The University of Melbourne

Faculty of Science

ON THE LOCAL LIMITING BEHAVIOUR OF SOME GROWING POINT PROCESSES

Thesis MSc Mathematics and Statistics

Author:
Jingyi Zhang

Supervisor:
Prof. Konstantin Borovkov

October 18, 2021

Acknowledgements

First and foremost, my deepest thanks to my supervisor Prof. Konstantin Borovkov. Thank you for your ideas, insights and patience. Thank you for your encouragement and dissatisfaction throughout my study. Thank you for believing that there exists a switch behind my clumsy brain. Thank you for your attention to my living habits and mental health. Thank you for reminding me to smile every day. To Prof. Aihua Xia, Dr Xi Geng and Dr Nathan Ross, thank you for the beneficial courses over the years. Thank Kostya, Aihua and Xi, for treating every question from dumb me seriously. Thank every author whose work appears in the reference list.

Contents

Chapter 1

Introduction

The first look into local limit behavior of sums of sequences of random variables was taken by A. de Moivre in 1718 [69] and P. S. de Laplace in 1812 [59]. But the study of Local Limit Theorems (LLTs) hadn't drew further attention until it was revisited by von Mises in 1931 [104].

B. V. Gnedenko [47], in 1948, proved a LLT for the sequences of i.i.d. lattice random variables and its refinements in several aspects. A multi-dimensional version of Gendenko's Theorem was proved by E. Rvačeva in 1954 [86]. The LLTs for sums of non-identically distributed independent random variables were studied by Yu. Prokhorov [83], Yu. Rozanov [84], N. Gamkrelidze [45], A. Mukhin [73], D. McDonald [64][65], and others from 1954 to recent years.

A remarkable refinement of the LLTs are the so-called Integro-Local Limit Theorems (I-LLTs), first obtained by C. Stone in 1965 [98]. The I-LLTs are superior in the sense that the CLTs and the LLTs can be deduced from the I-LLTs but not vice versa. In terms of precision, the I-LLTs are much stronger than the CLTs: the I-LLTs have the same accuracy as LLTs but without making any assumptions about the existence of densities. Most of the work done on the I-LLTs in the 20th century is due to C. Stone, and refinements were mainly studied by A. A. Borovkov starting from 1999. To the best of my knowledge, all studies for the I-LLTs that have appeared so far are restricted to the case of sums of i.i.d. random vectors, except for one work on the triangular array scheme by A. A. Borovkov in 2011 [20].

In this research project, we derived some Integro-local theorems for sums of non-identically distributed random vectors involved in a growing point process model developed from the Neyman contagious model. The process starts with an initial configuration of random points $\boldsymbol{X}_{0,1}, \ldots, \boldsymbol{X}_{0,k_0}$ in $\mathbb{R}^d$ and evolves in the following way. At each time step $n \geq 1$, a "mother" point will be chosen at random from the existing points with probabilities proportional to the weights of these points.

It is denoted by $\boldsymbol{X}^*_{n-1}$. Then, k_n new points with weights $w_{n,1}, \ldots, w_{n,k_n}$ will be added to the process displaced relative to $\boldsymbol{X}^*_{n-1}$ by the respective random vectors $\boldsymbol{Y}_{n,1}, \ldots, \boldsymbol{Y}_{n,k_n}$. Under some natural assumptions about the sequences of displacements and weights, we derive the integro-local limit behavior for $\boldsymbol{X}^*_n$. We obtained results for two different cases of the distribution of $\boldsymbol{Y}_{n,j}$: lattice and non-lattice. The derivations involve the use of several results obtained by the previously mentioned authors, such as B. Gnedenko, E. Rvačeva, C. Stone and D. McDonald.

Chapter 2

From Local to Integro-Local Limit Theorems

The study of Local Limit Theorems was originally inspired by B. Pascal and J. Bernoulli's works published in 1665 [76] and 1713 [10], respectively. They proved that the p.m.f. of the binomial distribution has the form $\binom{n}{m}p^m(1-p)^{n-m}$ for $p \in (0,1)$ and $m = 0, 1, \ldots, n$, based on which J. Bernoulli proved the first *Law of Large Numbers* (LLN) in history.

Inspired by J. Bernoulli's work, A. de Moivre in 1738 [70] and P. S. de Laplace in 1812 [59] proved the first *Local Limit Theorem* (LLT), and hence the first *Integral Limit Theorem* (ILT), both in the case of the binomial distribution. We will discuss these results in Section 2.1.

In 1948, B. V. Gnedenko [47] proved a LLT that holds for sums of independent and identically distributed (i.i.d.) random variables with a lattice distribution in the domain of attraction of a stable law. A version of this theorem in the multi-dimensional case was proved by E. Rvačeva [86]. We will discuss these results and their further refinements in Section 2.2.

As a remarkable refinement of the LLTs, C. Stone [98] proved in 1965 a LLT that is applicable to i.i.d. sequences with a non-lattice distribution and then extended the scope to any distribution in 1966 [99]. This is called the *Integro-local limt Theorem* (I-LLT) which is superior to both the LLT and ILT. We will discuss these results and their further refinements in Section 2.3.

Several results presented in Section 2.2 and 2.3 will play essential role in the derivations in Chapter 4.

2.1 The First LLT: De Moivre-Laplace

In this section, we consider a binomial distribution with parameters n and p. In 1738, A. de Moivre [70] was the first to study the behaviour of the local probabilities as a function of k, for large n, in a particular case when $p = 1/2$. In 1812, P. S. de Laplace [59] extended A. de Moivre's results to the general case of $p \in (0, 1)$. Although A. de Moivre and P. S. de Laplace didn't use the modern language term "Bernoulli trials", the basic model they set up was the same. These are best expressed in terms of random variables and of the probabilities of events connected with them.

Consider a (discrete) probabilistic model of n independent experiments with two outcomes, each or a *Bernoulli scheme*, constructed on the probability space

$$\begin{aligned} &(\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \quad \text{with} \quad \Omega_n = \{\omega : \omega = [a_1, \dots, a_n], a_i = 0, 1\}, \\ &\mathcal{A}_n = \{A : A \subseteq \Omega_n\}, \quad \mathbb{P}_n(\omega) = p^{\sum a_i} q^{n-\sum a_i} \quad (q = 1-p,\ \omega \in \Omega_n), \end{aligned} \tag{2.1}$$

where p is a constant in $(0, 1)$. We introduce Bernoulli random variables $\xi_{n1}, \dots, \xi_{nn}$ by taking $\xi_{ni}(\omega) = a_i$, $i = 1, \dots, n$, where $\omega = [a_1, \dots, a_n]$. It follows from (2.1) that the Bernoulli random variables $\xi_{ni}(\omega)$ are i.i.d. with $\mathbb{P}_n(\xi_{ni} = 1) = p$ and $\mathbb{P}_n(\xi_{ni} = 0) = q$, for $i = 1, \dots, n$. It is natural to think of ξ_{ni} as describing the result of an experiment at the i-th stage. We are interested in the sum of ξ_{ni} up to some stage $k \le n$. Put $S_{n0}(\omega) \equiv 0$ and $S_{nk} = \xi_{n1} + \dots + \xi_{nk}$, for $k = 1, \dots, n$. For notational simplicity, we will write S_n for S_{nn}, $\mathbb{P}$ for $\mathbb{P}_n$, and $\mathbb{E}$ for $\mathbb{E}_n$. As a sum of n i.i.d. Bernoulli random variables with success probability p, we know that S_n has distribution $Binomial(n, p)$ and hence has the p.m.f.[1]

$$P_n(k) := \mathbb{P}(S_n = k) = \binom{n}{k} p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k q^{n-k}, \qquad k = 0, \dots, n, \tag{2.2}$$

for any $n \ge 1$. From above, we know that $\mathbb{E}S_n = np$ and consequently

$$\mathbb{E}\,\frac{S_n}{n} = p. \tag{2.3}$$

In other words, the mean value of the frequency of "success" (i.e., S_n/n), coincides with the probability p of success. Hence we are led to ask the following question:

[Q1] How much does the frequency S_n/n of success differ from its probability p?

It turns out that this question is not simple: one cannot expect that there will exist a small $\varepsilon \in (0, 1)$ such that for all $\omega \in \Omega_n$, for sufficiently large n, the deviation of S_n/n

[1] The expression (2.2) for probabilities of $\{S_n = k\}$ was firstly derived by Pascal [76] in 1665 for a special case when $p = 1/2$. Jacob Bernoulli [10] in 1713 proved (2.2) is valid for any $p \in (0, 1)$.

from p is less than ε. One can see this is impossible simply from the observations $\mathbb{P}(S_n/n = 1) = p^n > 0$ and $\mathbb{P}(S_n/n = 0) = q^n > 0$ for any $n \in \mathbb{N}$ and $p \in (0,1)$. Nevertheless, for large n, it is natural to expect that the probability of the event $\{\omega : |S_n(\omega)/n - p| > \varepsilon\}$ is small. It was proved by J. Bernoulli [2] [10] in 1713 that, for any fixed $\varepsilon > 0$, as $n \to \infty$, this probability vanishes:

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) = \sum_{\{k:\ |k/n-p|>\varepsilon\}} P_n(k) \to 0. \tag{2.4}$$

This is a special case of what is nowadays called the *weak law of large numbers* (WLLN) in the Bernoulli scheme, denoted as $S_n/n \xrightarrow{p} p$ as $n \to \infty$.

But (2.4) hasn't yet fully answered question **[Q1]**. It is still unclear to us what region $\mathcal{I}_n$ will the probability mass of $S_n/n - p$ be concentrated in, for n sufficiently large. And concerning this region $\mathcal{I}_n$, as $n \to \infty$, we can further ask:

[Q2] How will the local probabilities $P_n(k)$ behave as a function of k?

The two above questions were firstly answered by A. de Moivre [70] in 1738 for a particular case when $p = 1/2$ and extended to the case of general $p \in (0,1)$ by P. S. de Laplace [59] in 1812. The idea is to first make an educated guess of the region $\mathcal{I}_n$ from the WLLN (2.4). Recall that, inspired by (2.3), J. Bernoulli proved (2.4) to be true. Under a similar idea, together with the observation that $\mathbb{E}\left(S_n/n - p\right)^2 = (pq)/n$, it is reasonable to expect that, for any $x \in \mathbb{R}$, the relation

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| > x\sqrt{\frac{pq}{n}}\right) \to \epsilon(x) \qquad \text{as } n \to \infty \tag{2.5}$$

will hold, where $\epsilon(x) \geq 0$ is small and the LHS, by (2.2), equals

$$\mathbb{P}\left(\left|\frac{S_n - \mathbb{E}S_n}{\sqrt{\operatorname{Var} S_n}}\right| > x\right) = \sum_{\left\{k:\ |(k-np)/\sqrt{npq}|>x\right\}} P_n(k). \tag{2.6}$$

The region of summation on the RHS of (2.6) indicates that a reasonable guess of the region $\mathcal{I}_n$ is the set $\{k : |k - np| = O\left(\sqrt{npq}\right)\}$.

For those k satisfying $|k - np| = O\left(\sqrt{npq}\right)$, A. de Moivre ($p = 1/2$) and P. S. de Laplace ($p \in (0,1)$) derived the asymptotic behaviour of $P_n(k)$. A proof can be found in W. Feller's book [39] or Y. G. Sinai's book [91, pp. 30–32].

As a refinement, it was shown (by asymptotic expansion) that the same asymptotic behavior of $P_n(k)$ will still hold for those k in the slightly enlarged region

[2] This is the first limit theorem in proabability theory. It appeared in the fourth part of *Ars conjectandi* [10] which later has many impacts on A. de Moivre's work. A English version see [11],

$\left\{k : |k - np| = o\,(npq)^{2/3}\right\}$. A proof can be found in A. N. Shiryaev's book [90, pp. 55–57]. One can see that the orginal version and the refined version can be proved by essentially the same argument. So, we will only present the refined version.

Theorem 2.1 (the De Moivre-Laplace Local Limit theorem)**.** *Let* $0 < p < 1$*, then*

$$\mathbb{P}(S_n = k) \equiv P_n(k) = \frac{1}{\sqrt{2\pi npq}} e^{-(k-np)^2/(2npq)} \left(1 + \varepsilon(n,k)\right), \tag{2.7}$$

where the remainder term $\varepsilon(n,k) \to 0$ *as* $n \to \infty$ *uniformly in* k *such that* $|k - np| = o\,(npq)^{2/3}$*. Or equivalently, for* $\Delta(n) = o\,(npq)^{2/3}$*, one has*

$$\sup_{\{k:|k-np|\le\Delta(n)\}} \left| \frac{P_n(k)}{\frac{1}{\sqrt{2\pi npq}} e^{-(k-np)^2/(2npq)}} - 1 \right| \to 0, \qquad \text{as } n \to \infty. \tag{2.8}$$

Proof. The proof starts with applying Stirling's formula [3]

$$n! = \sqrt{2\pi n} e^{-n} n^n \left(1 + R(n)\right), \qquad \text{where } \lim_{n\to\infty} R(n) = 0, \tag{2.9}$$

to the binomial coefficient $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ in (2.2). Then if $n \to \infty$, $k \to \infty$ and $n - k \to \infty$, we have

$$\begin{aligned}
\frac{n!}{k!(n-k)!} &= \frac{\sqrt{2\pi n} e^{-n} n^n}{\sqrt{2\pi k} e^{-k} k^k \cdot \sqrt{2\pi(n-k)} e^{-(n-k)} (n-k)^{n-k}} \frac{1 + R(n)}{(1 + R(k))(1 + R(n-k))} \\
&= \frac{1}{\sqrt{2\pi n \frac{k}{n}\left(1 - \frac{k}{n}\right)}} \frac{1}{\left(\frac{k}{n}\right)^k \left(1 - \frac{k}{n}\right)^{n-k}} \left(1 + \varepsilon_0(n,k)\right),
\end{aligned}$$

where $\varepsilon_0 = \varepsilon_0(n,k)$ satisfies $1 + \varepsilon_0(n,k) = \frac{1+R(n)}{(1+R(k))(1+R(n-k))}$. From (2.9), one has $\varepsilon_0 \to 0$ if $n \to \infty$, $k \to \infty$ and $n - k \to \infty$. Therefore, after changing of variable $\hat{p} = k/n$, (2.2) becomes

$$\begin{aligned}
P_n(k) &= \frac{1}{\sqrt{2\pi n\hat{p}(1-\hat{p})}} \left(\frac{p}{\hat{p}}\right)^k \left(\frac{1-p}{1-\hat{p}}\right)^{n-k} (1 + \varepsilon_0) \\
&= \frac{1}{\sqrt{2\pi n\hat{p}(1-\hat{p})}} \exp\left\{-nH\left(\hat{p}\right)\right\}(1 + \varepsilon_0),
\end{aligned}$$

[3] This formula was firstly discovered by A. de Moivre with an unsolved constant scale and named after Stirling for his contribution in showing that the constant is $\sqrt{2\pi}$. Stirling's formula is well-studied in the literature. For example, one can find classical proofs in W. Rudin's *Principles of Mathematical analysis* [85, pp. 195] or in R.C. Buck's *Advanced Calculus* [30, pp. 216–218]; a proof requiring very little mathematical background by Feller [38]; two easy proofs in [17]; a probabilistic proof using the classical CLT appeared in 1971 [56].

where $H(x) := x \ln \frac{x}{p} + (1-x) \ln \frac{1-x}{1-p}$.

Since we are considering values of k such that $|k - np| = o(npq)^{2/3}$ which guarantees $p - \hat{p} \to 0$, $n \to \infty$, it is natural to write $H(\hat{p})$ in the form $H(p + (\hat{p} - p))$ and use Taylor's expansion about p. One has

$$\begin{aligned} H(\hat{p}) &= H(p) + H'(p)(\hat{p} - p) + \frac{1}{2} H''(p)(\hat{p} - p)^2 + O\left(|\hat{p} - p|^3\right) \\ &= \frac{1}{2}\left(\frac{1}{p} + \frac{1}{q}\right)(\hat{p} - p)^2 + O\left(|\hat{p} - p|^3\right), \end{aligned}$$

where the last relation is obtained by substituting the derivatives $H'(x) = \ln(x/p) - \ln[(1-x)/(1-p)]$, $H''(x) = 1/x + 1/(1-x)$ and $H'''(x) = -1/x^2 + 1/(1-x)^2$ which hold for $0 < x < 1$. Consequently, one has

$$P_n(k) = \frac{1}{\sqrt{2\pi n \hat{p}(1-\hat{p})}} \exp\left\{-\frac{n}{2pq}(\hat{p} - p)^2\right\} \exp\left\{nO\left(|\hat{p} - p|^3\right)\right\}(1 + \varepsilon_0)$$

By noticing that $n(\hat{p} - p)^2 = n(k/n - p)^2 = (k - np)^2/n$, one has

$$P_n(k) = \frac{1}{\sqrt{2\pi npq}} e^{-(k-np)^2/(2npq)}(1 + \varepsilon(n,k))$$

where $1 + \varepsilon(n,k) = (1 + \varepsilon_0(n,k)) \exp\{nO(|\hat{p} - p|^3)\} \sqrt{p(1-p)/[\hat{p}(1-\hat{p})]}$.

To complete the proof, it suffices to observe that if $\Delta(n) = o(npq)^{2/3}$, then $\sup_{\{k:\ |k-np| \le \Delta(n)\}} |\varepsilon(n,k)| \to 0$ as $n \to \infty$, since for these values of k, all three quantities k, $n - k$ and n tend to infinity. □

An equivalent form of the De Moivre-Laplace LLT 2.1 can be obtained by changing of the variables setting $x := (k - np)/\sqrt{npq}$.

Corollary 2.2. *Let $0 < p < 1$, then for all $x \in \mathbb{R}$ such that $x = o(npq)^{1/6}$ and $np + x\sqrt{npq}$ is an integer from the set $\{0, 1, \ldots, n\}$, one has*

$$\mathbb{P}\left(\frac{S_n - \mathbb{E}S_n}{\sqrt{\operatorname{Var} S_n}} = x\right) = P_n(np + x\sqrt{npq}) = \frac{1}{\sqrt{2\pi npq}} e^{-x^2/2}(1 + \varepsilon(n)), \tag{2.10}$$

where $\varepsilon(n) \to 0$ as $n \to \infty$. Or, equivalently, for $\Lambda(n) = o(npq)^{1/6}$, one has

$$\sup_{\{x:\ |x| \le \Lambda(n)\}} \left| \frac{P_n(np + x\sqrt{npq})}{\frac{1}{\sqrt{2\pi npq}} e^{-x^2/2}} - 1 \right| \to 0, \quad \textit{as } n \to \infty. \tag{2.11}$$

As a natural consequence of the First Local Limit Theorem 2.1, Laplace [59] also proved the following corollary which is the first Integral Limit Theorem in

Probability theory.

Corollary 2.3 (the De Moivre-Laplace Integral Limit Theorem)**.** *Let $0 < p < 1$. Then*

$$\sup_{-\infty \le a < b \le \infty} \left| \sum_{a < x \le b} P_n(np + x\sqrt{npq}) - \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx \right| \to 0 \qquad \text{as } n \to \infty, \quad (2.12)$$

where the summation is over those x such that $np + x\sqrt{npq} \in \mathbb{Z}$. Or, equivalently,

$$\sup_{-\infty \le a < b \le \infty} \left| \mathbb{P}\left(a < \frac{S_n - \mathbb{E}S_n}{\sqrt{\operatorname{Var} S_n}} \le b \right) - \left[\Phi\left(b\right) - \Phi\left(a\right)\right] \right| \to 0 \qquad \text{as } n \to \infty, \quad (2.13)$$

where $S_n \sim Binomial(n, p)$ and Φ is the c.d.f. of the standard normal distribution.

Sketch of the proof. We start with considering the case when $(a, b]$ is bounded such that there exists a $T > 0$ such that $(a, b] \subseteq [-T, T]$. Then by summing up (2.10) over $x \in (a, b]$ satisfying $np+x\sqrt{npq} \in \mathbb{Z}$, and applying Riemann sum approximation with increment size $1/\sqrt{npq}$, one obtains that (2.13) holds uniformly in such a, b. Then, for the case when $(a, b]$ is unbounded, fix an $\epsilon > 0$, find a $T = T(\epsilon)$ such that $\Phi(T) - \Phi(-T) > 1-\epsilon/4$, and find an $N \ge 1$ such that $\left|\mathbb{P}\left(a < (S_n - \mathbb{E}S_n)/(\sqrt{\operatorname{Var} S_n}) \le b\right) - [\Phi\left(b\right) - \Phi\left(a\right)]\right| < \epsilon/4$ holds uniformly for $n \ge N$ and $-T \le a < b \le T$. By the triangle inequality, $\left|\mathbb{P}\left(a < (S_n - \mathbb{E}S_n)/(\sqrt{\operatorname{Var} S_n}) \le b\right) - [\Phi\left(b\right) - \Phi\left(a\right)]\right| \le \epsilon$ holds uniformly for $-\infty \le a \le -T < T \le b \le \infty$. But $\epsilon > 0$ was arbitrarily chosen, so combining these two will give us (2.13). □

Remark 2.4. It is natural to ask how fast the convergence in (2.13) is, as $n \to \infty$. In 1941-2, A. C. Berry [13] and C. G. Esseen [36] independently obtained the following result, which is called *the Berry–Esseen Theorem*. For any sum of independent random variables $X_1, \dots, X_n$, each with zero mean, and respective variance σ_k^2 and $\mathbb{E}|X_k|^3 < \infty$, one has

$$\sup_{-\infty \le x \le \infty} \left| \mathbb{P}\left(\frac{\sum_{k=1}^n X_k}{\sqrt{\sum_{k=1}^n \sigma_k^2}} \le x \right) - \Phi(x) \right| \le \frac{C \sum_{k=1}^n \mathbb{E}|X_k|^3}{(\sum_{k=1}^n \sigma_k^2)^{3/2}}. \quad (2.14)$$

A proof thereof can be found in V. V. Petrov's book [80, pp. 109–120]. One can see that, in the special case when X_k are i.i.d., the upper bound in (2.14) equals $\frac{C\mathbb{E}|X_1|^3}{\sigma_1^3\sqrt{n}}$. Applying (2.14) in our case of $S_n \sim Binomial(n, p)$ in Corollary 2.3, one has

$$\sup_{-\infty \le x \le \infty} \left| \mathbb{P}\left(\frac{S_n - \mathbb{E}S_n}{\sqrt{\operatorname{Var} S_n}} \le x \right) - \Phi(x) \right| \le \frac{C(p^2 + q^2)}{\sqrt{npq}}, \quad (2.15)$$

where the constant C was initially estimated to be 7.59 in 1942, improved to 0.4785 by I. S. Tyurin [103] in 2010. But it has been proved that this order of the estimate $(1/\sqrt{npq})$ cannot be improved, with the lower bound $\frac{D(p^2+q^2)}{\sqrt{npq}}$ where $D = 0.4097$ found by C. G. Esseen in 1956. This means that the approximation in Corollary 2.3 can be poor for values of p that are close to 0 or 1, even when n is large. It turns out that, for small values of p, instead of the normal distribution, the *Poisson distribution* provides a good approximation [90, p. 62].

Further refinements of the De Moivre-Laplace Limit Theorems can be found in [91, pp. 30–40] and [90, pp. 54–69] and quantitative examples can be found in W. Feller's book [39, pp. 187–190].

However, instead of the Local Limit Theorems, it was the area of the Integral Limit Theorems that was actively studied starting from P. S. de Laplace to the 20th century, in which period the classical probability theory had been on its way toward the "modern" one. Especially so after the appearance of A. M. Lyaponov's proofs [61] [62] in 1900–1 (translations in [1]); where for the first time the proof of the CLT using the method of characteristic functions systematically.

Later, Paul Lévy developed the theory of characteristic functions (ch.f.s) and used it to proof the CLT under very general conditions in around 1925. Due to the superiority of this "modern" approach to proving to CLTs for non-normal limit distributions and "non-classical" normings, people tend to use it to replace the "classical" Laplacian approximation method. See H. Fischer's book [40] for more on the history of the CLT.

In the following two sections, we will see how important the appearance of this method of ch.f.'s was for the development of local and integro-local theorems.

2.2 The Gnedenko and Rvačeva Local Limit theorems

The interest in the Local Limit Theorems (LLTs) for integer-valued random variables was revived from the 1930s. Von Mises [104] [105] and G. M. Bawly [6] in 1931 and 1937, respectively, generalized the De Moivre-Laplace Theorem to the case of sums of independent random variables that take integer values and under some additional conditions.

Based on von Mises and G. M. Bawly's work, A. N. Kolmogorov made a natural conjecture that, for a sequence of not only independent but also *identically distributed* random variables, more complete results can be obtained, but he didn't provide a proof. It was his student, B. V. Gnedenko [47], who proved in 1948 a LLT for the sequence of i.i.d. random variables with lattice distribution, finite mean and finite variance. An English version of this theorem and its proof can be found in Chapter 9 of the monograph *Limit Distributions for Sums of Independent Random Variables* [53] written by B. V. Gnedenko and A. N. Kolmogorov in 1949. In this book, not only the Gnedenko's LLT for normal limit laws (§49) and its refinements (§51) are discussed, but also an LLT for non-normal stable limit law is proved in §50 of the book. Another reference during the same period is Gnedenko's own text *Theory of probability* which was published in 1950 (an English translation appeared in 1962 [52]).

Before stating these Theorems, we first give definitions of lattice and arithmetic distributions.

Definition 2.5 (the lattice distribution)**.** A random variable ξ is said to have a *lattice distribution with shift c and span h* if there exist c and h such that

$$\sum_{k=-\infty}^{\infty} \mathbb{P}(\xi = c + kh) = 1, \tag{2.16}$$

where h is the greatest number satisfying (2.16) and c is in the interval $[0, h)$.

Remark 2.6. A lattice distribution with shift $c = 0$ and span $h = 1$ is called *arithmetic*. A random variable with such a distribution takes only integer values and the greatest common divisor (g.c.d.) of all the values it takes with positive probabilities equals 1.

Remark 2.7. For a lattice random variable ξ with arbitrary span h and shift c, and the sum S_n of i.i.d. random variables $\xi_1, \ldots, \xi_n$ with $\xi_i \overset{d}{=} \xi$, the transformations

$$\xi_m^{'} = \frac{\xi_m - c}{h}\,, \qquad S_n^{'} = \frac{S_n - nc}{h}\,, \qquad m \in \{1, \ldots, n\},\ n \in \mathbb{N}, \tag{2.17}$$

allow one to reduce the lattice case to the arithmetic one (cf. Remark 2.6). Hence, one can see that, in essence, *it is sufficient to only consider the arithmetic case* of summands ξ_m, where the sum S_n can only take arithmetic values too. So, from now on, we will only discuss the Local Theorems in their arithmetic versions. One can obtain the corresponding lattice versions trivially by transformations (2.17).

Here is the starting point in B. V. Gnedenko [47]. Suppose the independent summands ξ_m have a common arithmetic distribution, namely, each ξ_m can only take with integer values and the g.c.d. of all values that it takes positive probabilities equals 1. Then the sum $S_n = \xi_1 + \cdots + \xi_n$ takes only integer values too, i.e., the distribution function[4] $F_n(x)$ of the sum S_n is constant in each half-open interval $x \in (k, k+1]$, where $k \in \mathbb{Z}$. So, it is impossible to approximate such a distribution function F_n with continuous (not to say analytic) functions to within one-half of its maximum jump $P_n(k) := F_n(k) - F_n(k-)$. Therefore, the real interest lies only in the study of the values of the function $F_n(x)$ at the points $k \in \mathbb{Z}$ themselves, i.e., the sum $F_n(k) = \sum_{r \le k} P_n(r)$, where $P_n(r) = \mathbb{P}(S_n = r)$. However, it seems most natural to investigate the asymptotic behaviour of the local probabilities $P_n(r)$ from which one can obtain[5] the "integral" (sum) ones.

Theorem 2.8 (the Gnedenko Arithmetic Theorem). *Let i.i.d. random variables $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ have an arithmetic distribution, finite mean $m := \mathbb{E}\xi_1$ and finite variance $\sigma^2 := \operatorname{Var}\xi_1 \neq 0$. Let*

$$a_n := \mathbb{E}S_n = mn, \qquad b_n^2 := \operatorname{Var} S_n = n\sigma^2, \qquad z_{n,k} := \frac{k - a_n}{b_n} = \frac{k - mn}{\sigma\sqrt{n}}.$$

Then

$$\sup_{k \in \mathbb{Z}} \left| b_n P_n(k) - \frac{1}{\sqrt{2\pi}} e^{-\frac{z_{n,k}^2}{2}} \right| \to 0 \qquad \text{as } n \to \infty. \tag{2.18}$$

Or, equivalently, as $n \to \infty$

$$P_n(k) = \frac{1}{b_n\sqrt{2\pi}} e^{-\frac{z_{n,k}^2}{2}} + o\left(\frac{1}{b_n}\right) = \frac{1}{b_n}\phi\left(z_{n,k}\right) + o\left(\frac{1}{b_n}\right) \tag{2.19}$$

uniformly in $k \in \mathbb{Z}$, where ϕ is the univariate standard normal density.

Remark 2.9. We can see that the De Moivre-Laplace LLT 2.1 is a special case of this theorem when the ξ's have the common distribution *Bernoulli*(p).

Remark 2.10. It is natural to ask how fast the convergence in (2.18) is. As a special case of the result from Esseen's paper [37] in 1945, for such an i.i.d. sequence as in the setup of Theorem 2.8, if one further assumes that ξ has a finite third-order absolute moment, then the LHS of (2.18) will have order $O\left(1/\sqrt{n}\right)$. It means that the order $o(1/b_n) = o(1/\sqrt{n})$ in (2.19) can be improved to $O(1/n)$. It has also been shown that this estimate cannot be further improved. Proofs of these are also

[4] In Gnedenko's paper, he adopted $\mathbb{P}(X < x)$ as the definition of disrtribution function of the random varible X at x. But throughout this thesis, we follow the convention to use $\mathbb{P}(X \le x)$ as the definition of the disrtribution function of X at x.

[5] A quote from G&K [53]: "Generally speaking, such a method of proving integral theorems on the basis of local theorems may result in some loss in the precision of the estimates of the reminder terms as compared with direct methods of proof. However, for the first orientation in this problem the approach from the direction of local theorems seems preferable."

provided in Petrov's book [80, VII §3] and Gnedenko & Kolmogorov's monograph [53, §51].

Proof of Theorem 2.8. The proof uses the method of ch.f.'s. Let ψ_{S_n} be the ch.f. of S_n and ψ_{ξ_1} be the ch.f. of ξ_1. We start with using the inversion formula to give an integral expression for the probability $P_n(k)$. Since S_n is arithmetic, we prefer to use the inversion formula[6] for the p.g.f. $p_{S_n}(z) := \mathbb{E}z^{S_n}$ of S_n, which gives

$$P_n(k) = \frac{1}{2\pi i}\oint_{|z|=1} z^{-k-1}p_{S_n}(z)dz = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-itk}\psi_{S_n}(t)dt, \tag{2.20}$$

where the second relation is obtained by changing the variables $z = e^{it}$ and using relation $p_{S_n}(e^{it}) = \psi_{S_n}(t)$. But $\xi_1, \xi_2, \ldots, \xi_n$ are assumed to be i.i.d., which implies that $\psi_{S_n} = \psi_{\xi_1}^n$. So (2.20) becomes

$$P_n(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-itk}\psi_{\xi_1}^n(t)dt. \tag{2.21}$$

Note that for t away from 0, the integral is easy to approximate. So, we split the integral on the RHS of (2.21) into two subintegrals: over the domain $|t| < \gamma$ and over the complementary set, where $\gamma = \gamma(\sigma)$ is a small constant chosen in a way that will later become clear after (2.23). The first subintegral is

$$I_1 := \int_{|t|<\gamma} e^{-itk}\psi_{\xi_1}^n(t)dt = \int_{|t|<\gamma} e^{-ity}\left[e^{-itm}\psi_{\xi_1}(t)\right]^n dt$$

for $y = k - mn$. Recall that $m := \mathbb{E}\xi_1$. Put $u = \sigma\sqrt{n}t$, then

$$\begin{aligned} I_1 &= \int_{|u|<\gamma\sigma\sqrt{n}} e^{-i\frac{u}{\sigma\sqrt{n}}y}\left[e^{-i\frac{u}{\sigma\sqrt{n}}m}\psi_{\xi_1}\left(\frac{u}{\sigma\sqrt{n}}\right)\right]^n \frac{du}{\sigma\sqrt{n}} \\ &= \frac{1}{\sigma\sqrt{n}}\int_{|u|<\gamma\sigma\sqrt{n}} e^{-i\frac{u}{\sigma\sqrt{n}}y}\ \psi_{\frac{\xi_1-m}{\sigma}}^n\left(\frac{u}{\sqrt{n}}\right)du, \end{aligned} \tag{2.22}$$

where $\psi_{\frac{\xi_1-m}{\sigma}}$ is the ch.f. of $\frac{\xi_1-m}{\sigma}$ which has mean 0 and variance 1. Since $\mathbb{E}|\frac{\xi_1-m}{\sigma}|^2 = 1 < \infty$, in a proper neighbourhood $(-\gamma\sigma, \gamma\sigma)$ of the point 0, one has the expansion[7]

$$\psi_{\frac{\xi_1-m}{\sigma}}(s) = 1 + it\mathbb{E}\frac{\xi_1 - m}{\sigma} - \frac{s^2}{2}\mathbb{E}\left(\frac{\xi_1-m}{\sigma}\right)^2 + o\left(s^2\right) = 1 - \frac{s^2}{2} + o\left(s^2\right). \tag{2.23}$$

[6] The inversion formula for ch.f.'s is $P_n(k) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itk}\psi_{S_n}(t)dt$, one can see from the region of integration why (2.20) is easier to start with. Proofs of both inversion formulae can be found in Section 7.2.1 of [21].

[7] This expansion is a basic property of ch.f. whose proof can be found in [21, p.156] or [39].

Further, using relations $\ln(x^n) = n\ln x$ and $\ln(1+x) = x + o(x)$ as $x \to 0$, one has

$$\ln\psi^n_{\frac{\xi_1 - m}{\sigma}}(s) = n\ln\left[1 - \left(1 - \psi_{\frac{\xi_1 - m}{\sigma}}(s)\right)\right] = n\left(\psi_{\frac{\xi_1 - m}{\sigma}}(s) - 1\right) = -\frac{s^2}{2}n + o\left(s^2\right)n,$$

which applies well at points $s = u/\sqrt{n}$ with $|u| < \gamma\sigma\sqrt{n}$. Denote the remainder term in the above formula, after substitution $s = u/\sqrt{n}$, by $h_n(u)$. Substituting the above expression for $\ln\psi^n_{\frac{\xi_1 - m}{\sigma}}(u/\sqrt{n})$ back into (2.22), one has

$$I_1 = \frac{1}{\sigma\sqrt{n}}\int_{|u|<\gamma\sigma\sqrt{n}} e^{-i\frac{u}{\sigma\sqrt{n}}y}\exp\left\{-\frac{u^2}{2} + h_n(u)\right\}du, \tag{2.24}$$

where $h_n(u) \to 0$ for any fixed u as $n \to \infty$. Moreover, for γ small enough, in the domain $|u| < \gamma\sigma\sqrt{n}$, one has $|h_n(u)| \le u^2/6$, so $-u^2/2 + h_n(u)$ does not exceed $-u^2/3$. Hence, as $n \to \infty$, by applying the D.C.T. to the sequence

$$\mathbf{1}_{\{|u|<\gamma\sigma\sqrt{n}\}}e^{-i\frac{u}{\sigma\sqrt{n}}y}e^{-\frac{u^2}{2}+h_n(u)} = e^{-iu\frac{y}{\sigma\sqrt{n}}}e^{-\frac{u^2}{2}}\left(1 + o(1)\right),$$

one has

$$\sigma\sqrt{n}I_1 = \int_{\mathbb{R}} e^{-iu\frac{y}{\sigma\sqrt{n}}}e^{-\frac{u^2}{2}}du + o(1),$$

where the integral on the RHS is in the form of the inversion formula for the ch.f. of the univariate standard normal distribution evaluated at $y/\sigma\sqrt{n}$, up to the factor $1/2\pi$. Hence

$$b_n I_1 = \sqrt{2\pi}e^{-\frac{z^2_{n,k}}{2}} + o(1) \tag{2.25}$$

where $b_n := \sigma\sqrt{n}$ and $z_{n,k} := \frac{k - mn}{\sigma\sqrt{n}} = \frac{y}{\sigma\sqrt{n}}$.

The case of the remaining integral

$$I_2 := \int_{|t|\ge\gamma} e^{-itk}\psi^n_{\xi_1}(t)dt$$

is easier, since the range of integration $\{t : \gamma \le |t| \le \pi\}$ is compact and hence on that one has $|\psi_{\xi_1}| \le q(\gamma) < 1$. Since $q^n(\gamma) = o(n^{-1/2})$, one has

$$|I_2| \le \int_{|t|\ge\gamma} e^{-itk}q^n(\gamma)dt = o\left(\frac{1}{\sqrt{n}}\right)\int_{|t|\ge\gamma} e^{-itk}dt = o\left(\frac{1}{\sqrt{n}}\right),$$

which implies

$$b_n I_2 = o(1) \qquad \text{as } n \to \infty. \tag{2.26}$$

Combining (2.25) and (2.26), we have established that, as $n \to \infty$, the relation

$b_n(I_1+I_2)=\sqrt{2\pi}e^{-\frac{z_{n,k}^2}{2}}+o(1)$. Together with the relation $I_1+I_2=2\pi P_n(k)$ from (2.21), one has

$$b_nP_n(k)=\frac{1}{\sqrt{2\pi}}e^{-\frac{z_{n,k}^2}{2}}+o(1) \qquad \text{as } n\to\infty$$

uniformly in $k\in\mathbb{Z}$, by which we proved (2.18) and (2.19). □

Next, it is natural to ask whether the same type of behaviour as in Gnedenko's Arithmetic Theorem 2.8 will be obtained for non-normal stable limit distribution or not. It turns out that the answer is yes, proved by B. V. Gnedenko [51] in 1949.

Theorem 2.11 (the Gnedenko Arithmetic Non-Normal Theorem). *Let the i.i.d. random variables $\xi_1,\xi_2,\dots,\xi_n,\dots$ take only arithmetic values, let $g(y)$ be the density of a certain non-normal stable*[8] *distribution $G(y)$, and let a_n and b_n be certain constants. Then*

$$b_nP_n(k)-g\left(\frac{k-a_n}{b_n}\right)\to 0 \qquad \text{as } n\to\infty \tag{2.27}$$

uniformly in $k\in\mathbb{Z}$, if and only if the distribution of summands ξ_n belongs to the domain of attraction of the stable law $G(y)$ and $\mathbb{P}(S_n\le b_ny+a_n)\Rightarrow G(y)$ as $n\to\infty$.

Proof. It can be proved by the same argument as in the proof of *the Gnedenko Arithmetic Theorem 2.8*. □

One can find illustrations of Theorem 2.11 by two examples, one of which is the Cauchy law, in the monograph [53, pp. 236–237] by B. V. Gnedenko and A. N. Kolmogorov.

Another natural question is whether the same type of behaviour can be obtained for the non-lattice case. A complete discussion of that case was done a bit later, in 1965, by C. Stone [98]. Before C. Stone's work, B. V. Gnedenko [48, 49] proved a Local Limit Theorem for absolute continuous cases in 1953–4. The statement is as follows.

Theorem 2.12 (the Gnedenko Density Theorem). *Let the i.i.d. random variables $\xi_1,\xi_2,\dots,\xi_n,\dots$ have finite mean $m:=\mathbb{E}\xi_1$ and finite variance $\sigma^2:=\operatorname{Var}\xi_1\neq 0$. Denote the density of the normalized sum $Z_n:=\frac{1}{\sigma\sqrt{n}}\sum_{j=1}^n(\xi_j-m)$ by $f_n(x)$, if it exists. Then*

$$f_n(x)-\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\to 0 \qquad \text{as } n\to\infty \tag{2.28}$$

uniform in $x\in\mathbb{R}$, if and only if there exists an N such that Z_N has density $f_N(x)$ such that $\sup_{x\in\mathbb{R}} f_N(x)<\infty$.

[8]See the definition of *stable distribution* in [21, p. 233]

Remark 2.13. Under the same conditions of Theorem 2.12, i.e. assume an i.i.d. sequence with non-zero variance with $f_N(x)$ uniformly bounded for some N, if one further assumes that $\mathbb{E}|\xi_1|^k < \infty$ for some integer order $k \geq 3$, one can obtain a more accurate approximation (by asymptotic expansion)

$$f_n(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} + \sum_{\nu=1}^{k-2} \frac{q_\nu(x)}{n^{\nu/2}} + o\left(\frac{1}{n^{(k-2)/2}}\right)$$

uniformly in x, where $q_\nu(x) = N_{3\nu}(x) \exp\{-x^2/2\}$ with $N_{3\nu}(x)$ being a polynomial of degree 3ν in x. This is a result originally from the paper [54] written by Gnedenko with Kolmogorov in 1959, but also can be found in Petrov's book [80, VII §3].

One may also be interested in the multi-dimensional version of Theorem 2.8. In 1954, E. L. Rvačeva proved a version of Gnedenko's Arithmetic Theorem 2.8 in $\mathbb{R}^d$ (an English translation [86] appeared in 1962). The following are three necessary assumptions, where the third one can be treated as a multivariate anolog of the "arithmetic assumption" in Theorem 2.8.

Assumption 2.1. *The summands* $\boldsymbol{\xi}^{(n)} = \left[\xi_1^{(n)}, \xi_2^{(n)}, \ldots, \xi_d^{(n)}\right], n = 1, 2, \ldots$ *are i.i.d. random vectors in* $\mathbb{R}^d$.

Assumption 2.2. *All entries of the common mean vector* $\boldsymbol{m}$ *and covariance matrix* Σ *are finite, namely* $m_j := \mathbb{E}\xi_j^{(n)} < \infty$, $\Sigma_{ij} := \mathrm{Cov}\left(\xi_i^{(n)}, \xi_j^{(n)}\right) < \infty$, *and with non-zero determinant* $\det \Sigma$.

Let $P(\boldsymbol{k}) = P(k_1, k_2, \ldots, k_d)$ be the probability of the event $\{\boldsymbol{\xi}^{(n)} = \boldsymbol{k}\}$, and

$$\boldsymbol{S}^{(n)} = \left[S_1^{(n)}, S_2^{(n)}, \ldots, S_d^{(n)}\right] := \sum_{i=1}^{n} \boldsymbol{\xi}^{(i)} = \left[\sum_{i=1}^{n} \xi_1^{(i)}, \sum_{i=1}^{n} \xi_2^{(i)}, \ldots, \sum_{i=1}^{n} \xi_d^{(i)}\right]. \tag{2.29}$$

Assumption 2.3. *Let* S *be the class of all* d*-dimensional simplexes with vertices* $\boldsymbol{x} \in \mathbb{Z}^d$ *such that* $P(\boldsymbol{x}) > 0$ *and* $A := \{\mathrm{Vol}(B)d! : B \in S\}$, *then* $\mathrm{g.c.d.}(A) = 1$.

Theorem 2.14 (the Rvačeva Arithmetic Theorem)**.** *Let the* d*-dimensionmal random vectors* $\boldsymbol{\xi}^{(1)}, \boldsymbol{\xi}^{(2)}, \ldots, \boldsymbol{\xi}^{(n)}, \ldots$ *satisfy assumptions 2.1—2.3 with the components* $\xi_j^{(n)}$ *only taking integer values. Then, when* $n \to \infty$, *the local probability* $P_n(\boldsymbol{k}) := \mathbb{P}\left(\boldsymbol{S}^{(n)} = \boldsymbol{k}\right)$ *of the sum* $\boldsymbol{S}^{(n)}$ *defined in (2.29) behaves as*

$$P_n(\boldsymbol{k}) = \frac{1}{(2\pi n)^{d/2} (\det \Sigma)^{1/2}} \exp\left\{-\frac{(\boldsymbol{k} - n\boldsymbol{m})\Sigma^{-1}(\boldsymbol{k} - n\boldsymbol{m})^T}{2n}\right\} + o\left(n^{-d/2}\right), \tag{2.30}$$

where the remainder term is uniform in $\boldsymbol{k} \in \mathbb{Z}^d$.

Proof. It can be proved by following a multivariate version of the argument used in the proof of Theorem 2.8. Details can be found in E. Rvačeva's paper [86]. $\square$

Remark 2.15. An equivalent condition to Assumption 2.3 is: the parallelepiped lattice[9] generated by set $D := \{\boldsymbol{x} = \boldsymbol{x}_1 - \boldsymbol{x}_2 : P(\boldsymbol{x}_1) > 0, P(\boldsymbol{x}_2) > 0\}$ will contain all points with positive probability in $\mathbb{Z}^d$.

Remark 2.16. One can see that the Gnedenko Arithmetic Theorem 2.8 is a special case of the above Rvačeva Arithmetic Theorem 2.14 when $d = 1$.

Assuming the sequence being i.i.d. was found to be less useful in applications. A. Khintchine illustrated that what will be useful in applications is the LLTs for sums of non-identically distributed random variables, especially in the area of the classical statistical mechanics [57] and quantum statistics [58].

It turns out that, for the non-identical cases, bad things can happen in the periodicity asymptotically, as illustrated in N. G. Gamkrelidze's paper [45] in 1964. A simple example is as follows. Suppose random variables ξ_n has the p.m.f. $f_n(0) = 1/2$, $f_n(1) = 1/2^n$,$f_n(2) = 1/2 - 1/2^n$ and f_n equals zero elsewhere, then the support of this sequence is asymptotically concentrated only on the even integers, which can never coincide with the standard normal distribution.

The LLTs for sums of independent but non-identically distributed random variables, and refinements of these theorems, were studied by Yu. Prokhorov [83], Yu. Rozanov [84], P. Survila [100], N. Gamkrelidze [45], V. Statulevičius [96, 74, 77], A. Mitalauskas [74], V. Pipiras [77], A. Mukhin [73], B. Davis [32], D. McDonald [64, 65] and others from 1954 to 2005.

Here we will only present a relatively new one by D. McDonald [65, 3.1] for its superiority in helping our derivations in Chapter 4.

Theorem 2.17 (McDonald)**.** *Let $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ be a sequence of independent, but not necessarily identically distributed, random variables which only take integer values which p.m.f. p_n. Let $S_n = \xi_1 + \cdots + \xi_n$ and $P_n(k) := \mathbb{P}(S_n = k)$. Suppose there exist numbers a_n and b_n such that, for $t \in \mathbb{R}$, one has*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n - a_n}{b_n} \le t\right) = \Phi(t) \tag{2.31}$$

and

$$\limsup_{n\to\infty} \frac{b_n^2}{\sum_{m=1}^{n} \sum_{k\in\mathbb{Z}} [p_m(k) \wedge p_m(k+1)]} < \infty. \tag{2.32}$$

[9] A *lattice* generated by a set $D \in \mathbb{Z}^d$ is defined as the set $\{\sum_i z_i \boldsymbol{d}_i : z_i \in \mathbb{Z}, \boldsymbol{d}_i \in D\}$

Then, as $n \to \infty$

$$\sup_{k\in\mathbb{Z}} \left| b_n P_n(k) - \phi\left(\frac{k-a_n}{b_n}\right) \right| \to 0. \tag{2.33}$$

The proof of the above theorem largely relies on the results from Davis [32]. For details one can see [65].

We conclude this section with mentioning some further refinements of the LLTs in the case of independent but non-identically distributed random variables.

Sums of non-identically distributed independent random vectors in $\mathbb{R}^d$ have been considered by A. Mukhin [72] in 1984 and N. Gamkrelidze [46] in 2005.

Convergence to stable limit distributions for sums of non-identically distributed random variables have been considered by A. Mitalauskas [67] in 1962 and V. Statulevičius [97] in 1967.

Asymptotic expansions in local theorems in the metric L_p have been studied for various values of p by Yu. Prokhorov [82] in 1952, V. Petrov [78, 79] in 1961–4, S. Siraždinov and M. Mamatov [92, 93] in 1962–3, and by P. Survila [100] in 1963.

Estimates of the remainder term in local theorems for densities have been studied by S. Siraždinov [94] in 1965, V. Statulevičius [96] in 1965 and N. Šahaĭdarova [94, 87] in 1965–6.

A number-theoretic approach to the proof of local theorems for lattice distributions has been developed by G. Freĭman [42] in 1956 and by D. Moskvin, L. Postnikova, and A. Yudin [71] in 1970.

2.3 The Stone Integro-Local Limit Theorems

In this section, we will present a remarkable Integro-Local Limit Theorem (I-LLT) established by C. Stone [98] in 1965, and also its refinements in several aspects by C. Stone [99] in 1966 and A. A. Borovkov [18, 19, 20, 22, 23] from 1999 to 2017.

To study the asymptotic local behaviour for those distributions that are not lattice (Definition 2.5), L. A. Shepp [89](1964) and C. Stone [98](1965) established a version of Local Limit Theorem which is nowadays regarded as *Integro-Local*. The result obtained by Shepp is a special case of the one by Stone when x is fixed. In fact, Stone's result in [98] in 1965 and its refinements [99] in 1966 are applicable to the case of any stable limit law. But we will only focus on cases with normal limiting laws in this project.

Integro-local limit theorems describe the asymptotic behaviour of probabilities $\mathbb{P}(\boldsymbol{S}_n \in (x, x+\Delta])$ as $n \to \infty$ for a fixed $\Delta > 0$, where $\boldsymbol{S}_n$ is a sum of n random vectors in $\mathbb{R}^d$. For the sake of brevity, for $\boldsymbol{x} = [x_1, \ldots, x_d] \in \mathbb{R}^d$ and fixed $\Delta > 0$, we put

$$\Delta(\boldsymbol{x}] \equiv (\boldsymbol{x}, \boldsymbol{x} + \Delta] := \prod_{j=1}^{d} (x_j, x_j + \Delta] \equiv \prod_{j=1}^{d} \Delta(x_j] \text{ [10]} \tag{2.34}$$

and denote by $\phi(\cdot)$ the density of the univariate standard normal distribution. Define a *non-lattice distribution* as a distribution that is not lattice (cf. Definition 2.5).

The following two theorems present Stone's result [98] in the case of normal limit laws: Theorem 2.18 is for the one-dimensional case; Theorem 2.24 is for the multi-dimensional case.

Theorem 2.18 (the Stone Integro-local Theorem: non-lattice, one-dimensional)**.** *Let $\xi_1, \xi_2, \dots$ be a sequence of non-lattice i.i.d. random variables such that $\xi_k \overset{d}{=} \xi$ for all $k \geq 1$, $\mathbb{E}\xi = 0$ and $\mathbb{E}\xi^2 = \sigma^2 < \infty$. Let $S_n := \xi_1 + \dots + \xi_n$, $n \geq 1$. Then, as $n \to \infty$,*

$$\mathbb{P}\left(S_n \in \Delta(x]\right) = \frac{\Delta}{\sigma\sqrt{n}} \phi\left(\frac{x}{\sigma\sqrt{n}}\right) + o\left(\frac{1}{\sqrt{n}}\right), \tag{2.35}$$

where ϕ is the standard normal density and the remainder term $o\left(1/\sqrt{n}\right)$ is uniform in $x \in \mathbb{R}$ and in $\Delta \in [\Delta_L, \Delta_R]$ for any fixed $0 < \Delta_L < \Delta_R < \infty$.

Remark 2.19. If $\Delta = \Delta_n$ grows, then the asymptotics of $\mathbb{P}(S_n \in \Delta(x])$ can be obtained by summing the RHS of (2.35) for, say, $\Delta = 1$ (if $\Delta_n \to \infty$ is integer-valued). One can check that the error term, after summing up, will have size $o(1)$. Thus the integral theorem follows from the integro-local one but not vice versa.

Remark 2.20. Relation (2.35) was extended by A. A. Borovkov and A. A. Mogulskii [18, 19] in the multivariate setting to an asymptotic expansion in the powers of $n^{-1/2}$ and also to the case where Δ_L can be vanishing. By the properties of densities, the RHS of representation (2.35) has the same form as if the random variable $S_n/(\sigma\sqrt{n})$ had the density $\phi(\cdot)+o(1)$, although the existence of the density of S_n is not assumed in the theorem.

Remark 2.21. The proof we present here is slightly different from C. Stone's original proofs [98]. C. Stone firstly proved the theorem under the simplifying assumption $\limsup_{|t|\to\infty}|\psi(t)| < 1$ (the Cramér condition on the ch.f.) and then extends to the general case. His approach is more natural, but we will directly deal with the general case here, due to the page limit. If one is dealing with a case when the Cramér condition is satisfied, then by replacing the "smoothing sum" in (2.36) with

[10] Here we adopt the notation from A. A. Borovkov's papers instead of C. Stone's. C. Stone studies the integro-local theorems from the viewpoint of the n-fold convolution of laws which will be different from our setup in Chapter 4. Also, note that, instead of $\Delta[x]$, A. A. Borovkov uses the interval $\Delta[x) = [x, x + \Delta)$ in his work. This is because the definition he adopted [21] for the *c.d.f. of a random variable* X is $\mathbb{P}(X < x)$, but the definition we adopted in this thesis is $\mathbb{P}(X \leq x)$.

$Z_n = S_n + \eta_\delta$, where $\eta_\delta \sim \mathrm{Unif}[-\delta, 0]$ is independent of S_n, and following the same type of argument as in the following proof, one can obtain a technically simpler proof.

Proof. The proof uses the method of ch.f.'s. To use the inversion formula, we employ the "smoothing method" and consider, along with S_n, the sums

$$Z_n = S_n + \theta\eta, \tag{2.36}$$

where θ is a constant to be chosen later, and η is a random variable independent of $\xi_1, \xi_2, \ldots$ that has ch.f. equal to $\psi_\eta(t) = (1 - |t|)\,\mathbf{1}_{\{|t|\leq 1\}}$[11]. Denote the ch.f. of ξ by ψ and let $\psi_{\theta\eta}(t) = \max\{0, 1-\theta|t|\}$. Then the ch.f. of Z_n is $\psi_{Z_n}(t) = \psi^n(t)\psi_{\theta\eta}(t)$ which is integrable at infinity. Whence, we can apply the inversion formula to express the increments of the c.d.f. $G_n(x)$ of the random variable Z_n as

$$G_n(x+\Delta) - G_n(x) = \mathbb{P}\left(Z_n \in \Delta(x]\right) = \frac{1}{2\pi}\int \frac{e^{-itx} - e^{-it(x+\Delta)}}{it}\psi_{Z_n}(t)dt. \tag{2.37}$$

Let $\eta_\Delta \sim \mathrm{Unif}[-\Delta, 0]$ be a random variable with ch.f. $\psi_{\eta_\Delta}(t) = (1 - e^{-it\Delta})/(it\Delta)$, then the RHS above can be re-written as

$$\mathbb{P}\left(Z_n \in \Delta(x]\right) = \frac{\Delta}{2\pi}\int_{|t|\leq\frac{1}{\theta}} e^{-itx}\psi^n(t)\widehat{\psi}(t)dt, \tag{2.38}$$

where $\widehat{\psi}(t) := \psi_{\eta_\Delta}(t)\psi_{\theta\eta}(t)$ is the ch.f. of the sum of independent random variables η_Δ and $\theta\eta$. As in the proof of Theorem 2.8, here we again split the integral on the RHS of (2.38) into two sub-integrals: I_1 over the domain $|t| < \gamma$ and I_2 over the domain $\gamma \leq |t| \leq 1/\theta$ such that $\mathbb{P}\left(Z_n \in \Delta(x]\right) = \Delta/(2\pi)(I_1 + I_2)$, where γ is a constant to be chosen in a certain way which will become clear later.

Put $x = v\sigma\sqrt{n}$ in both I_1 and I_2, and consider first

$$I_1 := \int_{|t|<\gamma} e^{-itv\sigma\sqrt{n}}\psi^n(t)\widehat{\psi}(t)dt = \frac{1}{\sigma\sqrt{n}}\int_{|u|<\gamma\sigma\sqrt{n}} e^{-iuv}\psi^n\left(\frac{u}{\sigma\sqrt{n}}\right)\widehat{\psi}\left(\frac{u}{\sigma\sqrt{n}}\right)du,$$

where the last equality is by changed variables $u = \sigma\sqrt{n}t$. As in (2.23), since $\mathbb{E}|\xi/\sigma|^2 = 1 < \infty$, so in a neighbourhood of the point $s = 0$, one has the expansion $1 - \psi_{\xi/\sigma}(s) = \frac{s^2}{2} + o(s^2)$. This, together with the relation $\ln(1+x) = x + o(x)$ as $x \to 0$, gives

$$\ln\psi^n_{\xi/\sigma}(s) = n\ln\left[1 - \left(1 - \psi_{\xi/\sigma}(s)\right)\right] = -\frac{s^2}{2}n + o(s^2)n. \tag{2.39}$$

[11] This is also in form of the density of the convolution of two uniform distributions on $[-1/2, 1/2]$.

Relation (2.39) applies well to points in $\{s = u/\sqrt{n} : |u| < \gamma\sigma\sqrt{n}\}$ if γ was chosen appropriately so that $\gamma\sigma$ is close enough to zero. Substituting (2.39) into I_1, one has

$$I_1 = \frac{1}{\sigma\sqrt{n}} \int_{|u|<\gamma\sigma\sqrt{n}} e^{-iuv} \exp\left\{-\frac{u^2}{2} + h_n(u)\right\} \widehat{\psi}\left(\frac{u}{\sigma\sqrt{n}}\right) du, \tag{2.40}$$

where $|\widehat{\psi}(u/(\sigma\sqrt{n}))| \le 1$, $\widehat{\psi}(u/(\sigma\sqrt{n})) \to 1$ and $h_n(u) = o(u^2/n)n \to 0$ for any fixed u as $n \to \infty$. Moreover, for γ small enough, in the domain $|u| < \gamma\sigma\sqrt{n}$ we have $|h_n(u)| \le u^2/6$, so $-u^2/2 + h_n(u)$ does not exceed $-u^2/3$. Hence, we can apply the D.C.T. to the sequence of functions $\mathbf{1}_{\{|u|<\gamma\sigma\sqrt{n}\}} e^{-iuv} \exp\{-u^2/2 + h_n(u)\} \widehat{\psi}(u/(\sigma\sqrt{n}))$, as $n \to \infty$, to get

$$\sigma\sqrt{n} I_1 = \int_{\mathbb{R}} e^{-iuv} e^{-\frac{u^2}{2}} du + o(1) \tag{2.41}$$

uniformly in v, since the integral on the RHS of (2.40) is uniformly continuous in v. But the integral on the RHS of (2.41) is nothing but just, up to the factor $1/(2\pi)$, the result of applying the inversion formula to the ch.f. of the normal distribution, so that

$$I_1 = \frac{1}{\sigma\sqrt{n}} \sqrt{2\pi} e^{-\frac{v^2}{2}} + o\left(\frac{1}{\sqrt{n}}\right) = \frac{2\pi}{\sigma\sqrt{n}} \phi(v) + o\left(\frac{1}{\sqrt{n}}\right), \tag{2.42}$$

where $v = x/(\sigma\sqrt{n})$. As for the integral

$$I_2 := \int_{\gamma \le |t| \le \frac{1}{\theta}} e^{-itx} \psi^n(t) \widehat{\psi}(t) dt,$$

since the domain of integration in I_2 is compact, and so, by the non-latticeness of ξ, one has on it the upper bound $q := \sup_{\gamma \le |t| \le \frac{1}{\theta}} |\psi(t)| < 1$. Therefore

$$|I_2| \le q^n \int_{\gamma \le |t| \le \frac{1}{\theta}} |\widehat{\psi}(t)| dt \le q^n c(\Delta, \theta), \tag{2.43}$$

where $c(\Delta, \theta)$ is bounded and depends on Δ and θ only. Hence $I_2 = o(1/\sqrt{n})$ holds uniformly in x and $\Delta \in [\Delta_L, \Delta_R]$ for any fixed $0 < \Delta_L < \Delta_R < \infty$. We have now established that, for $Z_n = S_n + \theta\eta$, the relation

$$\mathbb{P}(Z_n \in \Delta(x]) = \frac{\Delta}{2\pi}(I_1 + I_2) = \frac{\Delta}{\sigma\sqrt{n}} \phi\left(\frac{x}{\sigma\sqrt{n}}\right) + o\left(\frac{1}{\sqrt{n}}\right) \tag{2.44}$$

holds uniformly in $x \in \mathbb{R}$ and $\Delta \in [\Delta_L, \Delta_R]$ for any fixed $0 < \Delta_L < \Delta_R < \infty$, as $n \to \infty$. It remains to show that (2.44) implies (2.35). Recall the latter states that

$$\mathbb{P}(S_n \in \Delta(x]) = \frac{\Delta}{\sigma\sqrt{n}} \phi\left(\frac{x}{\sigma\sqrt{n}}\right) + o\left(\frac{1}{\sqrt{n}}\right)$$

holds uniformly in $x \in \mathbb{R}$ and $\Delta \in [\Delta_L, \Delta_R]$ for any fixed $0 < \Delta_L < \Delta_R < \infty$, as $n \to \infty$.

This assertion can be proved as follows. Put $\theta := \delta^2\Delta$, where $\delta > 0$ will be chosen later, $\Delta_\pm := (1 \pm 2\delta)\Delta$, $\Delta_\pm(x] := (x, x + \Delta_\pm]$ and $\phi_0 := \max_y \phi(y)$. We first obtain an upper bound for $\mathbb{P}(S_n \in \Delta(x])$. We have

$$\mathbb{P}\left(Z_n \in \Delta_+(x - \Delta\delta]\right) \geq \mathbb{P}\left(Z_n \in \Delta_+(x - \Delta\delta]; |\eta| < 1/\delta\right).$$

On the event $\{|\eta| < 1/\delta\}$, one has $|\theta\eta| < \delta\Delta$, and hence $\{Z_n \in \Delta_+(x - \Delta\delta]\} \cap \{|\eta| < 1/\delta\} \supseteq \{S_n \in \Delta(x]\} \cap \{|\eta| < 1/\delta\}$. This implies

$$\mathbb{P}\left(Z_n \in \Delta_+(x - \Delta\delta]\right) \geq \mathbb{P}\left(S_n \in \Delta(x];\ |\eta| < 1/\delta\right) = \mathbb{P}\left(S_n \in \Delta(x]\right)\left(1 - h(\delta)\right),$$

where $h(\delta) := \mathbb{P}(|\eta| \geq 1/\delta) \to 0$ as $\delta \to 0$, and the last equality above holds by independence of η and S_n. But the LHS can be decomposed as

$$\mathbb{P}\left(Z_n \in \Delta_+(x - \Delta\delta]\right) = \mathbb{P}\left(Z_n \in \Delta_\delta(x - \Delta\delta]\right) + \mathbb{P}\left(Z_n \in \Delta(x]\right) + \mathbb{P}\left(Z_n \in \Delta_\delta(x + \Delta]\right),$$

where $\Delta_\delta := \Delta\delta$. Applying (2.44) to each of the probabilities on the RHS, we get

$$\mathbb{P}\left(Z_n \in \Delta_+(x - \Delta\delta]\right) \leq \frac{\Delta}{\sigma\sqrt{n}}\phi\left(\frac{x}{\sigma\sqrt{n}}\right) + \frac{2\delta\Delta\phi_0}{\sigma\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right), \tag{2.45}$$

where the term $(2\delta\Delta\phi_0)/(\sigma\sqrt{n}) + o\left(n^{-1/2}\right)$ comes from $\mathbb{P}\left(Z_n \in \Delta_\delta(x - \Delta\delta]\right) + \mathbb{P}\left(Z_n \in \Delta_\delta(x + \Delta]\right)$ by the uniform integrability of ϕ. Hence

$$\mathbb{P}\left(S_n \in \Delta(x]\right) \leq \left[\frac{\Delta}{\sigma\sqrt{n}}\phi\left(\frac{x}{\sigma\sqrt{n}}\right) + \frac{2\delta\Delta\phi_0}{\sigma\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right)\right](1 - h(\delta))^{-1}. \tag{2.46}$$

If, for a given $\epsilon > 0$, we choose $\Delta > 0$ such that $(1 - h(\delta))^{-1} \leq 1 + (\epsilon\Delta)/3$ and $2\delta\phi_0 \leq \epsilon/3$, then we derive from (2.46) that, for all n large enough and ϵ small enough,

$$\mathbb{P}\left(S_n \in \Delta(x]\right) \leq \frac{\Delta}{\sigma\sqrt{n}}\left[\phi\left(\frac{x}{\sigma\sqrt{n}}\right) + \epsilon\right]. \tag{2.47}$$

Now we will obtain a lower bound for $\mathbb{P}(S_n \in \Delta(x])$. Consider the event $A := \{Z_n \in \Delta_-(x + \Delta\delta]\}$ who has

$$\mathbb{P}(A) = \mathbb{P}\left(A;\ |\eta| < 1/\delta\right) + \mathbb{P}\left(A;\ |\eta| \geq 1/\delta\right). \tag{2.48}$$

We have $\{Z_n \in \Delta_-(x+\Delta\delta]\} \cap \{|\eta| < 1/\delta\} \subseteq \{S_n \in \Delta(x]\} \cap \{|\eta| < 1/\delta\}$ and hence

$$\mathbb{P}\left(A;\ |\eta| < 1/\delta\right) \leq \mathbb{P}\left(S_n \in \Delta(x]\right). \tag{2.49}$$

On the other hand, by the "idea" of LOTP, one as

$$\mathbb{P}\left(A;\ |\eta| \geq 1/\delta\right) = \mathbb{E}\left[\mathbb{P}(A \mid \eta);\ |\eta| \geq 1/\delta\right] = \mathbb{E}\left[\mathbb{P}(S_n \in \Delta_-(x+\theta\eta+\Delta\delta]);\ |\eta| \geq 1/\delta\right]$$

where the last equality is by independence of η and S_n. However, from (2.47) we know that $\mathbb{P}(S_n \in \Delta_-(x+\theta\eta+\Delta\delta]) \leq (\Delta/(\sigma\sqrt{n}))(\phi_0+\epsilon)$, whence $\mathbb{P}\left(A;\ |\eta| \geq 1/\delta\right) \leq (\Delta/(\sigma\sqrt{n}))(\phi_0+\epsilon)h(\delta)$. From this, together with (2.48) and (2.49), we get

$$\mathbb{P}(S_n \in \Delta(x]) \geq \mathbb{P}\left(Z_n \in \Delta_-(x+\Delta\delta]\right) - \frac{\Delta}{\sigma\sqrt{n}}(\phi_0+\epsilon)h(\delta). \tag{2.50}$$

We can employ the same idea as we did in (2.45) to get the lower bound

$$\mathbb{P}\left(Z_n \in \Delta_+(x-\Delta\delta]\right) \geq \frac{\Delta}{\sigma\sqrt{n}}\phi\left(\frac{x}{\sigma\sqrt{n}}\right) - \frac{2\delta\Delta\phi_0}{\sigma\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right) \tag{2.51}$$

Now, if we combine (2.50) and (2.51) together with choosing δ such that $\phi_0 h(\delta) < \epsilon/3$ and $2\delta\phi_0 < \epsilon/3$, we get

$$\mathbb{P}(S_n \in \Delta(x]) \geq \frac{\Delta}{\sigma\sqrt{n}}\left[\phi\left(\frac{x}{\sigma\sqrt{n}}\right) - \epsilon\right]. \tag{2.52}$$

Finally, since ϵ is arbitrarily small, inequalities (2.47) and (2.52) prove the required relation (2.35) . This completes the proof of Theorem 2.18. □

Remark 2.22. One can see from the proof above that the RHS of (2.35) remains the same even if the $\Delta(x]$ on the LHS is replaced by any of $\Delta(x) := (x, x+\Delta)$, $\Delta[x) := [x, x+\Delta)$ and $\Delta[x] := [x, x+\Delta]$, i.e., endpoints are not troublesome. The same holds in the multi-dimensional case, since $\Delta(\boldsymbol{x}]$ is just a direct product of intervals,

Before presenting the multi-dimensional version of Stone's theorem, we need to first state, without a proof, the following result from linear algebra, which will be used in Theorem 2.24 and other results in Chapter 4.

Proposition 2.23 (the Principal Square Root)**.** *Let M be a positive semidefinite real matrix. Then there is exactly one positive semidefinite (and hence symmetric) matrix S such that $M = S^T S$. This unique matrix is called the principal square root of M, denoted by $M^{1/2}$.*

Theorem 2.24 (the Stone Integro-local Theorem: non-lattice, multi-dimensional)**.** *Let* $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots$ *be a sequence of i.i.d. random vectors in* $\mathbb{R}^d$ *such that* $\boldsymbol{\xi}_k \overset{d}{=} \boldsymbol{\xi}$ *for all* $k \geq 1$, $\boldsymbol{a} := \mathbb{E}\boldsymbol{\xi}$ *and* $\Sigma := \operatorname{Cov} \boldsymbol{\xi} \equiv \mathbb{E}\left[(\boldsymbol{\xi} - \boldsymbol{a})^T(\boldsymbol{\xi} - \boldsymbol{a})\right]$. *Let* $\boldsymbol{S}_n := \boldsymbol{\xi}_1 + \cdots + \boldsymbol{\xi}_n$, $n \geq 1$. *Assume that* Σ *is positive definite (i.e.* $\boldsymbol{\xi}$ *is nondegenerate) and all components* $\xi_{(1)}, \ldots, \xi_{(d)}$ *of* $\boldsymbol{\xi}$ *are non-lattice. Then, as* $n \to \infty$,

$$\mathbb{P}\left(\boldsymbol{S_n} \in \Delta(\boldsymbol{x}]\right) = \frac{\Delta^d}{(2\pi n)^{d/2}\sqrt{\det \Sigma}} \exp\left\{-\frac{(\boldsymbol{x} - n\boldsymbol{a})\Sigma^{-1}(\boldsymbol{x} - n\boldsymbol{a})^T}{2n}\right\} + o\left(n^{-d/2}\right) \tag{2.53}$$

$$= \frac{\Delta^d}{n^{d/2}} \frac{(2\pi)^{-(d-1)/2}}{\sqrt{\det \Sigma}} \phi\left(\frac{\|(\boldsymbol{x} - n\boldsymbol{a})\Sigma^{-1/2}\|}{\sqrt{n}}\right) + o\left(n^{-d/2}\right) \tag{2.54}$$

as $n \to \infty$, *where* ϕ *is the univariate standard normal density. The remainder term* $o\left(n^{-d/2}\right)$ *is uniform in* $x \in \mathbb{R}^d$ *and* $\Delta \in [\Delta_L, \Delta_R]$ *with any fixed* $0 < \Delta_L < \Delta_R < \infty$. *The matrix* $\Sigma^{-1/2}$ *is the principal square root of* Σ^{-1} *whose existence is guaranteed by the positive definiteness of* Σ^{-1} *and Proposition 2.23.*

Proof. It can be proved using a multivariate analog of the same argument as in the one-dimensional version (Theorem 2.18) by the method of ch.f.'s. A proof can be found in C. Stone's original paper [98] in 1965. □

In 1966, C. Stone [99] established a refined version of the above theorem, which is applicable not only to non-lattice measures but also to lattice ones, or even to a combination of the two. Again, although C. Stone proved a version that applies to cases with any stable limit law, we are only interested in the normal limit case in this project.

To prepare for the refined Stone's Theorem, we need first to "tidy up" arbitrarily given distribution in the following sense.

Proposition 2.25. *For a random vector* $\boldsymbol{\xi} := [\boldsymbol{\xi}_{(1)}, \ldots, \boldsymbol{\xi}_{(d)}]$ *in* $\mathbb{R}^d$ *with an arbitrary distribution, it is always possible to reduce* $\boldsymbol{\xi}$ *to the following form by a suitable linear transformation (including a renumbering of coordinates and a shift of the origin):*

(i) *the first* l *coordinates* $\boldsymbol{\xi}_{(1)}, \ldots, \boldsymbol{\xi}_{(l)}$ *have non-lattice distribution, i.e., the ch.f.* $\psi(\boldsymbol{\lambda}) := \mathbb{E}\, e^{i\langle \boldsymbol{\lambda}, \boldsymbol{\xi}\rangle}$ *for* $\boldsymbol{\lambda} := [\lambda_1, \ldots, \lambda_d] \in \mathbb{R}^d$, *where* $\langle \boldsymbol{\lambda}, \boldsymbol{\xi}\rangle$ *is the scalar product of* $\boldsymbol{\lambda}$ *and* $\boldsymbol{\xi}$, *possesses the property* $|\psi(\boldsymbol{\lambda})| < 1$ *if* $l \geq 1$ *and* $[\lambda_1, \ldots, \lambda_l] \neq \mathbf{0}$;

(ii) *the remaining* $d - l$ *coordinates* $\boldsymbol{\xi}_{(l+1)}, \ldots, \boldsymbol{\xi}_{(d)}$ *are arithmetic; i.e., each coordinate is an integer, and the greatest common divisor of the differences between its possible values equals* 1.

By Proposition 2.25, we can restrict ourselves to the class of distributions satisfying (i) and (ii). C. Stone [99] proves the following refinement in 1966.

Theorem 2.26 (the Stone Integro-local Theorem: any distribution)**.** *Let $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots$ be a sequence of i.i.d. random vectors in $\mathbb{R}^d$ such that $\boldsymbol{\xi}_k \overset{d}{=} \boldsymbol{\xi}$ for all $k \geq 1$, $\boldsymbol{a} := \mathbb{E}\boldsymbol{\xi}$ ad $\Sigma := \operatorname{Cov} \boldsymbol{\xi} \equiv \mathbb{E}\left[(\boldsymbol{\xi} - \boldsymbol{a})^T(\boldsymbol{\xi} - \boldsymbol{a})\right]$. Let $\boldsymbol{S}_n := \boldsymbol{\xi}_1 + \cdots + \boldsymbol{\xi}_n$, $n \geq 1$. Assume that Σ is positive definite (i.e. $\boldsymbol{\xi}$ is nondegenerate), and all $\boldsymbol{\xi}_n$ are in the form stated in Proposition 2.25. Then, for a direct product of half-open intervals in $\mathbb{R}^d$*

$$\Delta_l(\boldsymbol{x}] := \prod_{j=1}^{l} (x_j, x_j + \Delta] \prod_{j=l+1}^{d} (x_j, x_j + 1] \quad \textit{for } \boldsymbol{x} = [x_1, \ldots, x_d],\ \Delta > 0,$$

one has

$$\begin{aligned} \frac{\mathbb{P}\left(\boldsymbol{S_n} \in \Delta_l(\boldsymbol{x}]\right)}{\Delta^l} &= \frac{1}{(2\pi n)^{d/2}\sqrt{\det \Sigma}} \exp\left\{-\frac{(\boldsymbol{x} - n\boldsymbol{a})\Sigma^{-1}(\boldsymbol{x} - n\boldsymbol{a})^T}{2n}\right\} + o\left(n^{-d/2}\right) \\ &= \frac{(2\pi)^{-(d-1)/2}}{n^{d/2}\sqrt{\det \Sigma}} \phi\left(\frac{\|(\boldsymbol{x} - n\boldsymbol{a})\Sigma^{-1/2}\|}{\sqrt{n}}\right) + o\left(n^{-d/2}\right) \end{aligned} \tag{2.55}$$

as $n \to \infty$, where ϕ is the density of the univariate standard normal. The remainder term $o\left(n^{-d/2}\right)$ is uniform in $\boldsymbol{x} \in \mathbb{R}^d$, and $\Delta \in [\Delta_L, \Delta_R]$ with any fixed $0 < \Delta_L < \Delta_R < \infty$. The matrix $\Sigma^{-1/2}$ is the principal square root of Σ^{-1} whose existence is guaranteed by the positive definiteness of Σ^{-1} and Proposition 2.23.

A proof can be found in [99]. C. Stone first proved it under the Cramér condition on the ch.f. and then extended to the general case.

Remark 2.27. If $l = 0$ (all components $\xi_{(j)}$ are arithmetic), then, WLOG, one may assume that $\boldsymbol{x}$ is a vector with arithmetic coordinates, and replace the LHS of (2.55) with $\mathbb{P}(\boldsymbol{S_n} = \boldsymbol{x})$. In this case, the integro-local theorem becomes the local theorem which was proved by E. Rvačeva (Theorem 2.14). If $l = d$, then (2.55) recovers the relation (2.53) we established in Theorem 2.24 for non-lattice laws. If $1 < l < d$, then $\boldsymbol{\xi}$ is a combination with some of the components being lattice and some being non-lattice.

It is natural to call this type of Limit Theorems *Integro-Local* since it contains some elements of *Integral* Theorems (the probability that the sum enters a set) and the *Local* Theorems (the set is *small*). This name makes it possible to distinguish this type of limit theorems from the classical *local* theorems, which assume the existence of densities and are about the asymptotics of the densities of the sum. And the classical Integral Theorems are much rougher than the integro-local ones in the sense that the former describe the asymptotics of the probability that the sum enters "large" sets. The integral theorems can be deduced from the integro-local theorems but not vice versa.

Hence the Integro-Local Theorem is perhaps the most perfect and precise version of the classical Limit Theorems. Indeed, it does not assume any additional conditions on top of the standard requirement of finite second moments (except for distinguishing between the lattice and non-lattice cases) but has basically got the same accuracy as the Local Limit Theorems without making any assumptions about the existence of the densities. That means that the Integro-Local Theorems are much more precise than the integral ones, and it is easy to see that one can derive the assertions of the latter from the former, but not the other way around.

The integro-local theorems are of interest in their own right and serve as a key element, for example, in computing the exact asymptotics of large deviation probabilities for sums of independent random variables [21, Ch. 9]; in proving the assertions about the first passage time of a remote curvilinear boundary by a random walk [24]; in establishing integro-local theorems for compound renewal processes [21, Ch. 10]; and in a number of other problems.

It is natural to ask if the reminder term $o(n^{-d/2})$ in Theorem 2.26 can be sharpened under minimal additional assumptions. The first step in that direction was made by A. A. Borovkov [23] in 2016. He proved that, if Cramér's strong non-lattice condition $\limsup_{|\lambda|<\infty}|\psi(\lambda)| < 1$ on the ch.f. is satisfied and $\mathbb{E}|\boldsymbol{\xi}|^{2+r} < \infty$ for some $r \in (0, 1]$, then the error term can be improved to $O(n^{-(d+r)/2})$ uniformly in $\boldsymbol{x} \in \mathbb{R}^d$ and $\Delta \in (q^n, cn^{(1-r)/2})$ for some $q \in (0, 1)$ and every fixed $c > 0$. Later in the same year, A. A. Borovkov and K. A. Borovkov [25] further developed the approach from [23] to derive the first term of asymptotic expansion for $\mathbb{P}(\boldsymbol{S}_n \in \Delta(\boldsymbol{x}])$, in the one-dimensional setting, with uniform bounds for the remainder term in the case when Cramér's strong non-lattice condition is met and $\mathbb{E}\,|\xi|^{2+r} < \infty$ for some $r \in [1, 2]$.

Extensions of the Integro-Local Limit Theorems to the case of non-identically distributed independent random variables were established, in the triangular array scheme, by A. A. Borovkov in [20] in 2011. In this project, for the processes (involving non-identical sequences) to be introduced in the next chapter, we derive their Integro-Local Limit behaviour. Our results will be presented in Chapter 4.

Chapter 3

Contagious Point Processes

This project is aimed at studying the local limit behaviour of the growing point processes considered in K. A. Borovkov's paper [28] which will be discussed in Section 3.3. Briefly speaking, the process considered in [28] is a dynamic version of the Neyman contagious point process. The latter was firstly introduced by J. Neyman [75] in 1939 with its wide applications in Entomology and Bacteriology. The class of Neyman contagious point processes, together with its history, will be discussed in Section 3.1. Prior to [28], the first step taken in studing such contagious growing point processes was taken by K. A. Borovkov and A. Motyer [26] in 2005, which will be discussed in Section 3.2.

3.1 On contagious distributions: historical viewpoint

The term "contagious" in the context of point processes goes back to G. Pólya and F. Eggenberger [34, 35, 81] in 1923 with the introduction of the well-known *Pólya-Eggenberger urn scheme* and its properties studied.

A brief description of the scheme is as follows. From an urn containing a red and b black balls. At each time step, one ball is removed from the urn; its colour is noted, and the ball is returned to the urn along with s additional balls of the same colour; the experiment is then repeated using the newly constituted urn, and in this way, a sequence of trials is performed. In fact, this model was firstly introduced by A. A. Markov [63] in 1906.

This model has been used extensively in studies of contagion and served as a prototype for many models discussed in the literature. One of them is the *B. Friedman urn model* [43] in 1949, whose elementary properties can be found in W. Feller [39, p. 119] and M. Fréchet [41]. More advanced properties of B. Friedman's urn model can be found in A. D. Friedman [44] and D. Blackwell [15]. Furthermore, in

1973, D. Blackwell and J. McQueen [16] described the construction of a Dirichlet prior distribution by a generalization of Pólya's urn scheme which is known as the *Blackwell-McQueen Pólya urn scheme*. Another well-known class of models motivated from the idea of the Pólya urn scheme is the *Neyman contagious distributions* [75] whose story can be told as follows.

Contagious distributions have been applied to biological problems. For instance, the spread and distribution of plants over an area, where we have initial or "parent" plants, and others ("offspring") associated with them in such a way as to render the use of the simple Poisson distribution as a description of the spreading process invalid. In such situations, new contagious distributions are in demand.

Neyman [75] developed a new class of contagious distributions in the fields of bacteriology and entomology in 1939. But they can be equally applicable to situations where individuals or items are supposed to initially be dispersed in randomly scattered groups — egg masses in the case of insects or clumps in the case of bacteria — which are subject to a chance fluctuation in size. It may be supposed that there occurs subsequently some spatial dispersion from the initial groups. Under such a scheme, the distribution of individuals, or items, cannot be random. Since the probability of an individual occurring in a given unit area is related to the probability of other individuals occurring in that unit area, the distribution may be called *contagious*.

Inspired by the idea of the Pólya urn, Neyman [75] has suggested three types of theoretical frequency distribution called Types A, B and C. These types all correspond well to observed data for cases where the mechanism of dispersion is roughly previously set forth. They proved to be better than the Poisson distribution and superior to the negative binomials. The *Neyman contagious distributions* are preferable in various biological applications, such as in the original work by Neyman [75] in 1939 and in later work by Beall [8] in 1940 and by Archibald [3] in 1948. Types A, B and C fit data progressively better. For illustration, one is the case of *Pyrausta nubilalis* Hubn. in 1937 [8, Table VII], where the types A, B and C give values of χ^2 with 0.011, 0.064 and 0.110. As a second illustration, there is the case of *P. nubilalis* [8, Table VII], where the types give values of χ^2 with 0.05, 0.48 and 0.57.

However, Neyman's types do not go far enough. It is indicated by their forms that they are simply early members of a great family of distributions. This can be seen from the following ch.f.'s of types A, B and C:

$$\varphi_A(t) = \exp\left\{m_1\left\{e^{m_2(e^{it}-1)} - 1\right\}\right\},$$
$$\varphi_B(t) = \exp\left\{m_1\left\{e^{m_2(e^{it}-1)} - 1 - m_2(e^{it}-1)\right\}/m_2(e^{it}-1)\right\},$$
$$\varphi_C(t) = \exp\left\{2m_1\left\{e^{m_2(e^{it}-1)} - 1 - m_2(e^{it}-1) - m_2^2(e^{it}-1)^2\right\}/m_2^2(e^{it}-1)^2\right\},$$

where m_1 and m_2 are parameters. This sequence of three forms suggests the following general form:

$$\varphi(t) = \exp\left\{m_1\Gamma(n+1)\sum_{s=0}^{\infty}\frac{m_2^{s+1}(e^{it}-1)^{s+1}}{\Gamma(n+s+2)}\right\},$$

where Types A, B and C are special cases when $n = 0, 1,$ and 2, respectively. For $n > 2$ we should have new distributions. For $n < 2$ and not an integer, we should have cases intermediate to Neyman's A, B and C Types. Further discussions about the general form of Neyman's contagious distributions can be found in [9, 1953].

Inspired by Neyman's contagious distributions and their shortcomings, more methods appeared afterwards. Thomas [101, 1949] invented the double Poisson distribution to describe non-random distributions of plants. Darwin [31, 1951] invented another method of obtaining contagious distributions, of wider applicability, and draws attention to some shortcomings of the two earlier methods. More aspects of Darwin's method were discussed by J. G. Skellam [95, 1952] and H. R. Thompson [102, 1954].

Motivated by Thompson's discussions in [102], *a simple dynamic version of the Neyman contagious point process* was considered by K. A. Borovkov and A. Motyer [26, 2005]. One can see from Section 3.2 that this simple point process is well-behaved asymptotically. Further inspired from this observation, *a general dynamic version of Neyman contagious point process* was considered in K. A. Borovkov [28]. One can see from Section 3.3 that the general version also possesses good asymptotic behaviour.

In this project, we derive more detail about the asymptotic behaviour of the growing point processes considered in [28] in the direction of local limit theorems. Results will be presented in Chapter 4.

3.2 On a simple growing point process model

This section describes the simple dynamic version of the Neyman contagious point process considered in [26, 2005] and its asymptotic behaviour. This process differs from the Neyman contagious process, which is a one-stage scheme, in the sense that it is multistage.

The model considers the evolving in discrete time $n = 0, 1, \dots$ of the simple point process

$$\{\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_{w(n)}\} \subseteq \mathbb{R}^d, \qquad w(n) := m + nk, \tag{3.1}$$

where $k \geq 1$ is a fixed integer, that starts with an arbitrary (random) initial configuration $\{\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_m\}$ at time $n = 0$ and then grows as follows. Let, for each time $n \geq 0$,

$$p_n = \{p_{n,1}, p_{n,2}, \dots, p_{n,w(n)}\}, \tag{3.2}$$

denote a probability distribution on $\{1, 2, \dots, w(n)\}$, and F_n a distribution on $\mathbb{R}^d$, and let $N_n, \boldsymbol{Y}_{n,1}, \dots, \boldsymbol{Y}_{n,k}$ denote independent (of each other and also of the past evolution of the process) random variables, with the distribution of N_n being p_n and all $\boldsymbol{Y}_{n,j}$, $j = 1, \dots, k$, following the law F_n. Then, given the configuration of the process at time $n \geq 0$, k new points are added to the process at time $n + 1$ at the locations

$$\boldsymbol{X}_{w(n)+j} := \boldsymbol{X}_{N_n} + \boldsymbol{Y}_{n,j}, \qquad j = 1, \dots, k. \tag{3.3}$$

Thus, first, we choose a "mother" at random from the existing points and then add a fixed number k of "daughter" points, randomly displaced relative to the mother's position. Note that, in the general case, the evolving point process is non-Markovian.

Remark 3.1. The process structure above also reminds that of a branching random walk, see, for example, Harris [55, 1963], Asmussen and Kaplan [4, 1976] and Biggins [14, 1997, p. 19–39]. In fact, if p_n is a uniform distribution, the process under consideration will be the embedded skeleton process for a k-nary continuous time branching random walk.

In [26], the authors consider the case when the distribution (3.2) has the following special form: for a fixed $a > 0$, for all $n \geq 0$, $j = 1, 2, \dots, w(n)$

$$p_{n,j} = C_n a^{j-1}, \qquad \text{where } C_n = \begin{cases} (1-a)/(1-a^{w(n)}), & a \neq 1, \\ 1/w(n), & a = 1, \end{cases} \tag{3.4}$$

and derive the limiting (as $n \to \infty$) behaviour of the common distribution P_n of the newly added points at time n:

$$P_n := \mathcal{L}(\boldsymbol{X}_{w(n-1)+1}) \equiv \cdots \equiv \mathcal{L}(\boldsymbol{X}_{w(n)}), \qquad n = 1, 2, \ldots, \tag{3.5}$$

where by $\mathcal{L}(\boldsymbol{X})$ we denote the distribution of $\boldsymbol{X}$. They also derive the limiting behaviour of the mean measure M_n of the process:

$$M_n(B) := \mathbb{E}\left[\frac{1}{w(n)}\sum_{j=1}^{w(n)} \mathbf{1}_{\{\boldsymbol{X}_j \in B\}}\right] = \frac{1}{w(n)}\sum_{j=1}^{m} \mathcal{L}(\boldsymbol{X}_j)(B) + \frac{nk}{w(n)}\frac{1}{n}\sum_{j=1}^{n} P_j(B). \tag{3.6}$$

The contribution of the first sum on the RHS of (3.6) is negligibly small as $n \to \infty$, and since $nk/w(n) \to 1$, the limiting behaviour of M_n will coincide with that of the distribution

$$E_n(B) := \frac{1}{n}\sum_{j=1}^{n} P_j(B). \tag{3.7}$$

Remark 3.2. The particular choice of the distribution class (3.4) of p_n is mainly due to a technical reason: it enables one to derive a simple product-form representation (relation (9) in [26]) for the ch.f. of P_n, which makes the derivation much easier. A wider class of distributions p_n is considered in [28], it will be discussed in Section 3.3.

It turns out that, under some moment conditions, the limiting behaviour of P_n and M_n will depend on the value of a in (3.4) in the following way. The above-mentioned moment conditions are the following two:

(C1) The family of random vectors $\{\|\boldsymbol{Y}_{n,1}\|\}_{n\geq 0}$ is uniformly integrable, and the sequence of means $\boldsymbol{\mu}_n := \mathbb{E}\boldsymbol{Y}_{n,1}$ has a Cesaro limit: for some $\boldsymbol{\mu} \in \mathbb{R}^d$, $\frac{1}{n}\sum_{j=1}^{n} \boldsymbol{\mu}_j \to \boldsymbol{\mu}$ as $n \to \infty$.

(C2) The family of random vectors $\{\|\boldsymbol{Y}_{n,1}\|^2\}_{n\geq 0}$ is uniformly integrable, and the sequence of the second-order moments matrices $\boldsymbol{m}_n := \mathbb{E}(\boldsymbol{Y}_{n,1}^T \boldsymbol{Y}_{n,1})$ has a Cesaro limit: for some matrix $\boldsymbol{m}$, $\frac{1}{n}\sum_{j=1}^{n} \boldsymbol{m}_j \to \boldsymbol{m}$ as $n \to \infty$.

Theorem 3.3. (i) *If $a < 1$ and $F_n \Rightarrow F$ as $n \to \infty$ for some distribution F, then there exists a probability distribution P such that $P_n \Rightarrow P$ and $M_n \Rightarrow P$ as $n \to \infty$.*

(ii) *If $a = 1$ and condition (C1) is met, then, as $n \to \infty$, $\boldsymbol{X}_{w(n)}/\ln n \xrightarrow{P} \boldsymbol{\mu}$ and also $K_n \Rightarrow \delta_{\boldsymbol{\mu}}$(the degenerate distribution at the point $\boldsymbol{\mu}$), where $K_n(B) := M_n(B \ln n)$.*

If, in addition, condition (C2) is satisfied, then, as $n \to \infty$, by letting $\boldsymbol{\nu}_n :=$

$\sum_{j=1}^{n} j^{-1}\boldsymbol{\mu}_j$ *and* $L_n(B) := M_n(\sqrt{\ln n}B+\boldsymbol{\nu}_n)$*, one has* $\mathcal{L}\left((\boldsymbol{X}_{w(n)} - \boldsymbol{\nu}_n)/\sqrt{\ln n}\right) \Rightarrow \mathcal{N}(\mathbf{0}, \boldsymbol{m})$*, and* $L_n \Rightarrow \mathcal{N}(\mathbf{0}, \boldsymbol{m})$*.*

If $\boldsymbol{\mu}_n - \boldsymbol{\mu} = o\left(\ln^{-1/2} n\right)$*, then* $\boldsymbol{\nu}_n$ *above can be replaced by* $\boldsymbol{\mu} \ln n$*.*

(iii) *Let* $a > 1$*. If condition* (C1) *is met, then, as* $n \to \infty$*, by letting* $b := (a^k - 1)/(a^k) \in (0, 1)$ *and* $U_n(B) := M_n(nB)$*, one has* $\boldsymbol{X}_{w(n)}/n \xrightarrow{P} b\boldsymbol{\mu}$ *and* $U_n \Rightarrow \mathrm{Unif}[\mathbf{0}, b\boldsymbol{\mu}]$*, where* $\mathrm{Unif}[\mathbf{0}, b\boldsymbol{\mu}]$ *denotes the uniform distribution over the straight line segment connecting the points* x *and* y *in* $\mathbb{R}^d$*.*

If, in addition, condition (C2) *is satisfied and the matrix sequence* $\{\boldsymbol{\mu}_n^T \boldsymbol{\mu}_n\}$ *has a Cesaro limit* $\boldsymbol{\mu}^2$*, then, as* $n \to \infty$*, by letting* $\boldsymbol{\kappa}_n := b\sum_{j=1}^{n} \boldsymbol{\mu}_j/(1 - a^{-m-kj})$*, one has* $\mathcal{L}\left((\boldsymbol{X}_{w(n)} - \boldsymbol{\kappa}_n)/\sqrt{n}\right) \Rightarrow \mathcal{N}(\mathbf{0}, b(\boldsymbol{m} - b\boldsymbol{\mu}^2))$*.*

If $\sum_{j=1}^{n}(\boldsymbol{\mu}_j - \boldsymbol{\mu}) = o(\sqrt{n})$*, then* $\boldsymbol{\kappa}_n$ *above can be replaced by* $b\boldsymbol{\mu}n$*. Moreover, if also* $\boldsymbol{\kappa}_n = o(\sqrt{n})$*, then the distribution* $Q_n(B) := M_n(\sqrt{n}B)$ *converges weakly as* $n \to \infty$ *to a mixture of normal distribution with the ch.f.* $q_n(\boldsymbol{t}) = 2[1 - \exp\{-b\boldsymbol{t}(\boldsymbol{m} - b\boldsymbol{\mu}^2)\boldsymbol{t}^T/2\}]/[b\boldsymbol{t}(\boldsymbol{m}^2 - b\boldsymbol{\mu}^2)\boldsymbol{t}^T]$*.*

Remark 3.4. The above theorem can be interpreted as follows.

If $a < 1$ (the older "mothers" are more likely to have offspring) and the displacement distributions F_n converge weakly to a limiting law F, then, roughly speaking, the point process remains "forever clustered" around the initial configuration, as with dominating probability any newly added point will have as its "mother" one of a relatively few initials points of the process, and its displacement relative to the "mother" will asymptotically follow the distribution F.

If $a = 1$ (all "mothers" are equally likely to produce offspring), then, under mild moment conditions (C1) and (C2), for $\boldsymbol{X}_n$ hold both the LLN (under scaling $1/\ln n$) and the CLT (after centring and under scale $1/\sqrt{\ln n}$), with similar results being true for M_n as well. Thus, the process gradually "crawls" along the direction of the asymptotic mean drift of the displacement vectors $\boldsymbol{Y}_n$, with this movement ever slowing down and the "cloud" of the points growing even slower around the mean trend value.

If $a > 1$ (the younger "mothers" are more productive), results for $\boldsymbol{X}_n$ are similar to those in the case $a = 1$ (but with scaling $1/n$ and $1/\sqrt{n}$, respectively), whereas M_n can be approximated by a uniform distribution along the "mean drift" line or, in the case of asymptotically negligible drift, by a mixture of normal distributions. In words, the process propagates at a constant rate along the direction of the asymptotic mean drift (if it is non-zero), with an "error" in positioning new points around the mean drift line being of the order $\sqrt{n}$, typical for a random walk (with a finite variance). One could observe that this kind of behaviour appears to be quite natural for the model, all the more so since, say, in the case $k = 1$, as $a \to \infty$, the limiting for our model process is a random walk indeed.

This simple model was further studied in [27, 2006], in the case where $k = 1$, $d = 1$, $\boldsymbol{Y}_{n,1} \geq 0$, and the weights w_n were assumed to be random.

3.3 On a dynamic version of Neyman contagious point process

This section describes the model considered in [28, 2014], which is generalized from the simple model discussed in Section 3.2, and its asymptotic behaviour.

The point process starts with an initial configuration of $k_0 \geq 1$ random points $\boldsymbol{X}_{0,1}, \dots, \boldsymbol{X}_{0,k_0} \in \mathbb{R}^d$, $d \geq 1$, labeled[1] with respective weights $w_{0,j} \geq 0$, $j = 1, \dots, k_0$. At time step $n \geq 1$, $k_n \geq 0$ new points are added to the process, with weights $w_{n,1}, \dots, w_{n,k_n}$.

Remark 3.5. The roles of the weights $w_{n,j} \geq 0$ will be to characterize the "fitness", or reproductive ability, of the individuals to be associated with the points to be added to the process at the n-th step (they can be used, for example, to model the aging of the individuals), as they will be used to determine the probability distributions for choosing new mothers in the future.

To specify the locations of the newly added points, let $W_{-1} := 0$, and , for $n \geq 0$, set

$$w_n := \sum_{l=1}^{k_n} w_{n,l}, \qquad W_n := \sum_{r=0}^{n} w_r. \tag{3.8}$$

Then, at step $n + 1 \geq 1$, an existing point $\boldsymbol{X}_{r,j}$, $0 \leq r \leq n$, $1 \leq j \leq k_r$, is chosen at random, according the probability distribution

$$\boldsymbol{p}_n := \{p_{n,r,j} : 0 \leq r \leq n, 1 \leq j \leq k_r\}, \qquad p_{n,r,j} := \frac{w_{r,j}}{W_n}. \tag{3.9}$$

Denote that point (chosen at time step $n + 1$) by $\boldsymbol{X}_n^*$ and add $k_{n+1} \geq 0$ new points to the process, at the locations

$$\boldsymbol{X}_{n+1,l} := \boldsymbol{X}_n^* + \boldsymbol{Y}_{n+1,l}, \qquad 1 \leq l \leq k_{n+1}, \tag{3.10}$$

[1] In fact, in [28], each random point $\boldsymbol{X}_{n,j}$ are equipped with not only a weight $w_{n,j}$ but also a quantity $u_{n,j} \geq 0$ which specifies the amount of a "resource" attached to the individuals (e.g., the body weight). But we will only focus on models with weights only in this thesis.

where we assume that the random vector $\left[\boldsymbol{Y}_{n+1,1},\ldots,\boldsymbol{Y}_{n+1,k_{n+1}}\right] \in (\mathbb{R}^d)^{k_{n+1}}$ is independent of $\{\boldsymbol{X}_{r,j}\}_{0\le r\le n, 1\le j\le k_r}$ and the choice of the "mother point" $\boldsymbol{X}_n^*$. In paper [28], and also in this thesis (Chapter 4), we assume nothing about the character of dependence between the components $\boldsymbol{Y}_{n+1,l}$ (in particular, some of them can coincide). Denote by

$$\psi_{n,j}(\boldsymbol{t}) := \mathbb{E}\exp\left\{i\boldsymbol{t}\boldsymbol{X}_{n,j}^T\right\}, \qquad \boldsymbol{t}\in\mathbb{R}^d, \tag{3.11}$$

the ch.f. of the distribution $P_{n,j}$ of the point $\boldsymbol{X}_{n,j}$, $n\ge 0$, $1\le j\le k_n$, and set, for $n=0,1,2,\ldots$ and $\boldsymbol{t}\in\mathbb{R}^d$,

$$\pi_n := \frac{w_n}{W_n} \equiv \sum_{j=1}^{k_n} p_{n,n,j}, \quad \psi_n(\boldsymbol{t}) := \frac{1}{w_n}\sum_{j=1}^{k_n} w_{n,j}\psi_{n,j}(\boldsymbol{t}) \equiv \frac{1}{\pi_n}\sum_{j=1}^{k_n} p_{n,n,j}\psi_{n,j}(\boldsymbol{t}), \tag{3.12}$$

where $\pi_0 := 1$. That is, ψ_n is the ch.f. of the mixture law $P_n := \frac{1}{w_n}\sum_{j=1}^{k_n} w_{n,j}P_{n,j}$, which coincides with the conditional distribution of the location $\boldsymbol{X}_n^*$ of the mother point for step $n+1$ in our process given that this point was chosen from the k_n points added at the previous step. For n such that $k_n=0$, we can leave $\phi_{n,j}$ and ϕ_n undefined. Further, denote by

$$f_{n,j}(\boldsymbol{t}) := \mathbb{E}\exp\left\{i\boldsymbol{t}\boldsymbol{Y}_{n,j}^T\right\}, \qquad \boldsymbol{t}\in\mathbb{R}^d, \tag{3.13}$$

the ch.f. of the distribution $Q_{n,j}$ of the displacement $\boldsymbol{Y}_{n,j}$ and set

$$f_n(\boldsymbol{t}) := \frac{1}{w_n}\sum_{j=1}^{k_n} w_{n,j} f_{n,j}(\boldsymbol{t}). \tag{3.14}$$

For $n\ge 0$, set $\boldsymbol{\mu}_{n,j} := \mathbb{E}\boldsymbol{Y}_{n,j}$, $\boldsymbol{m}_{n,j} := \mathbb{E}\boldsymbol{Y}_{n,j}^T\boldsymbol{Y}_{n,j}$ (when these expectations are finite), and let

$$\boldsymbol{\mu}_n := \frac{1}{w_n}\sum_{j=1}^{k_n} w_{n,j}\boldsymbol{\mu}_{n,j}, \qquad \boldsymbol{m}_n := \frac{1}{w_n}\sum_{j=1}^{k_n} w_{n,j}\boldsymbol{m}_{n,j}, \tag{3.15}$$

such that $\boldsymbol{\mu}_n$ is the mean vector of the distribution with ch.f. f_n, while $\boldsymbol{m}_n$ is the matrix of (mixed) second order moments of that mixture distribution.

The paper [26] considered in Section 3.2 is a special case of the current model where $k_n \equiv k =$ const, $n\ge 1$, and the weights being assumed were of the form $w_{n,j} = a^n$, $j=1,\ldots,k$, for some constant $a>0$, with $\boldsymbol{Y}_{n,j}$, $j=1,\ldots,k$, being i.i.d. One can check that the current model, under above constraints, reduces to the construction in (3.1), (3.2) and (3.4).

It turns out that, for the current generalized model, a version of Theorem 3.3 can be obtained. However, in this thesis, we will only be interested in the case when the total weight of points in the process, the W_n in (3.8), is unbounded as $n \to \infty$.

Theorem 3.6. *Assume that $w_n = \xi_n n^\alpha L(n)$, $n \geq 1$, where $\{\xi_n \geq 0\}$ is Cesàro summable to a positive value $\mathcal{S}_1(\xi)$*[2] *, $\alpha > -1$ and L is a slowly varying (at infinity) function*[3].

(C1*) *The family of random vectors $\{\|\boldsymbol{Y}_{n,j}\|\}_{(n,j)\in\mathcal{A}}$ is uniformly integrable, and the sequence $\{\xi_n \boldsymbol{\mu}_n\}$ is Cesàro summable with sum $\mathcal{S}_1(\xi\boldsymbol{\mu})$.*

(C2*) *The family of random vectors $\{\|\boldsymbol{Y}_{n,j}\|^2\}_{(n,j)\in\mathcal{A}}$ is uniformly integrable, and the sequence $\{\xi_n \boldsymbol{m}_n\}$ is Cesàro summable.*

(i) *If condition (C1*) holds, then*[4] *$w\text{-}\lim_{(n,j)} P_{n,j}(\cdot \ln n) = \delta_{\boldsymbol{\lambda}}(\cdot)$, where $\delta_{\boldsymbol{\lambda}}$ is the unit mass concentrated at the point $\boldsymbol{\lambda} := (\alpha+1)\mathcal{S}_1(\xi\boldsymbol{\mu})/\mathcal{S}_1(\xi) \in \mathbb{R}^d$ and one has $\boldsymbol{\nu}_n := (\ln n)^{-1}\sum_{r=1}^{n-1}\pi_r\boldsymbol{\mu}_r \to \boldsymbol{\lambda}$ as $n \to \infty$.*

(ii) *If condition (C2*) holds, then the limit $w\text{-}\lim_{(n,j)}$ of the distributions of $(\ln n)^{-1/2}(\boldsymbol{X}_{n,j} - \boldsymbol{\nu}_n \ln n)$ exists and is equal to the normal law in $\mathbb{R}^d$ with zero mean vector and covariance matrix $(\alpha+1)\mathcal{S}_1(\xi\boldsymbol{m})/\mathcal{S}_1(\xi)$.*

Remark 3.7. One can see from the proof of Theorem 3.6 (cf. relation (16) in [28]) that, under the above assumptions of w_n, the asymptotic behaviour of $\pi_n \equiv w_n/W_n$ as $n \to \infty$ is $\pi_n \asymp 1/n$ which means that the sum $e_n := \sum_{k=0}^{n}\pi_k \asymp \ln n$ as $n \to \infty$. From this observation, one can see that the integro-local limiting behaviour derived in Chapter 4 (cf. Theorems 4.4 and 4.5) is consistent with the integral limiting behaviour derived in the above theorem, and the former is stronger.

Remark 3.8. The paper [28] also deals with the asymptotic behaviour of mean measures and processes of the above type evolving in random environments. These are not considered in the current project.

The model we are dealing with belongs to the class of "preferential attachment processes". The most famous one of them was proposed in the much-cited paper [5] where new nodes were attached to already existing ones with probabilities proportional to the degrees of the latter, with the aim of generating random "scale-free" networks that are widely observed in natural and human-made systems, including the World Wide Web and some social networks. Since then, there appeared quite a

[2] This means, by definition, that $\lim_{n\to\infty}(1/n)\sum_{j=1}^{n}\xi_j = \mathcal{S}_1(\xi)$.

[3] This means, by definition, that for all $a > 0$, $\lim_{x\to\infty} L(ax)/L(x) = 1$.

[4] The notation $w\text{-}\lim_{(n,j)}$ stands for a limit over all pairs (n,j), $n \geq 0$, $1 \leq j \leq k_n$, in the topology of weak convergence of distributions instead of the ordinary component-wise (point-wise for functions) sense.

few publications on the topic, including a monograph [33]. In most cases, in the existing literature, the "attachment rule" depends on the degrees of the existing nodes (possibly with some variations as, e.g., in [2], where new nodes may link to a node only if they fall within the node's "influence region".

Models of such a kind can be used to describe the dynamics of the population of microorganisms in varying (in both time and space) environments and the process of distribution of individual insects (as in the larvae example from Neyman's paper [75]) or colonies of social insects (such as some bee species or ants). In particular, they can be of interest when modelling foreign species invasion scenarios. Further examples include dynamics of business development for companies employing multi-level marketing techniques (or referral marketing). One can think about the points in this process as nodes in a growing random forest, the roots thereof being the initial points, with edges connecting "daughters" to their "mothers".

In paper [28], only integral limiting behaviour was derived for the model. However, in this project, we derive the integro-local limiting behaviour of the model. Results will be presented in the next chapter.

Chapter 4

Results

In this chapter, we present the main results of this project studying the local limiting behaviour of the model described in Section 3.3. In particular, we will derive the limiting behaviour of the local probabilities of the chosen "mother" point $\boldsymbol{X}_n^*$ in (3.10) in $\mathbb{R}^d$, for some $d \geq 1$, when $n \to \infty$.

4.1 Preliminaries

To prepare for the main theorems, in this section we first derive a "cleaner" representation for the distribution of $\boldsymbol{X}_n^*$ and make three general assumptions that will be applicable to all of the remaining sections.

We first recall and explain several observations (i.e., relations (4.1)–(4.2)) from the paper [26]. First note that, at time step $n+1$, by the idea of the law of total probability (LOTP) formula, the ch.f. Ψ_n of $\boldsymbol{X}_n^*$ can be written as

$$\Psi_n = \sum_{r=0}^{n} \sum_{l=1}^{k_r} p_{n,r,l} \psi_{r,l}, \qquad \psi_{n+1,j} = \Psi_n f_{n+1,j}, \tag{4.1}$$

where we recall from (3.11) and (3.13) that $\psi_{r,l}$ is the ch.f. of $\boldsymbol{X}_{r,l}$ and $f_{n+1,j}$ is the ch.f. of $\boldsymbol{Y}_{n,j}$. The first relation above holds because $\boldsymbol{X}_n^*$ is chosen from the set of all existing points $\{\boldsymbol{X}_{r,l}\}_{0 \leq r \leq n, 1 \leq l \leq k_r}$ according to the probability distribution $\boldsymbol{p}_n$ in (3.9), and the second relation is due to (3.10).

Using (3.9) and the observation that $p_{n,r,j} = p_{n-1,r,j} W_{n-1}/W_n = p_{n-1,r,j}(1 - \pi_n)$

for all $r < n$, $1 \le j \le k_r$, we obtain for $n \ge 1$, employing (4.1) and (3.12), that

$$\begin{aligned}\Psi_n &= \sum_{r=0}^{n-1}\sum_{l=1}^{k_r} p_{n,r,l}\psi_{r,l} + \sum_{l=1}^{k_n} p_{n,n,l}\psi_{n,l} \\ &= (1-\pi_n)\Psi_{n-1} + \sum_{l=1}^{k_n} p_{n,n,l}\Psi_{n-1}f_{n,l} = (1-\pi_n)\Psi_{n-1} + \pi_n\Psi_{n-1}f_n,\end{aligned}$$

where f_n is the ch.f. of weighted average displacement at step n defined in (3.14). Then, by recursion, one has

$$\Psi_n = \prod_{r=0}^{n} \left[(1-\pi_r) + \pi_r f_r\right]. \tag{4.2}$$

Assumption 4.1. *At each time step $r \ge 1$, the weighted average displacements $\frac{1}{w_r}\sum_{j=1}^{k_r} w_{r,j}\boldsymbol{Y}_{r,j}$ of daughters are identically distributed. That is, in terms of the ch.f. $f_r := \frac{1}{w_r}\sum_{j=1}^{k_r} w_{r,j}f_{r,j}$ defined in (3.14), we assume $f_r = f$ from now on.*

Remark 4.1. A special case of Assumption 4.1 is when daughters' displacements $\boldsymbol{Y}_{r,j}$ are identically distributed for all r and j. Assumption 4.1 is better than assuming identically distributed daughters in the sense that the former includes the case when daughters within each step are non-identically distributed but the overall distribution of daughters in each step is identical, which is a natural case to appear in applications.

Under Assumption 4.1, let $\bar{\pi}_r := 1 - \pi_r$. Then (4.2) becomes $\Psi_n = \prod_{r=0}^{n}(\bar{\pi}_r + \pi_r f)$. This product is a polynomial in f. Denote the coefficient of f^k in this polynomial by $b_{n,k}$'s:

$$\Psi_n = \prod_{r=0}^{n}(\bar{\pi}_r + \pi_r f) = \sum_{k=0}^{n} b_{n,k} f^k, \tag{4.3}$$

which indicates that $\boldsymbol{X}_n^*$ has a mixture distribution which can be formulated as follows. Let $\boldsymbol{Y}_0, \boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ be a sequence of i.i.d. random vectors with the common ch.f. f and $I_r \sim \mathit{Bernoulli}(\pi_r)$, for $r = 1, \ldots, n$, such that all of $I_0, \ldots, I_n, \boldsymbol{Y}_0, \ldots, \boldsymbol{Y}_n$ are independent ($I_0 := 1$). Then the probabilistic interpretation of (4.3) is

$$\boldsymbol{X}_n^* \stackrel{d}{=} \sum_{r=0}^{n} I_r \boldsymbol{Y}_r, \qquad n \ge 1. \tag{4.4}$$

Denote the law of $\boldsymbol{X}_n^*$ by $\mathcal{L}_{\boldsymbol{X}_n^*}$ and let F be the law corresponding to the ch.f. f.

Then (4.3) implies

$$\mathcal{L}_{\boldsymbol{X}_n^*} = \sum_{k=0}^{n} b_{n,k} F^{*k}, \tag{4.5}$$

where F^{*k} is the k-fold convolution of F with itself.

Remark 4.2. For each $n \geq 0$, the collection of coefficients $\{b_{n,k}\}_{0\leq k\leq n}$ forms a discrete probability distribution. This can be verified as follows. Firstly, evaluating (4.3) at 0 gives $1 = \sum_{k=0}^{n} b_{n,k}$. Secondly, since π_r and $\bar{\pi}_r$ both in $[0,1]$, all possible coefficients in the sum expanded from $\prod_{r=0}^{n}(\bar{\pi}_r + \pi_r f)$ must be non-negative. Such a distribution is known as the *Poisson binomial distribution*, which is a special case of the so-called generalized binomial distribution (allows non-constant probabilities of success from trial to trial).

Let B_n be a random variable with $\mathbb{P}(B_n = k) = b_{n,k}$ for $k = 0, \dots, n$. Note that the existence of such a B_n is guaranteed by Remark 4.2. Then, from (4.3), one has

$$B_n \stackrel{d}{=} \sum_{k=0}^{n} I_k, \tag{4.6}$$

where the independent indicators I_k are those from (4.4).

It turns out that the mean and variance of B_n will be the key features in analyzing the local limiting behaviour of $\boldsymbol{X}_n^*$, and hence that of the mean measure M_n. It follows from (4.6) that

$$e_n := \mathbb{E}B_n = \sum_{k=0}^{n} \pi_k, \qquad q_n^2 := \operatorname{Var} B_n = \sum_{k=0}^{n} \pi_k(1-\pi_k), \tag{4.7}$$

where $\pi_k := w_k/W_k$, defined from (3.12), which represents the probability that the mother chosen at step $n+1$ is the point added to the process at step n. Recall from (3.8) that w_n is the total weight of those points added in step n and W_n is the total weight of all points, existing at that time.

Assumption 4.2. *The sequence of standard deviations diverges to infinity when time evolves: $q_n \to \infty$ (whence $e_n \to \infty$) as $n \to \infty$.*

Assumption 4.3. *There exist $\delta \in (0,1)$ and $N < \infty$ such that $\pi_k \leq 1-\delta$, $\forall\, k \geq N$. An immediate consequence is $\delta e_n \leq q_n^2 \leq e_n$ which means $q_n^2 \asymp e_n$ as $n \to \infty$.*

Remark 4.3. In the following three cases, we analyze the asymptotic behaviour of their sequences e_n and q_n. It turns out that Assumption 4.2 and 4.3 are satisfied in each case.

Case 1: If $w_n = \text{const}$ for all $n \geq 0$ (the overall weight added in each step is equal), then $\pi_n = \frac{1}{n}$ which means that $e_n = \sum_{k=0}^{n} \frac{1}{k} \asymp \ln n \asymp q_n^2$ as $n \to \infty$.

Case 2: If $w_n = cn^p$ for some $c > 0, p > -1$ (the overall weight added in each step increases as a power function), then[1] $W_n \asymp \frac{n^{p+1}}{p+1}$. Hence one can obtain $\frac{1}{n} < \pi_n < \frac{2}{3}\ \forall n \geq 2p+2$ and hence $q_n^2 \asymp e_n > \sum_{k=0}^{n} \frac{1}{k} \to \infty$ as $n \to \infty$.

Case 3: If $w_n = \alpha e^{\beta n}$ for some $\alpha, \beta > 0$ (the overall weight added in each step increases as an exponential function), then $\frac{1}{1+e^{-\beta}} < \pi_n < \frac{e^{\beta}-1}{e^{\beta}-e^{-2\beta}} < 1\ \forall n \geq 2$ and hence $q_n^2 \asymp e_n > \frac{n}{1+e^{-\beta}} \to \infty$ as $n \to \infty$.

Note that the cases with w_n increasing slowly so that $\lim_{n\to\infty} W_n < \infty$ are not considered in this chapter.

4.2 Local Limit Theorems for non-lattice X_n^*

For the location $\boldsymbol{X}_n^*$ of the "mother" chosen at time step $n+1$ in the non-lattice case, we derive the following results for the limiting behaviour of its integro-local probabilities: the one-dimensional case in Theorem 4.4; the multi-dimensional case in Theorem 4.5.

Recall that F is the law corresponding to the ch.f. f of the weighted average displacement in $\mathbb{R}^d$ (cf. Assumption 4.1). Recall e_n and q_n from (4.7). Recall from (2.34) that, for any $\Delta > 0$, $\boldsymbol{x} \in \mathbb{R}^d$, $\Delta(\boldsymbol{x}] := (\boldsymbol{x}, \boldsymbol{x} + \Delta]$ is a direct product of left open and right closed intervals.

Theorem 4.4 (the one-dimensional non-lattice case)**.** *Let the law F be non-lattice with mean $\mu < \infty$ and variance $\sigma^2 < \infty$, under Assumptions 4.1 — 4.3, the local law $\mathcal{L}_{\boldsymbol{X}_n^*}(\Delta(x]) = \mathbb{P}(\boldsymbol{X}_n^* \in \Delta(x])$ of $\boldsymbol{X}_n^* \in \mathbb{R}$ has the following asymptotic behaviour:*

(i) *for the case with $\mu = 0$, one has*

$$\mathcal{L}_{\boldsymbol{X}_n^*}(\Delta(x]) = \frac{\Delta}{\sigma\sqrt{e_n}}\phi\left(\frac{x}{\sigma\sqrt{e_n}}\right) + o\left(\frac{1}{\sqrt{e_n}}\right); \tag{4.8}$$

(ii) *for the case with $\mu \neq 0$, one has*

$$\mathcal{L}_{\boldsymbol{X}_n^*}(\Delta(x]) = \frac{\Delta}{\sqrt{\mu^2 q_n^2 + \sigma^2 e_n}}\phi\left(\frac{x - \mu e_n}{\sqrt{\mu^2 q_n^2 + \sigma^2 e_n}}\right) + o\left(\frac{1}{\sqrt{e_n}}\right), \tag{4.9}$$

where the remainder term $o\left(1/\sqrt{e_n}\right)$ in both cases is uniform in $x \in \mathbb{R}$, and in $\Delta \in [\Delta_L, \Delta_R]$ with fixed $0 < \Delta_L < \Delta_R < \infty$.

[1] It was firstly derived by J. Bernoulli in [10] for integer-valued p. Further studied by Karamata.

The proof of Theorem 4.4 will be given in Section 4.3.

We also derive a multivariate version of Theorem 4.4 for the case with higher dimensional locations $\boldsymbol{X}_n^* \in \mathbb{R}^d$ for $d \geq 1$. Denote the Euclidean norm by $\|\cdot\|$. Recall from (2.34) that $\Delta(\boldsymbol{x}] := \prod_{j=1}^d (x_j, x_j + \Delta]$ for $\Delta > 0$ and $\boldsymbol{x} = [x_1, \ldots, x_d] \in \mathbb{R}^d$.

Theorem 4.5 (the d-dimensional non-lattice case)**.** *Let every margin of the law F be non-lattice. Let the mean vector $\boldsymbol{\mu}$ and covariance matrix Σ (corresponding to the law F) have all entries finite. Assume that Σ is non-singular (i.e., Σ^{-1} exists), together with Assumptions 4.1 —4.3. Then $\mathcal{L}_{\boldsymbol{X}_n^*}(\Delta(\boldsymbol{x}]) = \mathbb{P}(\boldsymbol{X}_n^* \in \Delta(\boldsymbol{x}])$ has the following*[2] *asymptotic behaviour:*

(i) *for the case with $\mu = 0$, one has*

$$\mathcal{L}_{\boldsymbol{X}_n^*}(\Delta(\boldsymbol{x}]) = \frac{\Delta^d (2\pi)^{-(d-1)/2}}{e_n^{d/2}\sqrt{\det \Sigma}} \phi\left(\frac{\|\boldsymbol{x}\Sigma^{-1/2}\|}{\sqrt{e_n}}\right) + o\left(e_n^{-d/2}\right); \tag{4.10}$$

(ii) *for the case with $\mu \neq 0$, put $\boldsymbol{M} := {q_n}^2 \boldsymbol{\mu}^T \boldsymbol{\mu} + e_n \Sigma$, one has*

$$\mathcal{L}_{\boldsymbol{X}_n^*}(\Delta(\boldsymbol{x}]) = \frac{\Delta^d (2\pi)^{-(d-1)/2}}{\sqrt{\det \boldsymbol{M}}} \phi\left(\frac{\|(\boldsymbol{x} - \mu e_n)\boldsymbol{M}^{-1/2}\|}{\sqrt{e_n}}\right) + o\left(e_n^{-d/2}\right), \tag{4.11}$$

where the remainder term $o\left(e_n^{-d/2}\right)$ in both cases is uniform in $\boldsymbol{x} \in \mathbb{R}^d$, and in $\Delta \in [\Delta_L, \Delta_R]$ with fixed $0 < \Delta_L < \Delta_R < \infty$.

The proof of Theorem 4.5 is long and hence will be given seperately in Section 4.3.

Remark 4.6. One can think of Theorem 4.4 as an extended version of the one-dimensional Stone's Theorem 2.18 and think of Theorem 4.5 as an extended version of the multi-dimensional Stone's Theorem 2.24. Both are extended in the sense that they are applicable not only to a sum of i.i.d. terms, but also to the case where the summands are independent but non-identically distributed: $I_r \boldsymbol{Y}_r$, where $I_r \sim Bernoulli(\pi_r)$, $\boldsymbol{Y}_r$ are i.i.d. and all I_r and $\boldsymbol{Y}_r$ are independent, cf. (4.4).

Remark 4.7. Since Theorem 4.4 is valid for any fixed $\Delta > 0$, together with Remark 2.20, it is easy to see that Theorem 4.4 will also be valid when $\Delta = \Delta_n \to 0$ sufficiently slowly as $n \to \infty$. By the properties of densities, the RHS of representations (4.8) and (4.9) have the same form as if the random variables $(\boldsymbol{X}_n^* - \mu e_n) / \sqrt{\mu^2 q_n^2 + \sigma^2 e_n}$ had the density $\phi(\cdot) + o(1)$, although the existence of the density of $\boldsymbol{X}_n^*$ is not assumed in the theorem.

[2] Note that the ϕ is Theorem 4.5 is still the *univariate* standard normal density. We don't need to introduce multivariate standard normal density ϕ_{multi} since it can always be expressed through the relation $\phi_{\text{multi}}(\boldsymbol{t}) = (2\pi)^{(d-1)/2}\phi(\|\boldsymbol{t}\|)$, $\boldsymbol{t} \in \mathbb{R}^d$.

Remark 4.8. If $\Delta = \Delta_n$ grows, then the asymptotics of $\mathbb{P}(\boldsymbol{X}_n^* \in \Delta(\boldsymbol{x}])$ can be obtained by summing up the representations in Theorem 4.4 or Theorem 4.5 for, say, $\Delta = 1$ (if $\Delta_n \to \infty$ is integer-valued), cf. Remark 2.19. Thus the integral theorem follows from the integro-local one but not vice versa.

Remark 4.9. The integro-local limit behaviour derived in Theorem 4.4 is consistent with the integral limit behaviour derived in [28] (cf. Theorem 3.6) as it is expected to be. One can see this consistency from Remark 3.7 which explains that, under conditions of Theorem 3.6, we have $\pi_n \asymp 1/n$, whence $e_n \asymp \ln n$, as $n \to \infty$. Further, one can see that Theorem 4.4 and 4.5 are refined in the sense that they can imply Theorem 3.6 but not the vice versa.

4.3 Proofs of Theorem 4.4 and Theorem 4.5

This section contains two proofs: the former one for Theorem 4.4 and the latter one for Theorem 4.5. Note that in both proofs, we use $o(\cdot)$ to indicate the little-oh that is *uniform in x and $\Delta \in [\Delta_L, \Delta_R]$ for any fixed* $0 < \Delta_L < \Delta_R < \infty$, and use $o_x(\cdot)$ to indicate the little-oh that is *uniform in $\Delta \in [\Delta_L, \Delta_R]$ for any fixed $0 < \Delta_L < \Delta_R < \infty$, but depends on x.* The same convention applies to $O(\cdot)$.

We first state two auxiliary assertions that will be used in both proofs.

Lemma 4.10 (Bernstein's inequality)**.** *Let $\zeta_1, \dots, \zeta_n$ be a sequence of independent random variables with zero mean and there exists an $M < \infty$ such that $|\zeta_i| \le M$ a.s. for all $i \ge 1$. Then, for all $t > 0$, one has*

$$\mathbb{P}\left(\sum_{i=1}^{n} \zeta_i > t\right) \le \exp\left\{-\frac{\frac{1}{2}t^2}{\sum_{i=1}^{n} \mathbb{E}\zeta_i^2 + \frac{1}{3}Mt}\right\} \tag{4.12}$$

A proof of this equality can be found in [12].

Lemma 4.11 (Lindeberg's CLT)**.** *For a sequence of independent random variables $\zeta_1, \zeta_2, \dots$ with mean $\mu_i < \infty$ and variance $\sigma_i^2 < \infty$, let $q_n^2 := \sum_{i=1}^{n} \sigma_i^2$. If, for every $\tau > 0$, the following Lindeberg condition holds:*

$$\iota_n := \frac{1}{q_n^2} \sum_{i=1}^{n} \mathbb{E}\left[(\zeta_i - \mu_i)^2 \cdot \mathbf{1}_{\{|\zeta_i - \mu_i| > \tau q_n\}}\right] \to 0 \qquad \text{as } n \to \infty, \tag{4.13}$$

then $\lim_{n\to\infty} \mathbb{P}\left(B_n \le q_n t + \sum_{i=1}^{n} \mu_i\right) = \Phi(t)$ for all $t \in \mathbb{R}$, where Φ is the c.d.f. of the standard normal distribution.

Remark 4.12. One can check that the Lindeberg condition (4.13) holds in our case of $\zeta_k = I_k \sim \mathit{Bernoulli}(\pi_k)$: $0 \leq \iota_n \leq \frac{1}{\tau q_n^3}\sum_{i=1}^n \mathbb{E}\left[(I_i-\pi_i)^3\right] = \frac{1}{\tau q_n^3}\sum_{i=1}^n \pi_i(1-\pi_i)(1-2\pi_i) \leq \frac{1}{\tau q_n} \to 0$ as $n \to \infty$. Hence, we have $\lim_{n\to\infty} \mathbb{P}\left(B_n \leq q_n t + e_n\right) = \Phi(t)$ for all $t \in \mathbb{R}$.

Remark 4.13. The CLT for the sums B_n of independent random indicators is of interests of its own. Finner results were derived by, for example, Mikhailov [66], from which $\lim_{n\to\infty} \mathbb{P}\left(B_n \leq q_n t + e_n\right) = \Phi(t)$ can be obtained as a consequence.

Both proofs start with splitting the sum on the RHS of (4.5) into two sub-sums: over the domain $k \in \mathcal{C}_n^0 \cap \{0,\dots,n\}$ and over the complementary set, where $\mathcal{C}_n^0 := [e_n - a_n q_n, e_n + a_n q_n]$ for some $a_n > 0$. Recall from (4.7) that $e_n := \mathbb{E}B_n$ and $q_n^2 := \operatorname{Var} B_n$. The divergent sequence a_n will later be chosen such that $\mathbb{P}(B_n \in \mathcal{C}_n^0) = \sum_{k\in\mathcal{C}_n^0, 0\leq k\leq n} b_{n,k} \to 1$ as $n \to \infty$. That is, for any $\boldsymbol{x} \in \mathbb{R}^d$ and $\Delta > 0$, one has

$$\mathcal{L}_{\boldsymbol{X}_n^*}(\Delta(\boldsymbol{x})) = \sum_{k\in\mathcal{C}_n^0,\ 0\leq k\leq n} b_{n,k}F^{*k}(\Delta(\boldsymbol{x})) + \sum_{k\notin\mathcal{C}_n^0,\ 0\leq k\leq n} b_{n,k}F^{*k}(\Delta(\boldsymbol{x})). \tag{4.14}$$

Proof of Theorem 4.4. The proof of Theorem 4.4 takes six steps. In **Step 1** we approximate the first sum in (4.14) using Stone's Integral-local Theorem 2.18; in **Step 2** we show that the second sum in (4.14) is negligible compared to the first sum; in **Step 3** we "clean up" the summands and then obtain the desired result (4.8) for the case with $\mu = 0$; in **Step 4** we approximate the coefficients $b_{n,k}$ using Theorem 2.17 [65] (an extended version of Gnedenko's Arithmetic local limit Theorem 2.8 to the "non-identical case"); in **Step 5** we do Riemann sum approximation to get an integral expression; in **Step 6** we employ the idea of convolution integrals to obtain the desired result (4.9), which completes the proof.

Step 1. A natural first step is to approximate $F^{*k}(\Delta(x])$ in the summands of the first sum on the RHS of (4.14), since it is in the form of the local law in the Stone Integral-local Theorem 2.18. This will become clearer if we formulate it by considering a sequence of i.i.d. non-lattice random variables $\xi_1,\dots,\xi_k$ with law F such that $F^{*k}(\Delta(x]) = \mathbb{P}\left(\sum_{i=1}^k \xi_i \in \Delta(x]\right)$. Applying the Stone Integral-local Theorem 2.18 to the sequence of non-lattice i.i.d. random variables $\{\xi_i - \mu\}_{i\geq 1}$ gives

$$\mathbb{P}\left(\sum_{i=1}^k \xi_i \in \Delta(x]\right) = \frac{\Delta}{\sigma\sqrt{k}}\,\phi\left(\frac{x-k\mu}{\sigma\sqrt{k}}\right) + o\left(\frac{1}{\sqrt{k}}\right) \qquad \text{as } k \to \infty, \tag{4.15}$$

where the remainder term is uniform in $x \in \mathbb{R}$ and $\Delta \in [\Delta_L, \Delta_R]$ for any fixed

$0 < \Delta_L < \Delta_R < \infty$. Substituting (4.15) back into (4.14), the sum-up error term is

$$\sum_{k\in\mathcal{C}_n^0,\ 0\le k\le n} b_{n,k}o\left(\frac{1}{\sqrt{k}}\right) = \sum_{k\in\mathcal{C}_n^0,\ 0\le k\le n} b_{n,k}o\left(\frac{1}{\sqrt{e_n}}\right) = o\left(\frac{1}{\sqrt{e_n}}\right) \tag{4.16}$$

if a_n will be chosen properly (in the next step) such that $a_n q_n = o(e_n)$, from which one has $k = e_n\,(1+o(1))\ \forall k \in \mathcal{C}_n^0$ and hence the first equality in (4.16) holds. The second equality is simply due to $\sum_{0\le k\le n} b_{n,k} = 1$.

Step 2. A natural second step is to choose a_n properly such that the sum over $\{0,\ldots,n\}\backslash\mathcal{C}_n^0$ on the RHS of (4.14) has a consistent order with the error term from (4.16), i.e., $o\left(e_n{}^{-1/2}\right)$, when $n\to\infty$. First note that, for any $x\in\mathbb{R}$ and $\Delta>0$, one has $\sum_{k\notin\mathcal{C}_n^0,\ 0\le k\le n} b_{n,k}F^{*k}(\Delta(x]) \le \sum_{k\notin\mathcal{C}_n^0,\ 0\le k\le n} b_{n,k} = \mathbb{P}(B_n\notin\mathcal{C}_n^0)$. By this upper bound, by applying Bernstein's inequality (4.12) in Lemma 4.10 to the sequence $I_i - \pi_i$ with $t := a_n q_n$ and $M := 1$, one has

$$\mathbb{P}(B_n\notin\mathcal{C}_n^0) \le 2\exp\left(\frac{-a_n^2 q_n^2}{2q_n^2 + \frac{2}{3}a_n q_n}\right), \tag{4.17}$$

where the RHS will have the desired order $o\left(1/\sqrt{e_n}\right)$ if we choose

$$a_n := (1+\beta)\sqrt{\ln e_n} \qquad \text{for some } \beta>0. \tag{4.18}$$

One can check that, with this choice of a_n, the relation $a_n q_n = o(e_n)$ is satisfied and hence (4.16) always holds. Combining (4.14) – (4.18) gives

$$\mathcal{L}_{\boldsymbol{X}_n^*}(\Delta(x]) = \sum_{k\in\mathcal{C}_n^0,\ 0\le k\le n} b_{n,k}\frac{\Delta}{\sigma\sqrt{k}}\phi\left(\frac{x-k\mu}{\sigma\sqrt{k}}\right) + o\left(\frac{1}{\sqrt{e_n}}\right). \tag{4.19}$$

Step 3. Approximate the two k's in denominators on the RHS of (4.19) by e_n. It is reasonable to do so because, as $n\to\infty$, the width $2a_nq_n$ of $\mathcal{C}_n^0$ is negligible compared to the location e_n of its center. We end up with the following error term:

$$\varepsilon_1 := \sum_{k\in\mathcal{C}_n^0,\ 0\le k\le n} b_{n,k}\left[\frac{\Delta}{\sigma\sqrt{k}}\phi\left(\frac{x-k\mu}{\sigma\sqrt{k}}\right) - \frac{\Delta}{\sigma\sqrt{e_n}}\phi\left(\frac{x-k\mu}{\sigma\sqrt{e_n}}\right)\right].$$

Lemma 4.14. *The error term ε_1 has order $o\left(1/\sqrt{e_n}\right)$ as $n\to\infty$, uniform in $x\in\mathbb{R}$ and in $\Delta\in[\Delta_L,\Delta_R]$ for any fixed $0<\Delta_L<\Delta_R<\infty$.*

Proof of the Lemma. Suppose that we have already shown that the ratio $R := \frac{\Delta}{\sigma\sqrt{k}}\phi\left(\frac{x}{\sigma\sqrt{k}}\right)/\frac{\Delta}{\sigma\sqrt{e_n}}\phi\left(\frac{x}{\sigma\sqrt{e_n}}\right)$ is $1+o(1)$ for each $k\in\mathcal{C}_n^0$ as $n\to\infty$. Then $\varepsilon_1 =$

$\sum_{k\in\mathcal{C}_n^0,\ 0\le k\le n} b_{n,k}\frac{\Delta}{\sigma\sqrt{e_n}}\phi\left(\frac{x-k\mu}{\sigma\sqrt{e_n}}\right)o(1) = o\left(\frac{1}{\sqrt{e_n}}\right)\sum_{k\in\mathcal{C}_n^0,\ 0\le k\le n} b_{n,k}\phi\left(\frac{x-k\mu}{\sigma\sqrt{e_n}}\right)$, where the sum $\sum_{k\in\mathcal{C}_n^0,\ 0\le k\le n} b_{n,k}\phi\left(\frac{x-k\mu}{\sigma\sqrt{e_n}}\right) \le 1$ because $\sum_{k\in\mathcal{C}_n^0,\ 0\le k\le n} b_{n,k} = \mathbb{P}(B_n \in \mathcal{C}_n^0) \le 1$ with the coefficient of $b_{n,k}$ being $\phi\left(\frac{x-k\mu}{\sigma\sqrt{e_n}}\right) \le 1/\sqrt{2\pi} < 1$ for each k. Hence it suffices to show that $R = 1 + o(1)$ as $n \to \infty$. It turns out that the discussion of R will depend on the value of x. A natural way is to split it into two cases: when x is inside or outside the interval $[\mu e_n - a_n^2 q_n,\ \mu e_n + a_n^2 q_n]$. WLOG, we can assume that $\mu \neq 0$. Note that, for case when $\mu = 0$, the interval to be used is $[-a_n^2 q_n,\ +a_n^2 q_n]$, which is centered at 0, but nothing else in the following argument needs to be changed. Based on relations $k = e_n + O(a_n q_n)$ and $a_n q_n = o(e_n)$, we have

$$\begin{aligned}
R &= \sqrt{\frac{e_n}{k}}\exp\left\{\frac{(x-k\mu)^2}{2\sigma^2}\left(\frac{1}{e_n}-\frac{1}{k}\right)\right\} = \sqrt{\frac{e_n}{k}}\exp\left\{\frac{\mu^2}{2\sigma^2}\left(\frac{x/\mu}{k}-1\right)^2\frac{k}{e_n}(k-e_n)\right\}\\
&= \sqrt{\frac{e_n}{e_n+O(a_nq_n)}}\exp\left\{\frac{\mu^2}{2\sigma^2}\left(\frac{x/\mu}{e_n+O(a_nq_n)}-1\right)^2\frac{e_n+O(a_nq_n)}{e_n}O(a_nq_n)\right\}\\
&= (1+o(1))\exp\left\{\left(\frac{x/\mu}{e_n+O(a_nq_n)}-1\right)^2 O(a_nq_n)\right\},
\end{aligned}$$

where the last equality is due to $e_n/(e_n + O(a_n q_n)) = 1+o(1)$ and $(e_n+O(a_n q_n))/e_n = 1 + o(1)$ as $n \to \infty$.

Case 1: For $x \in [\mu e_n - a_n^2 q_n,\ \mu e_n + a_n^2 q_n]$, we call the ratio R in this region $R_\in$ and one has the relation $\left(\frac{x/\mu}{e_n+O(a_nq_n)}-1\right)^2 \le \left(\frac{e_n+a_n^2q_n/\mu-e_n-O(a_nq_n)}{e_n+O(a_nq_n)}\right)^2 = O_x\left(\frac{a_n^4q_n^2}{e_n^2}\right)$ which implies $R_\in = (1+o(1))\exp\left\{O_x\left(\frac{a_n^5q_n^3}{e_n^2}\right)\right\}$. From Assumption 4.3 and the choice (4.18) of a_n's, for any x and Δ, we will have $(a_n^5 q_n^3)/e_n^2 = O([\ln e_n]^{5/2}/\sqrt{e_n}) = o(1)$. Hence, for those $x \in [\mu e_n - a_n^2 q_n,\ \mu e_n + a_n^2 q_n]$, one has $R_\in = (1+o(1))\exp\{O_x(o(1))\} = (1 + o(1))(1 + o_x(1)) = 1 + o_x(1)$ as desired, where the second equality is due to $\exp\{o_x(1)\} = 1 + o_x(1)$.

Case 2: For the case when $x \notin [\mu e_n - a_n^2 q_n,\ \mu e_n + a_n^2 q_n]$, the discussion is simpler. Since those x satisfy $|x - \mu e_n|/q_n \ge a_n^2$, and hence $\phi\left(\frac{x-k\mu}{\sigma\sqrt{e_n}}\right) = \phi\left(\frac{x-k\mu}{\sigma q_n(1+o(1))}\right) \le \phi\left(\frac{a_n^2q_n/\mu-a_nq_n}{\sigma q_n(1+o(1))}\right) \le \phi\left(\frac{a_n^2}{\sigma\mu(1+o(1))}\right) = o_x(1)$. Similarly, one has $\phi\left(\frac{x-k\mu}{\sigma\sqrt{k}}\right) = o_x(1)$. Whence, we have $\epsilon_1 = o(1/\sqrt{e_n})$.

Combining these two cases gives us that $\epsilon_1 = o(1/\sqrt{e_n})$ uniformly in $x \in \mathbb{R}$ and in $\Delta > 0$. *This completes the proof of Lemma 4.14.*

By Lemma 4.14, relation (4.19) has now been approximated by

$$\mathcal{L}_{\boldsymbol{X}_n^*}(\Delta(x)) = \sum_{k\in\mathcal{C}_n^0,\ 0\le k\le n} b_{n,k}\frac{\Delta}{\sigma\sqrt{e_n}}\phi\left(\frac{x-k\mu}{\sigma\sqrt{e_n}}\right) + o\left(\frac{1}{\sqrt{e_n}}\right), \tag{4.20}$$

where the error term is proven to be uniform in $x\in\mathbb{R}$ and in $\Delta\in[\Delta_L,\Delta_R]$ for any fixed $0<\Delta_L<\Delta_R<\infty$.

One can observe from (4.20) that when $\mu=0$, the $\phi(\cdot)$ term will not depend on k. Hence the sum term becomes $\frac{\Delta}{\sigma\sqrt{e_n}}\phi\left(\frac{x}{\sigma\sqrt{e_n}}\right)\mathbb{P}(B_n\in\mathcal{C}_n^0)$. But when $n\to\infty$, $\mathbb{P}(B_n\in\mathcal{C}_n^0)=1+o(1)$. Hence, in the case $\mu=0$, we simply have $\mathcal{L}_{\boldsymbol{X}_n^*}(\Delta(x)) = \frac{\Delta}{\sigma\sqrt{e_n}}\phi\left(\frac{x}{\sigma\sqrt{e_n}}\right)+o\left(\frac{1}{\sqrt{e_n}}\right)$ by which we have obtained the result (4.8) in case (i).

From now on to the end of the proof, we will always assume that $\mu\neq 0$.

Step 4. Next we approximate coefficients $b_{n,k}=\mathbb{P}(B_n=k)$ in (4.20) using ϕ. Since B_n is a sum of independent but non-identically distributed random indicators, we apply Theorem 2.17 (with the shift parameter e_n and the scale parameter q_n). In order to obtain the limiting behaviour (2.33) from the Theorem, we need to firstly check its conditions.

The first condtion (2.31) (about the domain of attraction) has already been verified in Remark 4.12 and the second condition (2.32) can be verified as follows. Since the p.m.f. p_m in our case equals $\mathbb{P}(I_m=k)=\pi_m\mathbf{1}_{\{k=1\}}+(1-\pi_m)\mathbf{1}_{\{k=0\}}$, the denominator in (2.32) becomes $\sum_{m=1}^n\sum_{k\in\mathbb{Z}}[p_m(k)\wedge p_m(k+1)]=\sum_{m=1}^n[(1-\pi_m)\wedge\pi_m]$ for each n. But $(1-\pi_m)\pi_m\le(1-\pi_m)\wedge\pi_m$ for all m, whence $\frac{q_n^2}{\sum_{m=1}^n[(1-\pi_m)\wedge\pi_m]}\le 1$ for all $n\ge 1$ (recall that $q_n^2:=\sum_{m=0}^n(1-\pi_k)\pi_k$). By this we have verified condition (2.32). Now we can apply the result (2.33) from Theorem 2.17 to get

$$\mathbb{P}(B_n=k)=\frac{1}{q_n}\phi\left(\frac{k-e_n}{q_n}\right)+o\left(\frac{1}{q_n}\right)\qquad\text{as } n\to\infty, \tag{4.21}$$

where the error term is uniform in x. Combining (4.20) and (4.21) gives

$$\mathcal{L}_{\boldsymbol{X}_n^*}(\Delta(x)) = \sum_{k\in\mathcal{C}_n^0,\ 0\le k\le n}\frac{1}{q_n}\phi\left(\frac{k-e_n}{q_n}\right)\frac{\Delta}{\sigma\sqrt{e_n}}\phi\left(\frac{x-k\mu}{\sigma\sqrt{e_n}}\right)+\varepsilon_2+o\left(\frac{1}{\sqrt{e_n}}\right)$$

where the error term is $\varepsilon_2=o\left(\frac{1}{q_n}\right)\sum_{k\in\mathcal{C}_n^0,\ 0\le k\le n}\frac{\Delta}{\sigma\sqrt{e_n}}\phi\left(\frac{x-k\mu}{\sigma\sqrt{e_n}}\right)$.

Lemma 4.15. *The error term ε_2 has order $o\left(1/\sqrt{e_n}\right)$ as $n\to\infty$, which is uniform in $x\in\mathbb{R}$ and $\Delta\in[\Delta_L,\Delta_R]$ for any fixed $0<\Delta_L<\Delta_R<\infty$.*

Proof of Lemma 4.15. Let $\Xi_n:=\sum_{k\in\mathcal{C}_n^0,\ 0\le k\le n}\frac{1}{\sigma\sqrt{e_n}}\phi\left(\frac{x-k\mu}{\sigma\sqrt{e_n}}\right)$. Since $q_n\asymp\sqrt{e_n}$

(cf. Assumption 4.3), it suffices to show that $\Xi_n = O(1)$ as $n \to \infty$. By defining $h_n := \mu/(\sigma\sqrt{e_n})$ and $y_n := x/(\sigma\sqrt{e_n})$ one has

$$\Xi_n = \frac{1}{\mu} \sum_{k \in \mathcal{C}_n^0,\ 0 \le k \le n} h_n\, \phi\left(y_n - kh_n\right) \le \frac{1}{\mu} \sum_{k=-\infty}^{\infty} h_n\, \phi\left(y_n - kh_n\right),$$

where, for the inequality, we assumed that $\mu > 0$. (If $\mu < 0$, just take the absolute value of Ξ_n).The key in this proof is the following property: for any $z \in \mathbb{R}$ and $h \in (0, 1/2)$, one has

$$\phi(z) \le \min_{s \in [z-h,z]} \phi(s) + \min_{s \in [z,z+h]} \phi(s) \tag{4.22}$$

which can be proved as follows. There are three cases of z:

Case 1: When $z \ge h$, then $z > z - h \ge 0$. Since ϕ is monotonically decreasing in $[0, \infty)$, we know that $\phi(z) \le \phi(s)$ for each $s \in [z - h, z]$. Then $\phi(z) \le \min_{s \in [z-h,z]} \phi(s)$.

Case 2: When $z \le -h$, then by symmetry with Case 1 we have $\phi(z) \le \phi(s)$ for each $s \in [z, z + h]$. Then $\phi(z) \le \min_{s \in [z,z+h]} \phi(s)$.

Case 3: When $z \in (-h, h)$, then for any $s \in [z - h, z + h])$ one has $|s| \le 2h$. Hence $\phi(s) \ge \phi(2h) = \phi(0)e^{-2h^2} > \phi(0)/2$ since $e^{-2h^2} \ge 1/2$ when $|h| < 1/2$. Also $\phi(z) \le \phi(0)$ always. Now one has $\phi(z) < 2 \min_{s \in [z-h,z+h]} \phi(s) < \min_{s \in [z-h,z]} \phi(s) + \min_{s \in [z,z+h]} \phi(s)$. By this we proved the property (4.22).

Since $e_n \to \infty$ (cf. Assumption 4.2), there exists an $N \ge 1$ such that for all $n \ge N$, $h_n \in (0, 1/2)$. Hence

$$\begin{aligned} h_n \phi(y_n - kh_n) \le& h_n \min_{s \in [y_n - kh_n - h_n,\ y_n - kh_n]} \phi(s) + h_n \min_{s \in [y_n - kh_n,\ y_n - kh_n + h_n]} \phi(s) \\ \le& \int_{y_n-(k+1)h_n}^{y_n - kh_n} \phi(s) ds + \int_{y_n - kh_n}^{y_n-(k-1)h_n} \phi(s) ds, \end{aligned}$$

which implies

$$\begin{aligned} \sum_{k=-\infty}^{\infty} h_n \phi(y_n - kh_n) \le& \sum_{k=-\infty}^{\infty} \int_{y_n-(k+1)h_n}^{y_n - kh_n} \phi(s) ds + \sum_{k=-\infty}^{\infty} \int_{y_n - kh_n}^{y_n-(k-1)h_n} \phi(s) ds \\ =& 2 \int_{-\infty}^{\infty} \phi(s) ds = 2. \end{aligned}$$

Hence $|\Xi_n| \le 2/|\mu|$ for $n \ge N$. *This completes the proof of Lemma 4.15.*

By Lemma 4.15, the local probability has now been simplified to

$$\mathcal{L}_{\boldsymbol{X}_n^*}(\Delta(x]) = \frac{\Delta}{\sigma\sqrt{e_n}} \sum_{k\in\mathcal{C}_n^0,\ 0\le k\le n} \frac{1}{q_n}\phi\left(\frac{k-e_n}{q_n}\right)\phi\left(\frac{x-k\mu}{\sigma\sqrt{e_n}}\right) + o\left(\frac{1}{\sqrt{e_n}}\right). \tag{4.23}$$

Step 5. First notice that $\sum_{k=\lceil e_n-a_nq_n\rceil}^{\lfloor e_n+a_nq_n\rfloor}$ is equivalent to $\sum_{k\in\mathcal{C}_n^0\cap\{0,\dots,n\}}$, since $0 < e_n-a_nq_n < e_n+a_nq_n < n$ by choice of a_n in (4.18). Define $g\left(y\right) := \phi\left(y-\frac{e_n}{q_n}\right)\phi\left(\frac{x-\mu q_n y}{\sigma\sqrt{e_n}}\right)$ such that $g(k/q_n) = \phi\left(\frac{k-e_n}{q_n}\right)\phi\left(\frac{x-k\mu}{\sigma\sqrt{e_n}}\right)$ in summands of (4.23). Since g is Riemann integrable with repect to y, we can apply the left Riemann sum approximation of g over the interval $[\lceil e_n/q_n-a_n\rceil,\ \lfloor e_n/q_n+a_n\rfloor+1/q_n]$ with partition $P := \{[\lceil e_n/q_n-a_n\rceil,\ \lceil e_n/q_n-a_n\rceil+1/q_n),\ \dots,\ [\lfloor e_n/q_n+a_n\rfloor,\ \lfloor e_n/q_n+a_n\rfloor+1/q_n)\}$ gives

$$\sum_{k/q_n=\lceil e_n/q_n-a_n\rceil}^{\lfloor e_n/q_n+a_n\rfloor} g\left(\frac{k}{q_n}\right)\frac{1}{q_n} = \int_{\lceil e_n/q_n-a_n\rceil}^{\lfloor e_n/q_n+a_n\rfloor+1/q_n} g(y)dy + O\left(\frac{1}{q_n}\right) \qquad \text{as } n\to\infty, \tag{4.24}$$

where the error term comes from the fact that the error [30] [85] of the composite trapezoidal rule has magnitude $O(a_n^3/(a_nq_n)^2)g''(z)$ for some $z \in [\lceil e_n/q_n - a_n\rceil, \lfloor e_n/q_n+a_n\rfloor+1/q_n]$, and the second derivative of g is bounded by C/q_n^2 for some $C<\infty$. The integral on the RHS has nonnegative tail $\int_{\lfloor e_n/q_n+a_n\rfloor}^{\lfloor e_n/q_n+a_n\rfloor+1/q_n} g(y)dy \le 1/(2\pi q_n)$, so the RHS of (4.24) can be simplified to $\int_{\lceil e_n/q_n-a_n\rceil}^{\lfloor e_n/q_n+a_n\rfloor} g(y)dy + O\left(\frac{1}{q_n}\right)$.

To finish this step, the last thing we haven't discuss is the error

$$\varepsilon_3 := \sum_{k=\lceil e_n-a_nq_n\rceil}^{\lfloor e_n+a_nq_n\rfloor} g\left(\frac{k}{q_n}\right)\frac{1}{q_n} - \sum_{k/q_n=\lceil e_n/q_n-a_n\rceil}^{\lfloor e_n/q_n+a_n\rfloor} g\left(\frac{k}{q_n}\right)\frac{1}{q_n}$$

where $-1 \le \lceil e_n - a_nq_n\rceil/q_n - \lceil e_n/q_n - a_n\rceil \le 1/q_n$ and $-1/q_n \le \lfloor e_n + a_nq_n\rfloor/q_n - \lceil e_n/q_n + a_n\rceil \le 1$. As $n\to\infty$, there are at most two terms that can't be cancelled out in the above two sums: one with the lowest index and the other with the highest index. Since $g(y)\le 1/(2\pi)$, by the triangle inequality one has $|\varepsilon_3| \le \left(\frac{1}{2\pi}+\frac{1}{2\pi}\right)\frac{1}{q_n} = O\left(\frac{1}{q_n}\right)$ as $n\to\infty$. Then we have

$$\sum_{k=\lceil e_n-a_nq_n\rceil}^{\lfloor e_n+a_nq_n\rfloor} g\left(\frac{k}{q_n}\right)\frac{1}{q_n} = \int_{\lceil e_n/q_n-a_n\rceil}^{\lfloor e_n/q_n+a_n\rfloor} g(y)dy + O\left(\frac{1}{q_n}\right) \qquad \text{as } n\to\infty, \tag{4.25}$$

where the LHS equals $\sum_{k\in\mathcal{C}_n^0,\ 0\le k\le n}\frac{1}{q_n}\phi\left(\frac{k-e_n}{q_n}\right)\phi\left(\frac{x-k\mu}{\sigma\sqrt{e_n}}\right)$. Hence, as $n\to\infty$,

$$\mathcal{L}_{\boldsymbol{X}_n^*}(\Delta[x])=\frac{\Delta}{\sigma\sqrt{e_n}}\int_{\lceil e_n/q_n-a_n\rceil}^{\lfloor e_n/q_n+a_n\rfloor}\phi\left(y-\frac{e_n}{q_n}\right)\phi\left(\frac{x-\mu q_n y}{\sigma\sqrt{e_n}}\right)dy+o\left(\frac{1}{\sqrt{e_n}}\right). \quad (4.26)$$

Step 6. To simplify the integral in (4.26) we note that it is in the form of a convolution integral. To make it more obvious, we change the variables by letting $z:=\mu q_n y-\mu e_n$ such that (4.26) becomes

$$\mathcal{L}_{\boldsymbol{X}_n^*}(\Delta[x])=\frac{\Delta}{\sigma\sqrt{e_n}}\int_{\mathcal{C}_n}\phi\left(\frac{z}{\mu q_n}\right)\phi\left(\frac{x-z-\mu e_n}{\sigma\sqrt{e_n}}\right)\frac{dz}{\mu q_n}+o\left(\frac{1}{\sqrt{e_n}}\right), \quad (4.27)$$

where the region of integration is

$$\mathcal{C}_n:=[\lfloor e_n/q_n+a_n\rfloor\mu q_n-\mu e_n,\ \lceil e_n/q_n-a_n\rceil\mu q_n-\mu e_n]\ =\ [\lfloor a_n\rfloor\mu q_n,\ \lceil -a_n\rceil\mu q_n],$$

which is asymptotically centered at 0 and extends to $\mathbb{R}$ as $n\to\infty$. Let f_1 and f_2 be the p.d.f.'s of independent random variables $W_1\sim\mathcal{N}(0,\ \mu^2q_n^2)$ and $W_2\sim\mathcal{N}(\mu e_n,\ \sigma^2e_n)$, respectively. Then

$$f_1(z)=\frac{1}{\mu q_n}\phi\left(\frac{z}{\mu q_n}\right),\qquad f_2(x-z)=\frac{1}{\sigma\sqrt{e_n}}\phi\left(\frac{x-z-\mu e_n}{\sigma\sqrt{e_n}}\right).$$

Hence

$$\begin{aligned}\frac{1}{\sigma\sqrt{e_n}}\int_{\mathcal{C}_n}\phi\left(\frac{z}{\mu q_n}\right)\phi\left(\frac{x-z-\mu e_n}{\sigma\sqrt{e_n}}\right)\frac{dz}{\mu q_n}&=\int_{\mathcal{C}_n}f_1(z)\,f_2(x-z)dz\\&=\int_{\mathbb{R}}f_1(z)\,f_2(x-z)dz-\varepsilon_4.\end{aligned} \quad (4.28)$$

Recognize the integral in the first term as the convolution p.d.f. $f_1*f_2(x)$, which gives the p.d.f. of $W_1+W_2\sim\mathcal{N}(\mu e_n,\ \mu^2q_n^2+\sigma^2e_n)$. That is,

$$\int_{\mathbb{R}}f_1(z)\,f_2(x-z)dz=f_1*f_2(x)=\frac{1}{\sqrt{\mu^2q_n^2+\sigma^2e_n}}\phi\left(\frac{x-\mu e_n}{\sqrt{\mu^2q_n^2+\sigma^2e_n}}\right). \quad (4.29)$$

And the error term is

$$\varepsilon_4:=\int_{\mathbb{R}\setminus\mathcal{C}_n}f_1(z)\,f_2(x-z)dz\le\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}\setminus\mathcal{C}_n}f_1(z)dz\le\frac{2}{\sqrt{2\pi}}\mathbb{P}\left(W_1\ge(a_n-1)\mu q_n\right).$$

But the tail probability on the RHS can be estimated in the following way (W. Feller

[39], Section 7.1)[3] :

$$\mathbb{P}\left(W_1 \geq (a_n-1)\mu q_n\right) \leq \frac{1}{a_n-1}\frac{e^{-\frac{1}{2}(a_n-1)^2}}{\sqrt{2\pi}},$$

where the term $e^{-(a_n-1)^2/2} \asymp e^{-a_n^2/2} = e^{-(1+\delta)^2\ln e_n/2} = e_n^{-(1+\delta)^2/2} = o(1/e_n)$ as $n\to\infty$. Hence the error term in (4.28) is $\varepsilon_4 = o(1/\sqrt{e_n})$.

Combining (4.27) — (4.29) with our bound for ε_4, we have

$$\mathcal{L}_{\boldsymbol{X}_n^*}(\Delta[x]) = \frac{\Delta}{\sqrt{\mu^2q_n^2+\sigma^2e_n}}\phi\left(\frac{x-\mu e_n}{\sqrt{\mu^2q_n^2+\sigma^2e_n}}\right) + o\left(\frac{1}{\sqrt{e_n}}\right)$$

which is exactly (4.9). We proved both cases of Theorem 4.4 . □

The proof of the multi-dimensional case is as follows.

Proof of Theorem 4.5. This proof follows a similar argument to the previous proof. Again, we start with relation (4.14).

Step 1. To approximate $F^{*k}(\Delta(\boldsymbol{x}])$ in the summands of the first part on the RHS of relation (4.14), we can apply the multidimensional version of the Stone integro-local Theorem 2.24 since we are assuming that all margins of F are non-lattice. Consider a sequence of i.i.d. random vectors $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots$ with the common d-dimensional law $\boldsymbol{F}$ such that $\boldsymbol{F}^{*k}(\Delta(\boldsymbol{x}]) = \mathbb{P}\left(\sum_{i=1}^k \boldsymbol{\xi}_i \in \Delta(\boldsymbol{x}]\right)$. Applying the multi-dimensional Stone Integro-Local Theorem 2.24 gives

$$\mathbb{P}\left(\sum_{i=1}^k \boldsymbol{\xi}_i \in \Delta(\boldsymbol{x}]\right) = \frac{\Delta^d}{k^{d/2}}\frac{(2\pi)^{-(d-1)/2}}{\sqrt{\det\Sigma}}\,\phi\left(\frac{\|(\boldsymbol{x}-k\boldsymbol{\mu})\Sigma^{-1/2}\|}{\sqrt{k}}\right) + o\left(k^{-d/2}\right) \quad (4.30)$$

as $k\to\infty$, where the error term is uniform in $x\in\mathbb{R}$ and $\Delta\in[\Delta_L,\Delta_R]$ for any fixed $0<\Delta_L<\Delta_R<\infty$.

Combining (4.14) and (4.30) results in the error term

$$\sum_{k\in\mathcal{C}_n^0,\ 0\leq k\leq n} b_{n,k}\,o\left(k^{-d/2}\right) = \sum_{k\in\mathcal{C}_n^0,\ 0\leq k\leq n} b_{n,k}\,o\left(e_n^{-d/2}\right) = o\left(e_n^{-d/2}\right), \quad (4.31)$$

where the first relation will hold for a_n such that $a_nq_n = o(e_n)$ (by the same reason as we discussed for (4.16)).

Step 2. This step of choosing the sequence a_n properly can be argued in exactly the same way as in **Step 2** of the last proof. Recall from (4.18) that $a_n := (1+\beta)\sqrt{\ln e_n}$,

[3] A lower bound is also provided in [39]: $\mathbb{P}\left(W_1 \geq (a_n-1)\mu q_n\right) \geq \left(\frac{1}{a_n-1} - \frac{1}{(a_n-1)^3}\right)\frac{e^{-\frac{1}{2}(a_n-1)^2}}{\sqrt{2\pi}}$.

for some $\beta > 0$. Then one has

$$\mathcal{L}_{\boldsymbol{X}_n^*}(\Delta(\boldsymbol{x}]) = \frac{\Delta^d(2\pi)^{-(d-1)/2}}{\sqrt{\det\Sigma}} \sum_{k\in\mathcal{C}_n^0,\ 0\le k\le n} \frac{b_{n,k}}{k^{d/2}}\,\phi\left(\frac{\|(\boldsymbol{x}-k\boldsymbol{\mu})\Sigma^{-1/2}\|}{\sqrt{k}}\right) + o\left(e_n^{-d/2}\right). \tag{4.32}$$

Step 3. To further approximate the main term in (4.32), we replace both k's in the denominators by e_n, which leads to the error term

$$\epsilon_1 := \frac{\Delta^d(2\pi)^{-(d-1)/2}}{\sqrt{\det\Sigma}} \sum_{k\in\mathcal{C}_n^0,\ 0\le k\le n} \frac{b_{n,k}}{k^{d/2}}\,\phi\left(\frac{\|(\boldsymbol{x}-k\boldsymbol{\mu})\Sigma^{-1/2}\|}{\sqrt{k}}\right) - \frac{b_{n,k}}{{e_n}^{d/2}}\,\phi\left(\frac{\|(\boldsymbol{x}-k\boldsymbol{\mu})\Sigma^{-1/2}\|}{\sqrt{e_n}}\right).$$

Lemma 4.16. *The error term ϵ_1 has order $o\left(e_n^{-d/2}\right)$ as $n\to\infty$, uniform in $\boldsymbol{x}\in\mathbb{R}^d$ and in $\Delta\in[\Delta_L,\Delta_R]$ for any fixed $0<\Delta_L<\Delta_R<\infty$.*

Proof of the Lemma. Set $y := \|(\boldsymbol{x}-k\boldsymbol{\mu})\Sigma^{-1/2}\|$. It suffices to show that, for each $k\in\mathcal{C}_n^0$,

$$\widetilde{R} := \frac{k^{-d/2}\phi\left(y/\sqrt{k}\right)}{{e_n}^{-d/2}\phi\left(y/\sqrt{e_n}\right)} = 1+o(1) \qquad \text{as } n\to\infty.$$

Since if this hold, then

$$\begin{aligned}\epsilon_1 =& \frac{\Delta^d(2\pi)^{-(d-1)/2}}{\sqrt{\det\Sigma}} \sum_{k\in C_n^0,\ 0\le k\le n} \frac{b_{n,k}}{{e_n}^{d/2}}\,\phi\left(\frac{y}{\sqrt{e_n}}\right)\left(\widetilde{R}-1\right)\\ =& o\left({e_n}^{-d/2}\right) \sum_{k\in C_n^0,\ 0\le k\le n} b_{n,k}\phi\left(\frac{y}{\sqrt{e_n}}\right) = o\left({e_n}^{-d/2}\right)\end{aligned}$$

as $n\to\infty$, which is uniform in $\boldsymbol{x}$ and in $\Delta\in[\Delta_L,\Delta_R]$ for any fixed $0<\Delta_L<\Delta_R<\infty$ as desired. The last relation above is due to $\sum_{k\in C_n^0,\ 0\le k\le n} b_{n,k}\phi\left(y/\sqrt{e_n}\right) \le \sum_{k\in\mathcal{C}_n^0} b_{n,k}(2\pi)^{-1/2} \le (2\pi)^{-1/2}\mathbb{P}(B_n\in C_n^0) = O(1)$ as $n\to\infty$. However, as $n\to\infty$, the claim $\widetilde{R} = 1+o(1)$ can be directly obtained as a simple consequence of the fact $R = 1+o(1)$ that we proved before in the proof of Lemma 4.14. Since for any $\boldsymbol{x}\in\mathbb{R}^d$, the value of R evaluated at the corresponding $y := \|(\boldsymbol{x}-k\boldsymbol{\mu})\Sigma^{-1/2}\|$ equals

$$\widetilde{R}_y = \left(\frac{e_n}{k}\right)^{(d-1)/2} R_{\sigma y} = \left(\frac{e_n}{e_n+O(a_nq_n)}\right)^{(d-1)/2} R_{\sigma y} = (1+o(1))\,R_{\sigma y},$$

where the last equality is due to $a_nq_n = o(e_n)$ by Assumption 4.3 and the choice of a_n in **Step 2**. *This completes the proof of Lemma 4.16.*

By Lemma 4.16, relation (4.32) can be further approxiamted by

$$\mathcal{L}_{\boldsymbol{X}_n^*}(\Delta[\boldsymbol{x})) = \frac{\Delta^d(2\pi)^{-(d-1)/2}}{\sqrt{\det\Sigma}} \sum_{k\in\mathcal{C}_n^0,\ 0\le k\le n} \frac{b_{n,k}}{{e_n}^{d/2}}\,\phi\left(\frac{\|(\boldsymbol{x}-k\boldsymbol{\mu})\Sigma^{-1/2}\|}{\sqrt{e_n}}\right) + o\left(e_n^{-d/2}\right). \tag{4.33}$$

Observe from (4.33) that when $\mu = 0$, the ϕ component in the summands on the RHS will not depend on k. Then the local probability becomes

$$\mathcal{L}_{\boldsymbol{X}_n^*}(\Delta(\boldsymbol{x}]) = \frac{\Delta^d(2\pi)^{-(d-1)/2}}{e_n^{d/2}\sqrt{\det\Sigma}}\phi\left(\frac{\|\boldsymbol{x}\Sigma^{-1/2}\|}{\sqrt{e_n}}\right)\mathbb{P}(B_n\in\mathcal{C}_n^0) + o\left(e_n^{-d/2}\right).$$

When $n\to\infty$, $\mathbb{P}(B_n\in\mathcal{C}_n^0) = 1+o(1)$. Hence, for the case $\mu = 0$, we simply have

$$\mathcal{L}_{\boldsymbol{X}_n^*}(\Delta(\boldsymbol{x}]) = \frac{\Delta^d(2\pi)^{-(d-1)/2}}{e_n^{d/2}\sqrt{\det\Sigma}}\phi\left(\frac{\|\boldsymbol{x}\Sigma^{-1/2}\|}{\sqrt{e_n}}\right) + o\left(e_n^{-d/2}\right),$$

by which we have obtained the result (4.10) in case (i).

From now on to the end of the proof, we assume that $\mu \neq 0$.

Step 4. Next we approximate $b_{n,k} = \mathbb{P}(B_n = k)$ in (4.33) using ϕ (the 1-dimensional standard normal density). Since $b_{n,k}$ is independent of F, exactly the same argument as in the **Step 4** in the proof of the one-dimensional Theorem 4.4 applies here. Recall from (4.21) that, we have

$$\mathbb{P}(B_n = k) = \frac{1}{q_n}\phi\left(\frac{k-e_n}{q_n}\right) + o\left(\frac{1}{q_n}\right) \qquad \text{as } n\to\infty. \tag{4.34}$$

Combining (4.33) and (4.34) gives

$$\begin{aligned}\mathcal{L}_{\boldsymbol{X}_n^*}(\Delta(\boldsymbol{x}]) =& \frac{\Delta^d(2\pi)^{-(d-1)/2}}{\sqrt{\det\Sigma}} \sum_{k\in\mathcal{C}_n^0,\ 0\le k\le n} \frac{1}{q_n{e_n}^{d/2}}\phi\left(\frac{k-e_n}{q_n}\right)\phi\left(\frac{\|(\boldsymbol{x}-k\boldsymbol{\mu})\Sigma^{-1/2}\|}{\sqrt{e_n}}\right)\\ &+\epsilon_2 + o\left(e_n^{-d/2}\right)\end{aligned}$$

with the error term

$$\epsilon_2 := o\left(\frac{1}{q_n}\right)\frac{\Delta^d(2\pi)^{-(d-1)/2}}{\sqrt{\det\Sigma}} \sum_{k\in\mathcal{C}_n^0,\ 0\le k\le n} \frac{1}{{e_n}^{d/2}}\,\phi\left(\frac{\|(\boldsymbol{x}-k\boldsymbol{\mu})\Sigma^{-1/2}\|}{\sqrt{e_n}}\right).$$

Lemma 4.17. *The error term ϵ_2 has order $o\left(e_n^{-d/2}\right)$ as $n\to\infty$, uniform in $\boldsymbol{x}\in\mathbb{R}$ and $\Delta\in[\Delta_L,\Delta_R]$ for any fixed $0<\Delta_L<\Delta_R<\infty$.*

Proof of the Lemma. It suffices to show that, as $n \to \infty$,

$$U_n := \sum_{k \in \mathcal{C}_n^0,\ 0 \le k \le n} \frac{1}{\sqrt{e_n}} \phi \left(\frac{\|(\boldsymbol{x} - k\boldsymbol{\mu})\Sigma^{-1/2}\|}{\sqrt{e_n}} \right) = O(1),$$

uniformly in x. Since then $\epsilon_2 = e_n^{-d/2+1/2} o(q_n^{-1}) O(1) = o({e_n}^{-d/2})$ by Assumption 4.3, and is uniform in Δ in any bounded set. Comparing with the same step in the proof of the one-dimensional case, the sum U_n is nearly in the same form of S_n in Lemma 4.15, except for the numerator inside ϕ. It turns out that we can directly conclude that $U_n = O(1)$ from the proof of Lemma 4.15 if we can find an upper bound for U_n which is in the same form of the upper bound $\mu^{-1} \sum_{k \in \mathbb{Z}} h_n \phi(y_n - k h_n)$ of S_n in that proof. To find such an upper bound, we first use the reverse triangle inequality $\|(\boldsymbol{x} - k\boldsymbol{\mu})\Sigma^{-1/2}\| \ge \|\boldsymbol{x}\Sigma^{-1/2}\| - k\|\boldsymbol{\mu}\Sigma^{-1/2}\|$ and the shape of ϕ to obatin an upper bound which is in exactly the same form as S_n:

$$U_n \le \sum_{k \in \mathcal{C}_n^0,\ 0 \le k \le n} \frac{1}{\sqrt{e_n}} \phi \left(\frac{\|\boldsymbol{x}\Sigma^{-1/2}\|}{\sqrt{e_n}} - k \frac{\|\boldsymbol{\mu}\Sigma^{-1/2}\|}{\sqrt{e_n}} \right).$$

Now, by defining $\tilde{h}_n := \|\boldsymbol{\mu}\Sigma^{-1/2}\| / \sqrt{e_n}$ and $\tilde{y}_n := \|\boldsymbol{x}\Sigma^{-1/2}\| / \sqrt{e_n}$, the upper bound becomes $\|\boldsymbol{\mu}\Sigma^{-1/2}\|^{-1} \sum_{k \in \mathbb{Z}} \tilde{h}_n \phi(\tilde{y}_n - k\tilde{h}_n)$, for which all the arguments in the proof of Lemma 4.15 apply in exactly the same way. *This completes the proof of Lemma 4.17.*

By Lemma 4.17, the local probability is now simplified to

$$\mathcal{L}_{\boldsymbol{X}_n^*}(\Delta(\boldsymbol{x}]) = \frac{\Delta^d (2\pi)^{-\frac{d-1}{2}}}{{e_n}^{d/2} \sqrt{\det \Sigma}} \sum_{k \in \mathcal{C}_n^0,\ 0 \le k \le n} \frac{1}{q_n} \phi \left(\frac{k - e_n}{q_n} \right) \phi \left(\frac{\|(\boldsymbol{x} - k\boldsymbol{\mu})\Sigma^{-\frac{1}{2}}\|}{\sqrt{e_n}} \right) + o\left({e_n}^{-\frac{d}{2}} \right). \tag{4.35}$$

Step 5. We can now do the Riemann sum approximation. Again, we can make use of the ideas in **Step 5** in the proof of the one-dimensional case. Notice that in that **Step 5** (1-dim), the only special thing required about g is that g is Riemann integrable, with second derivative bounded by C/q_n^2 for some $C < \infty$, and $g(k/q_n)$ matches with the ϕ-product in summands. So if we define a function $\tilde{g}$ by

$$\tilde{g}(y) := \phi \left(y - \frac{e_n}{q_n} \right) \phi \left(\frac{\|(\boldsymbol{x} - \boldsymbol{\mu} q_n y)\Sigma^{-\frac{1}{2}}\|}{\sqrt{e_n}} \right),$$

so that $\tilde{g}$ is Riemann integrable over $\mathbb{R}$, with small second derivative in the sense described above, and $\tilde{g}(k/q_n)$ matches with the ϕ-product in the summands of (4.35), then all ideas in (4.24) — (4.25) in the one-dimensional proof appling in exactly the same way to $\tilde{g}$. Hence we will have

$$\mathcal{L}_{\boldsymbol{X}_n^*}(\Delta(\boldsymbol{x}]) = \frac{\Delta^d (2\pi)^{-\frac{d-1}{2}}}{{e_n}^{d/2}\sqrt{\det \Sigma}} \int_{\lceil e_n/q_n - a_n \rceil}^{\lfloor e_n/q_n + a_n \rfloor} \phi\left(y - \frac{e_n}{q_n}\right) \phi\left(\frac{\|(\boldsymbol{x} - \boldsymbol{\mu} q_n y)\Sigma^{-\frac{1}{2}}\|}{\sqrt{e_n}}\right) dy + o\left(e_n^{-\frac{d}{2}}\right). \tag{4.36}$$

Step 6. Convolution. By changing the variables $z := q_n y - e_n$, we normalize the region of integration in a way such that the region of integratoion $\widetilde{\mathcal{C}}_n = [\lceil -a_n \rceil q_n, \lfloor a_n \rfloor q_n]$ will expand to $\mathbb{R}$ as $n \to \infty$, and (4.36) becomes

$$\mathcal{L}_{\boldsymbol{X}_n^*}(\Delta(\boldsymbol{x}]) = \frac{\Delta^d (2\pi)^{-\frac{d-1}{2}}}{{e_n}^{d/2}\sqrt{\det \Sigma}} \int_{\widetilde{\mathcal{C}}_n} \phi\left(\frac{z}{q_n}\right) \phi\left(\frac{\|(\boldsymbol{x} - \boldsymbol{\mu} z - \boldsymbol{\mu} e_n)\Sigma^{-\frac{1}{2}}\|}{\sqrt{e_n}}\right) \frac{dz}{q_n} + o\left(e_n^{-\frac{d}{2}}\right). \tag{4.37}$$

Let $\widetilde{f}_1$ be the density of a random variable $\widetilde{W}_1 \sim \mathcal{N}(0, q_n^2)$, and let $\widetilde{f}_2$ be the density of a random vector $\widetilde{\boldsymbol{W}}_2 \sim \mathcal{N}_d(\boldsymbol{\mu} e_n, e_n \Sigma)$ that is independent of $\widetilde{W}_1$. Then

$$\widetilde{f}_1(z) = \frac{1}{q_n}\phi\left(\frac{z}{q_n}\right), \quad \widetilde{f}_2(\boldsymbol{x} - \boldsymbol{\mu} z) = \frac{(2\pi)^{-(d-1)/2}}{{e_n}^{d/2}\sqrt{\det \Sigma}}\phi\left(\frac{\|(\boldsymbol{x} - \boldsymbol{\mu} z - \boldsymbol{\mu} e_n)\Sigma^{-1/2}\|}{\sqrt{e_n}}\right),$$

and hence

$$\begin{aligned}\mathcal{L}_{\boldsymbol{X}_n^*}(\Delta(\boldsymbol{x}]) =& \Delta^d \int_{\widetilde{\mathcal{C}}_n} \widetilde{f}_1(z)\widetilde{f}_2(\boldsymbol{x} - \boldsymbol{\mu} z)dz + o\left(e_n^{-\frac{d}{2}}\right) \\ =& \Delta^d \int_{\mathbb{R}} \widetilde{f}_1(z)\widetilde{f}_2(\boldsymbol{x} - \boldsymbol{\mu} z)dz - \Delta^d \epsilon_4 + o\left(e_n^{-\frac{d}{2}}\right),\end{aligned} \tag{4.38}$$

where the error term

$$\begin{aligned}\epsilon_4 := \int_{\mathbb{R}\setminus\widetilde{\mathcal{C}}_n} \widetilde{f}_1(z)\widetilde{f}_2(\boldsymbol{x} - \boldsymbol{\mu} z)dz \leq& \frac{1}{(2\pi e_n)^{d/2}\sqrt{\det \Sigma}} \int_{\mathbb{R}\setminus\widetilde{\mathcal{C}}_n} \widetilde{f}_1(z)dz \\ \leq& \frac{2}{(2\pi e_n)^{d/2}\sqrt{\det \Sigma}} \mathbb{P}\left(\widetilde{W}_1 \geq (a_n - 1)q_n\right),\end{aligned}$$

where $(a_n - 1)q_n$ comes from the smallest possible right endpoint of $\widetilde{\mathcal{C}}_n$. Recall that $a_n := (1+\beta)\sqrt{\ln e_n}$ for some $\beta > 0$ from **Step 2**. That tail probability can be estimated in the following way (W. Feller [39], Section 7.1):

$$\mathbb{P}\left(\widetilde{W}_1 \geq (a_n - 1)q_n\right) \leq \frac{1}{a_n - 1}\frac{e^{-\frac{1}{2}(a_n-1)^2}}{\sqrt{2\pi}} \asymp \frac{e_n^{-(1+\delta)^2/2}}{\ln e_n} = o\left(\frac{1}{e_n}\right)$$

as $n \to \infty$. Hence $\epsilon_4 = o\left(e_n^{-d/2-1}\right) = o\left(e_n^{-d/2}\right)$. As for the integral in (4.38), it is in a "LOTP" form of the density of the random vector $\boldsymbol{\mu}\widetilde{W}_1 + \widetilde{\boldsymbol{W}}_2$ which is a sum of two independent normal random vectors of the same dimensionality d, where $\boldsymbol{\mu}\widetilde{W}_1 \sim \mathcal{N}(\mathbf{0}, {q_n}^2\boldsymbol{\mu}^T\boldsymbol{\mu})$. It can be proven by the method of ch.f.'s that the normality is preserved, and the mean and covariance matrix is as follows:

$$\boldsymbol{\mu}\widetilde{W}_1 + \widetilde{\boldsymbol{W}}_2 \sim \mathcal{N}_d\left(\boldsymbol{\mu}e_n, \boldsymbol{M}\right), \qquad \text{where } \boldsymbol{M} := {q_n}^2\boldsymbol{\mu}^T\boldsymbol{\mu} + e_n\Sigma.$$

Hence the integral in (4.38) becomes

$$\begin{aligned}\int_{\mathbb{R}} \widetilde{f}_1(z)\widetilde{f}_2(\boldsymbol{x} - \boldsymbol{\mu}z)dz =& \frac{1}{(2\pi)^{d/2}\sqrt{\det \boldsymbol{M}}} \exp\left\{-\frac{[\boldsymbol{x} - \boldsymbol{\mu}e_n]\boldsymbol{M}^{-1}[\boldsymbol{x} - \boldsymbol{\mu}e_n]^T}{2e_n}\right\} \\ =& \frac{1}{(2\pi)^{(d-1)/2}\sqrt{\det \boldsymbol{M}}}\phi\left(\frac{\|(x - \mu e_n)\boldsymbol{M}^{-1/2}\|}{\sqrt{e_n}}\right),\end{aligned}$$

where $\boldsymbol{M}^{-1/2}$ is the principle root of $\boldsymbol{M}^{-1}$ whose existence is guaranteed by the positive definiteness of $\boldsymbol{M}$. Hence (4.38) has been simplified to

$$\mathcal{L}_{\boldsymbol{X}_n^*}\left(\Delta(\boldsymbol{x}]\right) = \frac{\Delta^d}{(2\pi)^{(d-1)/2}\sqrt{\det \boldsymbol{M}}}\phi\left(\frac{\|(x - \mu e_n)\boldsymbol{M}^{-1/2}\|}{\sqrt{e_n}}\right) + o\left(e_n^{-d/2}\right)$$

by which we have proved the second case of Theorem 4.5. $\square$

4.4 Local Limit Theorems for lattice $\mathbf{X}_n^*$

For the location $\boldsymbol{X}_n^*$ of the "mother" chosen at time step $n+1$, in the case when each margin of F is lattice, we derive the following results for the limiting behaviour of its local probabilities: the one-dimensional case in Theorem 4.18; the higher-dimensional case in Theorem 4.19.

WLOG, we only state results for the arithmetic case of F, since one can always transform a lattice random variable to an arithmetic one by the simple linear transformation given in Remark 2.7.

Theorem 4.18 (the one-dimensional arithmetic case)**.** *Let the law F be arithmetic with mean $\mu < \infty$ and variance $\sigma^2 < \infty$. Under Assumptions 4.1 — 4.3, the local probability $P_{\boldsymbol{X}_n^*}(x) := \mathbb{P}(\boldsymbol{X}_n^* = x)$ of $\boldsymbol{X}_n^* \in \mathbb{R}$ has the following asymptotic behaviour:*

(i) *for the case with $\mu = 0$, one has*

$$P_{\boldsymbol{X}_n^*}(x) = \frac{1}{\sigma\sqrt{e_n}}\phi\left(\frac{x}{\sigma\sqrt{e_n}}\right) + o\left(\frac{1}{\sqrt{e_n}}\right); \tag{4.39}$$

(ii) *for the case with* $\mu \neq 0$, *one has*

$$P_{\mathbf{X}_n^*}(x) = \frac{1}{\sqrt{\mu^2 q_n^2 + \sigma^2 e_n}} \phi\left(\frac{x - \mu e_n}{\sqrt{\mu^2 q_n^2 + \sigma^2 e_n}}\right) + o\left(\frac{1}{\sqrt{e_n}}\right), \tag{4.40}$$

where the remainder term $o\left(1/\sqrt{e_n}\right)$ *in both cases is uniform in* $x \in \mathbb{Z}$.

Proof. The proof follows exactly the same argument as in the proof of Theorem 4.4 given in Section 4.3, except for **Step 1**. In the current case where we have arithmetic F, we need to start with applying Gendenko's Theorem 2.8 instead of the Stone's Theorem 2.18, which will give us

$$\mathbb{P}\left(\sum_{i=1}^{k} \xi_i = x\right) = \frac{1}{\sigma\sqrt{k}} \phi\left(\frac{x - k\mu}{\sigma\sqrt{k}}\right) + o\left(\frac{1}{\sqrt{k}}\right) \qquad \text{as } k \to \infty$$

uniformly in $x \in \mathbb{Z}$. If one compares the RHS of this relation with the RHS of (4.15), one finds that they are the same up to the positive constant Δ. Hence, every step after (4.15) in the previous proof can be equally applied to the current proof up to this factor Δ. Finally, one will obtain (4.8) and (4.9). □

Theorem 4.19 (the d-dimensional arithmetic case)**.** *Let the law F has each margin only taking integer values and the p.m.f. P corresponding to F satisfies Assumption 2.3. Let the mean vector $\boldsymbol{\mu}$ and covariance matrix Σ (corresponding to the law F) have all entries being finite. Assume Σ is non-singular (i.e., Σ^{-1} exists), together with Assumptions 4.1 —4.3, the local probability* $P_{\mathbf{X}_n^*}(\boldsymbol{x}) := \mathbb{P}(\boldsymbol{X}_n^* = \boldsymbol{x})$ *has the following asymptotic behaviour:*

(i) *for the case with* $\mu = 0$, *one has*

$$P_{\mathbf{X}_n^*}(\boldsymbol{x}) = \frac{(2\pi)^{-(d-1)/2}}{e_n^{d/2}\sqrt{\det \Sigma}} \phi\left(\frac{\|\boldsymbol{x}\Sigma^{-1/2}\|}{\sqrt{e_n}}\right) + o\left(e_n^{-d/2}\right); \tag{4.41}$$

(ii) *for the case with* $\mu \neq 0$, *put* $\boldsymbol{M} := {q_n}^2 \boldsymbol{\mu}^T \boldsymbol{\mu} + e_n \Sigma$, *one has*

$$P_{\mathbf{X}_n^*}(\boldsymbol{x}) = \frac{(2\pi)^{-(d-1)/2}}{\sqrt{\det \boldsymbol{M}}} \phi\left(\frac{\|(\boldsymbol{x} - \mu e_n)\boldsymbol{M}^{-1/2}\|}{\sqrt{e_n}}\right) + o\left(e_n^{-d/2}\right), \tag{4.42}$$

where the remainder term $o\left({e_n}^{-d/2}\right)$ *in both cases is uniform in* $\boldsymbol{x} \in \mathbb{Z}^d$.

Proof. The proof follows exactly the same argument as in the proof of Theorem 4.5 given in Section 4.3, except for **Step 1**. In the current case where we have F with each margin being arithmetic, we need to start with applying Rvačeva's Theorem 2.14 instead of the Stone's Theorem 2.24, which will give us

$$\mathbb{P}\left(\sum_{i=1}^{k} \boldsymbol{\xi}_i = \boldsymbol{x}\right) = \frac{(2\pi)^{-(d-1)/2}}{k^{d/2}\sqrt{\det \Sigma}}\, \phi\left(\frac{\|(\boldsymbol{x} - k\boldsymbol{\mu})\Sigma^{-1/2}\|}{\sqrt{k}}\right) + o\left(k^{-d/2}\right)$$

uniformly in $\boldsymbol{x} \in \mathbb{Z}^d$. If one compares the RHS of this relation with the RHS of (4.30), one finds that they are the same up to a positive constant scale Δ^d. Hence, every step after (4.30) in the previous proof can be equally applied to the current proof up to a scale Δ^d. Finally, one will obtain (4.10) and (4.11). □

4.5 Local Limit Theorems for combination of lattice and non-lattice X_n^*

Let $\boldsymbol{\xi}$ be a random vector in $\mathbb{R}^d$ with distribution F. In Section 4.2 and 4.4, we considered two cases: when $\boldsymbol{\xi}$ has all d coordinates being non-lattice, when $\boldsymbol{\xi}$ has all d coordinates being lattice. There is one more case can happen: when $\boldsymbol{\xi}$ has l coordinates being non-lattice and $d-l$ coordinates being lattice, where $1 \le l < d$.

WLOG, by virtue of Proposition 2.25, we can focus on the case when the first l coordinates of $\boldsymbol{\xi}$ are non-lattice and the remaining $d-l$ coordinates are arithmetic.

Theorem 4.20 (the combination of lattice and non-lattice)**.** *Let the law F has the distribution described above. Let the mean vector $\boldsymbol{\mu}$ and covariance matrix Σ (corresponding to the law F) have all entries being finite. Assume Σ is non-singular (i.e., Σ^{-1} exists), together with Assumptions 4.1 —4.3, then for the following direct product of half-open intervals in $\mathbb{R}^d$*

$$\Delta_l(\boldsymbol{x}] := \prod_{j=1}^{l}(x_j, x_j + \Delta] \prod_{j=l+1}^{d} (x_j, x_j + 1], \qquad \textit{for } \boldsymbol{x} = [x_1, \ldots, x_d],$$

the local law $\mathcal{L}_{\boldsymbol{X}_n^*}(\Delta_l(\boldsymbol{x}]) = \mathbb{P}(\boldsymbol{X}_n^* \in \Delta_l(\boldsymbol{x}])$ has the following asymptotic behaviour:

(i) *for the case with $\mu = 0$, one has*

$$\mathcal{L}_{\boldsymbol{X}_n^*}(\Delta_l(\boldsymbol{x}]) = \frac{\Delta^l (2\pi)^{-(d-1)/2}}{e_n^{d/2}\sqrt{\det \Sigma}}\phi\left(\frac{\|\boldsymbol{x}\Sigma^{-1/2}\|}{\sqrt{e_n}}\right) + o\left(e_n^{-d/2}\right); \tag{4.43}$$

(ii) *for the case with* $\mu \neq 0$*, put* $\boldsymbol{M} := {q_n}^2 \boldsymbol{\mu}^T \boldsymbol{\mu} + e_n \Sigma$*, one has*

$$\mathcal{L}_{\boldsymbol{X}_n^*}(\Delta_l(\boldsymbol{x}]) = \frac{\Delta^l (2\pi)^{-(d-1)/2}}{\sqrt{\det \boldsymbol{M}}} \phi\left(\frac{\|(\boldsymbol{x} - \mu e_n)\boldsymbol{M}^{-1/2}\|}{\sqrt{e_n}} \right) + o\left(e_n^{-d/2}\right), \quad (4.44)$$

where the remainder term $o\left(e_n^{-d/2}\right)$ *in both cases is uniform in* $\boldsymbol{x} \in \mathbb{R}^d$*, and in* $\Delta \in [\Delta_L, \Delta_R]$ *with fixed* $0 < \Delta_L < \Delta_R < \infty$*.*

Proof. The proof follows exactly the same argument as in the proof of Theorem 4.5 given in Section 4.3, except for **Step 1**. In the current case where F is a combination of lattice and non-lattice distribution, we need to start with applying the refined Stone's Theorem 2.26 instead of the original Stone's Theorem 2.24, which will give us

$$\mathbb{P}\left(\sum_{i=1}^{k} \boldsymbol{\xi}_i \in \Delta_l(\boldsymbol{x}] \right) = \frac{\Delta^l (2\pi)^{-(d-1)/2}}{k^{d/2}\sqrt{\det \Sigma}} \, \phi\left(\frac{\|(\boldsymbol{x} - k\boldsymbol{\mu})\Sigma^{-1/2}\|}{\sqrt{k}} \right) + o\left(k^{-d/2}\right)$$

uniformly in $\boldsymbol{x} \in \mathbb{R}^d$ and $\Delta \in [\Delta_L, \Delta_R]$ with fixed $0 < \Delta_L < \Delta_R < \infty$. If one compares the RHS of this relation with the RHS of (4.30), one finds that they are the same up to a positive constant factor. Hence every step after (4.30) in the previous proof can be equally applied to the current proof up to a constant factor. Finally, one will obtain (4.43) and (4.44). □

Bibliography

[1] Adams, W. J. (2009). *The Life and Times of the Central Limit Theorem* (2nd ed.). Providence, Rhode Island: American Mathematical Society.

[2] Aiello, W.; Bonato, A.; Cooper, C.; Janssen, J. and Pralat, P. (2009). A spatial web graphmodel with local influence regions. *Internet Mathematics*, **5**, 175–196.

[3] Archibald, E. E. A. (1948). Plant populations. I. A new application of Neyman's contagious distribution. *Ann. Bot. N. S.*, **12**, 221–235.

[4] Asmussen, S. and Kaplan, N. (1976). Branching random walks. I, II. *Stochastic Process. Appl.*, **4**, 1–13, 15–31.

[5] Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, **286**, 509–512.

[6] Bawly, G. M. (1937). O lokal'noĭ predel'noĭ teoreme teorii veroyatnosteĭ [On a local limit theorem in probability theory]. *Učenye zapiski Sverdlov. Univ.*, **2**, 7–24.

[7] Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. *Phil. Trans. R. Soc.*, **53**, 370–418.

[8] Beall, G. (1940). The fit and significance of contagious distibutions when applied to observations on larval insects. *Ecology*, **21**, 460–474.

[9] Beall, G. and Rescia, R. (1953). A Generalization of Neyman's Contagious Distributions. *Biometrics*, **9(3)**, 354–386.

[10] Bernoulli, J. (1713). *Ars conjectandi, opus posthumum. Accedit Tractatus de seriebus infinitis, et epistola gallicé scripta de ludo pilae reticularis*. Basel: Thurneysen Brothers.

[11] Bernoulli, J. (2005). *The Art of Conjecturing, together with Letter to a Friend on Sets in Court Tennis* (Edith Sylla, transl.). Baltimore: Johns Hopkins Univ Press.

[12] Bernstein, S. N. (1924). On a modification of Chebyshev's inequality and of the error formula of Laplace. *Ann. Sci. Inst. Sav. Ukraine, Sect. Math.* **1**.

[13] Berry, A. C. (1941). The Accuracy of the Gaussian Approximation to the Sum of Independent Variates. *Transactions of the American Mathematical Society.* **49**, 122–136.

[14] Biggins, J. D. (1997). *How fast does a general branching walk spread? In: Athreya, K.B., Jagers, P. (Eds.), Classical and Modern Branching Processes.*, Springer, New York.

[15] Blackwell, D. and Kendall, D. (1964). The Martin boundary for Pólya Urn scheme, and an Application to Stochastic Population Growth. *Journal of Applied Probability*, **1(4)**, 284–296.

[16] Blackwell, D. and McQueen, J. (1973). Ferguson Distributions via Polya Urn Schemes. *The Annals of Statistics*, **1(2)**, 353–355.

[17] Blyth, C. R. and Pathak, P. K. (1986). A Note on Easy Proofs of Stirling's formula. *Amer. Math. Monthly*, **93**, 376–379.

[18] Borovkov, A. A. and Mogulskii, A. A. (1999). Integro-local limit theorems including large deviations for for sums of random vectors. I. *Theory Probab. Appl.*, **43(1)**, 1–12.

[19] Borovkov, A. A. and Mogulskii, A. A. (2001). Integro-local limit theorems includeing large deviationsfor for sums of random vectors. II. *Theory Probab. Appl.*, **45(1)**, 3–22.

[20] Borovkov, A. A. (2011). Integro-Local and local theorems on normal and large deviations of the sums of nonidentically distributed random variables in the triangular array scheme. *Theory Probab. Appl.*, **54**, 571–587.

[21] Borovkov, A. A. (2013). *Probability theory* (5th ed.). London: Springer.

[22] Borovkov A. A. (2016). On the distribution of the first passage time of an arbitrary remote boundary by a random walk. *Theory Probab. Appl.*, **61(1)**, 1–24.

[23] Borovkov, A. A. (2017). Generalization and refinement of the integro-local Stone theorem for sums of random vectors. *Theory Probab. Appl.*, **61(4)**, 590–612.

[24] Borovkov, A. A. (2017). On the distribution of the first passage time of a random walk to an arbitrary remote boundary. *Theory Probab. Appl.*, **61**, , 235–254.

[25] Borovkov, A. A. and Borovkov, K. A. (2016). A refined version of the integro-local stone theorem. *submitted to Statistics and Probability Letters.*

[26] Borovkov, K. A. and Motyer, A. (2005). On the asymptotic behvaior of a simple growing point process. *Statist. Probab. Letters.*, **72**, 265–275.

[27] Borovkov, K. A. and Vatutin, V. A. (2006). On the asymptotic behaviour of random recursive trees in random environment. *Adv. Appl. Probab.*, **38**, 1047–1070.

[28] Borovkov, K. A. (2014). On the asymptotic behavior of a dynamic version of the Neyman contagious point process. *Stochastic Models*, **30**, 344–364.

[29] Breiman, L. (1968). *Probability*, Addison-Wesley, Reading, Mass.

[30] Buck, R. C. (1978). *Advanced Calculus* (3rd ed.). New York: McGraw-Hill.

[31] Darwin, J. H. (1951). Unpublished Ph.D. Thesis. Manchester University.

[32] Davis, B. and McDonald, D. (1995). An elementrary proof of the local central limit theorem. *Journal of the Theoratical Probability*, **8(3)**, 93–41.

[33] Durrett, R. (2007). *Random Graph Dynamics*. Cambridge University Press: Cambridge.

[34] Eggenberger, F. and Pólya, G. (1923). Über die Statistik verketer Vorgänge. *Zeitschirift Angewandte Mathematik and Mechanik*, **1**, 279–289.

[35] Eggenberger, F. and Pólya, G. (1928). Sur l'interprétation de certaines courbes de fréquence. *Comtes Rendus*, **187**, 870–872.

[36] Esseen, C. G. (1942). On the Liapounoff Limit of Error in the Theory of Probability. *Arkiv för Matematik, Astronomi och Fysik* A **28(9)**.

[37] Esseen, C. G. (1945). Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law. *Acta Math.* **77**, 1–125.

[38] Feller, W. (1967). A Direct Proof of Stirling's Formula. *The American Mathematical Monthly*, **74(10)**, 1223—1225.

[39] Feller, W. (1968). *An Introduction to Probability Theory and Its Applications* (Vol. 1, 3rd ed.). New York: Wiley.

[40] Fischer, H. (2011). *A History of the Central Limit Theorem: From Classical to Modern Probability Theory*, Sources and Studies in the History of Mathematics and Physical Sciences, New York: Springer.

[41] Fréchet, M. (1943). *Les Probabilités Associées à un Systéme d'Évenements Compatibles et Dépendants*. Paris: Hermann.

[42] Freĭman, G. A. (1956). An elementary method for the proof of limit theorems in probability theory. *Vestnik Leningrad. Univ.*, **1**, 57–73.

[43] Friedman, B. (1949). A simple urn model. *Comm. Pure Appl. Math.*, **2**, 59–70.

[44] Friedman, A. D. (1964). Bernard Friedman's Urn. *The Annals of Mathematical Statistics*, **36(3)**, 956–970.

[45] Gamkrelidze, N. G. (1964). On the local limit theorem for lattice random variables. *Theory of Prob. Appl.*, **9**, 662–665.

[46] Gamkrelidze, N. G. (2015). On local limit theorem for integer random vectors", *Theory Probab. Appl.*, **59(3)**, 494–499.

[47] Gnedenko, B. V. (1948). On the local limit theorem in the theory of probability. *Uspehi matem. nauk (N.S.)*, **3(3)**, 187–194.

[48] Gnedenko, B. V. (1953/54). On the local limit theorem for identically distributed independent terms. *Wiss. Z. Humboldt-Univ. Berlin. Math.-Naturwiss. Reihe*, **3(4)**, 287–293.

[49] Gnedenko, B. V. (1954). The local limit theorem for densities. *Dokl. AN SSSR*, **95(1)**, 5–7.

[50] Gnedenko, B. V. (1949). On the local limit theorem in the case of infinite variance. *Trudy mat. Inst. Akad. Nauk. Ukrain. SSR*, **12**, 22–30.

[51] Gnedenko, B. V. (1949). On the local limit theorem for stable limit distributions. *Ukrian. Mat. Zhurnal*, **4**, 3-15.

[52] Gnedenko, B. V. (1962). *Theory of Probability* [Translated from the 2nd Russian edition], Chelsea: New York.

[53] Gnedenko, B. V. and Kolmogorov, A.N. (1949/54). *Predel'nye raspredeleniya dlya summ nezavisimyhvelicin, Gostehizdat*, M.-L., 1949. *Limit Distributions for Sums of Independent Random Variables*[translated by K. L. Chung; with an appendix by J. L. Doob], MA: Addison-Wesley, 1954.

[54] Gnedenko, B. V. and Kolmogorov, A.N. (1959). *Grenzverteilungen von Summen unahhängiger ZufallsgröBen*, Berlin: Akademie-Verlag.

[55] Harris, T. E. (1963). *The Theory of BranchingProcesses.* Springer, Berlin.

[56] Khan, R. A. (1974). A Probabilistic Proof of Stirling's Formula. *The American Mathematical Monthly*, **81(4)**, 366–369.

[57] Khintchine, A. Ya. (1943). Mathematical foundations of statistical mechanics. *Gostehizdat, M.-L.*.

[58] Khintchine, A. Ya. (1951). Mathematical foundations of quantum statistics. *Gostehizdat, M.-L.*.

[59] Laplace, P. S. de (1812). *Théorie Analytique des Probabilités.* Paris: Courcier.

[60] Loève, M. (1963). *Probability theory* (3rd ed.). Van Nostrand, Princeton, N. J.

[61] Lyapunov, A. M. (1900). Sur une proposition de la théorie des probabilités. *Bulletin de l'Académie Impériale des Sciences de St.-Pétersbourg*, **13(5)**, 359–386.

[62] Lyapunov, A. M. (1901). Nouvelle forme du théoréme sur la limite de probabilité. *Mémoires de l'Académie Impériale des Sciences de St.-Pétersbourg VIIIe Série, Classe Physico-Mathématique*, **12**, 1–24.

[63] Markov, A. A. (1906). Extension of the law of large numbers to quantities depending on each other. *Izv. fizm.-mat. obsch. Kazanskom univ.* **2**, 135–156 (in Russian). [Reprinted in *J. Electron. Hist. Prob. Statist.* **2**, Article 10 (2006).]

[64] McDonald, D. (1979). On local limit theorems for integer valued random variables. *Theory of Prob. and Its Appl.*, **24**, 613–619.

[65] McDonald, D. (2005). The local central limit theorem: A historical perspective. *Journal of the Iranian Statistical Society*, **4(2)**, 73–86.

[66] Mikhailov, V. G. (1994). On a Refinement of the Central Limit Theorem for Sums of Independent Random Indicators. *Theory Probab. Appl.*, **38(3)**, 479–489.

[67] Mitalauskas, A. A. (1962). On the local limit theorem for stable limit distributions. *Theor. Probab. Appl.*, **7(2)**, 180–185.

[68] Moivre A. de (1711). De mensura sortis, seu, de probabilitate eventuum in ludis a casu fortuito pendentibus. *Phil. Trans. R. Soc.*, **27(329)**, 213—264.

[69] Moivre, A. de (1718). *The Doctrine of Chances.* London.

[70] Moivre, A. de (1738). *The Doctrine of Chances* (2nd ed.). London.

[71] Moskvin, D. A., Postnikova, L. P. and Yudin, A. A. (1970). On an arithmetic method for obtaining local limit theorems for lattice-distributed random variables. *Theor. Probab. Appl.*, **15(1)**, 87–98.

[72] Mukhin, A. (1984). Loc al limit theorems for distributions of sums of independent random vectors. *Theory Probab Appl.*, **29**, 369–375.

[73] Mukhin, A. B. (1991), Local limit theorems for lattice random variables. *Theory Probab. Appl.*, **35**, 698–713.

[74] Mitalauskas, A. A. and Statulevičius, V. A. (1966). Local limit theorems and asymptotic expansions for the sums of independent random variables with lattice distributions. *Litovskiĭ matem. sb.*, **6(4)**, 569–683.

[75] Neyman, J. (1939). On a new class of "contagious" distributions, applicable in entomology and bacteriology. *Ann. Math. Statist.*, **10**, 35–57.

[76] Pascal, B. (1665). *Traitédu triangle arithmétique: auec quelques autres petits traitez sur la mesme matiére.*

[77] Pipiras, V. and Statulevičius, V.A. (1966). Asymptotic expansions of sums of independent random variables. *Litovskiĭ matem. sb.*, **8(1)**, 137–151.

[78] Petrov, V. V. (1961). Asimptotičeskiĭ analiz nekotoryh predel'nyh teorem teorii veroyatnosteĭ. *Vestnik Leningrad. Univ.*, **7**, 51-61. An asymptotic analysis of some limit theorems in probability theory. (1965). *English transl. in "Selected Transl. on Math. Stat. and Probab."*, **5**, 179–190 .

[79] Petrov, V. V. (1964). On local limit theorems for sums of independent random variables. *Theor. Probab. Appl.*, **9(2)**, 312–320.

[80] Petrov, V. V. (1975). *Sums of Independent Random Variables.*, Springer: New York.

[81] Pólya, G. (1931). Sur quelques points de la théorie des probabilités. *Ann. Inst. Henri Poincaré*, **1**, 117–162.

[82] Prokhorov, Yu. V. (1952). A local theorem for densities. *Dokl. AN SSSR*, **83(6)**, 797–800.

[83] Prokhorov, Yu. V. (1954). On a local theorem for lattice distributions. *Dokl. AN SSSR*, **98(4)**, 535–538.

[84] Rozanov, Yu. (1957). On a local limit theorem for lattice distributions. *Theory Probab. Appl.*, **2**, 260–265.

[85] Rudin, W. (1976). *Principles of mathematical analysis.* McGraw-Hill: New York.

[86] Rvačeva, E. L. (1962). On domains of attraction of multi-dimensional distributions [Translated from the original 1954 Russian paper]. *Transl. Math. Statist. and Probability*, **2**, 183–205.

[87] Šahaĭdarova, N. (1966). Uniform local and global theorems for densities. *Izv. AN UzSSR, seriya fiz.-matem. nauk*, **5**, 90–91.

[88] Schneider, I. (2006). Direct and Indirect Influences of Jacob Bernoulli's Ars Conjectandi in 18th Century Great Britain. *Electronic Journal for the History of Probability and Statistics*, **2(1)**, 1–17.

[89] Shepp, L. A. (1964). A local limit theorem. *Ann. Math. Statist.*, **35**, 419–423.

[90] Shiryaev, A. N. (1984). *Probability-1.* New York: Springer.

[91] Sinai, Y. G. (1992). *Probability Theory.* Springer Textbook. Springer, Berlin, Heidelberg.

[92] Siraždinov, S. H. and Mamatov, M. (1962). On convergence in the mean for densities. *Theor. Probab. Appl.*, **7(4)**, 424–428.

[93] Siraždinov, S. H. and Mamatov, M. (1963). *On global limit theorems for densities and distributions.* Sb. "Predel'nye teoremy teorii veroyatnostei", Izd.-vo AN UzSSR, Taškent, 91–107.

[94] Siraždinov, S. H. and Šahaĭdarova, N. (1965). On a uniform local theorem for densities. *Izv. AN UzSSR, seriya fiz.-matem. nauk*, **6**, 30–36.

[95] Skellam, J. G. (1952). Studies in statistical ecology. I. Spatial pattern. *Biometrika*, **39**, 346.

[96] Statulevičius, V.A. (1965). Limit theorems for densities and asymptotic expansions of distributions of sums of independent random variables. *Theor. Probab. Appl.*, **10(4)**, 582–595.

[97] Statulevičius, V. A. (1967). On the limit theorems for stable limit distributions. *Litovskiĭ matem. sb.*, **7(2)**, 321–328.

[98] Stone, C. (1965). A local limit theorem for nonlattice multidimensional distribution functions. *Ann. math. Statist.*, **36**, 546–551.

[99] Stone, C. (1966). On local and ratio limit theorems. *In Proceedings of the Fifth Berkeley Symposium of Math. Statistics and Probability* (Vol. II, Part II, pp. 217–224). Berkeley, Los Angeles: University of California.

[100] Survila, P. (1963). Asymptotic expansions for densities. *Litovskiĭ matem. sb.*, **3(2)**, 177–191.

[101] Thomas, M. (1949). A generalization of Poisson's binomial limit for use in ecology. *Biometrika*, **36**, 18.

[102] Thompson, H. R. (1954). A note on contagious distributions. *Biometrika*, **41**, 268–271.

[103] Tyurin, I. S. (2010). Refinement of the upper bounds of the constants in Lyapunov's theorem. *Uspekhi Mat. Nauk*, **65(393)**, 201–202.

[104] von Mises, R. (1931). *Vorlesungen aus dem Gebiete der angewandten Mathematik* (Vol. 1). Leipzig–Wien.

[105] von Mises, R. (1934). Generalizzazione di un teorema sulla probabilita della somma di un numera illimitato di variabili casuali. *Giorn. Ist. Ital. Attuari*, **5**, 483–495.

Index of Notation

Spaces and Operations

$\mathbb{R}^d$	d-dimensional Euclidean space.
$\mathbb{N}$	set of positive integers $\{1, 2, 3, \ldots\}$.
$\boldsymbol{v}^T$	the transpose of (row) vector $\boldsymbol{v}$.
$\|\boldsymbol{v}\|$	the Euclidean norm of (row) vector $\boldsymbol{v}$.
$\det A$	the determinate of matrix A.
A^{-1}	the inverse matrix of matrix A.
$A^{1/2}$	the principal square root matrix of matrix A.

Distributions and Functions

$\delta_{\boldsymbol{a}}$	the degenerate distribution at point $\boldsymbol{a}$ in $\mathbb{R}^d$.
$\mathrm{Unif}[\boldsymbol{a}, \boldsymbol{b}]$	the (continuous) uniform distribution over the straight line segment connecting points $\boldsymbol{a}$ and $\boldsymbol{b}$ in $\mathbb{R}^d$.
$\mathcal{N}(\mu, \sigma^2)$	the normal (Gaussian) distribution with parameters $(\mu,\ \sigma^2)$.
$\Phi(\cdot)$	the distribution function of the standard normal law in $\mathbb{R}$.
$\phi(\cdot)$	the density of the standard normal law in $\mathbb{R}$.
F^{*k}	k-fold convolution of F with itself.
$F(k-)$	the limit of $F(x)$ when x approaches k from the left, written as $\lim_{x\to k-} F(x)$.
$\ln(\cdot)$	the natural logarithm function.
$\mathbf{1}_E$	indicator function of set E.
$a \wedge b$	notation $a \wedge b$ means take the minimum one from a and b.

Relations

$:=$	defining equality.
$\asymp$	notation $a_n \asymp b_n$ means that there $\exists\, C, D > 0$: $C\|a_n\| \leq \|b_n\| \leq D\|a_n\|$.
$o(\cdot)$	notation $a_n = o(b_n)$ means that $\lim_{n\to\infty} \frac{a_n}{b_n} = 0$.
$o_x(\cdot)$	notation $a_n = o_x(b_n)$ means that $\lim_{n\to\infty} \frac{a_n}{b_n} = 0$ non-uniformly in x.
$O(\cdot)$	notation $a_n = O(b_n)$ means that $\lim_{n\to\infty} \frac{a_n}{b_n} < \infty$.

$O_x(\cdot)$ notation $a_n = O_x(b_n)$ means that $\lim_{n\to\infty} \frac{a_n}{b_n} < \infty$ non-uniformly in x.

$\sim$ means that the random variable on the LHS is distributed as the distribution on the RHS.

$\xrightarrow{p}$ convergence of random variables in probability.

$\Rightarrow$ weak convergence of random variables (distributions).

Abbreviations

i.i.d. independent and identically distributed.

ch.f. characteristic function.

p.g.f. probability generating function.

p.m.f. probability mass function.

p.d.f. probability density function.

LOTP law of total probability.

LLN Law of Large numbers.

ILT Integral Limit Theorem.

LLT Local Limit Theorem.

I-LLT Integro-Local Limit Theorem.

MCT Monotone Convergence Theorem.

DCT Dominated Convergence Theorem.

WLOG without loss of generality.