<table>
<thead>
<tr>
<th>Title</th>
<th>Limit theorems in hidden Markov models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Han, G</td>
</tr>
<tr>
<td>Citation</td>
<td>IEEE Transactions on Information Theory, 2013, v. 59 n. 3, p. 1311-1328</td>
</tr>
<tr>
<td>Issued Date</td>
<td>2013</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10722/189143">http://hdl.handle.net/10722/189143</a></td>
</tr>
<tr>
<td>Rights</td>
<td>This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.; ©2013 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.</td>
</tr>
</tbody>
</table>
Limit Theorems in Hidden Markov Models
Guangyue Han

Abstract—In this paper, under mild assumptions, we derive a law of large numbers, a central limit theorem with an error estimate, an almost sure invariance principle, and a variant of the Chernoff bound in finite-state hidden Markov models. These limit theorems are of interest in certain areas of information theory and statistics. Particularly, we apply the limit theorems to derive the rate of convergence of the maximum likelihood estimator in finite-state hidden Markov models.

Index Terms—Entropy, hidden Markov models, limit theorem, Shannon-McMillan-Breiman theorem.

I. MAIN RESULTS AND RELATED WORK

Consider a discrete memoryless channel with a finite input alphabet \( \mathcal{Y} \) and a finite output alphabet \( \mathcal{Z} \). Assume that, at each time slot, the channel is characterized by the channel transition probability matrix \( \Pi = \{p(y|z)\} \). Let \( Y = \{Y_i : i \in \mathcal{Z}\} \) be the input process over \( \mathcal{Y} \), which is a stationary Markov chain with transition probability matrix \( \Delta \).

Let \( Z \) denote the output process over \( \mathcal{Z} \), which is often referred to as a hidden Markov chain (or hidden Markov processes) in information theory (for an excellent survey on hidden Markov processes, see [16]); alternatively, \( Z \) can be viewed as a probabilistic function of \( Y \), which is characterized by \( \Pi \). For some \( r > 0 \), assume that \( \{\Pi, \Delta\} \) is analytically parameterized by \( \theta \in \Omega = \mathbb{R}^m \), where \( \theta_0 \) is a true parameter, which is often assumed unknown in a statistical context, and \( \mathbb{R}^m \) denotes the closed \( r \)-ball around \( \theta_0 \) in \( \mathbb{R}^m \).

For any \( l \in \mathbb{N} \cup \{0\} \), we are interested in the limiting probabilistic behavior of the \( l \)th derivative of \( \log p^\theta(Z_1^n) \) with respect to any \( \theta \in \Omega \), denoted by \( D^l_\theta \log p^\theta(Z_1^n) \); here, \( Z_1^n \) is used to denote the sequence of random variables \( (Z_1, Z_2, \ldots, Z_n) \), and similar notational convention will be followed in the sequel. We will prove limit theorems for appropriately normalized versions of \( D^l_\theta \log p^\theta(Z_1^n) \), for any fixed \( l \) and any \( \theta \in \Omega \). Here, we remark that, only for notational convenience, we are treating \( \theta \) as a 1-D variable throughout this paper.

For given \( l \in \mathbb{N} \cup \{0\} \), consider the following two conditions.

(I) For any \( \theta \in \Omega \), \( \Pi^\theta \) is a strictly positive matrix and \( \Delta^\theta \) is irreducible and aperiodic;

(II) For any \( \theta \in \Omega \),

\[
\sigma_{n}^{(l)}(\theta) = \lim_{n \to \infty} \sqrt{\frac{\sigma_{n}^{(l)}(\theta)^2}{n}} > 0,
\]

where \( \sigma_{n}^{(l)}(\theta) = \text{Var}_{\Pi^\theta, \Delta^\theta} \left(D^l_\theta \log p^\theta(Z_1^n)\right) \) (the existence of this limit under Condition (I) will be established later).

And we define

\[
I^{(l)}(\theta) \triangleq \lim_{n \to \infty} E_{\theta_0} \left[D^l_\theta \log p^\theta(Z_1^n)/n \right]
\]

when the limit exists.

The following theorem is an analog of the law of large numbers (LLN).

Theorem 1.1: Fix \( l \in \mathbb{N} \cup \{0\} \) and assume Condition (I). Then, for any \( \theta \in \Omega \), \( I^{(l)}(\theta) \) is well defined, and

\[
\frac{D^l_\theta \log p^\theta(Z_1^n)}{n} \to I^{(l)}(\theta) \text{ with probability } 1.
\]

For the case \( l = 0 \), Theorem 1.1 has already been observed in [2], where the convergence is used to prove the consistency of the maximum likelihood estimator (MLE) in a hidden Markov model; the cases \( l = 0 \), 1 have been established in [32], where the consistency of the MLE for general hidden Markov models is considered. Note that when \( \theta = \theta_0 \), we have

\[
I^{(l)}(\theta_0) = -H^{\theta_0}(Z_1^n),
\]

where \( H^{\theta_0}(Z) \) denotes the entropy rate of the hidden Markov chain \( Z \) at the true parameter \( \theta_0 \). So, when \( l = 0 \), Theorem 1.1 is a (rather) special case of the celebrated Shannon–McMillan–Breiman theorem, which only assumes the stationarity and ergodicity of \( Z \). Entropy rate of a hidden Markov chain is of great importance in many areas of mathematics and physics; in particular, the computation of \( H^{\theta_0}(Z) \) is a first step to compute the capacity of a finite-state channel in information theory. Unfortunately, it is notoriously difficult to compute such a fundamental quantity (see [16], [31] and references therein). Recently, based on the Shannon–McMillan–Breiman theorem, efficient Monte Carlo methods for approximating \( H^{\theta_0}(Z) \) were proposed independently by Arnold and Loeliger [1], Pfister et al. [35], and Sharma and Singh [43].

We will prove the following central limit theorem (CLT) for \( D^l_\theta \log p^\theta(Z_1^n) \) with an error estimate, which is often referred to as Berry–Esseen bound [3], [17] in probability theory. Here, we remark that, in this paper, to avoid notational cumbersoness, while ensuring its dependence on various variables, we often use \( C \) to denote a constant, which may not be the same on each appearance.
Theorem 1.2: Fix \( i \in \mathbb{N} \cup \{0\} \) and assume Conditions (I) and (II). For any \( \varepsilon > 0 \), there exists \( C > 0 \) such that for any \( n \in \mathbb{N} \)

\[
\sup_{x} \left| \mathbb{P} \left( \frac{D_{h}^{i} \log p^{\theta}(Z_{n}^{i}) - n L^{(1)}(\theta)}{\sqrt{n \sigma^{(1)}(\theta)^{2}}} < x \right) - G(x) \right| \leq C n^{-1/4+\varepsilon}
\]

where \( G(x) = \int_{-\infty}^{x} (2\pi)^{-1/2} \exp(-y^{2}/2)dy \).

For the case \( i = 1 \), Theorem 1.2 (without the Berry–Esseen bound) has first been shown in [2], which, together with Theorem 1.1 for the case \( i = 2 \), can be further used to derive the asymptotic normality of the MLE for a hidden Markov model. This asymptotic normality result is of great importance to the statistical estimation aspects in hidden Markov models and has been generalized extensively in [5], [6], [13], [14], [29], [32], [40], and [41].

Theorem 1.2 for the case \( i = 0 \) and \( \theta = \theta_{c} \) (again without the Berry–Esseen bound) has been considered in more probabilistic settings as well: a CLT for \( \log p^{\theta}(Z_{n}^{i}) \) assuming \( Z \) is a Markov chain is first proven in [46]; this result is further generalized to obtain a refinement of the Shannon–McMillan–Breiman theorem in [27] under some mixing assumptions; under somewhat similar conditions, an almost sure invariance principle, a deep result which, among many other applications, implies a CLT, has been shown in [37]; the almost sure invariance principle is used to study the asymptotic behavior of the so-called recurrent and waiting times in [28], where a CLT for \( \log p^{\theta}(Z_{n}^{i}) \) is embedded in the main results.

In a more information theoretical context, a CLT [36] for \( \log p^{\theta}(Z_{n}^{i}) \) is derived as a corollary of a CLT for the top Lyapunov exponent of a product of random matrices; a functional CLT is also established in [26]. In essence, both of these two CLTs are proved using effective Martingale approximations of \( \log p^{\theta}(Z_{n}^{i}) \) (see [19] for this standard technique).

There is also a large body of work (see [25], [24] and references therein) on variants of the CLT for the empirical entropy of some ergodic mappings in the language of ergodic theory, among which of great relevance to this work are [25] and [24], where CLTs with Berry–Esseen bounds are derived. Here, we remark that there are minor mistakes in the proof of the main results in [25]; it appears that a modified proof, together with stronger assumptions, can only yield weaker results than claimed in [25].

Note that the error estimate in the CLTs is of great significance in many scenarios, such as characterizing the speed of convergence of the aforementioned Monte Carlo simulation in [1], [35], and [43] and deriving nonasymptotic coding theorems information theory [45], and so on. Among all the previously mentioned related work, only [25] and [24] give error estimates for the CLTs. Compared to these two works, where only some mixing conditions are assumed for \( Z \), our assumptions are rather strong. On the other hand, our CLT is considerably stronger in the sense that it is essential for a class of functions including \( \log p^{\theta}(Z_{n}^{i}) \) and its derivatives with tighter error estimate.

Following Phillip and Stout [37], we prove the following almost sure invariance principle.

\[ S(t) = \sum_{n \leq t} D_{h}^{i} \log p^{\theta}(Z_{n}^{i}) - n L^{(1)}(\theta). \]

Then, for any given \( \theta \in \Omega \), without changing the distribution of \( \{S(t), t \geq 0\} \), we can redefine the process \( \{S(t), t \geq 0\} \) on a richer probability space together with the standard Brownian motion \( \{B(t), t \geq 0\} \) such that for any \( \varepsilon > 0 \)

\[ S(t) - B((\sigma^{(1)}(\theta))^{2}t) = O(t^{1/3+\varepsilon}) \text{ a.s. as } t \to \infty. \]

As elaborated in [37], an almost sure invariance principle is fundamental for many applications, which includes, besides a CLT and some large deviation results, a law of iterated logarithm (LIL). The following LIL immediately follows from Theorem 1.3.

Theorem 1.4: Fix \( i \in \mathbb{N} \cup \{0\} \) and assume Conditions (I) and (II). Then, for any \( \theta \in \Omega \), we have

\[
\lim \sup_{n \to \infty} \frac{D_{h}^{i} \log p^{\theta}(Z_{n}^{i}) - n L^{(1)}(\theta)}{2n(\sigma^{(1)}(\theta)^{2})^{1/2} \log n (\sigma^{(1)}(\theta)^{2})^{1/2}} = 1 \text{ a.s.}
\]

Theorem 1.4 is somewhat “known”: the almost sure invariance principle in [37], which is established under much weaker conditions, implies Theorem 1.4 for the case \( i = 0 \). In [33], it has been shown that with reasonable assumptions, a CLT with a sharp enough error estimation term implies an LIL for i.i.d. sequences of random variables. For possibly dependent sequences of random variables, Petrov’s result may not be directly applied to derive an LIL; however, the spirit of the proof can be cautiously followed to establish Theorem 1.4 as an alternative approach (see [38]). Using this idea, an LIL (again for the case \( i = 0 \)) has also been noted in [25] and [24] under some mixing assumptions. The cases \( i = 0, 1, 2 \) have been derived [32] via a martingale approach (which, in principle, should be able to establish Theorem 1.4 for any generic \( i \)) and further used to derive an almost sure convergence rate of the MLE for a class of general-state hidden Markov models.

We also prove the following variant of the Chernoff bound (see [11]), giving a subexponentially decaying upper bound for the tail probability of \( S_{n} \). Let us note that, in contrast to the previous ones, the upper bound in the following theorem is uniform over all \( \theta \in \Omega \), and this uniformity will be used in the proof of Theorem 1.6. On the other hand, all the previous theorems can be strengthened to their uniform versions, which, however, are not needed in this paper.

Theorem 1.5: Fix \( i \in \mathbb{N} \cup \{0\} \) and assume Condition (I). For any \( x > 0 \) and any \( 0 < \varepsilon \leq 1 \), there exist \( C > 0, 0 < \gamma < 1 \) such that for any \( n \in \mathbb{N} \) and any \( \theta \in \Omega \)

\[
P \left( \frac{D_{h}^{i} \log p^{\theta}(Z_{n}^{i}) - n L^{(1)}(\theta)}{n} \geq x \right) \leq C \gamma^{n^{1-\varepsilon}}.
\]

Let \( \theta_{c} \in \Omega^{c} \), the interior of \( \Omega \), be an \( n \)-th-order MLE for the considered hidden Markov model, that is,

\[
\theta_{c} \in \arg \min_{\theta \in \Omega^{c}} \log p^{\theta}(Z_{n}^{i}).
\]
which is well defined if \( \arg \max_{\theta \in \Theta} \log p^\theta(Z^n) \) is nonempty. The consistency of the MLE in hidden Markov models have been extensively discussed in statistical contexts and various sufficient conditions guaranteeing the convergence of the MLE have been given (see representative work in [2], [29], [6], and [15]). As one of the principal applications of the limit theorems above, assuming the consistency of the MLE, the following theorem further gives the rate of convergence of the estimators to the true parameter \( \theta_0 \). Here, we remark that Theorem 1.2 with \( l = 1 \) and Theorem 1.5 with \( l = 2 \) are used in the proof.

**Theorem 1.6:** Assume Condition (I) and assume the consistency of the MLE (namely, \( \theta_n \to \theta_0 \) with probability 1). Let \( \tau \) be an integer-valued random time such that \( \arg \max_{\theta \in \Theta} \log p^\theta(Z^n_\tau) \) is nonempty for all \( n \geq \tau \). If \( L^{(2)}(\theta) \) is negative definite for all \( \theta \in \Theta \), then for any \( x, \epsilon > 0 \), there exists \( C > 0 \) such that for all \( n \geq \tau \)

\[
P(\|\theta_n - \theta_0\| \geq x) \leq Cn^{-1/4 + \epsilon}.
\]

II. LIMIT THEOREMS UNDER EXPONENTIAL MIXING AND FORGETTING CONDITIONS

A stationary stochastic process \( T = T^\infty_{\infty} \) is said to be \( \psi \)-mixing if

\[
\psi(n) \overset{\Delta}{=} \sup_{U \in \mathcal{B}(T^n_{\infty})} |P(VU) - P(V)P(U)| \to 0 \quad \text{as} \quad n \to \infty
\]

where \( \mathcal{B}(T^n_{\infty}) \) denotes the \( \sigma \)-field generated by \( \{T_k : k = i, i + 1, \ldots, j\} \). Let \( Z = (Z_n)_{n \in \mathbb{Z}} \) be a stationary \( \psi \)-mixing sequence of random variables over a finite alphabet \( \mathcal{Z} \) satisfying the following property.

(a) [exponential mixing] There exist \( C > 0, 0 < \lambda < 1 \) such that

\[
\psi(n) \leq C\lambda^n
\]

for all \( n \).

Let \( Z^* \) be the set of all finite words over \( Z \), and let \( f : Z^* \to \mathbb{R} \) be a function satisfying the following properties.

(b) There exist \( C', C'' > 0 \) such that for all \( z_{-n}^0 \in Z^* \)

\[
C' \leq f(z_0z_{-1}^{-1}) \leq C''.
\]

(c) [exponential forgetting] There exist \( C > 0, 0 < \rho < 1 \) such that for any two hidden Markov sequences \( z_{-m}^0, \tilde{z}_{-m}^0 \) with \( z_{-n}^0 = \tilde{z}_{-n}^0 \) (here \( m, \tilde{m} \geq n \geq 0 \), we have

\[
|f(z_0z_{-m}^0) - f(\tilde{z}_0\tilde{z}_{-m}^0)| \leq C\rho^n.
\]

Define

\[
X_i = f(Z_i|Z_i^{t-1}) - E[f(Z_i|Z_i^{t-1})]
\]

and

\[
S_n = \sum_{i=1}^{n} X_i, \quad \sigma_n^2 = Var(S_n).
\]

We will also consider

\[
d \overset{\triangle}{=} \lim_{n \to \infty} \frac{\sqrt{\sigma_n^2/n}}{n} > 0 \quad \text{(the existence of this limit under Conditions (a)–(c) will be established in Lemma 3.3 and Remark 3.4)}.
\]

We will prove the following theorems under Conditions (a)–(d). Not only can these theorems be used to prove the main results in Section I, but also they are of interest in their own right. The first theorem is an LLL.

**Theorem 2.1:** Assume Conditions (b) and (c). With probability 1

\[
X_1 + X_2 + \cdots + X_n \overset{n}{\to} 0 \quad \text{as} \quad n \to \infty.
\]

We will also prove the following CLT with a Berry–Esseen bound.

**Theorem 2.2:** Assume Conditions (a)–(d). For any \( \epsilon > 0 \), there exists \( C > 0 \) such that for any \( n \in \mathbb{N} \)

\[
\sup_x P(S_n/\sigma_n < x) - G(x) \leq Cn^{-1/4 + \epsilon},
\]

where

\[
G(x) = \int_{-\infty}^{x} (2\pi)^{-1/2} \exp(-y^2/2)dy.
\]

The following theorem is an almost sure invariance principle.

**Theorem 2.3:** Assume Conditions (a)–(d). Define a continuous parameter process \( \{S(t), t \geq 0\} \) by setting

\[
S(t) = \sum_{n \leq t} X_n.
\]

Then, for any given \( \theta \in \Theta \), without changing the distribution of \( \{S(t), t \geq 0\} \), we can redefine the process \( \{S(t), t \geq 0\} \) on a richer probability space together with the standard Brownian motion \( \{B(t), t \geq 0\} \) such that for any \( \epsilon > 0 \)

\[
S(t) - B(\sigma^2 t) = O(t^{1/2 + \epsilon}) \quad \text{a.s.} \quad \text{as} \quad t \to \infty.
\]

As one of many applications of Theorem 2.3, the following LIL immediately follows.

**Theorem 2.4:** Assume Conditions (a)–(d). Then, we have

\[
\lim_{n \to \infty} \sup_{n \in \mathbb{N}} \frac{S_n}{(2n\sigma^2 \log \log n \sigma^2)^{1/2}} = 1 \quad \text{a.s.}
\]

We also prove the following variant of the Chernoff bound (see [11]), giving a subexponentially decaying upper bound for the tail probability of \( S_n \).

**Theorem 2.5:** Assume Conditions (a)–(c). For any \( x > 0 \) and any \( 0 < \epsilon < 1 \), there exist \( C > 0, 0 < \gamma < 1 \) such that for any \( n \in \mathbb{N} \)

\[
P(S_n/n \geq x) \leq C\gamma^{n^\epsilon}.
\]
III. PROOFS OF THE THEOREMS IN SECTION II

A. Key Lemmas

From now on, we rewrite \( f(z_i|z_{i-1}^i) - E[f(Z_i|Z_{i-1}^i)] \) as \( g(z_i^i) \) for notational simplicity.

The following lemma shows that for a fixed \( j > 0 \), \( E[X_iX_{i+j}] \) converges exponentially as \( i \to \infty \), and for any \( i < j \), \( E[X_iX_j] \) decays exponentially in \( j - i \).

**Lemma 3.1:** Assume Conditions (a)–(c).
1) There exist \( C > 0, 0 < \rho < 1 \) (here \( \rho \) is as in Condition (c)) such that for all \( i, j \geq 0 \)

\[
E[X_{i+1}X_{i+1+j}] - E[X_iX_{i+j}] \leq C\rho^i.
\]

2) There exist \( C > 0, 0 < \epsilon < 1 \) such that for any positive \( \epsilon > 0 \)

\[
|E[X_iX_{i+j}]| \leq C\epsilon^{-j}.
\]

**Proof:**
1) Simple computations lead to

\[
E[X_{i+1}X_{i+1+j}] - E[X_iX_{i+j}] = \sum_{z_{i+1}^i} p(z_{i+1}^i|z_i^i)g(z_{i+1}^i) - \sum_{z_i^i} p(z_i^i)g(z_i^i) = \sum_{z_{i+1}^i} p(z_{i+1}^i|z_i^i)g(z_{i+1}^i) - \sum_{z_i^i} p(z_i^i)g(z_i^i)
\]

By Part 1 of Lemma 3.1 and Remark 3.2, for any \( i \),

\[
E[(S_{n+m} - S_m)^2] = (a_0 + 2 \sum_{j=1}^\infty a_j) \leq Cn^{-\epsilon_0}.
\]

Notice that the constants in \( O(\lambda^{-i}), O(\rho^{i-1}) \) above do not depend on \( z_i^i \). Part 2 then immediately follows.

**Remark 3.2:** By Part 1 of Lemma 3.1, for any fixed \( j \), the sequence \( E[X_iX_{i+j}], i = 1, 2, \ldots \), is a Cauchy sequence that converges exponentially. For any fixed \( j \), let \( a_j = \lim_{i \to \infty} E[X_iX_{i+j}] \). Then by Part 2, \( a_j \) decays exponentially as \( j \to \infty \); consequently, we deduce (for later use) that \( a_0 + 2 \sum_{j=1}^\infty a_j \) converges.

**Lemma 3.3:** Assume Conditions (a)–(c). For any \( l > 0 \),

\[
E[X_iX_{i+j}] = \sum_{z_i^i} p(z_i^i)g(z_i^i) = \sum_{z_i^i} p(z_i^i)g(z_i^i)|p(z_i^i)g(z_i^i)| + O(\rho^{-i}).
\]

**Proof:** Letting \( \beta = n^{-\epsilon_0} \) for a fixed \( 0 < \epsilon_0 < 1 \), we then have

\[
E[(S_{n+m} - S_m)^2] = \frac{E[(\sum_{i=m+1}^n X_i)^2]}{n} \leq \frac{E[X_iX_{i+j}] - \sum_{m+1 \leq i \leq n} E[X_iX_{i+j}]}{n}.
\]

By Part 1 of Lemma 3.1 and Remark 3.2, for any \( j > 0 \),

\[
E[X_iX_{i+j}] = (n-j)a_j + O(1).
\]

Here, the constant in \( O(1) \) does not depend on \( j \). Also, by Part 2 of Lemma 3.1 and Remark 3.2, there exists \( 0 < \beta < 1 \) such that
for all \( j > Bn \), \( E X_i X_{i+j} = O(e^{3n}) \) and, thus, \( a_j = O(e^{3n}) \). Continuing the computation, we have the first equation at the bottom of the page. The lemma then immediately follows if we let \( n \) go to infinity.

Remark 3.4: Choosing \( m \) in Lemma 3.3 to be 0, we deduce that \( \lim_{n \to \infty} \sigma_n^2/n \) exists and is equal to \( \sigma^2 = a_0 + 2 \sum_{j=1}^{\infty} a_j \).

Lemma 3.5: For any \( i \in \mathbb{N} \), there exists \( C > 0 \) such that for all \( m \) and \( n \)

\[
E[(S_{n+m} - S_m)^{2i}] \leq Cn^i.
\]

Proof: By Condition (c) and the stationarity of \( Z \), we observe that for any \( m, n \)

\[
E[(S_{n+m} - S_m)^{2i}] = E[\left( \sum_{i=m+1}^{n+m} g(Z_i^{(1)}) \right)^{2i}]
\]

\[
= E[\left( \sum_{i=m+1}^{n+m} (g(Z_i^{(1)}) + O(\rho^{i-m-1})) \right)^{2i}]
\]

\[
= E[S_n^{2i}] + O(E[S_n^{2i-1}]) + \cdots + O(1).
\]

Notice that for any \( j \)

\[
E[S_n^{2-1}] \leq E[S_n^{2-1}]^{1/3} E[S_n^{2i-3}]^{1/2}.
\]

So, in order to prove the lemma, it suffices to prove that for any \( l \in \mathbb{N} \), there exists \( C_l > 0 \) such that

\[
E[S_n^{2i}] - E[(X_1 + X_2 + \cdots + X_n)^2] \leq C_l n^l.
\]

Now, for any \( l \in \mathbb{N} \), consider the term \( X_{i_1} X_{i_2} X_{i_3} \cdots X_{i_k} \), where \( 1 \leq i_1 < i_2 < \cdots < i_k \leq n \) and \( i_j \)'s are all strictly positive satisfying \( i_1 + i_2 + \cdots + i_k \leq 2l \). Let \( n = n(i_1, i_2, \ldots, i_k) \) be the smallest index such that for all \( j = 1, 2, \ldots, k-1 \),

\[
i_{j+1} - i_j \geq i_{j+1} - i_j.
\]

Now, for any \( v + 1 \leq u \leq k \), recalling that

\[
X_{i_v} = f(Z_{i_v}, |Z_{i_v}| - 1) - f(Z_{i_v}, |Z_{i_v}| - 1),
\]

we define

\[
\tilde{X}_{i_v} = f(Z_{i_v}, |Z_{i_v}| - 1) - f(Z_{i_v}, |Z_{i_v}| - 1),
\]

Applying Condition (c), we have for some \( 0 < \rho < 1 \)

\[
X_{i_v} - \tilde{X}_{i_v} = \rho^{(i_v+1-i_v)/2}.
\]

We then have the decomposition shown at the bottom of the page, where \( r^{(1)} X_{i_1} X_{i_2} \cdots X_{i_k} \), \( r^{(2)} X_{i_1} X_{i_2} \cdots X_{i_k} \), and some intermediate terms produced during the decomposition and \( r^{(1)} X_{i_1} X_{i_2} \cdots X_{i_k} \) is the residual term resulted from the decomposition. Using (3) and Conditions (a)–(c), we can verify that for some \( 0 < \epsilon < 1 \)

\[
\sum_{i_1 > 0} \sum_{i_2 \leq 2} \sum_{i_3 \leq 2} \cdots \sum_{i_k \leq 2} O(e^{(2l-2)j_{i_1}}) = O(n)\quad\text{for some } j_{i_1} \geq 2, j_{i_2} + j_{i_3} + \cdots + j_{i_k} \leq 2l \text{ and } r_{i_1}, r_{i_2}, \ldots, r_{i_k}, \text{ are the residual terms resulted from the recursive decomposition. Then, similarly as in deriving (4), one checks that } E[S_n^{2i}] \text{ can be written as a sum of at most } 2^{2i} \text{ terms, each of which is upper bounded by }
\]

\[
\left( \sum_{i_1, i_2, \ldots, i_k} E[X_{i_1}^{(1)}] E[X_{i_2}^{(2)}] \cdots E[X_{i_k}^{(k)}] \right) O(n) \cdots O(n)\quad\text{for some } j_{i_1} \geq 2, j_{i_2} + j_{i_3} + \cdots + j_{i_k} \leq 2l \text{ and } r_{i_1}, r_{i_2}, \ldots, r_{i_k}, \text{ are the residual terms resulted from the recursive decomposition. Then, similarly as in deriving (4), one checks that } E[S_n^{2i}] \text{ can be written as a sum of at most } 2^{2i} \text{ terms, each of which is upper bounded by }
\]

\[
\sum_{i_1, i_2, \ldots, i_k} E[X_{i_1}^{(1)}] E[X_{i_2}^{(2)}] \cdots E[X_{i_k}^{(k)}] O(n) \cdots O(n)\quad\text{for some } j_{i_1} \geq 2, j_{i_2} + j_{i_3} + \cdots + j_{i_k} \leq 2l \text{ and } r_{i_1}, r_{i_2}, \ldots, r_{i_k}, \text{ are the residual terms resulted from the recursive decomposition. Then, similarly as in deriving (4), one checks that } E[S_n^{2i}] \text{ can be written as a sum of at most } 2^{2i} \text{ terms, each of which is upper bounded by }
\]

\[
\sum_{i_1, i_2, \ldots, i_k} E[X_{i_1}^{(1)}] E[X_{i_2}^{(2)}] \cdots E[X_{i_k}^{(k)}] O(n) \cdots O(n)\quad\text{for some } j_{i_1} \geq 2, j_{i_2} + j_{i_3} + \cdots + j_{i_k} \leq 2l \text{ and } r_{i_1}, r_{i_2}, \ldots, r_{i_k}, \text{ are the residual terms resulted from the recursive decomposition. Then, similarly as in deriving (4), one checks that } E[S_n^{2i}] \text{ can be written as a sum of at most } 2^{2i} \text{ terms, each of which is upper bounded by }
\]

\[
\sum_{i_1, i_2, \ldots, i_k} E[X_{i_1}^{(1)}] E[X_{i_2}^{(2)}] \cdots E[X_{i_k}^{(k)}] O(n) \cdots O(n)\quad\text{for some } j_{i_1} \geq 2, j_{i_2} + j_{i_3} + \cdots + j_{i_k} \leq 2l \text{ and } r_{i_1}, r_{i_2}, \ldots, r_{i_k}, \text{ are the residual terms resulted from the recursive decomposition. Then, similarly as in deriving (4), one checks that } E[S_n^{2i}] \text{ can be written as a sum of at most } 2^{2i} \text{ terms, each of which is upper bounded by }
\]

\[
\sum_{i_1, i_2, \ldots, i_k} E[X_{i_1}^{(1)}] E[X_{i_2}^{(2)}] \cdots E[X_{i_k}^{(k)}] O(n) \cdots O(n)\quad\text{for some } j_{i_1} \geq 2, j_{i_2} + j_{i_3} + \cdots + j_{i_k} \leq 2l \text{ and } r_{i_1}, r_{i_2}, \ldots, r_{i_k}, \text{ are the residual terms resulted from the recursive decomposition. Then, similarly as in deriving (4), one checks that } E[S_n^{2i}] \text{ can be written as a sum of at most } 2^{2i} \text{ terms, each of which is upper bounded by }
\]
where
\[ l_j' \geq 2, \quad l_1' + l_2' + \cdots + l_{k_2}' + 2k_2 \leq 2l_i \] (5)
and the summation is over all possible \( X_{i}^{l_1'} X_{i}^{l_2'} \cdots X_{i}^{l_{k_2}'} \)
satisfying (5), which can be estimated by
\[
\sum E[|X_{i}| |X_{i}^{l_1'}| |X_{i}^{l_2'}| \cdots |X_{i}^{l_{k_2}'}|] = O(n^{l-k_2}).
\]
It then follows that
\[
E[|S_{n+1}^{2i}|] = O(n^{l-k_2})O(n^{k_2}) = O(n^{l}).
\]
We then have established the lemma. \( \square \)

**Lemma 3.6:** For any \( i \in \mathbb{N} \), there exists \( C > 0 \) such that for all \( m, n \)
\[
E[|S_{n+m} - S_{m}|^{2i-1}] \leq Cn^{l-1/2}.
\]

**Proof:** The lemma immediately follows from Lemma 3.5 and the fact that for any \( m, n \)
\[
E[|S_{n+m} - S_{m}|^{2i-1}] < E[|S_{n+m} - S_{m}|^{2i}]^{1/2} E[|S_{n+m} - S_{m}|^{2i+1}]^{1/2}.
\]

**B. Proof of Theorem 2.1**

It follows from Condition (c) that there exists \( 0 < \rho < 1 \) such that for any \( j < i \)
\[
f(Z_i |Z_j^{l-1}) = f(Z_i |Z_j^{l-1}) = O(\rho^{-j})
\]
which implies that \( f(Z, Z_t^{l-1}) \rightarrow \lim_{j \rightarrow -\infty} f(Z_i |Z_j^{l-1}) \) exists, and
\[
|f(Z_i |Z_j^{l-1}) - f(Z_i |Z_k^{l-1})| = O(\rho^{-j})
\]
and furthermore
\[
E[f(Z_i |Z_j^{l-1})] - E[f(Z_i |Z_k^{l-1})] = O(\rho^{-j}).
\]
We then have
\[
\frac{\sum_{i=1}^{n} X_i}{n} = \frac{n}{n} f(Z_i |Z_i^{l-1}) - E[f(Z_i |Z_i^{l-1})] - \sum_{i=1}^{n} f(Z_i |Z_i^{l-1}) - E[f(Z_i |Z_i^{l-1})] + O(\rho^{-j}).
\]

Here, we remark that the constants in all the above \( O \)-terms are independent of \( i, j \). Note that the sequence \( f(Z_i |Z_i^{l-1}) - E[f(Z_i |Z_i^{l-1})] \) is stationary and ergodic. Applying the Birkhoff ergodic theorem, and using the fact that \( \sum_{i=1}^{n} \rho^{i} \rightarrow 0 \) as \( n \rightarrow \infty \), we then establish the theorem.

**C. Proof of Theorem 2.2**

For any fixed \( 0 < \beta < \alpha < 1 \), we consecutively partition the partial sum \( S_n \) into blocks \( \eta_1, \zeta_1, \eta_2, \zeta_2, \ldots \) such that each \( \eta_i \) is of length \( p = p(n) \leq n^{\alpha} \) and each \( \zeta_i \) is of length \( q = q(n) \leq n^{\beta} \). In other words, for any feasible \( i \),
\[
\eta_i = X_{(i-1)p+(i-1)+1} + \cdots + X_{iq+(i-1)p},
\]
and
\[
\zeta_i = X_{iq+(i-1)p+1} + \cdots + X_{iq+i p}.
\]
Then, \( S_n \) can be rewritten as a sum of \( \eta \)-“blocks” and \( \zeta \)-“blocks”
\[
S_n = S_n^{0} + S_n^{1} := \sum_{i=1}^{k} \eta_i + \sum_{i=1}^{k} \zeta_i
\]
where \( k = k(n) \leq n/(n^{\alpha} + n^{\beta}) \). The above so-called Bernstein blocking method [4] is a standard technique for proving limit theorems for a variety of mixing sequences. Roughly speaking, the partial sum \( S_n \) is partitioned into “short blocks” \( \eta_1, \eta_2, \ldots, \eta_k \) and “long blocks” \( \zeta_1, \zeta_2, \ldots, \zeta_k \). Under certain mixing conditions, all long blocks are “weakly dependent” on each other, while all short blocks are “negligible” in some sense.

Now, we will “truncate” \( \zeta_i \)’s to obtain \( \hat{\zeta}_i \)’s. In more detail, recall that for any \( j \) with \( iq + [i-1]p + 1 \leq j \leq iq + ip \), we have
\[
X_j = f(Z_i Z_j^{l-1}) - E[f(Z_i Z_j^{l-1})],
\]
We then define
\[
\hat{X}_j = f(Z_i Z_j^{l-1}) - f(Z_{iq+(i-1)p+1} + \cdots + X_{iq+i p} - E[f(Z_j^{l-1})].
\]
Applying Condition (c), we derive that
\[
\hat{X}_j - \hat{X}_j = O(\rho^{i(n)}/2).
\]
We then define
\[
\hat{S}_n = \sum_{i=1}^{k} \hat{\zeta}_i, \quad \hat{\sigma}_n = \sqrt{\text{Var}(\hat{S}_n)}.
\]
With lemmas in Section III-A established, the remainder of the proof of Theorem 2.2 becomes more or less standard, which can be roughly outlined as follows.

1. We first show \( E[\exp(i\hat{S}_n^{l}/\sigma_n)] \) and \( \prod_{j=1}^{k} E[\exp(i\hat{\zeta}_j/\sigma_n)] \) are “close” (see Lemma 3.9).
2. Then, by the standard Esseen’s Lemma, we show \( P(\hat{S}_n/\sigma_n < x) \) and \( G(x) \) are “close” (see Lemma 3.10).
3. Finally, since \( S_n^{0} \) are “negligible”, we conclude, in the proof of Theorem 2.2, that \( P(S_n/\sigma_n < x) \) and \( P(\hat{S}_n/\sigma_n < x) \) are “close,” and thus \( P(S_n/\sigma_n < x) \) and \( G(x) \) are “close”.
Before proceeding, we first remind the reader the classical Esseen’s inequality (see, e.g., [34, Lemma 5.1, p. 147]).

**Lemma 3.7 (Esseen’s Inequality):** Let $\zeta_1, \zeta_2, \ldots, \zeta_n$ be independent random variables with $E[\zeta_j] = 0$, $E[|\zeta_j|^3] < \infty$, $j = 1, 2, \ldots, n$. Let

$$\sigma_n^2 = \sum_{j=1}^n E[\zeta_j^2], \quad I_n = \sigma_n^{-3} \sum_{j=1}^n E[|\zeta_j|^3]$$

and let $\tilde{F}_n(x)$, $\tilde{F}_n(t)$ be the distribution, characteristic functions of the random variable $\sum_{j=1}^n \zeta_j/\sigma_n$, respectively. Then

$$|\tilde{F}_n(t) - e^{-t^2/2}| \leq 16L_n t^3 e^{-t^2/8} \quad (7)$$

for $t \leq 1/(4L_n)$.

The following lemma is a version of Esseen’s lemma, which gives an upper bound on the difference between two distribution functions using the difference between the two corresponding characteristic functions. We refer to [44, p. 314] for a standard proof.

**Lemma 3.8 (Esseen’s Lemma):** Let $F(x)$ and $G(x)$ be distribution functions with characteristic functions $\phi_F(t)$ and $\phi_G(t)$, respectively. Suppose that the distributions corresponding to $F(x)$ and $G(x)$ each have mean 0, and $G(x)$ is differentiable and for any $x$, $G'(x) \leq M$ for some $M > 0$. Then

$$\sup_x |F(x) - G(x)| \leq \frac{1}{\pi} \int_{-T}^T \left| \frac{\phi_F(t) - \phi_G(t)}{t} \right| dt + \frac{24M}{\pi T} \quad (8)$$

for every $T > 0$.

We will need the following lemma.

**Lemma 3.9:** There exist $\alpha$, $\beta$ such that for all $n$ and $|t| \leq n^{1/2}$

$$|E[\exp(it\tilde{S}_n/\sigma_n^2)] - \prod_{j=1}^k E[\exp(it\tilde{\zeta}_j/\sigma_n^2)]| \leq C\rho_1^{\alpha n}.$$  

**Proof:** Let $t = (k-1)p + (k-1)q + |q/2| + 1$. By Condition (a), there exists $0 < \lambda < 1$ such that the equation at the bottom of the page holds, where, again, $f(z_j|z_j^{-1}) = E[f(Z_j|Z_j^{-1})]$ is rewritten as $g(z_j)$. Noticing that $|E[\exp(it\tilde{\zeta}_j/\sigma_n^2)]| \leq 1$ and applying an inductive argument, we conclude that

$$E[\exp(it\tilde{S}_n/\sigma_n^2)] = E[\exp(it\sum_{j=1}^k \tilde{\zeta}_j/\sigma_n^2)]$$

$$= \prod_{j=1}^k E[\exp(it\tilde{\zeta}_j/\sigma_n^2)] + O(k\lambda^{q_1(n)^2})$$

which immediately implies the lemma.

Now, applying Lemma 3.8, we can derive the following lemma.

**Lemma 3.10:** There exists $C > 0$ such that for all $n$

$$\sup_x |P(\tilde{S}_n/\sigma_n < x) - G(x)| \leq Cn^{-1/2 + \alpha n/2}.$$  

**Proof:** Note that all $\tilde{\zeta}_j$’s have the same distribution. So, Lemma 3.9 in fact implies that

$$|E[\exp(it\tilde{S}_n/\sigma_n^2)] - (E[\exp(it\tilde{\zeta}_1/\sigma_n^2)])^k| = O(\rho_1^{\alpha n}) \quad (9)$$

for some $0 < \rho_1 < 1$. Consider a sequence of i.i.d. random variables $\tilde{\zeta}_j, j = 1, 2, \ldots, k$, each of which is distributed according to $\tilde{\zeta}_1$. It then follows from (8) that

$$|E[\exp(it\tilde{S}_n/\sigma_n^2)] - (E[\exp(it\tilde{\zeta}_1/\sigma_n^2)])^k| = O(\rho_1^{\beta n}). \quad (10)$$

Now, let

$$\sigma_n^2 = \sum_{j=1}^k E[\tilde{\zeta}_j^2].$$

It follows from Condition (a) that for some $0 < \lambda < 1$

$$(\sigma_n^2)^2 - \sigma_n^2 = O(k^2E[\tilde{\zeta}_1]^2\lambda^{q_1(n)})$$

which implies that

$$(E[\exp(it\tilde{\zeta}_1/\sigma_n^2)])^k - (E[\exp(it\tilde{\zeta}_1/\sigma_n^2)]) = O(\rho_1^{\alpha n}) \quad (11)$$

for some $0 < \rho_2 < 1$. Therefore, combining (9) and (11), we deduce that

$$(E[\exp(it\tilde{S}_n/\sigma_n^2)] - (E[\exp(it\tilde{\zeta}_1/\sigma_n^2)])^k = O(\rho_1^{\alpha n}) \quad (12)$$
for some $0 < \rho_3 < 1$. So, in the sense of (12), we can approximate $\hat{S}_j^i / \hat{\sigma}_n^i$, using the sum of i.i.d random variables $\zeta_{j,i} / \bar{\sigma}_n$, $j = 1, 2, \ldots, k$, each of which is distributed according to $\zeta_{1,i} / \bar{\sigma}_n$. Applying Lemma 3.7 to the i.i.d. sequence $\zeta_{j,i} / \bar{\sigma}_n$, we deduce that for $|t| \leq 1/(4L_n)$

$$
|E \exp(it\zeta_{1,i} / \bar{\sigma}_n)|^k e^{-|t|^2/2} \leq 16L_n t^3 e^{-|t|^2/2}
$$

(13)

where

$$
L_n = \sum_{j=1}^k E[\zeta_{j,i}^3 / \bar{\sigma}_n^3] = kE[\zeta_{1,i}^3] / \bar{\sigma}_n^3.
$$

Note that, by (10) and Lemma 3.3, we have

$$
\bar{\sigma}_n^3 = \Theta(n^{3/2}).
$$

Furthermore, by (6) and Lemma 3.6, we have

$$
kE[\zeta_{1,i}^3] = kO(p(n))^{3/2} = O(n^{1+\alpha/2}).
$$

It then follows that there exists $C_1 > 0$ such that for all $n$

$$
L_n \leq C_1 n^{-1/2+\alpha/2}.
$$

(14)

From now on, let $\phi_{F_n}(t)$, $\phi_{\hat{F}_n}(t)$ be the characteristic functions of the random variable $\sum_{j=1}^k \zeta_{j,i} / \bar{\sigma}_n$, $\hat{S}_j^i / \hat{\sigma}_n$, respectively. Then, by Lemma 3.8, we have

$$
\sup_x |P(\hat{S}_j^i / \hat{\sigma}_n < x) - G(x)|
\leq \frac{1}{\pi} \int_{-T}^T \left| \frac{\phi_{\hat{F}_n}(t) - \phi_{F_n}(t)}{t} \right| dt + \frac{24M_{\alpha}}{\pi T}
$$

for every $T > 0$. It then follows that for any $T > 0$

$$
\sup_x P(\hat{S}_j^i / \hat{\sigma}_n < x) - G(x)
\leq \frac{1}{\pi} \int_{-T}^T \left| \frac{\phi_{\hat{F}_n}(t) - \phi_{F_n}(t)}{t} \right| dt + \frac{24M_{\alpha}}{\pi T}
\leq \frac{1}{\pi} \int_{|t| < n^{-1/2}} \left| \frac{\phi_{\hat{F}_n}(t) - \phi_{F_n}(t)}{t} \right| dt
+ \frac{1}{\pi} \int_{n^{-1/2} < |t| < n} \left| \frac{\phi_{\hat{F}_n}(t) - \phi_{F_n}(t)}{t} \right| dt
+ \frac{1}{\pi} \int_{|t| < n} \left| \frac{\phi_{\hat{F}_n}(t) - \phi_{F_n}(t)}{t} \right| dt
+ \frac{24M_{\alpha}}{\pi T}
$$

Note that there exists $C_2 > 0$ such that for all $t$

$$
\phi_{\hat{F}_n}(t) - \phi_{F_n}(t) \leq C_2 t.
$$

(15)

Now, setting $T = 1/(4C_1 n^{-1/2+\alpha/2})$ and applying (13), (15) and (12), we then have

$$
\sup_x P(\hat{S}_n^i / \hat{\sigma}_n < x) - G(x) \leq \frac{2C_2}{\pi n^{1/2}} + \frac{2 - \alpha}{\pi} \left( \log n \right) \rho_3^{(n)}
+ \frac{2 - \alpha}{\pi} \left( \log(4C_1) \right) \rho_3^{(n)}
+ 16L_n \int_{-\infty}^\infty t^3 e^{-|t|^2/2} dt
+ \frac{96MC_{1,\alpha}}{\pi n^{-1/2+\alpha/2}}
$$

which immediately implies the lemma. □

We are now ready to prove Theorem 2.2. The key point is $P(S_n / \sigma_n \leq x)$ is “close” to $P(\hat{S}_n^i / \hat{\sigma}_n^i \leq x)$.

Proof of Theorem 2.2: Applying Conditions (a), (c), and Lemma 3.3, we deduce that for any small $\varepsilon_0 > 0$

$$
\sigma_n^2 = \sigma n + O(n^{\varepsilon_0})
$$

and

$$
(\hat{\sigma}_n^i)^2 = E[\sum_{i=1}^k \zeta_{1,i}^2]
- \sum_{i=1}^k E[\zeta_{1,i}^2] + 2 \sum_{i < j} E[\zeta_{1,i} \zeta_{1,j}]
= kE[\zeta_{1,i}^2] + 2 \sum_{i < j} E[\zeta_{1,i} \zeta_{1,j}]
= kE[\zeta_{1,i}^2] + O(kn^{2\alpha} \rho_3^{(n)}) + O(kn^\alpha E[\epsilon \lambda^{(n/2)}])
+ O(k^{1/2} E[\lambda^{(n/2)}])
= k(\sigma n^{\alpha} + O(n^{\varepsilon_0})) + O(kn^\alpha)
+ O(k^{1/2} E[\lambda^{(n/2)}])
= \sigma (n^{\alpha} + O(n^{\varepsilon_0})).
$$

It then follows that

$$
\sigma_n^2 - (\hat{\sigma}_n^i)^2 = O(n^{1-\alpha+\beta}).
$$

Note that there exists $C_2 > 0$ such that for all $t$

$$
\phi_{F_n}(t) - \phi_{\hat{F}_n}(t) \leq C_2 t.
$$

(15)

Next, applying Condition (b) and Lemma 3.3, we have, through simple computations, that

$$
\hat{S}_n^i - O\left( \frac{n^\alpha}{n^{\alpha} + n^{\beta}} \right) = O(n),
$$

$$
S_n - \hat{S}_n^i = O\left( \frac{n^{\beta}}{n^{\beta} + n^{\alpha}} \right) = O(n^{1+\beta-\alpha}),
$$

and

$$
\sigma_n = \Theta(n^{1/2}), \quad \hat{\sigma}_n = \Theta\left( \frac{n^{1/2}}{n^{\alpha} + n^{\beta}} \right) - \Theta(n^{1/2}).
$$
where here by $\sigma_n = \Theta(n^{1/2})$, we mean that there exist positive constants $C_1, C_2$ such that $C_1 n^{1/2} \leq \sigma_n \leq C_2 n^{1/2}$ for all $n$. We then observe that

$$S_n/\sigma_n - \hat{S}_n/\sigma_n = S_n/\sigma_n - \hat{S}_n/\sigma_n + \hat{S}_n/\sigma_n - \hat{S}_n/\sigma_n$$

$$= (S_n - \hat{S}_n)/\sigma_n + \hat{S}_n(1/\sigma_n - 1/\sigma_n)$$

$$= (S_n - \hat{S}_n)/\sigma_n + \hat{S}_n(\sigma_n - \sigma_n^2/\sigma_n)$$

$$= (S_n - \hat{S}_n)/\sigma_n + \hat{S}_n(\sigma_n^2 - \sigma_n^2/\sigma_n)$$

$$= (S_n - \hat{S}_n)/\sigma_n + \hat{S}_n(\sigma_n^2 - \sigma_n^2/\sigma_n)$$

$$= (S_n - \hat{S}_n)/\sigma_n + \hat{S}_n(\sigma_n^2 - \sigma_n^2/\sigma_n)$$

$$= (S_n - \hat{S}_n)/\sigma_n + \hat{S}_n(\sigma_n^2 - \sigma_n^2/\sigma_n)$$

For some $\tau < 0$, let $A_1$ denote the event that

$$|S_n - \hat{S}_n|/\sigma_n \geq n^\tau$$

and let $A_2$ denote the event that

$$\hat{S}_n(\sigma_n^2 - \sigma_n^2/\sigma_n) \geq n^\tau.$$

Then, by the Markov inequality, we have, for any $I \in \mathbb{N}$

$$P(A_1) = P(|S_n - \hat{S}_n|/\sigma_n \geq n^{\tau+1/2}) \leq E[|S_n - \hat{S}_n|^{2}] / n^{(\tau+1/2)/2}. $$

Note that there exist $0 < \epsilon_1, \epsilon_2 < 1$ such that

$$E[|S_n - \hat{S}_n|^{2}] = E[|S_n - \hat{S}_n| + O(\epsilon_1 n^{(1/2)})]^{2}\]

$$= E[\eta_1 + \eta_2 + \ldots + \eta_{n^{1/2}}] + O(\epsilon_2 n^{1/2})$$

$$= \sum O(E[\eta_1], E[\eta_2], \ldots, E[\eta_{n^{1/2}}]) + O(\epsilon_2 n^{1/2}).$$

Then, by Lemmas 3.5 and 3.6, we obtain, through some further computations, that

$$E[|S_n - \hat{S}_n|^{2}] = O(kn^d).$$

Now, applying Lemma 3.5, 3.6 and Conditions (a) and (c), one can verify that

$$P(A_1) = P(|S_n - \hat{S}_n|/\sigma_n \geq n^\tau) \leq E[|S_n - \hat{S}_n|^{2}] / n^{(\tau+1/2)/2}.$$ 

Similarly, applying Lemma 3.5, 3.6 and Conditions (a), (b), one can verify that

$$P(A_2) = P(|\hat{S}_n|/\sigma_n \geq n^{\tau+1/2-\beta+\alpha}) \leq E[|\hat{S}_n|^{2}] / n^{(\tau+1/2-\beta+\alpha)/2}. $$

and

$$P(S_n/\sigma_n \leq x) = P(S_n/\sigma_n \leq x, A_1' \cap A_2')$$

$$+ P(S_n/\sigma_n \leq x, A_1' \cup A_2')$$

and

$$P(S_n/\sigma_n \leq x) = P(S_n/\sigma_n \leq x, A_1' \cap A_2')$$

$$= P(S_n/\sigma_n \leq x, A_1' \cup A_2').$$

It then follows from (16) and (17) that for any $x > -a/2$, there exists $\beta > 0$ such that

$$P(S_n/\sigma_n \leq x, A_1' \cup A_2') \leq P(A_1) + P(A_2) = O(n^{-1/4})$$

and

$$P(S_n/\sigma_n \leq x, A_1' \cap A_2') \geq P(S_n/\sigma_n \leq x - C_1 n^\tau, A_1' \cap A_2')$$

$$\geq P(S_n/\sigma_n \leq x - C_1 n^\tau)$$

$$+ P(A_1' \cap A_2') - 1$$

$$= P(S_n/\sigma_n \leq x - C_1 n^\tau) - C_2 n^{-1/4}$$

for some $C_1, C_2 > 0$. On the other hand, it is easy to check that there exists $C_3 > 0$ such that

$$P(S_n/\sigma_n \leq x, A_1' \cap A_2') \leq P(S_n/\sigma_n \leq x + C_3 n^\tau).$$

Noticing the equation at the bottom of the page and applying Lemma 3.10, we derive

$$P(S_n/\sigma_n \leq x + C_3 n^\tau) - G(x)$$

$$\leq |P(S_n/\sigma_n \leq x + C_3 n^\tau) - G(x + C_3 n^\tau)|$$

$$+ |G(x + C_3 n^\tau) - G(x)|$$

$$= O(n^{-1/2+\alpha/2}) + O(n^\tau)$$

and similarly

$$|G(x) - P(S_n/\sigma_n \leq x - C_1 n^\tau) + C_2 n^{-1/4}|$$

$$= O(n^{-1/2+\alpha/2}) + O(n^\tau) + O(n^{-1/4}).$$

Setting $\alpha = 1/2$, $\tau$ slightly larger than $-1/4$, and choosing $\beta > 0$ sufficiently small, we then have established the theorem.

$$P(S_n/\sigma_n \leq x) - G(x) \leq \max\{P(S_n/\sigma_n \leq x + C_3 n^\tau) - G(x), G(x) - P(S_n/\sigma_n \leq x - C_1 n^\tau) + C_2 n^{-1/4}\}$$
D. Proof of Theorem 2.3

Consider the following Bernstein blocking method with variable block lengths: we consecutively partition the partial sum $S_n$ into blocks $\eta_1, \eta_2, \eta_3, \ldots$ such that $\eta_j$ is of length $g_j = q_j(n = j^\beta)$ and $\zeta_j$ is of length $p_j = p_j(n = j^\alpha)$. Similarly as in the proof of Theorem 2.2, we have

$$S_n = S_n^* + S_n'$$

where $S_n^*$ is the sum of all feasible $\eta$-blocks and $S_n'$ is the sum of all feasible $\zeta$-blocks. Let $\mathcal{L}_i$ denote the $\sigma$-algebra generated by all $X_j$’s contained in $\zeta_i$. It is well known (see, e.g., [19]) that $S_n'$ can be approximated using a Martingale in the following manner:

$$\zeta_i = \xi_i + \nu_i - \nu_{i+1}$$

where

$$\xi_i = \sum_{k=0}^{\infty} (E[\zeta_{i+k} | \mathcal{L}_i] - E[\zeta_{i+k} | \mathcal{L}_{i-1}])$$

is a Martingale difference sequence, and

$$\nu_i = \sum_{k=0}^{\infty} E[\zeta_{i+k} | \mathcal{L}_{i-1}].$$

Similarly as in the proof of Theorem 2.2, we truncate $\zeta$-blocks in the following way: Consider a $\zeta$-block taking the following form:

$$\zeta_i = X_{j_1} + X_{j_1+1} + \cdots + X_{j_k}.$$

For any $j = j_1, j_1+1, \ldots, j_2$, define

$$\hat{X}_j = f(Z_j | Z_{j_1-1}, q, j_2) - E[f(Z_j | Z_{j_1-1}, q, j_2)].$$

and further

$$\hat{\zeta}_i = \hat{X}_{j_1} + \hat{X}_{j_1+1} + \cdots + \hat{X}_{j_2}.$$

Before proving Theorem 2.3, we need to establish several lemmas. The following lemma states that $\nu_i$ is subexponentially small with respect to $i$.

**Lemma 3.11:** There exist $C > 0$, $0 < \epsilon < 1$, and $0 < \delta < \beta$ such that for all $i$

$$|\nu_i| \leq C r^i i^\epsilon.$$

**Proof:** Recall that for some $0 < \rho < 1$

$$\zeta_i - \hat{\zeta}_i = O(\rho^{i/2}).$$

We then have

$$\nu_i = \sum_{k=0}^{\infty} E[\zeta_{i+k} | \mathcal{L}_{i-1}] + O(\rho^{i+1})$$

Similarly as in the proof of Theorem 2.2, we truncate $\eta$-blocks in the following way: Consider a $\zeta$-block taking the following form:

$$\eta_i = X_{j_1} + X_{j_1+1} + \cdots + X_{j_k}.$$

For any $j = j_1, j_1+1, \ldots, j_2$, define

$$\hat{X}_j = f(Z_j | Z_{j_1-1}, q, j_2) - E[f(Z_j | Z_{j_1-1}, q, j_2)].$$

and further

$$\hat{\eta}_i = \hat{X}_{j_1} + \hat{X}_{j_1+1} + \cdots + \hat{X}_{j_2}.$$

Before proving Theorem 2.3, we need to establish several lemmas. The following lemma states that $\nu_i$ is subexponentially small with respect to $i$.

**Lemma 3.11:** There exist $C > 0$, $0 < \epsilon < 1$, and $0 < \delta < \beta$ such that for all $i$

$$|\nu_i| \leq C r^i i^\epsilon.$$

**Proof:** Recall that for some $0 < \rho < 1$

$$\zeta_i - \hat{\zeta}_i = O(\rho^{i/2}).$$

We then have

$$\nu_i = \sum_{k=0}^{\infty} E[\zeta_{i+k} | \mathcal{L}_{i-1}] + O(\rho^{i+1})$$

Similarly as in the proof of Theorem 2.2, we truncate $\eta$-blocks in the following way: Consider a $\zeta$-block taking the following form:

$$\zeta_i = X_{j_1} + X_{j_1+1} + \cdots + X_{j_k}.$$

For any $j = j_1, j_1+1, \ldots, j_2$, define

$$\hat{X}_j = f(Z_j | Z_{j_1-1}, q, j_2) - E[f(Z_j | Z_{j_1-1}, q, j_2)].$$

and further

$$\hat{\zeta}_i = \hat{X}_{j_1} + \hat{X}_{j_1+1} + \cdots + \hat{X}_{j_2}.$$

Before proving Theorem 2.3, we need to establish several lemmas. The following lemma states that $\nu_i$ is subexponentially small with respect to $i$.

**Lemma 3.11:** There exist $C > 0$, $0 < \epsilon < 1$, and $0 < \delta < \beta$ such that for all $i$

$$|\nu_i| \leq C r^i i^\epsilon.$$

**Proof:** Recall that for some $0 < \rho < 1$

$$\zeta_i - \hat{\zeta}_i = O(\rho^{i/2}).$$

We then have

$$\nu_i = \sum_{k=0}^{\infty} E[\zeta_{i+k} | \mathcal{L}_{i-1}] + O(\rho^{i+1})$$

Similarly as in the proof of Theorem 2.2, we truncate $\eta$-blocks in the following way: Consider a $\zeta$-block taking the following form:

$$\zeta_i = X_{j_1} + X_{j_1+1} + \cdots + X_{j_k}.$$

For any $j = j_1, j_1+1, \ldots, j_2$, define

$$\hat{X}_j = f(Z_j | Z_{j_1-1}, q, j_2) - E[f(Z_j | Z_{j_1-1}, q, j_2)].$$

and further

$$\hat{\zeta}_i = \hat{X}_{j_1} + \hat{X}_{j_1+1} + \cdots + \hat{X}_{j_2}.$$

Before proving Theorem 2.3, we need to establish several lemmas. The following lemma states that $\nu_i$ is subexponentially small with respect to $i$.

**Lemma 3.11:** There exist $C > 0$, $0 < \epsilon < 1$, and $0 < \delta < \beta$ such that for all $i$

$$|\nu_i| \leq C r^i i^\epsilon.$$

**Proof:** Recall that for some $0 < \rho < 1$

$$\zeta_i - \hat{\zeta}_i = O(\rho^{i/2}).$$

We then have

$$\nu_i = \sum_{k=0}^{\infty} E[\zeta_{i+k} | \mathcal{L}_{i-1}] + O(\rho^{i+1})$$

Similarly as in the proof of Theorem 2.2, we truncate $\eta$-blocks in the following way: Consider a $\zeta$-block taking the following form:

$$\zeta_i = X_{j_1} + X_{j_1+1} + \cdots + X_{j_k}.$$

For any $j = j_1, j_1+1, \ldots, j_2$, define

$$\hat{X}_j = f(Z_j | Z_{j_1-1}, q, j_2) - E[f(Z_j | Z_{j_1-1}, q, j_2)].$$

and further

$$\hat{\zeta}_i = \hat{X}_{j_1} + \hat{X}_{j_1+1} + \cdots + \hat{X}_{j_2}.$$

Before proving Theorem 2.3, we need to establish several lemmas. The following lemma states that $\nu_i$ is subexponentially small with respect to $i$.

**Lemma 3.11:** There exist $C > 0$, $0 < \epsilon < 1$, and $0 < \delta < \beta$ such that for all $i$

$$|\nu_i| \leq C r^i i^\epsilon.$$
or
\[
\frac{(M_N - 1)^{\alpha+1}}{\alpha + 1} + \frac{(M_N - 1)^{\beta+1}}{\beta + 1} \leq N
\]

\[
\leq \frac{M_N^{\alpha+1} - 1}{\alpha + 1} + \frac{(M_N + 1)^{\beta+1} - 1}{\beta + 1}.
\]

Apparently, we have, for either of the above cases
\[
M_N = \Theta(N^{1/(\alpha+1)}).
\]

As elaborated in [37], a somewhat standard procedure can be followed to establish an almost sure invariance principle. For Theorem 2.3 in this paper, it suffices to prove that
1) for any \( \varepsilon > 0 \)
\[
\sum_{i=1}^{M_N} \eta_i = O(N^{1/3+\varepsilon}) \quad \text{a.s.};
\]
(19)
2) for any \( \varepsilon > 0 \)
\[
\sum_{j=1}^{M_N} T_j - \sigma^2 N + O(N^{2/3+\varepsilon}) \quad \text{a.s.}
\]
(20)
as \( N \) tends to infinity.

We will establish (19) in Lemma 3.13. To establish (20), consider the following decomposition:
\[
\sum_{i=1}^{M_N} T_i - \sigma^2 N = \sum_{i=1}^{M_N} (T_i - E[T_i|\mathcal{L}_{t-1}]) + \sum_{i=1}^{M_N} (E[\xi_i^2|\mathcal{L}_{t-1}] - \xi_i^2) + (\sum_{i=1}^{M_N} \xi_i^2 - \sigma^2 N).
\]

It is then clear that we only need to prove all the above three terms are of \( O(N^{2/3+\varepsilon}) \), for any \( \varepsilon > 0 \).

We need the following well-known lemma, whose proof can be found in [37].

Lemma 3.12: Let \( \{X_j\} \) be a sequence of centered random variables with finite second moments. Suppose that there exists a constant \( s > 0 \) such that all integers \( k \geq j \)
\[
E[(\sum_{i=j}^{k} X_i)^2] = O(k^s - j^s).
\]

Then, for each \( \delta > 0 \), we have
\[
\sum_{i=1}^{N} X_j = O(N^{s/2} \log^{3/2} N) \quad \text{a.s.}
\]

From now on, we set \( \alpha = 1/2 \). The following lemma establishes (19).

Lemma 3.13: With probability 1
\[
\sum_{i=1}^{M_N} \eta_i = O(N^{1/3+\varepsilon})
\]
for any \( \varepsilon > 0 \).

Proof: Note that for any \( j, k \)
\[
E[(\sum_{i=j}^{k} \eta_i)^2] = \sum_{i=j}^{k} E[\eta_i^2] + \sum_{i=j}^{k} \sum_{j' \geq i} E[\eta_i \eta_{j'}].
\]

First, notice that an argument parallel to the proof for Part 2 of Lemma 3.1, together with Conditions (a) and (c), implies that there exists \( 0 < \varepsilon < 1 \) such that for any \( t_1 < t_2 \),
\[
E[\eta_{t_1} \eta_{t_2}] = O(e^{(t_2-t_1)/2}) = O(1).
\]
and thus
\[
\sum_{i_1=j}^{k} \sum_{i_2>i_1} E[\eta_{i_1} \eta_{i_2}] = \sum_{i_1=j}^{k} \sum_{i_2>i_1} O(e^{(t_2-t_1)/2}) = O(1).
\]

Applying Lemma 3.3, we have for some small \( \varepsilon_0 > 0 \)
\[
E[(\sum_{i=j}^{k} \eta_i)^2] = \left( \sum_{i=j}^{k} (\sigma^2 \eta_i^2 + O(\varepsilon_0^2)) \right) + O(1)
\]
\[
- O(k^{\varepsilon_0+1} - j^{\varepsilon_0+1}) + O(k^{\varepsilon_0+1} - j^{\varepsilon_0+1}).
\]

It then follows from Lemma 3.12 that for any \( \beta' > \beta, \varepsilon'_0 > \varepsilon_0 \)
\[
\sum_{i=1}^{M_N} \eta_i = O(M_N^{\beta'/2} + O(M_N^{\varepsilon'_0+1}/2))
\]
\[
- O(M_N^{\beta'/2} + O(\varepsilon_0)) + O(M_N^{\varepsilon'_0+1}/2))
\]
where we have applied the fact that \( M_N = \Theta(N^{1/(\alpha+1)}) \).

Choosing \( \beta, \varepsilon_0, \beta', \varepsilon'_0 > 0 \) sufficiently small and recalling \( \alpha = 1/2 \), the lemma then immediately follows.

The following three lemmas collectively establish (20).

Lemma 3.14: With probability 1
\[
\sum_{i=1}^{M_N} \xi_i^2 - \sigma^2 N = O(N^{2/3+\varepsilon})
\]
for any \( \varepsilon > 0 \).

Proof: Note that by Lemma 3.11, \( \zeta_i \) and \( \xi_i \) are subexponentially close. So, we only need to prove that
\[
\sum_{i=1}^{M_N} \zeta_i^2 - \sigma^2 N = O(N^{2/3+\varepsilon}) \quad \text{a.s.}
\]
for any \( \varepsilon > 0 \).

Depending on whether \( X_N \) is contained in a \( \zeta \)-block or a \( \eta \)-block, we have either
\[
-M_N^\beta + \sum_{i=1}^{M_N} i^\beta \leq N - \sum_{i=1}^{M_N} i^\alpha \leq \sum_{i=1}^{M_N} i^\beta
\]
which implies that
\[
-M_N^\beta + \frac{M_N^{\beta+1}}{\beta + 1} \leq N - \sum_{i=1}^{M_N} i^\alpha \leq \frac{M_N^{\beta+1}}{\beta + 1}
\]
or
\[
\sum_{i=1}^{M_N} i^\beta - M_N^\alpha \leq N - \sum_{i=1}^{M_N} i^\alpha \leq \sum_{i=1}^{M_N} i^\beta - M_N^\alpha
\]
which implies that

$$\frac{(MN-1)^{\beta+1}}{\beta+1} - M_N^\beta < N \sum_{i=1}^{M_N} \xi^a < \frac{(MN+1)^{\beta+1}-1}{\beta+1} - M_N^\beta.$$ 

In any case, applying Lemma 3.3, we have for some small \( \varepsilon_0 > 0 \)

$$E[\sum_{i=1}^{M_N} \xi_i^2] - \sigma^2 N = \sum_{i=1}^{M_N} (\sigma_i^a + O(\xi_i^a)) - \sigma^2 N$$

$$= \sigma^2 (\sum_{i=1}^{M_N} \xi^a - N) + O(\sum_{i=1}^{M_N} \xi_i^a)$$

$$= \sum_{i=1}^{M_N} (M_N^\beta + O(M_N^{\beta+1}) + O(M_N^{\beta+1}))$$

$$= O(N^{\alpha/(\alpha+1)}) + O(N^{(\beta+1)/(\alpha+1)})$$

where we have applied the fact that \( M_N = \Theta(N^{1/(\alpha+1)}) \).

Choosing \( \beta, \varepsilon_0 > 0 \) small enough and recalling \( \alpha = 1/2 \), we then have

$$E[\sum_{i=1}^{M_N} \xi_i^2] - \sigma^2 N = O(N^{2/3+\varepsilon})$$

for any \( \varepsilon > 0 \).

So, to prove the lemma, it suffices to prove that with probability 1

$$\sum_{i=1}^{M_N} (\xi_i^2 - E[\xi_i^2]) = O(N^{2/3+\varepsilon})$$

for any \( \varepsilon > 0 \). Using Conditions (a) and (c), we derive that with probability 1

$$|E[\xi_i^2] - E[\xi_i^2 | L_{i-1}^2]| = |E[\xi_i^2] - E[\xi_i^2 | L_{i-1}^2]| + O(E[\xi_i^2 | i^a \rho^y/2] + O(2^\alpha \rho^y))$$

$$= O(E[\xi_i^2 \xi_i^{-1}]) + O(E[\xi_i^2 | i^a \rho^y/2] + O(2^\alpha \rho^y)).$$

Applying Lemma 3.5, we then have for any \( j, k \)

$$E[(\sum_{i=j}^{k} (\xi_i^2 - E[\xi_i^2 | L_{i-1}^2])^2)] = E[(\sum_{i=j}^{k} (\xi_i^2 - E[\xi_i^2 | L_{i-1}^2]))^2]$$

$$+ O(1)$$

$$= \sum_{i=j}^{k} E[(\xi_i^2 - E[\xi_i^2 | L_{i-1}^2])^2] + O(1)$$

$$\leq \sum_{i=j}^{k} E[\xi_i^2] + O(1)$$

$$= O(j^{2\alpha + (j+1)2\alpha + \cdots + k2\alpha}) + O(1)$$

$$= O(k^{2\alpha+1} - j^{2\alpha+1})$$

where we have used the fact that for \( i_1 \neq i_2 \)

$$E[(\xi_{i_1}^2 - E[\xi_{i_1}^2 | L_{i_1-1}^2])(\xi_{i_2}^2 - E[\xi_{i_2}^2 | L_{i_2-1}^2])] = 0.$$

Applying Lemma 3.12, we then have, for any \( \alpha' > \alpha \),

$$\sum_{i=1}^{M_N} (\xi_i^2 - E[\xi_i^2]) = O(M_N^{2\alpha' + 1/2}) = O(N^{2\alpha' + 1/2 + \alpha'}).$$

Recalling \( \alpha = 1/2 \) and choosing \( \alpha' \) slightly larger than \( 1/2 \), we then have proven the lemma.

**Lemma 3.15:** With probability 1

$$\sum_{i=1}^{M_N} (\xi_i^2 - E[\xi_i^2]) = O(N^{2/3+\varepsilon})$$

for any \( \varepsilon > 0 \).

**Proof:** Note that by Lemma 3.11, \( \zeta_i \) and \( \xi_i \) are subexponentially close. So, we only need to prove that

$$\sum_{i=1}^{M_N} (\xi_i^2 - E[\xi_i^2]) = O(N^{2/3+\varepsilon})$$

for any \( \varepsilon > 0 \). But this has been established in the proof of the previous lemma.

**Lemma 3.16:** With probability 1

$$\sum_{i=1}^{M_N} (T_i - E[T_i | L_{i-1}^2]) = O(N^{2/3+\varepsilon})$$

for any \( \varepsilon > 0 \).

**Proof:** Similarly as in the proof of Lemma 3.14, we have that for any \( j, k \)

$$E[(\sum_{i=j}^{k} (T_i - E[T_i | L_{i-1}^2]))^2] = \sum_{i=j}^{k} E[(T_i - E[T_i | L_{i-1}^2])^2]$$

$$\leq \sum_{i=j}^{k} E[T_i^2]$$

$$\leq \sum_{i=j}^{k} E[\xi_i^2]$$

$$\leq O((j^{2\alpha + (j+1)2\alpha + \cdots + k2\alpha}))$$

$$= O(k^{2\alpha+1} - j^{2\alpha+1}).$$

Then, similarly as in the proof of Lemma 3.15, we deduce that

$$\sum_{i=1}^{M_N} (T_i - E[T_i | L_{i-1}^2]) = O(N^{2/3+\varepsilon})$$

for any \( \varepsilon > 0 \).

**E. Proof of Theorem 2.5**

In this proof, we assume the Bernstein blocking as in Theorem 2.2. Notice that for any \( x > 0 \)

$$P(S_n/n \geq x) = P(S_n \geq nx) = P(\hat{S}_n^2 + S_n - \hat{S}_n \geq nx)$$

$$= P(\hat{S}_n^2 \geq nx - (S_n - \hat{S}_n)).$$
Notice that \( S_n - \hat{S}_n = O(n^{1+\alpha} \gamma^{n-1}) \), so we have
\[
P(S_n/n \geq x) \leq P(S_n' \geq nx') = P(t \hat{S}_n'/p \geq nx'/p) \leq \frac{E[e^{t \hat{S}_n'}/p]}{e^{tnx'/p}}
\]
for some \( 0 < x' < x \). Applying Condition (a), we then have
\[
E[e^{t \hat{S}_n'}/p] = E[e^{t \sum_{i=1}^{n-1} \xi_i/p} e^{t \xi_n/p}] = (1 + O(\lambda^{(n/2)}))(E[e^{t \sum_{i=1}^{n-1} \xi_i/p}] E[e^{t \xi_n}].
\]
An iterative application of (22) gives us that for any \( 0 < t < 1 \)
\[
E[e^{t \xi_i/p}] = E[e^{t \xi_i/p}] - (1 + O(\lambda^{(n/2)}))^{k-1}(E[e^{t \xi_i/p}])^k
\]
as \( n \) goes to infinity. If Condition (d) holds, by Lemma 3.3, as \( n \) goes to infinity (and hence \( p, q \) go to infinity), we have
\[
E[\hat{S}_n^2/p^2] = o(1)
\]
which trivially holds when Condition (d) fails. It then follows that for any \( 0 < t < 1 \)
\[
E[e^{t \xi_i/p}] = 1 + o(1)t^2
\]
and furthermore, for \( t > 0 \) sufficiently small, we have
\[
\frac{E[e^{t \xi_i/p}]}{e^{tx'}} = \frac{1 + o(1)t^2}{1 + tx' + O(1)t^2} < 1
\]
Now, from (21), (23), and (24), we deduce that for any \( x > 0 \), there exists \( x' > 0 \) such that
\[
P(S_n/n \geq x) \leq \frac{E[e^{t \xi_i/p}]}{e^{tx'}} \leq (1 + O(\lambda^{(n/2)}))^{k}(E[e^{t \xi_i/p}] e^{tx'})^k.
\]
Notice that for sufficiently large \( n \), we have
\[
(1 + O(\lambda^{(n/2)}))E[e^{t \xi_i/p}] e^{tx'} < 1
\]
which, together with \( \alpha > 0 \) chosen sufficiently small, we conclude that for any \( x, x' > 0 \), there exists \( 0 < \gamma < 1 \) such that
\[
P(S_n/n \geq x) = O(\gamma^{n-1}).
\]
The proof is then complete.

**F. Alternatives for Condition (d)**

Note that for the case \( Z \) is in fact a Markov chain, a rather explicit alternative condition for Condition (d) has been derived in [46]. This section only assumes Conditions (a)–(c) and gives alternatives for Condition (d) provided Conditions (a)–(c) are satisfied.

Let \( (\Omega, \mathcal{F}, P) \) be the probability space on which \( Z \) is defined, and let \( H_0 = H(\mathcal{F}_k, k \in \mathcal{Z}) \) be the subspace of \( L^2(\mathcal{F}) \) spanned by the equivalence classes of the random variables \( Z_k, k \in \mathcal{Z} \), with inner product defined as
\[
\langle V, W \rangle = E[VW]
\]
for any \( V, W \in H_0 \).

**Lemma 3.17:** If \( \lim_{n \to \infty} E|S_n^2| < \infty \), then there exists a sequence of random variables \( \{V_i, i \in \mathbb{N}\} \) such that \( X_i - V_i - V_{i+1} \) with \( E[V_i^2] = O(1) \) uniformly for all \( i \), and thus \( \sup_{n} E[S_n^2] < \infty \).

**Proof:** Let \( Q \) be an infinite subset of \( \mathbb{N} \) such that \( \sup_{n \in Q} E[S_n^2] < \infty \). Applying Condition (c), we have for any \( n, m \)
\[
E[(S_{n+m} - S_n)^2] = E[\sum_{i=m+1}^{n+m} g(Z_i)^2]
\]
where \( S_0 \) is interpreted as 0. It follows from the Banach–Alaoglu theorem (which states that every bounded and closed set in a Hilbert space is weakly compact; see [39, Sec. 3.15]) that for any \( i \in \mathbb{N} \), there exists \( V_i \in H_0 \) with \( E[V_i^2] < C \), and \( Q_i \), an infinite subset of \( Q \) such that for all \( W \in H_0 \)
\[
\lim_{n \to \infty, n \in Q_i} \langle W, S_{n+i+1} - S_{i+1} \rangle = \langle W, V_i \rangle.
\]
Here, without loss of generality, we can assume that \( Q_i+1 \subset Q_i \) for all \( i \). Then, one verifies that for any \( W \in H_0 \), we have that for any \( i \)
\[
\langle W, X_i - V_i + V_{i+1} \rangle = \lim_{n \to \infty, n \in Q} \langle W, X_i - \sum_{i=1}^{n} (S_{n+i+1} + S_{i+1} - S_i) \rangle = \lim_{n \to \infty, \min \{ n, m \} \in Q} \langle W, X_{n+i+1} \rangle = 0
\]
where we have applied Lemma 3.1 for the last equality. Choosing \( W = X_i - V_i + V_{i+1} \), we then obtain that
\[
\|X_i - V_i + V_{i+1}\|_2 = 0
\]
which implies that
\[
X_i = V_i - V_{i+1}; \text{ a.s.}
\]
It then follows that
\[
E[S_n^2] = E[(V_i - V_{i+1})^2] = E[V_i^2] + E[V_{i+1}^2] - 2E[V_i V_{i+1}]
\]
which, together with \( E[V_i^2] < C \), implies the theorem. \( \square \)

A sequence of positive numbers, \( \{h(i), i \in \mathbb{N}\} \), is said to be slowly varying if for every positive integer \( m \)
\[
\lim_{n \to \infty} h(mn)/h(n) = 1
\]
and it is said to be slowly varying in the strong sense if
\[
\lim_{n \to \infty} \frac{\min_{m \leq n \leq 2m} h(n)}{\max_{m \leq n \leq 2m} h(n)} = 1.
\]
Lemma 3.18: If \( \lim_{n \to \infty} E[S_n^2] = \infty \), then \( E[S_n^2] = n h(n) \), where \( \{h(i), i \in \mathbb{N}\} \) is a sequence of slowly varying positive numbers.

Proof: We only need to show that for every positive integer \( l \)

\[
\lim_{n \to \infty} \frac{\sigma_n^2}{\sigma_n^2} = 1.
\]

Following [30], we use the Bernstein blocking method in the following way: We consecutively partition the partial sum \( S_{in} \) into blocks \( \zeta_1, \eta_1, \zeta_2, \eta_2, \ldots \), such that each \( \zeta_i \) is of length \( n \) and each \( \eta_i \) is of length \( r = \lfloor \log \sigma_n^2 \rfloor \). In other words, for any feasible \( i \)

\[
\zeta_i = \sum_{s=1}^{n} X_{(i-1)n+i_{r+1}} \eta_i = \sum_{s=1}^{r} X_{(i-1)n+i_{r+s}}.
\]

Now

\[
\sigma_n^2 - E[S_n^2]
\]

\[
\sum_{j=1}^{i} E[\zeta_j^2] + 2 \sum_{i \neq j} E[\zeta_i \zeta_j] + \sum_{i} E[\zeta_i \eta_j] + \sum_{i} E[\eta_i \eta_j].
\]

It follows from Lemma 3.3 that for any \( j \)

\[
E[\zeta_j^2] = \sigma_n^2 + O(\sigma_n).
\]

Using an argument similar to the proof for Part 2 of Lemma 3.1, one has that there exists \( 0 < \beta < 1 \) such that for \( i \neq j \)

\[
|E[\zeta_i \zeta_j]| = O(\theta^{|i-j|} \sigma_n^2 \cdot \sigma_n^2)
\]

where we also used (25). Using the Schwartz inequality and (25), we also have

\[
|E[\zeta_i \eta_j]| \leq E[\zeta_j^2]^{1/2} E[\eta_j^2]^{1/2} = O(\sigma_n \sigma_r) = O(\sigma_n \log \sigma_n),
\]

and

\[
|E[\eta_i \eta_j]| \leq O(\sigma_r^2) = O((\log \sigma_n)^2).
\]

It then follows that for any positive integer \( l \)

\[
\sigma_n^2 = l \sigma_n^2 + o(\sigma_n^2),
\]

which immediately implies the lemma. \( \square \)

Lemma 3.19: If \( \lim_{n \to \infty} E[S_n^2] = \infty \), then \( E[S_n^2] = n h(n) \), where \( \{h(i), i \in \mathbb{N}\} \) is a sequence of slowly varying positive numbers in the strong sense.

Proof: Note that by Lemma 3.3, we have that for any \( j \)

\[
\lim_{n \to \infty} \frac{E[(S_{n+1} - S_j)^2]}{E[S_j^2]} = 1
\]

uniformly in \( j \). The lemma then follows from (26), Lemma 3.18, and an almost the same proof for Theorem 8.13 of [10]. \( \square \)

The following lemma is well known (see, e.g., [10, Proposition 0.16]).

Lemma 3.20: Suppose \( \{h(n), n \in \mathbb{N}\} \) is a sequence of positive numbers which is slowly varying in the strong sense. Then, for every \( \varepsilon > 0 \), one has that \( n^\varepsilon h(n) \to \infty \) as \( n \to \infty \).

Lemma 3.21: If \( \lim_{n \to \infty} E[S_n^2] = \infty \), then \( \sigma > 0 \).

Proof: Assume, by contradiction, that \( \sigma = 0 \). Since \( \lim_{n \to \infty} E[S_n^2] = \infty \), we deduce, by Lemma 3.19, that \( E[S_n^2/n] \) is slowly varying in the strong sense. Then, by Lemma 3.20, for any \( \alpha > 0 \), \( n^\alpha E[S_n^2/n] \to \infty \) as \( n \to \infty \). However, by Lemma 3.3, when \( \sigma = 0 \), \( n^\alpha E[S_n^2/n] \to 0 \) as \( n \to \infty \) for any \( 0 < \alpha < 1 \), which is a contradiction. \( \square \)

The following theorem immediately follows from Lemmas 3.17 and 3.21, which gives alternatives for Condition (d) given Conditions (a)–(c) are satisfied.

Theorem 3.22: Under Conditions (a)–(c), the following statements are equivalent:

1) \( \sigma > 0 \).
2) \( \lim_{n \to \infty} \sigma_n = \infty \).
3) \( \lim \sup_{n \to \infty} \sigma_n = \infty \).

IV. PROOFS OF THE MAIN RESULTS

Unless specified otherwise, all the lemmas in this section only assume Condition (I).

For each \( z \in \mathbb{R} \), let \( \Delta_z \) denote the matrix such that \( \Delta_z(i,j) = \Delta_z(j,i) = \xi(j,i) \) for all feasible \( i, j \); obviously \( \sum_{z \in Z} \Delta_z = 0 \). One also observes that for any \( z^m_m \)

\[
p(z^m_m) = \pi \Delta_z \frac{\Delta_z}{m}
\]

where \( \pi \) is the stationary vector of \( Y \), \( 1 \) denotes the all-one column vector, and \( \Delta_z \) is strictly positive if \( m_2 - m_1 \) is large enough. Notice that \( \pi \) is strictly positive, by reblocking the process \( Z \) if necessary, we may assume that all \( \Delta_z \) are positive. It then follows from the argument in [21] and the quotient rule (for taking the derivatives) that

Lemma 4.1: For any \( l \geq 0 \), there exists \( C > 0 \) such that for any \( z^l_{-n} \) and any \( \theta \in \Omega \)

\[
|D_{\theta} \log p(z)_{-n}| < C.
\]

For \( \delta > 0 \), let \( C_{\delta + |\delta|} \) denote the "\( \delta \)-cone" of \( \mathbb{R}^+ \{x \in \mathbb{R} : x > 0\} \) within \( \mathbb{C} \), i.e.,

\[
C_{\delta + |\delta|} = \{x + iy : x \in \mathbb{R} : x > 0, -\delta x \leq y \leq \delta x\}.
\]

For any \( \tilde{r} > 0 \), let \( C_\theta^m(\tilde{r}) \) denote the closed \( \tilde{r} \)-neighborhood of \( \theta_0 \) in \( C^m \). It turns out that for some \( \tilde{r} > r \), \( p^\theta_0(z; z^{-1}_n) \) can be analytically continued to \( p^\theta_0(z; z^{-1}_n) \) for all \( \theta \in C_{\theta_0}^m(\tilde{r}) \), respectively. With the fact that an \( n \times n \) positive matrix induces a contraction mapping on the interior of the \( (n-1) \)-dimensional real simplex under the Hilbert metric [42], the following lemma has been established in [21] (see also a more direct proof in [22] using a complex Hilbert metric).

Lemma 4.2: For any \( \delta > 0 \), there exists \( \rho \) such that for any \( z^m_m \)

\[
p(z^m_m - \rho) \in C_{\delta + |\delta|}.
\]

For any \( \rho \), let \( C_\theta^m(\rho) \) denote the closed \( \rho \)-neighborhood of \( \theta_0 \) in \( C^m \). It turns out that for some \( \tilde{r} > r \), \( p^\theta_0(z; z^{-1}_n) \) can be analytically continued to \( p^\theta_0(z; z^{-1}_n) \) for all \( \theta \in C_{\theta_0}^m(\tilde{r}) \), respectively. With the fact that an \( n \times n \) positive matrix induces a contraction mapping on the interior of the \( (n-1) \)-dimensional real simplex under the Hilbert metric [42], the following lemma has been established in [21] (see also a more direct proof in [22] using a complex Hilbert metric).
2) There exist $C > 0$, $0 < \rho < 1$ and $\tilde{r} > r$ such that for any two hidden Markov sequences $z^0_n, z^\prime_n$ with $z^0_n - z^\prime_n$ (here $m, \hat{m} \geq n > 0$) and all $\theta \in C_{\theta^*}(\tilde{r})$, we have

$$p^\theta(z_0|z_{-n}) - p^\theta(\hat{z}_0|\hat{z}_{-n}) \leq C\rho^n$$

and thus

$$|\log p^\theta(z_0|z_{-n}) - \log p^\theta(\hat{z}_0|\hat{z}_{-n})| \leq C\rho^n$$

Together with the Cauchy integral formula, the above lemma immediately implies the following corollary.

**Corollary 4.3:** For any $l > 0$, there exist $C > 0$, $0 < \rho < 1$ such that for any two hidden Markov sequences $z^0_n, z^\prime_n$ with $z^0_n - z^\prime_n$ (here $m, \hat{m} > n > 0$) and any $\theta \in \Omega$

$$|D^l_\theta p^\theta(z_0|z_{-n}) - D^l_\theta p^\theta(\hat{z}_0|\hat{z}_{-n})| \leq C\rho^n$$

and thus

$$|D^l_\theta \log p^\theta(z_0|z_{-n}) - D^l_\theta \log p^\theta(\hat{z}_0|\hat{z}_{-n})| \leq C\rho^n$$

It is well known [9] that a finite-state irreducible and aperiodic Markov chain is a $\psi$-mixing sequence, and the corresponding $\psi(n)$ decays exponentially as $n \to \infty$. The following lemma asserts that under Condition (I), $Z$ is a $\psi$-mixing sequence and the corresponding $\psi(n)$ decays exponentially as $n \to \infty$. An excellent survey on various mixing sequences can be found in [9]; for a comprehensive exposition to the vast literature on this subject, we refer to [10].

**Lemma 4.4:** $Z^\alpha$ is a $\psi$-mixing sequence, and there exist $C > 0$ and any $\theta \in \Omega$

$$\psi(n) \leq C\lambda^n.$$

**Proof:** Note that for any positive $n$, $m$, $l$, and any $z^m_0, z^{m-1}_1$, we have

$$p(z^m_0|z^{m-1}_1) = \sum_{z^{m-1}_1} \frac{\pi \Delta z^{m-1}_0 \mathbf{1}}{\pi \Delta z^{m-1}_0 \mathbf{1}} = \frac{\pi \Delta z^{m-1}_0 \mathbf{1}}{\pi \Delta z^{m-1}_0 \mathbf{1}}\left(\sum_{z} \Delta z\right)^{-1} \Delta z^{m-1}_0 \mathbf{1}.$$

Let $\lambda_2$ denote the second largest (in modulus) eigenvalue of $\Delta$. By the Perron–Frobenius theory (see, e.g., [42]), $|\lambda_2| < 1$; furthermore, for any $\lambda$ with $\lambda_2 < \lambda < 1$, there exists $C_1 > 0$ such that for any probability vector $x$, we have

$$|x \Delta^n - \pi| \leq C_1 \lambda^n.$$

It then follows that

$$p(z^m_0|z^{m-1}_1) = \pi \Delta z^{m-1}_0 + O(\lambda^n) \Delta z^{m-1}_0 = p(z^m_0) + O(\lambda^n) p(z^m_0).$$

Noting that the constant in $O(\lambda^n)$ is independent of $n, m, l$ and $z^m_0, z^{m-1}_1$, we then conclude that for any $U \in B(\mathbb{Z}_{-\infty}^{-1})$, $V \in B(\mathbb{Z}^\infty)$

$$P(V|U) = P(V) + O(\lambda^n)P(V)$$

which immediately implies the lemma. □

**Remark 4.5:** Exponential forgetting and geometric ergodicity are well-established tools in hidden Markov models [18], [13]. Lemma 4.2 is a complex version of exponential forgetting property, which, roughly speaking, implies the exponential forgetting property of $\log p^\theta(Z_0|Z^{-1}_1)$ and all its derivatives (see Corollary 4.3). Lemma 4.4 establishes the exponential mixing property of the hidden Markov chain $Z$, which implies geometric ergodicity.

In the following, we shall establish the main results by invoking the limit theorems in Section II. Before doing so, we set

$$f(Z, Z^{-1}_1) = D_\theta \log p(Z_n|Z^{-1}_1),$$

$$X_i = f(Z_i, Z^{-1}_1) + E_{\theta_i}[f(Z_i|Z^{-1}_1)].$$

Then, by Corollary 4.3 and Lemma 4.4, Conditions (a)–(c) are satisfied. Theorem 3.22 then immediately translates to

$$S_n - \sum_{i=1}^n X_i = \sigma^2_n = Var\{S_n\} = \sigma = \lim_{n \to \infty} \sigma^2_n/n.$$ (27)

**Lemma 4.6:** Fix $l \in \mathbb{N} \cup \{0\}$ and assume Condition (I). Then, for any $\theta \in \Omega$, the following statements are equivalent:

1) $\sigma^{(l)}(\theta) > 0$.
2) $\lim_{n \to \infty} \sigma_n^{(l)}(\theta) = \infty$.
3) $\lim_{n \to \infty} \sup_{\theta \in \Omega} \sigma_n^{(l)}(\theta) = \infty$.

**A. Proof of Theorem 1.1**

Note that for any $i \geq 0$, applying Corollary 4.3, we have

$$E_{\theta_i}[D_\theta \log p^\theta(Z_i|Z^{-1}_1)] - E_{\theta_i}[D_\theta \log p^\theta(Z_i|Z^{-1}_1)] = O(\rho^{-i})$$

which implies that as $i \to -\infty$, $E_{\theta_i}[D_\theta \log p^\theta(Z_i|Z^{-1}_1)]$ converges to a limit, say $E_{\theta_i}[D_\theta \log p^\theta(Z_i|Z^{-1}_1)]$, such that

$$E_{\theta_i}[D_\theta \log p^\theta(Z_i|Z^{-1}_1)] - E_{\theta_i}[D_\theta \log p^\theta(Z_i|Z^{-1}_1)] = O(\rho^{-i}).$$

It then follows that

$$\frac{E_{\theta_i}[D_\theta \log p^\theta(Z_i|Z^{-1}_1)]}{n} = \frac{\sum_{i=1}^n E_{\theta_i}[D_\theta \log p^\theta(Z_i|Z^{-1}_1)]}{n} = \frac{\sum_{i=1}^n (E_{\theta_i}[D_\theta \log p^\theta(Z_i|Z^{-1}_1)])}{n} + O(\rho^i);$$

which converges to $E_{\theta_i}[D_\theta \log p^\theta(Z_i|Z^{-1}_1)]$ as $n$ tends to infinity. This implies the well definedness of $I^{(l)}(\theta)$.

Now, with (27) and (28), invoking Theorem 2.1, we have

$$\frac{D_\theta \log p^\theta(Z_n)}{n} - \frac{E_{\theta_i}[D_\theta \log p^\theta(Z_i|Z^{-1}_1)]}{n} \to 0 \quad \text{as} \quad n \to \infty$$

which, by the definition of $I^{(l)}(\theta)$, implies the theorem.
B. Proof of Theorem 1.2

We will need the following lemma, whose proof follows from Corollary 4.3 and Lemma 4.4 and a completely parallel argument as in the proof of Lemma 3.3, and thus omitted.

**Lemma 4.7:** Fix \( t \in \mathbb{N} \cup \{0\} \) and assume Conditions (I) and (II). For any \( 0 < \varepsilon_0 < 1 \), there exists \( C > 0 \) such that for any \( m, n \) and any \( \theta \in \Omega \)

\[
\frac{[\sigma_n^{(i)}(\theta)]^2}{n} - [\gamma_n^{(i)}(\theta)]^2 \leq Cn^{-\varepsilon_0}.
\]

Lemma 4.7 immediately implies that

\[
|\sigma_n^{(i)}(\theta)| = O(\sqrt{n})
\]

and furthermore, for any small \( \varepsilon > 0 \), any \( m, n \) and any \( \theta \in \Omega \),

\[
\sigma_n^{(i)}(\theta) - \sqrt{n}\sigma_n^{(i)}(\theta) = O(n^{-1/2+\varepsilon}).
\]

Notice that by Corollary 4.3

\[
E_{\theta_0} [D_\theta \log p(Z_1^n)] - nL^{(i)}(\theta) = O(\sum_{i=1}^\infty \rho^i) = O(1)
\]

and by Lemma 4.1

\[
|D_\theta \log p(Z_1^n) - E_{\theta_0} [D_\theta \log p(Z_1^n)]| = O(n).
\]

Applying (30), (31), (32), and (29), we then have, for some \( \varepsilon > 0 \) sufficiently small, (33) at the bottom of the page. Finally, with (27) and (28), invoking Theorem 2.2, we have the equation at the bottom of the page. It then follows from (33) that for any small \( \varepsilon > 0 \)

\[
P \left( \frac{D_\theta \log p(Z_1^n) - nL^{(i)}(\theta)}{\sqrt{n}\sigma_n^{(i)}(\theta)} < x \right) = G(x) + O(n^{-1/2+\varepsilon}) + O(n^{-1/4+\varepsilon}) = G(x) + O(n^{-1/4+\varepsilon}).
\]

We then have established the theorem.

C. Proof of Theorem 1.3

With (27) and (28), invoking Theorem 2.3, we can redefine the process \( \{S(t), t \geq 0\} \) on a richer probability space together with the standard Brownian motion \( \{B(t), t \geq 0\} \) such that for any \( \varepsilon > 0 \)

\[
\sum_{n \leq t} D_\theta \log p^n(Z_1^n) - E_{\theta_0} [D_\theta \log p^n(Z_1^n)] = O(t^{1/3+\varepsilon}) \text{ a.s. as } t \to \infty.
\]

The theorem then follows from (31).

D. Proof of Theorem 1.4

With (27) and (28), invoking Theorem 2.4, we have

\[
\limsup_{n \to \infty} \frac{D_\theta \log p^n(Z_1^n) - E_{\theta_0} [D_\theta \log p^n(Z_1^n)]}{(2n(\sigma_n^{(i)}(\theta))^2 \log n(\sigma_n^{(i)}(\theta))^2)^{1/2}} = 1 \text{ a.s.}
\]

The theorem then follows from (31).
E. Proof of Theorem 1.5

With (27) and (28), invoking Theorem 2.5 (more rigorously, we invoke the proof of Theorem 2.5, while paying special care to the uniformity over all \( \theta \in \Omega \)), we deduce that for any \( x, \varepsilon > 0 \), there exist \( C > 0, 0 < \gamma < 1 \) such that

\[
P \left( \frac{D^i_\theta \log p^\beta(Z^\alpha) - E_{\theta_0} [D^i_\theta \log p^\beta(Z^\alpha)]}{n} \geq x \right) = O(\gamma^{n-1}).
\]

The theorem then follows from (31).

F. Proof of Theorem 1.6

Again, in this proof, we treat \( \theta \) as a 1-D variable; then, we have \( L^{(2)}(\theta) < 0 \) for all \( \theta \in \Omega \). For simplicity, we assume that \( \tau = 1 \), in other words, with probability 1, \( \arg \max_{\theta \in \Omega} \log p^n(Z^\alpha) \) is nonempty for all \( n \in \mathbb{N} \).

By the mean value theorem, for any \( \theta_n \), there exists a \( \theta_n^* \), a convex combination of \( \theta_0 \) and \( \theta_n \), such that

\[
D^i_\theta \log p^{(\alpha)}(Z^\alpha_n) = D^i_\theta \log p^{(\alpha)}(Z^\alpha_n) + D^i_\theta \log p^{(\beta)}(Z^\alpha_n)(\theta_n - \theta_0).
\]

And, by the definition of \( \theta_n \)

\[
D^i_\theta \log p^{(\beta)}(Z^\alpha_n) = 0.
\]

It then follows that for any \( x > 0 \)

\[
P(\theta_n - \theta_0 \geq x) = P \left( \frac{D^i_\theta \log p^{(\beta)}(Z^\alpha_n) \sqrt{n}}{D^i_\theta \log p^{(\beta)}(Z^\alpha_n)/n} \geq \sqrt{n} x \right).
\]

It follows from nonnegativity of the relative entropy [12] that for all \( \theta \in \Omega \) and for all \( n \)

\[
E_{\theta_0} [D^i_\theta \log p^{(\beta)}(Z^\alpha_n)] \leq E_{\theta_n} [D^i_\theta \log p^{(\beta)}(Z^\alpha_n)]
\]

which implies that

\[
E_{\theta_n} [D^i_\theta \log p^{(\beta)}(Z^\alpha_n)] = 0, \text{ and thus } L^{(1)}(\theta_0) = 0.
\]

By Theorem 1.5, for any \( x_1, \varepsilon_1 > 0 \), there exist \( C_1 > 0, \varepsilon_2 < \gamma_1 < 1 \) such that for any \( n \) and any \( \theta \in \Omega \)

\[
P(\theta_n - \theta_0 \geq x) \leq C_1 \gamma^{n-1}.
\]

where \( \gamma \) denotes the standard normal random variable. It then follows that

\[
P(\theta_n - \theta_0 \geq x) = P(\theta_n - \theta_0 \geq x, F^{(\alpha)}(\tilde{\theta}_n, x_1)) + P(\theta_n - \theta_0 \geq x, F^{(\alpha)}(\tilde{\theta}_n, x_1))
\]

\[
\leq P(\tilde{\theta}_n \geq x) + P(\tilde{\theta}_n \geq x_1, F^{(\alpha)}(\tilde{\theta}_n, x_1))
\]

\[
\leq e^{-n \gamma^{(1)}(\tilde{\theta}_n)} + C_2 \gamma^{n-1}.
\]

where we have used the fact that for any \( y > 0 \)

\[
P(\tilde{\theta}_n \geq y) \leq e^{-y^2}.
\]

Case 2: \( \sigma^{(1)}(\tilde{\theta}_n) = 0 \). Then, by Lemma 4.6, there exists \( C_3 > 0 \) such that for any \( n \)

\[
\sigma^{(1)}(\tilde{\theta}_n) \leq C_3.
\]

It then follows that for any \( x_1 > 0 \)

\[
P(\theta_n - \theta_0 \geq x) = \frac{D^i_\theta \log p^{(\beta)}(Z^\alpha_n)}{\sqrt{n}} \geq \frac{n x_1}{\sigma^{(1)}(\tilde{\theta}_n)}
\]

\[
\leq \frac{\gamma^{(1)}(\tilde{\theta}_n)}{\sigma^{(1)}(\tilde{\theta}_n)}
\]

\[
\leq \gamma^{n-1}.
\]

V. CONCLUSION AND FUTURE WORK

In this paper, we derive several limit theorems for finite-state hidden Markov models. These limit theorems can be applied to a variety of scientific disciplines, particularly to relevant areas in information theory and statistics.

The celebrated Shannon–McMillan–Breiman theorem, which is often referred to as the asymptotic equipartition property (AEP) in information theory, plays a fundamental role in proving the classical channel and source coding theorems, the most fundamental results in information theory. Recently, the CLT with an error estimate and the Chernoff bound for i.i.d. sources have been employed [45] to derive the so-called nonasymptotic equipartition property (NEP), which can be used to establish nonasymptotic channel and source coding results for discrete memoryless channels with i.i.d. sources, in the same way as the AEP to all asymptotic coding theorems established so far. The results and techniques in this paper can be employed to establish NEP for finite-state channels with Markov sources; this is the subject of our forthcoming work.
which may further lead to nonasymptotic channel and source coding theorems for finite-state channels.

Hidden Markov models first originated from a seminal paper by Baum and Petrie [2], where the consistency of the MLE was established. This pioneering work has been extensively generalized in many directions via various approaches. We revisit the settings in [2] and derive the rate of convergence of the MLE using the derived limit theorems. One of the main tools we used is the exponential mixing property of hidden Markov processes, and techniques for proving limit theorems for mixing sequences, which are well developed in the probability theory, however, have not been closely examined in previous work. The framework of this paper, in principle, can be extended or modified to establish similar results for other estimators, such as maximum a posteriori estimator, conditional least square estimator, and so on. In a different direction, our first attempt [23] suggests that, despite a great level of technicality, it is possible to extend this work for general-state hidden Markov models.

REFERENCES


