Local alignment of Markov chains

Niels Richard Hansen

University of Copenhagen

We consider local alignments without gaps of two independent Markov chains from a finite alphabet, and we derive sufficient conditions for the number of essentially different local alignments with a score exceeding a high threshold to be asymptotically Poisson distributed. From the Poisson approximation a Gumbel approximation of the maximal local alignment score is obtained. The results extend those obtained by Dembo et al. (1994b) for independent sequences of iid variables.

1 Introduction

Local alignment of two biological sequences (DNA-molecules or proteins) is one of the most important and used tools in modern molecular biology for locating highly similar contiguous parts of the sequences. High similarity is usually interpreted as an evolutionary or functional relationship between the molecules. We show how the distribution of local alignment similarity scores behaves asymptotically when aligning independent Markov chains.

It is important to understand the distribution of local alignment scores for accessing the significance of e.g. the maximal scoring local alignment. Formally this is a test of the null hypothesis that two sequences are independent Markov chains against a somewhat unspecified alternative that they are not independent. The test statistic considered is the maximal local similarity score.

Usually when considering local alignments we are not only interested in the maximal scoring local alignment but also other essentially different local alignments that reach a score above a given threshold. It is therefore useful also to know the distribution of the number of local alignments of independent Markov chains that reach a score above a given threshold. In fact, it is this problem that we handle in the first place and the obtained Poisson approximation can easily be turned into a Gumbel approximation of the distribution of the maximal local alignment score.

The kind of local alignment we consider is gapless local alignment meaning that we search for (contiguous) parts of the two sequences that attain a high similarity when matched letter by letter. Similarity is measured by adding up a score for each pair.

AMS 1991 subject classification. Primary 60G70; secondary 60F10.

Key words and phrases. Chen-Stein method, extreme value theory, large deviations, local alignment, Markov additive processes, Markov chains, Poisson approximation.
of matched letters. In practice it is common to allow for the insertion of gaps in the sequences – each gap adding a suitably penalty to the similarity score – which usually increases the power of the test. The introduction of gaps does, however, make the problem of understanding the distribution of local alignment scores substantially more complicated although some recent developments for aligning independent iid sequences are promising (Siegmund & Yakir 2000), (Siegmund & Yakir 2003), and (Yakir & Grossmann 2001).

The main result is stated as Theorem 3.1. It says that if the expected similarity score under the null hypothesis is negative then there exist constants $\theta^*, K^* > 0$ such that if we let $s$ denote the maximal local alignment score obtained when aligning two independent Markov chains of length $n$ then the normalised score defined by

$$s' = \theta^* s - \log(K^* n^2)$$

is approximately a Gumbel distribution for $n \to \infty$. Moreover, the number of normalised local alignment scores exceeding the threshold $x$ is approximately Poisson distributed with mean $\exp(-x)$ for $n \to \infty$. We have ignored some details and there are certain regularity assumptions that need to be fulfilled for this to be a mathematically rigorous statement. We refer to Theorem 3.1 and its prerequisites.

It should be mentioned that the results are the expected generalisations of those obtained by Dembo et al. (1994b) for independent iid sequences, but the techniques of proof are not straightforward generalisations. Indeed, this author would like to emphasise the novelty of certain techniques developed in this paper. In particular the results achieved in Section 6.4 and Section 6.5 may be of independent interest. Moreover, the framework of Markov chains does not only provide a change of the null hypothesis but it also opens up for the possibility of choosing new types of score functions as we discuss in Remark 3.6. This can increase the power of the test. In addition, the results obtained in this paper cover many more null hypotheses than just the aligned sequences being Markov chains. Also higher order Markov dependency and hidden Markov models are covered by expanding the state space suitably.

In the present paper we develop most of the theory in sufficient generality to deal with other related problems. In particular, we present in the companion paper (Hansen 2004) a result of a conceptually different nature, but whose proof is along the very same lines as the proof presented in this paper.

2 Local gapless alignment

Let $(X_k)_{k \geq 1}$ and $(Y_k)_{k \geq 1}$ be two sequences of random variables taking values in a finite set $E$. We compare parts of one sequence with parts of the other using a score
function $f : E \times E \to \mathbb{Z}$, and we define the random variables

$$S_{i,j}^\Delta = \sum_{k=1}^\Delta f(X_{i+k}, Y_{j+k}),$$

for $i, j, \Delta \geq 0$. The variable $S_{i,j}^\Delta$ scores the local comparison of the sequence part $X_{i+1} \ldots X_{i+\Delta}$ with the sequence part $Y_{j+1} \ldots Y_{j+\Delta}$.

We make the assumption that $f$ takes integer values to emphasise the lattice nature of $f$ that is often met in practice. In this way our main results become more directly applicable.

**Remark 2.1.** The score function is often regarded as an $E \times E$ matrix. The matrix representation is convenient when writing down the values $f(x, y)$, but we will generally find it more useful to regard $f$ as an element in a vector space. Probability measures will then be regarded as elements in the dual space and we use the functional notation

$$\nu(f) = \sum_{x,y} f(x,y) \nu(x,y)$$

to denote the mean of $f$ evaluated under the probability measure $\nu$.

For $n \geq 1$ define

$$\mathcal{H}_n = \{(i, j, \Delta) \mid 0 \leq i \leq i + \Delta \leq n, 0 \leq j \leq j + \Delta \leq n\}.$$ 

The elements $(i, j, \Delta) \in \mathcal{H}_n$ are called alignments.

We want to understand the distribution of the collection

$$(S_{i,j}^\Delta)_{(i,j,\Delta) \in \mathcal{H}_n}$$

of local scores over all alignments. We will in particular be interested in the distribution of

$$\mathcal{M}_n = \max_{(i,j,\Delta) \in \mathcal{H}_n} S_{i,j}^\Delta,$$

the maximal local score over the set of alignments. We will also study the number, $C(t)$, say, of *essentially different* variables $S_{i,j}^\Delta$ exceeding some threshold $t \geq 0$. We will define essentially different precisely below.

The family of local scores are efficiently summarised in the matrix $(T_{i,j})_{0 \leq i,j \leq n}$ defined as follows. For $i = 0$ or $j = 0$ let $T_{i,j} = 0$ and define recursively

$$T_{i,j} = \max\{T_{i-1,j-1} + f(X_i, Y_j), 0\}$$

(3)
for \( i,j \geq 1 \). As we will show, cf. Remark 3.7 below, the maximum \( \mathcal{M}_n \) can be computed as
\[
\mathcal{M}_n = \max_{i,j} T_{i,j}.
\] (4)
This fact is the idea in the Smith-Waterman algorithm for computing the maximal local alignment score efficiently (Waterman 1995).

**Definition 2.2.** An alignment \((i,j,\Delta) \in \mathcal{H}_n\) is called an excursion if
\[
T_{i,j} = 0, \quad S^\delta_{i,j} > 0 \text{ for } 0 < \delta < \Delta
\]
and either \(S^\Delta_{i,j} \leq 0\), \(i + \Delta = n\) or \(j + \Delta = n\).

Let \( \mathcal{E}_n \) denote the set of all excursions.

Note that \( \mathcal{E}_n \) is a stochastic subset of \( \mathcal{H}_n \). It follows from the definition of the score matrix \((T_{i,j})\) and the definition of an excursion that if \((i,j,\Delta) \in \mathcal{E}_n\) and \(0 < \delta < \Delta\) then
\[
T_{i+\delta,j+\delta} = S^\delta_{i,j}.
\]

An excursion corresponds to a diagonal strip in the score matrix, for which the score starts at zero and then stays strictly positive along that diagonal strip until it either reaches zero or the indices hit the boundary of the score matrix.

The maximum over an excursion \(e = (i,j,\Delta) \in \mathcal{E}_n\) is denoted by
\[
\mathcal{M}_e = \max_{0 < \delta \leq \Delta} S^\delta_{i,j} = \max_{0 < \delta \leq \Delta} T_{i+\delta,j+\delta}.
\] (5)

**Definition 2.3.** The number of essentially different excesses over \(t\) is defined as
\[
C(t) = \sum_{e \in \mathcal{E}_n} 1(\mathcal{M}_e > t).
\] (6)

From (4) it follows that \((C(t) = 0) = (\mathcal{M}_n \leq t)\).

### 3 Alignment of independent Markov chains

Assume that the stochastic processes \((X_k)_{k \geq 1}\) and \((Y_k)_{k \geq 1}\) are independent Markov chains with transition probabilities \(P\) and \(Q\) respectively. Assume that \(P\) and \(Q\) are irreducible and aperiodic matrices with invariant left probability vectors \(\pi_P\) and \(\pi_Q\) respectively. Let \(\pi = \pi_P \otimes \pi_Q\). With
\[
\mu = \pi(f) = \sum_{x,y \in \mathcal{E}} f(x,y)\pi_P(x)\pi_Q(y)
\]
the (invariant) mean of \( f(X_1, Y_1) \) we will assume throughout that \( \mu < 0 \).

In the following, a cycle w.r.t. a matrix of transition probabilities \( P \) is a finite sequence \( x_1, \ldots, x_n \) such that

\[
P(x_i, x_{i+1} \pmod n) > 0
\]

for \( i = 1, \ldots, n \). We will assume that the following regularity conditions on \( f, P \) and \( Q \) are fulfilled: There exist cycles \( x_1, \ldots, x_n \) (w.r.t. \( P \)) and \( y_1, \ldots, y_n \) (w.r.t. \( Q \)) such that

\[
\sum_{k=1}^{n} f(x_k, y_k) > 0.
\]  

(7)

For any \( T \geq 1 \) there exist cycles \( x_1, \ldots, x_n \) (w.r.t. \( P \)) and \( y_1, \ldots, y_n \) (w.r.t. \( Q \)) such that

\[
\sum_{k=1}^{n} f(x_k, y_k) \neq \sum_{k=1}^{n} f(x_k, y_{k+T} \pmod n).
\]  

(8)

See Remark 3.3 and Remark 3.4 below for comments related to this somewhat strange looking condition.

For convenience we will assume that both Markov chains are stationary, though the results obtained hold even if they are not stationary. We denote by \( \mathbb{P}_\pi \) the probability measure under which \((X_n, Y_n)_{n \geq 1}\) is a stationary Markov chain with transition probabilities \( P \otimes Q \). For certain arguments it is in addition convenient to assume the existence of auxiliary stochastic variables \( X_0 \) and \( Y_0 \) such that \((X_n, Y_n)_{n \geq 0}\) is a stationary Markov chain too.

We define for \( \theta \in \mathbb{R} \) an \( E^2 \times E^2 \) matrix \( \Phi(\theta) \) with positive entries by

\[
\Phi(\theta)(x,y),(x',y') := \exp(\theta f(x', y')) P_{x,x'} Q_{y,y'},
\]

and we let \( \varphi(\theta) \) denote the spectral radius (the Perron-Frobenius eigenvalue) of this matrix. Then \( \varphi \) is a convex \( C^\infty \)-function in \( \theta \) and due to (7), \( \varphi(\theta) \rightarrow \infty \) for \( \theta \rightarrow \infty \). Furthermore, it holds that

\[
\partial_\theta \varphi(0) = \mu,
\]  

(9)

hence if \( \mu < 0 \) there exists a (by convexity unique) solution \( \theta^* > 0 \) to the equation \( \varphi(\theta) = 0 \).

In addition to the \( \Phi \)-matrix we introduce two \( E^3 \times E^3 \) matrices, \( \Phi_1(\theta) \) and \( \Phi_2(\theta) \), defined by

\[
\Phi_1(\theta)(x,y,z),(x',y',z') = \exp(\theta f(x', y') + \theta f(x', z')) P_{x,x'} Q_{y,y'} Q_{z,z'}
\]

\[
\Phi_2(\theta)(x,w,y),(x',w',y') = \exp(\theta f(x', y') + \theta f(w', y')) P_{x,x'} P_{w,w'} Q_{y,y'},
\]

and let \( \varphi_1(\theta) \) and \( \varphi_2(\theta) \) denote the corresponding spectral radii.
We define a process \( S_0 = 0 \) and for \( n \geq 1 \)
\[
S_n = \sum_{k=1}^{n} f(X_k, Y_k). \tag{10}
\]

In terms of this process a constant, \( K^* \), is defined by (1.26) in Theorem B in (Karlin & Dembo 1992). It is a non-trivial matter just to define \( K^* \) not to mention actually computing \( K^* \). We refer to (Karlin & Dembo 1992) for more details and in particular Section 5 for an algorithm for the computation of \( K^* \).

**Theorem 3.1.** Assume that \( \mu < 0 \), that the regularity conditions given by (7) and (8) are fulfilled, and that \( \theta^* \) and \( K^* \) are the constants defined above. Define for \( x \in \mathbb{R} \)
\[
t_n = \frac{\log K^* + \log n^2 + x}{\theta^*} \tag{11}
\]
and \( x_n \in [0, \theta^*) \) by \( x_n = \theta^*(t_n - \lfloor t_n \rfloor) \). Then if
\[
\varphi_1 \left( \frac{3\theta^*}{4} \right) < 1 \quad \text{and} \quad \varphi_2 \left( \frac{3\theta^*}{4} \right) < 1 \tag{12}
\]
it holds that
\[
||D(C(t_n)) - \text{Poi}(\exp(-x + x_n))|| \to 0 \tag{13}
\]
for \( n \to \infty \). In particular
\[
\mathbb{P}(M_n \leq t_n) - \exp(-\exp(-x + x_n)) \to 0 \tag{14}
\]
for \( n \to \infty \).

The theorem deserves a number of remarks.

**Remark 3.2.** The choice of \( x_n = \theta^*(t_n - \lfloor t_n \rfloor) \) assures that \( t_n - x_n/\theta^* = \lfloor t_n \rfloor \in \mathbb{Z} \).

Due to the lattice effect arising from \( f \) taking values in \( \mathbb{Z} \), it follows that
\[
(C(t_n) = m) = (C(t_n - x_n/\theta^*) = m)
\]
as well as
\[
(M_n \leq t_n) = (M_n \leq t_n - x_n/\theta^*),
\]
and this is the reason that we need to correct by \( x_n \) in the asymptotic formulas.
Remark 3.3. The regularity condition (8) doesn’t look particularly nice in general but is usually satisfied by quite trivial arguments. Essentially we want to avoid the situation where
\[ f(x, y) = f_1(x) + f_2(y) \] (15)
for two functions \( f_1, f_2 : E \to \mathbb{R} \). It is clear that if \( f \) is of the form (15) then (8) does not hold. It is easy to verify that if \( P \) and \( Q \) contain only strictly positive entries, condition (8) is equivalent to \( f \) not being of the form (15). In general, however, this author has not been able to prove that \( f \) not being of the form (15) is sufficient for (8) to hold. On the other hand, no counterexamples have been found either. In the proof we will explicitly need that (8) holds.

Remark 3.4. The fact that \( \varphi \) is convex goes back to Kingman (1961), but we will need strict convexity of \( \varphi \)-like functions. It can be shown (O’Cinneide 2000, Theorem 4) that if there exist cycles \( x_1, \ldots, x_n \) (w.r.t. \( P \)) and \( y_1, \ldots, y_n \) (w.r.t. \( Q \)) such that
\[ \sum_{i=1}^{n} f(x_i, y_i) \neq \mu \] (16)
then \( \varphi \) is strictly convex. In fact \( \log \varphi \) is strictly convex. This criterion for strict convexity, in a specific context in Section 6.5, is the reason behind (8). It should also be observed that \( \varphi \) as well as \( \varphi_1 \) and \( \varphi_2 \) are strictly convex by (7).

Remark 3.5. It is possible, and of practical relevans, to allow for the aligned sequences to have different lengths \( m \) and \( n \), say. In this case Theorem 3.1 holds for \( n, m \to \infty \) with
\[ t_{m,n} = \frac{\log K^* + \log(mn) + x}{\theta^*}. \]
Some restrictions on the simultaneous growth of \( m \) and \( n \) must be made in order for this to be true. It is sufficient that \( m \sim cn \) for some \( c > 0 \) but weaker assumptions will do.

Remark 3.6. For notational convenience Theorem 3.1 was stated and proved using a score function \( f \) that depends on a single pair of variables only. When aligning Markov chains it would be perfectly natural to use a score function that depends on pair-transitions instead, i.e. \( f : E^2 \times E^2 \to \mathbb{R} \) and
\[ S_{i,j}^\Delta = \sum_{k=1}^{\Delta} f(X_{i+k-1}, Y_{i+k-1}, X_{i+k}, Y_{i+k}). \]
With suitable modifications Theorem 3.1 holds for this kind of score function. For instance \( \Phi \) is defined as
\[ \Phi(\theta)(x, y, x', y') = \exp(\theta f(x, y, x', y')) P_{x', y'} Q_{y, y'}, \]
Φ₁ is defined as

\[ Φ₁(θ(x,y,z),(x',y',z')) = \exp(θf(x,y,x',y') + θf(x,z,x',z')) P_{x,x'}Q_{y,y'}Q_{z,z'}, \]

and Φ₂ is defined likewise.

In practice \( f \) is often chosen as a log-likelihood ratio. If the alternative to the null hypothesis for alignable sequences is that \( (X_n, Y_n)_{n≥1} \) is a Markov chain on \( E^2 \) governed by a \( E^2 \times E^2 \) matrix of transition probabilities \( R \), then we would choose

\[ f(x,y,x',y') = \log \frac{R_{x,y,x',y'}}{P_{x,x'}Q_{y,y'}}. \]

This score function does clearly not take integer values in general but choices such as \( \lfloor Nf \rfloor \) for suitably large \( N \) are common. If \( f \) is not a lattice function we don’t need the \( x_n \)-correction in the asymptotic formulas in Theorem 3.1.

We find that for this score function \( f \) and for \( θ = 1 \)

\[ Φ(1)(x,y) = \exp(f(x,y,x',y')) P_{x,x'}Q_{y,y'} = R_{x,y,x',y'}, \]

which has row sums equal to 1. Hence \( ϕ(1) = 1 \) implying that \( θ^* = 1 \).

**Remark 3.7.** The process \( (S_n)_{n≥0} \) defined by (10) is called a Markov controlled random walk or a Markov additive process (abbreviated MAP). The process \( (T_n)_{n≥0} \) defined by

\[ T_n = S_n - \min_{0≤k≤n} S_k \]

is called the reflection of the MAP \( (S_n)_{n≥0} \) at the zero barrier. It is straight forward to verify that \( (T_n)_{n≥0} \) satisfies the recursion

\[ T_n = \max\{T_{n-1} + f(X_n, Y_n), 0\} \]

for \( n ≥ 1 \). In addition

\[ \max_{1≤k≤m≤n} S_m - S_k = \max_{1≤m≤n} S_m - \min_{1≤k≤m} S_k = \max_{1≤m≤n} T_m. \]

Taking \( Δ = n \) and \( i = j = 0 \) we find that \( S_{0,0}^n = S_n \) and \( T_{n,n} = T_n \). Thus along the main diagonal in the score matrix \( (T_{i,j})_{0≤i,j≤n} \) we find the reflection of the MAP \( (S_n)_{n≥0} \). Along all other diagonals in the score matrix we find the reflections of MAPs too – these MAPs being defined by shifting the Markov chain \( (X_n)_{n≥1} \) along \( (Y_n)_{n≥1} \).

It follows from (18) that (4) indeed holds. Due to independence and stationarity
of the two Markov chains all the reflected MAPs along diagonals have the same
distribution, but they are dependent. The interpretation of Theorem 3.1 is that
asymptotically the number of excursions exceeding level $t_n$ has the same distribution
as if the reflected MAPs were independent. For this to be the case some assumptions
are necessary, cf. Example 3.8 below, and we found that (12) is sufficient.

**Example 3.8.** The Poisson approximation given in Theorem 3.1 in the framework
of independent sequences of iid variables was proved by Dembo et al. (1994b). When
comparing their result with the present theorem -- and especially the different conditions imposed -- some simplification of the setup in Theorem 3.1 is useful. Assume
in this example that $(X_n)_{n \geq 1}$ and $(Y_n)_{n \geq 1}$ are independent sequences of iid variables
with the $X$’s having distribution $\pi_1$ and the $Y$’s having distribution $\pi_2$. Then

$$\varphi(\theta) = \mathbb{E}(\exp(\theta f(X_1, Y_1)))$$

is the Laplace transform of $\pi = \pi_1 \otimes \pi_2$ and $\theta^* > 0$ solves $\varphi(\theta) = 1$. Also

$$\varphi_1(\theta) = \mathbb{E}(\exp(\theta f(X_1, Y_1) + \theta f(X_1, Y_2))) \quad \text{and}$$

$$\varphi_2(\theta) = \mathbb{E}(\exp(\theta f(X_1, Y_1) + \theta f(X_2, Y_1)))$$

are Laplace transforms. With $\pi^*$ the probability measure on $E \times E$ with point
probabilities $\pi^*(x, y) = \exp(\theta^* f(x, y))\pi_1(x)\pi_2(y)$, with $\pi_1^*$ and $\pi_2^*$ its marginals, and
with
\[ f_1(x) = \frac{1}{\theta^*} \log \sum_y \exp(\theta^* f(x, y)) \pi_2(y) \quad \text{and} \]
\[ f_2(y) = \frac{1}{\theta^*} \log \sum_x \exp(\theta^* f(x, y)) \pi_1(x), \]

Dembo et al. (1994b) find the condition (called \((E')\)) that if
\[ \pi^*(f) > 2 \max\{ \pi_1^*(f_1), \pi_2^*(f_2) \} \tag{19} \]
then the Poisson approximation (13) holds. This condition can be understood as a condition on relative entropies. It is unfortunately not clear how condition (19) is related directly to condition (12). As a simple example let \(E = \{0, 1\}\) and consider the symmetric score function
\[ f_a = \begin{pmatrix} 1 \\ -2 \end{pmatrix} \]
and the asymmetric score function
\[ f_b = \begin{pmatrix} 1 \\ -4 \end{pmatrix}. \]

Figure 1 shows the set of \((\pi_1, \pi_2)\) for which condition (12) as well as (19) are fulfilled. The figure shows that (12) is stronger than (19). We can also read off from the figure which measures \(\pi_1\) and \(\pi_2\) that imply \(\mu < 0\) for either score function. Turning to Dembo et al. (1994a), who shows that
\[ \frac{M_n}{\log n^2} \xrightarrow{a.s.} \frac{1}{\theta^*} \]
if and only if (19) is fulfilled with \(\geq\), we find that condition (19) must be very close to optimal. Hence the present condition must be stronger. The two examples above confirm that, but a direct proof of this has not been found.

4 The counting construction

The goal of this paper is to prove that \(C(t)\) is approximately Poisson distributed, but a direct proof does not seem feasible. The size and ‘shape’ of \(\mathcal{E}_n\) depend on the concrete realisation of the underlying stochastic variables and it is not clear how to handle the dependencies between the indicators \(1(M_e > t)\). Instead we
Figure 2: The index set is divided into vertical strips of width $l$ – on the figure $l = 3$. Along diagonals in each strip we look for local scores exceeding the threshold. On the figure two of these diagonals-within-a-strip at position $(2, l)$ and $(k, 2l)$ respectively are shown.

We define for $l > 0$ an index set

$$ I = \{(k, ql) \mid k \in \{0, \ldots, n\}, \; q \in \left\{0, \ldots, \left\lfloor \frac{n}{l} \right\rfloor \right\} \}. $$

One should think of this as a division of the score matrix into vertical strips of width $l$ with $I$ as an indexation of diagonals-within-a-strip, i.e. an index $a = (k, r) \in I$ represents the diagonal $(k + 1, r + 1), \ldots, (k + l, r + l)$, cf. Figure 2.

We will approximate the number of excursions exceeding $t$ by the number of diagonals-within-a-strip containing excursions exceeding $t$. To be precise, we will consider, for $a = (k, r) \in I$ and $t > 0$, the variable

$$ V_a = V_a(t) = 1 \left( \max_{1 \leq \delta \leq \Delta} \sum_{h=\delta}^{\Delta} f(X_{k+h}, Y_{r+h}) > t \right). $$
We shown in Section 6.7 that in the setup of the present paper
\[
P \left( \sum_{a \in I} V_a(t_n) \neq C(t_n) \right) \to 0 \tag{20}
\]
when \( n \to \infty \).

To prove that \( \sum_{a \in I} V_a \) is approximately Poisson distributed we apply a result obtained by Arratia et al. (1989) based on the Chen-Stein method. The counting construction presented will also be used in the companion paper (Hansen 2004) in a slightly different setup. Therefore we choose to formulate the following result for a subset \( I_0 \subseteq I \) of the index set \( I \). A formal proof of the result stated in Remark 3.5 can also be given based in using \( I_0 \subseteq I \) defined as
\[
I_0 = \left\{ (k, r) = (k, q l) \mid k \in \{0, \ldots, m\}, q \in \{0, \ldots, \left\lfloor \frac{n}{l} \right\rfloor \} \right\},
\]
for \( m \leq n \).

We assume that for all \( a \in I_0 \) a subset \( B_a \subseteq I_0 \) given. This set \( B_a \) is called the neighbourhood of strong dependence of \( V_a \) and it may take various shapes in practice. In Section 6.7 we make a concrete choice of \( B_a \), whereas in the companion paper (Hansen 2004) another choice of \( B_a \) is needed. Furthermore, for \( a \in I_0 \) let
\[
\mathcal{F}_a = \sigma(V_b \mid b \notin B_a)
\]
denote the \( \sigma \)-algebra generated by those variables \( V_b \) not in the neighbourhood of strong dependence of \( V_a \).

Rephrasing Theorem 1 in Arratia et al. (1989) gives:

**Theorem 4.1.** Suppose that \( l = l_n \) and \( t = t_n \) are chosen such that for some sequence \( (\lambda_n)_{n \geq 1} \)
\[
\beta_{1,n} = \left| \sum_{a \in I_0} \mathbb{E}(V_a) - \lambda_n \right| \to 0, \tag{21}
\]
for \( n \to \infty \), and suppose that
\[
\beta_{2,n} = \sum_{a \in I_0, b \in B_a} \mathbb{E}(V_a)\mathbb{E}(V_b) \to 0, \tag{22}
\]
\[
\beta_{3,n} = \sum_{a \in I_0, b \in B_a, b \neq a} \mathbb{E}(V_a V_b) \to 0, \tag{23}
\]
\[
\beta_{4,n} = \sum_{a \in I_0} \mathbb{E}[\mathbb{E}(V_a | \mathcal{F}_a) - \mathbb{E}(V_a)] \to 0, \tag{24}
\]
for $n \to \infty$, then

$$\left\| D \left( \sum_{a \in I_0} V_a \right) - \text{Poi}(\lambda_n) \right\| \to 0. \quad (25)$$

In fact, the bound

$$\left\| D \left( \sum_{a \in I_0} V_a \right) - \text{Poi}(\lambda_n) \right\| \leq \beta_{1,n} + 2(\beta_{2,n} + \beta_{3,n} + \beta_{4,n})$$

always hold.

As a direct consequence, using the coupling inequality, we obtain the following corollary.

**Corollary 4.2.** If (25) holds and (20) is fulfilled too, then

$$||D(C(t_n)) - \text{Poi}(\lambda_n)|| \to 0 \quad (26)$$

and

$$\mathbb{P}(M_n \leq t_n) - \exp(-\lambda_n) \to 0. \quad (27)$$

## 5 Useful mixing inequalities

When the aligned sequences are iid the set $B_a$ is usually chosen such that $V_a$ and $\mathcal{F}_a$ are independent, in which case $\mathbb{E} | \mathbb{E}(V_a | \mathcal{F}_a) - \mathbb{E}(V_a) | = 0$ and the term $\beta_{4,n}$ vanish. In the framework of Markov chains we need to control $\beta_{4,n}$ by using exponential $\beta$-mixing of stationary, finite state space Markov chains. To this end we need a few results on how to translate knowledge about the $\beta$-mixing coefficients into useful bounds on $\mathbb{E} | \mathbb{E}(V_a | \mathcal{F}_a) - \mathbb{E}(V_a) |$.

For two $\sigma$-algebras $\mathcal{F}$ and $\mathcal{G}$ the $\alpha$-mixing measure of dependence is

$$\alpha(\mathcal{F}, \mathcal{G}) = \sup_{A \in \mathcal{F}, B \in \mathcal{G}} | \mathbb{P}(A \cap B) - \mathbb{P}(A) \mathbb{P}(B) |.$$

The following lemma relates $\alpha$-mixing measures to mean values of the desired form.

**Lemma 5.1.** Let $\mathcal{F}$ and $\mathcal{G}$ be $\sigma$-algebras and let $A \in \mathcal{G}$. With $\eta = 1(A)$

$$\mathbb{E} | \mathbb{E}(\eta | \mathcal{F}) - \mathbb{E}(\eta) | \leq 2\alpha(\mathcal{F}, \mathcal{G}). \quad (28)$$
Proof: With $B = (\mathbb{E}(\eta|\mathcal{F}) \geq \mathbb{E}(\eta)) \in \mathcal{F}$ and $\xi = 1(B)$ we see that
\[
\mathbb{E}[\mathbb{E}(\eta|\mathcal{F}) - \mathbb{E}(\eta)] = \mathbb{E}(\xi(\mathbb{E}(\eta|\mathcal{F}) - \mathbb{E}(\eta))) - \mathbb{E}((1 - \xi)(\mathbb{E}(\eta|\mathcal{F}) - \mathbb{E}(\eta))) \\
= 2(\mathbb{E}(\xi\eta) - \mathbb{E}(\xi)\mathbb{E}(\eta)) \\
= 2(\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)) \leq 2\alpha(\mathcal{F}, \mathcal{G}).
\]

The $\beta$-mixing measure of dependence between the $\sigma$-algebras $\mathcal{F}$ and $\mathcal{G}$ is defined as
\[
\beta(\mathcal{F}, \mathcal{G}) = \mathbb{E}\left(\sup_{A \in \mathcal{F}}|\mathbb{P}(A|\mathcal{G}) - \mathbb{P}(A)|\right).
\]

For a stationary stochastic process $(Z_n)_{n \in \mathbb{Z}}$ taking values in some set $E$ and for a subset $I \subseteq \mathbb{Z}$, we define the corresponding $\sigma$-algebra $\mathcal{F}_I = \sigma(Z_n; n \in I)$. The $\beta$-mixing coefficient is defined as
\[
\beta(n) = \beta(\mathcal{F}_{[n, \infty)}, \mathcal{F}_{(-\infty, 0]}) = \mathbb{E}\left(\sup_{A \in \mathcal{F}_{[n, \infty)}}|\mathbb{P}(A|\mathcal{F}_{(-\infty, 0]}) - \mathbb{P}(A)|\right),
\]
for $n \geq 1$ and the process $(Z_n)_{n \in \mathbb{Z}}$ is called $\beta$-mixing if $\beta(n) \to 0$ for $n \to \infty$. For two subsets $I, J \subseteq \mathbb{Z}$, the distance, $d(I, J)$, between the sets is defined as
\[
d(I, J) = \inf_{n \in I, m \in J} |n - m|.
\]

We call $I \subseteq \mathbb{Z}$ an interval if $I$ is either empty, if $I = \{n, n+1, \ldots, m-1, m\}$ for some $n \leq m \in \mathbb{Z}$, in which case we write $I = [n, m]$, or if $I = \{\ldots, n-1, n\} = (-\infty, n]$ or $I = \{n, n+1, \ldots\} = [n, \infty)$. If $I, J \subseteq \mathbb{Z}$ are two subsets of integers, we write $I < J$ if $n < m$ for all $n \in I$ and $m \in J$.

**Lemma 5.2.** Assume that $I_1 < I_2 < \ldots < I_\kappa$ is an increasing sequence of intervals in $\mathbb{Z}$ with $\kappa \geq 2$. With $I = \bigcup_{i \text{ odd}} I_i$ and $J = \bigcup_{i \text{ even}} I_i$ it holds that
\[
\alpha(\mathcal{F}_I, \mathcal{F}_J) \leq (2\kappa - 3)\beta(d(I, J)).
\]

This result is Theorem 3.1 in (Takahata 1981) for $\kappa = 3$. An extension of his proof gives Lemma 5.2. In (Doukhan 1994, Theorem 1.3.3) a similar generalisation is provided with a sketch of proof, although the constant $(2\kappa - 3)$ appearing in Lemma 5.2 is better than the corresponding constant $3(\kappa - 1)$ appearing in (Doukhan 1994). We note that in this paper we will only need Lemma 5.2 for $\kappa = 3$. In (Hansen 2004) we will need the lemma for $\kappa = 5$. 
6 Proofs

The proof of Theorem 3.1 is divided into a number of lemmas. We need to verify the conditions in Theorem 4.1, and to this end we need bounds on the expectations $E(V_a V_b) = \mathbb{P}(V_a = 1, V_b = 1)$ for $b \in B_a$ and $a \neq b$. This is the subject of the following subsections and clearly the most difficult part of the proof. In Section 6.7 we collect the bounds obtained to prove that the conditions of Theorem 4.1 are fulfilled when aligning independent Markov chains under the assumptions given in Theorem 3.1. Finally we prove that (20) holds and the Poisson approximation of $\sum_{a \in I} V_a(t_n)$ can be translated into a Poisson approximation of $C(t_n)$.

For $a = (k, r), b = (i, j) \in I$ we always have that

$$E(V_a V_b) = \mathbb{P}\left( \max_{1 \leq h \leq \Delta} \sum_{h=\delta}^{\Delta} f(X_{i+h}, Y_{j+h}) > t, \max_{1 \leq h \leq \Delta} \sum_{h=\delta}^{\Delta} f(X_{k+h}, Y_{r+h}) > t \right) \leq l^4 \max_{\delta_1, \delta_2, \Delta_1, \Delta_2} \mathbb{P}\left( \sum_{h=\delta_1}^{\Delta_1} f(X_{k+h}, Y_{r+h}) > t, \sum_{h=\delta_2}^{\Delta_2} f(X_{i+h}, Y_{j+h}) > t \right).$$

(30)

To bound $E(V_a V_b)$ we thus need to bound the probability on the right hand side above. The same $X$- and $Y$- variables may enter both of the sums in two essentially different ways. Either there are variables from both sequences entering both sums or only variables from one sequence entering both sums. These two different cases need different treatment. To give an exhaustive treatment of the different ways that such a sharing of variables can be arranged becomes unnecessarily complicated, so we choose to treat the two essentially different cases for a specific arrangement of the sharing of variables in sufficient details for the reader to be able to convince himself that all other arrangements can be treated similarly.

6.1 Positive functionals of a Markov chain

We make a useful and general observation about bounding the expectation of positive functionals, e.g. probabilities, of a Markov chain. It allows us to assume parts of the same Markov chain to be independent, stationary versions at the expense of a constant factor.

**Lemma 6.1.** Let $Z = (Z_k)_{k \geq 0}$ be an irreducible Markov chain on a finite state space $E$ and let $0 = k_1 < \cdots < k_N < \infty$ be given. Then there exists a constant $\rho_N$ such that if $(Z^i_k)_{k=k_i}$ for $i = 1, \ldots, N$ ($k_{N+1} = \infty$) are $N$ independent stationary Markov chains with the same transition probabilities as $Z$, and $\tilde{Z} = (\tilde{Z}_k)_{k \geq 0}$ is given by $\tilde{Z}_k = Z^i_k$ if $k_i \leq k < k_{i+1}$ then for a positive functional $\Lambda: E^{\mathbb{N}_0} \to [0, \infty)$. 

\[ \Lambda : E^{\mathbb{N}_0} \to [0, \infty) \]
it holds that
\[ \mathbb{E}(\Lambda(Z)) \leq \rho_N \mathbb{E}(\Lambda(\tilde{Z})). \]  
(31)

The constant \( \rho_N \) does not depend on the actual initial distribution of \( Z \) nor on the functional \( \Lambda \).

**Proof:** Assume \( N = 2 \). The general result follows by induction. Assume first that \( Z \) is stationary and that \( (Z_1^1)_{k=0}^k \) and \( (Z_1^2)_{k \geq 2} \) are independent and stationary. Then \( Z \) has the same distribution as \( \tilde{Z} \) conditionally on \( Z_{k_2}^1 = Z_{k_2}^2 \), hence using that \( \Lambda \) is a positive functional
\[ \mathbb{E}(\Lambda(Z)) = \frac{\mathbb{E}(\Lambda(\tilde{Z}); Z_{k_2}^1 = Z_{k_2}^2)}{\mathbb{P}(Z_{k_2}^1 = Z_{k_2}^2)} \leq \rho \mathbb{E}(\Lambda(\tilde{Z})). \]

with \( \rho = (\sum_{x \in E} \pi_x^2)^{-1} \), where \( \pi \) is the invariant distribution.

If \( Z \) is non-stationary with initial distribution \( \nu \), say, we have that
\[ \mathbb{E}_\nu(\Lambda(Z)) = \sum_{x \in E} \frac{\nu_x \pi_x}{\pi_x} \mathbb{E}_x(\Lambda(Z)) \leq \frac{1}{\min_x \pi_x} \mathbb{E}_\pi(\Lambda(Z)). \]

So \( \rho_2 = \rho / \min_x \pi_x \) will do. In general \( \rho_N = \rho^{N-1} / \min_x \pi_x \) can be used. \( \Box \)

### 6.2 Exponential change of measure

Recall the definition of \( \Phi(\theta) \) and \( \varphi(\theta) \) and let \( r^\theta = (r^\theta(x,y)) \) denote the (right) eigenvector for \( \Phi(\theta) \) with eigenvalue \( \varphi(\theta) \). Due to irreducibility all the coordinates of \( r^\theta \) are strictly positive. We define a process \( (L^\theta_n)_{n \geq 0} \) for \( \theta \geq 0 \) by
\[ L^\theta_n = \frac{r^\theta(X_n, Y_n) \exp(\theta S_n)}{r^\theta(X_0, Y_0) \varphi(\theta)^n}. \]

Then with \( (\mathcal{F}_n)_{n \geq 0} \) the filtration of \( \sigma \)-algebras generated by the Markov chain \((X_n, Y_n)_{n \geq 0} \) it follows that
\[ \mathbb{E}(L_n \mid \mathcal{F}_{n-1}) = \frac{\exp(\theta S_{n-1})}{r^\theta(X_0, Y_0) \varphi(\theta)^n} \mathbb{E}(r^\theta(X_n, Y_n) \exp(\theta f(X_n, Y_n)) \mid (X_{n-1}, Y_{n-1})) \]
\[ = \frac{\exp(\theta S_{n-1})}{r^\theta(X_0, Y_0) \varphi(\theta)^n} (\Phi(\theta)^r(X_{n-1}, Y_{n-1})) \]
\[ = \frac{\exp(\theta S_{n-1})}{r^\theta(X_0, Y_0) \varphi(\theta)^n} \varphi(\theta) r^\theta(X_{n-1}, Y_{n-1}) = L^\theta_{n-1}. \]  
(32)
This shows that \((L^\theta_n, \mathcal{F}_n)_{n \geq 0}\) is a martingale, which satisfies that \(L^\theta_n > 0\) and \(\mathbb{E}(L_n(\theta)) = \mathbb{E}(L_0(\theta)) = 1\) for all \(\theta\). A probability measure \(\mathbb{P}^\theta_n\) on \(\mathcal{F}_n\) is then defined to have density \(L^\theta_n\) w.r.t. the restriction of \(\mathbb{P}\) to \(\mathcal{F}_n\). These measures can be extended to a single measure \(\mathbb{P}^\theta\), the exponentially changed or exponentially tilted measure, under which \((X_n, Y_n)_{n \geq 0}\) is a Markov chain with transition probabilities

\[
R^\theta(x, y; x', y') = \frac{r^\theta(x', y')}{r^\theta(x, x') \varphi(\theta) \Phi(\theta)} \Phi(x, y; x', y').
\]

One exponential changed measure of particular interest is given by \(\theta = \theta^*\) in which case we denote \(\mathbb{P}^{\theta^*}\) by \(\mathbb{P}^*\). The corresponding matrix of transition probabilities and the right eigenvector of \(\Phi(\theta^*)\) are likewise denoted \(R^*\) and \(r^*\) respectively.

Note that the exponential change of measure does not change the distribution \(\pi\) of \((X_0, Y_0)\) whereas the invariant measure for \(R^*\) typically differs from \(\pi\). Let \(\pi^*\) denote the invariant measure for \(R^*\) (which exists and is unique due to irreducibility of \(R\), which follows from irreducibility and aperiodicity of \(P\) and \(Q\)). The measure under which \((X_n, Y_n)_{n \geq 0}\) is a stationary Markov chain with transition probabilities \(R^*\) will be denoted \(\mathbb{P}^*_{\pi^*}\).

We should observe that since the eigenvector fraction is bounded below by a strictly positive constant (and trivially bounded above) then \(\mathbb{E}(L^\theta_n) = 1\) implies that

\[
\frac{1}{n} \log \mathbb{E}(\exp(\theta S_n)) \rightarrow \log \varphi(\theta)
\]

for \(n \rightarrow \infty\). The function \(\log \varphi(\theta)\) therefore serves as an asymptotic cumulant generating function for \(S_n\).

The exponential change of measure technique provides a number of useful inequalities. With \(a = (r, k)\) an index and

\[
A(a) = \{(r - \delta, k - \delta, \Delta) \in \mathcal{H}_n \mid 0 \leq \delta \leq \Delta\}
\]

the set of alignments ‘crossing’ the index the following inequality will turn out to be useful.

**Lemma 6.2.** For some constant \(K\)

\[
\mathbb{P} \left( \max_{(i, j, \Delta) \in A(a)} S^\Delta_{i,j} > t \right) \leq K \left\{ n^{-2} + (\log n)^2 \exp(-\theta^* t) \right\}.
\]

**Proof:** Observe that \(S^\Delta_{i,j} \overset{D}{=} S_{\Delta} \) so we find that for \(\theta \in (0, \theta^*)\)

\[
\mathbb{P} \left( S^\Delta_{i,j} > t \right) = \mathbb{P} \left( S_{\Delta} > t \right) = \mathbb{E}^\theta(L^\Delta_1; S_{\Delta} > t) \leq K_1 \varphi(\theta)^\Delta \exp(-\theta t)
\]
for some constant $K_1$. Then for any $\kappa > 0$

$$\mathbb{P}\left(\max_{(i,j,\Delta) \in A(a)} S_{i,j}^\Delta > t\right) = \mathbb{P}\left(\max_{(i,j,\Delta) \in A(a): \Delta > \kappa \log n} S_{i,j}^\Delta > t\right) + \mathbb{P}\left(\max_{(i,j,\Delta) \in A(a): \Delta \leq \kappa \log n} S_{i,j}^\Delta > t\right) \leq K_1 n^2 \varphi(\theta)^{\kappa \log n} + K_2 (\log n)^2 \exp(-\theta^* t),$$

since there are at most $n^2$ alignments crossing $a$ and at most $(\kappa \log n)^2$ of these that satisfy $\Delta \leq \kappa \log n$. With $\theta < \theta^*$ so that $\varphi(\theta) < 1$ and choosing $\kappa = -\frac{4}{4} - 1$ the result follows.

\[\square\]

### 6.3 Variables shared in one sequence

Let $r_i^\theta = (r_i^\theta(x, y, z))$ denote the right eigenvector for $\Phi_i(\theta)$ with eigenvalue $\varphi_i(\theta)$ for $i = 1, 2$ respectively. As above, due to irreducibility, all coordinates of these vectors are strictly positive.

In this section we derive a result corresponding to an overlap in the $X$-sequence only, and we thus work exclusively with the $\Phi_1$ matrix. Similar derivations for an overlap in the $Y$-sequence using $\Phi_2$ are possible.

Fix $i \leq j$ and $l \geq 1$ and define the functions

$$\sigma_1((x_k)_k, (y_k)_k) = \sum_{k=1}^i f(x_k, y_k)$$

$$\sigma_2((x_k)_k, (y_k)_k, (z_k)_k) = \sum_{k=i+1}^j f(x_k, y_k) + f(x_k, z_k)$$

$$\sigma_3((x_k)_k, (z_k)_k) = \sum_{k=j+1}^{i+l} f(x_k, z_k).$$

Note that the last sum is empty unless $i + l > j$.

For $\theta > 0$ define

$$\mathcal{L}^\theta = \frac{r_i^\theta(x, y) \exp(\theta \sigma_1)}{r_i^\theta(x, y) \varphi(\theta)^{i}} \times \frac{r_j^\theta(x, y, z) \exp(\theta \sigma_2)}{r_j^\theta(x, y, z) \varphi(\theta)^{j}} \times \frac{r_{i+l}^\theta(x, z) \exp(\theta \sigma_3)}{r_{i+l}^\theta(x, z) \varphi(\theta)^{i+l-j}},$$

likewise with the last factor absent unless $i + l > j$. Note that $\mathcal{L}^\theta : E^{n_0} \to [0, \infty)$ can be regarded as a function on the product space $E^{n_0}$.

Assume that $(Z_n)_{n \geq 1}$ is a stationary Markov chain with transition probabilities $Q$ independent of $(X_n, Y_n)_{n \geq 1}$. 
Lemma 6.3. It holds for all $\theta > 0$ that
\[
\mathbb{E}(L^\theta((X_k)_k, (Y_k)_k, (Z_k)_k)) = 1,
\]
and, furthermore, for $T + i \geq j$
\[
\mathbb{E}(L^\theta((X_k)_k, (Y_k)_k, (Y_{T+k})) \leq \rho
\]
for some $\rho$.

Proof: The first part of the lemma follows by repeating the arguments in (32) three times corresponding to making three different, successive exponential changes of measures. The second claim follows by Lemma 6.1.

We restrict our attention to the case where $T + i \geq j$, so that there is no overlap in the $Y$-sequence. We define three stochastic variables as follows
\[
S_1 = \sigma_1((X_k)_k, (Y_k)_k) = \sum_{k=1}^i f(X_k, Y_k),
\]
\[
S_2 = \sigma_2((X_k)_k, (Y_k)_{T+k}) = \sum_{k=i+1}^j f(X_k, Y_k) + f(X_k, Y_{T+k}),
\]
\[
S_3 = \sigma_3((Y_k)_k, (Y_k)_{T+k}) = \sum_{k=j+1}^{i+l} f(X_k, Y_{T+k}),
\]
together with $S = S_1 + S_2 + S_3$.

**Lemma 6.4.** If $\theta \in (0, \theta^*)$ and $\varphi_1(\theta) \leq 1$ there exists a constant $K$ such that
\[
\mathbb{P}(S > s) \leq K \exp \left( -\theta s \right).
\]

Proof: We have that
\[
\mathbb{P}(S > s) = \mathbb{E} \left( \frac{L^\theta((X_k)_k, (Y_k)_k, (Y_{T+k}))}{L^\theta((X_k)_k, (Y_k)_k, (Y_{T+k}))}; S > s \right).
\]
Note that $L^\theta = \gamma^\theta \exp(\theta S)$ with the $\gamma^\theta$-factor bounded below uniformly by $b$, say, since all the entries in the eigenvectors are strictly positive and since $\varphi(\theta) \leq 1$ by definition of $\theta^*$ together with $\varphi_1(\theta) \leq 1$ by assumption. Since we integrate over the set $(S > s)$, the exponential factor in $L^\theta$ can be bounded below by $\exp(\theta s)$. Hence the denominator is bounded below by $b \exp(\theta s)$. Using (36) in Lemma 6.3 we get that
\[
\mathbb{P}(S > s) \leq \rho b^{-1} \exp(-\theta s).
\]
Of course, if the overlap is in the $Y$ sequence instead Lemma 6.4 holds for those $\theta \in (0, \theta^*)$ satisfying $\varphi_2(\theta) \leq 1$. 

6.4 A uniform large deviation result

To handle the case with variables shared from both sequences we need a special large deviation result for Markov chains that we will derive in this section. We first state the useful Azuma-Hoeffding inequality for martingales with bounded increments, cf. Lemma 1.5 in (Ledoux & Talagrand 1991).

**Lemma 6.5.** If \((Z_m, \mathcal{F}_m)_{m \geq 0}\) is a mean zero martingale with \(Z_0 = 0\) such that for all \(m \geq 1\)

\[
|Z_m - Z_{m-1}| \leq c_m
\]

for some sequence \((c_m)_{m \geq 1}\), then for \(\lambda > 0\)

\[
\mathbb{P}(Z_m \geq \lambda) \leq \exp \left( -\frac{\lambda^2}{2\sum_{k=1}^{m} c_k^2} \right).
\]

Fix \(j \geq 1\) and let in this section \((X_k, Y_k)_{k \geq 1}\) be a stationary, aperiodic and irreducible Markov chain with transition probabilities given by \(R\) and invariant distribution \(\pi_R\). Let \((Y_k)_{k \geq j+1}\) be an independent, stationary, aperiodic and irreducible Markov chain with transition probabilities given by \(Q\) and invariant distribution \(\pi_Q\). For an \(E^2 \times E^2\)-matrix \(G\) define the norm of the matrix as

\[
||G||_\infty := \max_{(x,y)} \sum_{(z,w)} |G_{(x,y),(z,w)}|.
\]

With \(\mathbb{1}\) the column vector of ones, the matrix \(R^k\) converges to \(\mathbb{1}\pi_R\) due to irreducibility and aperiodicity, and since rate of convergence is sufficiently fast, in fact geometric, we have that

\[
\sum_{k=0}^{\infty} ||R^k - \mathbb{1}\pi_R||_\infty < \infty.
\]

For an \(E^2\) vector \(f\) we let \(||f||_\infty = \max_{(x,y)} |f(x,y)|\) denote the max-norm. Then clearly for any \(E^2 \times E^2\) matrix \(G\), with \(G(f)\) the matrix product of \(G\) with the vector \(f\), \(||G(f)||_\infty \leq ||f||_\infty ||G||_\infty\), and especially

\[
||R^k(f) - \mathbb{1}\pi_R(f)||_\infty \leq ||f||_\infty ||R^k - \mathbb{1}\pi_R||_\infty.
\]

For \(T \geq 1\) a fixed constant, we want to give an exponential bound of the probability

\[
\mathbb{P} \left( \sum_{k=1}^{j} f(X_k, Y_{k+T}) \geq \sum_{k=1}^{j} f(X_k, Y_k) \right)
\]
if $\mathbb{E}(f(X_k, Y_{k+T})) < \mathbb{E}(f(X_k, Y_k))$ all $k$. This is achieved by introducing a relevant martingale and then using the Azuma-Hoeffding inequality.

Let $\mathcal{F}_m = \sigma(X_1, Y_1, \ldots, X_m, Y_m)$ ($\mathcal{F}_0 = \emptyset, \Omega$),

$$S_{j,T} = \sum_{k=1}^{j} f(X_k, Y_{k+T}) - f(X_k, Y_k) \quad (S_{0,T} = 0),$$

and with $\xi_{j,T} = \mathbb{E}(S_{j,T})$ let

$$Z_m = \mathbb{E}(S_{j,T} - \xi_{j,T} | \mathcal{F}_m). \quad (39)$$

Then $(Z_m, \mathcal{F}_m)_{m=0}^{j}$ is a mean zero martingale with $Z_0 = 0$ (depending on $T$, though we have suppressed this in the notation). Notice that $Z_j = S_{j,T} - \xi_{j,T}$. The following lemma shows that the martingale differences

$$|Z_m - Z_{m-1}| = |\mathbb{E}(S_{j,T} | \mathcal{F}_m) - \mathbb{E}(S_{j,T} | \mathcal{F}_{m-1})|$$

are uniformly bounded by a constant.

**Lemma 6.6.** There exists a constant $\eta$ independent of $j$ and $T$ such that

$$|Z_m - Z_{m-1}| \leq \eta. \quad (40)$$

Here $\eta$ can be chosen as

$$\eta = 6 ||f||_{\infty} \sum_{k=0}^{\infty} ||R^k - \mathbb{I}_R||_{\infty}. \quad (41)$$

**Proof:** The Markov property gives that for $m \leq k \leq j$

$$\mathbb{E}(f(X_k, Y_k) | \mathcal{F}_m) = R^{k-m}(f)(X_m, Y_m),$$

and with $\bar{f}(x) = \sum_z f(x, z) \pi_Q(z)$ and $f_T(x, y) = R^T(f(x, \cdot))(x, y)$

$$\mathbb{E}(f(X_k, Y_{k+T}) | \mathcal{F}_m) = \begin{cases} R^{k-m}(f_T)(X_m, Y_m) & k \in C_1 \\ R^{k-m}(\bar{f})(X_m, Y_m) & k \in C_2 \\ R^{k+T-m}(f(X_k, \cdot))(X_m, Y_m) & k \in C_3 \\ \bar{f}(X_k) & k \in C_4 \end{cases},$$

where

$$C_1 = \{k \mid m \leq k < k + T \leq j\},$$
$$C_2 = \{k \mid m \leq k \leq j < k + T\},$$
$$C_3 = \{k \mid m - T \leq k < m \leq k + T \leq j\},$$
$$C_4 = \{k \mid m - T \leq k < m < j < k + T\}.$$
Observing that
\[ \mathbb{E}(S_{j,T}|\mathcal{F}_m) = \sum_{k=1}^{j} \mathbb{E}(f(X_k, Y_{k+T})|\mathcal{F}_m) - \sum_{k=1}^{j} \mathbb{E}(f(X_k, Y_k)|\mathcal{F}_m) \]

and subtracting \( \mathbb{E}(S_{j,T}|\mathcal{F}_{m-1}) \) from this, the martingale difference \( Z_m - Z_{m-1} \) is seen to be a sum of the following two terms

\[ t_1 = \sum_{k=m-T}^{j} \mathbb{E}(f(X_k, Y_{k+T})|\mathcal{F}_m) - \mathbb{E}(f(X_k, Y_{k+T})|\mathcal{F}_{m-1}) \]
\[ t_2 = \sum_{k=m}^{j} \mathbb{E}(f(X_k, Y_k)|\mathcal{F}_{m-1}) - \mathbb{E}(f(X_k, Y_k)|\mathcal{F}_m) \]

Since
\[ |\mathbb{E}(f(X_k, Y_k)|\mathcal{F}_m) - \pi_R(f)| = |R_{k-m}^k(f)(X_m, Y_m) - \pi_R(f)| \leq ||f||_{\infty} ||R_{k-m}^k - \mathbbm{1}\pi_R||_{\infty}, \]
the term \( t_2 \) is controlled by the following inequality
\[ |t_2| \leq 2||f||_{\infty} \sum_{k=m}^{j} ||R_{k-m}^k - \mathbbm{1}\pi_R||_{\infty} \leq 2||f||_{\infty} \sum_{k=0}^{\infty} ||R_{k}^k - \mathbbm{1}\pi_R||_{\infty}. \quad (42) \]

Noting that \( ||f_T||_{\infty}, ||\overline{f}||_{\infty}, ||f(x, \cdot)||_{\infty} \leq ||f||_{\infty} \) we observe that for \( k \in C_1 \)
\[ |\mathbb{E}(f(X_k, Y_{k+T})|\mathcal{F}_m) - \pi_R(f_T)| \leq ||f||_{\infty} ||R_{k-m}^k - \mathbbm{1}\pi_R||_{\infty}, \]
for \( k \in C_2 \)
\[ |\mathbb{E}(f(X_k, Y_{k+T})|\mathcal{F}_m) - \pi_R(\overline{f})| \leq ||f||_{\infty} ||R_{k-m}^k - \mathbbm{1}\pi_R||_{\infty}, \]
and for \( k \in C_3 \)
\[ |\mathbb{E}(f(X_k, Y_{k+T})|\mathcal{F}_m) - \pi_R(f(X_k, \cdot))| \leq ||f||_{\infty} ||R_{k+T-m}^{k+T} - \mathbbm{1}\pi_R||_{\infty}. \]

Since the three inequalities above also hold when conditioning on \( \mathcal{F}_{m-1} \), we obtain
\[ \sum_{k \in C_1 \cup C_2 \cup C_3} |\mathbb{E}(f(X_k, Y_{k+T})|\mathcal{F}_m) - \mathbb{E}(f(X_k, Y_{k+T})|\mathcal{F}_{m-1})| \]
\[ \leq 2||f||_{\infty} \sum_{k \in C_1 \cup C_2} ||R_{k-m}^k - \mathbbm{1}\pi_R||_{\infty} + 2||f||_{\infty} \sum_{k \in C_3} ||R_{k+T-m}^{k+T} - \mathbbm{1}\pi_R||_{\infty} \]
\[ \leq 4||f||_{\infty} \sum_{k=0}^{\infty} ||R_{k}^k - \mathbbm{1}\pi_R||_{\infty}. \quad (43) \]
Obviously $\mathbb{E}(f(X_k, Y_{k+T}) | \mathcal{F}_m) = \mathbb{E}(f(X_k, Y_{k+T}) | \mathcal{F}_{m-1})$ for $k \in C_4$, hence

$$|t_1| \leq 4\|f\|_\infty \sum_{k=0}^{\infty} \|R^k - 1\pi_R\|_\infty,$$

which together with (42) gives (40) with $\eta$ chosen as (41).

\textbf{Theorem 6.7.} If $\xi_{j,T} < 0$ it holds that

$$\mathbb{P}(S_{j,T} \geq 0) = \mathbb{P}(S_{j,T} - \xi_{j,T} \geq -\xi_{j,T}) \leq \exp \left(-\frac{\xi_{j,T}^2}{2j\eta^2}\right)$$

with $\eta$ chosen as in Lemma 6.6.

\textbf{Proof:} This follows directly from the Azuma-Hoeffding inequality for the mean zero martingale $(Z_m, \mathcal{F}_m)_{m=1}^j$, since it has increments uniformly bounded by $\eta$. \qed

\section{Mean value inequalities}

We will apply the result in the previous section by considering the Markov chain $(X_k, Y_k)_{k=1}^j$ under the exponentially tilted measure $\mathbb{P}_\pi^*$ and $(Y_k)_{k \geq j+1}$ under $\mathbb{P}_\pi$. To do so we will need to establish inequalities relating the mean of $f(X_k, Y_k)$ to the mean of $f(X_k, Y_{k+T})$ (or $f(X_{k+T}, Y_k)$). Let in the following

$$\mu^* = \mathbb{E}_{\pi^*}(f(X_k, Y_k)) = \pi^*(f)$$

denote the stationary mean of $f(X_k, Y_k)$ under the exponentially tilted measure and let

$$\mu_T^* = \mathbb{E}_{\pi^*}(f(X_k, Y_{k+T}))$$

denote the stationary mean when shifting the $Y$-sequence $T$ positions.

\textbf{Lemma 6.8.} With $\pi_1^*$ and $\pi_2^*$ the marginals of $\pi^*$ it holds that $\pi_1^* \otimes \pi_Q(f) < \mu^*$ as well as $\pi_P \otimes \pi_2^*(f) < \mu^*$.

\textbf{Proof:} We consider $(X_k, Y_k)_{k \geq 1}$ under the tilted measure and an independent stationary Markov chain $(Z_k)_{k \geq 1}$ with transition probabilities $Q$, thus $(Z_k)_{k \geq 1}$ has the same distribution as $(Y_k)_{k \geq 1}$ does under the original measure. Then

$$(X_k, Y_k, Z_k)_{k \geq 1}$$

is a Markov chain, and we define the score function $\tilde{f}$ on $E \times E \times E$ by

$$\tilde{f}(x, y, z) = f(x, z) - f(x, y).$$
The Markov chain has transition probabilities given by

\[ R^*(x,y),(x',y')Q_{z,z'} = \frac{r^*(x',y')}{r^*(x,y)} \exp(\theta f'(x',y'))P_{x,x'}Q_{y,y'}Q_{z,z'}, \]

with invariant measure \( \pi^* \otimes \pi_Q \). We also introduce the \( \tilde{\Phi}(\theta) \)-matrix

\[ \tilde{\Phi}(\theta)(x,y,z),(x',y',z') = \exp(\theta (f(x',z') - f(x',y')))R^*(x,y),(x',y')Q_{z,z'}. \]

With \( \bar{\varphi}(\theta) \) the spectral radius of \( \tilde{\Phi}(\theta) \) we have that \( \bar{\varphi}(0) = \bar{\varphi}(\theta^*) = 1 \) since \( \tilde{\Phi}(0) \) is stochastic and \( \tilde{\Phi}(\theta^*) \) has a right eigenvector with eigenvalue one having entries \( r^*(x,z)/r^*(x,y) \). Moreover, \( (8) \) together with \( (16) \) assures that \( \bar{\varphi} \) is strictly convex, and since

\[ \partial_\theta \bar{\varphi}(0) = \pi^* \otimes \pi_Q(f) - \pi^*(f) = \pi_1^* \otimes \pi_Q(f) - \mu^* \]

by \( (9) \), it follows that \( \pi_1^* \otimes \pi_Q(f) < \mu^* \). The second inequality follows similarly. \( \square \)

**Lemma 6.9.** The sequence \( (\mu_{T}^*)_{T \geq 1} \) is convergent and with

\[ \mu_{\infty}^* = \lim_{T \to \infty} \mu_{T}^* \]

it holds that \( \mu_{\infty}^* < \mu^* \).

**Proof:** We first observe that

\[ \mu_{T}^* = \mathbb{E}_{\pi^*}^* (f(X_1,Y_{1+T})) \to \pi_1^* \otimes \pi_2^*(f) \]

for \( T \to \infty \), where \( \pi_1^* \) and \( \pi_2^* \) are the marginals of \( \pi^* \).

We consider \((X_k,Y_k)_{k \geq 1}\) under the tilted measure and let \((W_k,Z_k)_{k \geq 1}\) be an independent copy with the same distribution. Then

\[ (X_k,Y_k,W_k,Z_k)_{k \geq 1} \]

is a Markov chain with transition probabilities \( R^*(x,y),(x',y')R_{w,z}^*(w',z') \) and invariant measure \( \pi^* \otimes \pi^* \). We define the score function \( f_{\infty} \) on \( E \times E \times E \times E \) by

\[ f_{\infty}(x,y,w,z) = f(x,z) + f(w,y) - f(x,y) - f(w,z). \]

Introduce the corresponding \( \tilde{\Phi}_{\infty}(\theta) \) matrix by

\[ \tilde{\Phi}_{\infty}(\theta)(x,y,w,z),(x',y',w',z') = \exp(\theta f_{\infty}(x',y',w',z'))R^*(x,y),(x',y')R_{w,z}^*(w',z') \]

and its spectral radius \( \bar{\varphi}_{\infty} \). By arguments analogous to those in Lemma 6.8 we conclude that \( \bar{\varphi}_{\infty}(0) = \bar{\varphi}_{\infty}(\theta^*) = 1 \), that \( \bar{\varphi}_{\infty}(\theta) \) is strictly convex, and that \( \partial_\theta \bar{\varphi}_{\infty}(0) = 2\mu_{\infty}^* - 2\mu^* \). Hence \( \mu_{\infty}^* < \mu^* \). \( \square \)

It is interesting and very useful that the inequality in Lemma 6.9 not only holds in the limit but in fact for all \( T \).
Lemma 6.10. For all \( T \geq 1 \) it holds that
\[
\mu_T^* < \mu^*.
\] (45)

Proof: With \( S_n^T = \sum_{k=1}^n f(X_k, Y_{k+T}) \) and \( S_n = \sum_{k=1}^n f(X_k, Y_k) \) we observe that \( S_n \overset{D}{=} S_n^T \), since under \( \mathbb{P} = \mathbb{P}_\pi \) the \( X \)- and \( Y \)-sequence are independent, stationary Markov chains. By (33) this implies that for \( \theta > 0 \)
\[
\frac{1}{n} \log \mathbb{E}(\exp(\theta S_n^T)) \rightarrow \log \varphi(\theta)
\] (46)
for \( n \to \infty \).

Consider first the case \( T = 1 \) and the Markov chain
\[(X_k, X_{k+1}, Y_k, Y_{k+1})_{k \geq 1},\]
which under the tilted measure has transition probabilities
\[
R^*_{(x,w,y,z),(x',w',y',z')} = \frac{r^*(w',z')}{r^*(w,z)} \exp(\theta^* f(w',z')) P_{w,w'} Q_{z,z'} \delta_{w,x'} \delta_{z,y'}.
\]

Introduce the matrix
\[
\tilde{\Phi}_1(\theta)_{(x,w,y,z),(x',w',y',z')} = \exp(\theta (f(x',z') - f(w',z'))) R^*_{(x,w,y,z),(x',w',y',z')}
\]
and its spectral radius \( \tilde{\varphi}_1(\theta) \). Clearly, \( \tilde{\varphi}_1(0) = 1 \) and we observe that
\[
\tilde{\Phi}_1(\theta^*)_{(x,w,y,z),(x',w',y',z')} = \frac{r^*(w',z')}{r^*(w,z)} \exp(\theta^* f(x',z')) P_{w,w'} Q_{z,z'} \delta_{w,x'} \delta_{z,y'}.
\]

The matrix \( \tilde{\Phi}(\theta^*) \) has the same spectrum if we remove the eigenvector fraction, hence (33) together with (46) imply that
\[
\log \tilde{\varphi}_1(\theta^*) = \lim_{n \to \infty} \frac{1}{n} \log \mathbb{E}(\exp(\theta^* S_n^1)) = \log \varphi(\theta^*) = 0,
\]
thus \( \tilde{\varphi}_1(\theta^*) = 1 \).

Furthermore, by (9) \( \partial_\theta \tilde{\varphi}_1(0) = \mu_1^* - \mu^* \). Using (8) (for \( T = 1 \)) together with (16) we find that \( \varphi_1 \) is strictly convex, hence
\[
\mu_1^* < \mu^*.
\]

A similar argument for \( T \geq 2 \) is possible by introducing the Markov chain
\[(X_k, \ldots, X_{k+T}, Y_k, \ldots, Y_{k+T})_{k \geq 1}\]
on \( E^{T+1} \times E^{T+1} \) and a score function \( f_T \) given by
\[
f_T(x_0, \ldots, x_T, y_0, \ldots, y_T) = f(x_0, y_T) - f(x_T, y_T).
\]
The spectral radius \( \tilde{\varphi}_T(\theta) \) of the corresponding matrix \( \tilde{\Phi}_T(\theta) \) fulfills that \( \tilde{\varphi}_T(0) = \tilde{\varphi}_T(\theta^*) = 1 \), \( \partial_\theta \tilde{\varphi}_T(0) = \mu_T^* - \mu^* \) and it is strictly convex by (8) and (16). Thus \( \mu_T^* < \mu^* \). \( \square \)
6.6 Variables shared in both sequences

We define for \(i, j, l, T \geq 1\) with \(i \leq j\)

\[
S_1 = \sum_{k=1}^{i} f(X_k, Y_k), \quad S_2 = \sum_{k=i+1}^{j} f(X_k, Y_k)
\]

\[
\tilde{S}_2 = \sum_{k=i+1}^{j} f(X_k, Y_{k+T}), \quad S_3 = \sum_{k=j+1}^{i+l} f(X_k, Y_{k+T}).
\]

**Theorem 6.11.** There exists an \(\varepsilon > 0\) and some \(K\) (both independent of \(T\)) such that

\[
\mathbb{P}(S_1 + S_2 > t, \tilde{S}_2 + S_3 > t) \leq K \exp(-\theta^*(1 + \varepsilon)t) \quad (47)
\]

for \(t \geq 0\).

**Proof:** Assume first that the number of variables \(j - i\) in the overlapping part is small, less than \(t(4\|f\|_{\infty})^{-1}\), say, in which case we obtain the estimate

\[
\mathbb{P}(S_1 + S_2 > t, \tilde{S}_2 + S_3 > t) \leq \mathbb{P}(S_1 > 3/4t, S_3 > 3/4t)
\]

\[
\leq \rho \mathbb{P}(S_1 > 3/4t) \mathbb{P}(S_3 > 3/4t)
\]

\[
\leq K \exp(-3/2\theta^* t),
\]

using Lemma 6.1 for the second inequality and then a standard exponential change of measure argument. This implies (47) with \(\varepsilon = 1/2\).

If instead \(j - i \geq t(4\|f\|_{\infty})^{-1}\) we observe that

\[
\mathbb{P}(S_1 + S_2 > t, \tilde{S}_2 + S_3 > t)
\]

\[
\leq \mathbb{P}(S_1 + S_2 > t, \tilde{S}_2 \geq S_2) + \mathbb{P}(\tilde{S}_2 + S_3 > t, S_2 \geq \tilde{S}_2). \quad (48)
\]

With \(L_j^* = r(X_j, Y_j)/r(X_0, Y_0) \exp(\theta^*(S_1 + S_2))\) we obtain

\[
\mathbb{P}_\pi(S_1 + S_2 > t, \tilde{S}_2 \geq S_2) = \mathbb{P}_\pi\left(\frac{L_j^*}{L_j}; S_1 + S_2 > t, \tilde{S}_2 \geq S_2\right)
\]

\[
\leq b^{-1} \exp(-\theta^* t) \mathbb{P}_{\pi, j}^*(\tilde{S}_2 \geq S_2)
\]

where \(\mathbb{P}_{\pi, j}^*\) denotes the tilted measure up to index \(j\). Using Lemma 6.1, we can, at the expense of a factor \(\rho\), assume that the sequence \((X_k, Y_k)_{k=i}^j\) is a stationary Markov chain under the tilted measure and that \((Y_k)_{k \geq j+1}\) is independent and stationary under the original measure. Under this assumption it follows that the mean of \(\tilde{S}_2 - S_2\)
equals \((j - T - i)\mu^*_T + T\pi^*_T \otimes \pi_Q(f) - (j - i)\mu^*\). Using Lemma 6.8, Lemma 6.9, and Lemma 6.10 we can find a \(\zeta > 0\), independent of \(T\), such that
\[
(j - T - i)\mu^*_T + T\pi^*_T \otimes \pi_Q(f) - (j - i)\mu^* < -(j - i)\zeta.
\]
Hence Theorem 6.7 gives that
\[
P_{\pi,j}^\pi(\tilde{S}_2 \geq S_2) \leq \rho \exp \left( -\frac{\zeta^2 (j - i)}{2\eta^2} \right) \leq \rho \exp \left( -\frac{\zeta^2 t}{8\|f\|_\infty \eta^2} \right)
\]
or, with \(\varepsilon = \zeta^2 (\theta^* 8\|f\|_\infty \eta^2)^{-1}\),
\[
P(S_1 + S_2 \geq t, \tilde{S}_2 \geq S_2) \leq \rho b^{-1} \exp(-\theta^*(1 + \varepsilon)t)
\]
and (47) follows. Of course, a similar argument takes care of the second term in (48).

\[\Box\]

### 6.7 Proof of the Poisson approximation

Recall from Remark 3.7 the definition of the Markov additive process \((S_n)_{n \geq 0}\) and the reflection \((T_n)_{n \geq 0}\) at the zero barrier. Define
\[
M_n = \max_{1 \leq k \leq m \leq n} S_m - S_k = \max_{1 \leq m \leq n} T_m.
\]
Then, for some \(x, y \in \mathbb{E}\), the stopping time
\[
\sigma = \inf\{n \geq 1 \mid T_n = 0, X_n = x, Y_n = y\}
\]
is an almost surely finite regeneration time for the process \((T_n)_{n \geq 0}\). With the constant \(K^*\) as defined by Karlin & Dembo (1992) a combination of Lemma B in (Karlin & Dembo 1992) and Proposition 10.1 in (Asmussen & Perry 1992) gives that
\[
\exp(\theta^* u)P_{x,y}^\pi \left( \max_{1 \leq m \leq \sigma} T_m > u \right) \rightarrow \mathbb{E}_{x,y}(\sigma)K^*
\]
for \(u \rightarrow \infty\) (within \(\mathbb{Z}\)). That \(\mathbb{E}_{x,y}(\sigma)K^*\) occurs as the right constant on the r.h.s. is a consequence of Wald’s identity for MAPs.

**Lemma 6.12.** Suppose that \(n(u)\) is a sequence of integers satisfying
\[
\frac{u}{n(u)} \rightarrow 0 \quad \text{and} \quad n(u) \exp(-\theta^* u) \rightarrow 0
\]
for \(u \rightarrow \infty\). Then
\[
P(M_{n(u)} > u) \sim n(u)K^* \exp(-\theta^* u)
\]
for \(u \rightarrow \infty\).
Figure 3: For each diagonal-within-a-strip given by $a$ we define a neighbourhood of strong dependence. This figure shows the neighbourhood for some $a = (k, r)$. The ‘arms’ of this cross correspond to diagonals-within-a-strip, $b$, sharing variables with $a$ from either the $X$- or $Y$-Markov chain but not both. In the intersection $b$ can share variables with $a$ from both the sequences. The dashed lines mark the strips into which the matrix is divided.

Proof: Let $G(u) = \mathbb{P}_{x,y}(M_{\sigma} \leq u)$ denote the distribution function for $M_{\sigma}$ when $X_0 = x$ and $Y_0 = y$ and let $G_{\pi}(x) = \mathbb{P}(M_{\sigma} \leq u)$ denote the distribution function when the Markov chains are stationary (recall that $\mathbb{P} = \mathbb{P}_{\pi}$). Introduce the sequence $(\sigma(m))_{m \geq 0}$ of stopping times by $\sigma(0) = 0$, $\sigma(1) = \sigma$, and for $m \geq 2$

$$\sigma(m) = \inf\{n > \sigma(m-1) \mid T_n = 0, X_n = x, Y_n = y\}. $$

Then

$$M_{\sigma(m)} = \max_{0 \leq k \leq \sigma(m)} T_k = \max_{1 \leq k \leq m, \sigma(k-1) \leq r \leq \sigma(k)} T_r, $$

which is a maximum of $m-1$ iid variables having distribution function $G$ and one variable having distribution function $G_{\pi}$. Thus $\mathbb{P}(M_{\sigma(m)} > u) = 1 - G(u)^{m-1}G_{\pi}(u)$.

Define, for $\delta > 0$ fixed, the sequences $m_-(u) = \gamma_-(u)n(u)/\mathbb{E}_{x,y}(\sigma)$ and $m_+(u) = \gamma_+(u)n(u)/\mathbb{E}_{x,y}(\sigma)$ with $\gamma_-(u) \leq 1 - \delta$ and $\gamma_+(u) \geq 1 + \delta$ chosen maximally and minimally such that $m_-(u), m_+(u) \in \mathbb{N}$. In particular this implies that $\gamma_-(u) \to 1 - \delta$
and $\gamma_+(u) \to 1 + \delta$ for $u \to \infty$. Since $(1 - G(u)) \exp(\theta^* u) \to \mathbb{E}_{x,y}(\sigma)K^*$ by (50) we have, using that $n(u) \exp(-\theta^* u) \to 0$, that

$$(1 - G(u)^{m_-(u)-1}G_\pi(u)) \frac{\exp(\theta^* u)}{n(u)} \to (1 - \delta)K^* \quad \text{and}$$

$$(1 - G(u)^{m_+(u)-1}G_\pi(u)) \frac{\exp(\theta^* u)}{n(u)} \to (1 + \delta)K^*.$$  

Since $M_n$ is clearly increasing the following inequalities hold

$$\mathbb{P}(M_{\sigma(m_-)} > u) - \mathbb{P}(\sigma(m_-) > n(u)) \leq \mathbb{P}(M_{n(u)} > u) \leq \mathbb{P}(M_{\sigma(m_+)} > u) + \mathbb{P}(\sigma(m_+) < n(u)). \tag{53}$$  

The stopping time $\sigma(m) = \sigma + \sum_{k=2}^m \sigma(k) - \sigma(k-1)$ is a sum of $m - 1$ iid variables and one independent variable $\sigma$ with $\mathbb{E}_{x,y}(\exp(\lambda \sigma)) < \infty$ as well as $\mathbb{E}(\exp(\lambda \sigma)) < \infty$ for some $\lambda > 0$, and it follows that we can find $\lambda^* > 0$ such that

$$\mathbb{P}(\sigma(m_-) > n(u)) \leq \exp(-\lambda^* m_-)u).$$

This implies that

$$\exp(\theta^* u)\mathbb{P}(\sigma(m_-) > n(u)) \to 0$$

by $u/n(u) \to 0$. And similarly $\exp(\theta^* u)\mathbb{P}(\sigma(m_+) < n(u)) \to 0$, which together with (53) imply that

$$(1 - \delta)K^* \leq \liminf_{u \to \infty} \mathbb{P}(M_{n(u)} > u) \frac{\exp(\theta^* u)}{n(u)} \leq \limsup_{u \to \infty} \mathbb{P}(M_{n(u)} > u) \frac{\exp(\theta^* u)}{n(u)} \leq (1 + \delta)K^*.$$  

Letting $\delta \to 0$ gives (52). \hfill \Box

We need to define the neighbourhood of strong dependence $B_a$ for $a = (k, r) \in I$. Introducing

$$B^1_a = \{k - l, \ldots, k + 2l\} \times \{0, l, 2l, \ldots, \left\lfloor \frac{n}{l} \right\rfloor l\}$$

$$B^2_a = \{0, \ldots, n\} \times \{r - l, r, r + l\}$$

we define $B_a = B^1_a \cup B^2_a$, cf. Figure 3. Note that $\max_a |B_a| = O(n)$. The set $B^1_a$ is a horizontal strip of strong dependence and $B^2_a$ is a vertical strip of strong dependence. The set $B_a$ as defined above is not necessarily contained in the index set $I$ and should be properly modified close to the boundaries of $I$. We could simply choose to always consider $B_a \cap I$. This boundary modification plays no role and will be ignored throughout.
Lemma 6.13. If we, for some \( x \in \mathbb{R} \), let

\[
\begin{align*}
l &= l_n \sim (\log n^2)^3 \quad \text{and} \quad t = t_n = \frac{\log K^* + \log n^2 + x}{\theta^*} \\
t &= t_n = \log K^* + \log n^2 + x \theta^*
\end{align*}
\]

and define \( x_n \in [0, \theta^*] \) by \( x_n = \theta^* (t_n - \lfloor t_n \rfloor) \), then under the assumptions in Theorem 3.1, the conditions in Theorem 4.1 are fulfilled with

\[
\lambda_n = \exp(-x + x_n).
\]

That is

\[
\left\| D \left( \sum_{a \in I} V_a \right) - \text{Poi}(\exp(-x + x_n)) \right\| \to 0.
\]

Proof: According to Lemma 6.12

\[
\mathbb{P}(V_{(0,0)} = 1) = \mathbb{P} \left( \max_{1 \leq \delta \leq \Delta} \sum_{k=\delta}^{\Delta} f(X_k, Y_k) > t - x_n/\theta^* \right) \sim ln^{-2} \exp(-x + x_n) (55)
\]

for \( n \to \infty \). Since \( |I| \sim n^2 l^{-1} \) and, due to stationarity, all the events \( (V_a = 1) \) for \( a \in I \) are equally probable, it follows from (55) that

\[
\sum_{a \in I} \mathbb{E}(V_a) = |I| \mathbb{E} \left( \max_{1 \leq \delta \leq \Delta} \sum_{k=\delta}^{\Delta} f(X_k, Y_k) > t - x_n/\theta^* \right) \sim \exp(-x + x_n),
\]

or, since \( \exp(-x + x_n) \) is bounded,

\[
\left| \sum_{a \in I} \mathbb{E}(V_a) - \exp(-x + x_n) \right| \to 0
\]

for \( n \to \infty \).

Furthermore, we observe that \( |I| \mathbb{E}(V_{(0,0)}) \) is bounded and that \( \max_a |B_a| = o(|I|) \) for \( n \to \infty \), so condition (22) is fulfilled by

\[
\sum_{a \in I, b \in B_a} \mathbb{E}(V_a) \mathbb{E}(V_b) \leq |I| \times \max_a |B_a| \times \mathbb{E}(V_{(0,0)})^2 = O \left( \frac{\max_a |B_a|}{|I|} \right) \to 0.
\]

We prove that (23) is fulfilled by splitting the set \( B_a \) into three disjoint sets and, depending on the set, give a bound of \( \mathbb{E}(V_a V_b) \) for \( b \) in each of these sets. For \( a \in I \) let

\[
B_a = C_a \cup D_{a1} \cup D_{a2}
\]
with $C_a$, $D^{1}_a$ and $D^{2}_a$ being the disjoint sets

$$C_a = B^{1}_a \cap B^{2}_a, \quad D^{1}_a = B^{1}_a \setminus C_a \quad \text{and} \quad D^{2}_a = B^{2}_a \setminus C_a$$

We note that $C_a$ is the centre of the set $B_a$, and $D^{1}_a$ and $D^{2}_a$ are the remaining horizontal and vertical ‘arms’.

Consider the case $b \in C_a$ and $b \neq a$. Using (30) together with Lemma 6.11 we can find an $\varepsilon > 0$ such that

$$\mathbb{E}(V_a V_b) \leq Kl^4 \exp(-\theta^* (1 + \varepsilon) t).$$

Hence, observing that $\sum_{a \in I} |C_a| \leq 9 |I| l \leq 9n^2$,

$$\sum_{a \in I, b \in C_a, b \neq a} \mathbb{E}(V_a V_b) \leq Kl^4 n^{-2(1+\varepsilon)} \sum_{a \in I} |C_a| \to 0$$

for $n \to \infty$.

For $b \in D^{i}_a$, Lemma 6.4 with $\theta = 3\theta^*(1 + \varepsilon)/4$ for some $\varepsilon > 0$ applies due to (12), and together with (30) we obtain that

$$\mathbb{E}(V_a V_b) \leq Kl^4 \exp(-3\theta^*(1 + \varepsilon) t/2),$$

in which case

$$\sum_{a \in I, b \in D^{i}_a} \mathbb{E}(V_a V_b) \leq Kl^4 n^{-3(1+\varepsilon)} \sum_{a \in I} |D^{i}_a| \leq Kl^4 n^{-3(1+\varepsilon)} 8n^3 \to 0.$$

The two-dimensional process $(X_k, Y_k)_{k \geq 1}$ is a stationary, irreducible Markov chain on a finite state space, hence we can extend it to a doubly infinite, stationary process $(X_k, Y_k)_{k \in \mathbb{Z}}$, which is exponentially $\beta$-mixing. The $\beta$-mixing coefficients therefore satisfy

$$\beta(k) \leq K_1 \exp(-K_2 k)$$

for some constants $K_1, K_2 > 0$. For $a = (r, r) \in I$ we define $I_1 = (-\infty, r - l]$, $I_2 = [r + 1, r + l]$, and $I_3 = [r + 2l + 1, \infty)$, for which $d(I_1 \cup I_3, I_2) = l + 1$. Then clearly with $I = I_1 \cup I_3$ and $J = I_2$, $\mathcal{F}_a \subseteq \mathcal{F}_I = \sigma(X_n, Y_n \mid n \in I_1 \cup I_3)$ and $V_a$ is measurable w.r.t. $\mathcal{F}_J = \sigma(X_n, Y_n \mid n \in I_2)$. By Lemma 5.1 and Lemma 5.2 it follows that

$$\mathbb{E} \mathbb{E}(V_a \mid \mathcal{F}_a) - \mathbb{E}(V_a) \leq 2\alpha(\mathcal{F}_I, \mathcal{F}_J) \leq 6\beta(l + 1) \leq K \exp(-K_2 l),$$
with $K = 6K_1 \exp(-K_2)$. For any non-diagonal $a = (k, r) \in I$ we can shift the $X$-process by stationarity to reduce the problem to the previous one and thus to obtain the same bound. This bound implies that

$$\sum_{a \in I} \mathbb{E} |E(V_a | F_a) - \mathbb{E}(V_a)| \leq Kn^2 \exp(-K_2(\log n)^3) \to 0$$

for $n \to \infty$. □

To finish the proof of Theorem 3.1 we need to show that

$$\mathbb{P} \left( \sum_{a \in I} V_a(t_n) \neq C(t_n) \right) \to 0, \quad (56)$$

since this together with Lemma 6.13 and Corollary 4.2 proves Theorem 3.1.

**Proof of (56):** Introduce for $a = (k, r) \in I$ the excursion $e(a)$ containing $a$, i.e.

$$e(a) = (i, j, \Delta) \quad \text{if} \quad a \in \{(i+1, j+1), \ldots, (i+\Delta, j+\Delta)\},$$

let $E_n^a$ denote the set of excursions contained in the diagonal-within-a-strip given by $a$, i.e. with $A = \{(k+1, r+1), \ldots, (k+l, r+l)\}$

$$e = (i, j, \Delta) \in E_n^a \quad \text{if} \quad (i+1, j+1), (i+\Delta, j+\Delta) \in A,$$

and finally let $\tilde{E}_n \subseteq E_{n+l}$ be defined by

$$(i, j, \Delta) \in \tilde{E}_n \text{ if } (i, j, \Delta) \in E_{n+l} \text{ and either } i \leq l, i \geq n-l, \text{ or } j \geq n-l.$$ Introduce also the three sets

$$A_1 = \left( \exists a \in I : \mathcal{M}_{e(a)} > t \right)$$

$$A_2 = \left( \exists a \in I \exists e, e' \in E_n^a : \mathcal{M}_e > t, \mathcal{M}_{e'} > t \right)$$

$$A_3 = \left( \exists e \in \tilde{E}_n : \mathcal{M}_e > t \right),$$

then

$$\left( \sum_{a \in I} V_a(t) \neq C(t) \right) \subseteq A_1 \cup A_2 \cup A_3. \quad (57)$$

It is easiest to see this by observing that on the complement of the r.h.s. the counting variable $C(t)$ equals $\sum_{a \in I} V_a(t)$.

With the notation in Lemma 6.2 it holds that

$$(\mathcal{M}_{e(a)} > t) = \left( \max_{(i,j,\Delta) \in A(a)} S_{i,j}^\Delta > t \right)$$
hence from (34) it follows that
\[
P(\exists a \in I : \mathcal{M}_{e(a)} > t_n) \leq K \{ |I| n^{-2} + |I| (\log n)^2 \exp(-\theta^* t_n) \} \leq \tilde{K} \{ (\log n)^{-3} + (\log n)^{-1} \} \to 0.
\]

For a given \(a \in I\) we get, using (52) and Lemma 6.1, that
\[
P(\exists e, e' \in \mathcal{E}_n^a : \mathcal{M}_e > t_n, \mathcal{M}_{e'} > t_n) \leq K l^3 \exp(-2\theta^* t_n).
\]

Hence
\[
P(\exists a \in I \exists e, e' \in \mathcal{E}_n^a : \mathcal{M}_e > t_n, \mathcal{M}_{e'} > t_n) \leq K l^2 n^{-2} \to 0
\]
for \(n \to \infty\).

Clearly
\[
P(\exists e \in \mathcal{E}_n : \mathcal{M}_e > t_n) \leq K n l \exp(-\theta^* t_n) \leq K_1 l n^{-1} \to 0.
\]

By (57) we get (56). \(\square\)

7 Concluding remarks

It is unfortunate that we haven’t been able to identify sufficient conditions for Theorem 3.1 that are equivalent to the condition \((E')\) in Dembo et al. (1994b) in the iid setup. Attempts of improvements should quite clearly be targeted at Lemma 6.4, which is the reason for assuming (12). The exponential change of measure technique used in the proof of Lemma 6.4 is believed to have been optimised as much as possible. Thus to achieve improvements it seems that another approach must be invented.

As mentioned in the introduction the result is a generalisation of that obtained by Dembo et al. (1994b) for aligning independent iid sequences. It might be worthwhile to emphasise the similarities and differences between their proof and the proof presented here. The overall strategy of using the Poisson approximation by Arratia et al. (1989) for proving a result like Theorem 3.1 is essentially identical to the strategy used by Dembo et al. (1994b). Indeed, Arratia et al. (1989) apply the Poisson approximation to a similar sequence comparison problem. However, the counting construction presented here for the approximation of \(C(t)\) differs from the counting construction in (Dembo et al. 1994b). Lemma 6.12 is a straightforward generalisation of Lemma 1 in (Dembo et al. 1994b) and Lemma 6.2 is similar to Lemma 1 in (Dembo et al. 1994a), which is also used in (Dembo et al. 1994b). Besides
these two Lemmas there is little similarity in the details of the two proofs. First of all an extra argument based on mixing inequalities was needed in order to take care of the $\beta_{4,n}$-term. The major challenge was, however, the generalisation of the part of the proof of Lemma 2 in (Dembo et al. 1994b) called case (c), where a smart permutation argument relying on exchangeability of iid-variables was used. The solution presented here, which works for Markov chains, is an application of the Azuma-Hoeffding inequality for martingales as described in Section 6.4 and 6.5.

References


DEPARTMENT OF APPLIED MATHEMATICS AND STATISTICS
UNIVERSITETSPARKEN 5
DK-2100 COPENHAGEN Ø
DENMARK
richard@math.ku.dk