<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Limit theorems for the sample entropy of hidden Markov chains</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Han, G</td>
</tr>
<tr>
<td><strong>Citation</strong></td>
<td>The 2011 IEEE International Symposium on Information Theory (ISIT), St. Petersburg, Russia, 31 July-5 August 2011. In Proceedings of ISIT, 2011, p. 3009-3013</td>
</tr>
<tr>
<td><strong>Issued Date</strong></td>
<td>2011</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10722/135893">http://hdl.handle.net/10722/135893</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>Proceedings of IEEE International Symposium on Information Theory. Copyright © IEEE.; This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.; ©2011 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 for the Sample Entropy of Hidden Markov Chains

Guangyue Han
University of Hong Kong
Email: ghan@hku.hk

Abstract—The Shannon-McMillan-Breiman theorem asserts that the sample entropy of a stationary and ergodic stochastic process converges to the entropy rate of the same process (as the sample size tends to infinity) almost surely. In this paper, we restrict our attention to the convergence behavior of the sample entropy of hidden Markov chains. Under certain positivity assumptions, we prove that a central limit theorem (CLT) with some Berry-Esseen bound for the sample entropy of a hidden Markov chain, and we use this CLT to establish a law of iterated logarithm (LIL) for the sample entropy.

I. INTRODUCTION AND NOTATIONS

Consider a bi-infinite stationary stochastic process \( Y = (Y_i, i \in \mathbb{Z}) \) on a finite alphabet \( \mathcal{Y} = \{1, 2, \cdots, B\} \). The entropy rate of \( Y \) is defined to be

\[
H(Y) = \lim_{n \to \infty} H(Y^n)/n,
\]

where

\[
H(Y^n) = -\sum_{y^n} p(y^n) \log p(y^n),
\]

here \( y^n := (y_1, y_2, \cdots, y_n) \) denotes an instance of \( Y^n := (Y_1, Y_2, \cdots, Y_n) \) and \( p(y^n) \) denotes the probability mass at \( y^n \). It is well known that \( H(Y) \) can also be written as

\[
H(Y) = \lim_{n \to \infty} H(Y_n|Y_1^{n-1}),
\]

where

\[
H(Y_n|Y_1^{n-1}) = -\sum_{y_i^{n-1}} p(y_i^{n-1}) \log p(y_n|y_i^{n-1}),
\]

here \( p(y_n|y_i^{n-1}) \) denotes the conditional probability mass at \( y_n \) given \( y_i^{n-1} \).

We call \(-\log P(Y^n)/n\) the \( n\)-th order sample entropy of \( Y \). If \( Y \) is also ergodic, the celebrated Shannon-McMillan-Breiman theorem asserts that the \( n\)-th order sample entropy of \( Y \) converges to \( H(Y) \) as \( n \to \infty \) almost surely. The Shannon-McMillan-Breiman theorem can be viewed as an analog of the law of large numbers, a fundamental limit theorem in probability theory. So, it is natural to ask if analogs of other limit theorems in probability theory, such as the central limit theorem (CLT) and the law of iterated logarithm (LIL), also hold for the sample entropy. It turns out that CLT and LIL do not hold when we assume \( Y \) is as general a process as stationary and ergodic; so, in this paper, we restrict our attention to hidden Markov chains (some special stochastic processes which will be defined later).

From now on, assume that \( Y \) is a stationary finite-state Markov chain with transition probability matrix \( \Delta \) with entries

\[
\Delta(i, j) = P(Y_1 = j|Y_0 = i), \quad 1 \leq i, j \leq B.
\]

A hidden Markov chain \( Z \) is a process of the form \( Z = \Phi(Y) \), where \( \Phi \) is a deterministic function defined on \( \mathcal{Y} \) with values from a finite alphabet \( \mathcal{Z} \). Often a hidden Markov chain is alternatively defined as a Markov chain observed when passing through a discrete-time memoryless noisy channel. It is well known that the two definitions are equivalent. For the Markov chain \( Y \), \( H(Y) \) has a simple analytic form:

\[
H(Y) = -\sum_{i, j} P(Y_0 = i) \Delta(i, j) \log \Delta(i, j).
\]

For the hidden Markov chain \( Z \), Blackwell [6] gave an integral formula for \( H(Z) \), however using a measure that is typically too complicated for effective computation of \( H(Z) \). So far, there is no simple and explicit formula for \( H(Z) \). So, many approaches have been adopted to compute and estimate \( H(Z) \) instead: Blackwell’s measure has been used to bound the entropy rate [22], a variation on the classical Birch bounds [5] can be found in [9] and a new numerical approximation of \( H(Z) \) has been proposed in [20]. Generalizing Blackwell’s idea, an integral formula for the derivatives of \( H(Z) \) has been derived in [27]. In another direction, [1], [18], [22], [31], [32], [21], [12], [14], [15], [16], [24], [27] have studied the variation of the entropy rate as parameters of the underlying Markov chain vary.

Another interesting approach, which has greatly motivated this work, is to use Monte Carlo simulation to approximate \( H(Z) \): Recently, based on the Shannon-McMillan-Breiman theorem, efficient Monte Carlo methods for approximating \( H(Z) \) were proposed independently by Arnold and Loeliger [2], Pfister, Soriaga and Siegel [25], Sharma and Singh [28]. The limiting behavior of the sample entropy of a hidden Markov chain, which governs the convergence behavior of such algorithms, is then of great interest. In this direction, a CLT [26] for the sample entropy 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 [17]. In essence, both of the two CLTs are proved using effective Martingale approximations of the sample entropy.

In this paper, adapting some standard techniques for proving limit theorems for mixing sequences, we further characterize the limiting behavior of the sample entropy of \( Z \) under...
certain positivity assumptions. Formally, define $X = (X_i, i = 1, 2, \ldots)$ as
$$X_i = -\log P(Z_i | Z_i^{(-1)}) - H(Z_i | Z_i^{(-1)}),$$
and
$$S_n = \sum_{i=1}^{n} X_i, \quad \sigma_n^2 = \text{Var}(S_n).$$

Unless specified otherwise, we assume, throughout the paper, that

(I) $\Delta$ is a strictly positive matrix; and

(II) $\sigma > 0$, where $\sigma^2 = \lim_{n \to \infty} \sigma_n^2/n$ (the existence of the limit under Condition (I) will be established in Lemma 2.5 and Remark 2.6).

Remark 1.1: It can be shown that given Condition (I) is satisfied, Condition (II) is equivalent to either one of the following

1. $\lim_{n \to \infty} E[S_n^2] = \infty.$
2. $\lim \sup_{n \to \infty} E[S_n^2] = \infty.$

We will prove the following central limit theorem with a Berry-Esseen bound (see [4], [10]; such bound, which is absent in [26], [17], characterizes rate of convergence of the CLT).

Theorem 1.2: Under Conditions (I) and (II), for any $\varepsilon > 0$, there exists $C > 0$ such that for any $n$
$$\sup_x |P(S_n/\sigma_n < x) - \int_{-\infty}^{x} (2\pi)^{-1/2} \exp(-y^2/2) dy| \leq C n^{-1/11+\varepsilon}.$$

We will use the above CLT to prove the following law of iterated logarithm.

Theorem 1.3: Under Conditions (I) and (II), we have
$$\lim_{n \to \infty} \sup \left( \frac{S_n}{2n\sigma^2 \log \log n \sigma^2} \right)^{1/2} = \frac{1}{\alpha} \text{ a.s.}$$

Most of the proofs in this paper are omitted due to space limit; we refer to [13] for a complete version of this paper.

II. KEY LEMMAS

This section includes several key lemmas, among which Lemmas 2.1, 2.2, 2.3 require Condition (I) only. Here, we remark that in this paper $C$ is used to denote a constant, which may not be the same on each appearance.

The following lemma can is well-known (see, e.g., [12] for a rigorous proof).

Lemma 2.1: There exist $C > 0$ and $0 < \rho < 1$ such that for any two hidden Markov sequences $z_{0-m}^{0}, z_{0-n}^{0}$ with $z_{0-n}^{0} = z_{0-n}^{0}$ (here $m, n \geq n \geq 0$), we have
$$|p(z_{0-n}^{0}) - p(z_{0-n}^{0})| \leq C \rho^n. $$

Consequently, there exists $C > 0$ and $0 < \rho < 1$ such that for any $n, t \geq 0$,
$$|\log p(z_{0-t-n}^{0}) - \log p(z_{0-t-n}^{0})| \leq C \rho^n,$$
$$|H(Z_{0-t-n}^{0}) - H(Z_{0-t-n}^{0})| \leq C \rho^n.$$

For a stationary stochastic process $T = T_\infty$, let $B(T_j)$ denote the $\sigma$-field generated by $T_k, k = i, i+1, \ldots, j$. Define
$$\psi(n) = \sup_{U \in B(T_\infty), V \in B(T_\infty)} \frac{|P(V | U) - P(V)|}{P(V)},$$
and $T$ is said to be a $\psi$-mixing sequence if $\psi(n) \to 0$ as $n \to \infty$.

It is well known [7] that a finite-state irreducible and aperiodic Markov chain is a $\psi$-mixing sequence, and the corresponding $\psi(n)$ exponentially decays as $n \to \infty$. The following lemma asserts that under Condition (I), $Z$ is a $\psi$-mixing sequence and the corresponding $\psi(n)$ exponentially decays as $n \to \infty$.

An excellent survey on various mixing sequences can be found in [7]; for a comprehensive exposition to the vast literature on this subject, we refer to [8].

Lemma 2.2: $Z$ is a $\psi$-mixing sequence and there exist $C > 0$ and $0 < \lambda < 1$ such that for any positive $n, \psi(n) \leq C n^\lambda$.

The following lemma shows that for a fixed $j > 0$, $E[X_i X_{i+j}]$ exponentially converges as $i \to \infty$; and for any $i < j$, $E[X_i X_j]$ exponentially decays in $j - i$.

Lemma 2.3: 1) There exist $C > 0$ and $0 < \rho < 1$ such that for all $i, j \geq 0$,
$$|E[X_{i+1} X_{i+1+j}] - E[X_{i} X_{i+j}]| \leq C \rho^i.$$

2) There exist $C > 0$ and $0 < \theta < 1$ such that for any positive $i < j$,
$$|E[X_i X_j]| \leq C \theta^{j-i}.$$

Remark 2.4: By Part 1) of Lemma 2.3, for any fixed $j$, the sequence $E[X_i X_{i+j}], i = 1, 2, \ldots,$ is a Cauchy sequence that exponentially converges. For any fixed $j$, let $a_j = \lim_{n \to \infty} E[X_i X_{i+j}]$. Then by Part 2), $|a_j|$ exponentially decays as $j \to \infty$; consequently, we deduce (for later use) that $a_0 = 2 \sum_{j=1}^{\infty} a_j$ converges.

Lemma 2.5: For any $0 < \alpha < 1$, there exists $C > 0$ such that for any $n$ and $m$,
$$\left| \frac{E[(S_{n+m} - S_m)^2]}{n} - (a_0 + 2 \sum_{j=1}^{\infty} a_j) \right| \leq C n^{-\alpha}$$

here, recall that, as defined in Remark 2.4, $a_j = \lim_{n \to \infty} E[X_i X_{i+j}]$.

Remark 2.6: Choosing $m$ in Lemma 2.5 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 2.7: There exists $C > 0$ such that for all $m$ and $n$
$$E[|S_{n+m} - S_m|^3] \leq C n^{3/2}.$$

III. CENTRAL LIMIT THEOREM

Fix an arbitrarily small $\varepsilon_0 > 0$, and let $p = p(n) = \left[ n^{3/11+\varepsilon_0} \right], q = q(n) = \left[ n^{\varepsilon} \right]$. Choose $k = k(n)$ such that $kp + (k - 1)q \leq n < (k + 1)p + kq$; one easily checks that $k = O(n^{8/11-\varepsilon_0})$. Then, for $1 \leq i \leq k$, define
$$\zeta_i = X_{(i-1)(p+q)+1} + \cdots + X_{ip+(i-1)q}.$$

For $1 \leq i \leq k - 1$, define
$$\eta_i = X_{ip+(i-1)q+1} + \cdots + X_{ip+q},$$
and define
$$\eta_k = \begin{cases} X_{kp+(k-1)q+1} + \cdots + X_n & \text{if } kp + (k - 1)q + 1 \leq n \\ 0 & \text{otherwise}. \end{cases}$$
Now $S_n$ can be rewritten as a sum of $\zeta$-“blocks” and $\eta$-
“blocks”:

$$S_n = S'_n + S''_n := \sum_{i=1}^{k} \zeta_i + \sum_{i=1}^{k} \eta_i.$$ 

The above so called “Bernstein blocking method” is a standard technique to
the proof of limit theorems for a variety of mixing sequences. Roughly speaking, the partial sum $S_n$
is partitioned into “long” blocks $\zeta_1, \zeta_2, \cdots, \zeta_k$ and “short”
blocks $\eta_1, \eta_2, \cdots, \eta_k$. Under certain mixing conditions, all long
blocks are “weakly dependent” on each other, while all short blocks are “negligible” in some sense.

With lemmas in Section II established, the remainder of the proof of Theorem 1.2 becomes more or less standard, which can be roughly outlined as follows:

1) We first show $\mathbb{E}[\exp(itS'_n/\sigma_n)]$ and $\prod_{j=1}^{k} \mathbb{E}[\exp(it\zeta_j/\sigma_n)]$ are “close” (see Lemma 3.1).

2) Standard analysis shows that $\prod_{j=1}^{k} \mathbb{E}[\exp(it\zeta_j/\sigma_n)]$ and $\exp(-t^2/2)$ are “close” (see Lemma 3.2).

3) Then by Esseen’s Lemma [10], $P(S'_n/\sigma_n < x)$ and $\int_{-\infty}^{\infty} (2\pi)^{-1/2} \exp(-y^2/2)dy$ are “close” (see Lemma 3.4).

4) Finally, since $S''_n$ are “negligible”, we conclude, in the proof of Theorem 1.2, that $P(S'_n/\sigma_n < x)$ and $P(S''_n/\sigma_n < x)$ are “close”; and thus $P(S_n/\sigma_n < x)$ is “close” (see Lemma 3.5).

**Lemma 3.1:** There exists $C > 0$ and $0 < \rho_1 < 1$ such that for all $n$ and $|t| \leq n^{1/11}$,

$$\mathbb{E}[\exp(itS'_n/\sigma_n)] - \prod_{j=1}^{k} \mathbb{E}[\exp(it\zeta_j/\sigma_n)] \leq C \rho_1^{g(n)}.$$

**Lemma 3.2:** There exists $C > 0$ such that for all $n$ and $|t| \leq n^{1/11}$,

$$\prod_{j=1}^{k} \mathbb{E}[\exp(it\zeta_j/\sigma_n)] - \exp(-t^2/2) \leq C n^{-1/11+\epsilon_0/2}.$$

The following lemma is a version of Esseen’s lemma, which gives upper bounds on the difference between two distribution functions using the difference between the two corresponding characteristic functions. We refer to page 314 of [29] for a standard proof.

**Lemma 3.3:** Let $F$ and $G$ be distribution functions with characteristic functions $\phi_F$ and $\phi_G$, respectively. Suppose that $F$ and $G$ each has mean 0, and $G$ has a derivative $g$ such that $|g| \leq M$. Then

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

for every $T > 0$.

**Lemma 3.4:** For any $\epsilon > 0$, there exists $C > 0$ such that for all $n$

$$\sup_{x} \left| P(S'_n/\sigma_n < x) - \int_{-\infty}^{x} (2\pi)^{-1/2} \exp(-y^2/2)dy \right| \leq C n^{-1/11+\epsilon}.$$

We are now ready to prove Theorem 1.2.

**Proof of Theorem 1.2:**

Let $A$ denote the event “$|S''_n|/\sigma_n \leq n^{-1/11+\epsilon}$”. Then

$$\mathbb{E}[(|S''_n|/\sigma_n - x)] \leq |P(S'_n/\sigma_n \leq x) - P(S'_n/\sigma_n \leq x)|$$

+ $|P(S''_n/\sigma_n \leq x, A) - P(S''_n/\sigma_n \leq x, A^c)|$

+ $|P(S''_n/\sigma_n \leq x, A^c) - P(S''_n/\sigma_n \leq x, A)|$,

where some necessary modifications are incorporated into the proof of Theorem 1.2.

**Remark 3.5:** If Condition (II) fails, i.e., $\lim_{n \to \infty} \sigma_n^2/n = 0$, then a CLT of degenerated form holds for $\{X_i, i \in \mathbb{N}\}$; more precisely, the distribution of $(X_1 + X_2 + \cdots + X_n)/\sqrt{n}$ converges to that of a centered normal distribution with variance 0, i.e., a point mass at 0, as $n \to \infty$. This is can be readily checked since for any $\epsilon > 0$, the Markov inequality, we have

$$\mathbb{P}(|(X_1 + X_2 + \cdots + X_n)/\sqrt{n} \geq \epsilon|) \leq \sigma_n^2/(\epsilon^2) \to 0 \text{ as } n \to \infty.$$

**IV. LAW OF ITERATED LOGARITHM**

From the central limit theorem with a Berry-Esseen bound (Theorem 1.2), we only need to follow a standard “track” to establish the law of iterated logarithm (Theorem 1.3). In particular, we closely follow the proof of Reznik’s law of the iterated logarithm for a stationary $\phi$-mixing sequence (see page 307 of [29]):

1) As an immediately corollary of Theorem 1.2, the following Lemma 4.1 gives bounds on the tail probability of $S_n$.

2) We then slightly modify Reznik’s maximal inequality to obtain our maximal inequality in Lemma 4.2.

3) Finally, we are ready for the proof of Theorem 1.3, where some necessary modifications are incorporated
into the original Reznik’s proof to deal with the complications stemming from the fact that $X$ is not a stationary mixing sequence.

**Lemma 4.1:** For any $|\delta| < 1$ and $\alpha > 0$, we have 
\[
(\log \sigma_n^2)^{-(1+\delta)^2(1-\alpha)} < P(S_n > (1+\delta)(2\sigma_n^2 \log \log \sigma_n^2)^{1/2}) < (\log \sigma_n^2)^{-(1+\delta)^2(1-\alpha)}
\]
for $n$ sufficiently large.

**Lemma 4.2:** For any $x > 0$, $0 < \alpha < 1/2$ and $C > 0$, we have 
\[
P(\max_{j \leq n} S_j > x) \leq 2P(S_n > x - 2\sigma_n) + Cn^{-\alpha},
\]
for sufficiently large $n$.

We are now ready to prove Theorem 1.3.

**Proof of Theorem 1.3:** We first show that 
\[
\limsup_{n \to \infty} \frac{S_n}{(2\sigma_n^2 \log \log \sigma_n^2)^{1/2}} \leq 1 \quad \text{a.s.};
\]
equivalently, we show that for any $\delta > 0$, 
\[
P\left(\frac{S_n}{(2\sigma_n^2 \log \log \sigma_n^2)^{1/2}} > 1 + \delta \quad \text{i.o.}\right) = 0,
\]
here we remind the reader that by Remark 2.6, $\sigma_n^2 = n(\sigma^2 + o(1))$ and “i.o.” means “infinitely often”.

Fox fixed $M > 1$, define $n_j = M^j$, $j = 1, 2, \ldots$. One then checks that 
\[
\frac{S_n}{(2\sigma_n^2 \log \log \sigma_n^2)^{1/2}} > 1 + \delta \quad \text{i.o.}
\]
\[
\subseteq \max_{n \leq n_{j+1}} \frac{S_n}{(2\sigma_n^2 \log \log \sigma_n^2)^{1/2}} > 1 + \delta \quad \text{i.o.}.
\]
So, to prove (2), it suffices (by the Borel-Cantelli Lemma) to show that 
\[
\sum_{j=1}^{\infty} P\left(\frac{S_n}{(2\sigma_n^2 \log \log \sigma_n^2)^{1/2}} > 1 + \delta\right) < \infty; \tag{3}
\]
in order to prove (3), by Lemma 4.2 and the fact that 
\[
\sum_{j=1}^{\infty} n_j^{-\alpha} = \sum_{j=1}^{\infty} M^{-(j+1)\alpha} < \infty,
\]
it suffices to prove that 
\[
\sum_{j=1}^{\infty} P(S_{n_{j+1}} > (1+\delta)(2\sigma_{n_j}^2 \log \log \sigma_{n_j}^2)^{1/2} - 2\sigma_{n_{j+1}}) < \infty. \tag{4}
\]
Note that there exists $0 < \delta_1 < \delta$ such that for $j$ sufficiently large, 
\[
(1+\delta)(2\sigma_{n_j}^2 \log \log \sigma_{n_j}^2)^{1/2} - 2\sigma_{n_{j+1}} > (1+\delta_1)(2\sigma_{n_j}^2 \log \log \sigma_{n_j}^2)^{1/2}. \tag{5}
\]
Applying Lemma 4.1 with $\alpha$ chosen such that $(1+\delta_1)^2(1-\alpha) > 1$, we deduce that 
\[
\sum_{j=1}^{\infty} P(S_{n_{j+1}} > (1+\delta_1)(2\sigma_{n_j}^2 \log \log \sigma_{n_j}^2)^{1/2}) \leq \sum_{j=1}^{\infty} (\log \sigma_{n_j}^2)^{-(1+\delta_1)^2(1-\alpha)} = \sum_{j=1}^{\infty} O(j^{-\delta_1^2(1-\alpha)}) < \infty. \tag{6}
\]
Immediately, (4) and then (3) and then (2) follow. Here, we remark that the same argument as above with $X_i$ replaced by $-X_i$ leads to 
\[
\liminf_{n \to \infty} \frac{S_n}{(2\sigma_n^2 \log \log \sigma_n^2)^{1/2}} \geq -1 \text{ a.s.} \tag{7}
\]
For the other direction, we next show that 
\[
\limsup_{n \to \infty} \frac{S_n}{(2\sigma_n^2 \log \log \sigma_n^2)^{1/2}} \geq 1 - \delta \quad \text{i.o.} = 1. \tag{8}
\]
For fixed $N > 1$ and $\delta > 0$, let $C_n(\delta)$ be the event 
\[
\{S_{N_n - N_n - 1 + N_{n_1} / 2} > (1 - \delta)g(N^n - N^n - 1 - N^n / 2)\}
\]
where $g(n) = (2\sigma_n^2 \log \log \sigma_n^2)^{1/2}$. With Lemmas 2.1 and 4.1, one checks that there exists $0 < \delta_2 < \delta$ such that for a given $\alpha > 0$, 
\[
P(C_n(\delta)) \geq P(S_{N_n - N_n - 1 + N_{n_1} / 2} > (1 - \delta_2)g(N^n - N^n - 1 - N^n / 2)) \geq \log(N^n - N^n - 1 - N^n / 2)^{- (1 - \delta_2)^2(1 - \alpha)/2}
\]
for sufficiently large $n$. From now on, we choose $\alpha > 0$ such that $(1 - \delta_2)^2(1 + \alpha) < 1$. If $n$ and $N$ are large enough, we have 
\[
N^n - N^n - 1 - N^n / 2 \geq N^n / 2,
\]
which, together with (9), implies that for any $\delta > 0$ 
\[
\sum_{n=1}^{\infty} P(C_n(\delta)) = \infty. \tag{10}
\]
Similarly, let $\hat{C}_n(\delta)$ be the event 
\[
\hat{C}_n(\delta^\prime) \supset C_n(\delta),
\]
which, together with (10), implies that for any $\delta^\prime > 0$ 
\[
\sum_{n=1}^{\infty} P(\hat{C}_n(\delta^\prime)) = \infty.
\]
Again, by Lemma 2.1, for any $\delta > 0$, there exists $0 < \delta'' < \delta$ such that for sufficiently large $n$, 
\[
\hat{C}_n(\delta'') \subset C_n(\delta).
\]
It then follows from an iterative application of Lemma 2.2 that there exists $0 < \theta < 1$ such that for any $n, l$,
\[
P(\cap_{m=n}^{n+l} C_m(\delta)) = \prod_{m=n}^{n+l} (1 - P(\hat{C}_m(\delta''))) + \sum_{m=n}^{n+l} O(\theta^{Nm/4})
\]
\[
\leq \exp(-\sum_{m=n}^{n+l} P(\hat{C}_m(\delta''))) + \sum_{m=n}^{n+l} O(\theta^{Nm/4}).
\]

(11)

So, as $l, n \to \infty$, $P(\cap_{m=n}^{n+l} C_m(\delta)) \to 0$, or equivalently, for any $\delta > 0$,
\[
P(C_n(\delta) \ i.o. = 1).
\]

(12)

Let $B_n$ be the event $“S_{N^{n-1}+N^{n/2}} > -2g(N^{n-1} + N^{n/2})”$. It then follows from (7) that
\[
P(B_n \ i.o. = 1,
\]

which, together with (12), implies that for any $\hat{\delta} > 0$
\[
P(B_n \cap C_n(\hat{\delta}) \ i.o. = 1).
\]

(13)

One then checks that for $\delta > 0$, there exists $0 < \hat{\theta} < \delta$ such that for sufficiently large $n$,
\[
“S_{N^n} > (1 - \hat{\delta})g(N^n) \ i.o.”
\]

$“S_{N^n} > (1 - \hat{\delta})g(N^n - N^{n-1} - N^{n/2}) - 2g(N^{n-1} + N^{n/2}) \ i.o.”$

$“B_n \cap C_n(\delta) \ i.o.”$

It then follows from (13) that
\[
P(S_{N^n} > (1 - \delta)g(N^n) \ i.o. = 1,
\]

which immediately implies (8).

REFERENCES