COMPARISON OF SEQUENCES GENERATED BY A HIDDEN MARKOV MODEL

A Dissertation
Presented to
The Academic Faculty

By

George Kerchev

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Mathematics

Georgia Institute of Technology

May 2019

COMPARISON OF SEQUENCES GENERATED BY A HIDDEN MARKOV MODEL

Approved by:

Prof. Houdré, Advisor
School of Mathematics
Georgia Institute of Technology

Prof. Damron
School of Mathematics
Georgia Institute of Technology

Prof. Koltchinskii
School of Mathematics
Georgia Institute of Technology

Prof. Tikhomirov
School of Mathematics
Georgia Institute of Technology

Prof. Foley
School of Industrial and Systems Engineering
Georgia Institute of Technology

Date Approved: March 12, 2019

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Houdré for introducing me to the many interesting questions regarding sequences comparison. His guidance and support made my Ph.D experience a very pleasant one.

Next I would like to thank Dr. Damron, Dr. Tikhomirov, Dr. Koltchinskii and Dr. Foley for serving as my committee members. Special thanks to Michael who has been an inspiration and a friend throughout my mathematical journey.

Finally I would like to thank my family for the love and support.

TABLE OF CONTENTS

LIST OF FIGURES

SUMMARY

The length LC_n of the longest common subsequences of two strings $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_n)$ is way to measure the similarity between X and Y. We study the asymptotic behavior of LC_n when the two strings are generated by a hidden Markov model $(Z, (X, Y))$. The latent chain Z is an aperiodic time-homogeneous and irreducible finite state Markov chain and the pair (X_i, Y_i) is generated according to a distribution depending of the state of Z_i for every $i \geq 1$. The letters X_i and Y_i each take values in a finite alphabet $\mathcal{A}$.

The goal of this work is to build upon asymptotic results for LC_n obtained for sequences of iid random variables. Under some standard assumptions regarding the model we first prove convergence results with rates for $\mathbb{E}[LC_n]$. Then, versions of concentration inequalities for the transversal fluctuations of LC_n are obtained. Finally, we have outlined a proof for a central limit theorem by building upon previous work and adapting a Stein's method estimate.

CHAPTER 1

INTRODUCTION AND BACKGROUND

There are many settings where one needs to compare two sequences and measure their similarity. For instance, comparison can be used for error-correction or for establishing relations. If there is a clear correspondence between the letters in the two sequences it is most sensible to develop measures that use it. Such measures between $a = (a_1, \ldots, a_n)$ and $b = (b_1, \ldots, b_n)$ are the Euclidean distance $\sqrt{\sum_{i=1}^{n}(a_i - b_i)^2}$, the city block distance $\sum_{i=1}^{n} |a_i - b_i|$ or the Hamming distance $\sum_{i=1}^{n} \mathbf{1}_{a_i=b_i}$.

However, often the correspondence between letters is not known in advance. A useful measure in this case which also respects the order of the elements in the sequences is the *length of the longest common subsequences*.

For two finite sequences $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_m)$ taking values in a finite alphabet $\mathcal{A}$, the object of study is $LCS(X_1, \ldots, X_n; Y_1, \ldots, Y_m)$, the length of the longest common subsequences of $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$, which is abbreviated as LC_n when $n = m$. Clearly LC_n is the largest k such that there exist $1 \leq i_1 < \cdots < i_k \leq n$ and $1 \leq j_1 < \cdots < j_k \leq n$ with

$$X_{i_s} = Y_{j_s}, \text{ for all } s = 1, 2, 3, \ldots, k.$$

Example 1.0.1. If $X = \{A, C, T, G, A, C, T, C, A, A, G, C, A, T, A\}$ and $Y = \{C, A, A, G, C, A, T, A, A, C, T, G, A, C, T\}$, the longest common subsequences have length 9, and so $LC_{15} = 9$. Here is a visual representation of this fact. The longest common subsequences can be seen in bold (and red).

A C T G **A C T** C A **A** G **C** A **T** A

C A **A** G **C** A **T** A **A C T** G **A C T**

Note that longest common subsequences do not necessarily consist of only contiguous letters. The realization is also not unique. Here is another set of longest common subsequences for the same strings X and Y.

A C T **G** A **C T** C **A A G** C **A T** A

C A **A G C** A **T A A** C T **G A** C **T**

1.1 Asymptotic results for LC_n

When the sequences are generated by a probabilistic model, the length of the longest common subsequence is a random variable as well. The distribution of LC_n depends on the model and cannot be determined for general n even for some of the simplest models. Some understanding of LC_n has been obtained through several asymptotic results. We now list some of the major ones.

For two independent words sampled independently and uniformly at random from the alphabet, Chvátal and Sankoff [7] proved that

$$\lim_{n\to\infty} \mathbb{E}[LC_n]/n = \gamma^*,$$

and provided upper and lower bounds on γ^*.

This was followed by Alexander [1] who obtained, for iid draws, the following generic rate of convergence result:

$$n\gamma^* - C\sqrt{n \log n} \leq \mathbb{E}[LC_n] \leq n\gamma^*, \tag{1.1}$$

where $C > 0$ is an absolute constant.

Then Houdré and Matzinger [18] showed a closeness to the diagonal result which is rather technical and will be stated more precisely in Chapter 3. Further work by Houdré and Işlak [16] establishes a central limit theorem for the length of the longest common subsequence in the iid case.

1.2 Hidden Markov models and thesis outline

From a practical point of view the independence assumptions, both between words and also among draws, has to be relaxed as they are often lacking. One such instance is in the field of computational biology where one compares similarities between two biological sequences. In particular, alignments of those sequences need to be qualified as occurring by chance or because of a structural relation. One way to generate alignments is with a hidden Markov model (HMM). The states of the hidden chain account for a match between two elements in X and Y or for an alignment of an element with a gap. Given X and Y one can find the most probable alignment using the Viterbi algorithm. This model is particularly useful when the similarity between X and Y is weak. In this case standard methods for pairwise alignment often fail to identify the correct alignment or test for its significance. With a hidden Markov model one can evaluate the total probability that X and Y are aligned by summing up over all alignments, and this sum can be efficiently computed with the Forward algorithm. For more information we refer the reader to Chapter 4 in [11].

There are very few results on the asymptotics of the longest common subsequences in a model exhibiting dependence properties. A rare instance is due to Steele [26] who showed the convergence of $\mathbb{E}[LC_n]/n$ when (X, Y) is a random sequence for which there is a stationary ergodic coupling, e.g., an irreducible, aperiodic, positive recurrent Markov chain. This thesis studies the longest common subsequences for strings exhibiting a different Markov relation, namely we study the case when (X, Y) is emitted by a latent Markov chain Z, i.e., when $(Z, (X, Y))$ is a hidden Markov model. A hidden Markov model consists of a Markov chain $Z = (Z_i)_{i\geq 1}$ which emits the observed variables $(X_i)_{i\geq 1}$. In our setting the Markov chain Z is defined on a finite state space $\mathcal{S}$ and is aperiodic, time homogeneous and irreducible. The possible states of Z are each associated with a distribution on the values of X. In other words, the observation $X = (X_i)_{i\geq 1}$ is a mixture model where the choice of mixture component of each observation depends on the component of the previous observation. The mixture components are given by the sequence Z. In our setting a single sequence of components Z gives rise to the two sequences (X, Y). In particular each Z_i generates a pair (X_i, Y_i), where X and Y take values in the same alphabet $\mathcal{A}$. Graphically, the model can be presented as

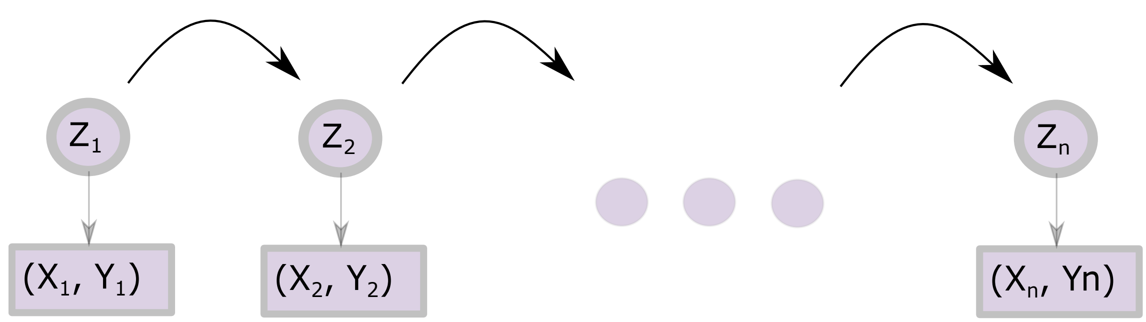

Figure 1.1: A hidden Markov model

Note that this framework includes the special case when (Z, X) and (Z', Y) are hidden Markov models, with the same parameters, while Z and Z' are independent. In our setting, mean convergence is quickly proved in Section 2.1 of Chapter 2. Then, a rate of convergence result, obtained in Section 2.2 of Chapter 2, recovers, in particular, (1.1). In Chapter 3 we generalize the technical result regarding closeness to the diagonal of Houdré and Matzinger [18] mentioned above. Chapter 4 establishes a Stein's method for functions of hidden Markov models by generalizing on a result by Chatterjee [6]. This result is used in Chapter 5 to outline a proof for the central limit theorem by following the approach in [16].

Throughout this manuscript our probability space $(\Omega, \mathcal{F}, \mathbb{P})$ *is assumed to be rich enough to consider all the random variables we are studying.*

CHAPTER 2
RATE OF CONVERGENCE

In this chapter we present a few results regarding the asymptotic behavior of $\mathbb{E}[LC_n]$ and thus generalizing the work of Chvátal and Sankoff [7] and Alexander [1].

2.1 Mean convergence

Recall that a hidden Markov model (Z, V) consists of a Markov chain $Z = (Z_n)_{n \geq 1}$ which emits the observed variables $V = (V_n)_{n \geq 1}$. The possible states in Z are each associated with a distribution on the values of V. In other words the observation V is a mixture model where the choice of the mixture component for each observation depends on the component of the previous observation. The mixture components are given by the sequence Z. Note also that given Z, V is a Markov chain. For such a model our first easy result asserts the mean convergence of LC_n.

Proposition 2.1.1. *Let Z be an aperiodic, irreducible, time homogeneous finite state space Markov chain. Let μ, P, and π be respectively the initial distribution, transition matrix and stationary distribution of Z. Let each Z_n, $n \geq 1$, generate a pair (X_n, Y_n) according to a distribution associated to the state of Z_n, i.e., let $(Z, (X, Y))$ be a hidden Markov model, where $X = (X_n)_{n \geq 1}$ and $Y = (Y_n)_{n \geq 1}$. Further, for all $i \geq 1$ and $j \geq 1$, let X_i and Y_j take their values in the common finite alphabet $\mathcal{A}$ and let there exists $a \in \mathcal{A}$, such that $\mathbb{P}(X_i = Y_j = a) > 0$, for some $i \geq 1$ and $j \geq 1$. Then,*

$$\lim_{n \to \infty} \frac{\mathbb{E}[LC_n]}{n} = \gamma^*,$$

where $\gamma^* \in (0, 1]$.

Proof. If $\mu = \pi$, the sequence (X, Y) is stationary and therefore by superadditivity and

Fekete's lemma or Kingman's subadditivity theorem (see [27]) imply:

$$\lim_{n\to\infty} \frac{\mathbb{E}[LC_n]}{n} = \sup_{k\geq 1} \frac{\mathbb{E}[LC_k]}{k} = \gamma^*, \tag{2.1}$$

for some $\gamma^* \in (0,1]$. When $\mu \neq \pi$, a coupling technique will prove the result. Let $\overline{Z}$ be a Markov chain with initial and stationary distribution π and having the same transition matrix P as the chain Z. Assume, further, that the emission probabilities are the same for Z and $\overline{Z}$ and denote by $(\overline{Z}, (\overline{X}, \overline{Y}))$ the corresponding HMM. Next consider the coupling $(Z, \overline{Z})$ where the two chains stay together after the first time i for which $Z_i = \overline{Z}_i$, and let τ be the meeting time of Z and $\overline{Z}$. Next, and throughout, let $X^{(n)} := (X_1, \dots, X_n)$ and similarly for $Y^{(n)}, \overline{X}^{(n)}$ and $\overline{Y}^{(n)}$. Since $LCS(X^{(n)}; Y^{(n)}) - LCS(\overline{X}^{(n)}; \overline{Y}^{(n)}) \leq n$, then for any $K > 0$,

$$\begin{aligned}
&|\mathbb{E}[LCS(X^{(n)}; Y^{(n)}) - LCS(\overline{X}^{(n)}; \overline{Y}^{(n)})]| \\
= & \left| \mathbb{E}\left[[LCS(X^{(n)}; Y^{(n)}) - LCS(\overline{X}^{(n)}; \overline{Y}^{(n)})] \mathbf{1}_{\tau > K} \right] \right. \\
& \left. \qquad + \mathbb{E}\left[[LCS(X^{(n)}; Y^{(n)}) - LCS(\overline{X}^{(n)}; \overline{Y}^{(n)})] \mathbf{1}_{\tau \leq K} \right] \right| \\
\leq & \; n\mathbb{P}(\tau > K) + K + \left| \mathbb{E}\left[[LCS^K(X^{(n)}; Y^{(n)}) - LCS^K(\overline{X}^{(n)}; \overline{Y}^{(n)})] \mathbf{1}_{\tau \leq K} \right] \right| \\
\leq & \; n\mathbb{P}(\tau > K) + K,
\end{aligned} \tag{2.2}$$

where $LCS^K(\cdot;\cdot)$ is now the length of the longest common subsequences restricted to the letters X_i and Y_i, for $i > K$, noting also that when $\tau \leq K$, then $LCS^K(X^{(n)}; Y^{(n)})$ and $LCS^K(\overline{X}^{(n)}; \overline{Y}^{(n)})$ are identically distributed. If $K \in (mk, m(k+1)]$, for some $m \geq 0$, by an argument going back to Doeblin [9] (see also [28]),

$$
\begin{aligned}
&\mathbb{P}(\tau > K) \\
&\quad \leq \quad \mathbb{P}(Z_k \neq \overline{Z_k}, Z_{2k} \neq \overline{Z_{2k}}, \ldots, Z_{mk} \neq \overline{Z_{mk}}) \\
&\quad = \quad \mathbb{P}(Z_k \neq \overline{Z_k})\mathbb{P}(Z_{2k} \neq \overline{Z_{2k}} | Z_k \neq \overline{Z_k}) \cdots \mathbb{P}(Z_{mk} \neq \overline{Z_{mk}} | Z_{(m-1)k} \neq \overline{Z_{(m-1)k}}) \\
&\quad \leq \quad (1-\epsilon)^{m-1} \\
&\quad \leq \quad c\alpha^K, \qquad (2.3)
\end{aligned}
$$

where $\alpha = \sqrt[k]{1-\epsilon} \in (0,1)$ and $c = 1/(1-\epsilon)^2$. Therefore, τ is finite with probability one. Choosing $K = \sqrt{n}$, yields $\mathbb{P}(\tau > K) + K/n \to 0$ and finally $\mathbb{E}[LC_n]/n \to \gamma^*$, as $n \to \infty$. Clearly, $\mathbb{E}[LC_n] \leq n$ and to see that $\gamma^* > 0$, note first that, by aperiodicity and irreducibility, $P^k \geq \epsilon$, for some fixed k and $\epsilon > 0$, i.e., all the entries of the matrix P^k are larger than some positive quantity ϵ. Therefore $\mathbb{P}(X_1 = Y_{k+1}) > p$, for some $p = p(k, \epsilon) > 0$. Now,

$$LC_{nk+1} \geq \mathbf{1}_{X_1 = Y_{k+1}} + \mathbf{1}_{X_{k+1} = Y_{2k+1}} + \cdots + \mathbf{1}_{X_{(n-1)k+1} = Y_{nk+1}}, \qquad (2.4)$$

hence

$$\frac{np}{nk+1} \leq \frac{\mathbb{E}[LC_{nk+1}]}{nk+1}.$$

Letting $n \to \infty$ implies that $\gamma^* \in [p/(k+1), 1] \subset (0,1]$, since $p > 0$.

□

Remark 2.1.2. *(i) Under a further assumption, one can show that* $\gamma^* > \mathbb{P}(X_1 = Y_1)$*. Indeed, assume that for all* $x, y \in \mathcal{A}$, $z \in \mathcal{S}$, $\mathbb{P}(X_i = x, Y_i = y | Z_i = z) = \mathbb{P}(X_i = y, Y_i =$

$x|Z_i = z) > 0$, and let Z be started at the stationary distribution. Then for any $n \geq 2$,

$$
\begin{aligned}
\mathbb{E}[LC_n] \geq \quad & \mathbb{E}[LC_{n-2}\mathbf{1}_{X_n=Y_n,X_{n-1}=Y_{n-1}}] + 2\mathbb{P}(X_n = Y_n, X_{n-1} = Y_{n-1}) \\
& + \mathbb{E}[LC_{n-2}\mathbf{1}_{X_n=Y_n,X_{n-1}\neq Y_{n-1}}] + \mathbb{P}(X_n \neq Y_n, X_{n-1} = Y_{n-1}) \\
& + \mathbb{E}[LC_{n-2}\mathbf{1}_{X_n\neq Y_n,X_{n-1}=Y_{n-1}}] + \mathbb{P}(X_n = Y_n, X_{n-1} \neq Y_{n-1}) \\
& + \mathbb{E}[LC_{n-2}\mathbf{1}_{X_n\neq Y_n,X_{n-1}\neq Y_{n-1}}] + \mathbb{P}(X_n \neq Y_n, X_{n-1} \neq Y_{n-1}, X_n = Y_{n-1}) \\
> \quad & \mathbb{E}[LC_{n-2}] + \mathbb{P}(X_n = Y_n) + \mathbb{P}(X_{n-1} = Y_{n-1}) \\
= \quad & \mathbb{E}[LC_{n-2}] + 2\mathbb{P}(X_1 = Y_1),
\end{aligned}
$$

by stationarity. Therefore, iterating, still using stationarity, and since $\mathbb{E}[LC_0] = 0$ while $\mathbb{E}[LC_1] = \mathbb{P}(X_1 = Y_1)$, it follows that for $n \geq 2$, $\mathbb{E}[LC_n] > n\mathbb{P}(X_1 = Y_1)$. Finally,

$$\gamma^* > \mathbb{P}(X_1 = Y_1) = \sum_{\alpha \in \mathcal{A}} \mathbb{P}(X_1 = \alpha)\mathbb{P}(Y_1 = \alpha),$$

and this inequality is strict since Fekete's lemma, e.g., see [27], ensures that

$$\gamma^* = \sup_n \mathbb{E}[LC_n]/n.$$

(ii) Steele's general result, see [26], asserts that Proposition 2.1.1 holds if there is a stationary ergodic coupling for (X, Y). Such an example is when the sequences X and Y are generated by two independent aperiodic, homogeneous and irreducible hidden Markov chains with the same parameters (and so the same emission probabilities). Indeed, at first, when the hidden chains Z_X and Z_Y generating respectively X and Y are started at the stationary distribution, convergence of $\mathbb{E}[LC_n]/n$ towards γ^*, follows from super-additivity and Fekete's lemma (see [27]). As previously, $\gamma^* > 0$, since the properties of the hidden chains imply (2.4)*. Then, when the initial distribution is not the stationary distribution, one can proceed with arguments as above. In particular let τ_1 and τ_2 be the respective meeting times of the chains $(Z_X, \overline{Z_X})$ and $(Z_Y, \overline{Z_Y})$, and let $\tau = \max(\tau_1, \tau_2)$. Then, equation* (2.2)

continues to hold:

$$\left|\mathbb{E}[LCS(X;Y) - LCS(\overline{X};\overline{Y})]\right| \leq n\mathbb{P}(\tau > K) + K$$
$$\leq 2n\mathbb{P}(\tau_1 > K) + K. \qquad (2.5)$$

Taking $K = \sqrt{n}$ and noting the exponential decay of $\mathbb{P}(\tau_1 > K)$ finishes the corresponding proof.

2.2 Rate of convergence

The previous section gives a mean convergence result, we now deal with its rate. Again let (X, Y) be the outcome of a hidden Markov chain Z with μ, P and π as initial distribution, transition matrix and stationary distribution respectively. In this section we impose the additional restriction that the emission distributions for all states in the hidden chain are symmetric (this is discussed further in Proposition 2.2.8 and in the Appendix), namely for all $x, y \in \mathcal{A}$ and all $z \in \mathcal{S}$, $\mathbb{P}(X_i = x, Y_i = y|Z_i = z) = \mathbb{P}(X_i = y, Y_i = x|Z_i = z)$. Symmetry clearly implies that the conditional law of X given Z and of Y given Z are the same since for all x, y and z,

$$\mathbb{P}(X_i = x|Z_i = z) = \sum_{y\in\mathcal{A}} \mathbb{P}(X_i = x, Y_i = y|Z_i = z) = \sum_{y\in\mathcal{A}} \mathbb{P}(X_i = y, Y_i = x|Z_i = z)$$
$$= \mathbb{P}(Y_i = x|Z_i = z).$$

In turn this implies that X_i and Y_i are identically distributed.

Moreover, one needs to control the dependency between X and Y and a way to do so is via the $\beta-$mixing coefficient, as given in Definition 3.3 of [5] which we now recall.

Definition 2.2.1. Let $\mathcal{F}_1$ and $\mathcal{F}_2$ be two $\sigma-$fields $\subset \mathcal{F}$, then the $\beta-$mixing coefficient, associated with these sub-σ-fields of $\mathcal{F}$, is given by:

$$\beta(\mathcal{F}_1, \mathcal{F}_2) \coloneqq \frac{1}{2} \sup \sum_{i=1}^{I} \sum_{j=1}^{J} |\mathbb{P}(A_i \cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j)|,$$

where the supremum is taken over all pairs of finite partitions $\{A_1, \dots, A_I\}$ and $\{B_1, \dots, B_J\}$ of Ω such that $A_i \in \mathcal{F}_1$, for all $i \in \{1, \dots, I\}$, $I \geq 1$ and $B_j \in \mathcal{F}_2$ for all $j \in \{1, \dots, J\}$, $J \geq 1$.

In our case the above notion of $\beta-$mixing coefficient is adopted for the $\sigma-$fields generated by *two* sequences. Moreover, by [5, Proposition 3.21], for a fixed $n \geq 1$, and since $X^{(n)} = (X_1, \dots, X_n)$ and $Y^{(n)} = (Y_1, \dots, Y_n)$ are discrete random vectors,

$$\begin{aligned} \beta(n) \coloneqq \quad & \beta\left(\sigma\left(X^{(n)}\right), \sigma\left(Y^{(n)}\right)\right) \\ = \quad & \frac{1}{2} \sum_{u \in \mathcal{A}^n} \sum_{v \in \mathcal{A}^n} \left|\mathbb{P}\left(X^{(n)} = u, Y^{(n)} = v\right) - \mathbb{P}\left(X^{(n)} = u\right)\mathbb{P}\left(Y^{(n)} = v\right)\right|, \end{aligned} \tag{2.6}$$

where $\sigma\left(X^{(n)}\right)$ and $\sigma\left(Y^{(n)}\right)$ are the $\sigma-$fields generated by $X^{(n)}$ and $Y^{(n)}$. Clearly $X^{(n)}$ and $Y^{(n)}$ are independent if and only if $\beta(n) = 0$. Further, set $\beta^* \coloneqq \lim_{n \to \infty} \beta(n)$, where the limit exists since $\beta(n)$ is non-decreasing, in n, and $\beta(n) \in [0, 1]$ (see Section 5 in [5]).

Remark 2.2.2. *(i) Another definition of $\beta-$mixing coefficient based on "past" and "future" is often studied in the literature, see, for instance, [4, Section 2]. For a single sequence of random variables $S = (S_k)_{k \in \mathbb{Z}}$ and for $-\infty \leq J \leq L \leq \infty$, let*

$$\mathcal{F}_J^L \coloneqq \sigma(S_k, J \leq k \leq L),$$

and for each $n \geq 1$, let

$$\beta_n \coloneqq \sup_{j \in \mathbb{Z}} \beta(\mathcal{F}_{-\infty}^j, \mathcal{F}_{j+n}^{\infty}).$$

In particular [4, Theorem 3.2] implies that if S is a strictly stationary, finite-state Markov

chain that is also irreducible and aperiodic, $\beta_n \to 0$ as $n \to \infty$. The mixing definition relevant to our approach is different and this limiting behavior does not follow. A further discussion of the values of $\beta(n)$ is included in Remark 2.2.6 (i).
(ii) One might also be interested to use the $\alpha-$mixing coefficient defined for $\sigma-$ fields $\mathcal{S}$ and $\mathcal{T}$ as:

$$\alpha(\mathcal{S},\mathcal{T}) = 2\sup\{|Cov(\mathbf{1}_S,\mathbf{1}_T)| : (S,T) \in \mathcal{S} \times \mathcal{T}\}$$

Suppose further that $\mathcal{T}$ has exactly N atoms. The following holds (see [4] and [3, Theorem 1]):

$$2\alpha(\mathcal{S},\mathcal{T}) \leq \beta(\mathcal{S},\mathcal{T}) \leq (8N)^{1/2}\alpha(\mathcal{S},\mathcal{T}).$$

However, for our setting the number of atoms N will be $|\mathcal{A}|^n$, and since $\alpha(n) := \alpha(\sigma(X^{(n)}),\sigma(Y^{(n)}))$ is increasing, a bound on $\beta(n)$ using the inequality above is useless.

The following rate of convergence is our main result:

Theorem 2.2.3. *Let $(Z,(X,Y))$ be a hidden Markov model, where the sequence Z is an aperiodic time homogeneous and irreducible Markov chain with finite state space $\mathcal{S}$. Let the distribution of the pairs (X_i,Y_i), $i = 1,2,3,\dots$, be symmetric for all states in Z. Then, for all $n \geq 2$,*

$$\frac{\mathbb{E}[LC_n]}{n} \geq \gamma^* - 2\beta^* - C\sqrt{\frac{\ln n}{n}} - \frac{2}{n} - (1-\mathbf{1}_{\mu=\pi})\left(\frac{1}{\sqrt{n}} + c\alpha^{\sqrt{n}}\right), \tag{2.7}$$

where $\alpha \in (0,1), c > 0$ are constants as in (2.3) *and $C > 0$. All constants depend on the parameters of the model but not on n. Moreover with the same α and c,*

$$\frac{\mathbb{E}[LC_n]}{n} \leq \gamma^* + (1-\mathbf{1}_{\mu=\pi})\left(\frac{1}{\sqrt{n}} + c\alpha^{\sqrt{n}}\right). \tag{2.8}$$

A key ingredient in proving Theorem 2.2.3 is a Hoeffding-type inequality for Markov chains, a particular case of a result due to Paulin [22], which is now recalled. It relies on the mixing time $\tau(\epsilon)$ of the Markov chain Z given by

$$\tau(\epsilon) := \min\{t \in \mathbb{N} : \overline{d}_Z(t) \le \epsilon\},$$

where

$$\overline{d}_Z(t) := \max_{1\le i\le N-t} \sup_{x,y\in\Lambda_i} d_{TV}(\mathcal{L}(Z_{i+t}|Z_i = x), \mathcal{L}(Z_{i+t}|Z_i = y)),$$

and where $d_{TV}(\mu, \nu) = \frac{1}{2}\sum_{x\in\Omega} |\mu(x) - \nu(x)|$ is the total variation distance between the two probability measures μ and ν on the finite set Ω.

Lemma 2.2.4. *Let $M := (M_1, \ldots, M_N)$ be a (not necessarily time homogeneous) Markov chain, taking values in a Polish space $\Lambda = \Lambda_1 \times \cdots \times \Lambda_N$, with mixing time $\tau(\epsilon)$, $0 \le \epsilon \le 1$. Let*

$$\tau_{min} := \inf_{0\le\epsilon<1} \tau(\epsilon)\left(\frac{2-\epsilon}{1-\epsilon}\right)^2,$$

and let $f : \Lambda \to \mathbb{R}$ be such that there is $c \in \mathbb{R}^N_+$ with $|f(u) - f(v)| \le \sum_{i=1}^N c_i \mathbf{1}_{u_i\ne v_i}$. Then for any $t \ge 0$,

$$\mathbb{P}(f(M) - \mathbb{E}f(M) \ge t) \le \exp\left(\frac{-2t^2}{\tau_{min}\sum_{i=1}^N c_i^2}\right). \tag{2.9}$$

For our purposes, the Hoeffding-type inequality used below follows directly from (3.10) once one notes that $(Z_i, X_i, Y_i)_{i\ge1}$ is jointly a Markov chain on a bigger state space. Let $\tau(\epsilon)$ be the mixing time of this chain. Taking f to be the length of the longest common subsequences of $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ we have $c = ((0, \ldots, 0), (1, \ldots, 1)) \in \mathbb{R}^n \times \mathbb{R}^{2n}$, since f is a function of Z, X and Y, whose values do not depend on Z. Letting $A := \sqrt{\tau_{min}/2}$, (3.10) becomes,

$$\mathbb{P}(LC_n - \mathbb{E}[LC_n] \ge t) \le \exp\left[\frac{-t^2}{A^2 n}\right], \tag{2.10}$$

for all $t \geq 0$.

Remark 2.2.5. *(i) When X and Y are generated by two independent hidden chains Z^X and Z^Y, the same reasoning yields* (2.10) *where now $\tilde{\tau}(\epsilon)$ is the mixing time of the chain $(Z_n^X, Z_n^Y, X_n, Y_n)_{n\geq 1}$.*
(ii) The mixing time $\tau(\epsilon)$ of $(Z_n, X_n, Y_n)_{n\geq 1}$ is the same as the mixing time $\tilde{\tau}(\epsilon)$ of the chain $(Z_n)_{n\geq 1}$. Two proofs of this fact are provided in the Appendix.

Proof of Theorem 2.2.3. First recall a result of Berbee [2], see also [10, Theorem 1, Section 1.2.1], [24, Chapter 5], and [14], asserting that on our probability space, which is rich enough, there exists $Y^{*(n)} := (Y_1^*, \ldots, Y_n^*)$, independent of $(Z, X)^{(n)} = ((Z_1, X_1), \ldots, (Z_n, X_n))$, having the same law as $Y^{(n)} = (Y_1, \ldots, Y_n)$ and such that

$$\mathbb{P}(Y^{(n)} \neq Y^{*(n)}) = \beta(n), \tag{2.11}$$

where $\beta(n) = \beta(\sigma((Z, X)^{(n)}), \sigma(Y^{(n)}))$ is the $\beta-$mixing coefficient of $(Z, X)^{(n)}$ and $Y^{(n)}$. Note also that if $(Y_i)_{i\geq 1}$ is stationary, then $(Y_1^*, \ldots, Y_k^*)$ and $(Y_\ell^*, \ldots, Y_{\ell+k-1}^*)$ are identically distributed, for every $\ell, k \geq 1$, and that if $(X^{(n)}, Y^{(n)})$ is symmetric, then so is $(X^{(n)}, Y^{*(n)})$ where $X^{(n)} = (X_1, \ldots, X_n)$. Note finally that this implies that $Y^{*(n)}$ is independent of both $X^{(n)}$ and $Z^{(n)} = (Z_1, \ldots, Z_n)$.
Next, fix $k \in \mathbb{N}$, the idea of the proof is to relate $\mathbb{E}[LC_{kn}]$ to $\mathbb{E}[LC_{2n}]$. For $k = 4$, this is done in the i.i.d case in [23]. However, we wish to take $k \to \infty$ and therefore follow arguments presented for the i.i.d case in [20]. Call $(\nu, \tau) := (\nu_1, \ldots, \nu_r, \tau_1, \ldots \tau_r)$ an $r-$partition with $k \leq r \leq \lceil 2kn/(2n-1) \rceil$ if

$$\begin{aligned}
&1 = \nu_1 \leq \nu_2 \leq \cdots \leq \nu_{r+1} = kn + 1, \\
&1 = \tau_1 \leq \tau_2 \leq \cdots \leq \tau_{r+1} = kn + 1, \\
&(\nu_{j+1} - \nu_j) + (\tau_{j+1} - \tau_j) \in \{(2n-1, 2n\}, \text{ for } j \in [1, r-1], \\
&(\nu_{r+1} - \nu_r) + (\tau_{r+1} - \tau_r) < 2n.
\end{aligned} \tag{2.12}$$

Let $\mathcal{B}^r_{k,n}$ be the set of all $r-$ partitions defined as above and let

$$\mathcal{B}_{k,n} = \bigcup_{r=k}^{\lceil 2kn/(2n-1) \rceil} \mathcal{B}^r_{k,n}.$$

If (ν, τ) is an $r-$partition, setting

$$LC_{kn}(\nu, \tau) \coloneqq \sum_{i=1}^{r} LCS(X_{\nu_i}, \dots, X_{\nu_{i+1}-1}; Y_{\tau_i}, \dots, Y_{\tau_{i+1}-1}),$$

then:

$$LC_{kn} = \max_{(\nu,\tau)\in\mathcal{B}(k,n)} LC_{kn}(\nu, \tau).$$

Let $\nu_{i+1} - \nu_i = n - m$, $\tau_{i+1} - \tau_i \leq n + m$ for $m \in (-n, n)$ and $\tau_i - \nu_i = \ell$. Then,

$$\begin{aligned}
&\mathbb{E}[LCS(X_{\nu_i}, \dots, X_{\nu_{i+1}-1}; Y_{\tau_i}, \dots, Y_{\tau_{i+1}-1})] \\
&= \quad \mathbb{E}[LCS(X_1, \dots, X_{n-m}; Y_\ell, \dots, Y_{\ell+n+m-1})] && (2.13) \\
&\leq \quad \mathbb{E}\left[LCS(X_1, \dots, X_{n-m}; Y^*_\ell, \dots, Y^*_{\ell+n+m-1})\mathbf{1}_{Y^{(kn)}=Y^{*(kn)}}\right] \\
&\qquad\qquad + \min(n-m, n+m)\mathbb{P}\left(Y^{(kn)} \neq Y^{*(kn)}\right) && (2.14) \\
&\leq \quad \mathbb{E}[LCS(X_1, \dots, X_{n-m}; Y^*_\ell, \dots, Y^*_{\ell+n+m-1})] + n\beta(kn). && (2.15)
\end{aligned}$$

In the last expression the LCS is now a function of two independent sequences. Stationarity implies (2.13) and $LCS(X_1, \dots, X_{n-m}; Y^*_\ell, \dots, Y^*_{\ell+n+m-1}) \leq \min(n-m, n+m)$ entails (2.14). The error term $n\beta(kn)$ in (2.15) follows from an application of Berbee's result (2.11). The same properties also imply

$$\begin{aligned}
&\mathbb{E}[LCS(X_1, \dots, X_{n-m}; Y^*_\ell, \dots, Y^*_{\ell+n+m-1})] \\
&= \quad \mathbb{E}[LCS(X_1, \dots, X_{n-m}; Y^*_1, \dots, Y^*_{n+m})] \\
&\leq \quad \mathbb{E}[LCS(X_1, \dots, X_{n-m}; Y_1, \dots, Y_{n+m})] + n\beta(kn), && (2.16)
\end{aligned}$$

and

$$
\begin{aligned}
&\mathbb{E}[LCS(X_1,\ldots,X_{n-m};Y^*_\ell,\ldots,Y^*_{\ell+n+m-1})] \\
&= \quad \mathbb{E}[LCS(X_1,\ldots,X_{n+m};Y^*_1,\ldots,Y^*_{n-m})] && (2.17)\\
&\leq \quad \mathbb{E}[LCS(X_{n-m+1},\ldots,X_{2n};Y_{n+m+1},\ldots,Y_{2n})] + n\beta(kn), && (2.18)
\end{aligned}
$$

where the symmetry of the distributions of X and Y^* is used to get (2.17). Next by super-additivity of the LCSs as well as (2.15), (2.16) and (2.18),

$$
\begin{aligned}
&\mathbb{E}[LCS(X_{\nu_i},\ldots,X_{\nu_{i+1}-1};Y_{\tau_i},\ldots,Y_{\tau_{i+1}-1})] \\
&\leq \quad \frac{1}{2}\Bigg(\mathbb{E}[LCS(X_1,\ldots,X_{n-m};Y_1,\ldots,Y_{n+m})] \\
&\qquad + \mathbb{E}[LCS(X_{n-m+1},\ldots,X_{2n};Y_{n+m+1},\ldots,Y_{2n})] + 2n\beta(kn)\Bigg) + n\beta(kn) \\
&\leq \quad \frac{1}{2}\Bigg(\mathbb{E}[LC_{2n}] + 2n\beta(kn)\Bigg) + n\beta(kn) \\
&= \quad \frac{1}{2}\mathbb{E}[LC_{2n}] + 2n\beta(kn). && (2.19)
\end{aligned}
$$

This inequality is key to the proof, since it yields an upper bound on $\mathbb{E}[LC_{kn}(\nu,\tau)]$ in terms of $\mathbb{E}[LC_{2n}]$, a quantity that does not depend on the partitioning (ν,τ). A similar result is central to the proof of the rate of convergence in the independent setting [1]. However, independence allows one to get (2.19) directly without the mere presence of or the need to introduce β-mixing coefficients. Moreover, our approach is more direct. Applying Hoeffding's inequality and summing over all partitions provide a relation between $\mathbb{E}[LC_{kn}]$ and $\mathbb{E}[LC_{2n}]$ which can be used to get the rate of convergence. Indeed,

$$
\mathbb{E}[LC_{kn}(\nu,\tau)] \leq \frac{r}{2}(\mathbb{E}[LC_{2n}] + 4n\beta(kn)) \leq \frac{1}{2}\left\lceil \frac{2kn}{2n-1} \right\rceil (\mathbb{E}[LC_{2n}] + 4n\beta(kn)).
$$

In addition, for $t > 0$,

$$\begin{aligned}
&\mathbb{P}\left(LC_{kn}(\nu,\tau) - \frac{1}{2}\left\lceil\frac{2kn}{2n-1}\right\rceil(\mathbb{E}[LC_{2n}] + 4n\beta(kn)) > tkn\right)\\
&\leq \quad \mathbb{P}\left(LC_{kn}(\nu,\tau) - \mathbb{E}[LC_{kn}(\nu,\tau)] > tkn\right)\\
&\leq \quad \exp\left[-\frac{t^2kn}{A^2}\right], \qquad (2.20)
\end{aligned}$$

where the second inequality follows from Lemma 3.2.2. Next note that:

$$\begin{aligned}
&\mathbb{P}\left(LC_{kn} - \frac{1}{2}\left\lceil\frac{2kn}{2n-1}\right\rceil(\mathbb{E}[LC_{2n}] + 4n\beta(kn)) > tkn\right)\\
&= \sum_{(\nu,\tau)\in\mathcal{B}_{k,n}} \mathbb{P}\left(LC_{kn}(\nu,\tau) - \frac{1}{2}\left\lceil\frac{2kn}{2n-1}\right\rceil(\mathbb{E}[LC_{2n}] + 4n\beta(kn)) > tkn\right)\\
&\leq |\mathcal{B}_{k,n}|\exp\left[-\frac{t^2kn}{A^2}\right].
\end{aligned}$$

The above can be rewritten as:

$$\mathbb{P}\left(\frac{LC_{kn}}{kn} > t + \frac{1}{k}\left\lceil\frac{2kn}{2n-1}\right\rceil\left(\frac{\mathbb{E}[LC_{2n}]}{2n} + 2\beta(kn)\right)\right) \leq |\mathcal{B}_{k,n}|\exp\left[-\frac{t^2kn}{A^2}\right].$$

Then, since $LC_{kn} \leq kn$,

$$\begin{aligned}
\mathbb{E}\left[\frac{LC_{kn}}{kn}\right] \leq \quad & t + \frac{1}{k}\left\lceil\frac{2kn}{2n-1}\right\rceil\left(\frac{\mathbb{E}[LC_{2n}]}{2n} + 2\beta(kn)\right)\\
&\qquad + \mathbb{P}\left(\frac{LC_{kn}}{kn} > t + \frac{1}{k}\left\lceil\frac{2kn}{2n-1}\right\rceil\frac{\mathbb{E}[LC_{2n}]}{2n}\right)\\
\leq \quad & t + \frac{1}{k}\left\lceil\frac{2kn}{2n-1}\right\rceil\left(\frac{\mathbb{E}[LC_{2n}]}{2n} + 2\beta(kn)\right) + |\mathcal{B}_{k,n}|\exp\left[-\frac{t^2kn}{A^2}\right]. \qquad (2.21)
\end{aligned}$$

Next a bound on $|\mathcal{B}_{k,n}|$ is obtained using methods as in [20]. Recall that $k \leq r \leq \lceil 2kn/(2n-1)\rceil$ and that $\mathcal{B}_{k,n} = \bigcup_{r=k}^{2kn/2n-1}\mathcal{B}^r_{k,n}$. Now

$$|\mathcal{B}^r_{k,n}| \leq 2^{r-1}2n\binom{nk+r-1}{r-1}. \qquad (2.22)$$

Indeed, the sum of sizes of the partition on the X side should sum to nk which gives a factor of less than $\binom{nk+r-1}{r-1}$. Also for each choice of the first $r-1$ elements of the partition on the X side we have at most 2 choices on the Y side. The last interval can take at most $2n$ values, as per (2.12). Recall Stirling's formula (see [12]), for $n \geq 1$,

$$n^n e^{-n}\sqrt{2\pi n} e^{1/(12n+1)} \leq n! \leq n^n e^{-n}\sqrt{2\pi n} e^{1/12n}.$$

Since in the end of the proof $k \to \infty$, this bound can be used in (2.22) to obtain:

$$\begin{aligned}
|\mathcal{B}^r_{k,n}| &\leq (2^{r-1}2n)\frac{(nk+r-1)^{nk+r-1}\sqrt{2\pi(nk+r-1)}e^{1/12(nk+r-1)}}{(r-1)^(r-1)\sqrt{2\pi(r-1)}e^{1/12(r-1)+1}(nk)^{nk}\sqrt{2\pi nk}e^{1/12(nk)+1}} \\
&\leq 2^r n\frac{(nk+r-1)^{nk+r-1}}{(r-1)^{r-1}(nk)^{nk}} \\
&\leq 2^r n\left(1+\frac{nk}{r-1}\right)^{r-1}\left(1+\frac{2}{2n-1}\right)^{nk} \\
&\leq 2^r n\left(1+n+\frac{n}{k-1}\right)^{\frac{2nk}{2n-1}}\left(\frac{2n+1}{2n-1}\right)^{nk}.
\end{aligned}$$

The last inequality in the above expression holds true since $k \leq r \leq \lceil 2kn/(2n-1)\rceil$. Then for $|\mathcal{B}_{k,n}|$ one gets:

$$\begin{aligned}
|\mathcal{B}_{k,n}| &\leq \left(\frac{2nk}{2n-1}-k+2\right)\max_r |\mathcal{B}^r_{k,n}| \\
&\leq \left(\frac{k}{2n-1}+2\right)2^r n\left(1+n+\frac{n}{k-1}\right)^{\frac{2nk}{2n-1}}\left(\frac{2n+1}{2n-1}\right)^{nk} \\
&\leq \exp\left(\left(\frac{\ln\left(\frac{k}{2n-1}+2\right)}{nk}+\frac{r\ln 2+\ln n}{nk}+\frac{2}{2n-1}\ln(2n)+\ln\left(\frac{2n+1}{2n-1}\right)\right)nk\right) \\
&\leq \exp\left(\left(\frac{\ln k}{k}+\frac{2}{2n-1}\ln 2+\frac{2}{2n-1}\ln(2n)+\ln\left(\frac{2n+1}{2n-1}\right)\right)nk\right) \\
&\leq \exp\left(\left(\frac{\ln k}{k}+\frac{4}{2n-1}\ln 2+\frac{2}{2n-1}\ln n+\ln\left(\frac{2n+1}{2n-1}\right)\right)nk\right) \\
&\leq \exp\left(10k\ln n\right),
\end{aligned}$$

where the last inequality holds for large k, in particular $k > n$, and since $\ln(1+x) \leq x$ for

$x > 0$. Let $t = 2A\sqrt{10}\sqrt{\ln n/n}$. Then,

$$\begin{aligned}|\mathcal{B}_{k,n}| \exp\left(-\frac{t^2kn}{A^2}\right) &\leq \exp\left(10k\ln n\right)\exp\left(-\frac{t^2kn}{A^2}\right)\\ &\leq \exp(-30k\ln n).\end{aligned}$$

Next, note that, as $k \to \infty$, $\mathbb{E}\left[LC_{kn}/(kn)\right] \to \gamma^*$ and that

$$\frac{1}{k}\left\lceil\frac{2kn}{2n-1}\right\rceil \leq \frac{1}{k}\left(\frac{2kn}{2n-1}+1\right) \to \frac{2n}{2n-1}.$$

Recall also that $\beta^* = \lim_{n\to\infty}\beta(n) = \lim_{k\to\infty}\beta(kn)$. Then (2.21) implies:

$$\frac{2n}{2n-1}\left(\frac{\mathbb{E}[LC_{2n}]}{2n}+2\beta^*\right) \geq \gamma^* - 2A\sqrt{10}\sqrt{\frac{\ln n}{n}}, \tag{2.23}$$

and finally:

$$\begin{aligned}\frac{\mathbb{E}[LC_{2n}]}{2n} &\geq \frac{2n-1}{2n}\left(\gamma^* - 2A\sqrt{10}\sqrt{\frac{\ln n}{n}}\right) - 2\beta^*\\ &\geq \gamma^* - 2\beta^* - 2A\sqrt{10}\sqrt{\frac{\ln n}{n}} - \frac{1}{2n}.\end{aligned} \tag{2.24}$$

To get the result for words of odd length note that by (2.23),

$$\begin{aligned}\frac{\mathbb{E}[LC_{2n+1}]}{2n+1} &\geq \frac{\mathbb{E}[LC_{2n}]}{2n+1}\\ &\geq \frac{2n-1}{2n+1}\left(\gamma^* - 2A\sqrt{10}\sqrt{\frac{\ln n}{n}}\right) - \frac{2n}{2n+1}2\beta^*\\ &\geq \gamma^* - 2\beta^* - 2A\sqrt{10}\sqrt{\frac{\ln n}{n}} - \frac{2}{2n+1}.\end{aligned}$$

Of course, these last bounds are only of interest, for n large enough, if $\gamma^* > 2\beta^*$. Otherwise, we get the trivial lower bound 0 (see Remark 2.2.6 below). One is then left with slightly modifying the constants to get (2.7). The extra term on the right hand side in (2.7) accounts for the difference in initial distributions (2.2).

The proof of the upper bound (2.8), where symmetry is not needed, follows by combining Fekete's lemma (see [27]) with (2.2) and (2.3).

□

Remark 2.2.6. *(i) Recall that the $\beta-$mixing coefficient $\beta(n)$ is a measure on the dependency between $(X_1, \dots, X_n)$ and $(Y_1, \dots, Y_n)$. The bounds in Theorem 2.2.3 rely on $\beta^* := \lim_{n\to\infty} \beta(n)$ which somehow quantifies a weak dependency requirement and $\beta^* \neq 0$ unless the sequences X and Y are independent. Note also that the lower bound in Theorem 2.2.3 is meaningful only if $2\beta^* < \gamma^*$. Besides the independent case, there are instances for which this condition is satisfied. For example, let X and Y be both Markov chains with L states and with the same transition matrix P, where some rows of P are equal to $(1, 1, 1, \dots, 1)/L$, i.e., such that there exists a set of states $\mathcal{L}$ such that the transition probability between each one of these states is uniform. Let the initial distribution of X_1 be μ with $\mu(x) = 0$ if $x \notin \mathcal{L}$ and assume that $Y_1 = X_1$. Then the sequence $\tilde{Y}$ defined, for all n, via $\tilde{Y}_i = Y_i$, for $i \geq 1$ while Y_1 is distributed according to μ will be such that $\tilde{Y}^{(n)}$ and $Y^{(n)}$ have the same distribution. Moreover for all n, $\tilde{Y}^{(n)}$ and $X^{(n)}$ will be independent and $\mathbb{P}(\tilde{Y}^{(n)} \neq Y^{(n)}) \geq \beta(n)$, but $\mathbb{P}(\tilde{Y}^{(n)} \neq Y^{(n)}) = \mathbb{P}(Y_1 \neq \tilde{Y}_1)$ which can be made as small as desired for a suitable choice of μ. Thus the lower bound in Theorem 2.2.3 holds and is meaningful.*

(ii) There are instances when the lower bound in Theorem 2.2.3 is vacuous. Such a case is when $X_i = Y_i$ for all $i \geq 1$ and the X_i are independent and uniformly distributed over the letters in $\mathcal{A}$. Then, it is clear that $\gamma^* = 1$ whereas one shows that

$$\beta(n) = 1 - \frac{1}{|\mathcal{A}|^n},$$

and so $\beta^* = 1$. In this case the lower bound in (2.7) *is a negative quantity.*

(iii) Theorem 2.2.3 continues to hold for Markov chains with a general state space Λ.

Indeed, the Hoeffding inequality (2.10) *is true when Λ is a Polish space. The exponential decay* (2.3) *holds when Λ is* petite*, i.e., when there exist a positive integer n_0, $\epsilon > 0$ and a probability measure ν on Λ such that $P^{n_0}(x, A) \geq \epsilon\nu(A)$, for every measurable A and $x \in \Lambda$, and where $P^{n_0}(x, A)$ is the n_0-step transition law of the Markov chain (see [25, Theorem 8]).*

When X and Y are generated by independent hidden Markov models. Then the following variant of Theorem 2.2.3 holds (for a sketch of proof, see the Appendix).

Corollary 2.2.7. *Let (Z_X, X) and (Z_Y, Y) be two independent hidden Markov models, where the latent chains Z_X and Z_Y have the same initial distribution, transition matrix and emission probabilities. Then, for all $n \geq 2$,*

$$\frac{\mathbb{E}[LC_n]}{n} \geq \gamma^* - C\sqrt{\frac{\ln n}{n}} - \frac{2}{n} - (1 - \mathbf{1}_{\mu=\pi})\left(\frac{1}{\sqrt{n}} + c\alpha^{\sqrt{n}}\right), \tag{2.25}$$

where $\alpha \in (0, 1), c > 0$ are constants as in (2.3) *and $C > 0$. All constants depend on the parameters of the model but not on n. Moreover with the same α and c,*

$$\frac{\mathbb{E}[LC_n]}{n} \leq \gamma^* + (1 - \mathbf{1}_{\mu=\pi})\left(\frac{1}{\sqrt{n}} + c\alpha^{\sqrt{n}}\right). \tag{2.26}$$

As mentioned in the end of the proof of Theorem 2.2.3, the symmetry of the distribution of (X_i, Y_i) is used only for proving the lower bound. Let

$$h(n) := \max_{m\in[-n,n]} \left(2\sum_{i=1}^{n-m} \mathbb{P}(X_i \neq Y_i) + \sum_{i=n-m+1}^{n+m} \mathbb{P}(X_i \neq Y_i)\right).$$

Then the following holds:

Proposition 2.2.8. *Let $(Z, (X, Y))$ be a hidden Markov model, where the sequence Z is an aperiodic time homogeneous and irreducible Markov chain with finite state space $\mathcal{S}$. Then,* *for all* $n \geq 2$,

$$\frac{\mathbb{E}[LC_n]}{n} \geq \gamma^* - \frac{h(n)}{n} - 2\beta^* - C\sqrt{\frac{\ln n}{n}} - \frac{2}{n} - (1 - \mathbf{1}_{\mu=\pi})\left(\frac{1}{\sqrt{n}} + c\alpha^{\sqrt{n}}\right). \tag{2.27}$$

For a sketch of proof of this proposition, and some comments on $h(n)$, we again refer the reader to the Appendix.

2.3 Appendix

First, as asserted in Remark 2.2.5 (ii), we provide two proofs of the fact that the mixing time $\tau(\epsilon)$ of $(Z_n, X_n, Y_n)_{n\geq 1}$ is the same as the mixing time $\tilde{\tau}(\epsilon)$ of the chain $(Z_n)_{n\geq 1}$.

Proof 1. Let $\tilde{T} = (\tilde{T}_n)_{n\geq 1}$ be a Markov chain with finite state space $\mathcal{S}$. Each $\tilde{T}_i$ emits an observed variable T_i according to some probability distribution that depends only on the state $\tilde{T}_i$. Let $T = (T_n)_{n\geq 1}$ and assume $T_i \in \mathcal{A}$ - a finite alphabet. Note that $(\tilde{T}, T)$ is a Markov chain; let $\tau(\epsilon)$ be its mixing time, and let $\tilde{\tau}(\epsilon)$ be the mixing time for the hidden chain $\tilde{T}$. Then,

$$\begin{aligned}
&d_{TV}(\mathcal{L}((\tilde{T}_{i+t}, T_{i+t})|(\tilde{T}_i, T_i) = (x,u)), \mathcal{L}((\tilde{T}_{i+t}, T_{i+t})|(\tilde{T}_i, T_i) = (y,v))) \\
&= \frac{1}{2}\sum_{(z,w)\in\mathcal{S}\times\mathcal{A}} \Big| \mathbb{P}((\tilde{T}_{i+t}, T_{i+t}) = (z,w)|(\tilde{T}_i, T_i) = (x,u)) - \\
&\qquad\qquad - \mathbb{P}((\tilde{T}_{i+t}, T_{i+t}) = (z,w)|(\tilde{T}_i, T_i) = (y,v) \Big| \\
&= \frac{1}{2}\sum_{(z,w)} \Big| \mathbb{P}(\tilde{T}_{i+t} = z|\tilde{T}_i = x)\mathbb{P}(z \to w) - \mathbb{P}(\tilde{T}_{i+t} = z|\tilde{T}_i = y)\mathbb{P}(z \to w) \Big| \\
&= \frac{1}{2}\sum_{(z,w)} \mathbb{P}(z \to w) \Big| \mathbb{P}(\tilde{T}_{i+t} = z|\tilde{T}_i = x) - \mathbb{P}(\tilde{T}_{i+t} = z|\tilde{T}_i = y) \Big| \\
&= \frac{1}{2}\sum_{z\in\mathcal{S}} \Big| \mathbb{P}(\tilde{T}_{i+t} = z|\tilde{T}_i = x) - \mathbb{P}(\tilde{T}_{i+t} = z|\tilde{T}_i = y) \Big| \\
&= d_{TV}(\mathcal{L}(\tilde{T}_{i+t}|\tilde{T}_i = x), \mathcal{L}(\tilde{T}_{i+t}|\tilde{T}_i = y)),
\end{aligned}$$

where $\mathbb{P}(z \to w) := \mathbb{P}(T_i = w|\tilde{T}_i = z)$, i.e., the probability that a state with value $z \in \mathcal{S}$ emits $w \in \mathcal{A}$. By definition of T and $\tilde{T}$ this last probability does not depend on i. Then $\sum_{w\in\mathcal{A}} \mathbb{P}(z \to w) = 1$. Therefore, $\overline{d}_{(\tilde{T},T)}(t) = \overline{d}_{\tilde{T}}(t)$ and $\tau(\epsilon) = \tilde{\tau}(\epsilon)$.

$\square$

Before we present the second proof, we recall the following classical result [21, Proposition 4.7]:

Lemma 2.3.1. *Let μ and ν be two probability distributions on Ω. Then,*

$$d_{TV}(\mu, \nu) = \inf\{\mathbb{P}(X \neq Y) : (X, Y) \textit{ is a coupling of } \mu \textit{ and } \nu\}.$$

Moreover, there is a coupling (X, Y) which attains the infimum and such a coupling is called optimal.

Proof 2. An alternative approach to proving the result of Remark 2.2.5 (ii) relies on coupling arguments and was kindly suggested by D. Paulin in personal communications with the authors.

Let $(\tilde{T}^1, \tilde{T}^2)$ be an optimal coupling according to $d_{TV}(\mathcal{L}(\tilde{T}_t|\tilde{T}_1 = x), \mathcal{L}(\tilde{T}_t|\tilde{T}_1 = y))$ for some $x, y \in \mathcal{S}$, i.e., $\tilde{T}^1$ and $\tilde{T}^2$ are Markov chains with the same transition probability as $\tilde{T}$, $\tilde{T}^1_0 = x$, $\tilde{T}^2_0 = y$, and

$$\mathbb{P}(\tilde{T}^1_t \neq \tilde{T}^2_t) = \quad d_{TV}(\mathcal{L}(\tilde{T}_t|\tilde{T}_1 = x), \mathcal{L}(\tilde{T}_t|\tilde{T}_1 = y)) \tag{2.28}$$

Next let T^1_t and T^2_t be respectively distributed according to the distributions associated with $\tilde{T}^1_t$ and $\tilde{T}^2_t$ and be independent of all the other random variables. In addition, if for some $t \geq 1$, $\tilde{T}^1_t = \tilde{T}^2_t$, then $T^1_t = T^2_t$. Then

$$\mathbb{P}(\tilde{T}^1_t \neq \tilde{T}^2_t) = \mathbb{P}\left((\tilde{T}^1_t, T^1_t) \neq (\tilde{T}^2_t, T^2_t)\right),$$

and by Lemma 2.3.1, for any $u, v \in \mathcal{A}$ and any $i \geq 1$,

$$\begin{aligned}
&\mathbb{P}\left((\tilde{T}_t^1, T_t^1) \neq (\tilde{T}_t^2, T_t^2)\right) \\
\geq \quad & d_{TV}\left(\mathcal{L}((\tilde{T}_{i+t}, T_{i+t})|(\tilde{T}_i, T_i) = (x, u)), \mathcal{L}((\tilde{T}_{i+t}, T_{i+t})|(\tilde{T}_i, T_i) = (y, v))\right).
\end{aligned}$$

Together with (2.28), the above yields

$$\begin{aligned}
&d_{TV}(\mathcal{L}(\tilde{T}_t|\tilde{T}_1 = x), \mathcal{L}(\tilde{T}_t|\tilde{T}_1 = y)) \geq \\
\geq \quad & d_{TV}\left(\mathcal{L}((\tilde{T}_{i+t}, T_{i+t})|(\tilde{T}_i, T_i) = (x, u)), \mathcal{L}((\tilde{T}_{i+t}, T_{i+t})|(\tilde{T}_i, T_i) = (y, v))\right)
\end{aligned}$$

Taking the sup over x, y, u, v gives $\overline{d}_{(\tilde{T},T)}(t) \leq \overline{d}_{\tilde{T}}(t)$.

For the reverse inequality, consider the optimal coupling $\left((\tilde{T}^1, T^1), (\tilde{T}^2, T^2)\right)$ according to $d_{TV}\left(\mathcal{L}((\tilde{T}_t, T_t)|(\tilde{T}_1, T_1) = (x, u)), \mathcal{L}((\tilde{T}_t, T_t)|(\tilde{T}_1, T_1) = (y, v))\right)$, for some $x, y \in \mathcal{S}$ and $u, v \in \mathcal{A}$. Then,

$$\mathbb{P}\left((\tilde{T}_t^1, T_t^1) \neq (\tilde{T}_t^2, T_t^2)\right) = d_{TV}\left(\mathcal{L}((\tilde{T}_t, T_t)|(\tilde{T}_1, T_1) = (x, u)), \mathcal{L}((\tilde{T}_t, T_t)|(\tilde{T}_1, T_1) = (y, v))\right), \tag{2.29}$$

and

$$\mathbb{P}\left((\tilde{T}_t^1, T_t^1) \neq (\tilde{T}_t^2, T_t^2)\right) \geq \mathbb{P}(\tilde{T}_t^1 \neq \tilde{T}_t^2).$$

However, by the Lemma 2.3.1, for any $i \geq 1$,

$$\mathbb{P}(\tilde{T}_t^1 \neq \tilde{T}_t^2) \geq d_{TV}(\mathcal{L}(\tilde{T}_{i+t}|\tilde{T}_i = x), \mathcal{L}(\tilde{T}_{i+t}|\tilde{T}_i = y)). \tag{2.30}$$

Taking the sup in (2.29) and (2.30) gives, $\overline{d}_{(\tilde{T},T)}(t) \geq \overline{d}_{\tilde{T}}(t)$, and then $\overline{d}_{(\tilde{T},T)}(t) = \overline{d}_{\tilde{T}}(t)$.

$\square$

Proof of Corollary 2.2.7. The Hoeffding inequality (2.10) holds as long as (Z, X, Y) is a Markov chain. In addition, (X, Y) has to be symmetric (see proof of Proposition 2.2.8) in order for (2.19) to hold. Again one such setting is when X and Y are two independent HMM with the same transition matrix for the latent chain and same emission probabilities. A rate of convergence result then follows from arguments as in Section 2.2. The bound on $\mathcal{B}_{k,n}$ is the same, and there is a Hoeffding type inequality for this model as per Remark 2.2.5 (i). One thing that differs is the bound (2.19), which is now much easier to get. When started at the stationary distribution, by stationarity, independence and symmetry, one has:

$$\begin{aligned} LCS(X_{\nu_i}, \ldots, X_{\nu_{i+1}-1}; Y_{\tau_i}, \ldots, Y_{\tau_{i+1}-1}) &\leq LCS(X_1, \ldots, X_{n-m}, Y_1, \ldots, Y_{n+m}) \\ &= LCS(X_1, \ldots, X_{n+m}; Y_1, \ldots, Y_{n-m}) \\ &\leq \frac{1}{2} LC_{2n}. \end{aligned}$$

In particular, there is no need to introduce mixing coefficients in this case ($\beta = 0$). When the hidden chains are not started at the stationary distribution one gets an error as in (2.5). Then Theorem 2.2.3 holds but with constants depending on the new model. Moreover, this setting reduces to the one where X and Y are independent Markov chains by letting each state of the hidden chains emit a unique letter, which can further recover the iid case originally obtained in [1].

$\square$

Proof of Proposition 2.2.8. The symmetry of the distribution of (X, Y) is only used to get (2.17), which entails that for any $m \in \{-n+1, \ldots, n-1\}$, $LCS(X_1, \ldots, X_{n-m}; Y_1, \ldots, Y_{n+m})$ and $LCS(X_1, \ldots, X_{n+m}; Y_1, \ldots, Y_{n-m})$ are identically distributed and upper bounded by half of LC_{2n}. Such a result yields a comparison between $\mathbb{E}[LC_{2n}]$ and $\mathbb{E}[LC_{kn}]$, leading as $k \to \infty$, to a lower bound on $\mathbb{E}[LC_{2n}]$ involving γ^*. Without assuming symmetry, the step (2.17) in obtaining (2.19) needs to be modified. One way to do so is to make use of the Lipschitz property of the LCS to get the following estimate:

$$
\begin{aligned}
&LCS(X_1,\dots,X_{n-m};Y_1,\dots,Y_{n+m})\\
&\quad= \quad LCS(X_1,\dots,X_{n+m};Y_1,\dots,Y_{n-m}) + \Big(LCS(X_1,\dots,X_{n-m};Y_1,\dots,Y_{n+m})\\
&\qquad\qquad - LCS(X_1,\dots,X_{n+m};Y_1,\dots,Y_{n-m})\Big)\\
&\quad\le \quad LCS(Y_1,\dots,Y_{n-m};X_1,\dots,X_{n+m}) + 2\sum_{i=1}^{n-m}\mathbf{1}_{X_i\neq Y_i} + \sum_{i=n-m+1}^{n+m}\mathbf{1}_{X_i\neq Y_i}.
\end{aligned}
$$

Taking expectations, then (2.19) becomes

$$\mathbb{E}[LCS(X_{\nu_i},\dots,X_{\nu_{i+1}-1};Y_{\tau_i},\dots,Y_{\tau_{i+1}-1})] \le \frac{1}{2}\Big(\mathbb{E}[LC_{2n}] + h(n)\Big) + 2n\beta(kn),$$

where $h(n) := \max_{m\in[-n,n]}\left(2\sum_{i=1}^{n-m}\mathbb{P}(X_i\neq Y_i) + \sum_{i=n-m+1}^{n+m}\mathbb{P}(X_i\neq Y_i)\right)$. This leads to a non-symmetric version of (2.7), namely,

$$\frac{\mathbb{E}[LC_n]}{n} \ge \gamma^* - C\sqrt{\frac{\ln n}{n}} - \frac{h(n)}{n} - 2\beta^* - \frac{1}{n-2} - (1-\mathbf{1}_{\mu=\pi})\left(\frac{1}{\sqrt{n}} + c\alpha^{\sqrt{n}}\right). \quad (2.31)$$

$\square$

If $h(n) = O(\sqrt{n\ln n})$, then the rate in (2.27) or (2.31) will be the same as in (2.7). Such will be the case when (Z', X) and (Z'', Y) are two independent hidden Markov models and $Z = (Z', Z'')$ is a coupling of the two latent chains such that if $Z_i' = Z_i''$, then $Z_j' = Z_j''$ for any $j > i$. Then, $(Z, (X, Y))$ is a hidden Markov model where $X_i = Y_i$ once the two latent chains have met, and by (2.3) $h(n) = O(\sqrt{n\log n})$.

However, $h(n)$ can be much larger, e.g, of order n. A case in hand is when the X_i and Y_i are iid Bernoulli random variables with parameters $1/3$ and $1/2$ respectively. Then $\mathbb{P}(X_i\neq Y_i) = \mathbb{P}(X_i = 0, Y_i = 1) + \mathbb{P}(X_i = 1, Y_i = 0) = 1/6 + 2/6 = 1/2$, for all $i \ge 1$,

and

$$\left(2\sum_{i=1}^{n-m}\mathbb{P}(X_i \neq Y_i) + \sum_{i=n-m+1}^{n+m}\mathbb{P}(X_i \neq Y_i)\right) = \big(2(n-m)1/2 + (2m)1/2\big) = n.$$

Note also that when $X = (X_i)_{i\geq 1}$ and $Y = (Y_i)_{i\geq 1}$ are independent sequences of random variables, the symmetry assumption is equivalent to X and Y being identically distributed.

CHAPTER 3

CLOSENESS TO THE DIAGONAL

3.1 Preliminaries

Let (Z, X, Y) be a hidden Markov model where Z a hidden irreducible aperiodic and time homogeneous Markov chain on a finite state $\mathcal{S}$. At each step Z_i the chain outputs a pair (X_i, Y_i) taking values in $\mathcal{A} \times \mathcal{A}$. Assume the output distribution is symmetric and there exist $0 < p_L \leq p_U < 1$ such that for all $x, y \in \mathcal{A}, s \in \mathcal{S}$, $\mathbb{P}(X_i = x, Y_i = y | Z_i = s) = \mathbb{P}(X_i = y, Y_i = x | Z_i = s) \in [p_L, p_U]$. Let Z be started at the stationary distribution.

Lemma 3.1.1. *There is $\epsilon_M > 0$ such that for $0 < \epsilon < \epsilon_M$,*

$$\mathbb{P}(LCS(X_1, \ldots, X_{2n-\epsilon n}; Y_1, \ldots, Y_{\epsilon n}) = \lfloor \epsilon n \rfloor) \geq 1 - c^n, \tag{3.1}$$

where $c \in (0, 1)$ depends on the parameters of the model.

Proof. Let $T_0 = \epsilon n$, and $T_i = \inf\{k > T_{i-1} : X_k = Y_i\}$ for $i = 1, \ldots, \epsilon n$. Then

$$\mathbb{P}(LCS(X_1, \ldots, X_{2n-\epsilon n}; Y_1, \ldots, Y_{\epsilon n}) = \lfloor \epsilon n \rfloor) = 1 - \mathbb{P}(T_{\epsilon n} \geq (2 - \epsilon)n).$$

Moreover,

$$\begin{aligned}
\mathbb{P}(T_{\epsilon n} \geq (2-\epsilon)n) = \quad & \mathbb{P}\left(T_{\epsilon n} - T_0 \geq (2 - 2\epsilon)n\right) \\
= \quad & \mathbb{P}\left(\sum_{k=1}^{\epsilon n} (T_k - T_{k-1}) \geq (2-2\epsilon)n\right) \\
\leq \quad & \min_{t>0} \frac{\mathbb{E}[\exp(t \sum_{k=1}^{\epsilon n} (T_k - T_{k-1}))]}{\exp(t(2-2\epsilon)n)} \\
= \quad & \min_{t>0} \frac{\mathbb{E}[\mathbb{E}[\exp(t \sum_{k=1}^{\epsilon n} (T_k - T_{k-1}))|Z]]}{\exp(t(2-2\epsilon)n)}.
\end{aligned}$$

Clearly, given Z, the variables $T_i - T_{i-1}$ are independent for $i = 1, \ldots, \epsilon n$ and

$$\mathbb{P}(T_k - T_{k-1} = x | Z) \leq (1 - p_L)^{x-1} p_U = (1 - p_L)^{x-1} p_L r,$$

where $r \coloneqq p_U / p_L \geq 1$ (see Remark 3.1.2). Then

$$\begin{aligned}
\frac{\mathbb{E}[\mathbb{E}[\exp(t \sum_{k=1}^{\epsilon n} t(T_k - T_{k-1}))|Z]]}{\exp(t(2-2\epsilon)n)} &= \frac{\mathbb{E}[\prod_{k=1}^{\epsilon n} \mathbb{E}[\exp t(T_k - T_{k-1})|Z]]}{\exp(t(2-2\epsilon)n)} \\
&\leq \frac{\mathbb{E}\left[\prod_{k=1}^{\epsilon n} \left(\sum_{x=0}^{\infty} e^{tx}(1-p_L)^{x-1} p_L r\right)\right]}{\exp(t(2-2\epsilon)n)} \\
&= \frac{r^{\epsilon n} p_L^{\epsilon n} e^{t\epsilon n}}{(1-(1-p_L)e^t)^{\epsilon n} e^{t(2n-2\epsilon n)}}. \qquad (3.2)
\end{aligned}$$

We now wish to minimize the right-hand side as a function of t where $t \in (0, -\ln(1-p_L))$. Computing the derivative and setting it equal to 0,

$$e^{t(3\epsilon n - 2n)}(3\epsilon n - 2n)(1-(1-p_L)e^t)^{\epsilon n} - e^{t(3\epsilon n-2n)}(1-(1-p_L)e^t)^{\epsilon n-1}\epsilon n(p_L - 1)e^t = 0.$$

Then,

$$3\epsilon - 2 - e^t(1-p_L)(3\epsilon - 2) - \epsilon(p_L - 1)e^t = 0,$$

or

$$e^t = \frac{3\epsilon - 2}{(2\epsilon - 2)(1-p_L)} = \frac{\epsilon}{(2\epsilon-2)(1-p_L)} + \frac{1}{1-p_L}.$$

The boundary points $t = 0$ and $t = -\ln(1-p_L)$ give bounds of 1 and ∞ respectively. So the minimum is indeed at $e^t = (3\epsilon - 2)/((2\epsilon-2)(1-p_L))$. Recall that $t > 0$ so $2 - 3\epsilon > (2-2\epsilon)(1-p_L)$, i.e., $\epsilon < 2p_L/(2p_L + 1)$.

The bound (3.2) then becomes:

$$\begin{aligned}\frac{\mathbb{E}[\exp(t\sum_{k=1}^{\epsilon n}(T_k - T_{k-1}))]}{\exp(t(2-\epsilon)n)} &\leq \frac{(rp_L)^{\epsilon n}\left(\frac{3\epsilon-2}{(2\epsilon-2)(1-p_L)}\right)^{3\epsilon n-2n}}{\left(\frac{\epsilon}{(2-2\epsilon)}\right)^{\epsilon n}} \\ &= \left(\frac{rp_L}{(1-p_L)^3}\frac{(2-3\epsilon)^3}{(2-2\epsilon)^2\epsilon}\right)^{\epsilon n}\left(\frac{2-3\epsilon}{(2-2\epsilon)(1-p_L)}\right)^{-2n}.\end{aligned}$$

Since p is a fixed constant, the above is a function of ϵ (and $\epsilon < 2p_L/(2p_L+1)$). We want to show that asymptotically it behaves like c^n, for $c < 1$. Then we need to show that $f(\epsilon) < 1$ where:

$$\begin{aligned}f(\epsilon) &\coloneqq \left(\frac{rp_L}{(1-p_L)^3}\frac{(2-3\epsilon)^3}{(2-2\epsilon)^2\epsilon}\right)^{\epsilon}\left(\frac{2-3\epsilon}{(2-2\epsilon)(1-p_L)}\right)^{-2} \\ &= \left(\frac{rp_L(2-2\epsilon)}{\epsilon}\right)^{\epsilon}\left(\frac{(1-p_L)(2-2\epsilon)}{(2-3\epsilon)}\right)^{2-3\epsilon}.\end{aligned}$$

The function f is continuous and positive for $\epsilon \in (0, 2p_L/(2p_L+1)]$. Moreover:

$$\begin{aligned}\frac{f'(\epsilon)}{f(\epsilon)} &= \ln\left(\frac{rp_L(2-2\epsilon)}{\epsilon}\right) + \epsilon\frac{\epsilon}{rp_L(2-2\epsilon)}\frac{-2\epsilon rp_L - rp_L(2-2\epsilon)}{\epsilon^2} \\ &\quad - 3\ln\left(\frac{(1-p_L)(2-2\epsilon)}{(2-3\epsilon)}\right) \\ &\quad + (2-3\epsilon)\frac{2-3\epsilon}{(1-p_L)(2-2\epsilon)}\frac{2(p_L-1)(2-3\epsilon)-(1-p_L)(2-2\epsilon)(-3)}{(2-3\epsilon)^2} \\ &= \ln\left(\frac{rp_L(2-2\epsilon)(2-3\epsilon)^3}{\epsilon(1-p_L)^3(2-2\epsilon)^3}\right) + \frac{-2}{2-2\epsilon} + \frac{2}{2-2\epsilon} \\ &= \ln\left(\frac{rp_L(2-2\epsilon)(2-3\epsilon)^3}{\epsilon(1-p_L)^3(2-2\epsilon)^3}\right).\end{aligned}$$

Now note that

$$\frac{(2-3\epsilon)^3}{(1-p_L)^3(2-2\epsilon)^3} = e^{3t} > 1,$$

and

$$\frac{rp_L(2-2\epsilon)}{\epsilon} > \frac{p_L(2-\frac{2p_L}{2p_L+1})}{\frac{2p_L}{2p_L+1}} = p_L + 1 > 1.$$

Therefore, $f'(\epsilon)/f(\epsilon) > \ln 1 = 0$ and so $f(\epsilon)$ is a strictly increasing function for $\epsilon \in (0, 2p_L/(2p_L+1))$. Finally note that $f(2p_L/(2p_L+1)) = r$ and $f(0^+) = (1-p_L)^2$. Then there is $\epsilon_M \in (0, 2p_L/(2p_L+1)]$ such that $f(\epsilon_M) = 1$. Hence for $\epsilon < \epsilon_M$,

$$\mathbb{P}\left(T_{\epsilon n} \geq (2-\epsilon)n\right) \leq f(\epsilon)^n = c^n,$$

where $c := f(\epsilon) < 1$ and then (3.1) follows.

□

Remark 3.1.2. *When X and Y are independent sequences of iid random variables, the variables $T_k - T_{k-1}$, $k = 1, \ldots, \epsilon n$ are independent geometric with parameters depending on Y_k (since the distribution for the letters in the X and Y sequences may not be uniform). Then since the moment generating function M is decreasing in p, the bound* (3.2) *holds with $r = 1$. Therefore in the independent case* (3.1) *holds whenever $\epsilon < 2p_L/(2p_L+1)$.*

Now for $p > 0$ and $q \in (-1, 1)$ let

$$\tilde{\gamma}_n(p) := \frac{\mathbb{E}[LCS(X_1, \ldots, X_n; Y_1, \ldots, Y_{np})]}{n(1+p)/2},$$

$$\gamma_n(q) := \frac{\mathbb{E}[LCS(X_1, \ldots, X_{n-nq}; Y_1, \ldots, Y_{n+nq})]}{n}.$$

One is interested in the limits $\tilde{\gamma}(p) := \lim_{n\to\infty} \tilde{\gamma}_n(p)$ and $\gamma(q) := \lim_{n\to\infty} \gamma_n(q)$. When $p = 1$ or $q = 0$, the limit exists, see [17], and is denoted by γ^*. Furthermore, $\gamma_n(0) \leq \gamma^*$ and $\tilde{\gamma}_n(1) \leq \gamma^*$ for all n. Note that $\tilde{\gamma}_n(p)$ (resp. $\gamma_n(q)$) is symmetric about 1 (resp. 0) since, by the symmetry assumption on the output distribution, each output (a, b) from a hidden state is equally likely to be (b, a).

Remark 3.1.3. *For $p \neq 1$ (respectively $q \neq 0$) Fekete's theorem fails to apply directly unless X and Y are generated by independent hidden chains. Then one can say further that γ_n (respectively $\tilde{\gamma}_n$) is concave and attains a maximum at 1 (respectively 0).*

Lemma 3.1.4. *The limit $\gamma^* \in (\gamma_\ell, \gamma_u]$, where*

$$\gamma_\ell \coloneqq \sum_{\alpha \in \mathcal{A}} \mathbb{P}(X_1 = \alpha)\mathbb{P}(Y_1 = \alpha) > 0,$$

and

$$\gamma_u \coloneqq 1 - \sum_{\substack{a,b,c,d \in \mathcal{A} \\ b \neq d, b \neq c, a \neq d}} \mathbb{P}(X_0 = a, X_1 = b, Y_0 = c, Y_1 = d) < 1.$$

Let $\delta \in [0, \gamma^*)$. For any $p_1 \in (0, (\gamma^* - \delta)/(2 - \gamma^* + \delta))$ and $p_2 > (2 - \gamma^* + \delta)/(\gamma^* - \delta)$, and any n,

$$\tilde{\gamma}_n(p_1) < \gamma^* - \delta, \qquad \tilde{\gamma}_n(p_2) < \gamma^* - \delta. \tag{3.3}$$

Proof. For any $n \geq 2$,

$$\begin{aligned}
\mathbb{E}[LC_n] \geq \quad & \mathbb{E}[LC_{n-2}\mathbf{1}_{X_n = Y_n, X_{n-1} = Y_{n-1}}] + 2\mathbb{P}(X_n = Y_n, X_{n-1} = Y_{n-1}) \\
& + \mathbb{E}[LC_{n-2}\mathbf{1}_{X_n = Y_n, X_{n-1} \neq Y_{n-1}}] + \mathbb{P}(X_n \neq Y_n, X_{n-1} = Y_{n-1}) \\
& + \mathbb{E}[LC_{n-2}\mathbf{1}_{X_n \neq Y_n, X_{n-1} = Y_{n-1}}] + \mathbb{P}(X_n = Y_n, X_{n-1} \neq Y_{n-1}) \\
& + \mathbb{E}[LC_{n-2}\mathbf{1}_{X_n \neq Y_n, X_{n-1} \neq Y_{n-1}}] + \mathbb{P}(X_n \neq Y_n, X_{n-1} \neq Y_{n-1}, X_n = Y_{n-1}) \\
> \quad & \mathbb{E}[LC_{n-2}] + \mathbb{P}(X_n = Y_n) + \mathbb{P}(X_{n-1} = Y_{n-1}) \\
= \quad & \mathbb{E}[LC_{n-2}] + 2\mathbb{P}(X_1 = Y_1).
\end{aligned}$$

Therefore by stationarity, for $n \geq 2$, $\mathbb{E}[LC_n] > n\mathbb{P}(X_1 = Y_1)$. Then

$$\gamma^* > \mathbb{P}(X_1 = Y_1) = \sum_{\alpha \in \mathcal{A}} \mathbb{P}(X_1 = \alpha)\mathbb{P}(Y_1 = \alpha).$$

The inequality is strict because Fekete's lemma implies that $\gamma^* = \sup_n \mathbb{E}[LC_n]/n$.

On the other hand, if A is the event $\{X_n \neq Y_n, X_{n-1} \neq Y_n, X_n \neq Y_{n-1}\}$,

$$\begin{aligned} \mathbb{E}[LC_n] \leq &\quad \mathbb{E}[LC_{n-1}\mathbf{1}_A] + \mathbb{E}[(LC_{n-1}+1)\mathbf{1}_{A^C}] \\ = &\quad \mathbb{E}[LC_{n-1}] + \mathbb{P}(A^C). \end{aligned}$$

Therefore by stationarity, for $n > 1$, $\mathbb{E}[LC_n] \leq n(1 - \mathbb{P}(X_1 \neq Y_1, X_0 \neq Y_1, X_1 \neq Y_0))$. Then

$$\begin{aligned} \gamma^* \leq &\quad 1 - \mathbb{P}(X_1 \neq Y_1, X_0 \neq Y_1, X_1 \neq Y_0) \\ = &\quad 1 - \sum_{\substack{a,b,c,d \in \mathcal{A} \\ b \neq d, b \neq c, a \neq d}} \mathbb{P}(X_0 = a, X_1 = b, Y_0 = c, Y_1 = d). \end{aligned}$$

Let $\delta \in [0, \gamma^*)$ and $q \in (-1, -1+\gamma^*-\delta)$. By definition, $\gamma_n(q) < 1+(-1+\gamma^*-\delta) = \gamma^*-\delta$ for all n. Then $\tilde{\gamma}_{n(1-q)}(p_1) < \gamma^* - \delta$ for $p_1 = (q+1)/(1-q)$ and all n, i.e., when $p_1 \in (0, (\gamma^*-\delta)/(2-\gamma^*+\delta)$. Note that $\tilde{\gamma}_n(p) = \tilde{\gamma}_n(1/p)$ by symmetry, so $\tilde{\gamma}_n(p_2) < \gamma^*-\delta$ for $p_2 > (2-\gamma^*+\delta)/(\gamma^*-\delta)$ and all n.

□

3.2 Main result

The main result provides a quantitative estimate of the fact that the positions in X and Y of the same letter in the longest common subsequence cannot be too far apart. Let $n = km$

and let the integers

$$r_0 = 0 \leq r_1 \leq r_2 \leq \cdots \leq r_{m-1} \leq r_m = n, \tag{3.4}$$

be such that

$$LC_n = \sum_{i=1}^{m} LCS(X_{k(i-1)+1} \dots X_{ki}; Y_{r_{i-1}+1} \dots Y_{r_i}). \tag{3.5}$$

When the above is satisfied, we identify $(r_0, \dots, r_m)$ as an *optimal alignment* for the sequences X and Y.

Let $\epsilon > 0$, $p_1 > 0$ and $p_2 > 0$ be constants. Let A^n_{ϵ,p_1,p_2} be the set of optimal alignments of $X_1, \dots, X_n$ and $Y_1, \dots, Y_n$, for which a proportion of at least $1 - \epsilon$ of the integer intervals $[r_{i-1} + 1, r_i]$ for $i = 1, \dots, m$ have their length between kp_1 and kp_2. More precisely, A^n_{ϵ,p_1,p_2} is the event that for all integer vectors $(r_0, r_1, \dots, r_m)$ satisfying (3.4) and (3.5),

$$\text{Card}(\{i \in 1, 2, \dots, m : kp_1 \leq r_i - r_{i-1} \leq kp_2\}) \geq (1 - \epsilon)m. \tag{3.6}$$

The main result states that A^n_{ϵ,p_1,p_2} holds with high probability. It uses the $\beta-$mixing coefficient, as given in Definition 3.3 of [5], to measure the dependence between X and Y.

Definition 3.2.1. Let $\mathcal{F}_1$ and $\mathcal{F}_2$ be two $\sigma-$fields $\subset \mathcal{F}$, then the $\beta-$mixing coefficient, associated with these sub-σ-fields of $\mathcal{F}$, is given by:

$$\beta(\mathcal{F}_1, \mathcal{F}_2) := \frac{1}{2} \sup \sum_{i=1}^{I} \sum_{j=1}^{J} |\mathbb{P}(A_i \cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j)|, \tag{3.7}$$

where the supremum is taken over all pairs of finite partitions $\{A_1, \dots, A_I\}$ and $\{B_1, \dots, B_J\}$ of Ω such that $A_i \in \mathcal{F}_1$, for all $i \in \{1, \dots, I\}$, $I \geq 1$ and $B_j \in \mathcal{F}_2$ for all $j \in \{1, \dots, J\}$, $J \geq 1$.

In our case, since $X^{(n)} = (X_1, \dots, X_n)$ and $Y^{(n)} = (Y_1, \dots, Y_n)$ are discrete random vectors for any fixed $n \geq 1$, by Proposition 3.21 in [5]:

$$\begin{aligned} \beta(n) \coloneqq & \quad \beta\left(\sigma\left(X^{(n)}\right), \sigma\left(Y^{(n)}\right)\right) \\ = & \quad \frac{1}{2} \sum_{u \in \mathcal{A}^n} \sum_{v \in \mathcal{A}^n} \left|\mathbb{P}\left(X^{(n)} = u, Y^{(n)} = v\right) - \mathbb{P}\left(X^{(n)} = u\right) \mathbb{P}\left(Y^{(n)} = v\right)\right|, \end{aligned} \tag{3.8}$$

where $\sigma\left(X^{(n)}\right)$ and $\sigma\left(Y^{(n)}\right)$ are the $\sigma-$fields generated by $X^{(n)}$ and $Y^{(n)}$. Clearly $X^{(n)}$ and $Y^{(n)}$ are independent if and only if $\beta(n) = 0$. Further, set $\beta^* \coloneqq \lim_{n\to\infty} \beta(n)$, where the limit exists since $\beta(n)$ is non-decreasing, in n, and $\beta(n) \in [0, 1]$ (see Section 5 in [5]). A key ingredient in proving Theorem 3.2.3 is a Hoeffding-type inequality for Markov chains, a particular case of a result due to Paulin [22], which is now recalled. It relies on the mixing time $\tau(\epsilon)$ of the Markov chain Z given by

$$\tau(\epsilon) \coloneqq \min\{t \in \mathbb{N} : \overline{d}_Z(t) \leq \epsilon\},$$

where

$$\overline{d}_Z(t) \coloneqq \max_{1 \leq i \leq N-t} \sup_{x,y \in \Lambda_i} d_{TV}(\mathcal{L}(Z_{i+t}|Z_i = x), \mathcal{L}(Z_{i+t}|Z_i = y)),$$

and where $d_{TV}(\mu, \nu) = \frac{1}{2} \sum_{x\in\Omega} |\mu(x) - \nu(x)|$ is the total variation distance between the two probability measures μ and ν on the finite set Ω.

Lemma 3.2.2. *Let $M \coloneqq (M_1, \ldots, M_N)$ be a (not necessarily time homogeneous) Markov chain, taking values in a Polish space $\Lambda = \Lambda_1 \times \cdots \times \Lambda_N$, with mixing time $\tau(\epsilon)$, $0 \leq \epsilon \leq 1$. Let*

$$\tau_{min} \coloneqq \inf_{0 \leq \epsilon < 1} \tau(\epsilon) \left(\frac{2 - \epsilon}{1 - \epsilon}\right)^2, \tag{3.9}$$

and let $f : \Lambda \to \mathbb{R}$ be such that there is $c \in \mathbb{R}^N_+$ with $|f(u) - f(v)| \leq \sum_{i=1}^N c_i \mathbf{1}_{u_i \neq v_i}$. Then for any $t \geq 0$,

$$\mathbb{P}(f(M) - \mathbb{E}f(M) \geq t) \leq \exp\left(\frac{-2t^2}{\tau_{min} \sum_{i=1}^N c_i^2}\right). \tag{3.10}$$

For our purposes, the Hoeffding-type inequality used below follows directly from the above result once one notes that $(Z_i, X_i, Y_i)_{i\geq 1}$ is jointly a Markov chain on a bigger state space with $\tau(\epsilon)$ - the mixing time of the chain

Theorem 3.2.3. *Let $\epsilon > 0$ and $\delta \in (4\beta^*, \gamma^*/2)$. Let $0 < p_1 < 1 < p_2$ be such that $\tilde{\gamma}_n(p_1) < \gamma^* - 2\delta$ and $\tilde{\gamma}_n(p_2) < \gamma^* - 2\delta$ for all n. Fix the integer k to be such that $(1+\ln(k+1))/k \leq (\delta - 4\beta^*)^2\epsilon^2/(8\tau_{\min})$, then*

$$\mathbb{P}(A^n_{\epsilon,p_1,p_2}) \geq 1 - \exp\left(-n\left(-\frac{1+\ln(k+1)}{k} + \frac{(\delta-4\beta^*)^2\epsilon^2}{8\tau_{\min}}\right)\right), \tag{3.11}$$

for all $n = n(\epsilon, \delta)$ large enough.

Remark 3.2.4. *Note that when X and Y are generated by independent hidden chains Z_X and Z_Y, (then $Z = Z_X \times Z_Y$), $\beta^* = 0$ and up to the mixing time τ_{min} in the expression above, we recover the independent case.*

3.3 Proof of Theorem 3.2.3

Let $0 = r_0 \leq r_1 \leq \ldots \leq r_m = n$ be a fixed set of integers. Again we align $X_{k(i-1)+1} \ldots X_{ki}$ with $Y_{r_{i-1}+1} \ldots Y_{ri}$ and get the alignment score:

$$\begin{aligned} L_n(r) :=& L_n(r_0, r_1, \ldots, r_m) \\ :=& \sum_{i=1}^{m} LCS(X_{k(i-1)+1} \ldots X_{ki}; Y_{r_{i-1}+1} \ldots Y_{r_i}). \end{aligned}$$

Note that $L_n(r) \leq LC_n$. Let $\mathcal{R}_{\epsilon,p_1,p_2}$ be the set of all integer vectors satisfying (3.4) and (3.6), while $\overline{\mathcal{R}}_{\epsilon,p_1,p_2}$ be the set of all integer vectors satisfying (3.4) but not (3.6). Note that if r^* is a random vector satisfying (3.4), $\mathbb{P}(r^* \in \mathcal{R}^c_{\epsilon,p_1,p_2}) = \mathbb{P}(r^* \in \overline{\mathcal{R}}_{\epsilon,p_1,p_2})$. Indeed, $\mathcal{R}^c_{\epsilon,p_1,p_2} \setminus \overline{\mathcal{R}}_{\epsilon,p_1,p_2}$ consists of the vectors not satisfying (3.4) and then $\mathbb{P}(r^* \in \mathcal{R}^c_{\epsilon,p_1,p_2} \setminus \overline{\mathcal{R}}_{\epsilon,p_1,p_2}) = 0$.

Lemma 3.3.1. *Let $\epsilon > 0$ and $\delta \in (4\beta^*, \gamma^*/2)$. Let $0 < p_1 < 1 < p_2$ be such that $\tilde{\gamma}_n(p_1) < \gamma^* - 2\delta$ and $\tilde{\gamma}(p_2) < \gamma^* - 2\delta$ for all n. Let $r \in \overline{\mathcal{R}}_{\epsilon,p_1,p_2}$, then*

$$\mathbb{E}[L_n(r) - LC_n] \leq -\frac{(\delta - 4\beta^*)\epsilon n}{2}, \tag{3.12}$$

for all $n = n(\epsilon, \delta)$ large enough.

Proof. For $p \notin [p_1, p_2]$,

$$\frac{\mathbb{E}[LCS(X_1 \ldots X_k; Y_1 \ldots Y_{kp})]}{k(1+p)/2} \leq \gamma^* - 2\delta.$$

Following a similar argument as in [17] and by stationarity, when $r_i - r_{i-1} = kp$,

$$\begin{aligned}
&\frac{2\mathbb{E}[LCS(X_{(i-1)k+1} \ldots X_{ik}; Y_{r_{i-1}+1} \ldots Y_{r_i})]}{k + r_i - r_{i-1}} \\
\leq \quad & \frac{2\mathbb{E}[LCS(X_{(i-1)k+1} \ldots X_{ik}; Y_{(i-1)k+1} \ldots Y_{(i-1)k+kp})]}{k(1+p)} + 4\beta(n) \\
\leq \quad & \frac{\mathbb{E}[LCS(X_1 \ldots X_k; Y_1 \ldots Y_{kp})]}{k(1+p)/2} + 4\beta^* \\
\leq \quad & \gamma^* - 2\delta + 4\beta^*.
\end{aligned}$$

Thus,

$$\gamma^* \left(\frac{k + r_i - r_{i-1}}{2}\right) - \mathbb{E}[LCS(X_{(i-1)k+1} \ldots X_{ik}; Y_{r_{i-1}+1} \ldots Y_{r_i})] \geq (\delta - 2\beta^*)k.$$

Let $\mathcal{M} = \{i : r_i - r_{i-1} \notin [kp_1, kp_2]\}$. Then

$$\begin{aligned}
&\sum_{i \in \mathcal{M}} \left(\gamma^* \left(\frac{k + r_i - r_{i-1}}{2}\right) - \mathbb{E}[LCS(X_{(i-1)k+1} \ldots X_{ik}; Y_{r_{i-1}+1} \ldots Y_{r_i})]\right) \\
\geq \quad & \sum_{i \in \mathcal{M}} (\delta - 2\beta^*)k \geq n(\delta - 2\beta^*)\epsilon.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
&\sum_{i\in\mathcal{M}}\left(\gamma^*\left(\frac{k+r_i-r_{i-1}}{2}\right)-\mathbb{E}[LCS(X_{(i-1)k+1}\dots X_{ik};Y_{r_{i-1}+1}\dots Y_{r_i})]\right)\\
\leq&\sum_{i=1}^{m}\left(\gamma^*\left(\frac{k+r_i-r_{i-1}}{2}\right)-\mathbb{E}[LCS(X_{(i-1)k+1}\dots X_{ik};Y_{r_{i-1}+1}\dots Y_{r_i})]\right)\\
=&\gamma^* n-L_n(r).
\end{aligned}$$

Therefore,

$$\gamma^* n-L_n(r)\geq n(\delta-2\beta^*)\epsilon, \tag{3.13}$$

when $r\in\overline{\mathcal{R}}_{\epsilon,p_1,p_2}$. Finally, since $\lim_{n\to\infty}\mathbb{E}[LC_n]/n=\gamma^*$,

$$0\leq\gamma^*-\frac{\mathbb{E}[LC_n]}{n}\leq\frac{\delta\epsilon}{2}, \tag{3.14}$$

for $n=n(\epsilon,\delta)$ large enough. Combining (3.13) and (3.14) recovers (3.12).

$\square$

We now proceed with the proof of Theorem 3.2.3. Clearly,

$$Card(\overline{\mathcal{R}}_{\epsilon,p_1,p_2})\leq\binom{n+m}{m}\leq\frac{(n+m)^m}{m!}\leq\left(\frac{e(n+m)}{m}\right)^m=(e(k+1))^m. \tag{3.15}$$

For the event A^n_{ϵ,p_1,p_2} to hold, there needs to exist at least one $r\in\overline{\mathcal{R}}_{\epsilon,p_1,p_2}$. Thus,

$$(A^n_{\epsilon,p_1,p_2})^c=\bigcup_{r\in\overline{\mathcal{R}}_{\epsilon,p_1,p_2}}\{L_n(r)-LC_n\geq 0\},$$

and

$$\mathbb{P}((A_{\epsilon,p_1,p_2})^c)\leq\sum_{r\in\overline{\mathcal{R}}_{\epsilon,p_1,p_2}}\mathbb{P}(L_n(r)-LC_n\geq 0). \tag{3.16}$$

When $r \in \overline{\mathcal{R}}_{\epsilon,p_1,p_2}$ it follows from Lemma 3.3.1 that:

$$\mathbb{E}[L_n(r) - LC_n] \leq -\frac{(\delta - 4\beta^*)\epsilon n}{2},$$

and so

$$\mathbb{P}(L_n(r) - LC_n \geq 0) \leq \mathbb{P}\left(L_n(r) - LC_n - \mathbb{E}[L_n(r) - LC_n] \geq \frac{(\delta - 4\beta^*)\epsilon n}{2}\right), \quad (3.17)$$

for all n large enough. Now, the difference $L_n(r) - LC_n$ changes by at most plus or minus two, when any of the entries $X_1, \dots, X_n, Y_1, \dots, Y_n$ are changed. Therefore, Lemma 3.2.2 applied to the right-hand side of (3.17), gives

$$\begin{aligned}\mathbb{P}(L_n(r) - LC_n \geq 0) \leq \quad & \mathbb{P}\left(L_n(r) - LC_n - \mathbb{E}[L_n(r) - LC_n] \geq \frac{(\delta - 4\beta^*)\epsilon n}{2}\right) \\ \leq \quad & \exp\left(-\frac{(\delta - 4\beta^*)^2\epsilon^2}{8\tau_{\min}} n\right).\end{aligned}$$

Combining the last inequality with (3.16), one obtains:

$$\mathbb{P}((A^n_{\epsilon,p_1,p_2})^C) \leq Card(\overline{\mathcal{R}}_{\epsilon,p_1,p_2}) \exp\left(-\frac{(\delta - 4\beta^*)^2\epsilon^2}{8\tau_{\min}} n\right).$$

But from (3.15),

$$\begin{aligned}\mathbb{P}((A^n_{\epsilon,p_1,p_2})^C) \leq & (e(k+1))^m \exp\left(-\frac{(\delta - 4\beta^*)^2\epsilon^2}{8\tau_{\min}} n\right) \\ = & \exp\left(-n\left(-\frac{1 + \ln(1+k)}{k} + \frac{(\delta - 4\beta^*)^2\epsilon^2}{8\tau_{\min}}\right)\right).\end{aligned}$$

Therefore, the proof of Theorem 3.2.3 is complete.

Before we move on we show the following corollary that makes use of the rate of convergence of $\mathbb{E}[LC_n]$ which we recall next, and holds for all $n \geq 1$. Our rate of convergence result, that is Theorem 2.2.3, states that there is a constant C_A such that

$$C_A\sqrt{\frac{\ln n}{n}} + 2\beta^* \geq \gamma^* - \frac{\mathbb{E}[LC_n]}{n},$$

for every n. The following variant of Theorem 3.2.3 holds

Corollary 3.3.2. *Let $\delta \in (6\beta^*, \gamma^*/2)$. Let $0 < p_1 < 1 < p_2$ be such that $\tilde{\gamma}_n(p_1) < \gamma^* - 2\delta$ and $\tilde{\gamma}_n(p_2) < \gamma^* - 2\delta$ for all n. Let $\alpha \in (0,1)$ and*

$$\epsilon := c_1\sqrt{\frac{1+\ln(1+n^\alpha)}{n^\alpha}},$$

where $c_1 > 0$ such that

$$c_1^2 \geq \frac{16\tau_{min}}{(\delta - 6\beta^*)^2},$$

and

$$c_1^2\left(\frac{1+\ln(1+n^\alpha)}{n^\alpha}\right) \geq \frac{4C_A^2}{\delta^2}\frac{\ln n}{n},$$

for all $n \geq 1$. Then,

$$\mathbb{P}(A^n_{\epsilon,p_1,p_2}) \geq 1 - \exp\left(-n^{1-\alpha}(1+\ln(1+n^\alpha))\right), \tag{3.18}$$

for all $n \geq 1$.

Proof. First show the following variant of Lemma 3.3.1.

Lemma 3.3.3. *$\delta \in (6\beta^*, \gamma^*/2)$. Let $0 < p_1 < 1 < p_2$ be such that $\tilde{\gamma}_n(p_1) < \gamma^* - 2\delta$ and $\tilde{\gamma}(p_2) < \gamma^* - 2\delta$ for all n. Let $\alpha \in (0,1)$, $\epsilon > 0$ and $c_1 > 0$ be such that*

$$\epsilon := c_1\sqrt{\frac{1+\ln(1+n^\alpha)}{n^\alpha}} \geq \frac{2C_A}{\delta}\sqrt{\frac{\ln n}{n}},$$

for all $n \geq 1$*. Let* $r \in \overline{\mathcal{R}}_{\epsilon,p_1,p_2}$*, then*

$$\mathbb{E}[L_n(r) - LC_n] \leq -\frac{(\delta - 6\beta^*)\epsilon n}{2}, \tag{3.19}$$

for all $n \geq 1$*.*

Proof of Lemma 3.3.3. The proof is very similar to Lemma 3.3.1. For $p \notin [p_1, p_2]$,

$$\frac{\mathbb{E}[LCS(X_1 \ldots X_k; Y_1 \ldots Y_{kp})]}{k(1+p)/2} \leq \gamma^* - 2\delta.$$

Following a similar argument as in [17] and by stationarity, when $r_i - r_{i-1} = kp$,

$$\begin{aligned} & \frac{2\mathbb{E}[LCS(X_{(i-1)k+1} \ldots X_{ik}; Y_{r_{i-1}+1} \ldots Y_{r_i})]}{k + r_i - r_{i-1}} \\ \leq \quad & \gamma^* - 2\delta + 4\beta^*. \end{aligned}$$

Thus,

$$\begin{aligned} & (\gamma^* - 2\beta^*)\left(\frac{k + r_i - r_{i-1}}{2}\right) - \mathbb{E}[LCS(X_{(i-1)k+1} \ldots X_{ik}; Y_{r_{i-1}+1} \ldots Y_{r_i})] \\ \geq \quad & (2\delta - 6\beta^*)\left(\frac{k + r_i - r_{i-1}}{2}\right) \\ \geq \quad & (\delta - 3\beta^*)k. \end{aligned}$$

Let $\mathcal{M} = \{i : r_i - r_{i-1} \notin [kp_1, kp_2]\}$. Then

$$\begin{aligned} & \sum_{i \in \mathcal{M}} \left((\gamma^* - 2\beta^*)\left(\frac{k + r_i - r_{i-1}}{2}\right) - \mathbb{E}[LCS(X_{(i-1)k+1} \ldots X_{ik}; Y_{r_{i-1}+1} \ldots Y_{r_i})] \right) \\ \geq \quad & \sum_{i \in \mathcal{M}} (\delta - 3\beta^*)k \\ \geq \quad & n(\delta - 3\beta^*)\epsilon. \end{aligned}$$

On the other hand,

$$\begin{aligned}
&\sum_{i\in\mathcal{M}}\left((\gamma^*-2\beta^*)\left(\frac{k+r_i-r_{i-1}}{2}\right)-\mathbb{E}[LCS(X_{(i-1)k+1}\dots X_{ik};Y_{r_{i-1}+1}\dots Y_{r_i})]\right)\\
\leq\ &\sum_{i=1}^{m}\left((\gamma^*-2\beta^*)\left(\frac{k+r_i-r_{i-1}}{2}\right)-\mathbb{E}[LCS(X_{(i-1)k+1}\dots X_{ik};Y_{r_{i-1}+1}\dots Y_{r_i})]\right)\\
=\ &(\gamma^*-2\beta^*)n-L_n(r).
\end{aligned}$$

Therefore,

$$(\gamma^*-2\beta^*)n-L_n(r)\geq n(\delta-3\beta^*)\epsilon, \tag{3.20}$$

when $r\in\overline{\mathcal{R}}_{\epsilon,p_1,p_2}$. Finally, Theorem 2.2.3 implies that for all $n\geq 1$,

$$\gamma^*-2\beta^*-\frac{\mathbb{E}[LC_n]}{n}\leq C_A\sqrt{\frac{\ln n}{n}}.$$

So in particular, let $\alpha\in(0,1)$, $\epsilon>0$ and $c_1>0$ be such that

$$\epsilon := c_1\sqrt{\frac{1+\ln(1+n^\alpha)}{n^\alpha}}\geq\frac{2C_A}{\delta}\sqrt{\frac{\ln n}{n}},$$

for all $n\geq 1$. Then

$$\gamma^*-2\beta^*-\frac{\mathbb{E}[LC_n]}{n}\leq\frac{\delta\epsilon}{2}, \tag{3.21}$$

for all $n\geq 1$. Combining (3.20) and (3.21) recovers (3.19).

□

The remainder of the proof of Corollary 3.3.2 follows from the same arguments that finish the proof of Theorem 3.2.3. □

3.4 Closeness to the diagonal

Let D^n_{ϵ,p_1,p_2} be the event that the points representing any optimal alignment of $X_1,\ldots,X_n$ with $Y_1,\ldots,Y_n$ are above the line $y = p_1x - p_1n\epsilon - p_1k$, and below the line $y = p_2x + p_2n\epsilon + p_2k$.

Theorem 3.4.1. *Let $\epsilon > 0$ and $\delta \in (4\beta^*, \gamma^*/2)$. Let $0 < p_1 < 1 < p_2$ be such that $\tilde{\gamma}_n(p_1) < \gamma^* - 2\delta$ and $\tilde{\gamma}_n(p_2) < \gamma^* - 2\delta$ for all n. Fix the integer k to be such that $(1+\ln(1+k))/k \leq (\delta - 4\beta^*)^2\epsilon^2/(8\tau_{\min})$, then*

$$\mathbb{P}(D^n_{\epsilon,p_1,p_2}) \geq 1 - 2\exp\left(-n\left(-\frac{1+\ln(1+k)}{k} + \frac{(\delta-4\beta^*)^2\epsilon^2}{8\tau_{\min}}\right)\right), \tag{3.22}$$

for all $n = n(\epsilon, \delta)$ large enough.

Proof. See the proof of Theorem 4.1 in [18].

□

3.5 Short string-lengths properties are generic

Let $\mathcal{P}$ be a relation assigning to every pair of strings (x, y) the value 1 if the pair (x, y) has a certain property, and 0 otherwise. Hence, if $\mathcal{A}$ is the alphabet we consider,

$$\mathcal{P} : \left(\cup_k \mathcal{A}^k\right) \times \left(\cup_k \mathcal{A}^k\right) \to \{0, 1\},$$

and if $\mathcal{P}(x, y) - 1$, the string pair (x, y) is said to have the property $\mathcal{P}$.

Let now $\epsilon > 0$, be fixed and let $r = (r_0, \ldots, r_m)$ satisfy condition (3.4). Let also $B^n_{\mathcal{P}}(r, \epsilon)$ be the event that there is a proportion of at least $1 - \epsilon$ of the string pairs

$$(X_{(i-1)k+1} \ldots X_{ik}; Y_{r_{i-1}+1} \ldots Y_i) \tag{3.23}$$

satisfying the property $\mathcal{P}$, i.e.,

$$B^n_{\mathcal{P}}(r,\epsilon) = \left\{ \sum_{i=1}^{m} \mathcal{P}(X_{(i-1)k+1}\dots X_{ik}; Y_{r_{i-1}+1}\dots Y_{ri}) \geq (1-\epsilon)m \right\}.$$

Next, let $B^n_{\mathcal{P}}(\epsilon)$ be the event that $B^n_{\mathcal{P}}(r,\epsilon)$ holds for all r satisfying (3.4) and $LC_n = L_n(r)$. There is a $q \in [0,1]$ such that

$$\mathbb{P}(\mathcal{P}(X_1\dots X_k; Y_s \dots Y_{s+\ell}) \geq 1-q,$$

and

$$\mathbb{P}(\mathcal{P}(X_s\dots X_{k+s}; Y_1 \dots Y_{\ell})) \geq 1-q,$$

for all $\ell \in [kp_1, kp_2]$ and all integers $s \geq 1$. If X and Y are independent sequences, by stationarity it is enough to consider only the case when $s = 1$.

We want to find minimum value of $q = q(k)$ such that a large proportion of the aligned pairs (3.23) has the property $\mathcal{P}$. Recall that $\mathcal{R}^n_{\epsilon,p_1,p_2}$ is the event that every optimal alignment aligns a proportion of at least $1-\epsilon$ of the sub-strings $X_{(i-1)k+1}\dots X_{ik}$ with sub-strings of Y with length in $[kp_1, kp_2]$.

Let $\tilde{B}^n_{\mathcal{P}}(r,\epsilon)$ be the event that among the aligned string pieces (3.23) there are no more than ϵm which do not satisfy the property $\mathcal{P}$ and have their length $r_i - r_{i-1} \in [kp_1, kp_2]$. For $\epsilon_1 > 0, \epsilon_2 > 0$,

$$A^n_{\epsilon,p_1,p_2} \cap \left(\bigcap_{r \in \mathcal{R}_{\epsilon,p_1,p_2}} \tilde{B}^n_{\mathcal{P}}(r,\epsilon_2) \right) \subset B^n_{\mathcal{P}}(\epsilon_1+\epsilon_2),$$

and so

$$\mathbb{P}((B^n_{\mathcal{P}}(\epsilon_1+\epsilon_2))^c) \leq \mathbb{P}((A^n_{\epsilon_1,p_1,p_2})^c) + \sum_{r \in \mathcal{R}_{\epsilon_1,p_1,p_2}} \mathbb{P}((\tilde{B}^n_{\mathcal{P}}(r,\epsilon_2))^c).$$

Next,

$$\mathbb{P}((\tilde{B}^n_{\mathcal{P}}(r,\epsilon_2))^c) \le \binom{m}{\epsilon_2 m} q^{\epsilon_2 m} \le \exp(H_e(\epsilon_2)m) q^{\epsilon_2 m},$$

where H_e is the base e entropy function. Hence,

$$\mathbb{P}((\tilde{B}^n_{\mathcal{P}}(r,\epsilon_2))^c) \le q^{\epsilon_2 m} \exp(H_e(\epsilon_2)m).$$

Therefore,

$$\mathbb{P}((B^n_{\mathcal{P}}(\epsilon_1+\epsilon_2))^c) \le \mathbb{P}((A^n_{\epsilon_1,p_1,p_2})^c) + (e(k+1))^m q^{\epsilon_2 m} \exp(H_e(\epsilon_2)m).$$

Taking $q(k) = 1/(2e(k+1))^{1/\epsilon_2}$, finally yields

$$\mathbb{P}((B^n_{\mathcal{P}}(\epsilon_1+\epsilon_2))^c) \le \mathbb{P}((A^n_{\epsilon_1,p_1,p_2})^c) + \exp((H_e(\epsilon_2) - \ln 2)m).$$

When $\epsilon_2 < 1/2$, $H_e(\epsilon_2) < \ln 2$, and then $\exp((H_e(\epsilon_2) - \ln 2)m)$ is exponentially small in n. Furthermore, Theorem 3.2.3 provides an exponentially small lower bound on $\mathbb{P}((A^n_{\epsilon_1,p_1,p_2})^c)$. Letting $\epsilon := \epsilon_1 = \epsilon_2$, we have obtained the following result.

Theorem 3.5.1. *Let $0 < \epsilon < 1$ and $\delta \in (4\beta^*, \gamma^*/2)$. Let $0 < p_1 < 1 < p_2$ be such that $\tilde{\gamma}_n(p_1) < \gamma^* - 2\delta$ and $\tilde{\gamma}_n(p_2) < \gamma^* - 2\delta$ for all n. Let the integer k be such that $(1+\ln(k+1))/k \le (\delta - 4\beta^*)^2\epsilon^2/(32\tau_{\min})$, and*

$$\min_{\ell\in[kp_1,kp_2],s\ge 1} \mathbb{P}(\mathcal{P}(X_1\dots X_k; Y_s \dots Y_{s+\ell}) \ge 1 - \frac{1}{(2e(k+1))^{2/\epsilon}},$$
$$\min_{\ell\in[kp_1,kp_2],s\ge 1} \mathbb{P}(\mathcal{P}(X_s\dots X_{k+s}; Y_1 \dots Y_{\ell})) \ge 1 - \frac{1}{(2e(k+1))^{2/\epsilon}}.$$

Then for any r satisfying (3.4) *and $L_n(r) = LC_n$, the proportion of string pairs $(X_{(i-1)k+1}\dots X_{ik};$*

$Y_{r_{i-1}+1} \dots Y_{r_i})$ *satisfying property* $\mathcal{P}$ *is at least* $1-\epsilon$ *with probability at least equal to:*

$$1 - \exp\left(-n\left(-\frac{1+\ln(k+1)}{k} + \frac{(\delta - 4\beta^*)^2\epsilon^2}{32\tau_{\min}}\right)\right) - \exp\left(\frac{n}{k}\left(H_e\left(\frac{\epsilon}{2}\right) - \ln 2\right)\right),$$

for all $n = n(\epsilon, \delta)$ *large enough.*

CHAPTER 4

STEIN'S METHOD

This chapter will focus on the development of tools that are used in the proof for the central limit theorem. In particular, we show a version of Stein's method for normal approximation that holds for hidden Markov models. In the process, the concept of a discrete derivative is realized and we give useful bounds for some expressions. Finally, we show an upper bound on the variance of LC_n that supports the main assumption in Theorem 5.1.1.

4.1 Preliminaries

Stein's method is a way to obtain normal approximation based on the observation that the standard normal distribution $\mathcal{N}$ is the only distribution that satisfies

$$\mathbb{E}[\mathcal{N} f(\mathcal{N})] = \mathbb{E}[f'(\mathcal{N})],$$

for all absolutely continuous f with a.e. derivative f' such that $\mathbb{E}[f'(\mathcal{N})] < \infty$. Then for another random variable W, the value of $|\mathbb{E}[Wf(W) - f'(W)]|$ can be thought of as a distance measuring the proximity of W and $\mathcal{N}$.

The distance between two probability measures μ and ν is often of the type:

$$d(\mu, \nu) = \sup_h \left\{ \left| \int h d\mu - \int h d\nu \right| : h \in \mathcal{H} \right\},$$

where $\mathcal{H}$ is a class of functions. (We write $d(X, Y)$ instead of $d(\mu, \nu)$ when $X \sim \mu$ and $Y \sim \nu$). Various $\mathcal{H}$ give rise to various distances. For instance, when $\mathcal{H}$ is the set of $1-$Lipschitz functions we get $d_W(\mu, \nu)$ - the Wasserstein distance, while for $\mathcal{H}$ is the set of $1_{(-\infty,x]}$ for $x \in \mathbb{R}$ we get $d_K(\mu, \nu)$ - the Kolmogorov distance. Now let $\mathcal{H}^*$ be the family of functions f that are solutions to the differential equation

$$h(W) - \mathbb{E}[h(\mathcal{N})] = f'(W) - Wf(W),$$

where $h \in \mathcal{H}$, $W \sim \mu$ and $\mathcal{N} \sim \nu = \mathcal{N}(0,1)$.

Bounding $|\mathbb{E}[h(W)] - \mathbb{E}[h(\mathcal{N})]|$ only depends on the properties of the solutions f and the law of W. More precisely the following result holds [23, Corollary 3.38]:

Proposition 4.1.1. *Let $\mathcal{N} \sim \mathcal{N}(0,1)$ and let W be a random variable with finite expectation. Then,*

$$d_K(W, \mathcal{N}) \leq \sup_{f \in \mathcal{H}^*} |\mathbb{E}(f'(W) - Wf(W))|,$$

where $\mathcal{H}^*$ is the set of all absolutely continuous $f : \mathbb{R} \to \mathbb{R}$ such that $||f||_\infty := \sup_{x \in \mathbb{R}} |f(x)| \leq \sqrt{\pi/2}$, $||f'||_\infty \leq 2$ and such that for all $u, v, w \in \mathbb{R}$,

$$|(w+u)f(w+u) - (w+v)f(w+v)| \leq \left(|w| + \sqrt{2\pi}/4\right)(|u| + |v|).$$

A simple optimization argument can yield the following result by Chatterjee [6, Proposition 1.1]:

Corollary 4.1.2. *Let Z be a standard normal random variable and W be any random variable with $\mathbb{E}[W] = 0$ and $\mathbb{E}[W^2] = 1$. Then*

$$d_K(W, \mathcal{N}) \leq 2 \left(\sup_{f \in \mathcal{H}^*} |\mathbb{E}(f'(W) - Wf(W))| \right)^{1/2}, \tag{4.1}$$

where $\mathcal{H}^*$ is the set of all $f : \mathbb{R} \to \mathbb{R}$ that are twice continuously differentiable, and $|f(x)| \leq 1$, $|f'(x)| \leq 1$, and $|f''(x)| \leq 1$ for all $x \in \mathbb{R}$.

In applications, $W = W(n)$, i.e., it is often a random quantity connected to some object of varying size n. A suitable bound on the right-hand side of (4.1) in terms of n can yield weak convergence of $W(n)$ to $\mathcal{N}$ and therefore a central limit theorem can be established for $W(n)$.

For instance if $W = g(X)$ - a function of n independent random variables $X = (X_1, \ldots, X_n)$ such bounds have been obtained by Chatterjee in [6]. His approach relies on introducing discrete derivatives on exchangeable pairs. In particular, let $X' = (X'_1, \ldots, X'_n)$ with X'_i - an independent copy of X_i and $W' = g(X')$. Then (W, W') is an exchangeable pair because it has the same joint distribution as (W', W). A perturbation $W^A = g^A(X) := g(X^A)$ of W is defined through the change X^A of X as follows:

$$X_i^A = \begin{cases} X'_i & \text{if } i \in A \\ X_i & \text{if } i \notin A. \end{cases}$$

for any $A \subseteq [n]$. Then a discrete derivative can be defined for any $A \subseteq [n]$ and $i \notin A$, as:

$$\Delta_i g^A = g(X^A) - g(X^{A \cup \{i\}}).$$

Chatterjee's result then reads as follows:

Theorem 4.1.3. *Let W be a function of n independent random variables as above with $\mathbb{E}[W] = 0$ and $\mathbb{E}[W^2] = 1$. Then:*

$$|\mathbb{E}(f'(W) - Wf(W))| \leq \sqrt{Var\mathbb{E}(T|W)} + \frac{1}{4}\sum_{i=1}^{n} \mathbb{E}|\Delta_i g|^3, \tag{4.2}$$

for every $f \in \mathcal{H}^*$ (as in Corollary 4.1.2) and where

$$T := \frac{1}{2}\sum_{i=1}^{n} \sum_{A \subseteq [n] \setminus \{i\}} \frac{1}{n\binom{n-1}{|A|}} \Delta_i g \Delta_i g^A. \tag{4.3}$$

Very little work has been done outside the independent case. Some notable exceptions include [8] where a rate of convergence of the $n-$step probability distribution to the stationary for the Ehrenfest urn model is obtained, and [13] and [15] where Markov chains appear in the generation of exchangeable pairs. I was able to show the following:

Theorem 4.1.4. *Let (Z, X) be a hidden Markov model with Z an aperiodic time homogeneous and irreducible Markov chain on a finite state space $\mathcal{S}$, and X taking values in a finite alphabet $\mathcal{A}$. Let $W := g(X_1, \dots, X_n)$ with $\mathbb{E}[W] = 0$ and $\mathbb{E}[W^2] = 1$. There exist a sequence of independent random variables $R_0, R_1, \dots, R_{|\mathcal{S}|(n-1)}$ and a function h such that $h(R_0, \dots, R_{|\mathcal{S}|(n-1)})$ has the same law as $g(X_1, \dots, X_n)$. Then:*

$$|\mathbb{E}(f'(W) - Wf(W))| \leq \sqrt{Var\mathbb{E}(T|h(R_0, \dots, R_{|\mathcal{S}|(n-1)}))} + \frac{1}{4}\sum_{i=1}^{|\mathcal{S}|(n-1)} \mathbb{E}|\Delta_i h|^3,$$

for every $f \in \mathcal{H}^*$ (as in Corollary 4.1.2) and where

$$T := \frac{1}{2}\sum_{i=1}^{|\mathcal{S}|n} \sum_{A \subseteq [|\mathcal{S}|n] \setminus \{i\}} \frac{1}{n\binom{n-1}{|A|}} \Delta_i h \Delta_i h^A.$$

The main idea of the proof of the above result is to think of $R = (R_0, \dots, R_{|\mathcal{S}|(n-1)})$ as stacks of independent random variables on the $|\mathcal{S}|$ possible states of the hidden chain that determine the next step in the process, with R_0 specifying the initial state. Each R_i takes values in $\mathcal{S} \times \mathcal{A}$ and is distributed according to the transition probability from the present hidden state. Then one can write $g(X_1, \dots, X_n) = h(R_0, \dots, R_{|\mathcal{S}|(n-1)})$ for $h = g \circ \gamma$ where the function γ translates between R and X. This construction is carried out in more detail in the following section.

4.2 HMM as stacks of independent random variables

Let (Z, X) be a hidden Markov model with Z an aperiodic time homogeneous and irreducible Markov chain on a finite state space $\mathcal{S}$, and X taking values in a finite alphabet $\mathcal{A}$. Let P be transition matrix of the hidden chain and Q be the $|\mathcal{S}| \times |\mathcal{A}|$ probability matrix for the observations, i.e., Q_{ij} is the probability of seeing output j if the latent state is in chain

i. Let the initial distribution of the hidden chain be μ. Then

$$\begin{aligned}&\mathbb{P}\Big((Z_1,\ldots,Z_n;X_1,\ldots,X_n)=(z_1,\ldots,z_n;x_1,\ldots,x_n)\Big)\\ =\quad &\mu(z_1)Q_{z_1,x_1}P_{z_1,z_2}\ldots P_{z_{n-1},z_n}Q_{z_n,x_n}.\end{aligned}$$

Our goal is to introduce a sequence of independent random variables $R_0, R_1, \ldots, R_{|\mathcal{S}|(n-1)}$ taking values on $\mathcal{S}\times\mathcal{A}$ and a function γ such that $\gamma(R_0,\ldots,R_{|\mathcal{S}|(n-1)}) = (Z_1,\ldots,Z_n;X_1,\ldots,X_n)$. For any $s, s' \in \mathcal{S}$, $x \in \mathcal{A}$ and $i \in [0, n-1]$, let

$$\begin{aligned}\mathbb{P}\big(R_0=(s,x)\big) &= \mu(s)Q_{s,x}\\ \mathbb{P}\big(R_{i|S|+s'}=(s,x)\big) &= P_{s',s}Q_{s,x}.\end{aligned}$$

The random variables R_i are well defined since $\sum_x Q_{s,x} = 1$ for any $s \in \mathcal{S}$, and $\sum_s P_{s',s} = \sum_s \mu(s) = 1$ for any $s' \in \mathcal{S}$. One can think of the variables R_i as a set of instructions on where the hidden Markov model goes next. The function γ reconstructs the realization (Z_i, X_i) sequentially from the sequence (R_i). In particular, γ captures the following relations

$$\begin{aligned}&(Z_1,X_1)=R_0\\ &(Z_{i+1},X_{i+1})=R_{i|S|+s}\text{ if } Z_i=s \text{ for } i\geq 1.\end{aligned}$$

One can think of the sequence (R_i) as $|\mathcal{S}|$ stacks of random variables on the $\mathcal{S}$ possible states of the latent Markov chain, and the values being rules for the next step in the model. Note that only one variable on the ith level of the stack will be used to determine the $(i+1)$-st hidden and observed pair. Furthermore, the distribution of the random variables R_i for $i \geq 1$ encodes the transition and output probabilities in the P and Q matrices of the original model.

Thus one can write $f(X_1,\ldots,X_n) = h(R_0,\ldots,R_{|\mathcal{S}|(n-1)})$ for $g := h\circ\gamma$ where the function

γ does the translation from $(R_i)_{i\geq 0}$ to $(Z_i, X_i)_{i\geq 1}$ as described above.

Let $R' = (R'_0, \ldots, R'_{|\mathcal{S}|(n-1)})$ be an independent copy of R. Let $A \subseteq \{0, 1, \ldots, |S|(n-1)\}$ and let the change R^A of R be defined as follows

$$R_i^A = \begin{cases} R'_i & \text{if } i \in A \\ R_i & \text{if } i \notin A. \end{cases}$$

Recall that the discrete derivative of h with a perturbation A is

$$\Delta_i h^A = h(R^A) - h(R^{A\cup\{i\}}).$$

Theorem 4.1.4 follows from Theorem 4.1.3 since when (Z, X) is a hidden Markov model one writes

$$W = g(X_1, \ldots, X_n) = h(R_0, \ldots, R_{|S|(n-1)}),$$

and the sequence $(R_i))_{i\geq 0}$ is a sequence of independent random variables and the conclusion of Theorem 4.1.3 holds for $(R_i)_{i\geq 0}$.

The remainder of the chapter establishes bounds on the discrete derivative $\Delta_i h(R^A)$ where $h = g \circ \gamma$ and g is a Lipschitz function. The following results will be used in the next chapter where we outline a proof to a central limit theorem for LC_n - which is a Lipschitz function of the observed variables of a hidden Markov model.

4.3 Bounds on $\Delta_i h^A$

The latent chain in the hidden Markov model is irreducible and aperiodic. Assume that each value in $\mathcal{A}$, the space of possible values for the observed variables X, can be reached through some state in $\mathcal{S}$. Assume further that the latent chain is started at the stationary distribution. Then there is K and $\epsilon > 0$ such that

$$\mathbb{P}(X_n = x, X_{n+K} = x') \geq \epsilon, \tag{4.4}$$

for all $x, x' \in \mathcal{A}$ and all $n \geq 1$. Such K and ϵ exist since the chain is irreducible and then

$$\mathbb{P}(Z_n = s, Z_{n+k} = s') > 0,$$

for all $n \geq 1$ and $s, s' \in \mathcal{S}$.

Proposition 4.3.1. *Let (Z, X) be a hidden Markov model as above and let $K > 0$ and $\epsilon > 0$ as in* (4.4)*. Let $g : \mathcal{A}^n \to \mathbb{R}$ be Lipschitz with constant $c > 0$. Let $R = (R_0, \dots, R_{|S|(n-1)})$ be a vector of independent random variables and h be the function such that*

$$g(X_1, \dots, X_n) = h(R_0, R_{|S|(n-1)}).$$

Then

$$\mathbb{P}(h(R) - h(R^i) \geq cKx(n)) \leq C(1-\epsilon)^{x(n)}, \tag{4.5}$$

where $x(n) > 0$ is some function of n. Furthermore, let $\alpha > 0$. Then for any $r > 0$,

$$\mathbb{E}|h(R) - h(R^i)|^r \leq C(\ln n)^r, \tag{4.6}$$

where $C = C(r)$ and n large enough.

Proof. The sequence of instructions R^i may give rise to a different realization (Z', X') of the hidden Markov model, as compared to (Z, X) - the one generated by R. However, the probability of seeing a mismatch between X and X' decays exponentially with the size of the mismatch. Assume R_i determines (Z_j, X_j) and R_i' determines (Z_j', X_j'). Then

$$\begin{aligned} & \mathbb{P}(X_{j+K} \neq X'_{j+K}) \\ = & \sum_{x \in \mathcal{A}} \mathbb{P}(X_{j+K} = x, X'_{j+K} \neq x) \\ = & \sum_{x \in \mathcal{A}} \mathbb{P}(X_{j+K} = x)\mathbb{P}(X'_{j+K} \neq x) \\ \leq & \ (1 - \epsilon), \end{aligned}$$

where we have used that if $X_{j+K} \neq X'_{j+K}$, then two variables have been generated by different instructions R_{j_1} and R_{j_2} and so X_{j+K} and X'_{j+K} are independent.
Let $r > 0$. Then

$$\mathbb{P}(X_{j+K} \neq X'_{j+K}, X_{j+2K} \neq X'_{j+2K}, \ldots, X_{j+x(n)K} \neq X'_{j+x(n)K}) \leq (1-\epsilon)^{x(n)}.$$

Note that once $X_\ell = X'_\ell$, $X_m = X'_m$ for all $m \geq l$. Let E be the event

$$E \coloneqq \{X_{j+K} \neq X'_{j+K}, X_{j+2K} \neq X'_{j+2K}, \ldots, X_{j+x(n)K} \neq X'_{j+x(n)K}\}.$$

Then

$$\begin{aligned} & \mathbb{P}(|h(R) - h(R^i)| \geq x(n)) \\ \leq & \ \mathbb{P}(E) \\ \leq & \ (1-\epsilon)^{x(n)}. \end{aligned}$$

This suffices for the proof of (4.5). Now,

$$\mathbb{E}|h(R) - h(R^i)|^r = \mathbb{E}|h(R) - h(R^i)|^r \mathbf{1}_E + \mathbb{E}|h(R) - h(R^i)|^r \mathbf{1}_{E^C},$$

Then

$$\mathbb{E}|h(R)-h(R^i)|^r \leq (Cn)^r(1-\epsilon)^{x(n)} + (CKx(n))^r. \tag{4.7}$$

Let $x(n) = -\frac{r\ln n}{\ln(1-\epsilon)} > 0$. Then,

$$\mathbb{E}|h(R)-h(R^i)| \leq \left(-\frac{Cr}{\ln(1-\epsilon)}\right)^r (\ln n)^r + (CK)^r. \tag{4.8}$$

The order of the bound is optimal. Indeed, assume that $x(n)$ is such that

$$(1-\epsilon)^{x(n)} \leq \left(\frac{\ln n}{n}\right)^r, \tag{4.9}$$

or

$$x(n) \geq -\frac{r(\ln n - \ln(\ln n))}{\ln(1-\epsilon)}.$$

Then

$$\mathbb{E}|h(R)-h(R^i)| \leq \left(-\frac{Cr}{\ln(1-\epsilon)}\right)^r (\ln n - \ln(\ln n))^r + C^r,$$

and the right-hand side has the same order of growth as (4.8).

If the order of growth of $(1-\epsilon)^{x(n)}$ is higher than the one in (4.9), the bound on the second term in (4.7) is of higher order as well.

$\square$

Using Efron-Stein's inequality one can produce the an upper bound on the variance of $f(X)$.

Corollary 4.3.2. *Let (Z, X) be a hidden Markov model as above. Let $g : \mathbb{A}^n \to \mathbb{R}$ be*

Lipschitz with constant c. Then

$$Var\left(g(X_1,\ldots,X_n) \le Cn^{1+\alpha},\right.$$

for any $\alpha > 0$, $C = C(|\mathcal{S}|)$ and n large enough.

Proof. As in Proposition 4.3.1 let $R = (R_0, \ldots, R_{|\mathcal{S}|(n-1)})$ be a vector of independent random variables and h be a function such that

$$g(X_1,\ldots,X_n) = h(R).$$

Let $R' = (R'_0, \ldots, R'_{|\mathcal{S}|(n-1)})$ be an independent copy of R. Applying Efron-Stein's inequality to $h(R)$ yields,

$$Var(h(R))) \le \frac{1}{2}\sum_{i=0}^{|\mathcal{S}|(n-1)} \mathbb{E}[(h(Z) - h(Z^i))^2],$$

with R^i defined as in Proposition 4.3.1.

By Proposition 4.3.1 there is $\alpha > 0$ and a constant $C = C(\alpha) > 0$, s.t

$$Var(h(R)) \le \frac{1}{2}(|S|(n-1)+1)Cn^{\alpha} \le C'n^{1+\alpha},$$

where $C' > 0$ is a function of $|S|$ and α. Finally, note that $Var(g(X_1,\ldots,X_n)) = Var(h(R))$. $\square$

Remark 4.3.3. *Note that the bound on the variance also follows from using an exponential bounded difference inequality for Markov chains proved by Paulin [22]. This holds for the general case when X is a Markov chain (not necessarily time homogeneous), taking values in a Polish space $\Lambda = \Lambda_1 \times \cdots \times \Lambda_n$, with mixing time τ_{min}. Then for any $t \ge 0$,*

$$\mathbb{P}(|f(X) - \mathbb{E}[f(X)]| \ge t) \le 2\exp\left(\frac{-2t^2}{||c||^2\tau_{min}}\right).$$

CHAPTER 5
CENTRAL LIMIT THEOREM

5.1 Introduction

So far we have obtained the rate of convergence of $\mathbb{E}[LC_n]$ and a closeness to the diagonal estimate. Furthermore we have outlined a way to modify the generalized perturbative approach introduced by Chatterjee for the case of sequences generated by hidden Markov models. Those are all the key results used in [16] to obtain a central limit theorem for LC_n in the independent case. In this chapter we follow the arguments in [16] to get the following result

Theorem 5.1.1. *Let $(Z,(X,Y))$ be a hidden Markov model. Assume $Var(LC_n) \geq Kn$ for some constant $K > 0$, independent of n. Then,*

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} \left| \mathbb{P}\left(\frac{LC_n - \mathbb{E}[LC_n]}{\sqrt{Var(LC_n)}} \leq x \right) - \mathbb{P}(\mathcal{N} \leq x) \right| = 0, \tag{5.1}$$

where $\mathcal{N}$ is the standard normal random variable.

Recall that there exists a vector $R = (R_0, \ldots, R_{|S|(n-1)})$ of independent random variables (though not identically distributed) taking values in $\mathcal{S} \times \mathcal{A}$ and a function $h : (\mathcal{S} \times \mathcal{A})^n \to \mathbb{R}$ such that

$$LC_n = h(R).$$

As in chapter 4, let R' be an independent copy of R, and if $A \subset \{0, \ldots, |S|(n-1)\}$, let

$$R_i^A = \begin{cases} R_i' & \text{if } i \in A \\ R_i & \text{if } i \notin A \end{cases}$$

Then the discrete derivative of g is defined for any $A \subsetneq \{0, \dots, |S|(n-1)\}$ and $i \notin A$, as

$$\Delta_i h^A = h(R^A) - h(R^{A \cup i}).$$

Let

$$T := \frac{1}{2} \sum_{A \subsetneq \{0, \dots, |S|(n-1)\}} \frac{1}{\binom{|S|(n-1)}{|A|}(|S|(n-1) - |A|)} \sum_{j \notin A} \Delta_j h \Delta_j h^A..$$

Theorem 4.1.3 implies that

$$d_W \left(\frac{h(R) - \mathbb{E}h(R)}{\sqrt{Var(h(R))}}, \mathcal{N} \right) \leq \frac{\sqrt{Var(T)}}{\sigma^2} + \frac{1}{2\sigma^3} \mathbb{E}|\Delta_j h|^3, \tag{5.2}$$

where $\sigma^2 := Var(LC_n) = Var(h(R))$ and d_W is the Monge-Kantorovich-Wasserstein distance. Note that here we use the variance $Var(T)$ as an upper bound of the conditional variance in Theorem 4.1.3, since it is a simpler object to study.

Assuming $\sigma^2 > Kn$ we show in Section 5.2 that the right-hand side of (5.2) converges to 0 as $n \to \infty$. This in turn will imply Theorem 5.1.1.

5.2 Proof of Theorem 5.1.1

We first note that $\mathbb{E}|\Delta_j g|^3/(2\sigma^3) \to 0$ as $n \to \infty$. Indeed, by Proposition 4.3.1 with $\alpha = 1$ and $r = 3$, there is a constant C such that

$$\mathbb{E}|\Delta_j h|^3 \leq Cn,$$

for all large n. Then $\sigma^3 > K^{3/2} n^{3/2}$, it follows that $\mathbb{E}|\Delta_j h|^3/(2\sigma^3) \to 0$ as $n \to \infty$.

We next turn to bounding Var(T) from above. It will be enough to show that Var$(T) = o(n^2)$ in order for Theorem 5.1.1 to hold.

To do so, we start by giving a variant of Theorem 3.2.3. Assume $n = vd$, and let the integers

$$0 = r_0 \leq r_1 \leq r_2 \leq \cdots \leq r_{d-1} \leq r_d = n, \tag{5.3}$$

be such that

$$LC_n = \sum_{i=1}^{d} |LCS(X_{v(i-1)+1} \ldots X_{vi}; Y_{r_{i-1}+1} \ldots Y_{r_i}|, \tag{5.4}$$

so the vector $r = (r_0, \ldots, r_d)$ determines an *optimal alignment* for the pair (X, Y).

As in Chapter 3 for any $p > 0$, let

$$\tilde{\gamma}_n(p) := \frac{\mathbb{E}[LCS(X_1, \ldots, X_n; Y_1, \ldots, Y_{np})]}{n(1+p)/2}.$$

Let $\epsilon > 0$, $p_1 > 0$ and $p_2 > 0$ be constants. Let A^n_{ϵ,p_1,p_2} be the set of optimal alignments of $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$, for which a proportion of at least $1 - \epsilon$ of the integer intervals $[r_{i-1}+1, r_i]$ for $i = 1, \ldots, d$ have their length between vp_1 and vp_2. More precisely, A^n_{ϵ,p_1,p_2} is the event that for all integer vectors $(r_0, r_1, \ldots, r_d)$ satisfying (5.3) and (5.4),

$$\text{Card}(\{i \in 1, 2, \ldots, d : vp_1 \leq r_i - r_{i-1} \leq vp_2\}) \geq (1 - \epsilon)d. \tag{5.5}$$

Let β^* and τ_{min} be the $\beta-$mixing coefficient and mixing time for the hidden Markov model as defined in Chapter 2.

Recall the following result regarding closeness to the diagonal from Chapter 3 (Corollary 3.3.2).

Proposition 5.2.1. *Let $\delta \in (6\beta^*, \gamma^*/2)$. Let $0 < p_1 < 1 < p_2$ be such that $\tilde{\gamma}_n(p_1) < \gamma^* - 2\delta$ and $\tilde{\gamma}_n(p_2) < \gamma^* - 2\delta$ for all n. Let $\alpha \in (0, 1)$, $d = n^{1-\alpha}$, $v = n^\alpha$ and*

$$\epsilon := c_1 \sqrt{\frac{1 + \ln(1 + n^\alpha)}{n^\alpha}},$$

where $c_1 > 0$ *such that*

$$c_1^2 \geq \frac{16\tau_{min}}{(\delta - 6\beta^*)^2},$$

and

$$c_1^2 \left(\frac{1+\ln(1+n^\alpha)}{n^\alpha}\right) \geq \frac{4C_A^2}{\delta^2}\frac{\ln n}{n},$$

for all $n \geq 1$. *Then,*

$$\mathbb{P}(A^n_{\epsilon,p_1,p_2}) \geq 1 - \exp\left(-n^{1-\alpha}(1+\ln(1+n^\alpha))\right),$$

for all $n \geq 1$.

We now focus on estimating the variance term in (5.2). Note that,

$$\begin{aligned} Var(T) = & \frac{1}{4} Var\left(\sum_{A \subsetneq \{0,\ldots,|S|(n-1)\}} \sum_{j \notin A} \frac{\Delta_j h(R)\Delta_j h(R^A)}{\binom{|S|(n-1)+1}{|A|}(|S|(n-1)+1-|A|)}\right) \\ = & \frac{1}{4} \sum_{(A,B,j,k)\in\mathcal{S}_1} \frac{Cov(\Delta_j h \Delta_j h^A, \Delta_k h \Delta_k h^B)}{\binom{|S|(n-1)+1}{|A|}(|S|(n-1)+1-|A|)\binom{|S|(n-1)+1}{|B|}(|S|(n-1)+1-|B|)}, \end{aligned}$$

where

$$\mathcal{S}_1 \coloneqq \{(A,B,j,k) : A \subsetneq \{0,\ldots,|S|(n-1)\}, B \subsetneq \{0,\ldots,|S|(n-1)\}, j \notin A, k \notin B\}.$$

The following proposition (see [16, Proposition 2.1]) will be useful at various points of the proof.

Proposition 5.2.2. *Let* $\mathcal{R}$ *be a subset of* $\{0,\ldots,|S|(n-1)\}^2$, *and let*

$$\mathcal{S}^* = \{(A,B,j,k) : A \subsetneq \{0,\ldots,|S|(n-1)\}, B \subsetneq \{0,\ldots,|S|(n-1)\}, j \notin A, k \notin B, (j,k) \in \mathcal{R}\}.$$

Then

$$\sum_{(A,B,j,k)\in\mathcal{S}^*} \frac{1}{\binom{|S|(n-1)+1}{|A|}(|S|(n-1)+1-|A|)\binom{|S|(n-1)+1}{|B|}(|S|(n-1)+1-|B|)} = |\mathcal{R}|.$$

Proof. Some elementary manipulations yield

$$\begin{aligned}
&\sum_{(A,B,j,k)\in\mathcal{S}^*} \frac{1}{\binom{|S|(n-1)+1}{|A|}(|S|(n-1)+1-|A|)\binom{|S|(n-1)+1}{|B|}(|S|(n-1)+1-|B|)} \\
= &\sum_{(j,k)\in\mathcal{R}} \sum_{s,r=0}^{|S|(n-1)} \sum_{\substack{A\not\ni j,|A|=s\\ B\not\ni k,|B|=r}} \frac{1}{\binom{|S|(n-1)+1}{s}(|S|(n-1)+1-s)\binom{|S|(n-1)+1}{r}(|S|(n-1)+1-r)} \\
= &\sum_{(j,k)\in\mathcal{R}} \sum_{s,r=0}^{|S|(n-1)} \frac{\binom{|S|(n-1)}{s}\binom{|S|(n-1)}{r}}{\binom{|S|(n-1)+1}{s}(|S|(n-1)+1-s)\binom{|S|(n-1)+1}{r}(|S|(n-1)+1-r)} \\
= &\sum_{(j,k)\in\mathcal{R}} \sum_{s,r=0}^{|S|(n-1)} \frac{1}{(|S|(n-1)+1)^2} \\
= &\ |R|
\end{aligned}$$

□

Note that by Proposition 4.3.1, if $\mathcal{R} = \{0,\dots,|S|(n-1)\}^2$,

$$\begin{aligned}
&\sum_{(A,B,j,k)\in\mathcal{S}_1} \left| \frac{Cov(\Delta_j h \Delta_j h^A, \Delta_k h \Delta_k h^B)}{\binom{|S|(n-1)+1}{|A|}(|S|(n-1)+1-|A|)\binom{|S|(n-1)+1}{|B|}(|S|(n-1)+1-|B|)} \right| \\
\leq &\ C(a)|\mathcal{R}|n^a \\
= &\ O(n^{2+a}),
\end{aligned}$$

where $a > 0$ and $C(a) > 0$ is a constant that depends on a and the parameters of the model. In particular, it is not enough to apply Proposition 5.2.2 to get that $Var(T) = o(n^2)$. To get the desired asymptotic behavior we use the optimal alignment introduced in (5.4).

For notational convenience, below we write $\sum_1$ for $\sum_{(A,B,j,k)\in\mathcal{S}_1}$. Also, for random variables U, V and a random variable Z taking its values in $\mathbb{R}$, we write $Cov_{Z=z}(U, V)$ for $\mathbb{E}[(U - \mathbb{E}[U])(V - \mathbb{E}[V])|Z = z]$,$z \in \mathbb{R}$. Moreover, let

$$k_{n,A,B} = \frac{1}{\binom{|S|(n-1)+1}{|A|}(|S|(n-1)+1-|A|)\binom{|S|(n-1)+1}{|B|}(|S|(n-1)+1-|B|)}.$$

Let, now, the random variable Z be the indicator function of the event A^n_{ϵ,p_1,p_2} where

$$\epsilon := c_1\sqrt{\frac{1+\ln(1+n^\alpha)}{n^\alpha}},$$

and c_1 as in Property 5.2.1. Then

$$\begin{aligned} Var(T) = & \sum_1 k_{n,A,B} Cov(\Delta_j h \Delta_j h^A, \Delta_k h \Delta_k h^B) \\ = & \sum_1 k_{n,A,B} Cov_{Z=0}(\Delta_j h \Delta_j h^A, \Delta_k h \Delta_k h^B)\mathbb{P}(Z=0) \\ & + \sum_1 k_{n,A,B} Cov_{Z=1}(\Delta_j h \Delta_j h^A, \Delta_k h \Delta_k h^B)\mathbb{P}(Z=1). \end{aligned} \tag{5.6}$$

To estimate the first term on the right-hand side of (5.6), note that $h(R) \leq n$ because $h(R) = LC_n$. Then by Proposition 5.2.2 and Proposition 5.2.1,

$$\begin{aligned} & \sum_1 k_{n,A,B} Cov_{Z=0}(\Delta_j h \Delta_j h^A, \Delta_k h \Delta_k h^B)\mathbb{P}(Z=0) \\ \leq \quad & \sum_1 k_{n,A,B} n^2 \mathbb{P}(Z=0) \\ \leq \quad & C n^4 e^{-n^{1-\alpha}(1+\ln(1+n^\alpha))}. \end{aligned} \tag{5.7}$$

For the second term in the right-hand side of (5.6) note that

$$
\begin{aligned}
&\sum_{1} k_{n,A,B} Cov_{Z=1}(\Delta_j h \Delta_j h^A, \Delta_k h \Delta_k h^B)\mathbb{P}(Z=1) \\
\leq &\quad \sum_{1} k_{n,A,B} \left| Cov_{Z=1}(\Delta_j h \Delta_j h^A, \Delta_k h \Delta_k h^B)\right| \mathbb{P}(Z=1) \\
\leq &\quad \sum_{1} k_{n,A,B} \left| Cov_{Z=1}(\Delta_j h \Delta_j h^A, \Delta_k h \Delta_k h^B)\right| . \qquad (5.8)
\end{aligned}
$$

Before we proceed with more careful estimates on right-hand side of (5.8) we show the following simple bound of the terms in the sum.

Proposition 5.2.3. *Let* $(A, B, j, k) \in \mathcal{S}_1$. *For any* $a > 0$*,*

$$
\left| Cov_{Z=1}(\Delta_j h \Delta_j h^A, \Delta_k h \Delta_k h^B)\right| \leq C n^a,
$$

where $C > 0$ *depends only on the parameters of the hidden Markov models and for all* n *large enough.*

Proof. Let $a > 0$. By Proposition 4.3.1, for all n large enough,

$$
\left| Cov(\Delta_j h \Delta_j h^A, \Delta_k h \Delta_k h^B)\right| \leq C n^a.
$$

Now note that

$$
\begin{aligned}
&\left| Cov(\Delta_j h \Delta_j h^A, \Delta_k h \Delta_k h^B)\right| \\
= &\quad \Big| Cov_{Z=0}(\Delta_j h \Delta_j h^A, \Delta_k h \Delta_k h^B)\mathbb{P}(Z=0) \\
&\qquad + Cov_{Z=1}(\Delta_j h \Delta_j h^A, \Delta_k h \Delta_k h^B)\mathbb{P}(Z=1)\Big| \\
\geq &\quad -\left| Cov_{Z=0}(\Delta_j h \Delta_j h^A, \Delta_k h \Delta_k h^B)\right| \mathbb{P}(Z=0) \\
&\qquad + \left| Cov_{Z=1}(\Delta_j h \Delta_j h^A, \Delta_k h \Delta_k h^B)\right| \mathbb{P}(Z=1),
\end{aligned}
$$

and then using (5.7),

$$\begin{aligned} & \left|Cov_{Z=1}(\Delta_j h \Delta_j h^A, \Delta_k h \Delta_k h^B)\right| \\ \leq \quad & Cn^a + Cn^4 e^{-n^{1-\alpha}(1+\ln(1+n^\alpha))} \\ \leq \quad & Cn^a, \end{aligned}$$

for all n large enough. □

Next we estimate the right-hand side of (5.8) using properties arising from the optimal alignment. Recall that for each pair of words (X, Y) there is an optimal alignment $\mathbf{r}$ satisfying (5.3) and (5.4). Note that conditionally on $\{Z = 1\}$, $\mathbf{r}$ satisfies (5.5).

Definition 5.2.4. For the optimal alignment $\mathbf{r}$, each of the sets

$$\{X_{n^\alpha(i-1)+1}, \ldots X_{n^\alpha i}; Y_{r_{i-1}+1} \ldots, Y_{r_i}\}, \qquad i = 1, \ldots, n^{1-\alpha}$$

is called a cell of $\mathbf{r}$.

Let P_j be the set of cells (zero, one or two) of the optimal alignment potentially modified by a change in the instruction h_j. Assume that h_j determines $X_{j'}$ and $Y_{j'}$. Note that P_j is empty if the instruction h_j is not used, and the other two cases correspond to $X_{j'}$ and $Y_{j'}$ being in the same or different cells respectively.

Define the following subsets of $\mathcal{S}_1$ with respect to the alignment $\mathbf{r}$:

$$\mathcal{S}_{1,1} = \{(A, B, j, k) \in \mathcal{S}_1 : h_j \text{ and } h_k \text{ influence the same cell in } \mathbf{r}\},$$

and

$$\mathcal{S}_{1,2} = \{(A, B, j, k) \in \mathcal{S}_1 : h_j \text{ and } h_k \text{ influence different cells in } \mathbf{r}\},$$

Clearly, $\mathcal{S}_{1,1} \cap \mathcal{S}_{1,2} = \emptyset$ and $\mathcal{S}_1 = \mathcal{S}_{1,1} \cup \mathcal{S}_{1,2}$. Now, for a given subset S of $\mathcal{S}_1$, and for $(A, B, j, k) \in \mathcal{S}_1$, define

$$Cov_{Z=1,(A,B,j,k),S}(U, V) = \mathbb{E}[(U - \mathbb{E}[U])(V - \mathbb{E}[V])\mathbf{1}_{(A,B,j,k)\in S}|Z = 1].$$

When (A, B, j, k) is understood from context we write $Cov_{Z=1,\mathcal{S}}(U, V)$ instead.
The right-hand side of (5.8) is then bounded by

$$\begin{aligned}
&\sum_1 k_{n,A,B} \left|Cov_{Z=1}(\Delta_j h\Delta_j h^A, \Delta_k h\Delta_k h^B)\right| \\
\leq \quad & \sum_1 k_{n,A,B} \left|Cov_{Z=1,\mathcal{S}_{1,1}}(\Delta_j h\Delta_j h^A, \Delta_k h\Delta_k h^B)\right| \\
& + \sum_1 k_{n,A,B} \left|Cov_{Z=1,\mathcal{S}_{1,2}}(\Delta_j h\Delta_j h^A, \Delta_k h\Delta_k h^B)\right| \qquad (5.9)
\end{aligned}$$

Next, note that by the definition of $Cov_{Z=1,\mathcal{S}_{1,1}}$ and Proposition 5.2.3,

$$\begin{aligned}
&\sum_1 k_{n,A,B} \left|Cov_{Z=1,\mathcal{S}_{1,1}}(\Delta_j h\Delta_j h^A, \Delta_k h\Delta_k h^B)\right| \\
\leq \quad & \sum_1 k_{n,A,B}\mathbb{E}\bigg[\Big|(\Delta_j h\Delta_j h^A - \mathbb{E}[\Delta_j h\Delta_j h^A]) \\
& \qquad (\Delta_k h\Delta_k h^B - \mathbb{E}[\Delta_k h\Delta_k h^B])\Big|\mathbf{1}_{(A,B,j,k)\in\mathcal{S}_{1,1}}\Big| Z = 1\bigg] \\
\leq \quad & Cn^a\mathbb{E}[|\mathcal{R}||Z = 1],
\end{aligned}$$

where

$$\mathcal{R} = \{(j, k) \in \{0, \dots, |\mathcal{S}|(n-1)\}^2 : R_j \text{ and } R_k \text{ influence the same cell in } \mathbf{r}\}.$$

Let $\mathcal{R}_i$ be the number of pairs of indices $(j, k) \in \{0, \dots, |\mathcal{S}|(n-1)\}^2$ that influence the ith-cell. Recall that given $Z = 1$, the number of cells with more than $n^\alpha(1 + p_2)$ letters is less than $\epsilon n^{1-\alpha}$. If $\alpha > 2/3$, then $\epsilon n^{1-\alpha} < 1$ for large n and then all cells have no more than $n^{\alpha}(1+p_2)$ letters given that $Z=1$.

To estimate $|\mathcal{R}_i|$ note that by Proposition 4.3.1 the probability that a change of an instruction R_j for a pair $(X_{j'}, Y_{j'})$ leads to changes in pairs $(X_{k'}, Y_{k'})$ decays exponentially with the distance $|j'-k'|$. Therefore,

$$\mathbb{E}[|\mathcal{R}_i||Z=1] \leq (Cn^a + (1+p_2)n^{\alpha})^2 \leq Cn^{2\alpha}.$$

Now since

$$\mathbb{E}[|\mathcal{R}||Z=1] = \sum_{i=1}^{n^{1-\alpha}} \mathbb{E}[|\mathcal{R}_i||Z=1],$$

it follows that

$$\begin{aligned} & \sum_{1} k_{n,A,B} \left| Cov_{Z=1,\mathcal{S}_{1,1}}(\Delta_j h \Delta_j h^A, \Delta_k h \Delta_k h^B) \right| \\ \leq \quad & Cn^a n^{1-\alpha} n^{2\alpha} \\ = \quad & Cn^{1+\alpha+a}. \end{aligned} \tag{5.10}$$

We move next to the estimation of the second term of the right-hand side of (5.9) which is given by

$$\sum_{1} k_{n,A,B} \left| Cov_{Z=1,\mathcal{S}_{1,2}}(\Delta_j h \Delta_j h^A, \Delta_k h \Delta_k h^B) \right|.$$

To estimate the covariance terms we need to decompose them in such a way that independence of certain random variables leads to a simplified expression. For this purpose, for each $i \in \{0, \ldots, |\mathcal{S}|(n-1)\}$ let P_i be the set of cells in the optimal alignment that contains the letter determined by the R_i instruction. Then let

$$\tilde{\Delta}_i h = h(P_i) - h(P_i'),$$

where P_i' is the same as P_i except that the instruction R_i is replaced by the independent copy R_i'. While the change of R_i can lead to multiple changes among the realization (X, Y), the variable $\tilde{\Delta}_i h$ is the difference between the length of the longest common subsequence before and after the modification restricted to the cell P_i. Now for $(A, B, j, k) \in \mathcal{S}_1$,

$$\begin{aligned}
& Cov_{Z=1,\mathcal{S}_{1,2}}(\Delta_j h \Delta_j h^A, \Delta_k h \Delta_k h^B) \\
= \quad & Cov_{Z=1,\mathcal{S}_{1,2}}((\Delta_j h - \tilde{\Delta}_j h)\Delta_j h^A, \Delta_k h \Delta_k h^B) \\
& + Cov_{Z=1,\mathcal{S}_{1,2}}(\tilde{\Delta}_j h(\Delta_j h^A - \tilde{\tilde{\Delta}}_j h^A), \Delta_k h \Delta_k h^B) \\
& + Cov_{Z=1,\mathcal{S}_{1,2}}(\tilde{\Delta}_j h \tilde{\tilde{\Delta}}_j h^A, (\Delta_k h - \tilde{\Delta}_k h)\Delta_k h^B) \\
& + Cov_{Z=1,\mathcal{S}_{1,2}}(\tilde{\Delta}_j h \tilde{\tilde{\Delta}}_j h^A, \tilde{\Delta}_k h(\Delta_k h^B - \tilde{\tilde{\Delta}}_k h^B)) \\
& + Cov_{Z=1,\mathcal{S}_{1,2}}(\tilde{\Delta}_j h \tilde{\tilde{\Delta}}_j h^A, \tilde{\Delta}_k h \tilde{\tilde{\Delta}}_k h^B),
\end{aligned} \tag{5.11}$$

where for any $i \notin A$, we also set $\tilde{\tilde{\Delta}}_j h^A = h^A|_{P_i} - h^{A\cup\{i\}}|_{P_i}$, i.e., the restriction is the same as in $\tilde{\Delta}_i h$ but there is a set A of instruction flipped as well.

The first technical assumption we make (which will be addressed in future work) concerns the contribution of the last term in (5.11). In particular,

Assumption 1. Assume

$$\sum_1 k_{n,A,B} |Cov_{Z=1,\mathcal{S}_{1,2}}(\tilde{\Delta}_j h \tilde{\tilde{\Delta}}_j h^A, \tilde{\Delta}_k h \tilde{\tilde{\Delta}}_k h^B)| \leq C n^{1+\alpha}.$$

for all n large enough.

The second assumption we make concerns the distribution of $\Delta_j h - \tilde{\Delta}_j h$.

Assumption 2. If $(A, B, j, k) \in \mathcal{S}_{1,2}$, then

$$|\Delta_j h^A - \tilde{\tilde{\Delta}}_j h^A| =_d |\Delta_j h - \tilde{\Delta}_j h|.$$

We first estimate the first term in (5.11). It is given by

$$Cov_{Z=1,\mathcal{S}_{1,2}}((\Delta_j h - \tilde{\Delta}_j h)\Delta_j h^A, \Delta_k h \Delta_k h^B).$$

Let

$$U := (\Delta_j h - \tilde{\Delta}_j h)\Delta_j h^A,$$

and

$$V := \Delta_k h \Delta_k h^B,$$

so that the first term is equivalent to $Cov_{Z=1,\mathcal{S}_{1,2}}(U,V)$. Note that

$$\begin{aligned}
& |Cov_{Z=1,\mathcal{S}_{1,2}}(U,V)| \\
= \quad & |\mathbb{E}[(U-\mathbb{E}[U])(V-\mathbb{E}[V])\mathbf{1}_{(A,B,j,k)\in\mathcal{S}_{1,2}}|Z=1]| \\
\leq \quad & \mathbb{E}[|UV|\mathbf{1}_{(A,B,j,k)\in\mathcal{S}_{1,2}}|Z=1] \\
& +\mathbb{E}|V|\mathbb{E}[|U|\mathbf{1}_{(A,B,j,k)\in\mathcal{S}_{1,2}}|Z=1] \\
& +\mathbb{E}|U|\mathbb{E}[|V|\mathbf{1}_{(A,B,j,k)\in\mathcal{S}_{1,2}}|Z=1] \\
& +\mathbb{E}|U|\mathbb{E}|V|\mathbb{E}[\mathbf{1}_{(A,B,j,k)\in\mathcal{S}_{1,2}}|Z=1] \\
:= \quad & T_1+T_2+T_3+T_4.
\end{aligned}$$

Note that

$$\begin{aligned}
T_1 :=& \mathbb{E}[|UV|\mathbf{1}_{(A,B,j,k)\in\mathcal{S}_{1,2}}|Z=1] \\
\leq & \quad \frac{\mathbb{E}|UV|}{\mathbb{P}(Z=1)} \\
\leq & \quad C\mathbb{E}|UV|,
\end{aligned}$$

where we have used that $P(Z = 0)$ is exponentially small for large n by Theorem 3.2.3. Now since $|\Delta_j h - \tilde{\Delta}_j h| \leq Cn$, Property 4.3.1 implies that

$$\mathbb{E}|UV| \leq Cn^a \mathbb{E}|\Delta_j h - \tilde{\Delta}_j h| + Cn^4 e^{-n^{1-\alpha}(1+\ln(1+n^\alpha))}$$

for large n. Similar approach yields the same upper bound for T_2, T_3 and T_4. Then,

$$|Cov_{Z=1,\mathcal{S}_{1,2}}(U,V)| \leq Cn^a \mathbb{E}|\Delta_j h - \tilde{\Delta}_j h| + Cn^4 e^{-n^{1-\alpha}(1+\ln(1+n^\alpha))}.$$

Now by Assumption 2, and using the symmetry between A and B and between j and k, each of the first four terms in (5.11) is bounded above by

$$Cn^a \mathbb{E}|\Delta_j h - \tilde{\Delta}_j h| + Cn^4 e^{-n^{1-\alpha}(1+\ln(1+n^\alpha))}.$$

Together with Assumption 1, we can bound second term of the right-hand side of (5.9)

$$\begin{aligned} & \sum_1 k_{n,A,B} |Cov_{Z=1,\mathcal{S}_{1,2}}(\Delta_j h \Delta_j h^A, \Delta_k h \Delta_k h^B)| \\ \leq \quad & n^{1+\alpha} + 4 \sum_1 k_{n,A,B} Cn^a \mathbb{E}|\Delta_j h - \tilde{\Delta}_j h| + Cn^4 e^{-n^{1-\alpha}(1+\ln(1+n^\alpha))}. \end{aligned}$$

Next by superadditivity, we immediately get

$$\Delta_j h \leq \tilde{\Delta}_j h.$$

By independence, $\mathbb{E}[\Delta_j h] = 0$, so

$$\begin{aligned} & \sum_1 k_{n,A,B} |Cov_{Z=1,\mathcal{S}_{1,2}}(\Delta_j h \Delta_j h^A, \Delta_k h \Delta_k h^B)| \\ \leq \quad & n^{1+\alpha} + 4 \sum_1 k_{n,A,B} Cn^a \mathbb{E}[\tilde{\Delta}_j h] + Cn^4 e^{-n^{1-\alpha}(1+\ln(1+n^\alpha))}. \end{aligned} \tag{5.12}$$

Our final assumption is regarding a bound on the middle term above.

Assumption 3. Assume

$$4\sum_{1} k_{n,A,B} C n^{a} \mathbb{E}[\tilde{\Delta}_j h] \leq C n^{1+a+\alpha}.$$

Then (5.7), (5.10), and (5.12) imply

$$\begin{aligned} Var(T) \leq & C n^{1+a+\alpha} + C n^{4} e^{-n^{1-\alpha}(1+\ln(1+n^{\alpha}))} \\ \leq & \quad C n^{1+a+\alpha}. \end{aligned}$$

Therefore, Theorem 4.1.3 ensures that

$$d_W\left(\frac{LC_n - \mathbb{E}LC_n}{\sqrt{Var LC_n}}, \mathcal{N}\right) \leq C\frac{1}{n^{\frac{1-a-\alpha}{2}}}.$$

for all $n \geq 1$ and with $C > 0$ a constant independent of n.

CHAPTER 6
CONCLUSION

In this thesis we have studied the asymptotic behavior of the length of the longest common subsequences of two strings generated by a hidden Markov model. Under some standard assumptions regarding the model we have generalized results of Chvátal and Sankoff [7] and Alexander [1] for the convergence of $\mathbb{E}[LC_n]$. We have also obtained versions for the closeness of the diagonal estimate of Houdré and Matzinger [18] and we have outlined a proof for a central limit theorem by building upon work by Houdré and Işlak [16] and adapting a Stein's method estimate by Chatterjee [6].

For the central limit theorem proof to be complete one needs to provide a linear lower bound on $Var(LC_n)$ - a result that is not yet obtained even for the iid case. Another point that needs to be addressed are the two technical assumptions in our outline for the proof.

A second direction for future work concerns other models with a dependence structure. For instance one can try to obtain similar results for the rate of convergence or even a central limit theorem by only assuming a suitable mixing condition for the sequences.

Finally, the new version of Stein's method for dependent sequences, that is Theorem 4.1.3, can be applied to Markovian equivalents of other models, like the ones discussed in [19]. Some instances are the stochastic coverage process, and the problem of set approximation with random tessellations. Furthermore, a Stein's method for functions of dependent random variables should be obtainable directly using the dependency structure, especially if the variance of the function in question is of higher order. Such would be the case when the underlying model is the two dimensional Ising model at critical temperature.

REFERENCES

[1] K. S. Alexander, "The rate of convergence of the mean length of the longest common subsequence," *Ann. Appl. Probab.*, vol. 4, no. 4, pp. 1074–1082, 1994.

[2] H. C. P. Berbee, *Random walks with stationary increments and renewal theory*, ser. Mathematical Centre Tracts. Mathematisch Centrum, Amsterdam, 1979, vol. 112, pp. iii+223, ISBN: 90-6196-182-3.

[3] R. C. Bradley, "Approximation theorems for strongly mixing random variables," *Michigan Math. J.*, vol. 30, no. 1, pp. 69–81, 1983.

[4] ——, "Basic properties of strong mixing conditions. A survey and some open questions," *Probab. Surv.*, vol. 2, pp. 107–144, 2005, Update of, and a supplement to, the 1986 original.

[5] ——, *Introduction to strong mixing conditions. Vol. 1*. Kendrick Press, Heber City, UT, 2007, pp. xviii+539, ISBN: 0-9740427-6-5.

[6] S. Chatterjee, "A short survey of Stein's method," in *Proceedings of the International Congress of Mathematicians—Seoul 2014. Vol. IV*, Kyung Moon Sa, Seoul, 2014, pp. 1–24.

[7] V. Chvátal and D. Sankoff, "Longest common subsequences of two random sequences," *J. Appl. Probability*, vol. 12, pp. 306–315, 1975.

[8] P. Diaconis, "Stein's method for Markov chains: First examples," in *Stein's method: expository lectures and applications*, ser. IMS Lecture Notes Monogr. Ser. Vol. 46, Inst. Math. Statist., Beachwood, OH, 2004, pp. 27–43.

[9] W. Doeblin, "Exposé de la théorie des chaînes simples constantes de markoff à un nombre fini d'états," *Revue Math. de l'Union Interbalkanique*, vol. 2, pp. 77–105, 1938.

[10] P. Doukhan, *Mixing*, ser. Lecture Notes in Statistics. Springer-Verlag, New York, 1994, vol. 85, pp. xii+142, Properties and examples, ISBN: 0-387-94214-9.

[11] R. Durbin, S. Eddy, and A. Krogh, *Biological sequence analysis*. Cambridge University Press, 1998.

[12] W. Feller, *An introduction to probability theory and its applications. Vol. I*, ser. Third edition. John Wiley & Sons, Inc., New York-London-Sydney, 1968, pp. xviii+509.

[13] J. Fulman, "Stein's method and non-reversible Markov chains," in *Stein's method: expository lectures and applications*, ser. IMS Lecture Notes Monogr. Ser. Vol. 46, Inst. Math. Statist., Beachwood, OH, 2004, pp. 69–77.

[14] S. Goldstein, "Maximal coupling," *Z. Wahrsch. Verw. Gebiete*, vol. 46, no. 2, pp. 193–204, 1978/79.

[15] S. Holmes, "Stein's method for birth and death chains," in *Stein's method: expository lectures and applications*, ser. IMS Lecture Notes Monogr. Ser. Vol. 46, Inst. Math. Statist., Beachwood, OH, 2004, pp. 45–67.

[16] C. Houdré and U. Işlak, "A central limit theorem for the length of the longest common subsequences in random words," 2018. arXiv: `1408.1559 [math.PR]`.

[17] C. Houdré and G. Kerchev, "On the rate of convergence for the length of the longest common subsequences in hidden markov models," 2018. arXiv: `1712.09881 [math.PR]`.

[18] C. Houdré and H. Matzinger, "Closeness to the diagonal for longest common subsequences in random words," *Electron. Commun. Probab.*, vol. 21, Paper No. 36, 19, 2016.

[19] R. Lachièze-Rey and G. Peccati, "New Berry-Esseen bounds for functionals of binomial point processes," *Ann. Appl. Probab.*, vol. 27, no. 4, pp. 1992–2031, 2017.

[20] J. Lember, H. Matzinger, and F. Torres, "The rate of the convergence of the mean score in random sequence comparison," *Ann. Appl. Probab.*, vol. 22, no. 3, pp. 1046–1058, 2012.

[21] D. A. Levin and Y. Peres, *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2017, pp. xvi+447, Second edition of [MR2466937], With contributions by Elizabeth L. Wilmer, With a chapter on "Coupling from the past" by James G. Propp and David B. Wilson, ISBN: 978-1-4704-2962-1.

[22] D. Paulin, "Concentration inequalities for Markov chains by Marton couplings and spectral methods," *Electron. J. Probab.*, vol. 20, no. 79, 32, 2015.

[23] W. T. Rhee, "On rates of convergence for common subsequences and first passage time," *Ann. Appl. Probab.*, vol. 5, no. 1, pp. 44–48, 1995.

[24] E. Rio, *Asymptotic theory of weakly dependent random processes*, ser. Probability Theory and Stochastic Modelling. Springer, Berlin, 2017, vol. 80, pp. xviii+204, Translated from the 2000 French edition [MR2117923], ISBN: 978-3-662-54322-1; 978-3-662-54323-8.

[25] G. O. Roberts and J. S. Rosenthal, "General state space Markov chains and MCMC algorithms," *Probab. Surv.*, vol. 1, pp. 20–71, 2004.

[26] J. M. Steele, "Long common subsequences and the proximity of two random strings," *SIAM J. Appl. Math.*, vol. 42, no. 4, pp. 731–737, 1982.

[27] ——, *Probability theory and combinatorial optimization*, ser. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997, vol. 69, pp. viii+159, ISBN: 0-89871-380-3.

[28] H. Thorisson, *Coupling, stationarity, and regeneration*, ser. Probability and its Applications (New York). Springer-Verlag, New York, 2000, pp. xiv+517, ISBN: 0-387-98779-7.