Nonlinearity

Geometric law for multiple returns until a hazard

Yuri Kifer and Ariel Rapaport

Institute of Mathematics, The Hebrew University, Jerusalem 91904, Israel

E-mail: kifer@math.huji.ac.il and ariel.rapaport@mail.huji.ac.il

Received 21 February 2018, revised 16 October 2018
Accepted for publication 3 January 2019
Published 21 March 2019

Recommended by Professor Konstantin M Khanin

Abstract

For a \( \psi \)-mixing stationary process \( \xi_0, \xi_1, \xi_2, \ldots \) we consider the number \( N_N \) of multiple recurrences \( \{ \xi_{q_i(n)} \in \Gamma_N, i = 1, \ldots, \ell \} \) to a set \( \Gamma_N \) for \( n \) until the moment \( \tau_N \) (which we call a hazard) when another multiple recurrence \( \{ \xi_{q_i(n)} \in \Delta_N, i = 1, \ldots, \ell \} \) takes place for the first time where \( \Gamma_N \cap \Delta_N = \emptyset \) and \( q_i(n) < q_{i+1}(n), i = 1, \ldots, \ell \) are nonnegative increasing functions taking on integer values on integers. It turns out that if \( P\{\xi_0 \in \Gamma_N\} \) and \( P\{\xi_0 \in \Delta_N\} \) decay in \( N \) with the same speed then \( N_N \) converges weakly to a geometrically distributed random variable. We obtain also a similar result in the dynamical systems setup considering a \( \psi \)-mixing shift \( T \) on a sequence space \( \Omega \) and study the number of multiple recurrences \( \{ T^{q_i(n)}\omega \in A^m_n, i = 1, \ldots, \ell \} \) until the first occurrence of another multiple recurrence \( \{ T^{q_{i+1}(n)}\omega \in A^m_n, i = 1, \ldots, \ell \} \) where \( A^m_n, A^n_m \) are cylinder sets of length \( m \) and \( n \) constructed by sequences \( a, b \in \Omega \), respectively, and chosen so that their probabilities have the same order. This work is motivated by a number of papers on asymptotics of numbers of single and multiple returns to shrinking sets, as well as by the papers on open systems studying their behavior until an exit through a ‘hole’.

Keywords: geometric distribution, poisson distribution, multiple returns, stationary process, shifts

Mathematics Subject Classification numbers: Primary: 60F05 Secondary: 37D35, 60J05
1. Introduction

Let \(\xi_0, \xi_1, \xi_2, \ldots\) be a sequence of independent identically distributed (i.i.d.) random variables and \(\Gamma_N, \Delta_N\) be a sequence of sets such that \(\Gamma_N \cap \Delta_N = \emptyset\) and

\[
\lim_{N \to \infty} N^{-1} P\{\xi_0 \in \Gamma_N\} = \lim_{N \to \infty} \nu^{-1} N^{-1} P\{\xi_0 \in \Delta_N\} = 1.
\]

Then by the classical Poisson limit theorem

\[
S_N^{(\lambda)} = \sum_{n=0}^{N-1} \mathbb{I}_{\Gamma_n}(\xi_n) \text{ and } S_N^{(\nu)} = \sum_{n=0}^{N-1} \mathbb{I}_{\Delta_n}(\xi_n),
\]

where \(\mathbb{I}_\Gamma\) is the indicator of a set \(\Gamma\), converge in distribution to Poisson random variables with parameters \(\lambda\) and \(\nu\), respectively. On the other hand, if

\[
\tau_N = \min\{n \geq 0 : \xi_n \in \Gamma_N\}
\]

then it turns out that the sum

\[
S_{\tau_N} = \sum_{n=0}^{\tau_N-1} \mathbb{I}_{\Delta_n}(\xi_n),
\]

which counts returns to \(\Delta_N\) until arriving at \(\Gamma_N\), converges in distribution to a geometric random variable \(\zeta\) with the parameter \(p = (1 - \lambda + \nu)^{-1}\), i.e.

\[
P\{\zeta = k\} = (1 - p)^{k-1} p.
\]

Next, consider a more general setup which includes increasing functions \(q_i(n), i = 1, \ldots, \ell, n \geq 0\) taking on integer values on integers and satisfying 

\[
0 \leq q_1(n) < q_2(n) < \cdots < q_\ell(n)
\]

with all differences \(q_i(n) - q_{i-1}(n)\) tending to \(\infty\) as \(n \to \infty\). Here we deal with ‘nonconventional’ sums

\[
S_{\tau_N} = \sum_{n=0}^{\tau_N-1} \prod_{i=1}^{\ell} \mathbb{I}_{\Delta_n}(\xi_{q_i(n)})
\]

(1.1)

defining now

\[
\tau_N = \min\{n \geq 0 : \prod_{i=1}^{\ell} \mathbb{I}_{\Gamma_n}(\xi_{q_i(n)}) = 1\}
\]

(1.2)

and setting \(\tau_N = \infty\) if the set in braces is empty. Now \(S_{\tau_N}\) equals the number \(X_N\) of multiple returns to \(\Delta_N\) until the first multiple return to \(\Gamma_N\). It turns out that if

\[
\lim_{N \to \infty} N^{-1} P\{\xi_0 \in \Gamma_N\}^\ell = \lim_{N \to \infty} \nu^{-1} N^{-1} N\{\xi_0 \in \Delta_N\}^\ell = 1
\]

(1.3)

then, again, \(S_{\tau_N}\) converges in distribution to a geometric random variable with the parameter

\[
p = \lambda(\lambda + \nu)^{-1}.
\]

In the most general case in this setup we consider \(\xi_0, \xi_1, \xi_2, \ldots\) forming a \(\psi\)-mixing (see section 2) stationary sequence of random variables with \(S_{\tau_N}\) defined again by (1.1). We will show that if (1.3) holds true then, as in the i.i.d. case, \(S_{\tau_N}\) will converge in distribution to a geometric random variable with the parameter \(p = \lambda(\lambda + \nu)^{-1}\). In fact, we will obtain estimates for the total variation distance between the distribution of \(S_{\tau_N}\) and the geometric distribution with the parameter \(p = \lambda(\lambda + \nu)^{-1}\). On the other hand, if the second equality in (1.3) holds true and \(S_N\) is the sum in (1.1) taken up to \(N - 1\) instead of \(\tau_N - 1\) then the distribution of \(S_N\) converges in total variation to the Poisson distribution with the parameter \(\nu\).
When $\ell = 1$ the sum $S_{\nu_\ell}$ describes the number of returns to $\Delta_N$ by the sequence $\{\xi_n\}$ before reaching $\Gamma_N$ which can be interpreted as a ‘hole’ through which the system (particle) exits and the count stops. When $\ell > 1$ the sum $S_{\nu_\ell}$ describes the number of multiple returns to $\Delta_N$ taking place at the moments $q_i(n)$, $i = 1, \ldots, \ell$, until the system performs first multiple return to another set $\Gamma_N$ (disjoint with $\Gamma_N$) which we designate as a ‘hazard’.

We consider in this paper also another setup which comes from dynamical systems but has a perfect probabilistic sense, as well. Let $\zeta_k$, $k = 0, 1, 2, \ldots$ be a $\psi$-mixing discrete time process evolving on a finite or countable state space $\mathcal{A}$. For each sequence $a = (a_0, a_1, a_2, \ldots) \in \mathcal{A}^\infty$ of elements from $\mathcal{A}$ and any $m \in \mathbb{N}$ denote by $a^{(m)}$ the string $a_0, a_1, \ldots, a_{m-1}$ which determines also an $m$-cylinder set $A^m_a$ in $\mathcal{A}^\infty$ consisting of sequences whose initial $m$-string coincides with $a_0, a_1, \ldots, a_{m-1}$. Let $\tau^m_a$ be the first $l$ such that starting at the times $q_1(l), q_2(l), \ldots, q_k(l)$ the process $\zeta_k = \zeta_k(\omega)$, $k \geq 0$ repeats the string $a^{(m)} = (a_0, \ldots, a_{m-1})$. Let $b = (b_0, b_1, \ldots) \in \mathcal{A}^\infty$, $b \neq a$. We are interested in the number of $j < \tau^m_a$ such that process $\zeta_k$ repeats the string $b^{(m)} = (b_0, \ldots, b_{n-1})$ starting at the times $q_1(j), q_2(j), \ldots, q_k(j)$. Employing the left shift transformation $T$ on the sequence space $\mathcal{A}^\infty$ we can represent the number in question as a random variable on $\Omega = \mathcal{A}^\infty$ given by the sum

$$\Sigma_{n,m}^b(\omega) = \sum_{j=0}^{\tau^m_a-1} \prod_{i=1}^{\ell} l_{A^i_j}(T^n_j(\omega)). \tag{1.4}$$

We will show that for any $T$-invariant $\psi$-mixing probability measure $P$ on $\Omega$ and $P$-almost all $a, b \in \Omega$ the distribution of random variables $\Sigma_{n,m}^b$ approaches in the total variance distance as $n \to \infty$ the geometric distribution with the parameter

$$(P(A^m_a))^\ell ((P(A^m_a))\ell + (P(A^m_b))\ell)^{-1}$$

provided the ratio $P(A^m_a)/P(A^m_b)$ stays bounded away from zero and infinity. In particular, if this ratio tends to $\lambda$ when $m = m(n)$ and $n \to \infty$ then the distribution of $\Sigma_{n,m}^b$ converges in total variation distance to the geometric distribution with the parameter $(1 + \lambda^\ell)^{-1}$.

Our results are applicable to larger classes of dynamical systems and not only to shifts. Among such systems are smooth expanding endomorphisms of compact manifolds and Axiom A (in particular, Anosov) diffeomorphisms which have symbolic representations via Markov partitions (see [4]). Then, in place of cylinder sets we can count multiple returns to an element of a Markov partition until first multiple return to another element of this partition. If for such dynamical systems we consider Sinai–Ruelle–Bowen type measures then the results can be extended to returns to geometric balls in place of elements of Markov partitions using approximations of the former by unions of the latter (see, for instance, the proof of theorem 3 in [8]). The results remain true for some systems having symbolic representations with infinite alphabet, for instance, for the Gauss map $Tx = \frac{1}{2} (\text{mod } 1)$, $x \in (0, 1]$, $T0 = 0$ of the unit interval considered with the Gauss measure $G(\Gamma) = \frac{1}{\log 2} \int_{\Gamma} \frac{dt}{1+\tau}$ which is known to be $T$-invariant and $\psi$-mixing with an exponential speed ([7]). It seems that our geometric distribution results are new even for single return cases, i.e. when $\ell = 1$.

The motivation for the present paper is two-fold. On one hand, it comes from the series of papers deriving Poisson type asymptotics for distributions of numbers of single and multiple returns to appropriately shrinking sets (see, for instance, [1, 2, 10] and references there). On the other hand, our motivation was influenced by works on open dynamical systems which study dynamics of such systems until they exit the phase space through a ‘hole’ (see, for instance, [6] and references there). In our setup the number of multiple returns is studied until
a ‘hazard’ which is interpreted as certain ℓ-tuple visits to a set which can be also viewed as a ‘hole’. Then we can think on a system as a cluster of ℓ particles which move together and loose one particle upon visiting a ‘hole’ at prescribed times.

The structure of this paper is as follows. In the next section we will describe precisely our setups and formulate main results. In section 3 we will prove our geometric limit theorem for the case of stationary processes and in section 4 this result will be derived for shifts.

2. Preliminaries and main results

2.1. Stationary processes

Our first setup includes a stationary sequence of random variables \( \xi_0, \xi_1, \xi_2, \ldots \) defined on a complete probability space \( (\Omega, \mathcal{F}, P) \) and a two parameter family of \( \sigma \)-algebras \( \mathcal{F}_{mn} = \sigma(\xi_m, \xi_{m+1}, \ldots, \xi_n) \), \( m \leq n \), i.e. \( \mathcal{F}_{mn} \) is the minimal \( \sigma \)-algebra for which \( \xi_m, \xi_{m+1}, \ldots, \xi_n \) are measurable. Recall, that the \( \psi \)-dependence (mixing) coefficient between two \( \sigma \)-algebras \( \mathcal{G} \) and \( \mathcal{H} \) can be written in the form (see [5]),

\[
\psi(\mathcal{G}, \mathcal{H}) = \sup_{\Gamma \in \mathcal{G}, \Delta \in \mathcal{H}} \left\{ \left| \frac{P(\Gamma \cap \Delta)}{P(\Gamma)P(\Delta)} - 1 \right|, P(\Gamma)P(\Delta) \neq 0 \right\} = \sup \{ ||E(g(\mathcal{G})) - E(g))||_{L^\infty} : g \text{ is } \mathcal{H} \text{ measurable and } E|g| \leq 1 \}.
\]  

Set also

\[
\psi(n) = \sup_{m \geq 0} \psi(\mathcal{F}_{0,m}, \mathcal{F}_{m+n,\infty}).
\]  

The sequence (process) \( \xi_1, \xi_2, \ldots \) is called \( \psi \)-mixing if \( \psi(1) < \infty \) and \( \psi(n) \to 0 \) as \( n \to \infty \).

Our multiple recurrence setup includes also strictly increasing functions \( q_i, i = 1, \ldots, \ell \) taking on integer values on integers and satisfying

\[
0 \leq q_1(n) < q_2(n) < \ldots < q_\ell(n) \text{ for all } n \geq 0
\]

and

\[
q(n) = \min_{k \geq n} \min_{1 \leq i \leq \ell} (q_{i+1}(k) - q_i(k)) \to \infty \text{ as } n \to \infty.
\]  

Set

\[
X_{n,\alpha} = \prod_{i=1}^\ell I_{\alpha_i}(\xi_{q_i(n)}), \quad \alpha = 0, 1
\]

where \( I_0 \) and \( I_1 \) are disjoint Borel sets. The sum \( S_M = \sum_{n=0}^{M-1} X_{n,1} \) counts the number of multiple returns of the sequence \( \xi_1, \xi_2, \ldots \) to \( I_1 \) at times \( q_1(n), q_2(n), \ldots, q_\ell(n) \) for \( 0 \leq n < M \).

Statistical properties of such sums were studied in [9, 10]. Statistical properties of such sums were studied in [9, 10].

Set

\[
\tau = \min \{ n \geq 0 : X_{n,0} = 1 \}
\]

writing \( \tau = \infty \) if the set in braces above is empty. We will describe below the statistical properties of sums \( S_\tau \) (setting \( S_0 = 0 \)) which count the number of multiple returns of the sequence \( \xi_1, \xi_2, \ldots \) to \( I_1 \) at times \( q_1(n), q_2(n), \ldots, q_\ell(n) \) until the random time \( \tau \) which we call a hazard.

For any two random variables or random vectors \( Y \) and \( Z \) of the same dimension denote by \( \mathcal{L}(Y) \) and \( \mathcal{L}(Z) \) their distribution and by

\[
d_{TV}(\mathcal{L}(Y), \mathcal{L}(Z)) = \sup_G |\mathcal{L}(Y)(G) - \mathcal{L}(Z)(G)|
\]
the total variation distance between $\mathcal{L}(Y)$ and $\mathcal{L}(Z)$ where the supremum is taken over all Borel sets. We denote also by $\text{Geo}(\rho)$, $\rho \in (0, 1)$ the geometric distribution with the parameter $\rho$, i.e.

$$\text{Geo}(\rho)\{k\} = \rho(1 - \rho)^k \text{ for each } k \in \mathbb{N} = \{0, 1, \ldots\}.$$ 

Denote by $Q$ the distribution of $\xi_0$, i.e. $P\{\xi_0 \in \Gamma\} = Q(\Gamma)$ for any Borel $\Gamma \subset \mathbb{R}$.

**Theorem 2.1.** Let $\xi_0, \xi_1, \xi_2, \ldots$ be a $\psi$-mixing stationary process and assume that the condition (2.3) holds true. Then for any disjoint Borel sets $\Gamma_0, \Gamma_1$ with $Q(\Gamma_0) > 0$, $\alpha = 0, 1$ and positive integers $M, R$ with $\psi(R) < 2^{\frac{1}{\alpha+1}} - 1$ we have

$$d_{TV}(\mathcal{L}(S_n), \text{Geo}(\rho)) \leq C \left( (1 - Q(\Gamma_0)^\alpha)^M + (Q(\Gamma_0)^\alpha + Q(\Gamma_1)^\alpha) \right) \times \left( Q(\Gamma_0) + Q(\Gamma_1) \right)^MR + \sum_{n=0}^{M} \psi(q(n))) \right) + 2Q(\Gamma_1)^\alpha$$

(2.5)

where $\rho = \frac{Q(\Gamma_0)^\alpha}{Q(\Gamma_0)^\alpha + Q(\Gamma_1)^\alpha}$ and the constant $C > 0$ does not depend on $Q(\Gamma_0)$, $Q(\Gamma_1)$, $M$ and $R$.

Next, let $\Gamma_N, \Delta_N, N = 1, 2, \ldots$ be a sequence of pairs of disjoint sets such that

$$Q(\Gamma_N), Q(\Delta_N) \to 0 \text{ as } N \to \infty \text{ and } 0 < C^{-1} \leq \frac{Q(\Gamma_N)}{Q(\Delta_N)} \leq C < \infty$$

(2.6)

for some constant $C$. Set

$$X^{(N)}_{n,0} = \prod_{i=1}^{\ell} 1_{\Gamma_N}(\xi_{q_i(n)}), \quad X^{(N)}_{n,1} = \prod_{i=1}^{\ell} 1_{\Delta_N}(\xi_{q_i(n)}).$$

$$\tau_N = \min\{n \geq 0 : X^{(N)}_{n,0} = 1\} \text{ and } S^{(N)}_M = \sum_{n=0}^{M-1} X^{(N)}_{n,1}.$$ 

**Corollary 2.2.** Suppose that the conditions of theorem 2.1 concerning the stationary process $\xi_0, \xi_1, \xi_2, \ldots$ and the functions $q(n), i = 1, \ldots, \ell$ are satisfied. Let $\Gamma_N, \Delta_N, N = 1, 2, \ldots$ be Borel sets satisfying (2.6). Then

$$d_{TV}(\mathcal{L}(S^{(N)}_{\tau_N}), \text{Geo}(\rho_N)) \to 0 \text{ as } N \to \infty$$

(2.7)

where $\rho_N = Q(\Gamma_N)^\alpha(Q(\Delta_N)^\alpha + Q(\Gamma_N)^\alpha)^{-1}$. In particular, if

$$\lim_{N \to \infty} \frac{Q(\Delta_N)}{Q(\Gamma_N)} = \lambda$$

(2.8)

then the distribution of $S^{(N)}_{\tau_N}$ converges in total variation as $N \to \infty$ to the geometric distribution with the parameter $(1 + \lambda)^{-1}$.

The arguments in the proof of theorem 2.1 and corollary 2.2 will yield also the following multiple recurrence results which generalize some of the results from [9] (where only independent and Markov sequences $\xi_n, n \geq 0$ were considered).

**Theorem 2.3.** Let the conditions of theorem 2.1 concerning the stationary process $\xi_0, \xi_1, \xi_2, \ldots$ and the functions $q(n), i = 1, \ldots, \ell$ hold true. Let $\Gamma$ be a Borel set, $X_{n} = \prod_{i=1}^{\ell} 1_{\Gamma}(\xi_{q_i(n)})$ and $S_{n} = \sum_{n=0}^{N-1} X_{n}$. Then
\[
d_{\text{TV}}(\mathcal{L}(S_N), \text{Pois}(\lambda)) \leq C N Q(\Gamma)^{\ell} (R Q(\Gamma) + \psi(R)) + \psi(C Q(\Gamma)^{\ell} \sum_{n=0}^{N} \psi(q(n)))
\]
(2.9)

where \( \lambda = N Q(\Gamma)^{\ell} \), \( \psi(x) = xe^{-x} \), \( R < N \) is an arbitrary positive integer with \( \psi(R) < 2^{\frac{1}{2m}} - 1 \), \( C > 0 \) is a constant which does not depend on \( Q(\Gamma) \), \( N \) and \( R \) and \( \text{Pois}(\lambda) \) denotes the Poisson distribution with the parameter \( \lambda \).

**Corollary 2.4.** Under the conditions of theorem 2.3 suppose that in place of one set \( \Gamma \) we have a sequence of Borel sets \( \Gamma_N \) such that
\[
0 < C^{-1} \leq N Q(\Gamma_N)^{\ell} \leq C < \infty
\]
(2.10)
for some constant \( C \). Set
\[
X_n^{\ell(N)} = \prod_{i=1}^{\ell} \| \mathbb{P}_n(\xi_{h_i}(n)) \| \text{ and } S_N = \sum_{n=0}^{N-1} X_n^{\ell(N)}.
\]
Then
\[
d_{\text{TV}}(\mathcal{L}(S_N), \text{Pois}(\lambda_N)) \rightarrow 0 \text{ as } N \rightarrow \infty
\]
(2.11)
where \( \lambda_N = N Q(\Gamma_N)^{\ell} \). In particular, if
\[
\lim_{N \rightarrow \infty} N Q(\Gamma_N)^{\ell} = \lambda
\]
(2.12)
then the distribution of \( S_N \) converges in total variation as \( N \rightarrow \infty \) to the Poisson distribution with the parameter \( \lambda \).

### 2.2. Shifts

Our second setup consists of a finite or countable set \( \mathcal{A} \), the sequence space \( \Omega = \mathcal{A}^{\mathbb{N}} \), the \( \sigma \)-algebra \( \mathcal{F} \) on \( \Omega \) generated by cylinder sets, the left shift \( T : \Omega \rightarrow \Omega \), and a \( T \)-invariant probability measure \( P \) on \( (\Omega, \mathcal{F}) \). We assume that \( P \) is \( \psi \)-mixing with the \( \psi \)-dependence coefficient given by (2.1) and (2.2) considered with respect to the \( \sigma \)-algebras \( \mathcal{F}_{mn} \), \( n \geq m \) generated by the cylinder sets \( \{ \omega = (\omega_0, \omega_1, \ldots) \in \Omega : \omega_i = a_i \text{ for } m \leq i \leq n \} \) for some \( a_m, a_{m+1}, \ldots, a_n \in \mathcal{A} \).

Clearly, \( \mathcal{F}_{mn} = T^{-m} \mathcal{F}_{0,n-m} \) for \( n \geq m \). For each word \( a = (a_0, a_1, \ldots, a_{n-1}) \in \mathcal{A}^n \) we will use the notation \( [a] = \{ \omega = (\omega_0, \omega_1, \ldots) : \omega_i = a_i, i = 0, 1, \ldots, n-1 \} \) for the corresponding cylinder set. Write \( \Omega_P \) for the support of \( P \), i.e.
\[
\Omega_P = \{ \omega \in \Omega : P[\omega_0, \ldots, \omega_n] > 0 \text{ for all } n \geq 0 \}.
\]
For \( n \geq 1 \) set \( \mathcal{C}_n = \{ [w] : w \in \mathcal{A}^n \} \). Since \( P \) is \( \psi \)-mixing it follows (see [10], lemma 3.1) that there exists \( \nu > 0 \) such that
\[
P(A) \leq e^{-\nu n} \text{ for all } n \geq 1 \text{ and } A \in \mathcal{C}_n.
\]
(2.13)
For \( n, m \geq 1 \), \( A \in \mathcal{C}_n \) and \( B \in \mathcal{C}_m \) set \( n \lor m = \max\{n, m\} \), \( n \land m = \min\{n, m\} \),
\[
\pi(A) = \min\{1 \leq k \leq n : A \cap T^{-k}A \neq \emptyset\}
\]
(2.14)
and
\[ \pi(A, B) = \min \{ 0 < k < n \land m : A \cap T^{-k}B \neq \emptyset \lor B \cap T^{-k}A \neq \emptyset \} . \]

Let strictly increasing functions \( q_1, \ldots, q_\ell : \mathbb{N} \to \mathbb{N} \) satisfy (2.3) with \( q(n) \) defined there. For each \( n \in \mathbb{N} \) define also
\[ \gamma(n) = \min \{ k \geq 0 : q(k) \geq 2n \} . \]

For \( \eta \in \Omega \) and \( n \geq 1 \) write \( A_\eta^n = [\eta_0 \ldots \eta_{n-1}] \in C_n \). Let \( \tau_n^\eta : \Omega \to \mathbb{N} \) be with
\[ \tau_n^\eta(\omega) = \inf \{ k \geq 1 : T^\eta(k) \omega \in A_\eta^n \ \text{for all} \ 1 \leq i \leq \ell \} . \]

For \( \eta, \omega \in \Omega \) and \( n, m \geq 1 \) define \( \Sigma_n^{\omega, \eta} : \Omega \to \mathbb{N} \) by
\[ \Sigma_n^{\omega, \eta} = \sum_{k=0}^{\tau_n^\eta - 1} \prod_{i=1}^\ell \| A_\psi^\eta \circ T^\eta(k) \]
and write
\[ \kappa_n^{\omega, \eta} = \min \{ \pi(A_\alpha^\eta, A_\beta^\eta), \pi(A_\alpha^\eta), \pi(A_\alpha^\eta) \} . \]

**Theorem 2.5.** There exists a constant \( C = C(\ell, \psi(1)) \geq 1 \) such that for every \( (\omega, \eta) \in \Omega_\ell \times \Omega_\ell \) and \( n, m \geq 1 \) with \( \psi(m) < (3/2)^{1/(\ell + 1)} - 1 \),
\[ d_{TV}(\mathcal{L}(\Sigma_n^{\omega, \eta}), \text{Geo}(P(A_\eta^n)^\ell, P(A_\eta^n)^\ell + P(A_\eta^n)^\ell)) \leq C \left( e^{-\frac{1}{2} \kappa_n^{\omega, \eta}} \gamma(n \lor m) + \left( 1 + \left( \frac{P(A_\eta^n)^\ell}{P(A_\eta^n)^\ell + P(A_\eta^n)^\ell} \right)^\ell \right) (n \lor m) e^{-\frac{1}{2} \kappa_n^{\omega, \eta}} + \psi(m)^{1/2} \right) . \]

**Example 2.6.** Let us consider an explicit example. Assume \( A = \{ 0, 1, 2 \} \), let \( \phi : \Omega \to \mathbb{R} \) be Hölder continuous, and assume \( P \) is the Gibbs measure corresponding to \( \phi \). There exist constants \( C > 1 \) and \( \Pi \in \mathbb{R} \) such that for each \( \omega \in \Omega \) and \( n \geq 1 \),
\[ C^{-1} \leq \exp(-\Pi n + \sum_{j=0}^{n-1} \phi(T^j \omega)) \leq C . \] (2.14)

Additionally, it is well known that \( P \) is \( \psi \)-mixing and that \( \psi(m) \to 0 \) as \( m \to \infty \) at an exponential speed (see [4]).

Assume \( \ell = 2 \) and that \( q_1(n) = n \) and \( q_2(n) = 2n \) for each \( n \in \mathbb{N} \). This implies that \( \gamma(n) = 2n \) for each \( n \in \mathbb{N} \). Let \( \omega, \eta \in \Omega \) be such that \( \omega_0 = 1, \eta_0 = 2, \text{ and } \omega_j = \eta_j = 0 \) for each \( j \geq 1 \). It is easy to see that \( \kappa_n^{\omega, \eta} = n \) for all \( n \geq 1 \). Also, since \( T^j \omega = T^j \eta \) for \( j \geq 1 \), it follows by (2.14) that
\[ \sup \{ \frac{P(A_\eta^n)}{P(A_\eta^n)} : n \geq 1 \} < \infty . \]

Hence, from theorem 2.5 it follows that
\[ d_{TV}(\mathcal{L}(\Sigma_n^{\omega, \eta}), \text{Geo}(P(A_\eta^n)^\ell, P(A_\eta^n)^\ell + P(A_\eta^n)^\ell)) \to 0 \text{ exponentially fast as } n \to \infty . \]
We now return to our general setup. The following corollary deals with the limit behaviour of $\mathcal{L}^{\omega, n}_{\eta}$ for $P \times P$-typical pairs $(\omega, \eta) \in \Omega \times \Omega$, where $|m(n) - n| = o(n)$. By $o(n)$ we mean an unspecified function $f : \mathbb{N} \to \mathbb{N}$ with $\frac{f(n)}{n} \to 0$ as $n \to \infty$.

**Corollary 2.7.** Let $\{m(n)\}_{n \geq 1} \subset \mathbb{N} \setminus \{0\}$ be with $|m(n) - n| = o(n)$ as $n \to \infty$. Assume that there exists $\beta \in (0, 1)$ and $k \geq 1$ such that $\psi(n) = O(\beta^n)$ and $\gamma(n) = O(n^k)$ for $n \geq 1$. Impose also the finite entropy condition - $\sum_{a \in A} P([a]) \ln P([a]) < \infty$. Then for $P \times P$-a.e. $(\omega, \eta) \in \Omega \times \Omega$,

$$
\lim_{n \to \infty} d_{TV}(\mathcal{L}^{\omega, n}_{\eta}), \text{Geo}(P(A^n_{m(n)})) = 0,
$$

In particular, if

$$
\lim_{n \to \infty} \frac{P(A^n_{\omega})}{P(A^n_{\eta})} = \lambda
$$

then $\mathcal{L}^{\omega, n}_{\eta}$ converges in total variation as $n \to \infty$ to the geometric distribution with the parameter $(1 + \lambda^{-1})^{-1}$.

We observe that, in general (in fact, ‘usually’), the ratio $\frac{P(A^n_{\omega})}{P(A^n_{\eta})}$ will be unbounded for distinct $\omega, \eta \in \Omega$, and so in order to obtain nontrivial limiting geometric distribution it is necessary to choose cylinders $A^n_{\omega}$ and $A^n_{\eta}$ with appropriate lengths. In order to have the ratio $\frac{P(A^n_{\omega})}{P(A^n_{\eta})}$ bounded away from zero and infinity our condition $|m(n) - n| = o(n)$ is, essentially, necessary (at least, in the finite entropy case) which follows from the Shannon–McMillan–Breiman theorem (see [11]).

A number theory (combinatorial) application of our results can be described in the following way. Let $a, b \in (0, 1)$ have base $k$ or continued fraction expansions with digits $a_0, a_1, \ldots$ and $b_0, b_1, \ldots$, respectively. For each point $\omega \in (0, 1)$ with base $k$ or continued fraction expansion with digits $\omega_0, \omega_1, \ldots$, let $\gamma_{a}^{m}(\omega)$ be the smallest $l \geq 0$ such that the $m$-string $a_0, a_1, \ldots, a_{m-1}$ is repeated by the sequence $\omega$ starting from all places $q_i(l), i = 1, \ldots, \ell$. Now, count the number $N_{a}^{b,n}(\omega)$ of those $j < \gamma_{a}^{m}(\omega)$ for which the $n$-string $b_0, b_1, \ldots, b_{n-1}$ is repeated starting from all places $q_i(j), i = 1, \ldots, \ell$. Considering on $(0, 1)$ the Lebesgue measure we conclude from our results that for almost all pairs $a, b$ the distribution of $N_{a}^{b,n}$ in the base $k$ expansion case converges in total variation as $n \to \infty$ to the geometric distribution with the parameter $1/2$.

In the continued fraction case let $G$ be the Gauss measure $G(\Gamma) = \frac{1}{2\pi} \int_{\Gamma} \frac{dt}{t^2 + 1}$ and denote by $[c_0, c_1, \ldots, c_{n-1}]$ the interval of points $\omega \in (0, 1)$ having continued fraction expansion starting with $c_0, \ldots, c_{n-1}$. Then assuming that

$$
\lim_{n \to \infty} \frac{G[b_0, \ldots, b_{n-1}]}{G[a_0, \ldots, a(m(n)-1)]} = \lambda
$$

and that $\kappa_{\omega}^{n,m(n)} \to \infty$ fast enough, we obtain that the distribution of $N_{a}^{b,n}$ converges in total variation to the geometric distribution with the parameter $(1 + \lambda)^{-1}$.
3. Multiple returns for a stationary process

3.1. A lemma

We will need the following result which is, essentially, an exercise in elementary probability.

**Lemma 3.1.** Let \( Y = \{ Y_k : k \geq 0 \text{ and } l \in \{0, 1\} \} \) be independent Bernoulli random variables such that \( 1 > P\{ Y_{k,0} = 1 \} = p = 1 - P\{ Y_{k,0} = 0 \} > 0 \) and \( 1 > P\{ Y_{k,1} = 1 \} = q = 1 - P\{ Y_{k,1} = 0 \} > 0 \). Set \( \tau = \min \{ l \geq 0 : Y_{l,0} = 1 \} \). Then \( S = \sum_{l=0}^{\tau-1} Y_{l,1} \) is a geometric random variable with the parameter \( p(p + q - pq) \).

**Proof.** Clearly
\[
\{ S = m \} = \bigcup_{n=m}^{\infty} \{ \tau = n \} \cap \{ \sum_{l=0}^{n-1} Y_{l,1} = m \}.
\]
Since the processes \( \{ Y_{l,1} \}_{l \geq 0} \) and \( \{ Y_{l,0} \}_{l \geq 0} \) are independent of each other, the events \( \{ \tau = n \} \) and \( \{ \sum_{l=0}^{n-1} Y_{l,1} = m \} \) are independent, as well. Moreover, \( \tau \) and \( \sum_{l=0}^{n-1} Y_{l,1} \) have geometric and binomial distributions, respectively. Thus,
\[
P\{ S = m \} = \sum_{n=m}^{\infty} P\{ \tau = n \} P\{ \sum_{l=0}^{n-1} Y_{l,1} = m \}
= \sum_{n=m}^{\infty} (1-p)^n p \binom{n}{m} q^m (1-q)^{n-m}
= q^m (1-p)^m \sum_{n=0}^{\infty} \binom{n+m}{m} (1-p)^n (1-q)^n.
\]
Set \( r = (1-p)(1-q) \) then
\[
\sum_{n=0}^{\infty} \binom{n+m}{m} r^n (1-r)^{m+1} = 1 \tag{3.1}
\]
since we are summing the probability density (mass) function of the negative binomial distribution with the parameters \( (m+1, r) \). Since \( \binom{n+m}{m} = \binom{n+m+m}{m} \) we obtain from (3.1) and (3.2) that
\[
P\{ S = m \} = \left( \frac{q(1-p)}{1-r} \right)^m \frac{p}{1-r} = \left( \frac{q(1-p)}{p+q-pq} \right)^m \frac{p}{p+q-pq} \tag{3.3}
\]
and taking into account that \( 1 = \frac{q(1-p)}{p+q-pq} = \frac{p}{p+q-pq} \) the proof of the lemma is complete. \( \square \)

3.2. Proof of theorem 2.1

Let \( X'_{n,\alpha}, n = 0, 1, ..., \alpha = 0, 1 \) be a sequence of independent random variables such that \( X'_{n,\alpha} \) has the same distribution as \( X_{n,\alpha} \). Set \( \tau_M = \min(\tau, M) \).
\[ S'_M = \sum_{n=0}^{M-1} X'_{n,1}, \quad \tau' = \min\{n \geq 0 : X'_{n,0} = 1\} \text{ and } \tau'_M = \min(\tau', M). \]

Next, let \( Y_{0,0} \) and \( Y_{1,1} \), \( n = 0, 1, \ldots \) be two independent of each other sequences of i.i.d. random variables such that

\[ P\{Y_{0,0} = 1\} = Q(\Gamma_0)^e = 1 - P\{Y_{0,0} = 0\}, \quad \alpha = 0, 1. \tag{3.4} \]

We can and will assume that all above random variables are defined on the same (sufficiently large) probability space. Set also

\[ S''_M = \sum_{n=0}^{M-1} Y_{n,1}, \quad \tau^* = \min\{n \geq 0 : Y_{n,0} = 1\} \text{ and } \tau''_M = \min(\tau^*, M). \]

Now observe that \( S''_M \) has by lemma 3.1 the geometric distribution with the parameter

\[ \rho = \frac{Q(\Gamma_0)^e}{Q(\Gamma_0)^e + Q(\Gamma_1)^e(1 - Q(\Gamma_0)^e)}. \tag{3.5} \]

Next, we can write

\[ d_{TV}(\mathcal{L}(S), \text{Geo}(\rho)) \leq A_1 + A_2 + A_3 + A_4 + A_5 \tag{3.6} \]

where \( A_1 = d_{TV}(\mathcal{L}(S), \mathcal{L}(S_{\tau})) \), \( A_2 = d_{TV}(\mathcal{L}(S_{\tau}), \mathcal{L}(S'_{\tau})) \), \( A_3 = d_{TV}(\mathcal{L}(S'_{\tau}), \mathcal{L}(S''_{\tau})) \), \( A_4 = d_{TV}(\mathcal{L}(S''_{\tau}), \mathcal{L}(S''_{\tau})) \) and \( A_5 = d_{TV}(\text{Geo}(\rho), \text{Geo}(\rho)). \)

Introduce random vectors \( \mathbf{X}_{M, \alpha} = \{X_{n,\alpha}, 0 \leq n \leq M\}, \quad \alpha = 0, 1, \quad \mathbf{X}_M = \{X_{M,0}, X_{M,1}\}, \quad \mathbf{X}'_M = \{X'_{n,\alpha}, 0 \leq n \leq M\}, \quad \alpha = 0, 1, \quad \mathbf{X}'_M = \{X'_{M,0}, X'_{M,1}\}, \quad \mathbf{Y}_M = \{Y_{M,0}, Y_{M,1}\}, \quad \mathbf{Y}'_M = \{Y'_{n,\alpha}, 0 \leq n \leq M\}, \quad \alpha = 0, 1 \) and \( Y_M = \{Y_{M,0}, Y_{M,1}\}. \) Observe that the event \( \{S' \neq S_{\tau}\} \) can occur only if \( \tau > M \). Also, we can write \( \{\tau > M\} = \{X_{0,0} = 0 \text{ for all } n = 0, 1, \ldots, M\} \) and \( \{\tau' > M\} = \{X'_{0,0} = 0 \text{ for all } n = 0, 1, \ldots, M\} \) Hence,

\[ A_1 \leq P\{\tau > M\} \leq P\{\tau' > M\} + |P\{X_{0,0} = 0 \text{ for } n = 0, 1, \ldots, M\} - P\{X_{0,0} = 0 \text{ for } n = 0, 1, \ldots, M\}| \leq P\{\tau' > M\} + d_{TV}(\mathcal{L}(X_{M,0}), \mathcal{L}(X'_{M,0})). \tag{3.7} \]

and similarly,

\[ P\{\tau' > M\} \leq P\{\tau^* > M\} + d_{TV}(\mathcal{L}(X'_{M,0}), \mathcal{L}(Y_{M,0})). \tag{3.8} \]

Since \( Y_{0,0}, n = 0, 1, \ldots \) are i.i.d. random variables we obtain that

\[ P\{\tau^* > M\} = (P\{Y_{0,0} = 0\})^{M+1} = (1 - Q(\Gamma_0)^e)^{M+1}. \tag{3.9} \]

Next, we claim that

\[ d_{TV}(\mathcal{L}(\mathbf{X}_{M,0}), \mathcal{L}(\mathbf{Y}_{M,0})) \leq d_{TV}(\mathcal{L}(\mathbf{X}_M), \mathcal{L}(\mathbf{Y}_M)) \]

\[ \leq \sum_{0 \leq n \leq M, \alpha = 0, 1} d_{TV}(\mathcal{L}(X'_{n,\alpha}), \mathcal{L}(Y_{n,\alpha})). \tag{3.10} \]

The first inequality above is clear and the second one holds true in view of the following general argument. Let \( \mu_1, \mu_2 \) and \( \tilde{\mu}_1, \tilde{\mu}_2 \) be Borel probability measures on Borel measurable spaces \( \mathcal{X} \) and \( \tilde{\mathcal{X}} \), respectively. Then for any product Borel sets \( U_i \times \tilde{U}_i \subset \mathcal{X} \times \tilde{\mathcal{X}}, \ i = 1, \ldots, k \) such that \( U_i \subset \mathcal{X}, \ U_i \subset \tilde{\mathcal{X}}, \ i = 1, \ldots, k \) and \( U_1, \ldots, U_k \) are disjoint we have
\[ |\mu_1 \times \tilde{\mu}_1((\bigcup_{i=1}^k U_i \times \tilde{U}_i)) - \mu_2 \times \tilde{\mu}_2((\bigcup_{i=1}^k U_i \times \tilde{U}_i))| \leq B_1 + B_2 \]

where

\[ B_1 = \left| \sum_{i=1}^k \mu_1(U_i)(\tilde{\mu}_1(\tilde{U}_i) - \tilde{\mu}_2(\tilde{U}_i)) \right| \]

and

\[ B_2 = \left| \sum_{i=1}^k \tilde{\mu}_2(\tilde{U}_i)(\mu_1(U_i) - \mu_2(U_i)) \right|. \]

Since \( U_1, \ldots, U_k \) are disjoint then

\[ B_1 \leq \sum_{i=1}^k \mu_1(U_i)|\tilde{\mu}_1(\tilde{U}_i) - \tilde{\mu}_2(\tilde{U}_i)| \leq d_{TV}(\tilde{\mu}_1, \tilde{\mu}_2) \]

and

\[ B_2 \leq \max((\mu_1 - \mu_2)(H_+), (\mu_2 - \mu_1)(H_-)) \leq d_{TV}(\mu_1, \mu_2) \]

where \( \mathcal{X} = H_+ \cup H_- \) is the Hahn decomposition of \( \mathcal{X} \) into positive and negative part with respect to the signed measure \( \mu_1 - \mu_2 \). Thus,

\[ |\mu_1 \times \tilde{\mu}_1(W) - \mu_2 \times \tilde{\mu}_2(W)| \leq d_{TV}(\mu_1, \mu_2) + d_{TV}(\tilde{\mu}_1, \tilde{\mu}_2) \]

for any \( W \subset \mathcal{X} \times \tilde{\mathcal{X}} \) having the form \( W = \bigcup_{i \in I \subset \mathcal{I}} (U_i \times \tilde{U}_i) \) with disjoint Borel \( U_1, \ldots, U_k \subset \mathcal{X} \) and arbitrary Borel \( \tilde{U}_1, \ldots, \tilde{U}_k \subset \tilde{\mathcal{X}} \). But any finite union of disjoint Borel subsets of \( \mathcal{X} \times \tilde{\mathcal{X}} \) can be represented in this form, whence the above inequality holds true for all such unions which form an algebra of sets. This inequality is preserved under monotone limits of sets, and so it remains true for any Borel set \( W \subset \mathcal{X} \times \tilde{\mathcal{X}} \) yielding (3.10) by induction on \( M \).

Now,

\[ d_{TV}(\mathcal{L}(X_{n,\alpha}'), \mathcal{L}(Y_{n,\alpha})) = |P[X_{n,\alpha}' = 1] - P[Y_{n,\alpha} = 1]| \]

\[ = |P[\xi_{\psi}(\alpha) \in \Gamma_\alpha \text{ for } i = 1, \ldots, \ell] - Q(\psi)^{\ell}| \leq ((1 + \psi(q(n)))^{\ell} - 1)Q(\Gamma_\alpha)^{\ell} \]  \hspace{1cm} (3.11)

where the last inequality follows from lemma 3.2 in [10] and it is based on standard properties of the \( \psi \)-mixing coefficient. For any positive integer \( M \) we can write

\[ d_{TV}(\mathcal{L}(X_{M}), \mathcal{L}(Y_{M})) \leq (Q(\Gamma_0)^{\ell} + Q(\Gamma_1)^{\ell}) \sum_{n=0}^{M-1} ((1 + \psi(q(n)))^{\ell} - 1). \]  \hspace{1cm} (3.12)

Observe that

\[ d_{TV}(\mathcal{L}(X_{M,0}), \mathcal{L}(X_{M,0}')) \leq d_{TV}(\mathcal{L}(X_{M}), \mathcal{L}(X_{M}')) \text{ and } A_2 \leq d_{TV}(\mathcal{L}(X_{M}), \mathcal{L}(X_{M}')). \]  \hspace{1cm} (3.13)

The first inequality in (3.13) is clear and the second one follows from the fact that \( S_{\gamma_0} = f(X_{M}) \) and \( S_{\gamma_1}' = f(X_{M}') \) for a certain function \( f : \{0, 1\}^{2(M+1)} \to \{0, 1, \ldots, M\} \). We will estimate next \( d_{TV}(\mathcal{L}(X_{M}), \mathcal{L}(X_{M}')) \) relying on [3] warning the reader first that in section 2 we defined \( d_{TV} \) in a more standard way than in [3] where this quantity is multiplied by the factor 2, and so we adjust estimates from there accordingly.

By theorem 3 in [3],
\[
d TV \left( \mathcal{L}(X_M), \mathcal{L}(X'_M) \right) \leq 2b_1 + 2b_2 + b_3 + 2 \sum_{0 \leq n \leq M, \alpha = 0, 1} p_{n,\alpha}^2
\]
(3.14)
where for $\alpha = 0, 1$,
\[
p_{n,\alpha} = P\{X_{n, \alpha} = 1\} = P\{\xi_{q(n)} \in \Gamma_\alpha \text{ for } i = 1, \ldots, \ell \} \leq (1 + \psi(q(n)))^\ell Q(\Gamma_\alpha)^\ell
\]
(3.15)
with the latter inequality satisfied by lemma 3.2 in [10]. In order to define $b_1, b_2$ and $b_3$ we introduce the distance between positive integers
\[
\delta(k, l) = \min_{1 \leq i \leq \ell} |q_i(k) - q_i(l)|
\]
and the set
\[
B_{n,\alpha}^{M,R} = \{ (l, \alpha), (l, 1 - \alpha) : 0 \leq l \leq M, \delta(l, n) \leq R \}
\]
(which, in fact, does not depend on $\alpha$) where an integer $R > 0$ is another parameter. Set also $I_M = \{ (n, \alpha) : 0 \leq n \leq M, \alpha = 0, 1 \}$. Then
\[
b_1 = \sum_{(n, \alpha) \in I_M} \sum_{(l, \beta) \in B_{n,\alpha}^{M,R}} p_{n,\alpha} p_{l,\beta},
\]
(3.16)
\[
b_2 = \sum_{(n, \alpha) \in I_M} \sum_{(l, \beta) \in B_{n,\alpha}^{M,R}} p_{n,\alpha, l,\beta},
\]
(3.17)
where $p_{n,\alpha, l,\beta} = E(X_{n,\alpha} X_{l,\beta})$, and
\[
b_3 = \sum_{(n, \alpha) \in I_M} \sum_{n,\alpha}
\]
(3.18)
where
\[
s_{n,\alpha} = E[E(X_{n,\alpha} - p_{n,\alpha})| \sigma \{ X_{l,\beta} : (l, \beta) \in I_M \setminus B_{n,\alpha}^{M,R} \}]
\]
Since the functions $q_i, i = 1, \ldots, \ell$ are strictly increasing, for any $i, j, n$ and $k$ there exists at most one $l$ such that $q_i(n) = q_j(l) = k$. It follows from here that
\[
|B_{n,\alpha}^{M,R}| \leq 8 \ell^2 (R + 1)
\]
(3.19)
where $|U|$ denotes the cardinality of a finite set $U$. It follows from (3.15), (3.16) and (3.19) that
\[
b_1 \leq 4(M + 1)\ell^2 (R + 1)(1 + \psi(1))^\ell Q(\Gamma_0)^{2\ell} + Q(\Gamma_1)^{2\ell}.
\]
(3.20)
Next,
\[
p_{(n,\alpha), l,\beta} = P(X_{n,\alpha} = X_{l,\beta} = 1) = 0
\]
(3.21)
if $n = l$ and $\beta = 1 - \alpha$ since $\Gamma_0 \cap \Gamma_1 = \emptyset$. If $n \neq l$ then assuming, for instance, that $l > n$ we obtain by lemma 3.2 in [10] that
\[
p_{(n,\alpha), l,\beta} = P(X_{n,\alpha} = X_{l,\beta} = 1)
\leq P(X_{n,\alpha} = 1 \text{ and } \xi_{q_i(l)} \in \Gamma_\beta) \leq (1 + \psi(1))^i Q(\Gamma_\alpha)^i Q(\Gamma_\beta).
\]
(3.22)
Hence,
\[
b_2 \leq 2(M + 1)\ell^2 (R + 1)(1 + \psi(1))^i Q(\Gamma_0)^i + Q(\Gamma_1)^i (Q(\Gamma_0) + Q(\Gamma_1)).
\]
(3.23)
Next, we claim that
\[
s_{n, \alpha} \leq 2^{2(l+2)}(2 - (1 + \psi(R))^\ell + 1)^{-2}\psi(R)|X_{n, \alpha} - p_{n, \alpha}|
\leq 2^{2l+5}(2 - (1 + \psi(R))^\ell + 1)^{-2}\psi(R)p_{n, \alpha}
\]  
(3.24)
where \(s_{n, \alpha}\) is the same as in (3.18). Indeed, let \(G\) be the \(\sigma\)-algebra generated by all \(\xi_{q(j)}\), \(i = 1, \ldots, \ell\) such that \((l, 0) \in I_M \setminus B_{n, \alpha}^M\) and \(H\) be the \(\sigma\)-algebra generated by \(\xi_{q(j)}\), \(i = 1, \ldots, \ell\). Since \(|q_i(n) - q_i(l)| > R\) for all \(i, j = 1, \ldots, \ell\) and \(l\) such that \((l, 0) \in I_M \setminus B_{n, \alpha}^M\) we conclude from (3.24) from (2.1) and (3.25). Now by (3.15), (3.18) and (3.24),
\[
b_3 \leq 2^{2l+5}(M + 1)(2 - (1 + \psi(R))^\ell + 1 - 2\psi(1))^\ell \psi(R)(Q(\Gamma_0)^\ell + Q(\Gamma_1)^\ell).
\]  
(3.26)
Next, in the same way as in the estimate of \(A_2\) we conclude that
\[
A_3 \leq d_{TV}(L(X_Q), L(Y_Q))
\]  
(3.27)
which together with (3.12) estimates \(A_3\).

As in the estimate of \(A_1\) we see that
\[
A_4 \leq \mathcal{P}\{\tau^* > M\} \leq (1 - Q(\Gamma_0)^\ell)^{M+1}
\]  
(3.28)
since \(Y_{i,0}, n = 0, 1, \ldots\) are i.i.d. random variables.

Since \(\varrho > \rho\) we obtain
\[
A_3 \leq \sum_{k=0}^{\infty} |\varrho(1 - \varrho)^k - \rho(1 - \rho)^k| \leq 2\sum_{k=1}^{\infty} ((1 - \rho)^k - (1 - \varrho)^k)
= 2(1 - \rho)\varrho^{-1} - 2(1 - \varrho)\rho^{-1} = 2\frac{\varrho - \rho}{\rho\varrho} = 2Q(\Gamma_1)^\ell.
\]  
(3.29)
Collecting (3.6)–(3.15), (3.20), (3.23), (3.24) and (3.26)–(3.29) we derive (2.5).}

In order to prove corollary 2.2 we rely on the estimate (2.5) with \(\Gamma_0 = \Gamma_N\) and \(\Gamma_1 = \Delta_N\) choosing \(M = MN \to \infty\) and \(R = RN \to \infty\) as \(N \to \infty\) so that
\[
\lim_{N \to \infty} MNQ(\Gamma_N)^\ell = \infty, \lim_{N \to \infty} Q(\Gamma_N)^\ell \sum_{n=0}^{MN} \psi(q(n)) = 0,
\]
\[
\lim_{N \to \infty} MN\psi(R_N)Q(\Gamma_N)^\ell = 0 \text{ and } \lim_{N \to \infty} MN R_N Q(\Gamma_N)^\ell + 1 = 0
\]  
(3.30)
which is clearly possible since \(\psi(n) \to 0\) and \(q(n) \to \infty\) as \(n \to \infty\). This together with (2.5) yields (2.7).

3.3. Returns until a fixed time

Now we will prove theorem 2.3. By theorem 1 in [3],
\[
d_{TV}(L(S_N), \text{Pois}(ES_N)) \leq b_1 + b_2 + b_3
\]  
(3.31)
where \(b_1, b_2\) and \(b_3\) are defined by (3.16)–(3.18) with the sums there taken only in \(n\) and \(l\) (but not in \(\alpha\)), taking \(N\) in place of \(M\) and replacing there \(p_{n, \alpha}\) by
\( p_n = P\{X_n = 1\} \), \( p_{\alpha, \beta} \) by \( p_{\alpha, \beta} = E(X_n X_l) \), \( B_{\alpha, \beta}^R \) by \( B_{\alpha, \beta}^R = \{ l : 0 \leq l \leq N, \delta(l, n) \leq R \} \) and \( s_{\alpha, \beta} \) by \( s_{\alpha, \beta} = E(E(X_n - p_n)\{X_l \in I_n \setminus B_{\alpha, \beta}^R\}) \) where \( I_n = \{0, 1, \ldots, N\} \). Then all right hand side estimates (3.20), (3.23) and (3.26) remain valid but we will have to consider only one set \( \Gamma_0 = \Gamma \) (deleting terms with \( Q(\Gamma_1) \) there) and in order to complete the proof of theorem 2.3 it remains to show that

\[
\text{d}_{\text{TV}}(\text{Pois}(ES_N), \text{Pois}(NQ(\Gamma)^f)) \leq \psi(CQ(\Gamma)^f \sum_{n=0}^{N} \psi(q(n))).
\]

(3.32)

Indeed, for any \( \lambda_1, \lambda_2 > 0 \) by lemma 3.4 in [10],

\[
d_{\text{TV}}(\text{Pois}(\lambda_1), \text{Pois}(\lambda_2)) \leq \sum_{n=0}^{\infty} [e^{-\lambda_1} \frac{\lambda_1^n}{n!} - e^{-\lambda_2} \frac{\lambda_2^n}{n!}] = 2|\lambda_1 - \lambda_2| e^{-|\lambda_1 - \lambda_2|}.
\]

(3.33)

Now,

\[
ES_N = \sum_{n=0}^{N-1} EX_n \text{ and } EX_n = P\{X_n = 1\} = P\{\xi_{p(n)} \in \Gamma \text{ for } i = 1, \ldots, \ell\}.
\]

By lemma 3.2 in [10] (which is an easy application of the definitions (2.1) and (2.2) of the \( \psi \)-dependence coefficient) together with stationarity of the sequence \( \{\xi_n\} \) we obtain

\[
|P\{\bigcap_{i=1}^{\ell} \{\xi_{p(n)} \in \Gamma\}\} - Q(\Gamma)^f| \leq ((1 + \psi(q(n)))^f - 1) Q(\Gamma)^f.
\]

(3.34)

Hence,

\[
|ES_N - NQ(\Gamma)^f| \leq Q(\Gamma)^f \sum_{n=0}^{N} ((1 + \psi(q(n)))^f - 1) \leq CQ(\Gamma)^f \sum_{n=0}^{N} \psi(q(n)),
\]

(3.35)

where \( C > 0 \) does not depend on \( N \) or \( \Gamma \), and (3.32) follows.

In order to prove corollary 2.4 we rely on (2.9) with \( \Gamma = \Gamma_N \) and choosing \( R = R_N \to \infty \) as \( N \to \infty \) so that \( \lim_{N \to \infty} R_N Q(\Gamma_N) = 0 \). In view of (2.10) and taking into account that \( \psi(n) \to 0 \) and \( q(n) \to \infty \) as \( n \to \infty \) we obtain that

\[
\lim_{N \to \infty} \psi(CQ(\Gamma_N)^f \sum_{n=0}^{N} \psi(q(n))) = 0
\]

which together with (2.9) yields (2.11).

\section{4. Returns to cylinder sets for shifts}

\subsection{4.1. Preliminary lemmas and corollary 2.7}

First we prove corollary 2.7 while relying on theorem 2.5, for which we need the following lemma. In what follows \( \{m(n)\}_{n \geq 1} \) is a sequence of positive integers with \( |m(n) - n| = o(n) \) as \( n \to \infty \). For \( n \geq 1 \) we write \( b(n) = n \wedge m(n) \).

**Lemma 4.1.** Set \( c = 3\nu^{-1} \) and let \( \mathcal{E} \) be the set of all \( (\omega, \eta) \in \Omega \times \Omega \) for which there exists \( N = N(\omega, \eta) \geq 1 \) such that \( m(n) \geq b(n) - c \ln b(n) \) for all \( n \geq N \), then \( P \times P(\Omega^2 \setminus \mathcal{E}) = 0 \).
Proof. For $\omega \in \Omega$ and $n \geq 1$ set
\[ B_{\omega,n} = \{ \eta \in \Omega : \pi(A_{m(n)}^n, A^\eta_{m(n)}) \leq b(n) - c \ln b(n) \}. \]
Assume $b(n) - c \ln b(n) \geq 1$ and set $d = [b(n) - c \ln b(n)]$, then
\[ P(B_{\omega,n}) \leq \sum_{r=0}^{d} P(\eta : T^{-r}A^\eta_{m(n)} \cap A^\eta_{m(n)} \neq \emptyset) + \sum_{r=0}^{d} P(\eta : T^{-r}A^\eta_{m(n)} \cap A^\eta_{m(n)} \neq \emptyset). \]

For $0 \leq r \leq d$,
\[ \{ \eta : T^{-r}A^\eta_{m(n)} \cap A^\eta_{m(n)} \neq \emptyset \} = T^{-r}[\omega_0, ..., \omega_{n \wedge (m(n)-r)-1}] \]
and
\[ \{ \eta : T^{-r}A^\eta_{m(n)} \cap A^\eta_{m(n)} \neq \emptyset \} = [\omega_r, ..., \omega_{n \wedge (m(n)+r)-1}]. \]

Hence by (2.13),
\[ P(B_{\omega,n}) \leq \sum_{r=0}^{d} e^{-(b(n)+r)} + \sum_{r=0}^{d} e^{-(b(n)+r) \wedge m(n)} \leq 2 \sum_{r=0}^{d} e^{-b(n)-r} \leq \frac{2b(n)^{-3}}{1 - e^{-b(n)}} \]

From this and since $|b(n) - n| = o(n)$ it follows that $\sum_{n=1}^{\infty} P(B_{\omega,n}) < \infty$, and so by the Borel–Cantelli lemma
\[ P(\eta : \#\{n \geq 1 : \eta \in B_{\omega,n}\} = \infty) = 0. \]

From Fubini’s theorem we now get,
\[ P \times P(\{\omega, \eta : \#\{n \geq 1 : \pi(A_{m(n)}^\omega, A^\eta_{m(n)}) \leq b(n) - c \ln b(n)\} = \infty\} \]
\[ = \int_{\Omega} P(\eta : \#\{n \geq 1 : \eta \in B_{\omega,n}\} = \infty) \, dP(\omega) = 0. \]

In a similar manner (see [10, corollary 2.2]) it can be shown that
\[ P(\omega : \#\{n \geq 1 : \pi(A_{m(n)}^\omega) \leq b(n) - c \ln b(n)\} = \infty) = 0 \]
and
\[ P(\eta : \#\{n \geq 1 : \pi(A^\eta_{m(n)}) \leq b(n) - c \ln b(n)\} = \infty) = 0. \]

This completes the proof of the lemma. \qed

Proof of corollary 2.7. Let $c$ and $E$ be as in the statement of lemma 4.1. Denote by $h$ the entropy of the system $(\Omega, P, T)$ which is finite under our assumptions. Let $E_0$ be the set of all $(\omega, \eta) \in E \cap (\Omega_P \times \Omega_P)$ for which

\[ 1539 \]
\[ - \lim_{n \to \infty} \frac{\log P(A^\omega_n)}{n} = - \lim_{n \to \infty} \frac{\log P(A^\eta_n)}{n} = h. \]

Observe (see, for instance, [5]) that under \( \psi \)-mixing (even under \( \alpha \)-mixing) condition the shift \( T \) is mixing in the ergodic theory sense with respect to the invariant measure \( P \), and so \( T \) is ergodic. By the Shannon–McMillan–Breiman theorem (see, for instance, [11]) and lemma 4.1 it follows that \( P \times P(\Omega^2 \setminus \mathcal{E}_0) = 0 \) and \( h \geq \nu > 0 \) by (2.13).

Let \((\omega, \eta) \in \mathcal{E}_0\), then for every \( n \geq 1 \) large enough
\[ e^{-\nu \omega_n^{\omega, \eta}/2} \leq \exp\left(-\frac{\nu(b(n) - c \log b(n))}{2}\right) \leq b(n)^2 \cdot e^{-\nu b(n)/2}. \]

By our assumption \( \gamma(n) \) grows at most polynomially, hence by \(|m(n) - n| = o(n)\),
\[ e^{-\nu \omega_n^{\omega, \eta}/2} \gamma(n \vee m(n)) \xrightarrow{\mathcal{D}} 0 \text{ as } n \to \infty. \]

From
\[ \lim_{n \to \infty} \log \left( \frac{P(A^\omega_n)}{P(A^\eta_{m(n)})} \right)^{\ell} (\nu \vee m(n))b(n)^2e^{-\nu b(n)/2} \]
\[ = \lim_{n \to \infty} n \cdot \left( \ell \log P(A^\omega_n) - \ell \log P(A^\eta_{m(n)}) \right) + \frac{\log(n \vee m(n))}{n} + 2 \log b(n)/n - \frac{\nu b(n)}{2n} \]
\[ = \left( \ell h - \ell h - \frac{\nu}{2} \right) \lim_{n \to \infty} n = -\infty, \]
it follows that
\[ \left( 1 + \left( \frac{P(A^\omega_n)}{P(A^\eta_{m(n)})} \right)^{\ell} \right) (\nu \vee m(n))e^{-\nu \omega_n^{\omega, \eta}/2} \to 0 \text{ as } n \to \infty. \]

By our assumption \( \psi(n) \) tends to 0 at an exponential rate as \( n \to \infty \), hence we also have
\[ \left( 1 + \left( \frac{P(A^\omega_n)}{P(A^\eta_{m(n)})} \right)^{\ell} \right) \psi(m(n))^{1/2} \to 0 \text{ as } n \to \infty. \]

The corollary now follows directly from theorem 2.5. \( \square \)

In what follows we will consider \( \ell \) and \( \psi(1) \) as global constants. Hence, whenever we use the big-O notation we consider the implicit constant may depend on these parameters. We will need also the following result.

**Lemma 4.2.** Let \( \ell \geq 1 \) and \( n \geq 1 \) be such that \( \psi(n) < (3/2)^{1/(\ell+1)} - 1 \). Then for every \( \eta \in \Omega_P \),
\[ |P \left\{ P(A^\eta_n) \gamma_n > t \right\} - \text{Pois}(t) \{0\}| = O \left( te^{-\nu \pi(A^\eta_n)} (n + \gamma(n)) + \nu \psi(n) \right), \]
where, recall, \( \text{Pois}(t) \) is the Poisson distribution with the parameter \( t \).
Proof. For $n \geq 1$ set $N_n = \lfloor tP(A_n^0)^{-t} \rfloor$ and

$$S_n = \sum_{k=1}^N \prod_{i=1}^\ell I_{A_k^i} \circ T^n_{\psi(k)},$$

then

$$P\{S_n = 0\} = P \left\{ P(A_n^0)^{-t} r_n^0 > t \right\} .$$

(4.1)

By [10, theorem 2.1],

$$|P\{S_n = 0\} - \text{Pois}(t)\{0\}| = O \left( e^{-\psi(A_n^0)} (n + \gamma(n)) + t\psi(n) \right).$$

This together with (4.1) proves the lemma.

4.2. Proof of theorem 2.5

Let $(\omega, \eta) \in \Omega_p \times \Omega_p$ and $n, m \geq 1$ with $\psi(m) < (3/2)^{1/(\ell+1)} - 1$. Set $\kappa = \kappa_{n,m}$, then we can clearly assume that

$$e^{-\psi/2} (n \lor m + \gamma(n \lor m)) \leq \frac{1}{2} .$$

(4.2)

Set $\epsilon = \max\{e^{-\psi}, \psi(m)\}$ and $t = e^{-1/2}$, then $0 < \epsilon \leq 1$ and $e^{-t} < \epsilon$.

Set $p_0 = P(A_n^0)^t$, $p_\omega = P(A_n^w)^t$, $L = [t : p_n^{-1}]$, $I_0 = \{\gamma(n \lor m), \ldots, L \times \{0\}$, $I_1 = \{\gamma(n \lor m), \ldots, L \times \{1\}$, and $I = I_0 \cup I_1$. For $\gamma(n \lor m) \leq k \leq L$ define random variables $X_{k,0}$ and $X_{k,1}$ on $(\Omega, F, \mathbb{P})$ by

$$X_{k,0} = \prod_{i=1}^\ell I_{A_k^i} \circ T^n_{\psi(k)} \text{ and } X_{k,1} = \prod_{i=1}^\ell I_{A_k^i} \circ T^n_{\psi(k)},$$

and denote the Bernoulli process $\{X_{k,l} : (k, l) \in I\}$, i.e. $X_{k,l}$ takes values 0 or 1 only, by $\mathbf{X}$. Let $\mathbf{X}' = \{X_{k,l}' : (k,l) \in I\}$ be a Bernoulli process, with $\mathcal{L}(X_{k,l}') = \mathcal{L}(X_{k,l})$ for each $(k,l) \in I$, such that the $X_{k,l}'$ are all mutually independent.

Let $Y = \{Y_{k,l} : k \geq 0, l = 0, 1\}$ be a collection of independent Bernoulli random variables such that $\mathbb{P}\{Y_{k,0} = 1\} = p_0 = 1 - \mathbb{P}\{Y_{k,0} = 0\}$ and $\mathbb{P}\{Y_{k,1} = 1\} = p_\omega = 1 - \mathbb{P}\{Y_{k,1} = 0\}$. Write also,

$$Y' = \{Y_{k,l}' : (k,l) \in I\} .$$

For $y \in \{0, 1\}^{\mathbb{N} \times \{0,1\}}$ set

$$\tilde{f}(y) = \inf\{k \geq 0 : y_{k,0} = 1\} \text{ and } \tilde{g}(y) = \sum_{j=0}^{\tilde{f}(y)-1} y_{j,1},$$

and for $y \in \{0, 1\}^I$ set
\[ f(y) = \min \{ \gamma(n \lor m) \leq k \leq L : y_{k,0} = 1 \text{ or } k = L \} \] and \( g(y) = \sum_{j=\gamma(n \lor m)}^{f(y)-1} y_{j,1} \).

Let \( S \subseteq \mathbb{N} \), then
\[
\left| P\left\{ \Sigma_{n,m}^w \in S \right\} - \text{Geo}\left( \frac{p_\eta}{p_\eta + p_w} \right)(S) \right| \leq \left| P\left\{ \Sigma_{n,m}^w \in S \right\} - P\left\{ g(X) \in S \right\} \right| \\
+ \left| P\left\{ g(X) \in S \right\} - P\left\{ g(X') \in S \right\} \right| + \left| P\left\{ g(X') \in S \right\} - P\left\{ g(Y) \in S \right\} \right| \\
+ \left| P\left\{ g(Y) \in S \right\} - P\left\{ \tilde{g}(Y) \in S \right\} \right| + \left| P\left\{ \tilde{g}(Y) \in S \right\} - \text{Geo}\left( \frac{p_\eta}{p_\eta + p_w} \right)(S) \right| \\
= \sigma_1 + \sigma_2 + \sigma_3 + \sigma_4 + \sigma_5. 
\] (4.3)

Let us estimate \( \sigma_1 \) from above. The event \( \left\{ \Sigma_{n,m}^w \neq g(X) \right\} \) is contained in the union of the events \( \left\{ \tau^w_m > L \right\} \) and
\[
E = \left\{ \prod_{i=1}^\ell \tilde{a}^w_i \circ T^{\theta(k)} = 1 \text{ or } \prod_{i=1}^\ell \tilde{a}^w_i \circ T^{\theta(k)} = 1 \text{ for some } 0 < k < \gamma(n \lor m) \right\}. 
\]

By (2.13),
\[
P(E) = O(\gamma(n \lor m)e^{-\psi(n \lor m)}). 
\]

Since \( \text{Pois}(t)\{0\} = e^{-\varepsilon} < \varepsilon \), this together with lemma 4.2 yields
\[
\sigma_1 = O\left( \varepsilon + te^{-\psi \varepsilon}(m + \gamma(n \lor m)) + t\psi(m) \right). 
\]

Note that \( \sigma_2 \leq d_{TV}(\mathcal{L}(X), \mathcal{L}(X')) \), hence the following lemma gives an upper estimate on \( \sigma_2 \).

**Lemma 4.3.** It holds that,
\[
d_{TV}(\mathcal{L}(X), \mathcal{L}(X')) = O\left( t \left( 1 + \frac{p_w}{p_\eta} \right) (\gamma(n \lor m)e^{-\psi \varepsilon} + \psi(n \lor m)) \right). 
\]

**Proof of lemma 4.3.** For \((k, l) \in I\) set
\[
B_{k,l} = \bigcup_{j=1}^\ell \left\{ (r, s) \in I : |q_i(r) - q_j(k)| \leq 2(n \lor m) \right\}
\]
and
\[
G_{k,l} = \sigma \left\{ X_{r,s} : (r, s) \in I \setminus B_{k,l} \right\}. 
\]

By theorem 3 in [3],
\[
d_{TV}(\mathcal{L}(X), \mathcal{L}(X')) \leq O(b_1 + b_2 + b_3), 
\]
where
\[
b_1 = \sum_{(k, l) \in I} \sum_{(r, s) \in B_{k,l}} P\{X_{k,l} = 1\} P\{X_{r,s} = 1\}, 
\]
\[ b_2 = \sum_{(k,l) \in I} \sum_{(r,s) \in B_{k,l} \backslash (k,l)} P\{X_{k,l} = 1 = X_{r,s}\}, \]

and

\[ b_3 = \sum_{(k,l) \in I} E |E[X_{k,l} - E[X_{k,l}] | \mathcal{G}_{k,l}]|. \]

Let us estimate \( b_1 \) from above. By (2.13),
\[ P\{X_{k,l} = 1\} \leq e^{-\varphi(n \wedge m)} \text{ for } (k,l) \in I. \]

Also, by the \( \psi \)-mixing assumption (see [10, lemma 3.2]),
\[ P\{X_{k,l} = 1\} = \begin{cases} O(p_\eta) & \text{if } l = 0 \\ O(p_\omega) & \text{if } l = 1 \end{cases} \text{ for } (k,l) \in I. \]

Since \(|B_{k,l}| = O(n \lor m)\) for \((k,l) \in I\),
\[ b_1 = O \left( L(n \lor m) \cdot e^{-\varphi(n \wedge m)}(p_\eta + p_\omega) \right). \]

Hence by \( Lp_\eta \leq t \),
\[ b_1 = O \left( t(n \lor m) \cdot e^{-\varphi(n \wedge m)}(1 + \frac{p_\omega}{p_\eta}) \right). \]

We shall now estimate \( b_2 \). Let \((k,l) \in I\) and \((r,s) \in B_{k,l} \backslash (k,l)\) be given. Assume without loss of generality that \( k \geq r \). If \(|q_1(r) - q_1(k)| < \kappa\) then \( \{X_{k,l} = 1 = X_{r,s}\} = \emptyset \) by the definition of \( \kappa \). Otherwise, by the \( \psi \)-mixing assumption and (2.13),
\[ P\{X_{k,l} = 1 = X_{r,s}\} = O \left( P\{X_{k,l} = 1\} \cdot e^{-\varphi \kappa} \right). \]

Hence, by the considerations made for bounding \( b_1 \),
\[ b_2 = O \left( t(n \lor m) \cdot e^{-\varphi \kappa}(1 + \frac{p_\omega}{p_\eta}) \right). \]

Finally, we estimate \( b_3 \) from above. Given \((k,l) \in I\) it follows, by the argument given in the proof of [10, theorem 2.1] in order to estimate \( b_3 \), that
\[ E \left| E\left( X_{k,l} - EX_{k,l} \right) | \mathcal{G}_{k,l} \right| = O \left( \psi(n \lor m)P\{X_{k,l} = 1\} \right). \]

Hence,
\[ b_3 = O \left( t\psi(n \lor m)(1 + \frac{p_\omega}{p_\eta}) \right). \]

Now by the estimates on \( b_1, b_2 \) and \( b_3 \) the lemma follows. \( \square \)

We now resume the main proof and estimate \( \sigma_3 \). As explained in section 3, given probability distributions \( \mu_1, \mu_2, \nu_1, \nu_2 \), on the same measurable space, it holds that
\[ d_{TV}(\mu_1 \times \nu_1, \mu_2 \times \nu_2) \leq d_{TV}(\mu_1, \mu_2) + d_{TV}(\nu_1, \nu_2). \]

From this and since \( X' \) and \( Y' \) are independent Bernoulli processes,
\[
\sigma_3 \leq d_{TV}(\mathcal{L}(X'), \mathcal{L}(Y')) = \sum_{(k,l) \in I} d_{TV}(\mathcal{L}(X'_{kl}), \mathcal{L}(Y_{kl})) = \sum_{(k,l) \in I} |P[X'_{kl} = 1] - P[Y_{kl} = 1]|.
\]

For \( \gamma(n \lor m) \leq k \leq L \) it follows by the \( \psi \)-mixing assumption (see [10, lemma 3.2]) that
\[
|P[X'_{k,0} = 1] - P[Y_{k,0} = 1]| = \left| P \left( \prod_{i=1}^k \delta_{\omega_i} \circ T^{\eta(k)} = 1 \right) - p_\eta \right| = O(\psi(n \lor m) p_\eta),
\]
and similarly
\[
|P[X'_{k,1} = 1] - P[Y_{k,1} = 1]| = O(\psi(n \lor m) p_\omega).
\]

Hence,
\[
\sigma_3 = O(L\psi(n \lor m)(p_\eta + p_\omega)) = O(t\psi(n \lor m)(1 + \frac{p_\omega}{p_\eta})).
\]

In order to bound \( \sigma_4 \), note that the event \( \{g(Y') \neq \tilde{g}(Y)\} \) is contained in the union of the events
\[
E_1 = \{ Y_{k,l} = 1 \text{ for some } 0 \leq k < \gamma(n \lor m) \text{ and } l = 0 \text{ or } 1 \}
\]
and
\[
E_2 = \{ Y_{k,0} = 0 \text{ for each } \gamma(n \lor m) \leq k \leq L \}.
\]

By (2.13),
\[
P(E_1) = O(\gamma(n \lor m)e^{-\psi(n \land m)}).
\]

Since \( Y \) is an independent Bernoulli process with \( P[Y_{k,0} = 1] = p_\eta \) for \( k \geq 0 \),
\[
P(E_2) = O \left( (1 - p_\eta)^{L - \gamma(n \lor m)} \right).
\]

Since \( x \geq \log(1 + x) \) for \( x > -1 \),
\[
(1 - p_\eta)^{L - \gamma(n \lor m)} = \exp \left( (\gamma(n \lor m) - L) \log \left( 1 + \frac{p_\eta}{1 - p_\eta} \right) \right) \leq \exp \left( (\gamma(n \lor m) - L) p_\eta \frac{1}{1 - p_\eta} \right).
\]

Now by (4.2) we get \( P(E_2) = O(e^{-t}) = O(\epsilon) \) which gives
\[
\sigma_4 = O \left( \gamma(n \lor m)e^{-\psi(n \land m) + \epsilon} \right).
\]

Next, observe that \( \tilde{g}(Y) \) has by lemma 3.1 the geometric distribution with the parameter \( p_\eta(p_\omega + p_\eta - p_\omega p_\eta)^{-1} < p_\eta(p_\eta + p_\omega)^{-1}. \) Hence, in the same way as in (3.29) we obtain
\[
\sigma_5 \leq d_{TV}(\mathcal{L}(\tilde{g}(Y)), \mathcal{L}(\text{Geo}(\frac{p_\eta}{p_\eta + p_\omega}))) \leq 2p_\omega \leq 2e^{-\psi n}.
\]
Combining all of our bounds we obtain,
\[
\sum_{j=1}^{5} \sigma_j = O\left(te^{-\nu_k/2}(n \vee m) + t \left(1 + \frac{p_\omega}{p_\eta}\right)(n \vee m) \cdot e^{-\nu_k} + t\psi(m)(1 + \frac{p_\omega}{p_\eta}) + \epsilon \right).
\]

Recall that
\[
\epsilon = \max\{e^{-\nu_k}, \psi(m)\}
\]
and \(t = \epsilon^{-1/2}\), hence
\[
\sum_{j=1}^{5} \sigma_j = O\left(e^{-\nu_k/2}(n \vee m) + \left(1 + \frac{p_\omega}{p_\eta}\right)(n \vee m) \cdot e^{-\nu_k/2} + \psi(m)^{1/2}(1 + \frac{p_\omega}{p_\eta})\right).
\]

Now since \(S\) from (4.3) is an arbitrary subset of \(\mathbb{N}\) the theorem follows. 

References

[1] Abadi M and Vergne N 2008 Sharp errors for point-wise Poisson approximations in mixing processes Nonlinearity 21 2871–85
[2] Abadi M and Saussol B 2011 Hitting and returning into rare events for all alpha-mixing processes Stoch. Process. Appl. 121 314–23
[3] Arratia R, Goldstein L and Gordon L 1989 Two moments suffice for Poisson approximations: the Chen–Stein method Ann. Probab. 17 9–25
[4] Bowen R 1975 Equilibrium States, the Ergodic Theory of Anosov Diffeomorphisms (Lecture Notes in Mathematics vol 470) (Berlin: Springer)
[5] Bradley R C 2007 Introduction to Strong Mixing Conditions (Heber City: Kendrick Press)
[6] Demers M, Wright P and Young L-S 2012 Entropy, Lyapunov exponents and escape rates in open systems Ergod. Theor. Dynam. Syst. 30 1270–301
[7] Heinrich L 1996 Mixing properties and central limit theorem for a class of non-identical piecewise monotonic \(C^2\)-transformations Math. Nachr. 181 185–214
[8] Haydn N T A and Psiloyenis Y 2014 Return times distribution for Markov towers with decay of correlations Nonlinearity 27 1323–49
[9] Kifer Y 2013 Nonconventional poisson limit theorems Isr. J. Math. 195 373–92
[10] Kifer Y and Rapaport A 2014 Poisson and compound poisson approximations in conventional and nonconventional setups Probab. Theory Relat. Fields 160 797–831
[11] Petersen K 1983 Ergodic Theory (Cambridge: Cambridge University Press)