1. Introduction

The classical De Moivre-Laplace theorem says that if $\xi_1, \xi_2, \ldots$ are independent identically distributed (i.i.d.) 0–1 Bernoulli random variables taking on 1 with probability $p$ and $S_N = \sum_{n=1}^{N} \xi_n$ then the probability $P(S_N = k)$ is equivalent as $N \to \infty$ to $(2\pi Npq)^{-1/2} \exp\left(-\frac{(k - Np)^2}{2Npq}\right)$, $q = 1 - p$ uniformly in $k$ such that $|k - Np| = o(Npq)^{2/3}$. It turns out that this type of results can be extended to "nonconventional" sums of the form

$$S_N = \sum_{n=1}^{N} \prod_{j=1}^{\ell} F(\xi_n, \xi_{2n}, \ldots, \xi_{\ell n})$$

where $\{\xi_n\}$ are either an i.i.d. random variables or they form a sufficiently fast mixing Markov chain.

In fact, we will deal here with more general sums of the form

$$(1.1) \quad S_N = \sum_{n=1}^{N} F(\xi_n, \xi_{2n}, \ldots, \xi_{\ell n}).$$

When $F(x_1, \ldots, x_{\ell}) = \prod_{j=1}^{\ell} \mathbb{1}_{\Gamma}(x_j)$, where $\mathbb{1}_{\Gamma}(x) = 1$ if $x \in \Gamma$ and $= 0$ if $x \not\in \Gamma$, then $S_N$ counts the number of events $\{\xi_n, \xi_{2n}, \ldots, \xi_{\ell n} \in \Gamma\}$, $j \leq N$ representing multiple returns to $\Gamma$. The name "nonconventional" comes from [9] where $L^2$ ergodic theorems for such sums with $F(x_1, \ldots, x_{\ell}) = \prod_{j=1}^{\ell} f_j(x_j)$ were studied. Recently, strong laws of large numbers and the central limit theorem type results for sums (1.1) were obtained in [15] and [17], respectively, while other related limit theorems
for such sums were studied in a series of papers of the second author and co-authors. Nonconventional limit theorems is probably the first topic in probability whose original motivation comes from the ergodic theory.

Modern proofs of local limit theorems are usually based on the perturbation theory of Fourier operators approach going back to [21] which together with quasi-compactness (or spectral gap, non-arithmeticity) assumptions yields appropriate estimates of characteristic functions of sums under consideration implying the result (see, for instance, [11]). The characteristic functions of the sums (1.1) cannot be studied via iterates of one Fourier operator, and so we cannot benefit here from the full strength of the above method though some elements of existing proofs of local limit theorems will be employed here, as well. We observe that a local limit theorem can only be formulated if the variance in the corresponding central limit theorem is positive and since the latter question was not studied extensively before in the nonconventional setup we deal with it here, as well. Usually, proofs of local limit theorems rely, in particular, on a corresponding central limit theorem result and we employ this argument also here but, in fact, we consider a slightly different from [17] version of a nonconventional central limit theorem which will be introduced in the next section.

This paper is organized as follows. In the next section we state our main results and comment on them. In Section 3 we derive our results on positivity of the limiting variance while in Sections 4–6 our nonconventional local limit theorems will be proved.

2. Preliminaries and main results

Our setup consists of a probability space $(\Omega, \mathcal{F}, P)$ together with a stationary Markov chain $\xi_0, \xi_1, \xi_2, \ldots$ evolving on a Polish space $\mathcal{X}$ equipped with a Borel $\sigma$-algebra $\mathcal{B}$. Let $P(x, \Gamma) = P(\xi_1 \in \Gamma | \xi_0 = x)$ be the transition probability of the Markov chain $\{\xi_n\}$, $\mu$ be its stationary probability so that $\int d\mu(x)P(x, \Gamma) = \mu(\Gamma)$ for any $\Gamma \in \mathcal{B}$, and let $F = F(x_1, \ldots, x_\ell)$, $\ell \geq 1$ be a Borel function on $\mathcal{X}^\ell = \mathcal{X} \times \cdots \times \mathcal{X}$ such that

$$b^2 = \int F^2(x_1, \ldots, x_\ell) d\mu(x_1) \ldots d\mu(x_\ell).$$

Our main goal in this paper is to derive a local limit theorem for sums given by (1.1) and in order to simplify formulas we assume also a centering condition

$$\bar{F} = \int F(x_1, \ldots, x_\ell) d\mu(x_1) \ldots d\mu(x_\ell) = 0$$

which is not a restriction since we always can replace $F$ by $F - \bar{F}$.

Introduce $\sigma$-algebras $\mathcal{F}_n = \sigma\{\xi_j, j \leq n\}$ and $\mathcal{F}^m = \sigma\{\xi_j, j \geq m\}$ and define the $\psi$-mixing (dependence) coefficient by

$$\psi(m) = \sup \left\{ \left| \frac{P(\Gamma \cap \Psi)}{P(\Gamma)P(\Psi)} - 1 \right| : \Gamma \in \mathcal{F}_n, \Psi \in \mathcal{F}^{n+m}; P(\Gamma), P(\Psi) > 0 \right\}.$$

The process $\{\xi_n\}$ is called $\psi$-mixing if all $\psi(m)$ are finite and $\psi(m) \to 0$ as $m \to \infty$. To avoid excessive technicalities we will work here under

2.1. Assumption. For some $\alpha > 0$ and any $m \geq 1$,

$$\psi(m) \leq \alpha^{-1} e^{-\alpha m}.$$
Local Limit Theorem 3

The exponentially fast decay (2.1) of the \( \psi \)-mixing coefficient can be relaxed to some polynomial decay but it is known from [5] that in the Markov chains case any decay \( \psi(m) \to 0 \) as \( m \to \infty \) yields already an exponentially fast decay of \( \psi(m) \).

It is known also (see, for instance, [3], Ch. 7 and 21) that (2.4) holds true if \( \{ \xi_n \} \) is a finite state irreducible and aperiodic Markov chain and, in fact, [5] provides necessary and sufficient conditions for (2.4) to take place. In particular, (2.4) will be satisfied if there exists a positive integer \( n_0 \), a probability measure \( \eta \) on \( \mathcal{X} \) and a number \( \gamma \in (0, 1] \) such that for any \( x \in \mathcal{X} \) and a Borel set \( \Gamma \subset \mathcal{X} \),

\[
\gamma^{-1}\eta(\Gamma) \geq P(n_0, x, \Gamma) \geq \gamma\eta(\Gamma),
\]

where \( P(k, x, \cdot) \) is the \( k \)-step transition probability of the Markov chain \( \{ \xi_n \} \). Observe that employing the technique from [17] we can obtain our results under weaker mixing conditions on expense of other assumptions on the function \( F \). Observe that the right hand side of (2.5) implies also the geometric ergodicity condition

\[
\| P(n, x, \cdot) - \mu \| \leq \beta^{-1}e^{-\beta n}, \beta > 0,
\]

where \( \| \cdot \| \) is the total variation norm, and under (2.6) the results of the present paper can be obtained for any initial distribution of the Markov chain \( \{ \xi_n \} \) and not only for the stationary one.

As usual, our local limit theorem will rely on a version of a nonconventional central limit theorem which will be presented in a more general form than needed here and under slightly different assumptions than in [17]. It will be convenient to represent the function \( F = F(x_1, \ldots, x_\ell) \) in the form

\[
F = F_1(x_1) + \cdots + F_\ell(x_1, \ldots, x_\ell)
\]

where

\[
F_\ell = F(x_1, \ldots, x_\ell) - \int F(x_1, \ldots, x_\ell) d\mu(x_\ell)
\]

and for \( i < \ell \),

\[
F_i(x_1, \ldots, x_i) = \int F(x_1, \ldots, x_i) d\mu(x_{i+1}) \cdots d\mu(x_\ell) - \int F(x_1, \ldots, x_\ell) d\mu(x_i) \cdots d\mu(x_\ell)
\]

which ensures, in particular, that for all \( x_1, \ldots, x_{i-1} \in \mathcal{X} \),

\[
\int F_i(x_1, \ldots, x_{i-1}, x_i) d\mu(x_i) = 0.
\]

Now we write

\[
S_N(t) = \sum_{i=1}^\ell S_{i,N}(t) \quad \text{where} \quad S_{i,N}(t) = \sum_{n=1}^{[N/t]} F_i(\xi_n, \ldots, \xi_{in})
\]

and we abbreviate \( S_N = S_N(1) \) and \( S_{i,N} = S_{i,N}(1) \).

2.2. Theorem. Suppose that (2.1), (2.2) and Assumption 2.1 hold true. Then the \( \ell \)-dimensional process \( \{ N^{-1/2}S_{j,N}(t/j) : 1 \leq j \leq \ell \} \) converges in distribution to a Gaussian process \( \{ \zeta_j(t) : 1 \leq j \leq \ell \} \) with stationary independent increments, zero means and covariances having the form \( E(\zeta_i(s)\zeta_j(t)) = D_{i,j} \min(s, t), i, j = 1, \ldots, \ell \).
The process $N^{-1/2}S_N(\cdot)$ itself converges in distribution to the Gaussian process $\zeta(\cdot)$ having a representation in the form

$$\zeta(t) = \sum_{i=1}^{\ell} \zeta_i(it)$$

which may have dependent increments.

In spite of different assumptions on the function $F$ here in comparison to [17] this theorem follows in the same way as the main result of [17] which will be explained at the beginning of Section 3.

A local limit theorem can only be meaningful if the variance of a limiting Gaussian distribution is strictly positive. This question was addressed in [14] only in a very particular case (which corresponds to $\ell = 1$ in our setup) and it was not dealt with at all in [17]. Here we will establish some sufficient conditions for positivity of the limiting variance in our Markov chains setup.

2.3. Theorem. (i) Set $\sigma^2_N = \text{var}S_N = E(S_N - ES_n)^2$ and suppose that (2.4) holds true. Then the limit

$$\lim_{N \to \infty} \frac{1}{N} \sigma^2_N = \sigma^2 = \lim_{N \to \infty} \frac{1}{N} ES_N^2$$

exists.

(ii) Let $\{\xi_n^{(1)}\}, \{\xi_n^{(2)}\}, \ldots, \{\xi_n^{(\ell)}\}$ be $\ell$ independent copies of the stationary Markov chain $\{\xi_n\}$ with the initial distribution $\mu$ and set $U_{\ell,N} = \sum_{n=0}^{N} F_\ell(\xi_n^{(1)}, \xi_n^{(2)}, \ldots, \xi_n^{(\ell)})$. Then

$$s_\ell^2 = \lim_{N \to \infty} \frac{1}{N} \text{var}U_{\ell,N} = \lim_{N \to \infty} \frac{1}{N} EU_{\ell,N}^2$$

exists and

$$\sigma^2 \geq \frac{1}{2\ell} s_\ell^2.$$

Of course, if $F_\ell = 0$ $\mu^\ell$-a.e. then $s_\ell = 0$ but in this case $F$ depends essentially only on $\ell - 1$ variables and we can apply the same arguments with $\ell - 1$ in place of $\ell$. Next, observe that specifying a bit our estimates it is possible to improve the right hand side of (2.13) to $\frac{1}{\ell} s_\ell^2$. Nevertheless, the main purpose of Theorem 2.3(ii) is to obtain a sufficient condition for positivity of $\sigma^2$ relying on the well known conditions of positivity of the limiting variance of the stationary Markov chain $\Xi_n^{(\ell)} = (\xi_n^{(1)}, \xi_n^{(2)}, \ldots, \xi_n^{(\ell)})$ which is $\psi$-mixing with the same $\psi(n)$-coefficient as $\{\xi_n\}$ itself. Namely, $s_\ell^2 > 0$ unless $F_\ell$ can be represented in the form $F_\ell(\Xi_1^{(\ell)}(\omega)) = g(\theta \omega) - g(\omega)$ where $\theta$ is the $P$-preserving paths’ shift transformation for the stationary process $\{\Xi_n\}$ so that $\Xi_n^{(\ell)} = \Xi_0^{(\ell)}(\theta^n \omega)$ (see Chapter 18 in [12]). We observe also that the assertion of Theorem 2.3 remain valid for any stationary (and not only Markov) process with a $\psi$-mixing coefficient decaying much slower than in (2.4) (see the proof in the next section). Furthermore, these assertions hold also true for sufficiently fast $\psi$-mixing dynamical systems (such as subshifts of finite type, Anosov diffeomorphisms, expanding transformations etc.) where in place of independent copies of the process $\{\xi_n\}$ we should consider the product of $\ell$ copies of the corresponding dynamical system.
Introduce correlation coefficients
\begin{equation}
\rho_k = \sup \{ \text{corr}(X, Y) : X \in L_2(\Omega, \mathcal{F}_n, \mu), Y \in L_2(\Omega, \mathcal{F}_{n+k}, P) \} = \|Q\|_{L_2^2(\mu)} \tag{2.15}
\end{equation}
where \(Qg(x) = \int_X P(x, dy)g(y)\) is the transition operator and \(L_2^0(\mathcal{X}, \mathcal{B}, \mu) = \{g \in L_2(\mathcal{X}, \mathcal{B}, \mu) : \int g d\mu = 0\}\). We consider also contraction coefficients
\begin{equation}
\delta_k = \sup_{x,y \in \mathcal{X}, \Gamma \in \mathcal{B}} |P(k, x, \Gamma) - P(k, y, \Gamma)|. \tag{2.16}
\end{equation}

### 2.4. Assumption
Either
\begin{equation}
\rho_1 < 1 \quad \text{or} \quad \delta_1 < 1. \tag{2.17}
\end{equation}

Next, we provide additional more specific sufficient conditions for the positivity of \(\sigma^2\).

#### 2.5. Theorem
(i) Let \eqref{2.4} and either \(\rho_1 < 1\) or \(\delta_1 < 1\) holds true. If \(\sigma = 0\) then \(F_\ell = 0 \mu^\ell\text{-a.e.}\).

(ii) Suppose that \eqref{2.4} and either \eqref{2.17}(i) or \eqref{2.17}(ii) hold true and assume that \(P(x, \cdot)\) is absolutely continuous with respect to \(\mu\) for all \(x \in \text{supp} \mu\). Then \(\sigma^2 = 0\) if and only if \(F = 0 \mu^\ell = \mu \times \cdots \times \mu\text{-almost everywhere (a.e.)}\

(iii) Suppose that the conditions of (ii) hold true except that in place of the absolute continuity requirement there we assume here that \(F\) is continuous on \((\text{supp} \mu)^\ell = \text{supp} \mu \times \cdots \times \text{supp} \mu\). Then \(\sigma^2 = 0\) if and only if \(F = 0\) identically on \((\text{supp} \mu)^\ell\).

#### 2.6. Remark
In principle, we could exclude \(F_\ell = 0 \mu^\ell\text{-a.e.}\) from the beginning since if this happens then \(F\) can be viewed as a function of \(\ell - 1\) variables and we could study then the problem with \(\ell - 1\) in place of \(\ell\). Thus, from the beginning we could assume that \(\ell\) is the maximal number such that \(F_\ell = 0 \mu^\ell\text{-a.e.}\) does not happen.

Observe that when \(\xi_0, \xi_1, \xi_2, \ldots\) are independent identically distributed (i.i.d.) random variables with values in \(\mathcal{X}\) then \(\rho_k = \delta_k = \psi(k) = 0\) for all \(k \geq 1\), and so \eqref{2.4}, \eqref{2.17}(i) and \eqref{2.17}(ii) hold true. The inequality \eqref{2.17}(i) is known to hold true, for instance, when the Markov chain \(\{\xi_n\}\) is reversible and geometrically ergodic (see \[23\]). Some spectral conditions for \eqref{2.17}(i) to hold true can be found in \[8\]. The validity of \eqref{2.17}(ii) can be ensured assuming, for instance, the right hand side of \eqref{2.5} with \(n_0 = 1\). When \(\mathcal{X}\) is a finite state space conditions for \eqref{2.17}(i) and \eqref{2.17}(ii) to be valid can be easily written (see, for instance, \[6\], \[3\] and references there). We will need either of assumptions \eqref{2.17}(i) or \eqref{2.17}(ii) in order to rely on lower bounds for variances of sums of (different) functions of Markov chains from \[22\] or from \[24\], respectively.

As in many expositions of the (conventional) local limit theorem we distinguish between a lattice and a nonlattice cases which take in our circumstances the following form. For any \(x_1, \ldots, x_{\ell-1} \in \mathcal{X}\) set
\begin{equation}
A_{x_1, \ldots, x_{\ell-1}} = \{h \geq 0 : F(x_1, \ldots, x_{\ell-1}, x) \in \{kh : k \in \mathbb{Z} \} \text{ for } \mu - \text{almost all } x \in \mathcal{X}\} \tag{2.18}
\end{equation}
and
\begin{equation}
B_{x_1, \ldots, x_{\ell-1}} = \{h \geq 0 : F(x_1, \ldots, x_{\ell-1}, x) - F(x_1, \ldots, x_{\ell-1}, y) \in \{kh : k \in \mathbb{Z} \} \text{ for } \mu^2 - \text{almost all } (x, y) \in \mathcal{X}^2\} \tag{2.19}
\end{equation}
If \( B_{x_1,\ldots,x_{\ell-1}} \neq \emptyset \) we define
\[
(2.20) \quad h(x_1,\ldots,x_{\ell-1}) = \sup\{h : h \in B_{x_1,\ldots,x_{\ell-1}}\}.
\]
We call the case a lattice one if there exists \( h > 0 \) such that
\[
(2.21) \quad h(x_1,\ldots,x_{\ell-1}) = h \in A_{x_1,\ldots,x_{\ell-1}} \quad \text{for } \mu^{\ell-1} - \text{almost all } (x_1,\ldots,x_{\ell-1}).
\]
If
\[
(2.22) \quad \mu^{\ell-1}\{(x_1,\ldots,x_{\ell-1}) : B_{x_1,\ldots,x_{\ell-1}} = \emptyset\} > 0
\]
then we call the case a non-lattice one. Observe, that there are other cases beyond what we designated as a lattice and a non-lattice case. For instance, \( h(x_1,\ldots,x_{\ell-1}) \) may take on several (or countably many, or continuum) values on subsets of \( X^{\ell-1} \) having positive (or zero in the continuum values case) \( \mu^{\ell-1} \)-measure. Some of these cases can be treated too but in order not to overload the exposition we do not consider them here.

Now we can state our nonconventional local limit theorems which are considered both in a lattice and a non-lattice cases.

2.7. **Theorem.** Suppose that \( \xi_0,\xi_1,\xi_2,\ldots \) is a Markov chain having transition probabilities satisfying (2.23) with \( n_0 = 1 \). Assume (2.2) and (2.2) with \( b \neq 0 \) and that the equality \( F_{\ell} = 0 \) \( \mu^{\ell} \)-a.e. does not hold true. Then \( \sigma^2 \) in (2.12) is positive and for any real continuous function \( g \) on \( \mathbb{R} \) with a compact support
\[
(2.23) \quad \lim_{N \to \infty} \sup_u |\sigma N \sqrt{\mathbb{E}g(S_N - u)} - e^{-\frac{\sigma^2}{2}} \int g \mathcal{L}_h | = 0
\]
where the supremum is taken over \( u \in \{kh : k \in \mathbb{Z}\} \) in the lattice case (2.21) and over all real \( u \) in the non-lattice case. In the latter case we set \( h = 0 \) and \( \mathcal{L}_0 \) is then the Lebesgue measure on \( \mathbb{R} \) while in the lattice case \( \mathcal{L}_h, h > 0 \) is the measure assigning mass \( h \) to each point \( kh, k \in \mathbb{Z} \).

The proof of Theorem 2.7 will rely on estimates of the characteristic function \( \varphi_N(t) \) of \( S_N \) of the form \( e^{-qN} \) for \( \theta \) belonging to a compact set disjoint from 0 and of the form \( e^{-rN\theta^2} \) for \( \theta \) close to 0. These are obtained employing some large deviations results from [7], [1] and [16]. From these estimates Theorem 2.7 will follow essentially in a standard way. When \( X \) is a finite state space Theorem 2.7 requires that the transition matrix of the Markov chain \( \{\xi_n\} \) consists of positive entries only. Actually, in this case we can obtain the nonconventional local limit theorem in a bit more general situation.

2.8. **Theorem.** Let \( \{\xi_n\} \) be a Markov chain with a finite state space \( X \) and a matrix of transition probabilities \( \Pi = (p_{ij})_{i,j \in X} \) such that for some \( k \) the matrix \( \Pi^k = (p_{ij}^{(k)})_{i,j \in X} \) consists of all positive entries and
\[
(2.24) \quad \min_{i,j} \sum_k \min(p_{ik}^{(t)}, p_{jk}^{(t)}) > 0 \quad \text{and} \quad \min_{i,j} \sum_k \min(p_{ki}^{(t)}, p_{kj}^{(t)}) > 0.
\]
Assume (2.23) and that \( F_{\ell} \) does not equal zero identically. Then both in the lattice and the non-lattice cases the local limit theorem in the form (2.23) takes place.

We observe that (2.17)(ii) implies that \( \delta_\ell < 1 \) and the latter is equivalent to the first inequality in (2.24) (see [9]). The first and the second inequalities in (2.24) mean that the matrices \( \Pi^{(t)}(\Pi^{(t)})^* \) and \( (\Pi^{(t)})^*\Pi^{(t)} \), respectively, have all positive
3. Positivity of variance

In this section we explain first why the proof of Theorem 2.2 goes through in the same way as in [17] and after that we derive Theorem 2.5.

We will rely on the following result whose proof goes through in the same way as in Lemma 3.1 from [16]. Let \( h = h(y, z) \), \( y \in \mathcal{X}^k \), \( z \in \mathcal{X}^l \) be a Borel measurable function on \( \mathcal{X}^{k+l} \). Let \( X = (X_1, \ldots, X_k) : \Omega \to \mathcal{X}^k \) and \( Y = (Y_1, \ldots, Y_l) : \Omega \to \mathcal{X}^l \) be sequences of \( \mathcal{X} \)-valued random variables (random "vectors") such that \( X \) is \( \mathcal{F}_n \)-measurable and \( Y \) is \( \mathcal{F}^{n+m} \)-measurable. Suppose that \( E|h(X, Y)| < \infty \) and \( r(x) = E|h(x, Y)| < \infty \) for any \( x \in \mathcal{X}^k \). Then with probability one, \(|E(h(X, Y)|\mathcal{F}_n) - g(X)| \leq \psi(m)r(X)\),

where \( g(x) = Eh(x, Y) \).

Now observe that special Hölder continuity assumptions imposed on \( F \) in [17] were only needed because we did not assume there measurability of \( \xi_n \) with respect to \( \mathcal{F}_n \) but required only some estimates on the error when replacing \( \xi_n \) by \( E(\xi_n|\mathcal{F}_{n+m}) \). To use this we had to estimate what happens when \( \xi_n \) is replaced by \( E(\xi_n|\mathcal{F}_{n+m}) \) inside \( F \) which could be done only relying on Hölder continuity (or better regularity) of \( F \). In the present paper \( \mathcal{F}_n \) is generated by \( \xi_0, \xi_1, \ldots, \xi_n \), i.e. \( \xi_n \) is \( \mathcal{F}_n \)-measurable, and we can apply (3.1) directly to \( F \) without any need of Hölder continuity as exhibited below.

We will show next how to obtain variance and covariance estimates needed to employ the method from [17] to the setup of Theorem 2.2 using directly (3.1) without any need of regularity assumptions on \( F \). Define inductively for \( i = \ell - 1, \ell - 2, \ldots, 0 \),

\[
(3.2) \quad G_i^2(x_1, \ldots, x_i) = EG_{i+1}^2(x_1, \ldots, x_i, \xi_1) = \int G_{i+1}^2(x_1, \ldots, x_i, y)d\mu(y)
\]

where \( G_1^2 = F^2 \) and \( G_0^2 = EG_1^2(\xi_1) \). Then by (3.1) considered with \( X = (\xi_n, \xi_2, \ldots, \xi_{(i-1)n}) \) and \( Y = \xi_{in}, i = \ell, \ell - 1, \ldots, 1 \) we obtain

\[
(3.3) \quad EF(\xi_{in}, \xi_{2n}, \ldots, \xi_{tn}) \leq (1 + \psi(n))EG_{i-1}^2(\xi_n, \ldots, \xi_{(i-1)n})
\]

\[
\leq \cdots \leq (1 + \psi(n))^iG_0^2 = (1 + \psi(n))^ib^2
\]

where \( b^2 \) is given by (2.1). Next, observe that by the Cauchy-Schwarz inequality

\[
(3.4) \quad F_{i}^2(x_1, \ldots, x_i) \leq 2(\int F(x_1, \ldots, x_i)d\mu(x_{i+1})\ldots d\mu(x_{\ell}))^2
\]

\[
+2(\int F(x_1, \ldots, x_i)d\mu(x_{i+1})\ldots d\mu(x_{\ell}))^2 \leq 2 \int F^2(x_1, \ldots, x_i)d\mu(x_{i+1})\ldots d\mu(x_{\ell})
\]

\[
+2 \int F^2(x_1, \ldots, x_i)d\mu(x_{i+1})\ldots d\mu(x_{\ell}) = 2G_i^2(x_1, \ldots, x_i) + 2G_{i-1}^2(x_1, \ldots, x_{i-1}),
\]

and so, as in (3.3),

\[
(3.5) \quad EF_{i}^2(\xi_{i}, \ldots, \xi_{n}) \leq 2(1 + \psi(n))^{i-1}(2 + \psi(n)b^2).
\]

Now we estimate covariances for \( j \geq i \) and \( m > k \) which by (3.1) considered with \( X = \xi_k, \ldots, \xi_m; \xi_{m+1}, \ldots, \xi_{(j-1)m} \) and \( Y = \xi_{jm} \) together with (2.10) yields

\[
(3.6) \quad A_{i,j,k,m} = |EF_i(\xi_k, \ldots, \xi_{ik})F_j(\xi_{m}, \ldots, \xi_{jm})|
\]

\[
\leq \psi(\min(m, jm - ik))ER(\xi_k, \ldots, \xi_{ik}; \xi_m, \ldots, \xi_{(j-1)m})
\]
where

\[ R(x_1, \ldots, x_i; y_1, \ldots, y_{j-1}) = |F_i(x_1, \ldots, x_i)|E|F_j(y_1, \ldots, y_{j-1}, \xi_1)| \leq \frac{1}{2} F_i^2(x_1, \ldots, x_i) + \frac{1}{2} E F_j^2(y_1, \ldots, y_{j-1}, \xi_1). \]

Hence, by (3.10), (3.1) and the Cauchy-Schwarz inequality for all \( i = 1, \ldots, \ell, \)

\[ \text{Where} \]

\[ (3.10) \]

\[ \lim \text{inf} \sum_{m=N_{\ell}}^{N_{\ell}-1} F_i(\xi_n, \ldots, \xi_{\ell_n}) \]

Thus,

\[ (3.11) \]

\[ \sigma_N^2 = \text{var} S_N = \text{var} \left( \sum_{m=1}^{N_{\ell}} S_{i,N} + S_{i,N}^{(1)} \right) + \text{var} S_{i,N}^{(2)} + 2 \sum_{k=1}^{N_{\ell}} \sum_{m=N_{\ell}}^{N_{\ell}-1} B_{i,k,m} + 2 \sum_{k=1}^{N_{\ell}-1} \sum_{m=N_{\ell}}^{N_{\ell}} B_{i,k,m} \]

where for \( i = 1, 2, \ldots, \ell, \)

\[ B_{i,k,m} = E \{ F_i(\xi_k, \ldots, \xi_{\ell}) - E F_i(\xi_k, \ldots, \xi_{\ell}) \} \]

In order to prove Theorem 2.3(ii) we will show first that the multiple sums in (3.11) are bounded after which it will remain only to obtain that

\[ \lim \inf \frac{1}{N} \text{var} S_{i,N}^{(2)} > 0. \]

Clearly,

\[ (3.12) \]

\[ |B_{i,k,m}| \leq A_{i,k,m} + |E F_i(\xi_k, \ldots, \xi_{\ell})||E F_i(\xi_m, \ldots, \xi_{\ell_m})|. \]

Observe that

\[ (3.13) \]

\[ \ell N_{\ell} - (\ell - 1)N \geq N/2, \]

and so when \( m \geq N_{\ell}, i \leq \ell - 1 \) and \( k \leq N \) we can estimate the right hand side of (3.13) by means of (2.3) and (3.1)–(3.3) which yields that two double sums in (3.11) are bounded by constants independent of \( N \).
Next, we study
\begin{equation}
B_{\ell,m,n} = E(F_\ell(\xi_m, \xi_{2m}, \ldots, \xi_{\ell m}) F_\ell(\xi_n, \xi_{2n}, \ldots, \xi_{\ell n}))
- E F_\ell(\xi_m, \xi_{2m}, \ldots, \xi_{\ell m}) E F_\ell(\xi_n, \xi_{2n}, \ldots, \xi_{\ell n})
\end{equation}
for \(m, n \geq N_\ell\) in more details. The product of expectations in (3.15) is small by (3.3) and (3.9), and so we have to deal mainly with the expectation of the product here. Set
\[H_{m,n}^{(\ell)}(x_1, \ldots, x_\ell; y_1, \ldots, y_\ell) = F_\ell(x_1, \ldots, x_\ell) F_\ell(y_1, \ldots, y_\ell)\]
and define recursively for \(j = \ell - 1, \ell - 2, \ldots, 1,\)
\[H_{m,n}^{(j-1)}(x_1, \ldots, x_{j-1}; y_1, \ldots, y_{j-1}) = \int H_j(x_1, \ldots, x_j; y_1, \ldots, y_j) d\mu_{j|n-m}(x_j, y_j)\]
where \(\mu_j\) is the distribution of the pair \((\xi_0, \xi_k)\).

Next, observe that
\begin{equation}
H_{m,n}^{(0)} = E(F_\ell(\xi_m, \xi_{2m}, \ldots, \xi_{\ell m}) F_\ell(\xi_n, \xi_{2n}, \ldots, \xi_{\ell n}))
\end{equation}
and that \(E F_\ell(\xi_m, \xi_{2m}, \ldots, \xi_{\ell m}) = 0\) by (2.2). Applying (3.1) with \(X = (\xi_m, \ldots, \xi_{(j-1)m}; \xi_n, \ldots, \xi_{(j-1)n})\) and \(Y = (\xi_{jm}, \xi_{jn})\) we obtain that for any \(m, n \geq N_\ell,\)
\begin{equation}
|E H_{m,n}^{(j)}(\xi_m, \xi_{2m}, \ldots, \xi_{jm}; \xi_n, \xi_{2n}, \ldots, \xi_{jn})|
- E H_{m,n}^{(j-1)}(\xi_m, \xi_{2m}, \ldots, \xi_{(j-1)m}; \xi_n, \xi_{2n}, \ldots, \xi_{(j-1)n})|
\leq \psi(\left\lceil \frac{N}{2} \right\rceil) EH_{m,n}^{(j-1)}(\xi_m, \xi_{2m}, \ldots, \xi_{(j-1)m}; \xi_n, \xi_{2n}, \ldots, \xi_{(j-1)n})
\end{equation}
where
\[\hat{H}_{m,n}^{(j-1)}(x_1, \ldots, x_{j-1}; y_1, \ldots, y_{j-1}) = E[H_{m,n}^{(j)}(x_1, \ldots, x_{j-1}, \xi_{jm}; y_1, \ldots, y_{j-1}, \xi_{jn})].\]

Since
\begin{equation}
\hat{H}_{m,n}^{(j)}(x_1, \ldots, x_{j}; y_1, \ldots, y_{j}) = |F_\ell(x_1, \ldots, x_\ell)||F_\ell(y_1, \ldots, y_\ell)|
\leq \frac{1}{2} F_\ell^2(x_1, \ldots, x_\ell) + \frac{1}{2} F_\ell^2(y_1, \ldots, y_\ell)
\end{equation}
we obtain by induction that
\begin{equation}
\hat{H}_{m,n}^{(j)}(x_1, \ldots, x_{j}; y_1, \ldots, y_{j}) \leq \frac{1}{2} G_j^2(x_1, \ldots, x_\ell) + \frac{1}{2} G_j^2(y_1, \ldots, y_\ell)
\end{equation}
with \(G_j^2\) defined by (3.2). In the same way as in (3.3) we see that
\begin{equation}
EG_j^2(\xi_m, \ldots, \xi_{jn}) \leq (1 + \psi(n))^{j/2}.
\end{equation}

Now, combining (3.10) and (3.19) we obtain that
\begin{equation}
|E(F_\ell(\xi_m, \xi_{2m}, \ldots, \xi_{\ell m}) F_\ell(\xi_n, \xi_{2n}, \ldots, \xi_{\ell n})) - H_{m,n}^{(0)}| \leq \psi(\left\lceil \frac{N}{2} \right\rceil) \ell (1 + \psi(N_\ell))^j b^2.
\end{equation}

Set
\[U_{\ell,N}^{(2)} = \sum_{n=N_\ell}^N F_\ell(\xi_n^{(1)}, \xi_{2n}^{(2)}, \ldots, \xi_{\ell n}^{(\ell)}).
\]

Then by (3.8), (3.9), (3.15), (3.16) and (3.20),
\begin{equation}
|\text{var} U_{\ell,N}^{(2)} - \text{var} U_{\ell,N}^{(2)}| \leq \frac{N^2}{4 \ell^2} \psi(\left\lceil \frac{N}{2} \right\rceil)^j (2 + \psi(\left\lceil \frac{N}{2} \right\rceil))^j b^2 (\ell + 2 \psi(\left\lceil \frac{N}{2} \right\rceil)).
\end{equation}
Next, observe that since $\Xi_n^{(\ell)} = (\xi_n^{(1)}, \xi_n^{(2)}, \ldots, \xi_n^{(l)})$ is a stationary process then
\begin{equation}
\var U_{\ell,N}^{(2)} = \var U_{\ell,N-N_{\ell}}.
\end{equation}
It is a standard fact that under (2.4) (and even much slower $\psi$-mixing) the limit (2.13) exists (see, for instance, [12] or [3]), and so
\begin{equation}
\lim_{N \to \infty} \var U_{\ell,N}^{(2)} = \frac{1}{2\ell^2} S_{\ell}^2.
\end{equation}
Now, (3.12) follows from (3.21) and (3.23) so that (2.14) holds true in view of (3.11) and (3.13), completing the proof of Theorem 2.3. □

Next we prove Theorem 2.5. In order to obtain (3.12) we set for each $\bar{x} = (x_{N_\ell}, x_{N_\ell+1}, \ldots, x_{(\ell-1)N})$,
\begin{equation}
S_{\ell,N}^{(2)}(\bar{x}) = \sum_{n=N_\ell}^{N} F_\ell(x_n, x_{2n}, \ldots, x_{(\ell-1)n}, \xi_n)
\end{equation}
where in the case $\ell = 1$ we take $S_{\ell,N}^{(2)}(\bar{x}) = S_{\ell,N}^{(2)}$. Put also $\sigma_{\ell,N}^{2}(\bar{x}) = \var S_{\ell,N}^{(2)}(\bar{x})$. Then applying (3.11) with $X = \bar{\xi} = (\xi_N, \xi_{N+1}, \ldots, \xi_{(\ell-1)N})$ and $Y = (\xi_{N_\ell}, \ldots, \xi_{N})$ we obtain taking into account (3.3) and (3.14) that
\begin{equation}
|E \sigma_{\ell,N}^{2}(\bar{\xi}) - \var S_{\ell,N}^{(2)}| \leq 2N^2(2 + \psi(1))^{\ell+1}\psi([N/2])b^2.
\end{equation}

Next, we apply to $S_{\ell,N}^{(2)}(\bar{x})$ either Proposition 13 from [22] or Proposition 4.1 from [24], depending on whether (2.17)(i) or (2.17)(ii) is assumed, in order to conclude that
\begin{equation}
\var S_{\ell,N}^{(2)}(\bar{x}) \geq c_\ell \sum_{n=N_\ell}^{N} \var F_\ell(x_n, \ldots, x_{(\ell-1)n}, \xi_n)
\end{equation}
where $c_\ell = 1 - \rho_2$ if we assume (2.17)(i) and $c_\ell = \frac{1}{1+\rho_1}$ if we assume (2.17)(ii). It is easy to see that $\rho_2 \leq \rho_1$ and $\delta_2 \leq \delta_1$ so that if (2.17)(i) or (2.17)(ii) holds true then the corresponding coefficient $c_\ell$ is positive but for this either $\rho_2 < 1$ or $\delta_2 < 1$ is also sufficient.

Observe that by (2.11) and the stationarity, for any $y_1, \ldots, y_{\ell-1}$,
\begin{align*}
q(y_1, \ldots, y_{\ell-1}) &= \var F_\ell(y_1, \ldots, y_{\ell-1}, \xi_n) \\
&= \var F_\ell(y_1, \ldots, y_{\ell-1}, \xi_1) = EF_\ell^2(y_1, \ldots, y_{\ell-1}, \xi_1).
\end{align*}
Applying $\ell - 1$ times (3.11) and using (3.6) we obtain easily that
\begin{align}
|E q(\xi_N, \xi_{2n}, \ldots, \xi_{(\ell-1)n}) - \int q(x_1, \ldots, x_{\ell-1})d\mu(x_1)\ldots d\mu(x_{\ell-1})| \\
&\leq 2(\ell - 1)\psi(n)(2 + \psi(n))^{\ell+1}b^2.
\end{align}
This together with (2.4), (3.11), (3.13), (3.24) and (3.25) yields that
\begin{align}
\liminf_{N \to \infty} \frac{1}{N} \var S_{\ell,N}^{(2)} &\geq c_\ell \frac{1}{\ell} \int q(x_1, \ldots, x_{\ell-1})d\mu(x_1)\ldots d\mu(x_{\ell-1}).
\end{align}
If $\sigma^2 = 0$ then
\begin{align}
\int q(x_1, \ldots, x_{\ell-1})d\mu(x_1)\ldots d\mu(x_{\ell-1}) &= \int F_\ell^2(x_1, \ldots, x_\ell)d\mu(x_1)\ldots d\mu(x_\ell) = 0,
\end{align}
i.e.
\begin{align}
F_\ell(x_1, \ldots, x_\ell) &= 0 \quad \mu - \text{a.e.}
\end{align}
which proves the assertion (i) of Theorem 2.5

Next, introduce probability measures \( \mu_{i,n} \) on \( \mathcal{X}^i \) such that for any Borel \( \Gamma \subset \mathcal{X}^i \),
\[
\mu_{i,n}(\Gamma) = P\{ (\xi_n, \xi_{2n}, ..., \xi_{in}) \in \Gamma \}.
\]
If \( \Gamma = \Gamma_1 \times \Gamma_2 \times \cdots \times \Gamma_i \) is a product set with Borel \( \Gamma_1, ..., \Gamma_i \subset \mathcal{X} \) then
\[
\mu_{i,n}(\Gamma) = \int_{\Gamma_1} d\mu(x_1) \int_{\Gamma_2} P(n, x_1, dx_2) ... \int_{\Gamma_i} P(n, x_{i-1}, dx_i).
\]
Now assume that we are in the circumstances of Theorem 2.5(ii), i.e. that \( \mu \) is absolutely continuous with respect to \( \mu \).

\( \text{Kolmogorov formula} \)

\( P \) all \( x \) \( \mu \) and \( x \) \( \Gamma \)

which proves the assertion (i) of Theorem 2.5.

\[ S(3.30) \]

(3.30)

(3.30)

Also, in view of Theorem 10.7 from [4] (see §3, Ch. 6 in [2]) and the Markov property we conclude that
\[
\mu_{j+1,n}(\Gamma) = \int_{\chi^j} d\mu_{j,n}(x_1, ..., x_j) P(n, x_j, \Gamma_{x_1, ..., x_j}) = 0
\]

since \( P(n, x, \cdot) \prec \mu \), completing the induction step.

Now, since \( \mu_{j,n} \prec \mu \) then \( F_{\ell}(\xi_n, \xi_{2n}, ..., \xi_{\ell n}) = 0 \) with probability one for all \( n \), and so \( S_{\ell,N} = 0 \) P-a.s. Hence,
\[
S_N = \sum_{i=1}^{\ell-1} S_{i,N} \quad \text{P-a.s.}
\]

Repeating the above argument for \( \ell - 1 \) in place of \( \ell \) we obtain that \( F_{\ell-1}(x_1, ..., x_{\ell-1}) = 0 \) \( \mu^{\ell-1} \)-a.e. and \( F_{\ell-1}(\xi_n, \xi_{2n}, ..., \xi_{(\ell-1)n}) = 0 \) P-a.s. for all \( n \), and so \( S_{\ell-1,N} = 0 \) P-a.s., as well. Continuing in the same way we obtain finally from (2.22) that \( F(x_1, ..., x_{\ell}) = 0 \) \( \mu^{\ell} \)-a.e. proving the assertion (ii).

In the assertion (iii) we assume that \( F \) is continuous on \( (\text{supp}\mu)^{\ell} \) and then the absolute continuity arguments above are not needed since from (3.29) we obtain directly in this case that \( F_{\ell}(x_1, ..., x_{\ell}) = 0 \) for all \( x_1, ..., x_{\ell} \in \text{supp}\mu \). Since
\[
P\{ (\xi_n, ..., \xi_{\ell n}) \notin (\text{supp}\mu)^{\ell} \} \leq \sum_{i=1}^{\ell} P\{ \xi_{in} \notin \text{supp}\mu \} = \ell \mu(\chi \setminus \text{supp}\mu) = 0
\]

we obtain again that \( F_{\ell}(\xi_n, ..., \xi_{\ell n}) = 0 \) P-a.s. and proceeding as above we conclude that \( F(x_1, ..., x_{\ell}) = 0 \) not only \( \mu^{\ell} \)-a.e. but for all \( x_1, ..., x_{\ell} \in \text{supp}\mu \) in view of continuity of \( F \) here, completing the proof.

4. LOCAL LIMIT THEOREM

In this and the next two sections we will prove Theorems 2.6 and 2.8. The arguments of the present section are rather standard and they are valid in a more general framework while the ones in the next sections will be heavily affected by our setups. The proof of (2.23) will consist mainly of the study of the characteristic function of the sum \( S_N \). First, observe that in view of Theorem 10.7 from [4] (see
also Section 10.4 there and Lemma IV.5 together with arguments of Section VI.4 in [11] it suffices to prove (2.23) for all continuous complex-valued functions \( g \) on \( \mathbb{R} = (-\infty, \infty) \) such that

\[
\int_{-\infty}^{\infty} |g(x)| \, dx < \infty
\]

and its Fourier transform

\[
\hat{g}(\lambda) = \int_{-\infty}^{\infty} e^{-i\lambda x} g(x) \, dx, \quad \lambda \in \mathbb{R}
\]

vanishes outside of a finite interval \([-L, L]\). In particular, the inversion formula

\[
g(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\lambda x} \hat{g}(\lambda) \, d\lambda
\]

holds true.

Hence,

\[
Eg(S_N - u) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi_N(\lambda)e^{-i\lambda u} \hat{g}(\lambda) \, d\lambda
\]

where \( \varphi_N(\lambda) = e^{\lambda S_N} \) is the characteristic function of \( S_N \). Changing variables \( s = \lambda \sigma \sqrt{N} \) we obtain

\[
\sigma \sqrt{2\pi N}Eg(S_N - u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \varphi_N(\frac{s}{\sigma \sqrt{N}})e^{-i\frac{s}{\sigma \sqrt{N}} \hat{g}(\frac{s}{\sigma \sqrt{N}})} ds.
\]

On the other hand, from the formula for the characteristic function of the Gaussian distribution and the Fourier inversion formula it follows that

\[
e^{-\frac{u^2}{2\pi \sigma^2}} \int gdL_h = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(-\frac{i\lambda u}{\sigma \sqrt{N}} - \frac{\lambda^2}{2})d\lambda.
\]

Now, in the non-lattice case we write by (4.5) and (4.6),

\[
|\sigma \sqrt{2\pi N}Eg(S_N - u) - e^{-\frac{u^2}{2\pi \sigma^2}} \int gdL_0| \leq I_1(N, T) + I_2(N, T) + I_3(N, \delta) + I_4(N, \delta, T)
\]

where

\[
I_1(N, T) = \frac{1}{\sqrt{2\pi}} \int_{-T}^{T} |\varphi_N(\frac{\lambda}{\sigma \sqrt{N}})\hat{g}(\frac{\lambda}{\sigma \sqrt{N}})| - e^{-\frac{\lambda^2}{2}} \int gdL_0|d\lambda,
\]

\[
I_2(T) = \frac{1}{\sqrt{2\pi}} \int_{|\lambda| > T} e^{-\frac{\lambda^2}{2}} d\lambda,
\]

\[
I_3(N, \delta) = \frac{\|\hat{g}\|}{\sqrt{2\pi}} \int_{|\lambda| > \delta \sigma \sqrt{N}} |\varphi_N(\frac{\lambda}{\sigma \sqrt{N}})| d\lambda,
\]

\[
I_4(N, \delta, T) = \frac{\|\hat{g}\|}{\sqrt{2\pi}} \int_{|\lambda| > \delta \sigma \sqrt{N} > T} |\varphi_N(\frac{\lambda}{\sigma \sqrt{N}})| d\lambda,
\]

\( \|\hat{g}\| = \sup_{\lambda} |\hat{g}| \) and in writing \( I_3(N, \delta) \) we used the fact that \( \hat{g}(s) = 0 \) for \( s \notin [-L, L] \).

In the lattice case (2.21) we proceed in a slightly different way. First, observe that then

\[
\varphi_N(\frac{\lambda}{\sigma \sqrt{N}} + \frac{2\pi k}{h}) = \varphi_N(\frac{\lambda}{\sigma \sqrt{N}}) \quad \text{for all} \quad k \in \mathbb{Z}.
\]
Set
\[ r(v) = \sum_{k=-\infty}^{\infty} \hat{g}(v + \frac{2\pi k}{h}). \]

Taking into account that here \( u \in \{kh : k \in \mathbb{Z}\} \) we can rewrite (4.5) in the following way
\[ (4.9) \quad \sigma \sqrt{2\pi N} E g(S_N - u) = \frac{1}{\sqrt{2\pi}} \int_{-\frac{x\sqrt{2\pi}}{2\sigma N}}^{\frac{x\sqrt{2\pi}}{2\sigma N}} \varphi_N(\frac{\lambda}{\sigma \sqrt{N}}) e^{-i \frac{\lambda u}{\sigma \sqrt{N}}} r(\frac{\lambda}{\sigma \sqrt{N}}) d\lambda. \]

This together with (4.6) yields
\[ (4.10) \quad |\sigma \sqrt{2\pi N} E g(S_N - u) - e^{-\frac{u^2}{2\sigma^2}}| \int gd\mathcal{L}_h| \leq J_1(N, T) + J_2(N, T) + J_3(N, \delta) + J_4(N, \delta, T) \]
where
\[ \begin{align*}
J_1(N, T) &= \frac{1}{\sqrt{2\pi}} \int_{-T}^{T} |\varphi_N(\frac{\lambda}{\sigma \sqrt{N}}) r(\frac{\lambda}{\sigma \sqrt{N}}) - e^{-\frac{\lambda^2}{2}}| \int gd\mathcal{L}_h| d\lambda, \\
J_2(T) &= \frac{1}{\sqrt{2\pi}} \int_{|\lambda| > T} \lambda^{-\frac{\lambda^2}{2}} d\lambda, \\
J_3(N, \delta) &= \frac{1}{\sqrt{2\pi}} \int_{|\lambda| \geq \delta \sigma \sqrt{N}} |\varphi_N(\frac{\lambda}{\sigma \sqrt{N}}) ||r(\frac{\lambda}{\sigma \sqrt{N}})|| d\lambda, \\
J_4(N, \delta, T) &= \frac{1}{\sqrt{2\pi}} \int_{\delta \sigma \sqrt{N} > |\lambda| > T} |\varphi_N(\frac{\lambda}{\sigma \sqrt{N}}) ||r(\frac{\lambda}{\sigma \sqrt{N}})|| d\lambda. 
\end{align*} \]

By (the central limit) Theorem 2.2 for any \( \lambda \),
\[ (4.11) \quad \lim_{N \to \infty} \varphi_N(\frac{\lambda}{\sigma \sqrt{N}}) = e^{-\frac{\lambda^2}{2}}. \]

Furthermore, since \( g \) is continuous and integrable on \( \mathbb{R} \),
\[ (4.12) \quad \lim_{N \to \infty} \hat{g}(\frac{\lambda}{\sigma \sqrt{N}}) = \int \hat{g} \mathcal{L}_0. \]

In the lattice case (2.21) it follows from Ch. 10 in [4] that
\[ (4.13) \quad \lim_{N \to \infty} r(\frac{\lambda}{\sigma \sqrt{N}}) = \int \mathcal{L}_h. \]

Now by (4.11)–(4.13) and the dominated convergence theorem we obtain that for any \( T > 0 \),
\[ (4.14) \quad \lim_{N \to \infty} I_1(N, T) = \lim_{N \to \infty} J_1(N, T) = 0. \]

Next, clearly,
\[ (4.15) \quad \lim_{T \to \infty} I_2(T) = \lim_{T \to \infty} J_2(T) = 0. \]

Thus, it remains to estimate \( J_3(N, \delta) \), \( I_4(N, \delta, T) \), \( J_3(N, \delta) \) and \( J_4(N, \delta, T) \). In order to do this we will need two following results which will be proved in the next sections for the setups of Theorems 2.7 and 2.8.
4.1. **Lemma.** There exists an integer $N_0$ such that for all $N \geq N_0$,
\begin{equation}
|\varphi_N(\theta)| \leq e^{-qN}
\end{equation}
which holds true in the non-lattice case for any $\theta \in [\delta, \delta^{-1}]$ and $\delta > 0$, while in the lattice case (2.21) the inequality (4.16) holds true for any $\theta \in [-\pi/\bar{h}, \pi/\bar{h}] \setminus [-\delta, \delta]$ and $\delta > 0$ with some $q = q_\delta > 0$ depending in both cases only on $\delta$.

4.2. **Lemma.** There exists an integer $N_0$ such that for all $N \geq N_0$,
\begin{equation}
|\varphi_N(\theta)| \leq e^{-rN\theta^2}
\end{equation}
which holds true whenever $|\theta| \leq \delta$ with some $r = r_\delta > 0$ depending only on $\delta$ which is supposed to be small enough.

Before proving Lemmas 4.1 and 4.2 we will rely on them in order to estimate $I_3(N, \delta)$, $I_4(N, \delta, T)$, $J_3(N, \delta)$, $J_4(N, \delta, T)$ and to derive (2.23). Indeed, with $\delta$ small enough we estimate $I_3(N, \delta)$ and $J_3(N, \delta)$ by Lemma 4.1 to obtain
\begin{equation}
I_3(N, \delta) \leq \|\hat{g}\| \sqrt{2\pi} (L - \delta) \sigma \sqrt{N} e^{-qN} \to 0 \quad \text{as} \quad N \to \infty
\end{equation}
and
\begin{equation}
J_3(N, \delta) \leq e^{-q_\delta N} \sigma \sqrt{N} \int_{-\pi/\bar{h}}^{\pi/\bar{h}} |r(s)| ds \to 0 \quad \text{as} \quad N \to \infty.
\end{equation}

Next, we estimate $I_4(N, \delta, T)$ and $J_4(N, \delta, T)$ by Lemma 4.2 to obtain
\begin{equation}
I_4(N, \delta, T) \leq \|\hat{g}\| \int_{|\lambda| > T} e^{-r\lambda^2} d\lambda \to 0 \quad \text{as} \quad T \to \infty
\end{equation}
and
\begin{equation}
J_4(N, \delta, T) \leq R \sqrt{2\pi} \int_{|\lambda| > T} e^{-r\lambda^2} d\lambda \to 0 \quad \text{as} \quad T \to \infty
\end{equation}
where
\[ R = \sup_{|u| \leq \delta} |r(u)| \leq \sup_{|v| \leq L + \delta} |\hat{g}(v)| < \infty. \]

Thus, letting first $N \to \infty$ and then $T \to \infty$ we derive (2.23) from (4.14), (4.15) and (4.18)–(4.21). □

5. **Positive transition densities case**

First, as it is easy to understand, (2.5) implies $\psi$-mixing (see, for instance, [5] and Ch.7.21 in [3]). Next, (2.5) with $n_0 = 1$ yields that
\[ P(x, \Gamma) = \int_{\Gamma} p(x, y) d\eta(y) \quad \text{and} \quad \mu(\Gamma) = \int_{\Gamma} p(y) d\eta(y) \]
with densities $p(x, y)$ and $p(y)$ satisfying $\gamma^{-1} \geq p(x, y) \geq \gamma$ and $\gamma^{-1} \geq p(y) \geq \gamma$ for all $x, y \in \mathcal{X}$. Now,
\[ |P(x, \Gamma) - P(y, \Gamma)| \leq \int_{\Gamma} |p(x, z) - p(y, z)| d\eta(z) \leq 1 - \gamma^2, \]
and so (2.17) (ii) follows. Hence, the conditions of Theorem 2.5 (ii) are satisfied and since $F_\ell = 0 \mu^\ell$-a.e. is excluded we conclude that the limiting variance $\sigma^2$ is positive.
Next, for each sequence \( \bar{x} = (x^{(1)}, \ldots, x^{(\ell-1)}) \in \mathcal{X}^{\ell-1} \) and a real number \( \theta \) we define the (Fourier) operator

\[
\Phi_\theta(y) = E_y \exp(i\theta F(x^{(1)}, \ldots, x^{(\ell-1)}, \xi_\ell))f(\xi_\ell)
\]

\[
= \int P(\ell, y, dz) \exp(i\theta F(x^{(1)}, \ldots, x^{(\ell-1)}, z))f(z)
\]

and the function

\[
\rho_\theta(\bar{x}) = \sup_{y \in \mathcal{X}} \int \eta(dv) \left| \int p^{(\ell)}(y, z) p^{(\ell)}(z, v) \exp(i\theta F(x^{(1)}, \ldots, x^{(\ell-1)}, z)) d\eta(z) \right|
\]

where \( p^{(\ell)} \) is the \( \ell \)-step transition density of the Markov chain \( \{\xi_n\} \). Then for any \( \bar{x}_1 = (x^{(1)}, \ldots, x^{(\ell-1)}) \), \( \bar{x}_2 = (x^{(2)}, \ldots, x^{(\ell-1)}) \in \mathcal{X}^{\ell-1} \),

\[
(\Phi_\theta(\bar{x}_1)\Phi_\theta(\bar{x}_2)\| = \sup_{\|f\| = 1} \sup_y \left| \int \eta(dv) \exp(i\theta F(x^{(1)}, \ldots, x^{(\ell-1)}, v))f(v) \right|
\]

\[
\times \left| \int p^{(\ell)}(y, z) p^{(\ell)}(z, v) \exp(i\theta F(x^{(1)}, \ldots, x^{(\ell-1)}, z)) d\eta(z) \right| \leq \rho_\theta(\bar{x}_1).
\]

Consider

\[
\varphi_\theta(x^{(1)}, \ldots, x^{(\ell-1)})(y) = \exp(i\theta F(x^{(1)}, \ldots, x^{(\ell-1)}, y))
\]

as a function of \( y \). Then either \( \varphi_\theta(x^{(1)}, \ldots, x^{(\ell-1)})(y) \) does not depend on \( y \) \( \eta \)-a.e. or

\[
p^{(2\ell)}(y, v) - |\int p^{(\ell)}(y, z) p^{(\ell)}(z, v) \exp(i\theta F(x^{(1)}, \ldots, x^{(\ell-1)}, z)) d\eta(z)|
\]

\[
\geq \bar{c}_\theta(x^{(1)}, \ldots, x^{(\ell-1)}) > 0
\]

where \( \bar{c}_\theta(x^{(1)}, \ldots, x^{(\ell-1)}) \) depends on \( x^{(1)}, \ldots, x^{(\ell-1)} \) but not on \( y \) and \( v \) taking into account that \( \gamma \leq p^{(\ell)} \leq \gamma^{-1} \). But if (5.2) holds true for all \( y \) and \( v \) then

\[
(5.3) \quad \rho_\theta(\bar{x}) \leq 1 - \bar{c}_\theta(\bar{x}), \quad \bar{x} = (x^{(1)}, \ldots, x^{(\ell-1)}).
\]

Observe that (5.3) holds true for any \( \theta \neq 0 \) in the non-lattice case if \( B_\bar{x} = \emptyset \) (with \( B_\bar{x} \) defined by (2.19)) and for any \( \theta \in [-\frac{\pi}{\alpha}, \frac{\pi}{\alpha}] \) \( \emptyset \) in the lattice case (2.21).

Taking into account in addition that \( \rho_\theta(\bar{x}) \) is continuous in \( \theta \) we conclude that for any \( \delta > 0 \) there exists \( c_\delta(\bar{x}) > 0 \) such that

\[
(5.4) \quad \rho_\theta(\bar{x}) \leq 1 - c_\delta(\bar{x})
\]

whenever \( \delta \leq |\theta| \leq \frac{1}{\alpha} \) and \( B_\bar{x} = \emptyset \) in the non-lattice case and whenever \( \theta \in [-\frac{\pi}{\alpha}, \frac{\pi}{\alpha}] \) \( \emptyset \) in the lattice case (2.21).

On the other hand, since \( F_\ell = 0 \) \( \mu^{\ell-1} \)-a.e. is excluded then \( g_\theta(x^{(1)}, \ldots, x^{(\ell-1)})(y) \) cannot be constant in \( y \) for \( \mu^{\ell-1} \)-almost all \( (x^{(1)}, \ldots, x^{(\ell-1)}) \) whenever \( \delta \leq |\theta| \leq \frac{1}{\alpha} \) in the non-lattice case and whenever \( \theta \in [-\frac{\pi}{\alpha}, \frac{\pi}{\alpha}] \) \( \emptyset \) in the lattice case (2.21).

It follows that for each such \( \theta \) the inequality (5.4) holds true with \( c_\delta(\bar{x}) > 0 \) for \( \bar{x} \in \mathcal{X}^{\ell-1} \) belonging to a set having positive \( \mu^{\ell-1} \)-measure. Hence, for any \( \delta > 0 \) there exist \( c_\delta > 0 \) and a Borel set \( G_1 \subset \mathcal{X}^{\ell-1} \) such that for all \( \theta \) from the corresponding ranges above

\[
(5.5) \quad \rho_\theta(\bar{x}) \leq 1 - c_\delta \quad \text{for all } \bar{x} \in G_1 \text{ and } \mu^{\ell-1}(G_1) = \varepsilon > 0.
\]

Next,

\[
(5.6) \quad |\varphi_N(\theta)| \leq E[E(\exp(i\theta(S_N - S_M))|\mathcal{F}_M)]
\]
where $M = M(N) = N - 2 \lceil \frac{N}{2} \rceil$ and, recall, $N_\ell = \lceil N(1 - \frac{1}{2\ell}) \rceil + 1$. By the Markov property we obtain that with probability one,

$$(5.7) \quad E(\exp(i\theta(S_N - S_M))|\mathcal{F}_{IM}) = \prod_{k=M+1}^{N} \Phi_{\xi_k, \xi_{2k}, \ldots, \xi_{(\ell-1)k}}(\theta)1(\xi_{IM})$$

where $1$ is the function equal $1$ identically and we took into account that

$$(5.8) \quad jM - (j-1)N \geq \lceil \frac{N}{2} \rceil \quad \text{for all } \quad j = 1, 2, \ldots, \ell.$$

Let $\{\xi_n^{(1)}\}, \{\xi_n^{(2)}\}, \ldots, \{\xi_n^{(\ell-1)}\}$ be $\ell - 1$ independent copies of the stationary Markov chain $\{\xi_n\}$ with the initial distribution $\mu$. Applying (3.1) subsequently $\ell - 1$ times with $X = (\xi_k, k \leq jM)$ and $Y = (\xi_{jn}, n = M+1, M+2, \ldots, N)$, $j = \ell - 1, \ell - 2, \ldots, 1$ and using (5.8) we obtain that

$$(5.9) \quad |E| \prod_{k=M+1}^{N} \Phi_{\xi_k, \xi_{2k}, \ldots, \xi_{(\ell-1)k}}(\theta)1| - E| \prod_{k=M+1}^{N} \Phi_{\xi_k^{(1)}, \xi_{2k}^{(2)}, \ldots, \xi_{(\ell-1)k}^{(\ell-1)}}(\theta)1| \leq (\ell - 1)\psi\left(\frac{N}{2}\right) - 2\ell.$$

Next, introduce the Markov chain $\Xi_n = (\xi_n^{(1)}, \xi_n^{(2)}, \ldots, \xi_n^{(\ell-1)})$ on $\mathcal{X}^{\ell-1}$ with the transition probability

$$P_\Xi(\bar{x}, \bar{\Gamma}) = \prod_{j=1}^{\ell-1} P(j, x^{(j)}, \Gamma_j)$$

where $\bar{x} = (x^{(1)}, \ldots, x^{(\ell-1)}) \in \mathcal{X}^{\ell-1}$ and $\bar{\Gamma} = \Gamma_1 \times \cdots \times \Gamma_{\ell-1}$ is a Borel subset of $\mathcal{X}^{\ell-1}$. Observe that $\mu^{\ell-1} = \mu \times \cdots \times \mu$ is the stationary distribution of $\{\Xi_n\}$ and that its transition probability $P_\Xi(\bar{x}, \cdot)$ has a transition density $p_\Xi$ with respect to $\eta^{\ell-1}$ satisfying

$$\gamma^{(\ell-1)} \leq p_\Xi(\bar{x}, \cdot) \leq \gamma^{-(\ell-1)}$$

where $\gamma$ is the same as in (2.25).

Now, by (5.1) and the submultiplicativity of norms of operators,

$$(5.10) \quad E\left| \prod_{k=M+1}^{N} \Phi_{\xi_k^{(1)}, \xi_{2k}^{(2)}, \ldots, \xi_{(\ell-1)k}^{(\ell-1)}}(\theta)1 \right| \leq E \prod_{j=1}^{N - N_\ell} \left| \prod_{k=M+1}^{M+2j} \Phi_{\xi_k^{(1)}, \xi_{2k}^{(2)}, \ldots, \xi_{(\ell-1)k}^{(\ell-1)}}(\theta) \right| \leq E \prod_{j=1}^{N - N_\ell} \rho_\theta(\Xi(M + 2j - 1))$$

It follows from (2.4), (5.1), (5.6), (5.7), (5.9) and (5.10) that in order to obtain (4.16) it suffices to show that for some $N_0$ and all $N \geq N_0$,

$$(5.11) \quad E \prod_{n=1}^{N - N_\ell} \rho_\theta(\Xi(2n - 1)) \leq e^{-\beta N}$$

for some $\beta = \beta_\ell > 0$ depending on $\delta$ which determines corresponding domains for $\theta$ in the non-lattice and the lattice cases. It is easy to see that $\rho_\theta$ is Lipschitz continuous in $\theta$,

$$(5.12) \quad |\rho_\theta(\bar{x}) - \rho_\theta(\tilde{x})| \leq K|\theta - \tilde{\theta}|$$
with \( K \) satisfying
\[
K \leq R(\bar{x}) = \int |F(x^{(1)}, \ldots, x^{(\ell-1)}, y)|d\mu(y)
\leq \left( \int F^2(x^{(1)}, \ldots, x^{(\ell-1)}, y)d\mu(y) \right)^{1/2}
\]
where \( \bar{x} = (x^{(1)}, \ldots, x^{(\ell-1)}) \). In view of (2.1) we can choose \( L > 0 \) such that the set
\[
G_2 = \{ \bar{x} \in \mathcal{X}^{\ell-1} : R(\bar{x}) \leq L \}
\]
satisfies
\[
(5.13) \quad P\{\Xi(1) \in G_2\} \geq 1 - \frac{\varepsilon}{2}
\]
where \( \varepsilon \) is the same as in (5.5). Set \( G = G_1 \cap G_2 \), then by (5.5) and (5.13),
\[
(5.14) \quad P\{\Xi(n) \in G\} = P\{\Xi(1) \in G\} \geq \varepsilon^2.
\]

Introduce the counting function
\[
V_{\theta, L}(N) = \sum_{n=1}^{[N^2(1-N_0)]} \mathbb{1}_G(\Xi(2n - 1)),
\]
where \( \mathbb{1}_G(x) = 1 \) if \( x \in G \) and = 0, otherwise, and the events
\[
\Gamma_{\theta, L}(N) = \{ V_{\theta, L}(N) < \frac{\varepsilon N}{20\ell} \}.
\]
Since \( [N^2(1-N_0)] \geq N^2 - 2 \) it follows from (5.14) and the large deviations results from [1] together with [7] that
\[
(5.15) \quad P(\Gamma_{\theta, L}(N)) \leq \kappa^{-1} e^{-\kappa N}
\]
for some \( \kappa > 0 \) independent of \( N \). Namely, by Lemma 2.5 in [7] the rate function on the second level of large deviations (for occupational measures) of the Markov chain \( \Xi(n), n \geq 0 \) has zero only at the invariant measure. According to [1] (see also [16]) large deviations in the \( \tau \)-topology (generated by convergencies on bounded Borel functions) have the same rate function. Thus we can apply the contraction principle (see [8]) in the \( \tau \)-topology to conclude that for any bounded Borel function \( g \) on \( \mathcal{X}^{2(\ell-1)} \) the rate function on the first level of large deviations for sums of the form
\[
\sum_{k=1}^{n} g(\Xi(k))
\]
has a unique zero at the integral of \( g \) with respect to the invariant measure. Now, taking \( g = \mathbb{1}_{G(\theta, L)} \) we will arrive at (5.15).

Next, let \( \theta_1, \ldots, \theta_k \) be a minimal \( \frac{\varepsilon}{2\ell} \)-net in the interval \( [\delta, \delta^{-1}] \) or in \( [-\frac{\pi}{h}, \frac{\pi}{h}] \setminus [-\delta, \delta] \) depending on whether we consider the non-lattice case or the lattice case, respectively. If \( \Xi(n) \in G \) then \( \rho_\theta(\Xi(n)) \leq 1 - \frac{c_3}{2} \) whenever \( |\theta - \hat{\theta}| \leq \frac{\varepsilon}{2\ell} \). Thus, for all \( \theta \in [\delta, \delta^{-1}] \) or \( \theta \in [-\frac{\pi}{h}, \frac{\pi}{h}] \setminus [-\delta, \delta] \) depending on the case under consideration
\[
\prod_{n=1}^{[N^2(1-N_0)]} \rho_\theta(\Xi(n)) \leq \mathbb{1}_{\Gamma_{\theta, L}(N)} + (1 - \frac{c_3}{2}) \frac{2\ell}{N^2}
\]
and (5.8) follows from (5.15), completing the proof of Lemma 4.1 in the setup of Theorem 2.7. \( \square \)
Next, we derive Lemma 4.2. Employing the Taylor formula we can write for any Borel function $f$ on $X$ with $\|f\| = 1$ and $|\theta|$ small enough (cf. Ch.8 in [4]) that

\begin{equation}
\|\Phi_{x_1^{(1)}, \ldots, x_1^{(\ell-1)}(\theta)} \Phi_{x_2^{(1)}, \ldots, x_2^{(\ell-1)}(\theta)} f\|
= \sup_{x \in X} | \int p^{(\ell)}(x, y) p^{(\ell)}(y, z) \exp(i\theta F(x_1^{(1)}, \ldots, x_1^{(\ell-1)}, y))
\times \exp(i\theta F(x_2^{(1)}, \ldots, x_2^{(\ell-1)}, z)) f(z) d\eta(y) d\eta(z)|
\leq \sup_{x \in X} \int d\eta(z) \int p^{(\ell)}(x, y) p^{(\ell)}(y, z) \exp(i\theta F(x_1^{(1)}, \ldots, x_1^{(\ell-1)}, y)) d\eta(y) d\eta(z)
\leq 1 - \frac{o(1)}{\Theta^2} \inf_{x, z} \int p^{(\ell)}(x, y) p^{(\ell)}(y, z) D(x_1^{(1)}, \ldots, x_1^{(\ell-1)}, x, y, z) d\eta(y) d\eta(z)
\times \exp(\Theta^2 o_{x_1^{(1)}, \ldots, x_1^{(\ell-1)}(1)}(1))
\end{equation}

where $o_{x_1^{(1)}, \ldots, x_1^{(\ell-1)}(1)}(1) \to 0$ as $\theta \to 0$ and

$$D(x_1, \ldots, x_{\ell-1}, x, y, z) = (F(x_1, \ldots, x_{\ell-1}, y)
- \frac{1}{p^{(\ell)}(x,z)} \int p^{(\ell)}(x, y) p^{(\ell)}(y, z) F(x_1, \ldots, x_{\ell-1}, y) d\eta(y))^2.$$ 

Since $\gamma \leq p^{(\ell)} \leq \gamma^{-1}$ it is clear that

$$\inf_{x, z} \int p^{(\ell)}(x, y) p^{(\ell)}(y, z) D(x_1^{(1)}, \ldots, x_1^{(\ell-1)}, x, y, z) d\eta(y) = 0 \mu^{\ell-1} \text{ a.e.}$$

if and only if

$$\inf_{x, z} D(x_1^{(1)}, \ldots, x_1^{(\ell-1)}, x, y, z) = 0 \mu^{\ell-1} \times \eta \text{ a.e.}$$

and the latter holds true if and only if $F_{\ell} = 0 \mu^{\ell}$-a.e. which is excluded by our assumptions. Observe also that by (2.1),

$$\{(x_1^{(1)}, \ldots, x_1^{(\ell-1)}): \sup_{x, z} \int p^{(\ell)}(x, y) p^{(\ell)}(y, z)
\times D(x_1^{(1)}, \ldots, x_1^{(\ell-1)}, x, y, z) d\eta(y) \leq L\} \uparrow Y \text{ as } L \uparrow \infty$$

where $\mu^{\ell-1}(Y) = 1$. It follows that the right hand side of (5.16) is less than $1 - c\Theta^2 \leq e^{-c\Theta^2}$ for some $c > 0$, all small enough $\Theta^2$ and all collections $x_1^{(1)}, \ldots, x_1^{(\ell)}$ from a Borel set $G \subset X^{\ell-1}$ such that $\mu^{\ell-1}(G) = \epsilon > 0$.

Hence, for small $|\theta|$, (5.17)

$$\rho_\theta(\bar{x}_1) \leq e^{-c\Theta^2} \text{ whenever } \bar{x}_1 \in G.$$

Introduce, again, the counting function

$$W(N) = \sum_{n=1}^{[\frac{N}{8\epsilon}]} I_G(\Xi(2n - 1))$$

and the events

$$\Gamma(N) = \{W(N) < \epsilon(\frac{N}{8\epsilon} - 1)\}.$$

Then, since $\rho_\theta(\bar{x}) \leq 1$, we obtain by (5.17) that

\begin{equation}
\prod_{n=1}^{[\frac{N}{8\epsilon}]} \rho_\theta(\Xi(2n - 1)) \leq \exp(-c\epsilon\Theta^2(\frac{N}{8\epsilon} - 2)) + I_{\Gamma(n)}.
\end{equation}
Relying on the large deviations results from \cite{7} and \cite{1} as explained above we conclude that

\begin{equation}
P(\Gamma(N)) \leq \kappa^{-1} e^{-\kappa N}
\end{equation}

for some $\kappa > 0$ independent of $N$. Finally, (4.17) follows from (2.4), (5.1), (5.6), (5.7), (5.9), (5.10), (5.18) and (5.19), completing the proof of Lemma 4.2 in the setup of Theorem 2.7. \hfill \Box

5.1. Remark. Observe that a weaker estimate in Lemma 4.1 of the form $|\varphi_N(\theta)| \leq e^{-q\sqrt{N}}$ is also quite sufficient for our purposes in (4.18) and (4.19). This estimate can be easily obtained without any use of large deviations. Indeed, we can estimate the left hand side of (5.8) by

\begin{equation}
E \prod_{n=1}^{\lfloor N/2\rfloor} \rho_\theta(\Xi(n[\sqrt{N}])).
\end{equation}

Now the gaps of order $\sqrt{N}$ enable us to employ (5.1) which yields that up to an error of order $\sqrt{N}\psi(\lfloor \sqrt{N} \rfloor)$ the expression (5.20) can be estimated by

\begin{equation}
\prod_{n=1}^{\lfloor N/2\rfloor} E\rho_\theta(\Xi(n[\sqrt{N}])) \leq (1 - c_3) \frac{N-N}{\sqrt{N}}.
\end{equation}

Observe also that under certain conditions the estimate of Lemma 4.2 can be derived from \cite{10} without any use of large deviations results.

5.2. Remark. Several of more recent works on (conventional) local limit theorems rely on some assumptions (like non-arithmeticity condition in \cite{11}) which ensure that contraction conditions needed in Lemmas 4.1 and 4.2 hold true. We could also impose certain contraction conditions on compositions (products) of several Fourier operators without specifying Markov chains we are dealing with but in order to obtain from these Lemmas 4.1 and 4.2 we would have to assume that the combined Markov chain $\Psi(n) = (\Xi(n), \Xi(n+1), \ldots, \Xi(n+m))$ satisfy some large deviations bounds. Such set of assumptions would look somewhat unwieldy but, still, we could derive a local limit theorem from it and then to check the assumptions for more specific Markov chains considered in Theorems 2.7 and 2.8.

6. Finite state space case

First, observe that our assumption that $\Pi^k$ for some $k$ has all positive entries yields $\psi$-mixing of $\{\xi_n\}$ with (2.4) satisfied (see, for instance, \cite{3}, Ch.7). Clearly, $\Pi^k$ has all positive entries for all $n \geq k$ and we take $m$ so that $(m-2)\ell \geq k$. Furthermore, there exists a unique stationary distribution $\mu$ which gives a positive mass to each state $j$ in $\mathcal{X}$. Since $F_\ell = 0$ $\mu'$-a.e. is excluded then Theorem 2.5(ii) guarantees positivity of the limiting variance $\sigma^2$ in (2.12) provided $\delta_\ell < 1$ which is equivalent to the first inequality in (2.24) (see \cite{6}).

Now, for each $\bar{x} = (x^{(1)}, \ldots, x^{(\ell-1)}) \in \mathcal{X}^{\ell-1}$ and a function $f$ on $\mathcal{X}$ we can write

$$
\Phi_{\bar{x}}(\theta)f(a) = \sum_{b \in \mathcal{X}} p^{(\ell)}_{ab} e^{i\theta F(x^{(1)}, \ldots, x^{(\ell-1)}, b)} f(b).
$$
where \( \bar{x} \) corresponds to the transition matrix, then clearly, \( \Pi \) is a probability matrix. 

Next, for \( \bar{G} \) a Borel set in \( \mathbb{R}^d \), 

\[
\rho(\bar{G}) = \max_{a} \sum_{x \in \bar{G}} P_{ab}(\bar{x}) \sum_{c} \sum_{x} P_{bc}(\bar{x}) \delta(x) = \max_{a} \sum_{x \in \bar{G}} P_{ab}(\bar{x}) \sum_{c} \sum_{x} P_{bc}(\bar{x}) \delta(x)
\]

where for each \( \bar{x} = (x^{(1)},...,x^{(\ell - 1)}) \in X^{\ell - 1} \) we set 

\[
\rho(\bar{x}) = \max_{a} \sum_{x \in \bar{G}} P_{ab}(\bar{x}) \sum_{c} \sum_{x} P_{bc}(\bar{x}) \delta(x)
\]

Since \( P_{ab}^{(m-2)\ell} > 0 \) for any \( a, b \in X \) then \( \rho(\bar{x}) = 1 \) if and only if for any \( b \in X \), 

\[
0 \leq \sum_{x \in \bar{G}} P_{bc}(\bar{x}) \sum_{x} P_{cd}(\bar{x}) = 1.
\]

This will hold true if and only if \( g_{\bar{x}}^{(\theta)}(c) = e^{\theta F(x^{(1)},...,x^{(\ell - 1)},c)} \delta(c) \) does not depend on \( c \) on the set \( \{ c \in X : P^{(\ell)}_{bc}, P^{(\ell)}_{cd} > 0 \} \). But, in view of the assumption \( (2.21) \) for any \( c, c' \in X \) there exist \( b, d \in X \) such that \( P^{(\ell)}_{bc}, P^{(\ell)}_{cd}, P^{(\ell)}_{d} > 0 \). Hence, if \( (6.2) \) holds true then \( g_{\bar{x}}^{(\theta)}(c) \) does not depend on \( c \) on the whole \( X \). Since we assume that \( F_{\ell} \) is not zero identically then \( g_{\bar{x}}^{(\theta)}(c) \) cannot be constant in \( y \) for \( \mu^{\ell - 1} \)-almost all \( (x^{(1)},...,x^{(\ell - 1)}) \) whenever \( \delta \leq |\theta| \leq \frac{1}{2} \) in the non-lattice case and whenever \( \theta \in [-\frac{\pi}{2}, \frac{\pi}{2}] \) in the lattice case \( (2.21) \). It follows that \( \rho(\bar{x}) < 1 \) for each such \( \theta \) and for \( \bar{x} \in X^{\ell - 1} \) belonging to a set having positive \( \mu^{\ell - 1} \)-measure. Since \( \rho(\theta) \) is continuous in \( \theta \) this holds true uniformly in \( \theta \) in the corresponding compact domains above. Hence, for any \( \delta > 0 \) there exist \( \epsilon > 0 \) and a Borel set \( G \subset X^{\ell - 1} \) such that for all \( \theta \) from the corresponding ranges above 

\[
|\rho(\bar{x})| \leq 1 - \epsilon \quad \text{for all} \quad \bar{x} \in G \quad \text{and} \quad \mu^{\ell - 1}(G) = \epsilon > 0.
\]

Next, take \( M = M(N) = N - m[\frac{N - N_{k}}{2}] \) and proceed as in \( (5.6) \). Introduce again the Markov chain \( \Xi_n = (x^{(1)}_n, x^{(2)}_n, ..., x^{(\ell - 1)}_n) \) on \( X^{\ell - 1} \) with the transition probabilities

\[
p_{\Xi}(\bar{a}, \bar{b}) = \prod_{j=1}^{\ell - 1} p_{\xi_j \xi_j}(\bar{a}_{(j)} \bar{b}_{(j)})
\]

where \( \bar{a} = (a^{(1)}, ..., a^{(\ell - 1)}) \), \( \bar{b} = (b^{(1)}, ..., b^{(\ell - 1)}) \in X^{(\ell - 1)} \). If \( \Xi = (p_{\Xi}(\bar{a}, \bar{b})) \) is the corresponding transition matrix then, clearly, \( \Xi_{\Xi}^{(m-2)\ell} \) has all positive entries. Now similarly to \( (5.10) \),

\[
|\prod_{k=1}^{N} \prod_{j=M+1}^{M+2j} \Phi_{\xi^{(1)}_k \xi^{(2)}_k \cdots \xi^{(\ell - 1)}_k}(\theta) | \\
\leq |\prod_{j=1}^{N - N_{k}} \prod_{k=M+M(j-1)+1}^{M+2j} \Phi_{\xi^{(1)}_k \xi^{(2)}_k \cdots \xi^{(\ell - 1)}_k}(\theta) | \\
\leq E \prod_{j=1}^{N - N_{k}} \rho(\Xi(M + mj - 1))
\]
Thus, in order to obtain (1.10) it suffices to show that for some \( N_0 \) and all \( N \geq N_0 \),
\[
E \prod_{n=1}^{\lfloor (N-N_1) \rfloor} \rho_\theta(\Xi(mn-1)) \leq e^{-\beta N}
\]
(6.5)
for some \( \beta = \beta_0 > 0 \) depending on \( \delta \) which determines corresponding domains for \( \theta \) in the non-lattice and the lattice cases. It is easy to see that \( \rho_\theta \) is Lipschitz continuous in \( \theta \) with a constant \( K \) satisfying
\[
K \leq R(\bar{x}) = \sum_y |F(x^{(1)}, \ldots, x^{(\ell-1)}, y)| \leq L
\]
for some \( L < \infty \), where \( \bar{x} = (x^{(1)}, \ldots, x^{(\ell-1)}) \), since \( F \) is a function on the finite state space \( \mathcal{X} \).

Introduce, again, the counting function
\[
W_\theta(N) = \sum_{n=1}^{\lfloor (N-N_0) \rfloor} \mathbb{I}_G(\Xi(mn-1)).
\]
where \( \mathbb{I}_G(x) = 1 \) if \( x \in G \) and = 0, otherwise, and the events
\[
\Gamma_\theta(N) = \{ V_\theta(N) < \frac{\varepsilon N}{20\ell} \}.
\]
Since \( \lfloor (N-N_0) \rfloor \geq \frac{N}{\ell} - 2 \) it follows from (6.3) and the large deviations results from Section 3.1 in [8] together with [7] that
\[
P(\Gamma_{\theta,L}(N)) \leq \kappa^{-1} e^{-\kappa N}
\]
for some \( \kappa > 0 \) independent of \( N \). Namely, by Lemma 2.5 in [7] the rate function on the second level of large deviations (for occupational measures) of the Markov chain \( \Xi(n), n \geq 0 \) has zero only at the integral of \( g \) with respect to the invariant measure. Now, taking \( g = \mathbb{I}_{G(\theta)} \) we will arrive at (6.6). In fact, in our circumstances we can obtain (6.6) directly from [20] and [19]. Now we conclude the proof of (6.5) and of the whole Lemma 4.1 (in the setup of Theorem 2.8) in the same way as in the previous section.

Next, we derive Lemma 4.2 in the present setup. Employing the Taylor formula we can write for \( |\theta| \) small enough (cf. Ch.8 in [4]) that
\[
\rho_\theta(\bar{x}) \leq 1 - \frac{\theta^2}{2} \inf_X \sum_{b \in X} P_{ab}((m-2)\ell) \sum_{d \in X} \sum_{c \in X_{bd}} p_{bc}^{(\ell)}
\times D(x^{(1)}, \ldots, x^{(\ell-1)}, b, c, d) p_{cd}^{(\ell)} + \theta^2 a_{x^{(1)}, \ldots, x^{(\ell-1)}}(1)
\]
where \( \bar{x} = (x^{(1)}, \ldots, x^{(\ell-1)}) \), \( X_{bd} = \{ c : p_{bc}^{(\ell)}, p_{cd}^{(\ell)} > 0 \}, a_{x^{(1)}, \ldots, x^{(\ell-1)}}(1) \to 0 \) as \( \theta \to 0 \) and
\[
D(x^{(1)}, \ldots, x^{(\ell-1)}, b, c, d)
= (F(x^{(1)}, \ldots, x^{(\ell-1)}, c) - \frac{1}{p_{bc}^{(\ell)} p_{cd}^{(\ell)}} \sum_{e \in X_{bd}} p_{bc}^{(\ell)} p_{cd}^{(\ell)} F(x^{(1)}, \ldots, x^{(\ell-1)}, e))^2
\]
while we set $\sum_{c \in X_{bd}} = 0$ whenever $X_{bd} = \emptyset$. Observe that
\[
(6.8) \quad \sum_{c \in X_{bd}} p_{bc}^{(\ell)} D(x^{(1)}, \ldots, x^{(\ell-1)}, b, c, d) p_{cd}^{(\ell)} = 0
\]
if and only if $F(x^{(1)}, \ldots, x^{(\ell-1)}, c)$ does not depend on $c \in X_{bd}$. But in view of (2.24) for any $c, c' \in X$ there exist $b, d \in X$ such that $c, c' \in X_{bd}$. It follows that (6.8) holds true for any $b, d \in X$ if and only if $F(x^{(1)}, \ldots, x^{(\ell-1)}, c)$ does not depend on $c$ on the whole $X$, i.e. $F_{\ell}(x^{(1)}, \ldots, x^{(\ell-1)}, c) = 0$ for all $c \in X$. Since the latter equality cannot hold true identically by our assumptions we conclude that there exists a nonempty $G \subset X^{\ell-1}$ such that for some $r > 0$ and all $\bar{x} \in G$,
\[
(6.9) \quad \rho_{\theta}(\bar{x}) \leq 1 - r\theta^2 \leq e^{-r\theta^2}
\]
provided $|\theta|$ is small enough. In present circumstances $\mu^{\ell-1}(G) > 0$ as for any nonempty subset of $X^{\ell-1}$ and we complete the proof of Lemma 4.2 (in the present setup) proceeding in the same way as in the proof of Lemma 4.1 above and at the end of the previous section by introducing a counting function of arrivals of $\Xi(mn - 1)$ to $G$, using (6.4) and (6.9), and relying on the same large deviations argument.  

\[\square\]

References

[1] E. Bolthausen, \textit{Markov process large deviations in $\tau$-topology}, Stoch. Proc. Appl. 25(1987), 95–108.
[2] N. Bourbaki, \textit{Integration I}, Springer-Verlag, Berlin, 2004.
[3] R.C. Bradley, \textit{Introduction to Strong Mixing Conditions}, Kendrick Press, Heber City, UT, 2007.
[4] L. Breiman, \textit{Probability}, SIAM, Philadelphia, 1992.
[5] J.R. Blum, D.L. Hanson, L.H. Koopmans, \textit{On the strong law of large numbers for a class of stochastic processes}, Z. Wahrsch. verw. Geb. 2(1963), 1–11.
[6] R.L. Dobrushin, \textit{Central limit theorem for nonstationary Markov chains I, II}, Theory Probab. Appl. 1(1956), 65–80, 329–383.
[7] M.D. Donsker and S.R.S. Varadhan, \textit{Asymptotic evaluation of certain Markov process expectations for large time, I}, Comm. Pure. Appl. Math. 28(1975), 1–47.
[8] A. Dembo and O. Zeitouni, \textit{Large Deviations Techniques and Applications}, Springer-Verlag, Berlin, 2010.
[9] H. Furstenberg, \textit{Nonconventional ergodic averages}, Proc. Symp. Pure Math. 50, 43–56 (1990).
[10] N. Giesbrecht, \textit{Bounds for sums of random variables over a Markov chain}, Stoch. Proc. Appl. 53(1994), 269–283.
[11] H. Hennion and L. Hervé, \textit{Limit Theorems for Markov Chains and Stochastic Properties of Dynamical Systems by Quasi-Compactness}, Lect. Notes Math. 1766, Springer-Verlag, Berlin, 2001.
[12] I.A. Ibragimov and Yu.V. Linnik, \textit{Independent and Stationary Sequences of Random Variables}, Wolters-Noordhoff, Groningen (1971).
[13] Yu. Kifer, \textit{Optimal stopping and strong approximation theorems}, Stochastics 79(2007), 253–273.
[14] Yu. Kifer, \textit{Nonconventional limit theorems}, Probab. Th. Rel. Fields 148 (2010), 71–106.
[15] Yu. Kifer, \textit{A nonconventional strong law of large numbers and fractal dimensions of some multiple recurrence sets}, Stoch. Dynam. 12(2012), 1150023.
[16] I. Kontoyiannis and S.P. Meyn, \textit{Large deviations asymptotics and the spectral theory of multiplicative regular Markov processes}, Electr. J. Probab. 10(2005), 61–123.
[17] Yu. Kifer and S.R.S. Varadhan, Nonconventional limit theorems in discrete and continuous time via martingales, Ann. Probab. 42(2014), 649–688.

[18] J.G. Kemeny, J.L. Snell, A.W. Knapp, Denumerable Markov Chains, Springer-Verlag, New York, 1976.

[19] A. Kontorovich and R. Weiss, Uniform Chernoff and Dvoretzky-Kiefer-Wolfowitz-type inequalities for Markov chains and related problems, arXiv: 1207.4678 (2013).

[20] P. Lezaud, Chernoff-type bound for finite Markov chains, Ann. Appl. Probab. 8(1998), 849–867.

[21] S.V. Nagaev, Some limit theorems for stationary Markov chains, Theory Probab. Appl. 2(1957), 378–406.

[22] M. Peligrad, Central limit theorem for triangular arrays of non-homogeneous Markov chains, Probab. Th. Rel. Fields 154(2012), 409–428.

[23] G.O. Roberts and J.S. Rosenthal, Geometric ergodicity and hybrid Markov chains, Electron. Commun. Probab. 2(1997), 13–25.

[24] S. Sethuraman and S.R.S. Varadhan, A martingale proof of Dobrushin’s theorem for non-homogeneous Markov chains, Electron. J. Probab. 10(2005), 1221–1235.