deviations and refined central limit theorem for affine point processes, using the framework of mod-φ convergence. Our results extend the large-time limit theorems in [Zhang et al. 2015. Math. Oper. Res. 40(4), 797-819]. The resulting explicit approximations for large deviation probabilities and tail expectations can be used as an alternative to importance sampling Monte Carlo simulations. We carry out numerical studies to illustrate our results, and present applications in credit risk analysis.

1 Introduction

Affine point processes, as described in [16], are versatile Markovian models used to capture the “clustering” feature of event arrivals. The intensity of an affine point process is an affine function of an affine jump diffusion in the sense of [14]. The Poisson process and Markovian Hawkes processes ([24]) are special cases of affine point processes. This family of point processes is tractable as their characteristic functions have an exponential-affine form, i.e., they are exponentials of affine functions of the state vector, and the coefficients of these affine functions generally solve certain ordinary differential equations (ODEs); see e.g. [14 16]. The components of affine point processes are self- and mutually-exciting, hence they (and the special case Markovian Hawkes processes) have become popular models in applications where occurrences of events exhibit clustering. Examples include finance and economics ([1 6 16 25]), social networks ([18 34]), queueing ([10 19 30]) and many others.

In a recent work, Zhang et al. [36] analyzed the large-time asymptotics of affine point processes. Specifically, they consider a multi-dimensional affine point process \((L_1, \ldots, L_n)\), and they establish a central limit theorem and a large deviation principle for \(V(t) := \sum_{i=1}^n L_i(t)\) as \(t \to \infty\). The large deviations result in [36] is of logarithmic asymptotics type, i.e., they obtained the limit \(\mathcal{I}(R) := -\lim_{t \to \infty} \frac{1}{t} \log P(V(t) \geq Rt)\) for \(R\) being a suitably

1Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, N.T. Hong Kong; xfgao@se.cuhk.edu.hk

2Department of Mathematics, Florida State University, 1017 Academic Way, Tallahassee, FL-32306, United States of America; zhu@math.fsu.edu.
large number. Relying on the large deviations result, they developed an asymptotically optimal importance sampling algorithm to estimate small tail probabilities \( \mathbb{P}(V(t) \geq Rt) \).

In this paper, we derive explicit asymptotic expansions of large deviation probabilities \( \mathbb{P}(V(t) \geq Rt) \) and tail expectations \( \mathbb{E}[g(V(t) - Rt) \cdot 1_{V(t) \geq Rt}] \) for a wide class of real-valued functions \( g \) as \( t \to \infty \). See Theorem 4. Such results improve the logarithmic-scale large deviations asymptotics in [36]. By truncating the asymptotic expansions, we can develop explicit approximations for large deviation probabilities and tail expectations for affine point processes. These approximations can be useful when exact computations or numerical inversion methods have difficulties, e.g., when the ODEs governing the transform of an affine point process do not have closed-form solutions, and the probability of the event \( \{V(t) \geq Rt\} \) is very small. Our numerical studies and applications in credit risk analysis (see Section 6) illustrate that these analytical approximations can be faster to evaluate than importance sampling simulations in [36], and they can perform quite well. So these approximations can potentially provide alternatives to Monte Carlo simulation which is computer resource extensive.

Deriving asymptotic expansions of large deviation probabilities and tail conditional expectations for affine point processes in the large–time regime is a difficult problem, mainly because such point processes have complex self- and mutually–exciting dependence structures, and also we aim to obtain asymptotic expansions beyond the logarithmic asymptotics. In the literature, asymptotic expansions for large deviation probabilities date back to [3, 9], where they derived expansions for tail probabilities of sums of independent and identically distributed (i.i.d) random variables, which can be applied to Poisson processes. See also [12, 33]. These asymptotic expansions go beyond the logarithmic asymptotics known as the large deviation principle in [35], also known as the Donsker–Varadhan type large deviations. Beyond the i.i.d case, there are some very general and relatively easy-to-check conditions to guarantee the large deviation principle, e.g., Gärtner-Ellis theorem, which is used in [36] to obtain the large deviation principle for affine point processes. Nevertheless, there are not many general results for asymptotic expansions beyond the logarithmic asymptotics. Chaganty and Sethuraman [7] extended the Gärtner-Ellis theorem and obtained a more refined large deviations result, the so-called strong large deviation, but their results do not contain higher–order expansions. See also [28]. While some prior work (see e.g. [5, 26, 37] and the references therein) have studied large deviations of certain point processes, they also do not obtain asymptotic expansions. Higher-order large deviations expansions are only known for some special models, see e.g. [4].

To overcome this difficulty, in this paper we develop a new approach that is based on the mod-\( \phi \) convergence theory studied recently in the probability literature, see e.g. [11, 27, 31]. In particular, the authors of [31] used the framework of mod-\( \phi \) convergence to obtain precise estimates of \( \mathbb{P}(X_n \in t_nB) \) for large \( t_n \) where \( B \) is a Borel set, instead of the usual estimates for the rate of exponential decay \( \frac{1}{n} \log \mathbb{P}(X_n \in t_nB) \), for quite general sequences of random variables \((X_n)_{n \in \mathbb{N}} \). This framework allows one to obtain precise large deviations, and refined central limit theorems together for \((X_n)_{n \in \mathbb{N}} \) if one can verify
that \((X_n)_{n \in \mathbb{N}}\) converges in the mod-\(\phi\) sense where \(\phi\) is an infinitely divisible distribution. The convergence speed determines the order of asymptotic expansions of the deviation probabilities \(P(X_n \in t_n B)\). See Section 4 for a short introduction and further details.

Specifically, our approach is as follows. First, we extend and sharpen the results in [31] for quite general sequence of random variables. See Propositions 8 and 11. The paper [31] focused on asymptotic expansions of deviation probabilities, and we extend their results to obtain asymptotic expansions of tail expectations in the context of mod-\(\phi\) convergence. In addition, in the case when \(\phi\) is non-lattice distributed, the paper [31] obtained expansions of deviation probabilities of zeroth order. With additional assumptions, we obtain high order expansions of deviation probabilities and tail expectations, by some delicate applications of Esseen’s smoothing inequality, Laplace’s method and Faà di Bruno’s formula (see Appendix C). These results hold for quite general sequence of random variables, so they are of independent interest and can be useful in other applications, e.g. establish limit theorems for other stochastic processes.

Second, we prove that for affine point processes, the sequence of random variables \((V(t))_{t>0}\) converges mod-\(\phi\) exponentially fast as \(t \to \infty\), see Theorem 13. This step is also nontrivial as one needs to identify the infinitely divisible distribution \(\phi\), establish the exponential convergence and characterize the limiting function. Our proof relies on the measure-change technique in [36] and a careful analysis of the affine structure of the point process as well as the ODEs governing the transform of an affine point process.

With these two steps, we can then apply the general extended results in the first step to \((V(t))_{t>0}\), and obtain explicit asymptotic expansions of large deviation probabilities and tail expectations for affine point processes. Moreover, from [31], the mod-\(\phi\) convergence of \((V(t))_{t>0}\) we establish also implies a refined central limit theorem for affine point processes (see Theorem 5), which extends the central limit theorem in [36]. While we have limited the discussions to affine point processes in this paper due to their analytical tractability, we remark that the tools developed here can be potentially used in other settings.

The rest of the paper is organized as follows. Section 2 introduces affine point processes and existing large-time asymptotics results in [36]. Section 3 presents our main results. In Section 4 we review the mod-\(\phi\) convergence framework in [31] and present related new results for quite general sequence of random variables. In Section 5 we establish the mod-\(\phi\) convergence of \((V(t))_{t>0}\) and prove the main results in Section 3. Section 6 presents numerical studies and financial applications of our theoretical results. Finally, proofs of other technical results and the computations of expansion coefficients for affine point processes are collected in the Appendix.

2 Affine point processes and existing results

To make the paper self-contained, we follow [36] to introduce affine point processes in this section and review their results on large-time asymptotics of such point processes.
We fix a complete probability space \((\Omega, \mathbb{P}, \mathbb{F})\) and a filtration \(\{\mathbb{F}_t : t \geq 0\}\) satisfying the usual conditions of right continuity and completeness (see, e.g. [29]). We write \(\mathbb{R}^d_+ = \{y \in \mathbb{R}^d : y_i \geq 0, i = 1, \ldots, d\}\). Let \(W = (W(t) : t \geq 0)\) be a standard \(d\)-dimensional Brownian motion. Let \(X = (X_1, \ldots, X_d)\) be an affine jump diffusion satisfying the stochastic differential equation:

\[
dX(t) = \mu(X(t))dt + \sigma(X(t))dW(t) + \sum_{i=1}^{n} \gamma_i \int_{\mathbb{R}_+} zN_i(dt,dz),
\]

with \(X(0) = x_0\), where the drift and volatility functions are given by

\[
\mu(x) = b - \beta x, \quad b \in \mathbb{R}^d, \quad \beta \in \mathbb{R}^{d \times d},
\]

\[
\sigma(x)\sigma(x)^T = a + \sum_{j=1}^{d} \alpha^j x_j, \quad a \in \mathbb{R}^{d \times d}, \quad \alpha^j \in \mathbb{R}^{d \times d}, \quad j = 1, \ldots, d.
\]

Here \(\gamma_i \in \mathbb{R}^d\) and \(N_i(dt,dz)\) is a random counting measure on \([0, \infty) \times \mathbb{R}_+\) with compensator measure \(\Lambda_i(X(t))dt \varphi_i(dz)\), where \(\varphi_i\) is a probability measure on \(\mathbb{R}_+\) and

\[
\Lambda_i(x) = \lambda_i + \sum_{j=1}^{d} \kappa_{i,j} x_j, \quad \text{for} \quad i = 1, \ldots, n.
\]

We use \(Z_i\) to denote a random variable having distribution \(\varphi_i\). An \(n\)-dimensional affine point process \(L = (L_1, \ldots, L_n)\) is given by

\[
L_i(t) := \int_0^t \int_{\mathbb{R}_+} zN_i(ds,dz).
\]

We note that the term \(\sum_{i=1}^{n} \gamma_i \int_{\mathbb{R}_+} zN_i(dt,dz)\) in (2.1) introduces self- and mutual-excitation into \(L\). Without this term, such effects are absent. Examples of affine point processes will be illustrated in Section 6.

The study [36] obtained the large-time asymptotics (See Theorems 2 and 3 below) of

\[
V(t) := \sum_{i=1}^{n} L_i(t).
\]

We aim to sharpen their results. To this end, we follow [36] to impose the following assumption throughout the paper. We use the following notations: \(A_{I,J} = (A_{ij} : i \in I, j \in J)\) for a matrix \(A\) and two index sets \(I\) and \(J\), \(A^T\) denotes the transpose, and \(Id(i)\) denotes a matrix with all entries equal to 0 except the \(i\)-th diagonal entry, which is 1.
Assumption 1. (I) There exist index sets $I = \{1, \ldots, m\}$ and $J = \{m + 1, \ldots, d\}$ such that
(1) $a$ is a symmetric positive semi-definite matrix with $a_{I,I} = 0$.
(2) $\alpha_i$ is symmetric positive semi-definite and $\alpha_i^{I,I} = \alpha_i^{i,i}I_{d_i}$ for each $i \in I$; $\alpha_i = 0$ for $i \in J$.
(3) $b \in \mathbb{R}_+^m \times \mathbb{R}^{d-m}$.
(4) $\beta_{I,J} = 0$ and $\beta_{I,I}$ has nonpositive off-diagonal elements.
(5) $\lambda = (\lambda_i) \in \mathbb{R}_+^n$, $\kappa \in \mathbb{R}^{n \times d}$ with $\kappa_{i,J} = 0$ for $i = 1, \ldots, n$.
(6) $\gamma_i \in \mathbb{R}_+^m \times \mathbb{R}^{d-m}$, for $i = 1, \ldots, n$.
(II) $\alpha_i^{i,i} > 0$ and $b_i > 0$ for each $i = 1, \ldots, m$; $\lambda_i + \sum_{j=1}^m \kappa_{i,j} > 0$ for each $i = 1, \ldots, n$.
(III) $\beta - \sum_{i=1}^n \mathbb{E}[Z_i] \gamma_i \kappa_i^T$ is positive stable, where $\kappa_i^T$ is the $i$-th row of $\kappa = (\kappa_{i,j})$, $i = 1, \ldots, n$.

This assumption on the parameters $(\alpha, a, b, \beta, \lambda, \kappa, \gamma)$ of the SDE in (2.1) essentially ensures that the affine point process is properly defined and $X(\cdot)$ in (2.1) is ergodic. We refer the readers to [36] for further discussions on this assumption.

We next summarize two main mathematical results in [36]. The first result is a central limit theorem for $V(t)$ as $t \to \infty$.

**Theorem 2** (Theorem 1 in [36]). Assume $\mathbb{E}[(Z_i)^{2+\epsilon}] < \infty$ for some $\epsilon > 0$ for all $i = 1, \ldots, n$. Under Assumption 1, we have as $t \to \infty$,

$$\frac{V(t) - rt}{\sqrt{t}} \to N(0, \sigma^2),$$

in distribution, where

$$r = A^T b + \sum_{i=1}^n \lambda_i \mathbb{E}[Z_i](1 + A^T \gamma_i),$$

$$\sigma^2 = A^T a A + C^T \lambda + (A^T \alpha A + C^T \kappa) B,$$

$$A^T = \left( \sum_{i=1}^n \mathbb{E}[Z_i] \kappa_i^T \right)^{-1} \left( \beta - \sum_{i=1}^n \mathbb{E}[Z_i] \gamma_i \kappa_i^T \right),$$

$$B = \left( \beta - \sum_{i=1}^n \mathbb{E}[Z_i] \gamma_i \kappa_i^T \right)^{-1} \left( b + \sum_{i=1}^n \lambda_i \mathbb{E}[Z_i] \gamma_i \right),$$

$$C_i = (1 + A^T \gamma_i)^2 \mathbb{E}(Z_i)^2, \quad i = 1, 2, \ldots, n.$$

The second result is a large deviation principle for $V(t)$ as $t \to \infty$.

**Theorem 3** (Theorem 2 in [36]). Assume $R > r$ and suppose $\sup \{ \theta \in \mathbb{R} : \mathbb{E}[e^{\theta Z_i}] < \infty \} > 0$ for each $i = 1, \ldots, n$. Under Assumption 1, we have

$$\lim_{t \to \infty} \frac{1}{t} \log \mathbb{P}(V(t) \geq Rt) = -I(R),$$

5
where $\mathcal{I}(R) = \sup_{\theta \in \mathbb{R}} \{\theta R - \eta(\theta)\}$, and

$$
\eta(\theta) = u^*(\theta)^T b + \sum_{i=1}^d \lambda_i \left( \mathbb{E} \left[ e^{(\theta^* + u^*(\theta)^T \gamma_i)Z_i} \right] - 1 \right),
$$

(2.3)

where $u^*(\theta) : \mathbb{R} \to \mathbb{R}^n$ is the implicit function defined as the unique solution branch with $u^*(0) = 0$ of the system of nonlinear equations:

$$
\sum_{i=1}^d u_i \beta_{i,j} - \frac{1}{2} u^T \alpha^j u - \sum_{i=1}^n \left( \mathbb{E} \left[ e^{(\theta + u^T \gamma_i)Z_i} \right] - 1 \right) \kappa_{i,j} = 0, \quad j = 1, 2, \ldots, d.
$$

(2.4)

Note that there exists some $\theta_c > 0$ so that $u^*(\theta)$ and $\eta(\theta)$ in Theorem 3 above are well-defined for any $\theta \leq \theta_c$. For the definition of $\theta_c$ and related discussions, we refer to Section 4.3 in [36].

Before we present our results, we remark that [36] also considered the limit theorems for weighted combinations of $L_i(t)$ and the left tail of $V(t)$ as $t \to \infty$. For the simplicity of the presentation, in this paper we restrict our discussions on the refining the above two results (Theorems 2 and 3) only, although similar refinements can be also obtained for weighted combinations of $L_i(t)$ and the left tail of $V(t)$.

### 3 Main results: refined large-time asymptotics

In this section, we present our main results on refinements of large–time asymptotics of affine point processes. Recall that for two real-valued functions $f_1, f_2$, we write $f_1 = O(f_2)$ as $t \to \infty$ if there are constants $c_1$ and $c_2 > 0$ such that $|f_1(t)| \leq c_1 |f_2(t)|$ whenever $t > c_2$. We write $f_1 = o(f_2)$ as $t \to \infty$ if $f_1(t)/f_2(t) \to 0$ as $t \to \infty$.

Our first result is a precise large deviation for $V(t)$ for $t$ goes to infinity, which extends Theorem 3 (Theorem 2 in [36]). Recall the quantities $r, \mathcal{I}(R), \eta$ given in Theorem 3

**Theorem 4.** Let $R > r$ and $h \in \mathbb{R}$ defined by $\eta'(h) = R$. Under the same assumptions as in Theorem 3 we have the following.

1. If the random variables $\sum_{i=1}^n \sum_{j=1}^{n_i} \xi_{ij}$, where $n_i \in \mathbb{N} \cup \{0\}$ and $Z_{ij}$ are i.i.d. distributed as $Z_i$, are supported on $b\mathbb{N} \cup \{0\}$ for some parameter $b > 0$, then there exist constants $(c_k)_{k=0}^\infty$ such that for $Rt \in b\mathbb{N}$, as $t \to \infty$,

$$
\mathbb{P}(V(t) \geq Rt) = \frac{e^{-t\mathcal{I}(R)}}{\sqrt{2\pi t \eta''(h)}} \left( c_0 + \sum_{k=1}^v \frac{c_k}{t^k} + O \left( \frac{1}{t^v} \right) \right),
$$

(3.1)

where $v$ is any positive integer. More generally, assume $g : \mathbb{R}_+ \to \mathbb{R}$ satisfies $|g(x)| \leq \hat{a} e^{h \hat{a}}$ for some $\hat{h} < h$ and $\hat{a} > 0$, then there exist constants $(\hat{c}_k)_{k=0}^\infty$ such that for any
\[ R_t \in bN, \text{ as } t \to \infty, \text{ we have} \]
\[ \mathbb{E} \left[ g(V(t) - R_t) \cdot 1_{V(t) \geq R_t} \right] = \frac{e^{-t\mathbb{I}(R)}}{\sqrt{2\pi \eta'(h)}} \cdot \left( \hat{c}_0 + \sum_{k=1}^{v} \frac{\hat{d}_k}{t^k} + O \left( \frac{1}{t^{v+1}} \right) \right). \tag{3.2} \]

(2) Otherwise, there exist constants \((d_k)_{k=0}^{\infty}\) such that as \(t \to \infty\),
\[ \mathbb{P}(V(t) \geq R_t) = \frac{e^{-t\mathbb{I}(R)}}{\sqrt{2\pi \eta'(h)}} \cdot \left( d_0 + \sum_{k=1}^{v} \frac{d_k}{t^k} + o \left( \frac{1}{t^v} \right) \right), \]
where \(v\) is any positive integer. In addition, assume that \(g : \mathbb{R}_+ \to \mathbb{R}\) admits the expansion \(g(x) = \sum_{k=0}^{\infty} g_k x^{k+\Delta}\), where \(\Delta \in [0,1)\) and there exist some \(\bar{a} > 0\) and \(0 < \bar{h} < h\), such that \(g_k \leq \frac{\bar{a}^k}{k!}\) for every \(k \in \mathbb{N}\), then there exist constants \((\hat{d}_k)_{k=0}^{\infty}\) such that as \(t \to \infty\),
\[ \mathbb{E} \left[ g(V(t) - R_t) \cdot 1_{V(t) \geq R_t} \right] = \frac{e^{-t\mathbb{I}(R)}}{\sqrt{2\pi \eta'(h)}} \cdot \left( \hat{d}_0 + \sum_{k=1}^{v} \frac{\hat{d}_k}{t^k} + o \left( \frac{1}{t^v} \right) \right). \]

The coefficients \((c_k, \hat{c}_k, d_k, \hat{d}_k)_{k=0}^{\infty}\) in the above result have explicit formulas (see Propositions 8 and 11, and Appendix B), and can be numerically computed. Hence, based on Theorem 4, we can develop approximations for large deviation tail probabilities and tail expectations for \(V(t)\) by truncations of the asymptotic expansions. Numerical results on the quality of approximations will be illustrated in Section 6.

Our next result is a refined central limit theorem (CLT) for \(V(t)\), which improves Theorem 2 (i.e. Theorem 1 in [36]) under a stronger assumption.

**Theorem 5** (Refined CLT). Let \(y = o(t^{1/6})\). Assume the random variable \(V(t)\) is either lattice distributed or it has a non-lattice law that is absolutely continuous with respect to Lebesgue measure. Under the same assumptions as in Theorem 3, we have as \(t \to \infty\),
\[ \mathbb{P} \left( V(t) \geq rt + \sigma y \sqrt{t} \right) = \int_{y}^{\infty} \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}} du \cdot (1 + o(1)), \]
where \(r\) and \(\sigma\) are given in Theorem 3.

To prove Theorems 4 and 5, we use an approach that is based on the mod-\(\phi\) convergence framework which we discuss in the next section. The stronger assumption needed in Theorem 5 is required in using this approach. On the other hand, in [36], one needs extra effort to prove the large deviation principle in addition to the central limit theorem, while the mod-\(\phi\) convergence provides a unified approach to establish both precise large deviations and refined CLT. We give the details of the proofs of Theorems 4 and 5 in Section 5.
4 Mod-\(\phi\) convergence framework and some new results

In this section, we briefly review the mod-\(\phi\) convergence framework in \([31]\), and present some related new results (Propositions 8 and 11) for quite general sequence of random variables. These new results extend the results on precise estimates of large deviations probabilities in \([31]\), and they are of independent interest.

We first recall the definition of mod-\(\phi\) convergence, where one considers a renormalization of the characteristic functions of random variables. Let \((X_n)_{n \in \mathbb{N}}\) be a sequence of real-valued random variables and \(E[e^{zX_n}]\) exist in a strip \(S_{(c,d)} := \{z \in \mathbb{C} : c < R(z) < d\}\), with \(c < d\) being extended real numbers, (i.e. we allow \(c = -\infty\) and \(d = +\infty\)) and \(R(z)\) denotes the real part of \(z \in \mathbb{C}\).

**Definition 6** (Definition 1.1 in \([31]\)). We say that \((X_n)_{n \in \mathbb{N}}\) converges mod-\(\phi\) on \(S_{(c,d)}\) with parameters \((t_n)_{n \in \mathbb{N}}\) and the limiting function \(\psi\), if there exists a non-constant infinitely divisible distribution \(\phi\) with \(\int_{\mathbb{R}} e^{zx} \phi(dx) = e^{\eta(z)}\), which is well defined on \(S_{(c,d)}\), and an analytic function \(\psi(z)\) that does not vanish on the real part of \(S_{(c,d)}\) such that as \(n \to \infty\), we have \(t_n \to +\infty\), and

\[
e^{-t_n \eta(z)} \cdot E[e^{zX_n}] \to \psi(z), \quad \text{locally uniformly in } z \in S_{(c,d)}.
\]

(4.1)

In addition, we say that \((X_n)_{n \in \mathbb{N}}\) converges mod-\(\phi\) at speed \(O((t_n)^{-v})\) (respectively, exponentially fast) if the difference of the two sides of Equation (4.1) is uniformly bounded by \(C_K(t_n)^{-v}\) (respectively, \(C_K e^{-t \bar{C}_K}\)) for \(z\) in a compact subset \(K\) of \(S_{(c,d)}\), for some constants \(C_K\) and \(\bar{C}_K\).

An informal interpretation of the convergence in (4.1) is that \(X_n\) can be represented as the sum of \(t_n\) independent copies of the (non-constant) infinitely divisible distribution \(\phi\) plus a perturbation encoded in the limiting function \(\psi\). Note in the above definition, we have slightly abused the notation by using the same notation \(\eta\) (as in Sections 2 and 3 for affine point processes) here in the general setting.

The mod-\(\phi\) convergence framework allows one to obtain precise estimates for tail probabilities and one needs to consider two separate cases: lattice distributed \(\phi\), and non-lattice distributed \(\phi\). We next review the results in \([31]\) and present our new results for these two cases. To facilitate the presentation, we write

\[
S_n := \{(m_1, \ldots, m_n) : 1 \cdot m_1 + 2 \cdot m_2 + \cdots + n \cdot m_n = n, \text{with each } m_i \in \mathbb{N} \cup \{0\}\}.
\]

(4.2)

This set \(S_n\) appears whenever we apply Faà di Bruno’s formula (see Lemma 17 in Appendix C.2) for derivatives of composite functions for \(n \geq 1\), and whenever \(S_0\) appears, we simply consider the 0-th order derivative.

8
4.1 Lattice case

We first discuss the case where \( \phi \) is lattice distributed. For a given \( x \), define \( h \) (with a slight abuse of notations) and \( F(x) \) by

\[
\eta'(h) = x, \quad \text{and} \quad F(x) = \sup_{\theta \in \mathbb{R}} \{ \theta x - \eta(\theta) \}. \tag{4.3}
\]

The following result from [31] yields an expansion for the tail probabilities when \( \phi \) has a lattice distribution, which generalizes the Buhadur–Rao theorem for sums of i.i.d random variables in large deviations theory [3, 12]. Without loss of generality, as in [31], we assume that \( X_n \)'s and the infinite divisible distribution \( \phi \) both take values in \( \mathbb{Z} \) and \( \mathbb{Z} \) is the minimal lattice for \( \phi \).

**Theorem 7** (Theorem 3.4. [31]). Suppose \( \phi \) is lattice distributed, and \((X_n)_{n \in \mathbb{N}}\) converges mod-\( \phi \) at speed \( O((t_n)^{-v}) \) on a band \( S(c,d) \) (\( c < 0 < d \)) with parameters \((t_n)_{n \in \mathbb{N}}\) and the limiting function \( \psi \). Then for \( x \in (\eta'(0), \eta'(d)) \) and \( t_n x \in \mathbb{N} \), we have as \( n \to \infty \),

\[
P(X_n \geq t_n x) = \frac{e^{-t_n F(x)}}{\sqrt{2\pi t_n \eta''(h)}} \left( c_0 + \frac{c_1}{t_n} + \ldots + \frac{c_{v-1}}{(t_n)^{v-1}} + O \left( \frac{1}{(t_n)^v} \right) \right),
\]

where \( c_0 = \frac{\psi(h)}{1-e^{-h}} \) and \((c_k)_{k=1}^{v-1}\) can be computed as follows:

\[
c_k = \sum_{m+\ell+n=2k} \sum_{q=0}^{\infty} e^{-qh} \frac{(-q)^m}{m!} \frac{\psi(\ell)(h)}{\ell!} \cdot \sum_{S_n} \frac{(-1)^{m_1+\cdots+m_n}}{m_1!m_2!2!^{m_2} \cdots m_n!n!m_n} \cdot \prod_{j=1}^{n} \left( \frac{\eta^{(j+2)}(h)}{\eta''(h) (j+2)(j+1)} \right)^{m_j} (-1)^k \frac{(2(k + m_1 + \cdots + m_n) - 1)!!}{(\eta''(h))^k},
\]

where \( S_n \) is defined in [1.2].

We extend the above theorem to obtain precise estimates of tail expectations. The proof of the following result is given in the appendix.

**Proposition 8.** Suppose the assumptions in Theorem 7 hold. In addition, assume \( g : \mathbb{R}_+ \to \mathbb{R} \) satisfies \(|g(x)| \leq a e^{hx} \) for some \( h < h \) and \( a > 0 \). Then as \( n \to \infty \), we have

\[
\mathbb{E}[g(X_n - t_n x) 1_{X_n \geq t_n x}] = \frac{e^{-t_n F(x)}}{\sqrt{2\pi t_n \eta''(h)}} \left( c_0 + \frac{\hat{c}_1}{t_n} + \ldots + \frac{\hat{c}_{v-1}}{(t_n)^{v-1}} + O \left( \frac{1}{t_n} \right) \right),
\]
where \( h \) and \( F \) are defined in \([4,3]\). Here, \( c_0 = \sum_{k=0}^{\infty} g(k)e^{-kh} \cdot \psi(h) \), and \((\hat{c}_k)_{k=1}^{\nu-1}\) can be computed using \( g \) and the higher order derivatives of \( \eta \) and \( \psi \) at \( h \):

\[
\hat{c}_k = \sum_{m+\ell+n=2k} \sum_{q=0}^{\infty} g(q)e^{-qh} \frac{(-q)^m}{m!} \cdot \frac{\psi^{(\ell)}(h)}{\ell!} \cdot \sum_{s_n} \frac{(-1)^{m_1+\cdots+m_n}}{m_1!m_2!m_3!\cdots m_n!n!m} \cdot \prod_{j=1}^{n} \left( \frac{1}{\eta^{(j+2)}(h)} \right)^{m_j} \frac{(-1)^k(2(k + m_1 + \cdots + m_n) - 1)!!}{(\eta''(h))^k}, \tag{4.4}
\]

where \( S_n \) is defined in \([4,2]\).

**Remark 9.** From the above two results, one can readily find that

\[
c_1 = -\frac{1}{2} \psi(h) e^{-h} + e^{-2h} \frac{1}{(1-e^{-h})^3} \frac{1}{\eta''(h)} - \frac{1}{2} \frac{1}{1-e^{-h}} \psi''(h) \frac{1}{\eta''(h)} \\
+ \frac{1}{1-e^{-h}} \psi'(h) \left[ \frac{1}{8} \eta^{(4)}(h) \frac{1}{(\eta''(h))^2} - \frac{5}{24} (\eta^{(3)}(h))^2 \frac{1}{(\eta''(h))^3} \right] \\
+ \frac{e^{-h}}{(1-e^{-h})^2} \psi'(h) \frac{1}{\eta''(h)} + \frac{1}{2} \frac{1}{1-e^{-h}} \psi'(h) \eta^{(3)}(h) \frac{1}{(\eta''(h))^2} \\
+ \frac{1}{2} \frac{e^{-h}\psi'(h)}{(1-e^{-h})^2} \eta^{(3)}(h) \frac{1}{(\eta''(h))^2},
\]

and

\[
\hat{c}_1 = -\frac{1}{2} \psi(h) \sum_{k=0}^{\infty} g(k) e^{-kh} k^2 \frac{1}{\eta''(h)} - \frac{1}{2} \sum_{k=0}^{\infty} g(k) e^{-kh} \psi''(h) \frac{1}{\eta''(h)} \\
+ \sum_{k=0}^{\infty} g(k) e^{-kh} \psi'(h) \left[ \frac{1}{8} \eta^{(4)}(h) \frac{1}{(\eta''(h))^2} - \frac{5}{24} (\eta^{(3)}(h))^2 \frac{1}{(\eta''(h))^3} \right] \\
+ \sum_{k=0}^{\infty} g(k) e^{-kh} k \psi'(h) \frac{1}{\eta''(h)} + \frac{1}{2} \sum_{k=0}^{\infty} g(k) e^{-kh} \psi'(h) \eta^{(3)}(h) \frac{1}{(\eta''(h))^2} \\
- \frac{1}{2} \sum_{k=0}^{\infty} g(k) e^{-kh} k \psi'(h) \eta^{(3)}(h) \frac{1}{(\eta''(h))^2}.
\]

We will use and compute these coefficients in numerical experiments in Section 6.

**4.2 Non-lattice case**

We next discuss the case where \( \phi \) is non-lattice distributed. We first present the result from \([31]\).
Theorem 10 (Theorem 4.3. [31]). Suppose $\phi$ is non-lattice, and $(X_n)_{n \in \mathbb{N}}$ converges mod-$\phi$ with parameters $(t_n)_{n \in \mathbb{N}}$ and the limiting function $\psi$ on a band $\mathcal{S}_{(c,d)}$ with $c < 0 < d$. If $x \in (\eta'(0), \eta'(d))$, then

$$
P(X_n \geq t_n x) = \frac{e^{-t_n F(x)}}{2\pi t_n \eta''(h)} \left( \frac{\psi(h)}{h} + o(1) \right),$$

where $h$ is defined via $\eta'(h) = x$ and $F(x) := \sup_{\theta \in \mathbb{R}} \{\theta x - \eta(\theta)\}$.

In Proposition 11 below, we sharpen the above result and extend it to obtain precise estimates for tail expectations under additional assumptions. The proof relies on some delicate applications of Esseen’s smoothing inequality, Laplace’s method and Faà di Bruno’s estimates for tail expectations under additional assumptions. The proof relies on some delicate applications of Esseen’s smoothing inequality, Laplace’s method and Faà di Bruno’s formula which is given in the appendix.

Proposition 11. Suppose $\phi$ is non-lattice distributed, and $(X_n)_{n \in \mathbb{N}}$ converges mod-$\phi$ with parameters $(t_n)_{n \in \mathbb{N}}$ and the limiting function $\psi$ at the speed $O((t_n)^{-v})$ on a band $\mathcal{S}_{(c,d)}$ with $c < 0 < d$. Assume that $x \in (\eta'(0), \eta'(d))$.

(i) We have as $t_n \to \infty$,

$$
P(X_n \geq t_n x) = \frac{e^{-t_n F(x)}}{\sqrt{2\pi t_n \eta''(h)}} \left[ d_0 + \frac{d_1}{t_n^1} + \frac{d_2}{t_n^2} + \cdots + \frac{d_{v-1}}{t_n^{v-1}} + o(t_n^{-v+1}) \right].$$

Here, $(d_k)_{k=0}^{v-1}$ are given by

$$
d_k = \frac{1}{h} \cdot \sum_{p=0}^{2k} \sum_{\ell=0}^{p} \frac{\psi(p-\ell)(h)}{(p-\ell)!} \\
\cdot \frac{1}{S_{\ell}m_1!m_2!m_3! \cdots m_{\ell}!} \prod_{j=1}^{\ell} \frac{\eta(j+2)(h)}{\eta''(h)(j+2)(j+1)} \cdot \frac{1}{(\eta''(h))^{p/2}} \\
\cdot \sum_{m=0}^{\lfloor \frac{p}{2} + m_1 + \cdots + m_\ell \rfloor} \frac{2^m(-1)^m(p + 2(m_1 + \cdots + m_\ell))!}{m!(p + 2(m_1 + \cdots + m_\ell) - 2m)!} \\
\cdot \frac{\Gamma(k + (m_1 + \cdots + m_\ell) - m + 1)a_{k+m-p-(m_1+\cdots+m_\ell)}}{(h^2 \eta''(h))^{k+1}},$$

where $S_{\ell}$ is defined in [4,2] and $(a_k)_{k=0}^{\infty}$ is defined recursively as:

$$a_k = (-1)^k - \sum_{j=0}^{k-1} \frac{(j+s)k-j}{(k-j)!2^{k-j}}a_j,$$

and $(\cdot)_i$ is the Pochhammer symbol.\footnote{Pochhammer symbol is defined as $(x)_n = \Gamma(x+n)/\Gamma(x)$.}
(ii) Furthermore, assume that $g : \mathbb{R}_+ \to \mathbb{R}$ admits the expansion $g(x) = \sum_{k=0}^{\infty} g_k x^{k+\Delta}$ where $\Delta \in [0,1)$ and there exist some $\hat{a} > 0$ and $0 < \hat{h} < h$ so that $g_k \leq \frac{\hat{a} x^{p_k}}{k!}$ for every $k \in \mathbb{N}$. Then we have

$$
\mathbb{E}[g(X_n - t_n x)1_{X_n \geq t_n x}] = \frac{e^{-t_n F(x)}}{\sqrt{2\pi} t_n \eta''(h)} \left[ \hat{d}_0 + \frac{\hat{d}_1}{t_n} + \frac{\hat{d}_2}{t_n^2} + \cdots + \frac{\hat{d}_{n-1}}{t_n^{n-1}} + o(t_n^{-v+1}) \right],
$$
as $t_n \to \infty$, where $(\hat{d}_k)_{k=0}^{v-1}$ are given by

$$
\hat{d}_k = \sum_{q=0}^{\infty} g_q \cdot \frac{1}{h^{\Delta+q+1}} \cdot \prod_{p=0}^{2k} \psi(p-\ell)(h) \cdot \prod_{m=0}^{\lfloor \frac{k}{2} \rfloor} \frac{\eta''(h)}{m!} \cdot \frac{\eta''(h)}{(\eta''(h))^{p/2}} \cdot \frac{\eta''(h)}{m!} \\
\sum_{m=0}^{\infty} \frac{m! (-1)^m (p + 2(m_1 + \cdots + m_\ell))!}{m!(p + 2(m_1 + \cdots + m_\ell) - 2m)!} \cdot \frac{\eta''(h)}{m!(p + 2(m_1 + \cdots + m_\ell) - 2m)!} \cdot \Gamma(k + \Delta + q + (m_1 + \cdots + m_\ell) - m + 1) a_{k+m-p-(m_1+\cdots+m_\ell)}.
$$

**Remark 12.** From Part (i) of the above result, one can find that $d_0 = \frac{\psi(h)}{h}$, and

$$
d_1 = -\frac{\psi(h)}{h^2 \eta''(h)} + \psi(h) \left( \frac{\psi''(h)}{\psi(h)} - \frac{\eta'''(h)}{\eta''(h)} \right) \frac{1}{h^2 \eta''(h)} + \psi(h) \left[ \frac{1}{8} \frac{\eta^{(4)}(h)}{(\eta''(h))^2} - \frac{5}{24} \frac{(\eta^{(3)}(h))^2}{(\eta''(h))^3} - \frac{1}{2} \frac{\eta''(h)}{\psi(h) \eta''(h)} + \frac{1}{2} \frac{\eta''(h)}{\psi(h) (\eta''(h))^2} \right].
$$

Similarly, from Part (ii) of the above result, one can check that when $g(x) = x^\gamma$ for $\gamma \geq 0$, we have

$$
\hat{d}_0 = \frac{\Gamma(\gamma + 1) \psi(h)}{h^{\gamma+1}},
$$

and

$$
\hat{d}_1 = -\frac{\psi(h) \Gamma(\gamma + 2)(1 + \frac{\gamma}{h})}{h^{\gamma+1} \eta''(h)} + \psi(h) \left( \frac{\psi''(h)}{\psi(h)} - \frac{\eta'''(h)}{\eta''(h)} \right) \frac{\Gamma(\gamma + 2)}{h \eta''(h)} + \psi(h) \left[ \frac{1}{8} \frac{\eta^{(4)}(h)}{(\eta''(h))^2} - \frac{5}{24} \frac{(\eta^{(3)}(h))^2}{(\eta''(h))^3} - \frac{1}{2} \frac{\eta''(h)}{\psi(h) \eta''(h)} + \frac{1}{2} \frac{\eta''(h)}{\psi(h) (\eta''(h))^2} \right].
$$

We will use and compute these coefficients in numerical experiments in Section 6.
5 Mod-\(\phi\) convergence for affine point processes and proofs of results in Section \([\text{3}]\)

In this section, we prove the main results in Section \([\text{3}]\). The key step is to establish the mod-\(\phi\) convergence of the sequence of random variables \((V(t))_{t>0}\) in \((2.2)\). We summarize the result below in Section \([\text{5.1}]\) and present its proof in Section \([\text{5.3}]\).

5.1 Mod-\(\phi\) convergence of \((V(t))_{t>0}\)

Recall from Theorem \([\text{3}]\) that \(\eta(\theta)\) and \(u^*(\theta)\) are well defined for a real number \(\theta \leq \theta_c\) with \(\theta_c > 0\). One can naturally extends their definitions to a complex number \(\theta \in \mathbb{C}\) with the real part \(R(\theta) \leq \theta_c\), so that they fit in the mod-\(\phi\) convergence framework (Definition \([\text{6}]\) which deals with characteristic functions.

**Theorem 13.** The sequence of random variables \((V(t))_{t>0}\) converges mod-\(\phi\) exponentially fast as \(t \to \infty\) along any lattice with limiting function \(\psi\), where

\[
\psi(\theta) = e^{u^*(\theta)^Tx_0+B(\infty;\theta,u^*(\theta))},
\]

and \(\psi\) is analytic for \(R(\theta) < \theta_c\). Here,

\[
B(t;\theta,\delta) = \int_0^t b^T A(s;\theta,\delta) ds + \int_0^t \frac{1}{2} A^T a A ds + \sum_{i=1}^n \lambda_i \int_0^t \int_{\mathbb{R}_+} \left(e^{A^T \gamma_i z} - 1\right) e^{(\theta + \delta^T \gamma_i)z} \varphi_i(dz) ds,
\]

is a scalar function, where \(A(t;\theta,\delta) = (A_1(t;\theta,\delta), \cdots, A_n(t;\theta,\delta))\) is a matrix function, with \(A(0;\theta,\delta) = -\delta, B(0;\theta,\delta) = 0\) and

\[
\frac{d}{dt} A_j(t;\theta,\delta) = -\sum_{i=1}^n A_i \beta^*_{i,j} + \frac{1}{2} A^T \alpha^* A + \sum_{i=1}^n \int_{\mathbb{R}_+} \left(e^{A^T \gamma_i z} - 1\right) e^{(\theta + \delta^T \gamma_i)z} \varphi_i(dz) \kappa_{i,j},
\]

for \(j = 1, 2, \ldots, n\), where \(\beta^*\) is defined in \((5.8)\).

Several remarks are in order. First, the infinite divisible distribution \(\phi\) will be identified and given in Section \([\text{5.3.2}]\). Second, note that \((V(t))_{t>0}\) is a continuous-time process, so to be consistent with Definition \([\text{3}]\) we have stated the mod-\(\phi\) convergence of \(V(t)\) along any lattice (i.e. \(t = b^* n\) for any \(b^* > 0\) and \(n \in \mathbb{N}\) with \(n \to \infty\)) in the above result. Finally, we use the notation \(B(\infty;\theta,u^*(\theta))\) in \((5.1)\) to denote \(\lim_{t \to \infty} B(t;\theta,u^*(\theta))\) which will be shown to exist in the proof in Section \([\text{5.3}]\).

Relying on Theorem \([\text{13}]\) we are ready to prove the results in Section \([\text{3}]\) in the next section.
5.2 Proofs of Theorems 4 and 5 in Section 3

Proof. We first prove Theorem 4. For Part (1), without loss of generality, assume that the random variables in the set

\[
\{ \sum_{i=1}^{n} \sum_{j=1}^{n} Z_{ij}, \text{where } n_i \in \mathbb{N} \cup \{0\} \text{ and } Z_{ij} \text{ are i.i.d. distributed as } Z_i \}
\]

are supported on \( \mathbb{N} \cup \{0\} \). Then one can readily see that \( V(t) \) takes values in \( \mathbb{N} \cup \{0\} \). In addition, it will be clear from Section 5.3.2 that the infinite divisible distribution \( \phi \) in Theorem 13 also takes values in \( \mathbb{N} \cup \{0\} \). Hence to expand \( \mathbb{P}(V(t) \geq Rt) \), we can assume \( Rt \in \mathbb{N} \), and define \( X_n = V(\frac{t}{n}) \) and \( t_n = \frac{1}{n} \) to apply the mod-\( \phi \) convergence theory. Then, Part (1) of Theorem 4 directly follows from Theorem 13, Theorem 7 and Proposition 8. Similarly, Part (2) of Theorem 4 follows from Theorem 13, Proposition 11 and Theorem A in 22, which allows one to extend the convergence along any lattice (i.e. \( \{b^* n \} \) for any \( b^* > 0 \) and \( n \in \mathbb{N} \) with \( n \to \infty \) so that we can define \( X_n = V(b^* n) \) and \( t_n = b^* n \) to apply the mod-\( \phi \) convergence theory) to any \( t \to \infty \).

We next prove Theorems 5. It follows from Theorems 3.9 and 4.8 in [31], and Theorem 13 that for \( y = o(t^{1/6}) \), as \( t \to \infty \),

\[
\mathbb{P}\left( V(t) \geq t \eta'(0) + \sqrt{t} \eta''(0) y \right) = \int_{y}^{\infty} e^{-\frac{a^2}{2}} du \cdot (1 + o(1)),
\]

where \( \eta(\cdot) \) is defined in (2.3). So it suffices to show \( \eta'(0) = r \) and \( \eta''(0) = \sigma^2 \). From (2.3), we have

\[
\eta'(0) = (u^*)'(0)^T b + \sum_{i=1}^{n} \lambda_i \mathbb{E} \left[ \left( 1 + (u^*)'(0)^T \gamma_i \right) Z_i \right],
\]

\[
\eta''(0) = (u^*)''(0)^T b + \sum_{i=1}^{n} \lambda_i \mathbb{E} \left[ \left( 1 + (u^*)'(0)^T \gamma_i \right)^2 Z_i \right] + \sum_{i=1}^{n} \lambda_i \mathbb{E} \left[ (u^*)''(0)^T \gamma_i Z_i \right].
\]

To compute \( (u^*)'(0) \) and \( (u^*)''(0) \), we can use (2.3) and find that

\[
\beta^T (u^*)'(0) - \kappa^T \mathbb{E} [(1 + (u^*)'(0)^T \gamma)] Z = 0,
\]

and

\[
\beta^T ((u^*)''(0) - (u^*)'(0)^T \alpha (u^*)'(0) - \kappa^T \mathbb{E} [(1 + (u^*)'(0)^T \gamma)^2 Z] - \kappa^T \mathbb{E} [(u^*)''(0)^T \gamma Z] = 0.
\]

Then it is straightforward to verify that \( \eta'(0) = r \) and \( \eta''(0) = \sigma^2 \), where \( r, \sigma^2 \) are given in Theorem 2. Hence, the conclusion follows.

\footnote{If these random variables are supported on \( b \mathbb{N} \cup \{0\} \) for some \( b > 0 \), then in \( \mathbb{P}(V(t) \geq Rt) \) etc., with \( Rt \in b \mathbb{N} \), it is equivalent to \( \mathbb{P}(V'(t) \geq R't) \) with \( R't \in \mathbb{N} \) and \( V'(t) = V(t)/b \) and \( R' = R/b \).}
5.3 Proof of Theorem 13

We prove Theorem 13 in this section. By Definition 6, it suffices to show

(i) \[ \sup_{\theta \in K} \left| e^{-\eta(\theta)t} \cdot \mathbb{E} \left[ e^{\theta V(t)} \right] - \psi(\theta) \right| \leq C_{K} e^{-\bar{C}_{K}t}, \] \hspace{0.5cm} (5.4)

where \( K \) is a compact subset of the strip \( S_{(0, \theta_{c})} \) for some \( \theta_{c} > 0 \) \(^5\) and \( C_{K}, \bar{C}_{K} > 0 \) are two constants.

(ii) For \( \eta(\cdot) \) given in (2.3) with \( \theta \in \mathbb{C} \) and \( \mathcal{R}(\theta) \leq \theta_{c} \), we have \( e^{\eta(\theta)} = \mathbb{E}[e^{\theta Y}] \) for some (non-constant) infinitely divisible random variable \( Y \).

We present the proofs of (i) and (ii) in the next two sections. Before we proceed, we recall from [36] that they introduced \( Q_{\alpha}^{\theta} \), the equivalent probability measure induced by the martingale \( M_{\theta}^{a}(t) = \exp[\theta V(t) - \eta(\theta)t + u^{*}(\theta)^{T}(X(t) - X(0))] \), such that for given \( X(0) = x_{0} \),

\[ \mathbb{E} \left[ e^{\theta V(t) - \eta(\theta)t} \right] = e^{u^{*}(\theta)^{T}X(0)} \mathbb{E}_{Q_{\alpha}^{\theta}} \left[ e^{-u^{*}(\theta)^{T}X(t)} \right]. \] \hspace{0.5cm} (5.5)

By Girsanov’s theorem, under the measure \( Q_{\alpha}^{\theta} \), the process \( X \) is still affine and satisfies the SDE in (2.1) with parameters \((\alpha, a, b, \beta^{*}, \lambda^{*}, \kappa^{*}, \gamma)\), and measure \( \varphi^{*}_{i} \), where

\[ \lambda_{i}^{*} = \lambda_{i} \int_{\mathbb{R}^{+}} e^{(\theta + u^{*}(\theta)^{T} \gamma_{i})z} \varphi_{i}(dz), \] \hspace{1cm} (5.6)

\[ \kappa_{i}^{*} = \kappa_{i} \int_{\mathbb{R}^{+}} e^{(\theta + u^{*}(\theta)\gamma_{i})z} \varphi_{i}(dz), \] \hspace{1cm} (5.7)

\[ \beta^{*} = \begin{pmatrix} \beta_{1,1} - \text{diag}(\alpha_{11}^{T} u^{*}_{1}(\theta), \ldots, \alpha_{m m}^{T} u^{*}_{m}(\theta)) & 0 \\ \beta_{j,1} & \beta_{j,j} \end{pmatrix}, \] \hspace{1cm} (5.8)

\[ \varphi^{*}_{i}(dz) = \frac{e^{(\theta + u^{*}(\theta)\gamma_{i})z} \varphi_{i}(dz)}{\int_{\mathbb{R}^{+}} e^{(\theta + u^{*}(\theta)\gamma_{i})z} \varphi_{i}(dz)}. \] \hspace{1cm} (5.9)

and \( \mathcal{R}(\theta) \leq \theta_{c} \). We will use these facts in the proofs of (i) and (ii).

5.3.1 Locally Uniform Convergence at Exponential Speed in (5.4)

In this section, we prove (5.4). First, we infer from (5.5) that with \( X(0) = x_{0} \),

\[ \lim_{t \to \infty} \mathbb{E} \left[ e^{\theta V(t) - \eta(\theta)t} \right] = e^{u^{*}(\theta)^{T} x_{0}} \lim_{t \to \infty} \mathbb{E}_{Q_{\alpha}^{\theta}} \left[ e^{-u^{*}(\theta)^{T} X(t)} \left| X(0) = x_{0} \right. \right] \]

\[ = e^{u^{*}(\theta)^{T} x_{0}} \mathbb{E}_{Q_{\alpha}^{\theta}} \left[ e^{-u^{*}(\theta)^{T} X(\infty)} \left| X(0) = x_{0} \right. \right] \] \hspace{0.5cm} (5.10)

\(^5\)As we only study the right tail of \( V(t) \), we only need to consider \( \theta \in S_{(0, \theta_{c})} \). For the expression of \( \theta_{c} \) and related discussions, see Section 4 in [36].
where the random vector $X(\infty)$ follows the stationary distribution of $X(\cdot)$ under the measure $Q^*_\theta$, and the second equality follows from the ergodicity of $X(\cdot)$, see Proposition 11 in [36].

We next compute the right hand side of (5.10) and show it is exactly $\psi$ given in (5.1). We can compute that $v(t, x) := \mathbb{E}Q^*_\theta[e^{-\delta t}X(t)\mid X(0) = x]$ satisfies the Kolmogorov equation, which is a partial integro-differential equation in our context:

$$\frac{\partial v}{\partial t} = (b - \beta^* x) \cdot \nabla v(t, x) + \frac{1}{2} \sum_{i, j=1}^d \left( a_{ij} + \sum_{k=1}^m \alpha^k_{ij} x_k \right) \frac{\partial^2 v}{\partial x_i \partial x_j}(x)$$

$$+ \sum_{i=1}^n (\lambda^*_i + \kappa^*_i \cdot x) \int_{\mathbb{R}^+} [v(t, x + \gamma_i z) - v(t, x)] \varphi^*_i(dz),$$

with the initial condition $v(0, x) = e^{-\delta x}$. Since the process $X$ is still affine under the measure $Q^*_\theta$, one can readily obtain that

$$v(t, x) = e^{A(t; \theta, \delta)^T x + B(t; \theta, \delta)},$$

(5.11)

where $A(t; \theta, \delta)$ and $B(t; \theta, \delta)$ are given in (5.3) and (5.2) respectively. By ergodicity of $X(\cdot)$, we know that $\mathbb{E}Q^*_\theta \left[ e^{-u^*(\theta)^T X(\infty)} \right] X(0) = x_0$ is independent of the initial position $x_0$ and thus $A(\infty; \theta, u^*(\theta)) = 0$ and we conclude that

$$e^{u^*(\theta)^T x_0} \cdot \mathbb{E}Q^*_\theta \left[ e^{-u^*(\theta)^T X(\infty)} \right] X(0) = x_0 = e^{u^*(\theta)^T x_0 + B(\infty; \theta, u^*(\theta))} = \psi(\theta).$$

It remains to show the convergence in (5.10) is exponentially fast in $t$ and locally uniformly in $\theta$. By (5.11), we need to show $e^{A(t; \theta, u^*(\theta))^T x_0 + B(t; \theta, u^*(\theta)) + u^*(\theta)^T x_0}$ converges exponentially fast to $e^{B(\infty; \theta, u^*(\theta)) + u^*(\theta)^T x_0}$. So it suffices to show that

$$A(t; \theta, u^*(\theta))^T x_0 + B(t; \theta, u^*(\theta)) \to B(\infty; \theta, u^*(\theta)),$$

exponentially fast in $t \to \infty$ locally uniformly in $\theta$.

Since $A(t; \delta) \to 0$ as $t \to \infty$, for sufficiently large $t$, we can infer from (5.3) that

$$\frac{d}{dt} \mathcal{R}(A_j)(t; \theta, \delta) \leq \sum_{i=1}^n \mathcal{R}(A_i)(-\beta^*_{i,j} + \epsilon_{i,j}) + \sum_{i=1}^n \int_{\mathbb{R}^+} \mathcal{R}(A^T) \gamma_{i} z e^{(\theta^* + u^*(\theta)^T) x_0} \varphi_i(dz) \kappa_{i,j},$$

$$\frac{d}{dt} \mathcal{I}(A_j)(t; \theta, \delta) \leq \sum_{i=1}^n \mathcal{I}(A_i)(-\beta^*_{i,j} + \epsilon_{i,j}) + \sum_{i=1}^n \int_{\mathbb{R}^+} \mathcal{I}(A^T) \gamma_{i} z e^{(\theta^* + u^*(\theta)^T) x_0} \varphi_i(dz) \kappa_{i,j},$$

where $\epsilon_{i,j} > 0$ is sufficiently small, $\mathcal{R}(A_j)$ and $\mathcal{I}(A_j)$ take the real and imaginary parts of $A_j$. Thus, for sufficiently large $t$, we have

$$\frac{d\mathcal{R}(A)}{dt} \leq -((\beta^*)^T + (\kappa^*)^T \mathbb{E}[\gamma Z^*])^T \mathcal{R}(A),$$

$$\frac{d\mathcal{I}(A)}{dt} \leq -((\beta^*)^T + (\kappa^*)^T \mathbb{E}[\gamma Z^*])^T \mathcal{I}(A),$$

16
where $Z^* = (Z^*_i)$ with the random variable $Z^*_i$ following the distribution $\varphi^*_i(dz)$, and we have used (5.7) and (5.9). Since the matrix $(\beta^*)^T - (\kappa^*)^T\mathbb{E}[\gamma Z^*]^T$ is positive stable (see the proof of Proposition 11 in [36]), we deduce that $A(t; \theta, u^*(\theta)) \to 0$ exponentially fast in $t$, which together with (5.2) implies that $B(t; \theta, u^*(\theta)) \to B(\infty; \theta, u^*(\theta))$ also exponentially fast in $t$. The convergence holds locally uniformly in $\theta$, and thus we have proved the desired result.

5.3.2 Infinite divisibility

In this section, we prove that for $\eta(\cdot)$ given in (2.3), we have $e^{\eta(\theta)} = \mathbb{E}[e^{\theta Y}]$ for some infinitely divisible, non-constant random variable $Y$. Recall from (2.3) that

$$\eta(\theta) = u^*(\theta)^Tb + \sum_{i=1}^{n} \lambda_i \left( \mathbb{E} \left[ e^{(\theta + u^*(\theta)^T\gamma_i)Z_i} \right] - 1 \right).$$

(5.12)

It suffices to show that we can find two independent random variables $Y_1$ and $Y_2$ such that $Y = Y_1 + Y_2$ is (non-constant) infinitely divisible and

$$\mathbb{E} \left[ e^{\theta Y_1} \right] = e^{u^*(\theta)^Tb},$$

(5.13)

$$\mathbb{E} \left[ e^{\theta Y_2} \right] = e^{\sum_{j=1}^n \lambda_j \left( \mathbb{E} \left[ e^{(\theta + u^*(\theta)^T\gamma_j)Z_j} \right] - 1 \right)}.$$  

(5.14)

We first show (5.13). Let $L_i(X(0), \lambda, a, b; t)$ and $V(X(0), \lambda, a, b; t)$ denote the process $L_i(t)$ and $V(t)$ respectively with emphasis on its dependence on $X(0)$, $\lambda$, $a$ and $b$. Note that by the definition of $u^*(\theta)$ in Equation (2.3), $u^*(\theta)$ is independent of the parameters $X(0), \lambda, a$ and $b$. Now if $a, b = 0$ and $\lambda = 0$, then it follows from Equation (5.12) that $\eta(\theta) = 0$. In addition, when $a = 0$, $b = 0$ and $\lambda = 0$, we get from the definition of $B$ in Equation (5.2) and the initial condition $B(0; \theta, \delta) = 0$ that $B(\infty; \theta, u^*(\theta)) = 0$. Then with $\theta$ in the strip $S(0, \theta_0)$, we obtain for given deterministic $X(0), \lambda, \delta, a, b$,

$$\lim_{t \to \infty} \mathbb{E} \left[ e^{\theta V(X(0),0,0,0; t)} \right] = \lim_{t \to \infty} \mathbb{E} \left[ e^{\theta V(X(0),0,0,0; t)} \right] e^{-\eta(\theta)t}$$

$$= e^{u^*(\theta)^T X(0) + B(\infty; \theta, u^*(\theta))} = e^{u^*(\theta)^T X(0)} < \infty,$$

(5.15)

where the second equality follows from the results in Section 5.3.1. Moreover, since $V(., 0, 0, 0; t)$ is monotone in $t$, we then infer from Fatou’s lemma that for real valued $\theta \in (0, \theta_0)$,

$$\mathbb{E} \left[ \lim_{t \to \infty} e^{\theta V(.,0,0,0; t)} \right] \leq \lim_{t \to \infty} \mathbb{E} \left[ e^{\theta V(.,0,0,0; t)} \right] < \infty.$$  

\footnote{In Part (II) of Assumption \[1\] we have followed [36] to assume $b_i > 0$, $1 \leq i \leq m$, so that the large deviations and central limit theorems for $V(t)$ are non-trivial. But we can indeed let $a = 0$, $b = 0$ and $\lambda = 0$ as in part (I) of Assumption \[1\] so that the parameters are admissible, see [36].}
Hence, the random variable \( V(\cdot, 0, 0, 0; \infty) := \lim_{t \to \infty} V(\cdot, 0, 0, 0; t) < \infty \) with probability one, so is \( L_i(\cdot, 0, 0, 0; \infty) := \lim_{t \to \infty} L_i(\cdot, 0, 0, 0; t), \ 1 \leq i \leq n \). In addition, we obtain that for \( \theta \) in the strip \( S_{(0, \theta_c)} \),

\[
E \left[ e^{\theta \sum_{i=1}^{n} L_i(\gamma_j \hat{Z}_j, 0, 0, 0; \infty)} \right] = e^{u^*(\theta)^T X(0)},
\]

where we have used (5.15) and the dominated convergence theorem. Now we define

\[
Y_1 = \sum_{i=1}^{n} L_i(b, 0, 0, 0; \infty), \quad \text{in distribution.}
\]

Then from Equation (5.16), we get Equation (5.13).

To establish (5.14), we define

\[
Y_2 = \sum_{j=1}^{n} Y_2^j, \quad \text{in distribution,}
\]

where \( Y_2^j \) are independent and

\[
Y_2^j = \sum_{k=1}^{N_{\lambda_j}} (Z_{j,k} + X_{j,k}).
\]

Here \( Z_{j,k} \) has the same distribution as \( Z_j \) and \( Z_{j,k} \) are i.i.d., independent of \( X_{j,k} \) where \( X_{j,k} \) are i.i.d. and distributed as \( \sum_{i=1}^{n} L_i(\gamma_j Z_j, 0, 0, 0; \infty) \) and finally \( N_{\lambda_j} \) is a Poisson random variable with mean \( \lambda_j \) and independent of \( Z_{j,k} \) and \( X_{j,k} \). Then, from Equation (5.16), we get

\[
E \left[ e^{\theta Y_2} \right] = \exp \left\{ \sum_{j=1}^{n} \lambda_j \left( E \left[ e^{\theta (Z_j + \sum_{i=1}^{n} L_i(\gamma_j \hat{Z}_j, 0, 0, 0; \infty))} \right] - 1 \right) \right\} - 1,
\]

where \( \hat{Z}_j \) is an independent copy of \( Z_j \). This verifies Equation (5.14).

Hence, we have defined a random variable \( Y = Y_1 + Y_2 \) so that \( E[e^{\theta Y}] = e^{\nu(\theta)} \). To see that \( Y \) is infinitely divisible, we can write it as i.i.d. \( m \) copies of the same random variable with parameters \( b \) replaced by \( b/m \) and \( \lambda \) replaced by \( \lambda/m \). That is, we can define
\[ Y = \sum_{\ell=1}^{m} Y_{\ell, (m)}, \text{ where } Y_{\ell, (m)} \text{ are i.i.d. with the same distribution as } Y_{1, (m)}^{(m)} + Y_{2, (m)}^{(m)}, \text{ where } Y_{1, (m)}^{(m)} \text{ and } Y_{2, (m)}^{(m)} \text{ are independent and} \]

\[ Y_{1, (m)}^{(m)} = \sum_{i=1}^{n} L_i(b/m, 0, 0; \infty), \quad \text{in distribution,} \]

and

\[ Y_{2, (m)}^{(m)} = \sum_{j=1}^{n} \sum_{k=1}^{\lambda_j/m} (Z_{j,k} + X_{j,k}), \quad \text{in distribution,} \]

so that

\[
\mathbb{E}\left[e^{\theta \sum_{\ell=1}^{m} Y_{\ell, (m)}}\right] = \left(\mathbb{E}\left[e^{\theta(Y_{1, (m)}^{(m)} + Y_{2, (m)}^{(m)})}\right]\right)^m = \left(e^{u^* (\theta)^T b + \sum_{j=1}^{n} \lambda_j m\left(\mathbb{E}\left[e^{(\theta+u^* (\theta)^T \gamma_j)Z_j}\right] - 1\right)}\right)^m = e^{\theta Y}.
\]

This completes the proof of the infinite divisibility of \( Y \).

Finally, let us show that \( Y \) is lattice distributed if and only if the random variables in the set

\[ Z := \left\{ \sum_{i=1}^{n} \sum_{j=1}^{n_i} Z_{ij}, \text{ where } n_i \in \mathbb{N} \cup \{0\} \text{ and } Z_{ij} \text{ are i.i.d. distributed as } Z_i \right\}, \]

are supported on \( b\mathbb{N} \cup \{0\} \) for some \( b > 0 \). First, we prove the ‘if’ part. If the random variables in \( Z \) are supported on \( b\mathbb{N} \cup \{0\} \) for some \( b > 0 \), then so is each \( Z_{ij} \) and by the definition of \( Y \) above, \( Y \) is lattice distributed. Next, we prove the ‘only if’ part. If \( Y \) is lattice distributed, then \( Y \) is supported on \( a + b\mathbb{Z} \) for some \( a \in \mathbb{R} \) and \( b > 0 \). By the definition of \( Y \) above, we infer that each random variable in \( Z \) is also supported on \( a + b\mathbb{Z} \). Therefore, \( Z_{ij} \) is also supported on \( a + b\mathbb{Z} \). If \( Z_{ij} = a + b z_{ij} \) for some \( z_{ij} \in \mathbb{Z} \), then \( Z_{i1} + Z_{i2} = 2a + b(z_{i1} + z_{i2}) \) also belongs to \( a + b\mathbb{Z} \), which implies that \( a \in b\mathbb{Z} \), and hence we can take \( a = 0 \). Finally, if each \( Z_{ij} \) is supported on \( b\mathbb{Z} \), then so are the random variables in the set \( Z \).

### 6 Applications and numerical experiments

Our theoretical results can be potentially useful in applications where affine point processes are relevant models to capture the clustering or self- and mutually-exciting effects and one is interested in large deviation probabilities or tail expectations. In this section, we illustrate the application of our results (Theorem 4) to credit risk analysis and present numerical studies.
6.1 Financial application: credit risk analysis

We consider credit risk assessment of synthetic CDO (Collateralized Debt Obligations) tranches. See e.g. [13, 15, 20] for background. Following [20], we consider a stylized synthetic CDO which is made up of a portfolio of \( n \) single-name credit default swaps with notional 1 and maturity \( T \) that usually ranges from six months to ten years. The constituent credit default swaps are referenced on bonds issued by rated firms. When a bond issuer defaults, the portfolio incurs a loss associated with the default event.

The risk of loss on the reference portfolio is divided into tranches of increasing seniority. Losses will first affect the “equity” or “first loss” tranche, next the “mezzanine” tranche(s), finally the “senior” and then the “super-senior” tranches. Specifically, each tranche is specified by an attachment point \( K_A \) and a detachment point \( K_D > K_A \) that determine the amount of portfolio loss, and correspondingly the number of defaults, the tranche can absorb. Mathematically, write \( L(t) \) for the cumulative portfolio loss up to time \( t \), then the tranche loss \( U(t) \) is given by

\[
U(t) = (L(t) - K_A)^+ - (L(t) - K_D)^+.
\]

That is, the tranche starts to suffer losses when the cumulative portfolio loss \( L(\cdot) \) exceeds the attachment point \( K_A \), and the tranche would be wiped out if the loss exceeds \( K_D \). In general, senior and super-senior tranches may have a low probability to suffer losses. See e.g. [13, 15, 20].

As each tranche has a different degree of credit exposure, it has its own credit rating at issuance. Standard and Poor’s ratings are based on the tranche (real-world) default probability given by \( \mathbb{P}(L(T) > K_A) \), and Moody’s ratings are based on the (real-world) expected tranche loss \( \mathbb{E}(U(T)) \) at maturity \( T \). These quantities are important as investors often rely on the credit ratings of tranches for investment and risk management purposes in structured finance markets.

To compute the tranche default probability or the expected loss, one needs a model for the distribution of the portfolio loss \( L(T) \). There are several different classes of models in the literature. We follow the approach of [2, 20] and consider using an one-dimensional affine point process to directly model the aggregate portfolio loss \( L(\cdot) \) to capture the empirical feature of default contagion, that is, the default of one firm can trigger the default of other related firms. Specifically, as an illustrative example, we assume that under the real-world probability measure, the default events arrive according to the intensity \( \lambda(\cdot) \) given by

\[
\lambda(t) = 1 + 0.5\gamma \cdot \sum_{0<T_i\leq t} \exp(-(t - \tau_i)),
\]

where \( \gamma \geq 0 \) and \( \tau_i \) is the random default timing of \( i \)-th event. Equivalently,

\[
d\lambda(t) = (1 - \lambda(t))dt + \gamma dL(t), \quad \text{where} \quad \lambda(0) = 1 \quad \text{and} \quad L(t) = 0.5N(t).
\]

\(^7\)See, e.g., [22], for large deviations analysis of the loss distribution in the Gaussian copula model of portfolio credit risk.
Here \( N(t) \) counts the total number of defaults in the portfolio up to time \( t \), and each default generates a fixed loss 50%. In this model, the initial default rate is one default on average per year. When defaults occur in the portfolio, the default intensity \( \lambda(t) \) jumps up by \( 0.5\gamma \) at each default event, and then decays with speed one. The parameter \( \gamma \) indicates the level of default clustering as a large \( \gamma \) leads to significant default contagion. To address concerns about possible model or parameter misspecification, we will consider three different cases for \( \gamma \) (with \( \gamma = 0.04, 0.2 \) and 1) in the numerical experiments.

We focus on computing \( P(L(T) \geq K) \) and \( E[(L(T) - K)^+] \) for this one-dimensional affine point process \( L \), where \( K = xT \). These two quantities are approximately the default probability and the expected loss for a super-senior tranche with attachment point \( K \), for a large well-diversified portfolio.

We now present a set of representative numerical results in Tables 1-3 for common maturity points \( T = 1, 3, 5, 7, 10 \) years. The 20- and 30-year results are also reported for illustrations. We compare the results from first order approximations and those from Monte Carlo simulations. The first order approximations refer to truncating the asymptotic series in (3.1) and (3.2) to retain the first two terms. When computing the expansion coefficients, we need to numerically solve ODEs (see e.g. Theorem 13) and we use Runge-Kutta methods. We do not report numerical results from zeroth order approximations as the errors are not small. For the simulations, as plain Monte Carlo is very slow in computing small probabilities and tail expectations, we use the importance sampling algorithm in [21] with sample size 500k. The computations are performed using Python on a Mac computer with Processor 1.4 GHz Intel Core i5.

We find with weak or intermediate level of default clustering (with \( \gamma = 0.04 \) in Table 1 or \( \gamma = 0.2 \) in Table 2), our first order approximations for the default probability and the expected loss can be quite accurate for large maturity \( T \), e.g. \( T = 10 \) years. These analytical approximations are much faster to evaluate than Monte Carlo simulation via importance sampling, and they perform numerically stable. Hence, these approximations can provide alternatives in these scenarios to simulation which is computer resource extensive.

On the other hand, with significant default clustering (\( \gamma = 1 \)), the first order approximations do not perform well for realistic maturities \( T \leq 10 \) years, see Table 3. The intuitive reason is that significant default clustering shift the probability mass of the loss \( L(T) \) to the right tail, so the default probability is not very small (compared with the model with a low \( \gamma \) for a fixed \( K = xT \)) and it is not in the large deviation regime. In this case, in order for the first order approximation to produce reasonably accurate results, we need to increase the maturity \( T \) to, say, 200-years which is not realistic.

A natural question is whether higher order approximations can produce better results, say, for \( \gamma = 1 \). To address this question, we present in Table 4 the numerical results using second order approximations, i.e. we retain the first three terms in the asymptotic

\[8\] We remark that the loss process \( L(t) \) and default counting process \( N(t) \) are not bounded from above with this model. One can cap them based on the portfolio size \( n \) if necessary. For a large diversified portfolio, this effect is minimal in our approximations. See [16] for a related discussion.

21
expansions in (3.1) and (3.2). As one can read from this table and Table 3, second order approximations do not lead to more accurate results than first order approximations. This is not surprising. An asymptotic series is usually a non-convergent series. There is no general rule for selecting the “right” order of the truncation in developing approximations, and adding additional higher order terms does not necessarily improve the approximation (see e.g. [17]).

| T   | P(IS) [95% CI] | P(1-Order) [R.E.] | E(IS) [95% CI] | E(1-Order) [R.E.] |
|-----|----------------|------------------|----------------|------------------|
| 1Y  | 8.31E-02 ±0.36% | 6.05E-02 [-27.2%] | 2.53E-02 ±0.40% | -2.79E-02 [-210.1%] |
| 3Y  | 4.76E-03 ±0.42% | 4.52E-03 [-5.04%] | 2.05E-03 ±0.49% | 1.48E-03 [-27.8%] |
| 5Y  | 3.56E-04 ±0.56% | 3.48E-04 [-2.25%] | 1.68E-04 ±0.60% | 1.51E-04 [-10.1%] |
| 7Y  | 2.87E-05 ±0.70% | 2.83E-05 [-1.39%] | 1.43E-05 ±0.70% | 1.35E-05 [-5.59%] |
| 10Y | 7.07E-07 ±0.57% | 6.97E-07 [-1.41%] | 3.65E-07 ±0.55% | 3.55E-07 [-2.74%] |
| 20Y | 3.74E-12 ±0.80% | 3.73E-12 [-0.27%] | 2.04E-12 ±0.98% | 2.03E-12 [-0.49%] |
| 30Y | 2.27E-17 ±0.88% | 2.27E-17 [+0.00%] | 1.25E-17 ±0.80% | 1.26E-17 [+0.80%] |

Table 1: First order approximations are compared with simulations using importance sampling (IS) when the clustering parameter $\gamma = 0.04$ in (6.2). $P$ stands for $P(L(T) \geq xT)$ and $E$ stands for $E[(L(T) - xT)^+]$ with fixed $x = 3$. CI stands for the confidence interval and R.E. stands for the relative errors of approximations compared with simulation results.

| T   | P(IS) [95% CI] | P(1-Order) [R.E.] | E(IS) [95% CI] | E(1-Order) [R.E.] |
|-----|----------------|------------------|----------------|------------------|
| 1Y  | 9.39E-02 ±0.32% | 1.55E-02 [-83.5%] | 3.43E-02 ±0.29% | -1.69E-01 [-593.1%] |
| 3Y  | 1.05E-02 ±0.95% | 8.76E-03 [-16.6%] | 6.24E-03 ±0.48% | 1.54E-03 [-75.3%] |
| 5Y  | 1.66E-03 ±0.60% | 1.56E-03 [-6.02%] | 1.17E-03 ±0.85% | 8.77E-04 [-25.0%] |
| 7Y  | 2.86E-04 ±0.70% | 2.77E-04 [-3.15%] | 2.19E-04 ±0.46% | 1.93E-04 [-11.9%] |
| 10Y | 2.17E-05 ±0.92% | 2.14E-05 [-1.38%] | 1.77E-05 ±0.56% | 1.68E-05 [-5.08%] |
| 20Y | 4.89E-09 ±0.82% | 4.86E-09 [-0.61%] | 4.34E-09 ±0.69% | 4.31E-09 [-0.69%] |
| 30Y | 1.52E-12 ±1.60% | 1.24E-12 [-0.80%] | 1.14E-12 ±0.88% | 1.13E-12 [-0.88%] |

Table 2: First order approximations are compared with simulations using importance sampling (IS) when the clustering parameter $\gamma = 0.2$ in (6.2). $P$ stands for $P(L(T) \geq xT)$ and $E$ stands for $E[(L(T) - xT)^+]$ with fixed $x = 3$. CI stands for the confidence interval and R.E. stands for the relative errors of approximations compared with simulation results.

**Remark.** The intensity model (6.1) is highly stylized for the purpose of illustrating the application of our theoretical results. It does not take into account that firms are exposed to observable and latent systematic economic factors as in [2]. However, it captures default contagion which is found to be a significant source for default clustering. In (6.1), before any firm defaults, the “baseline intensity” of default arrivals is constant one per year. In practice, this baseline intensity may not be constant, especially over a long horizon. Let us briefly discuss the extension to the non-homogeneous (deterministic) baseline intensity. If we have a non-constant deterministic baseline intensity $\lambda(t)$, i.e., $\lambda(t) = v(t) + 0.5\gamma \sum_{0<T_i \leq t} \exp(-(t - \tau_i))$, then the affine structure of the point process
Table 3: First order approximations are compared with simulations using importance sampling (IS) when the clustering parameter $\gamma = 1$ in (6.2). $P$ stands for $P(L(T) \geq xT)$ and $E$ stands for $E[(L(T) - xT)^+]$ with fixed $x = 3$. CI stands for the confidence interval and R.E. stands for the relative errors of approximations compared with simulation results.

| $T$ | $P(1)$ [95% CI] | $P(1)$ [R.E.] | $E(1)$ [95% CI] | $E(1)$ [R.E.] |
|-----|----------------|--------------|----------------|--------------|
| 1Y  | 1.49E-01 ±0.67% | -1.24E+01 [-8395%] | 1.07E-01 ±0.93% | -3.72E+02 [-347764%] |
| 3Y  | 1.03E-01 ±0.97% | -1.87E+00 [-1915%] | 1.97E-01 ±1.02% | -6.18E+01 [-31471%] |
| 5Y  | 9.56E-02 ±0.52% | -6.68E-01 [-799%] | 2.82E-01 ±0.71% | -2.47E+01 [-8859%] |
| 7Y  | 8.97E-02 ±0.56% | -2.99E-01 [-433%] | 3.37E-01 ±0.59% | -1.28E+01 [-3898%] |
| 10Y | 7.96E-02 ±0.50% | -6.68E-01 [-799%] | 2.82E-01 ±0.71% | -2.47E+01 [-8859%] |
| 20Y | 4.72E-02 ±0.64% | 1.63E-02 [-65.5%] | 3.07E-01 ±0.65% | -8.57E-01 [-379%] |
| 30Y | 7.53E-02 ±0.73% | 1.86E-02 [-32.4%] | 2.05E-01 ±0.49% | -1.39E-01 [-168%] |
| 50Y | 9.37E-03 ±0.53% | 8.14E-03 [-31.3%] | 7.96E-02 ±0.50% | 3.00E-02 [-62.3%] |
| 100Y| 7.03E-04 ±0.57% | 6.76E-04 [-3.84%] | 6.78E-03 ±0.44% | 5.69E-03 [-16.1%] |
| 200Y| 4.80E-06 ±0.63% | 4.77E-06 [-0.62%] | 5.04E-05 ±0.60% | 4.85E-05 [-3.77%] |
| 500Y| 2.40E-12 ±0.83% | 2.41E-12 [+0.42%] | 2.69E-11 ±0.74% | 2.67E-11 [-0.74%] |

Table 4: Second order approximations results and relative errors compared with simulations using importance sampling (IS) when the clustering parameter $\gamma = 1$ in (6.2). $P$ stands for $P(L(T) \geq xT)$ and $E$ stands for $E[(L(T) - xT)^+]$ with fixed $x = 3$. R.E. stands for the relative errors of approximations compared with simulation results.

| $T$ | $P(2)$ [R.E.] | $E(2)$ [R.E.] |
|-----|--------------|--------------|
| 1Y  | 7.78E+03 [5220638.26%] | -3.72E+02 [99739152.34%] |
| 3Y  | 4.53E+02 [439802.91%] | -6.18E+01 [3143047.21%] |
| 5Y  | 1.15E+02 [119983.68%] | -2.47E+01 [553800.71%] |
| 7Y  | 4.50E+01 [50067.22%] | -1.28E+01 [180908.99%] |
| 10Y | 1.00E+01 [20000.50%] | -5.90E+00 [57654.01%] |
| 20Y | 1.80E+00 [3713.56%] | -8.57E-01 [7619.87%] |
| 30Y | 4.23E-01 [1438.18%] | -1.39E-01 [2548.78%] |
| 50Y | 5.22E-02 [457.10%] | 3.00E-02 [700.25%] |
| 100Y| 1.42E-03 [101.99%] | 5.69E-03 [334.51%] |
| 200Y| 5.97E-06 [24.38%] | 4.85E-05 [29.17%] |
| 500Y| 2.50E-12 [4.17%] | 2.67E-11 [4.09%] |
model is still preserved. When \( v(t) \) converges to a constant sufficiently fast as \( t \to \infty \), then the results we derive in this paper will still hold. On the other hand, if \( v(t) \) converges very slowly to a constant or when \( v(t) \) is not asymptotically constant as \( t \to \infty \), then we may not have the desired convergence speed in the mod-\( \phi \) convergence, and the mod-\( \phi \) convergence method will no longer be applicable in these cases, and it is out of the scope of the present paper to study these cases.

6.2 Two additional test cases

We also illustrate the performance of our approximations (from Theorem 4) for two test cases on three-dimensional affine point processes where the parameters are taken from [36]. Both lattice and non-lattice cases will be studied.

The model specification of the three-dimensional affine point process \((L_1, L_2, L_3)\) is as follows. The components of \(X = (X_1, X_2, X_3)\) satisfy

\[
dX_j(t) = (b_j - \beta_j X_j(t))dt + \sigma_j \sqrt{X_j(t)}dW_j(t) + \sum_{i=1}^{3} \gamma_i dL_i(t), \quad j = 1, 2, 3,
\]

where \(b_j, \beta_j, \sigma_j, \gamma_i > 0\). The jump intensity is \(\Lambda_i(X(t)) = \lambda_i + \kappa_i X_i(t)\) for some \(\lambda_i, \kappa_i > 0\). In both test cases below, \(\beta = (\beta_1, \beta_2, \beta_3) = (2, 2.1, 2.2), b = (6, 6.1, 6.2), \sigma = (0.5, 0.6, 0.7), \gamma = (0.2, 0.3, 0.4), \kappa = (1, 1.1, 1.2), \lambda = (0, 0, 0)\). The two test cases differ only in the distribution of \(Z_i\). In the first test case, \(Z_i\) is assumed to be constant one for each \(i\); In the second test case 2, \(Z_i\) is assumed to follow an exponential distribution with mean one.

In Tables 5 and 6 we report the results of the first order approximations for \(\mathbb{P}(V(t) \geq xt)\) and \(\mathbb{E}[(V(t) - xt)^+]\) for different values of \(x\) and \(t\) for the two test cases, where \(V(t) = \sum_{i=1}^{3} L_i(t)\). We can observe that the relative errors for the first order approximations compared with Monte Carlo simulations (via importance sampling) become small when \(t\) becomes large.

Acknowledgements

We thank Jim Dai, Jayaram Sethuraman, S. R. S. Varadhan, Xunyu Zhou for helpful comments and special thank goes to Shujian Liao for his assistance on numerical experiments. Xuefeng Gao acknowledges support from Hong Kong RGC Grants 24207015 and 14201117. Lingjiong Zhu is grateful to the support from NSF Grant DMS-1613164.
Table 6: First order approximations are compared with simulations using importance sampling (IS) when \( Z_i = 1 \) for \( i = 1, 2, 3 \). \( P \) stands for \( P(V(t) \geq x) \) and \( E \) stands for \( E[(V(t) - xt)^+] \) with \( V(t) = \sum_{i=1}^{3} L_i(t) \). CI stands for the confidence interval and R.E. stands for the relative errors of approximations compared with simulation results.

| \( t \) | \( P(\text{IS}) [95\% \text{ CL.}] \) | \( P(\text{1-Order}) \) [R.E.] | \( E(\text{IS}) [95\% \text{ CL.}] \) | \( E(\text{1-Order}) \) [R.E.] |
|---|---|---|---|---|
| **\( x=25 \)** | | | | |
| 5 | 1.85E-02 [±1.08\%] | -5.48E-03 [−130\%] | 1.78E-01 [±1.12\%] | -6.12E-01 [−444\%] |
| 10 | 4.62E-03 [±1.30\%] | 2.04E-03 [−55.8\%] | 5.30E-02 [±1.32\%] | -2.68E-02 [−151\%] |
| 20 | 3.55E-04 [±1.69\%] | 2.59E-04 [−27.0\%] | 4.59E-03 [±1.31\%] | 1.99E-03 [−56.6\%] |
| 30 | 2.93E-05 [±2.39\%] | 2.42E-05 [−17.4\%] | 4.09E-04 [±1.96\%] | 2.68E-04 [−34.5\%] |
| 50 | 2.32E-07 [±2.59\%] | 2.06E-07 [−11.2\%] | 3.38E-06 [±2.07\%] | 2.72E-06 [−19.5\%] |
| 100 | 1.72E-12 [±2.91\%] | 1.55E-12 [−9.88\%] | 2.52E-11 [±2.38\%] | 2.25E-11 [−10.7\%] |
| 200 | 1.13E-22 [±3.54\%] | 1.12E-22 [−0.88\%] | 1.75E-21 [±2.86\%] | 1.69E-21 [−3.43\%] |
| **\( x=30 \)** | | | | |
| 5 | 7.52E-04 [±1.86\%] | 1.66E-04 [−77.9\%] | 5.95E-03 [±1.40\%] | -4.96E-03 [−183\%] |
| 10 | 2.69E-05 [±1.86\%] | 1.75E-05 [−35.3\%] | 2.53E-04 [±1.28\%] | 7.49E-05 [−70.4\%] |
| 20 | 4.49E-08 [±3.25\%] | 3.96E-08 [−17.8\%] | 4.38E-07 [±1.90\%] | 3.10E-07 [−29.2\%] |
| 30 | 8.21E-11 [±2.92\%] | 7.33E-11 [−10.7\%] | 8.47E-10 [±2.01\%] | 6.85E-10 [−19.1\%] |
| 50 | 3.31E-16 [±3.32\%] | 3.04E-16 [−8.16\%] | 3.45E-15 [±2.17\%] | 3.04E-15 [−11.9\%] |
| 100 | 1.38E-29 [±3.62\%] | 1.28E-29 [−7.25\%] | 1.47E-28 [±2.99\%] | 1.33E-28 [−9.52\%] |
| 200 | 3.03E-56 [±3.96\%] | 2.98E-56 [−1.65\%] | 3.18E-55 [±3.26\%] | 3.17E-55 [−0.31\%] |

Table 5: First order approximations are compared with simulations using importance sampling (IS) when \( Z_i = 1 \) for \( i = 1, 2, 3 \). \( P \) stands for \( P(V(t) \geq x) \) and \( E \) stands for \( E[(V(t) - xt)^+] \) with \( V(t) = \sum_{i=1}^{3} L_i(t) \). CI stands for the confidence interval and R.E. stands for the relative errors of approximations compared with simulation results.

| \( t \) | \( P(\text{IS}) [95\% \text{ CL.}] \) | \( P(\text{1-Order}) \) [R.E.] | \( E(\text{IS}) [95\% \text{ CL.}] \) | \( E(\text{1-Order}) \) [R.E.] |
|---|---|---|---|---|
| **\( x=25 \)** | | | | |
| 5 | 6.22E-02 [±1.29\%] | -9.06E-02 [−246\%] | 1.01E+00 [±0.99\%] | -9.60E+00 [−1051\%] |
| 10 | 2.83E-02 [±1.41\%] | 8.51E-04 [−96.9\%] | 5.79E-01 [±1.04\%] | -1.28E+00 [−321\%] |
| 20 | 6.54E-03 [±1.53\%] | 3.94E-03 [−39.8\%] | 1.53E-01 [±1.31\%] | -1.54E-02 [−110\%] |
| 30 | 1.65E-03 [±1.82\%] | 1.26E-03 [−23.6\%] | 4.29E-02 [±1.40\%] | 1.68E-02 [−60.8\%] |
| 50 | 1.19E-04 [±1.68\%] | 1.03E-04 [−13.5\%] | 3.17E-03 [±1.58\%] | 2.27E-03 [−28.4\%] |
| 100 | 2.00E-07 [±2.50\%] | 1.89E-07 [−5.50\%] | 5.62E-06 [±1.60\%] | 5.11E-06 [−9.07\%] |
| 200 | 7.89E-13 [±2.79\%] | 7.57E-13 [−4.06\%] | 2.36E-11 [±1.69\%] | 2.20E-11 [−6.78\%] |
| **\( x=30 \)** | | | | |
| 5 | 1.04E-02 [±1.92\%] | -1.83E-03 [−118\%] | 1.54E-01 [±1.30\%] | -3.37E-01 [−319\%] |
| 10 | 1.68E-03 [±1.79\%] | 8.17E-04 [−51.4\%] | 2.92E-02 [±1.37\%] | -3.47E-03 [−112\%] |
| 20 | 5.17E-05 [±2.13\%] | 3.97E-05 [−23.2\%] | 9.93E-04 [±1.61\%] | 5.58E-04 [−43.8\%] |
| 30 | 1.78E-06 [±2.25\%] | 1.52E-06 [−14.6\%] | 3.47E-05 [±1.73\%] | 2.63E-05 [−24.2\%] |
| 50 | 2.42E-09 [±3.31\%] | 2.21E-09 [−8.68\%] | 5.02E-08 [±2.19\%] | 4.28E-08 [−14.7\%] |
| 100 | 2.17E-16 [±3.69\%] | 2.02E-16 [−6.91\%] | 4.63E-15 [±2.38\%] | 4.20E-15 [−9.29\%] |
| 200 | 2.27E-30 [±3.96\%] | 2.18E-30 [−3.96\%] | 4.91E-29 [±2.65\%] | 4.66E-29 [−5.09\%] |
References

[1] Aït-Sahalia, Y., Cacho-Díaz, J. and Laeven, R.J., 2015. Modeling financial contagion using mutually exciting jump processes. Journal of Financial Economics, 117(3), pp.585-606.

[2] Azizpour, S., Giesecke, K., and Schwenkler, G. 2018. Exploring the sources of default clustering. Journal of Financial Economics. 129, 154-183.

[3] Bahadur, R.R. and Rao, R.R., 1960. On deviations of the sample mean. Ann. Math. Statist, 31(4), pp.1015-1027.

[4] Bercu, B. and A. Rouault, 2002. Sharp large deviations for the Ornstein-Uhlenbeck process. SIAM Theory of Probability and Its Applications. 46, 1-19.

[5] Bordenave, C., and Torrisi, G. L., 2007. Large deviations of Poisson cluster processes. Stochastic Models, 23(4), 593-625.

[6] Bowsher, C.G., 2007. Modelling security market events in continuous time: Intensity based, multivariate point process models. Journal of Econometrics, 141(2), pp.876-912.

[7] Chaganty, N. R. and J. Sethuraman, 1993. Strong large deviation and local limit theorems. Annals of Probability. 21, 1671-1690.

[8] Constantine, G. M. and T. H. Savits, 1996. A multivariate Faa di Bruno formula with applications. Transactions of the American Mathematical Society. 348. 503-520.

[9] Cramér, H., 1938. Sur un nouveau théoreme-limite de la théorie des probabilités. Actualités scientifiques et industrielles, 736(5-23), p.115.

[10] Daw, A. and Pender, J., 2018. Queues driven by Hawkes processes. Stochastic Systems, 8(3), pp.192-229.

[11] Delbaen, F., Kowalski, E. and Nikeghbali, A., 2014. Mod-ϕ convergence. International Mathematics Research Notices, 2015(11), pp.3445-3485.

[12] Dembo, A. and O. Zeitouni., 1998. Large Deviations Techniques and Applications. 2nd Edition, Springer, New York.

[13] Duffie, D., and Garleanu, N, 2001. Risk and valuation of collateralized debt obligations. Financial Analysts Journal, 57(1), 41-59.

[14] Duffie, D., Pan, J. and Singleton, K., 2000. Transform analysis and asset pricing for affine jump–diffusions. Econometrica, 68(6), pp.1343-1376.
[15] Duffie, D. and Singleton, K.J., 2012. *Credit Risk: Pricing, Measurement, and Management*. Princeton University Press.

[16] Errais, E., K. Giesecke, and Goldberg, L., 2010. Affine point processes and portfolio credit risk, *SIAM J. Financial Math.* 1, 642-665.

[17] Field, C. A., and Ronchetti, E., 1990. *Small Sample Asymptotics*. Hayward, CA: Institute of Mathematical Statistics.

[18] Fox, E.W., Short, M.B., Schoenberg, F.P., Coronges, K.D. and Bertozzi, A.L., 2016. Modeling e-mail networks and inferring leadership using self-exciting point processes. *Journal of the American Statistical Association*, 111(514), pp.564-584.

[19] Gao, X. and Zhu, L., 2018. A functional central limit theorem for stationary Hawkes processes and its application to infinite-server queues. *Queueing Systems*. 90, 161-206.

[20] Giesecke, K., and Kim, B., 2011. Risk analysis of collateralized debt obligations. *Operations Research*, 59(1), 32-49.

[21] Giesecke K, Kim B and Zhu S., 2011. Monte Carlo algorithms for default timing problems. *Management Science*. 57(12), 2115-2129.

[22] Glasserman, P., Kang, W. and Shahabuddin, P., 2007. Large deviations in multifactor portfolio credit risk. *Mathematical Finance*, 17(3), pp.345-379.

[23] Harrington, D. P., 1978. Almost sure convergence of continuous time branching processes. *Stochastic Processes and their Applications*, 6(3), 317-322.

[24] Hawkes, A. G., 1971. Spectra of some self-exciting and mutually exciting point processes, *Biometrika*. 58, 83-90.

[25] Hewlett. P., 2006. Clustering of order arrivals, price impact and trade path optimization, in: Workshop on Financial Modeling with Jump Processes, Ecole Polytechnique, pp. 6-8.

[26] Hult, H., and Samorodnitsky, G., 2010. Large deviations for point processes based on stationary sequences with heavy tails. *Journal of Applied Probability*. 47(1), 1-40.

[27] Jacod, J., Kowalski, E. and Nikeghbali, A., 2011. Mod-Gaussian convergence: new limit theorems in probability and number theory. In: *Forum Mathematicum* (Vol. 23, No. 4, pp. 835-873).

[28] Joutard, C., 2013. Strong large deviations for arbitrary sequences of random variables. *Ann. Inst. Stat. Math.* 65, 49-67.
[29] Karatzas, I., and Shreve, S., 2012. *Brownian Motion and Stochastic Calculus.* (Vol. 113). Springer New York.

[30] Koops, D. T., Mayank Saxena, O. J. Boxma, and Michel Mandjes., 2018. ”Infinite-server queues with Hawkes input.” *Journal of Applied Probability.* 55, no. 3: 920-943.

[31] Feray, V., Meliot, P.-L., and A. Nikeghbali., 2013. Mod-$\phi$ convergence I: Normality zones and precise deviations. *arXiv:1304.2934*.

[32] Olver, F., 1997. *Asymptotics and Special Functions.* 2nd Edition, CRC Press.

[33] Petrov, V.V., 1975. *Sums of Independent Random Variables.* Springer-Verlag. Berlin.

[34] Pinto, J.C.L., Chahed, T. and Altman, E., 2015. Trend detection in social networks using Hawkes processes. *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining.* pp. 1441-1448. ACM.

[35] Varadhan, S. R. S., 1966. Asymptotic probability and differential equations. *Comm. Pure Appl. Math.* 19, 261-286.

[36] Zhang, X., Blanchet, J., Giesecke, K., and P. W. Glynn., 2015. Affine point processes: Approximation and efficient simulation. *Mathematics of Operations Research.* 40(4), 797-819.

[37] Zhu, L., 2015. Large deviations for Markovian nonlinear Hawkes Processes. *Annals of Applied Probability.* 25, 548-581.
A    Proofs of Propositions 8 and 11

This appendix collects the proofs of Propositions 8 and 11.

A.1 Proof of Proposition 8

Proof. We follow the proof of Theorem 3.4. (Theorem 7 in our paper) and Remark 3.7 in [31]. Before we proceed, since we assume that $|g(x)| \leq \tilde{a}e^{\tilde{h}x}$ for some $\tilde{h} < h$ and $\tilde{a} > 0$, we have

$$
E[g(X_n - t_n x)1_{X_n \geq t_n x}] \leq \tilde{a}E[e^{\tilde{h}X_n}] < \infty,
$$

since $\tilde{h} < h \leq d$, where we recall that $E[e^{xX_n}]$ exists on $S(c,d)$. Moreover, since we assume that $|g(x)| \leq \tilde{a}e^{\tilde{h}x}$ for some $\tilde{h} < h$ and $\tilde{a} > 0$, we can check that $\hat{c}_k$ whose definition involves the summation of $g(q)e^{-q(h)}(-q)^m$ over $q$ from 0 to $\infty$ is well-defined.

Next, let us define:

$$
H_n(w) := \sum_{k=0}^{\infty} g(k)e^{-kh}e^{-k\frac{iw}{\sqrt{t_n \eta''(h)}}} \cdot \psi \left( h + \frac{iw}{\sqrt{t_n \eta''(h)}} \right) \cdot e^{tn(\psi(h) - \eta(h)) + \frac{w^2}{2t_n}}
$$

Since we assume that $|g(x)| \leq \tilde{a}e^{\tilde{h}x}$ for some $\tilde{h} < h$ and $\tilde{a} > 0$, $H_n(w)$ is well-defined. By Taylor expansion, we define $\beta_j(w)$ via the expansion:

$$
H_n(w) = \sum_{k=0}^{2v-1} \beta_k(w) + O(t_n^{-v}).
$$

Following the proof of Theorem 3.4. and Remark 3.7. in [31], we can readily obtain

$$
E[g(X_n - t_n x)1_{X_n \geq t_n x}] = \frac{e^{-t_n F(x)}}{\sqrt{2\pi t_n \eta''(h)}} \left( \hat{c}_0 + \frac{\hat{c}_1}{t_n} + \ldots + \frac{\hat{c}_{v-1}}{t_n^{v-1}} + O \left( \frac{1}{t_n^v} \right) \right),
$$

where

$$
\hat{c}_k = \int_{\mathbb{R}} \beta_k(w) \frac{e^{-w^2/2}}{\sqrt{2\pi}} dw. \quad (A.2)
$$

We next compute the coefficients $\hat{c}_k$. As $\beta_j(w)$ is defined via the expansion of $g_n(w)$, we proceed to expand the three terms in (A.1).
First, we expand the term $\sum_{k=0}^{\infty} g(k) e^{-kh} e^{-k \frac{iw}{\sqrt{t_n \eta''(h)}}}$. We can compute that

$\sum_{k=0}^{\infty} g(k) e^{-kh} e^{-\frac{k}{\sqrt{t_n \eta''(h)}}} = \sum_{k=0}^{\infty} g(k) e^{-kh} \sum_{m=0}^{\infty} \left(\frac{-k}{m!}\right)^m \left(\frac{iw}{\sqrt{t_n \eta''(h)}}\right)^m$

$= \sum_{m=0}^{\infty} \left[ \sum_{q=0}^{\infty} g(q) e^{-qh} \left(\frac{-q}{m!}\right)^m \left(\frac{iw}{\sqrt{t_n \eta''(h)}}\right)^m \right].$

Next, as $\psi$ is analytic, we expand the term $\psi\left(h + \frac{iw}{\sqrt{t_n \eta''(h)}}\right)$ as

$\psi\left(h + \frac{iw}{\sqrt{t_n \eta''(h)}}\right) = \sum_{\ell=0}^{\infty} \psi^{(\ell)}(h) \left(\frac{iw}{\sqrt{t_n \eta''(h)}}\right)^\ell.$

Finally, let us expand the term $e^{-\frac{w^2}{\eta''(h)} \sum_{\ell=1}^{\infty} \eta^{(\ell+2)}(h) \left(\frac{iw}{\sqrt{t_n \eta''(h)}}\right)^\ell}$. Let us define $f(x) = e^{-x}$ and

$\tilde{f}(x) = \frac{w^2}{\eta''(h)} \sum_{k=1}^{\infty} \eta^{(k+2)}(h) k!x^k.$

Then

$f\left(\tilde{f}(x)\right) = \sum_{n=0}^{\infty} \frac{d^n}{dx^n} f\left(\tilde{f}(0)\right) \frac{x^n}{n!},$

where we can compute that $\tilde{f}(0) = 0$ and for every $j \in \mathbb{N},$

$\tilde{f}^{(j)}(0) = \frac{w^2}{\eta''(h)} \frac{\eta^{(j+2)}(h)}{(j+2)(j+1)},$

which implies that by Faà di Bruno’s formula (see Appendix C.2)

$f\left(\tilde{f}(x)\right) = \sum_{n=0}^{\infty} \sum_{S_n} n! \frac{(-1)^{m_1+\cdots+m_n}}{m_1!m_2!^2\cdots m_n!n!^m} \prod_{j=1}^{n} \left(\frac{w^2}{\eta''(h)} \frac{\eta^{(j+2)}(h)}{(j+2)(j+1)}\right)^{m_j} \frac{x^n}{n!}.$

Hence,

$e^{-\frac{w^2}{\eta''(h)} \sum_{\ell=1}^{\infty} \eta^{(\ell+2)}(h) \left(\frac{iw}{\sqrt{t_n \eta''(h)}}\right)^\ell}$

$= \sum_{n=0}^{\infty} \sum_{S_n} \frac{(-1)^{m_1+\cdots+m_n}}{m_1!m_2!^2\cdots m_n!n!^m} \cdot \prod_{j=1}^{n} \left(\frac{w^2}{\eta''(h)} \frac{\eta^{(j+2)}(h)}{(j+2)(j+1)}\right)^{m_j} \left(\frac{iw}{\sqrt{t_n \eta''(h)}}\right)^n.$
Hence, by (1), (2) and (3), we conclude that

\[ \beta_k(w) = \sum_{m+\ell+n=k} \sum_{q=0}^{\infty} g(q)e^{-q\beta} \frac{(-q)^m}{m!} \cdot \frac{\psi^{(\ell)}(h)}{\ell!} \]

\[ \cdot \sum_{S_n} \frac{(-1)^{m_1+\ldots+m_n}}{m_1!m_2!m_2! \ldots m_n!n^n} \cdot \prod_{j=1}^{n} \left( \frac{1}{\eta''(h)} \frac{(j+2)(h)}{(j+2)(j+1)} \right) \cdot \frac{(i)^{k}w^{k+2(m_1+\ldots+m_n)}}{(\eta''(h))^{k/2}}. \]

By the property of standard normal random variable, we have \( \int_{-\infty}^{\infty} w^m e^{-\frac{w^2}{2}} dw = (m-1)!! \) if \( m \) is even, and 0 if \( m \) is odd. Hence, we obtain from (A.2) the formula of \( \hat{d}_k \) given in (1.1). The proof is therefore complete.

**A.2 Proof of Proposition 11**

To prove Proposition 11 it suffices to prove Part (ii). Before we proceed to the proof, let us first show that \( E[g(X_n - t_n x)1_{X_n \geq t_n x}] \) and \( \beta_k \) are well-defined and finite. Since we assume that \( g(x) = \sum_{k=0}^{\infty} g_k x^{k+\Delta} \), where \( \Delta \in [0, 1) \) and \( g_k \leq \tilde{a} \frac{k!}{\bar{h}^k} \) for some \( \bar{a} > 0 \) and \( 0 < \bar{h} < h \) for every \( k \in \mathbb{N} \), we infer that \( g(x) \leq \bar{a} e^{\bar{h}x} x^\Delta \) for any \( x \geq 0 \), and hence there exist some \( \bar{a}_0 > 0 \) and \( \tilde{h}_0 < \bar{h} < h \) such that \( g(x) \leq \bar{a}_0 e^{\tilde{h}_0 x} \) for any \( x \geq 0 \). Therefore

\[ E[g(X_n - t_n x)1_{X_n \geq t_n x}] \leq \bar{a}_0 E[e^{\tilde{h}_0 X_n}] < \infty, \]

since \( \tilde{h}_0 < h \leq d \), where we recall that \( E[e^{\tilde{h}_0 X_n}] \) exists on \( S_{(c,d)} \). Moreover, we assume that \( g(x) = \sum_{k=0}^{\infty} g_k x^{k+\Delta} \), where \( \Delta \in [0, 1) \) and \( g_k \leq \tilde{a} \frac{k!}{\bar{h}^k} \) for some \( \bar{a} > 0 \) and \( 0 < \bar{h} < h \) for every \( k \in \mathbb{N} \), we can check that \( \beta_k \) whose definition involves the summation of \( g_k \Gamma(k+\Delta+q+(m_1+\ldots+m_n)-m-1) \) over \( q \) from 0 to \( \infty \) is well-defined.

We next introduce a lemma. It calculates certain Gaussian integrals which will be used in our computations in proving Proposition 11. The proof of this lemma will be deferred to the end of this section.

**Lemma 14.** Fix any \( s \geq 0 \). We have as \( t_n \to \infty \),

\[ \int_{y=0}^{\infty} y^s e^{-\frac{(y+h\sqrt{2\eta''(h))})^2}{2}} dy \sim e^{-h^2 t_n \eta''(h)} \frac{1}{2} \Gamma(k+1) \sum_{k=0}^{\infty} \frac{\Gamma(k+s+1)a_k}{(h^2 t_n \eta''(h))^{\frac{k}{2}+\frac{1}{2}+k}}, \]

where \( (a_k)_{k=0}^{\infty} \) are determined recursively as:

\[ a_k = (-1)^k - \sum_{j=0}^{k-1} \frac{(j+s)_{k-j}}{(k-j)!2^{k-j}a_j}, \]

where \( (\cdot)_i \) is the Pochhammer symbol.
We are now ready to prove Proposition 11 below.

By the expansion $g(z) = \sum_{k=0}^{\infty} g_k z^{\Delta+k}$, it suffices to show our results for $g(z) = z^\gamma$ for every $\gamma \geq 0$. Fix $h$ where $\eta'(h) = x$. As in Lemma 4.7 of [31], we denote $\tilde{X}_n$ as the random variable which follows the law $Q(dy) = \frac{e^{h_{\tilde{X}_n(h)}(y)}}{\phi_{\tilde{X}_n(h)}(y)} \mathbb{P}(X_n \in dy)$. Denote $F_n(\cdot)$ the cumulative distribution function of the random variable

$$\frac{\tilde{X}_n - t_n \eta'(h)}{\sqrt{t_n \eta''(h)}}$$

Then as in the proof of Theorem 4.3 in [31], one can readily compute that

$$\mathbb{E} [(X_n - t_n x)^\gamma 1_{X_n \geq t_n x}]$$

$$= \varphi_{X_n}(h) \int_{y=t_n x}^{\infty} (y - t_n x)^\gamma e^{-hy} Q(dy)$$

$$= \varphi_{X_n}(h) \int_{u=0}^{\infty} \left( \sqrt{t_n \eta''(h)} u \right)^\gamma e^{-h(t_n \eta'(h) + \sqrt{t_n \eta''(h)} u)} dF_n(u)$$

$$= \psi_n(h) e^{-t_n F(x)} \left( \sqrt{t_n \eta''(h)} \right)^\gamma \int_{u=0}^{\infty} u^\gamma e^{-h \sqrt{t_n \eta''(h)} u} dF_n(u)$$

$$= \psi_n(h) e^{-t_n F(x)} \left( \sqrt{t_n \eta''(h)} \right)^\gamma$$

$$\cdot \int_{u=0}^{\infty} \left[ h \sqrt{t_n \eta''(h)} u^\gamma - \gamma u^{\gamma-1} \right] e^{-h \sqrt{t_n \eta''(h)} u} (F_n(u) - F_n(0)) du,$$

(A.4)

where we used integration by parts in the last equality.

Suppose we can find $G_n(x) = \int_{-\infty}^{x} g_n(y) dy$, such that

$$\sup_{x \in \mathbb{R}} |F_n(x) - G_n(x)| = o(t_n^{-u}),$$

(A.5)

where

$$g_n(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \left[ 1 + \sum_{j=1}^{2} Q_j(y) \right],$$

(A.6)
with \(Q_j(y)\) being polynomials of order \(j\) in \(y\). Then we obtain from (A.4) that
\[
\mathbb{E}[(X_n - t_n x)^{\gamma} 1_{X_n \geq t_n x}] \\
= \psi_n(h) e^{-t_n F(x)} \left( \frac{1}{2} \right)^{\gamma} \\
\cdot \int_{0}^{\infty} \left[ h \frac{1}{2} \sqrt{t_n \eta''(h) u^\gamma - \gamma u^{\gamma - 1}} \right] e^{-h \frac{1}{2} \sqrt{t_n \eta''(h) u^\gamma - \gamma u^{\gamma - 1}}} (G_n(u) - G_n(0) + o(t_n^{-v})) \, du \\
= \psi_n(h) e^{-t_n F(x)} \left( \frac{1}{2} \right)^{\gamma} \\
\cdot \int_{0}^{\infty} \left[ h \frac{1}{2} \sqrt{t_n \eta''(h) u^\gamma - \gamma u^{\gamma - 1}} \right] e^{-h \frac{1}{2} \sqrt{t_n \eta''(h) u^\gamma - \gamma u^{\gamma - 1}}} g_n(y) \, dy \\
+ \psi_n(h) e^{-t_n F(x)} \left( \frac{1}{2} \right)^{\gamma} \left( \int_{0}^{\infty} y^\gamma e^{-h \frac{1}{2} \sqrt{t_n \eta''(h) y^2 - \gamma y^{\gamma - 1}}} g_n(y) \, dy + 1_{\gamma = 0} \cdot o(t_n^{-v}) \right),
\]
where we used
\[
\int_{0}^{\infty} \left[ h \frac{1}{2} \sqrt{t_n \eta''(h) u^\gamma - \gamma u^{\gamma - 1}} \right] e^{-h \frac{1}{2} \sqrt{t_n \eta''(h) u^\gamma - \gamma u^{\gamma - 1}}} \, du = y^\gamma e^{-h \frac{1}{2} \sqrt{t_n \eta''(h) y^2}} \bigg|_{y = 0}^{\infty} = 1_{\gamma = 0}.
\]
To calculate and approximate the integral in (A.7), in view of the expression of \(g_n(\cdot)\) in (A.6), it suffices to note that
\[
\int_{0}^{\infty} y^\gamma e^{-h \frac{1}{2} \sqrt{t_n \eta''(h) y^2}} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \, dy = e^{h^2 t_n \eta''(h) \frac{1}{2}} \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} y^\gamma e^{-(y + h \sqrt{t_n \eta''(h)})^2/2} \, dy,
\]
and by Lemma 13 for any \(s \geq 0,
\[
\int_{0}^{\infty} y^s e^{-(y + h \sqrt{t_n \eta''(h)})^2/2} \, dy \sim e^{-h^2 t_n \eta''(h) \frac{1}{2} \sum_{k=0}^{\infty} \Gamma(k + s + 1) a_k} (h^2 t_n \eta''(h))^{\frac{1}{2} + \frac{1}{2} + k} = e^{-h^2 t_n \eta''(h) \frac{1}{2} O \left( t_n^{-\frac{1}{2} - \frac{1}{4}} \right)}
\]
as this implies that
\[
\int_{0}^{\infty} y^\gamma e^{-h \frac{1}{2} \sqrt{t_n \eta''(h) y^2}} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \frac{Q_j(y)}{t_n^{1/2}} \, dy = O \left( t_n^{\frac{1}{2} - \frac{1}{4} - \frac{1}{4} - \frac{1}{4}} \right) = O \left( t_n^{-\frac{3}{2} - \frac{1}{4} - j} \right).
\]
It then follows from (A.7) and the fact that \(\psi_n(h)\) converges to \(\psi(h)\) with speed \(O(t_n^{-v})\) locally uniformly that for \(g(x) = x^\gamma\), we have the expansion of the form
\[
\mathbb{E} [g(X_n - t_n x) 1_{X_n \geq t_n x}] = \frac{e^{-t_n F(x)}}{\sqrt{2\pi} \sqrt{t_n \eta''(h)}} \left[ \hat{d}_0 + \frac{\hat{d}_1}{t_n} + \frac{\hat{d}_2}{t_n^2} + \cdots + \frac{\hat{d}_{n-1}}{t_n^{v-1}} + o(t_n^{-v+1}) \right].
\]
To complete the proof, it remains to find the function \(G_n(\cdot)\) so that (A.3) holds and identify the coefficients \(\hat{d}_k\) in the above expansion. This will be done in the next section.
A.2.1 Finding $G_n(\cdot)$ and computing $\delta_k$

In this section we show how to find the function $G_n(\cdot)$ (or equivalently $g_n(\cdot)$) so that (A.5) holds, and we also compute the expansion coefficients $\delta_k$.

To find $G_n(\cdot)$ which approximates the distribution function $F_n(\cdot)$, we approximate the Fourier transform of the distribution $F_n(\cdot)$, and then take the inverse Fourier transform to obtain $g_n(\cdot)$. To this end, we write the Fourier transform $f_n^*(\zeta) = \int_\mathbb{R} e^{i\zeta x} dF_n(x)$, and

$$\tilde{\psi}(z) = \frac{\psi(z + h)}{\psi(h)}, \quad \text{and} \quad \tilde{\eta}(z) = \eta(z + h) - \eta(h).$$

Then for every $k \in \mathbb{N}$,

$$\tilde{\psi}^{(k)}(0) = \frac{\psi^{(k)}(h)}{\psi(h)}, \quad \text{and} \quad \tilde{\eta}^{(k)}(0) = \eta^{(k)}(h).$$

We can compute that for $z = i\zeta$,

$$f_n^*(\zeta) := E \left[ e^{\frac{\tilde{\psi}(z\sqrt{t_n\eta''(h)})}{\sqrt{t_n\eta''(h)}}} \right] = e^{t_n \left( \tilde{\eta}\left(\frac{z}{\sqrt{t_n\eta''(h)}}\right) - \eta'(\frac{z}{\sqrt{t_n\eta''(h)}}) \right)} \cdot \tilde{\psi}_n^* \left(\frac{z}{\sqrt{t_n\eta''(h)}}\right), \quad (A.9)$$

where

$$\tilde{\psi}_n(z) := e^{-t_n \tilde{\eta}(z)} \cdot E \left[ e^{zX_n} \right].$$

By Lemma 4.7 in [31], $\tilde{\psi}_n$ converges to $\tilde{\psi}$ locally uniformly with speed $O(t_n^{-1})$, and together with Taylor expansion, one obtains

$$\tilde{\psi}_n^* \left(\frac{z}{\sqrt{t_n\eta''(h)}}\right) = \sum_{i=0}^{2v} \frac{\psi^{(i)}(0)}{i!} \left(\frac{z}{\sqrt{t_n\eta''(h)}}\right)^i + o \left(\frac{z}{\sqrt{t_n}}\right)^{2v}. \quad (A.10)$$

In addition, as in (A.3), one can expand the following term in (A.9)

$$e^{t_n \left( \tilde{\eta}\left(\frac{z}{\sqrt{t_n\eta''(h)}}\right) - \eta'(\frac{z}{\sqrt{t_n\eta''(h)}}) \right)} = e^{\frac{z^2}{\eta''(h)} \sum_{k=1}^{\infty} \frac{\eta^{(k+2)}(h)}{(k+2)!} \left(\frac{z}{\sqrt{t_n\eta''(h)}}\right)^k}$$

by applying the Faà di Bruno’s formula (see Section C.2). This leads to the following:

$$f_n^*(\zeta) = g_n^*(\zeta) + e^{t_n^2} P_{2v+1}(|z|) \left(\frac{|z|}{\sqrt{t_n}}\right)^{2v} \varepsilon \left(\frac{z}{\sqrt{t_n}}\right), \quad (A.10)$$

where $\lim_{t \to 0} \varepsilon(t) = 0$, and $P_i(|z|)$ is some polynomial of $|z|$ of order $i$, and for $z = i\zeta$,

$$g_n^*(\zeta) = \sum_{k=0}^{2v} e^{t_n^2} \sum_{\ell=0}^{k} \psi(h)(k-\ell) \sum_{S_1} \frac{1}{m_1!m_2!\cdots m_{\ell}!l^!} \cdot \prod_{j=1}^{\ell} \left(\frac{1}{\eta''(h) + 1} \right) \frac{z^{k+2(m_1+\cdots+m_{\ell})}}{(\eta''(h))^{k/2} i_n^{k/2}}. \quad (A.11)$$
It is known that
\[ \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-it\gamma} e^{-t^2/2} (it)^k dt = H_k(y) \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}, \]  
(A.12)
where \( H_k(y) \) are Hermite polynomials given by
\[ H_k(y) = k! \sum_{m=0}^{[k/2]} \frac{(-1)^m}{m!(k-2m)!} y^{k-2m}. \]  
(A.13)

Therefore, by (A.11), (A.12) and \( z = i\zeta \), the inverse Fourier transform of \( g_n^*(\cdot) \) is given by
\[ g_n(y) := \frac{1}{2\pi} \int_{-\infty}^{\infty} g_n^*(\zeta) e^{-iy\zeta} d\zeta \]
\[ = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \sum_{k=0}^{2v} \sum_{\ell=0}^{k} \frac{\psi^{(k-\ell)}(h)}{(k-\ell)!} \sum_{S_\ell} m_1! m_2! m_2! \cdots m_\ell! \ell! m_\ell \cdot \prod_{j=1}^{\ell} \left( \frac{1}{\eta''(h)} \right)^{m_j} \frac{1}{(\eta''(h))^{k/2}} \cdot \frac{1}{t_n^{k/2}}. \]  
(A.14)

We will show later that for \( G_n(x) = \int_{-\infty}^{x} g_n(y) dy \) with \( g_n(\cdot) \) given in (A.14), we have
\[ \sup_{x \in \mathbb{R}} |F_n(x) - G_n(x)| = o(t_n^{-v}). \]

By using the expansion for Hermite polynomials in (A.13), one can readily express \( g_n(\cdot) \) in the form of (A.6). Then similar arguments in (A.7)-(A.8) yield that
\[ \mathbb{E} [(X_n - t_n x)^\gamma 1_{X_n \geq t_n x}] \]
\[ = e^{-t_n F(x)} \frac{1}{\sqrt{2\pi t_n \eta''(h)} h^{\gamma+1}} \sum_{k=0}^{2v} \sum_{\ell=0}^{k} \frac{\psi^{(k-\ell)}(h)}{(k-\ell)!} \sum_{S_\ell} m_1! m_2! m_2! \cdots m_\ell! \ell! m_\ell \cdot \prod_{j=1}^{\ell} \left( \frac{1}{\eta''(h)} \right)^{m_j} \frac{1}{(\eta''(h))^{k/2}} \cdot \frac{[\gamma + m_1 + \cdots + m_\ell]}{m!(k+2m_1 + \cdots + m_\ell - 2m)!} \cdot \sum_{m=0}^{[\gamma + m_1 + \cdots + m_\ell]} \frac{\Gamma(q + \gamma + k + 2(m_1 + \cdots + m_\ell) - 2m + 1)a_q}{(h^2 \eta''(h))^{2m+1} \prod_{j=1}^{\ell} (2m_\ell)!} + o(t_n^{-v+1}). \]
By comparing the above equation with the expansion

$$\mathbb{E} [(X_n - t_n x)^{\gamma} 1_{X_n \geq t_n x}] = \frac{e^{-t_n F(x)}}{\sqrt{2\pi t_n \eta'(h)}} \left( \hat{d}_0 + \frac{\hat{d}_1}{t_n} + \frac{\hat{d}_2}{t_n^2} + \cdots + \frac{\hat{d}_{v-1}}{t_n^{v-1}} + O \left( \frac{1}{t_n^v} \right) \right),$$

we conclude that for every $k \leq v - 1$,

$$\hat{d}_k = \frac{1}{k^{\gamma+1}} \sum_{p=0}^{2k} \sum_{\ell=0}^{p} \psi^{(p-\ell)}(h) \prod_{j=1}^{\ell} \left( \frac{1}{\eta''(h) (j+2)(j+1)^{m_j}} \right)^m \frac{1}{(\eta''(h))^{p/2}}.$$ 

So we have obtained the formula of $\hat{d}_k$ when $g(z) = z^\gamma$. Finally, assume $g(z)$ has the expansion $g(z) = \sum_{q=0}^{\infty} g_q z^{q+\Delta}$, then the formula of $\hat{d}_k$ in Part (ii) of Proposition 11 readily follows.

To complete the proof, it only remains to show that

$$\sup_{x \in \mathbb{R}} |F_n(x) - G_n(x)| = o(t_n^{-v}).$$

where $F_n(x) = \int_{-\infty}^{x} g_n(y) dy$ with $g_n(\cdot)$ given in (A.11). We apply the Esseen’s smoothing inequality in Appendix C.1 and use a similar argument as the proof of Proposition 4.1 in [31].

Recall from (A.10) that

$$f_n^*(\zeta) = g_n^*(\zeta) + e^{-\frac{2}{\pi}} H_{2v+1}(|\zeta|) \left( \frac{|\zeta|}{\sqrt{t_n}} \right)^{2v} \varepsilon \left( \frac{z}{\sqrt{t_n}} \right),$$

where $f_n^*$ and $g_n^*$ are the corresponding Fourier transforms of $F_n$ and $G_n$, $\lim_{t \to 0} \varepsilon(t) = 0$, and $P_i(|\zeta|)$ is a polynomial of $|\zeta|$ of order $i$. Write $k = 2v$. We can check that $F_n - G_n$ vanish at $\pm \infty$, since $F_n$ and $G_n$ both vanish at $-\infty$ and $\lim_{x \to +\infty} F_n(x) = \lim_{x \to +\infty} G_n(x) = 1$ as $F_n$ is a cumulative distribution function and $g_n^*(0) = 1$ from (A.11). In addition, we have

$$\sup_{n \in \mathbb{N}} \sup_{x \in \mathbb{R}} |G_n'(x)| = \sup_{n \in \mathbb{N}} \sup_{x \in \mathbb{R}} g_n(x) =: m < \infty,$$

from the definition of $g_n$ in (A.14) and the fact that $e^{-\frac{2}{\pi} H_{k+2(m_1 + \cdots + m_\ell)}(y)} \frac{1}{t_n^{k+2}}$ is uniformly bounded in $n \in \mathbb{N}$, $y \in \mathbb{R}$ and $m_1 \cdot 1 + \cdots + m_\ell \ell = \ell$, where $\ell \leq k \leq 2v$, where $v \in \mathbb{N}$ is
fixed. By taking $T = Mt_n^{k/2}$ and fixing $\delta \in (0, M)$, we deduce from Esseen’s smoothing inequality that

$$|F_n(x) - G_n(x)| \leq \frac{1}{\pi} \int_{-Mt_n^{k/2}}^{Mt_n^{k/2}} \left| \frac{f_n^*(\xi) - g_n^*(\xi)}{\xi} \right| d\xi + r(\pi) \frac{m}{Mt_n^{k/2}}$$

$$\leq \frac{1}{\pi} \int_{-\delta t_n^{1/2}}^{\delta t_n^{1/2}} \left| \frac{f_n^*(\xi) - g_n^*(\xi)}{\xi} \right| d\xi + r(\pi) \frac{m}{Mt_n^{k/2}}$$

$$+ \frac{1}{\pi} \int_{[-Mt_n^{k/2}, Mt_n^{k/2}] \setminus [-\delta t_n^{1/2}, \delta t_n^{1/2}]} \left| \frac{f_n^*(\xi) - g_n^*(\xi)}{\xi} \right| d\xi. \quad (A.15)$$

We next estimate each of the three terms in the right hand side of the above inequality. We can estimate that

$$\frac{1}{\pi} \int_{-\delta t_n^{1/2}}^{\delta t_n^{1/2}} \left| \frac{f_n^*(\xi) - g_n^*(\xi)}{\xi} \right| d\xi = \frac{1}{t_n^{k/2}} \frac{1}{\pi} \int_{-\delta t_n^{1/2}}^{\delta t_n^{1/2}} e^{-\xi^2/2} P_{k+1}(|\xi|) |\xi|^{k-1} \varepsilon \left( \frac{\xi}{\sqrt{t_n}} \right) d\xi$$

$$\leq \frac{C_0}{t_n^{k/2}} \max_{-\delta \leq \xi \leq \delta} \varepsilon(t),$$

for some constant $C_0$ (that depends on $k$). For fixed small $\epsilon > 0$, one can choose $\delta$ so that $\max_{-\delta \leq \xi \leq \delta} \varepsilon(t) < \epsilon$ and $m\delta < \epsilon$, and hence we have for large $t_n$

$$\frac{1}{\pi} \int_{-\delta t_n^{1/2}}^{\delta t_n^{1/2}} \left| \frac{f_n^*(\xi) - g_n^*(\xi)}{\xi} \right| d\xi = o \left( \frac{1}{t_n^{k/2}} \right),$$

uniformly in $x$. In addition, by taking $M = 1/\delta$ we get

$$r(\pi) \frac{m}{Mt_n^{k/2}} = o \left( \frac{1}{t_n^{k/2}} \right).$$

We can also estimate that

$$\frac{1}{\pi} \int_{[-Mt_n^{k/2}, Mt_n^{k/2}] \setminus [-\delta t_n^{1/2}, \delta t_n^{1/2}]} \left| \frac{f_n^*(\xi) - g_n^*(\xi)}{\xi} \right| d\xi$$

$$\leq \frac{1}{\pi \delta t_n^{1/2}} \left[ \int_{\delta t_n^{1/2} \leq |\xi| \leq Mt_n^{k/2}} |f_n^*(\xi)| d\xi + \int_{\delta t_n^{1/2} \leq |\xi| \leq Mt_n^{k/2}} |g_n^*(\xi)| d\xi \right], \quad (A.16)$$
and we can compute that

\[
\int_{\delta t_n^{1/2} \leq |\zeta| \leq M t_n^{k/2}} |f_n^*(\zeta)| d\zeta \\
\leq \int_{\delta t_n^{1/2} \leq |\zeta| \leq M t_n^{k/2}} \left| \psi_n \left( \frac{i\zeta}{\sqrt{t_n \eta''(0)}} \right) \right| e^{t_n \eta''(0)} \left| \frac{i\zeta}{\sqrt{t_n \eta''(0)}} \right| d\zeta \\
\leq \sqrt{t_n} \int_{\delta \leq |\zeta| \leq M t_n^{(k-1)/2}} \psi_n \left( \frac{i\zeta}{\sqrt{\eta''(0)}} \right) d\zeta \cdot \left( \max_{|\zeta| \geq \delta} \left| \eta \left( \frac{i\zeta}{\sqrt{\eta''(0)}} \right) \right| \right) t_n.
\]

Since \( \phi \) is non-lattice with \( \int_{\mathbb{R}} e^{\nu x} \phi(dx) = e^{\nu(x)} \), by Lemma 4.9. in [31], we obtain

\[
\max_{|\zeta| \geq \delta} \left| \exp \left\{ \eta \left( \frac{i\zeta}{\sqrt{\eta''(0)}} \right) \right\} \right| < 1.
\]

Then it follows that

\[
\int_{\delta t_n^{1/2} \leq |\zeta| \leq M t_n^{k/2}} |f_n^*(\zeta)| d\zeta \\
\leq \limsup_{n \to \infty} \sup_{x \in \mathbb{R}} |\psi_n (ix)| \cdot 2 M t_n^{k/2} \cdot \left( \max_{|\zeta| \geq \delta} \left| \eta \left( \frac{i\zeta}{\sqrt{\eta''(0)}} \right) \right| \right) t_n,
\]

which goes to zero faster than any power of \( t_n \). Finally, the term \( \int_{\delta t_n^{1/2} \leq |\zeta| \leq M t_n^{k/2}} |g_n^*(\zeta)| d\zeta \) in (A.16) can be estimated similarly. Therefore, we deduce from (A.15) that for large \( t_n \),

\[
\sup_{x \in \mathbb{R}} |F_n(x) - G_n(x)| = o \left( \frac{1}{t_n^{k/2}} \right) = o \left( \frac{1}{t_n^v} \right).
\]

The proof is hence complete.

**A.2.2 Proof of Lemma 14**

**Proof of Lemma 14.** For any \( s \geq 0 \), it is readily seen that

\[
\int_{y=0}^{\infty} y^s e^{-\left( y+h \sqrt{t_n \eta''(h)} \right)^2} dy = \left( h \sqrt{t_n \eta''(h)} \right)^{s+1} \int_{y=0}^{\infty} y^s e^{-y^2 t_n \eta''(h) \frac{(y+1)^2}{2}} dy.
\]

For large \( t_n \), we can apply the Laplace’s method (see Lemma 18 in Appendix C.3) to estimate the above integral. To this end, we use the notation in Lemma 18 and we set

\[
p(y) = \frac{1}{2} (y+1)^2, \quad q(y) = y^s.
\]
and \( a = 0, b = \infty \) and \( \lambda = h^2 t, \eta''(h) \). We can compute that

\[
p(y) = \frac{1}{2} + y + \frac{1}{2}y^2,
\]

which implies that \( \mu = 1 \). Since \( q(y) = y^s \), we have \( \sigma = s + 1 \). Moreover,

\[
v = p(y) - p(0) = y + \frac{1}{2}y^2,
\]

and \((a_k)_{k=0}^\infty\) are determined from:

\[
\frac{q(y)}{p'(y)} = \frac{y^s}{1 + y} \sim \sum_{k=0}^{\infty} a_k \left(y + \frac{1}{2}y^2\right)^{k+s}, \quad y \to 0^+.
\]

Hence, as \( t_n \to \infty \),

\[
\int_{y=0}^{\infty} y^s e^{-h^2 t_n \eta''(h) \frac{(y+1)^2}{2}} dy \sim e^{-h^2 t_n \eta''(h) \frac{1}{2}} \sum_{k=0}^{\infty} \frac{\Gamma(k + s + 1)}{(h^2 t_n \eta''(h))^{s+1+k}} a_k.
\]

Finally, we can compute that

\[
\frac{y^s}{1 + y} \sim \sum_{k=0}^{\infty} a_k \left(y + \frac{1}{2}y^2\right)^{k+s} = \sum_{k=0}^{\infty} a_k y^{s+k} \left(1 + \frac{1}{2}y\right)^{k+s},
\]

which implies that for \( y \to 0^+ \),

\[
\frac{1}{1 + y} = \sum_{k=0}^{\infty} (-1)^k y^k \sim \sum_{j=0}^{\infty} a_j y^j \left(1 + \frac{1}{2}y\right)^{j+s} = \sum_{j=0}^{\infty} a_j y^j \sum_{i=0}^{\infty} \frac{1}{2^i} y^i \frac{(j + s)_i}{i!},
\]

where \((\cdot)_i\) is the Pochhammer symbol. This implies that

\[
(-1)^k = \sum_{j=0}^{k} \frac{(j + s)_{k-j}}{(k-j)!2^{k-j}} a_j = \sum_{j=0}^{k-1} \frac{(j + s)_{k-j}}{(k-j)!2^{k-j}} a_j + \sum_{j=0}^{k-1} \frac{(j + s)_{k-j}}{(k-j)!2^{k-j}} a_j.
\]

Hence, the conclusion follows.

\[\square\]

### B Computations of expansion coefficients for affine point processes

In this section, we compute \( \hat{c}_k \) and \( \hat{d}_k \) for affine point processes. Their general formulas are given in Propositions 8 and 11. It is clear that these coefficients depend on the derivatives of \( \eta \) and \( \psi \) at \( h \). Recall that for affine point processes, the function \( \psi(\cdot) \) is given in (5.1), \( \eta(\cdot) \) is given in (2.3), and the real number \( h \) is defined by \( \eta'(h) = R \). We next discuss how to compute high order derivatives for functions \( \psi \) and \( \eta \) that are associated with affine point processes.
B.1 Computations of $\psi^{(k)}(\theta)$

From (5.1), we have $\psi(\theta) = e^{u^*(\theta)^T x_0 + B(\infty; \theta, u^*(\theta))}$. By applying Faà di Bruno’s formula (see Section [C.2]), we get the $k$–th derivative of $\psi$ is given by

$$
\psi^{(k)}(\theta) = \sum_{S_k} \frac{k!\psi(\theta)}{m_1!^{|m_1|} m_2!^{|m_2|} \ldots m_k!^{|m_k|}} \prod_{j=1}^{k} \left( \left( u^*(\theta)^{s_j} \right) x_0 + \frac{d^{j}}{d\theta^j} B(\infty; \theta, u^*(\theta)) \right)^{m_j}.
$$

Hence, it suffices for us to compute the derivatives $\frac{d^{j}}{d\theta^j} B(\infty; \theta, u^*(\theta))$ which will be presented in Section [B.3] and the derivatives of $u^*(\theta)$ with respect to $\theta$ (up to order $k$), which will be presented in Section [B.4].

B.2 Computations of $\eta^{(k)}(\theta)$

We recall from (2.3) that

$$
\eta(\theta) = u^*(\theta)^T b + \sum_{i=1}^{n} \lambda_i \left( E[e^{(\theta + u^*(\theta)^T \gamma_i)Z_i}] - 1 \right).
$$

For any $k \geq 1$, by Leibnitz formula, we can compute the $k$–th derivative of $\eta$

$$
\eta^{(k)}(\theta) = (u^*)^{(k)}(\theta)^T b + \sum_{i=1}^{n} \lambda_i \mathbb{E} \left[ \frac{d^k}{d\theta^k} e^{\theta Z_i u^*(\theta)^T \gamma_i Z_i} \right]
$$

and by Faà di Bruno’s formula,

$$
\frac{d^j}{d\theta^j} e^{u^*(\theta)^T \gamma_i Z_i} = \sum_{S_j} \frac{j! e^{u^*(\theta)^T \gamma_i Z_i}}{m_1!^{|m_1|} m_2!^{|m_2|} \ldots m_j!^{|m_j|}} \prod_{k=1}^{n} \left( u^*(\theta)^{s_k} \right)^{m_k},
$$

which yields that

$$
\eta^{(k)}(\theta) = (u^*)^{(k)}(\theta)^T b + \sum_{i=1}^{n} \lambda_i \mathbb{E} \left[ \frac{d^k}{d\theta^k} e^{\theta Z_i u^*(\theta)^T \gamma_i Z_i} \right]
$$

$$
= (u^*)^{(k)}(\theta)^T b + \sum_{i=1}^{n} \lambda_i \sum_{j=0}^{k} \binom{k}{j} \sum_{S_j} \frac{j! e^{u^*(\theta)^T \gamma_i Z_i}}{m_1!^{|m_1|} m_2!^{|m_2|} \ldots m_j!^{|m_j|}} \prod_{k=1}^{n} \left( u^*(\theta)^{s_k} \right)^{m_k}.
$$

Hence, to find the $k$–th derivative of $\eta$, we need to compute the derivatives of $u^*(\theta)$ with respect to $\theta$ (up to order $k$), which we present in Section [B.4].
B.3 Computations of $\frac{d^j}{d\theta^j} B(\infty; \theta, u^*(\theta))$

By applying multivariate Faà di Bruno’s formula and using the notations in Appendix C.2, we have

$$\frac{d^j}{d\theta^j} B(\infty; \theta, u^*(\theta)) = \frac{\partial^j}{\partial\theta^j} B(\infty; \theta, u^*(\theta)) + \sum_{1 \leq |\lambda| \leq j} D^\lambda_\delta B(\infty; \delta) \sum_{p(j,\lambda)} j! \prod_{m=1}^j \left[ \frac{(u^*)^m(\theta)}{(k_m!) [\ell_m!]^{k_m}} \right],$$

where $\delta = u^*(\theta)$.

We first compute $\frac{\partial^j}{\partial\theta^j} B(\infty; \theta, u^*(\theta))$. Recall from Theorem 13 that,

$$B(\infty; \theta, \delta) = \int_0^\infty b^T Adt + \int_0^\infty \frac{1}{2} A^T a Adt + \sum_{i=1}^n \lambda_i \int_0^\infty \int_{\mathbb{R}_+} (e^{A^T \gamma_i z} - 1) e^{(\theta + \delta^T \gamma_i) z} \phi_i(dz)dt,$$

where for notational simplicity, we write $A := A(t; \theta, \delta) = (A_1(t; \theta, \delta), \ldots, A_n(t; \theta, \delta))$, with $A(0; \theta, \delta) = -\delta$ and

$$\frac{d}{dt} A_j(t; \theta, \delta) = -\sum_{i=1}^n A_i \beta^*_{i,j} + \frac{1}{2} A^T \alpha^j A + \sum_{i=1}^n \int_{\mathbb{R}_+} (e^{A^T \gamma_i z} - 1) e^{(\theta + \delta^T \gamma_i) z} \phi_i(dz)dt,$$

for $j = 1, 2, \ldots, n$, where

$$\beta^* = \begin{pmatrix} \beta_{1,1} - \text{diag}(\alpha_{11}^1 u^*_1(\theta), \ldots, \alpha_{mm}^m u^*_m(\theta)) & 0 \\ \beta_{j,1} & \beta_{j,j} \end{pmatrix}.$$
and by further applying the univariate Faà di Bruno’s formula (see Appendix C.2), we get

\[
\frac{\partial^j}{\partial \theta^j} B(\infty; \theta, \delta) = \int_0^\infty b^T \frac{\partial^j}{\partial \theta^j} \eta dt + \int_0^\infty \sum_{\ell=0}^j \binom{j}{\ell} \frac{1}{2} \frac{\partial^\ell}{\partial \theta^\ell} A^T \frac{\partial^{j-\ell}}{\partial \theta^{j-\ell}} \eta dt
\]

\[
+ \sum_{i=1}^n \lambda_i \int_0^\infty \int_{\mathbb{R}^+} \sum_{\ell=1}^j \binom{j}{\ell} \sum_{m_1!m_1!m_2!m_2! \ldots m_\ell!m_\ell!} \prod_{k=1}^\ell \left( \gamma_{i,z} \frac{\partial^k}{\partial \theta^k} A^T \right)^{m_k} \cdot z^{-\ell} e^{(\theta + \delta^T \gamma_{i,z})} \varphi_i(dz) dt
\]

\[
+ \sum_{i=1}^n \lambda_i \int_0^\infty \int_{\mathbb{R}^+} \left( e^{A^T \gamma_{i,z}} - 1 \right) z^{-\ell} e^{(\theta + \delta^T \gamma_{i,z})} \varphi_i(dz) dt, \quad (B.3)
\]

Moreover, for any \(1 \leq j \leq n\),

\[
\frac{d}{dt} \frac{\partial^q}{\partial \theta^q} A(t; \theta, \delta) = -\sum_{i=1}^n \sum_{\ell=0}^q \binom{q}{\ell} \frac{\partial^\ell}{\partial \theta^\ell} A_i \frac{\partial^{q-\ell}}{\partial \theta^{q-\ell}} \beta^*_i,j + \sum_{\ell=0}^q \binom{q}{\ell} \frac{1}{2} \frac{\partial^\ell}{\partial \theta^\ell} A^T \alpha_j \frac{\partial^{q-\ell}}{\partial \theta^{q-\ell}} A
\]

\[
+ \sum_{i=1}^n \int_{\mathbb{R}^+} \sum_{\ell=1}^q \binom{q}{\ell} \sum_{m_1!m_1!m_2!m_2! \ldots m_\ell!m_\ell!} \prod_{k=1}^\ell \left( \gamma_{i,z} \frac{\partial^k}{\partial \theta^k} A^T \right)^{m_k} \cdot z^{-\ell} e^{(\theta + \delta^T \gamma_{i,z})} \varphi_i(dz)
\]

\[
+ \sum_{i=1}^n \int_{\mathbb{R}^+} \left( e^{A^T \gamma_{i,z}} - 1 \right) z^{-\ell} e^{(\theta + \delta^T \gamma_{i,z})} \varphi_i(dz) \kappa_{i,j}, \quad (B.4)
\]

with \(\frac{\partial^q}{\partial \theta^q} A(0; \theta, \delta) = 0\) for any \(q \geq 1\), and

\[
\frac{\partial^{q-\ell}}{\partial \theta^{q-\ell}} \beta^*_i,j = \begin{pmatrix}
\beta_{1,i} \cdot 1_{q-\ell=0} - \text{diag}(\alpha_1 \frac{\partial^{q-\ell}}{\partial \theta^{q-\ell}} u_1^*(\theta), \ldots, \alpha_m \frac{\partial^{q-\ell}}{\partial \theta^{q-\ell}} u_m^*(\theta)) & 0 \\
\beta_{j,i} \cdot 1_{q-\ell=0} & \beta_{j,j} \cdot 1_{q-\ell=0}
\end{pmatrix};
\]

Next, let us compute \(D^\nu B(\infty; \theta, \delta)\) for any multi-index \(\nu = (\nu_1, \ldots, \nu_d) \neq 0\). By Applying Leibnitz formula and multivariate Faà di Bruno’s formula (see Appendix C.2),
we get

\[
D_\delta^\nu B(\infty; \theta, \delta) = \int_0^\infty b^T D_\delta^\nu Adt + \int_0^\infty \frac{1}{2} \sum_{\lambda \leq \nu} \left( \frac{\nu}{\lambda} \right) (D_\delta^\lambda A)^T a(D_\delta^{\nu-\lambda} A) dt
\]

\[
+ \sum_{i=1}^n \lambda_i \int_0^\infty \int_{\mathbb{R}} \left[ \sum_{0 \leq \tilde{\nu} \leq \nu} \left( \frac{\nu}{\tilde{\nu}} \right) \sum_{1 \leq \lambda \leq |\tilde{\nu}|} e^{A^T \gamma_i z} \sum_{\nu(p, \lambda)} \tilde{\nu}! \prod_{j=1}^{|\tilde{\nu}|} \left[ D_\delta^{e_j} A^{T \gamma_i z} \right] \frac{k_j}{(k_j!)} e^{\theta + \delta^T \gamma_i z} D_\delta^{-\nu} \right] \varphi_i (dz) dt, \quad (B.5)
\]

where for two multi-indices \( \lambda, \nu \), we write \( \lambda \leq \nu \) if \( \lambda_i \leq \nu_i \) for all \( i \).

So to compute \( D_\delta^\nu B(\infty; \theta, \delta) \), it remains to compute \( D_\delta^\nu A_j(t; \theta, \delta) \) for each \( j \) for a multi-index \( \nu \). From the ODE for \( A_j(t; \theta, \delta) \) in (B.2), one can readily obtain the following ODE for \( D_\delta^\nu A_j(t; \theta, \delta) \):

\[
\frac{d}{dt} D_\delta^\nu A_j(t; \theta, \delta) = -\sum_{i=1}^n D_\delta^\nu A_i \beta_{i,j} + \int_0^\infty \frac{1}{2} \sum_{\lambda \leq \nu} \left( \frac{\nu}{\lambda} \right) (D_\delta^\lambda A)^T \alpha^j (D_\delta^{\nu-\lambda} A)
\]

\[
+ \sum_{i=1}^n \int_{\mathbb{R}} \sum_{1 \leq \lambda \leq |\tilde{\nu}|} e^{A^T \gamma_i z} \left[ \sum_{0 \leq \tilde{\nu} \leq \nu} \left( \frac{\nu}{\tilde{\nu}} \right) \sum_{\nu(p, \lambda)} \tilde{\nu}! \prod_{j=1}^{|\tilde{\nu}|} \left[ D_\delta^{e_j} A^{T \gamma_i z} \right] \frac{k_j}{(k_j!)} e^{\theta + \delta^T \gamma_i z} D_\delta^{-\nu} \right] \varphi_i (dz) \kappa_{i,j}, \quad (B.6)
\]

for \( j = 1, 2, \ldots, n \), with the initial condition \( D_\delta^\nu A(0; \theta, \delta) = -D_\delta^\nu \delta \) as \( A(0; \theta, \delta) = -\delta \).

In summary, one can first numerically solve the ODEs in (B.4) to obtain \( \frac{\partial}{\partial \theta} A(\infty; \theta, \delta) \) and then (B.3) to obtain \( \frac{\partial}{\partial \theta} B(\infty; \theta, \delta) \). Next, we solve numerically the ODEs in (B.6) to obtain \( D_\delta^\nu A_j(t; \theta, \delta) \) for each \( j \), next numerically compute \( D_\delta^\nu B(\infty; \theta, \delta) \) using (B.5), and finally obtain \( \frac{\partial}{\partial \theta} B(\infty; \theta, u^*(\theta)) \) using (B.1).

**B.4 Computations of \((u^*)^{(k)}(\theta)\)**

We recall from Theorem 3 that \( u^*(\cdot) : \mathbb{R} \to \mathbb{R}^n \) is the implicit function defined as the unique solution branch with \( u^*(0) = 0 \) of the system of nonlinear equations:

\[
\beta^T u^*(\theta) - \frac{1}{2} u^*(\theta)^T \kappa^* u^*(\theta) - \kappa^T \left( E \left[ e^{(\theta + (u^*(\theta))^T \gamma)} Z \right] - 1 \right) = 0. \quad (B.7)
\]
To obtain the $k$–th derivative of $u^*(\theta)$ with respect to $\theta$, that is, $(u^* (k))(\theta)$, we can apply Leibnitz formula and Faà di Bruno’s formula (see Appendix C.2) to the above equation and get

$$
\beta^T(u^*)^{(k)}(\theta) - \frac{1}{2} \sum_{j=0}^{k} \binom{k}{j} ((u^*)(k-j))(\theta))^T \alpha(u^*)^{(j)}(\theta)
$$

$$
= \kappa^T \sum_{j=0}^{k} \binom{k}{j} \sum_{S_j} j! \prod_{\ell=1}^{j} m_\ell! \prod_{i=1}^{j} \frac{m_i! m_{i+1}! \cdots m_{j}!}{m_{j+1}! \cdots m_{k}!} f(\theta)^{\gamma \ell} Z \prod_{\ell=1}^{j} ((u^*)^{(\ell)}(\theta))^T \gamma Z^{m_\ell}.
$$

(B.8)

Therefore, given the parameters, one can solve the nonlinear equation in (B.7) to obtain $u^*(\theta)$, and then successively solve the nonlinear equations in (B.8) to obtain the derivatives $(u^*)(k)(\theta)$ for any $k \geq 1$.

C  Smoothing inequality, Faà di Bruno’s formula and Laplace’s Method

For completeness, this section collects three known results that are used in our proofs.

C.1 Esseen’s smoothing inequality

**Lemma 15** (Esseen’s smoothing inequality, cf. Theorem 2 in Chapter V of [33]). Let $F$ be a non-decreasing function and $G$ be a differentiable function of bounded variation on the real line with respective Fourier-Stieltjes transforms $f^*$ and $g^*$. Suppose that $F - G$ vanishes at $\pm \infty$ and that $G$ is $m$-Lipschitz with sup $|G''(x)| \leq m$. Then for every $T > 0$,

$$
\sup_{x \in \mathbb{R}} |F(x) - G(x)| \leq \frac{1}{\pi} \int_{-T}^{T} \left| \frac{f^*(\zeta) - g^*(\zeta)}{\zeta} \right| d\zeta + r(\pi) \frac{m}{T},
$$

where $r(\pi)$ is a positive constant depending only on $\pi$.

C.2 Faà di Bruno’s formula

(Multivariate) Fa di Bruno’s formula gives an explicit equation for the higher order (partial) derivatives of the composition $h(x_1, \ldots, x_d) = f(g_1(x_1, \ldots, x_d), \ldots, g_m(x_1, \ldots, x_d))$, where $g_i : \mathbb{R}^d \to \mathbb{R}$ and $f : \mathbb{R}^m \to \mathbb{R}$ are differentiable a sufficient number of times.

To introduce the formula, we recall some multivariate notations. For $\nu \in (\mathbb{N} \cup \{0\})^d$ and $\mathbf{x} \in \mathbb{R}^d$, we define $|\nu| = \sum_{i=1}^{d} \nu_i$, $\nu! = \prod_{i=1}^{d} \nu_i!$, $D_\nu = \frac{\partial^{\nu}}{\partial x_1^{\nu_1} \cdots \partial x_d^{\nu_d}}$ for $|\nu| > 0$, and $\mathbf{x}^\nu = \prod_{i=1}^{d} x_i^{\nu_i}$. In addition, if $\mu = (\mu_1, \ldots, \mu_d)$ and $\nu = (\nu_1, \ldots, \nu_d)$ are both in $(\mathbb{N} \cup \{0\})^d$, we write $\mu \prec \nu$ provided one of the following holds: (i) $|\mu| < |\nu|$; (ii) $|\mu| = |\nu|$ and $\mu_1 < \nu_1$ or (iii) $|\mu| = |\nu|$, $\mu_1 = \nu_1, \ldots, \mu_k = \nu_k$ and $\mu_{k+1} < \nu_{k+1}$ for some $1 \leq k < d$. 

44
Let \( \nu = (\nu_1, \ldots, \nu_d) \neq (0, \ldots, 0) \). Setting \( h_\nu = D^\nu_x h(x) \), \( f_\lambda = D^\lambda_y f(y) \), where \( y = (g_1(x), \ldots, g_m(x)) \) and \( \nu, \lambda \in (\mathbb{N} \cup \{0\})^d \), and \( g_\mu = (D^\mu_x g_1(x), \ldots, D^\mu_x g_m(x)) \). Then, the multivariate Faà di Bruno’s formula is given as follows.

**Lemma 16** (Multivariate Faà di Bruno’s formula, cf. (2.4) in [8]).

\[
\begin{align*}
  h_\nu &= \sum_{1 \leq |\lambda| \leq |\nu|} f_\lambda \sum_{p(\nu, \lambda)} \nu! \prod_{j=1}^{\nu} \frac{[g_{\ell_j}]^{k_j}}{(k_j!)^{\ell_j} |k_j|!},
\end{align*}
\]

where

\[
p(\nu, \lambda) = \left\{ (k_1, \ldots, k_{|\nu|}; \ell_1, \ldots, \ell_{|\nu|}) : \text{for some } 1 \leq s \leq |\nu|,
  \begin{align*}
    k_i &= 0 \text{ and } \ell_i = 0 \text{ for } 1 \leq i \leq |\nu| - s; \ |k_i| > 0 \text{ for } |\nu| - s + 1 \leq i \leq |\nu|;
  \end{align*}
  \text{and } 0 \prec \ell_{|\nu| - s + 1} \prec \cdots \prec \ell_{|\nu|} \text{ are such that } \sum_{i=1}^{\nu} k_i = \lambda, \sum_{i=1}^{\nu} |k_i| \ell_i = \nu \right\}.
\]

In the above formula, the vectors \( k \) are \( m \)-dimensional, the vectors \( \ell \) are \( d \)-dimensional, and we always set \( 0^0 = 1 \).

For the special case \( d = m = 1 \), we have the one-dimensional Faà di Bruno’s formula, which has a more explicit expression:

**Lemma 17** (Univariate Faà di Bruno’s formula).

\[
\frac{d^n}{dx^n} f(g(x)) = \sum_{S_n} \frac{n!}{m_1! m_2! \cdots m_n!} \cdot f^{(m_1 + \cdots + m_n)}(g(x)) \cdot \prod_{j=1}^{n} \left( g^{(j)}(x) \right)^{m_j},
\]

where the sum is over the set \( S_n \) consisting of all the \( n \)-tuples of non-negative integers \( (m_1, \ldots, m_n) \) satisfying the following constraint:

\[
1 \cdot m_1 + 2 \cdot m_2 + 3 \cdot m_3 + \cdots + n \cdot m_n = n.
\]

### C.3 Laplace’s Method

Laplace’s Method is useful in estimating integrals of the form \( I(\lambda) = \int_a^b e^{-\lambda p(t)} q(t) dt \) as \( \lambda \to \infty \). For two functions \( f_1, f_2 \), we write \( f_1(\lambda) \sim f_2(\lambda) \) if \( f_1(\lambda)/f_2(\lambda) \) tends to unity as \( \lambda \to \infty \). The following result can be found in e.g. [32, Section 3.7].

**Lemma 18** (Laplace’s Method). Suppose that

1. \( p(t) > p(a) \) for \( t \in (a, b) \) and the minimum of \( p(t) \) is only approached at \( t = a \).
2. \( p'(t), q'(t) \) are continuous in a neighborhood of \( t = a \) except possibly at \( t = a \).

45
(3) As $t \to a^+$,

$$p(t) \sim p(a) + \sum_{k=0}^{\infty} p_k(t-a)^{k+\mu}, \quad q(t) \sim \sum_{k=0}^{\infty} q_k(t-a)^{k+\sigma-1},$$

where $\mu, \sigma > 0$, $p_0 \neq 0$ and $q \neq 0$. Also assume that we can differentiate $p(t)$ and

$$p'(t) \sim \sum_{k=0}^{\infty} (k+\mu)p_k(t-a)^{k+\mu-1}.$$

(4) $\int_a^b e^{-\lambda p(t)} q(t) dt$ converges absolutely for all sufficiently large $\lambda$.

Then, we have

$$I(\lambda) = \int_a^b e^{-\lambda p(t)} q(t) dt \sim e^{-\lambda p(a)} \sum_{k=0}^{\infty} \frac{a_k}{\lambda^{k+\sigma}} \Gamma\left(\frac{k+\sigma}{\mu}\right), \quad \text{as } \lambda \to \infty,$$

where $v = p(t) - p(a)$ and

$$f(v) = \frac{q(t)}{p'(t)} \sim \sum_{k=0}^{\infty} a_k v^{\frac{k+\sigma-\mu}{\mu}}, \quad \text{as } v \to 0^+.$$