

DIPARTIMENTO DI MATEMATICA
POLITECNICO DI MILANO

**Lower bounds for moments of global scores of
pairwise Markov chains**

Lember, J.; Matzinger, H., Sova, J.; Zucca, F.

Collezione dei *Quaderni di Dipartimento*, numero **QDD 220**
Inserito negli *Archivi Digitali di Dipartimento* in data 21-03-2016



Piazza Leonardo da Vinci, 32 - 20133 Milano (Italy)

LOWER BOUNDS FOR MOMENTS OF GLOBAL SCORES OF PAIRWISE MARKOV CHAINS

JÜRI LEMBER, HEINRICH MATZINGER, JOONAS SOVA, AND FABIO ZUCCA

ABSTRACT. Let X_1, X_2, \dots and Y_1, Y_2, \dots be two random sequences so that every random variable takes values in a finite set \mathbb{A} . We consider a global similarity score $L_n := L(X_1, \dots, X_n; Y_1, \dots, Y_n)$ that measures the homology (relatedness) of words (X_1, \dots, X_n) and (Y_1, \dots, Y_n) . A typical example of such score is the length of the longest common subsequence. We study the order of central absolute moment $E|L_n - EL_n|^r$ in the case where the two-dimensional process $(X_1, Y_1), (X_2, Y_2), \dots$ is a Markov chain on $\mathbb{A} \times \mathbb{A}$. This is a very general model involving independent Markov chains, hidden Markov models, Markov switching models and many more. Our main result establishes a general condition that guarantees that $E|L_n - EL_n|^r \asymp n^{\frac{r}{2}}$. We also perform simulations indicating the validity of the condition.

Keywords: Random sequence comparison, longest common sequence, fluctuations, Waterman conjecture.
AMS subject classification: 60K35, 41A25, 60C05

1. INTRODUCTION

1.1. **Sequence comparison setting.** Throughout this paper $X = (X_1, X_2, \dots, X_n)$ and $Y = (Y_1, Y_2, \dots, Y_n)$ are two random strings, usually referred as sequences, so that every random variable X_i and Y_i take values on a finite alphabet \mathbb{A} . Since the sequences X and Y are not necessarily independent nor identically distributed, it is convenient to consider the two-dimensional sequence $Z = ((X_1, Y_1), \dots, (X_n, Y_n))$. The sample space of Z will be denoted by \mathcal{Z}_n . Clearly $\mathcal{Z}_n \subseteq (\mathbb{A} \times \mathbb{A})^n$ but, depending on the model, the inclusion can be strict. The problem of measuring the similarity of X and Y is central in many areas of applications including computational molecular biology [10, 15, 40, 42, 46] and computational linguistics [33, 34, 36, 37]. In this paper, we adopt the same notation as in [31], namely we consider a general scoring scheme, where $S : \mathbb{A} \times \mathbb{A} \rightarrow \mathbb{R}^+$ is a *pairwise scoring function* that assigns a score to each couple of letters from \mathbb{A} . An *alignment* is a pair (ρ, τ) where $\rho = (\rho_1, \rho_2, \dots, \rho_k)$ and $\tau = (\tau_1, \tau_2, \dots, \tau_k)$ are two increasing sequences of natural numbers, i.e. $1 \leq \rho_1 < \rho_2 < \dots < \rho_k \leq n$ and $1 \leq \tau_1 < \tau_2 < \dots < \tau_k \leq n$. The integer k is the number of aligned letters, $n - k$ is the number of non-aligned letters. Given the pairwise scoring function S the score of the alignment (ρ, τ) when aligning X and Y is defined by

$$U_{(\rho, \tau)}(X, Y) := \sum_{i=1}^k S(X_{\rho_i}, Y_{\tau_i}) + \delta(n - k),$$

where $\delta \in \mathbb{R}$ is another scoring parameter. Typically $\delta \leq 0$ so that many non-aligned letters in the alignment reduce the score. If $\delta \leq 0$, then its absolute value $|\delta|$ is often called the *gap penalty*. Given S and δ , the optimal alignment score of X and Y is defined to be

$$L_n := L(X, Y) = L(Z) := \max_{(\rho, \tau)} U_{(\rho, \tau)}(X, Y), \tag{1.1}$$

where the maximum above is taken over all possible alignments. Sometimes, when we talk about a *string comparison model*, we refer to the study of L_n for given sequences X and Y , score function S and gap penalty δ . It is important to note that for any constant gap price $\delta \in \mathbb{R}$, changing the value of one of the $2n$ random variables $X_1, \dots, X_n, Y_1, \dots, Y_n$ changes the value of L_n by at most Δ , where

$$\Delta := \max_{u, v, w \in \mathbb{A}} |S(u, v) - S(u, w)|. \tag{1.2}$$

When $\delta = 0$ and the scoring function assigns one to every pair of similar letters and zero to all other pairs, i.e.

$$S(a, b) = \begin{cases} 1, & \text{if } a = b; \\ 0, & \text{if } a \neq b. \end{cases} \quad (1.3)$$

then $L(Z)$ is just the maximal number of aligned letters, also called the length of the *longest common subsequence* (abbreviated by LCS) of X and Y . In this article, to distinguish the length of LCS from another scoring schemes, we shall denote it via $\ell_n := \ell(Z) = \ell(X, Y)$. In other words $\ell(Z)$ is the maximal k so that there there exists an alignment (ρ, τ) such that $X_{\rho_i} = Y_{\tau_i}$, $i = 1, \dots, k$. Note that the optimal alignment (ρ, τ) as well as the longest common subsequence $X_{\rho_1}, \dots, X_{\rho_k}$ is not typically unique. The length of LCS is probably the most important and the most studied measure of global similarity between strings.

1.2. History and overview. The problem of measuring the similarity of two strings is of central importance in many applications including computational molecular biology, linguistics etc. For instance, in computational molecular biology, the similarity of two sequences (for example DNA- or proteins) is used to determine their homology (relatedness) – similar strings are more likely to be the decedents of the same ancestor. Out of all possible similarity measures, the global score $L(X, Y)$, in particular the length of LCS, is probably the most common measure of similarity. Its popularity is partially due to the well-known dynamic programming algorithms (so-called Needleman-Wunsch algorithm) that allows to calculate the optimal alignment with complexity $O(n^2)$ and the score with complexity $O(n)$ [10, 15, 40, 46, 9].

Unfortunately, although easy to apply and define, it turns out that the theoretical study of L_n is very difficult. It is easy to see that the global score is superadditive. This implies that when Z is taken from an ergodic process, by Kingman’s subadditive ergodic theorem, there exists a constant γ^* such that

$$\frac{L_n}{n} \rightarrow \gamma^* \quad \text{a. s. and in } L_1, \quad \text{as } n \rightarrow \infty. \quad (1.4)$$

In the LCS case, the existence of γ^* was first shown by Chvátal and Sankoff [11], but its exact value (or an expression for it), although well estimated, remains unknown even for i.i.d. Bernoulli sequences. Alexander [1] established the rate of convergence of the left hand side of equation (1.4) in the LCS case, a result extended by Lember, Matzinger and Torres [30] to general scoring functions.

In their leading paper [11], Chvátal and Sankoff first studied the the asymptotic order of $\text{Var}(\ell_n)$ and based on some simulations, they conjectured that $\text{Var}(\ell_n) = o(n^{2/3})$, for X and Y independent i.i.d. symmetric Bernoulli. In the case of independent i.i.d. sequences, it follows from Efron-Stein inequality (see, e.g. [7]) that

$$\text{Var}(L_n) \leq C_2 n, \quad \text{for all } n \in \mathbb{N}, \quad (1.5)$$

where $C_2 > 0$ is an universal constant, independent of n . For the LCS case, this result was proved by Steele [43]. In [45], Waterman asked whether or not the linear bound on the variance can be improved, at least in the LCS case. His simulations showed that, in some special cases (including the LCS case), $\text{Var}(L_n)$ should grow linearly in n . These simulations suggest the linear lower bound $\text{Var}(\ell_n) \geq c \cdot n$, which would invalidate the conjecture of Chvátal and Sankoff. In the past ten years, the asymptotic behavior of $\text{Var}(\ell_n)$ has been investigated by Bonetto, Durringer, Houdré, Lember, Matzinger and Torres, under various choices of independent sequences X and Y (cf. [5], [18], [22], [29], [31],[32] [20], [25] etc). In particular, in [25] and [31] a general approach for obtaining the lower bound for moments $E\Phi(L_n - EL_n)$, where Φ is a convex increasing function we worked out. For more detailed history of the problem as well as the connection between the rate of central absolute moments of L_n and the central limit theorem $\sqrt{n}(L_n - EL_n)$, we refer to [25].

In this paper, we follow the general approach developed in [31] and [25], but unlike all previous papers, we apply it for sequences that are not necessarily independent and i.i.d. Indeed, in the present paper, we assume that Z consists of n observations of aperiodic stationary Markov chain. Following W. Pieczynski, we call such a model as *pairwise Markov chain* (PMC) [41, 14, 16]. It is important to realize that the Markov

property of Z does not imply the one of marginal processes X and Y . On the other hand, it is not hard to see that conditionally on X , the Y process is a Markov chain and, obviously, vice versa [41]. Hence the name – pairwise Markov chains. Thus, we do not assume that X and Y are both Markov chains, although often this is the case. So, our model is actually rather general one including as a special case hidden Markov models (HMM's), Markov switching models, HMM's with dependent noise [16] and also the important case where X and Y are independent Markov chains or even i.i.d.. Except [31], where among others also some specific independent non-i.i.d. sequences were considered, all previous articles cited above deal with the case when X and Y are independent i.i.d. sequences.

The paper is organized as follows. In Section 2, we state a very general theorem – Theorem 2.1 – for obtaining the lower bound of $E\Phi(|L_n - EL_n|)$ for any model (Theorem 2.1). The proof of Theorem 2.1 is a generalization of Theorem 3.2 in [25] and therefore, we prove it in the appendix. Theorem 2.1 does not assume any particular stochastic model Z , instead it requires a specific *random transformation* \mathcal{R} and random vectors U, V so that several general assumptions listed as **A1** – **A4** are satisfied. The objective of the present paper is to show that under PMC-model, a random transformation \mathcal{R} as well as U and V can be constructed so that the assumptions **A2** – **A4** hold. In this paper, we do not formally prove the assumption **A1** – the proof of that assumption is rather technical and beyond the scope of the present paper. Instead of the proof, we present some heuristic reasoning and computer simulations to convince the reader that it holds in many cases. The computer simulations also allow us to estimate constants in the lower bound. Finally, in Section 4 we present an upper bound of $E|L_n - EL_n|^r$. The upper bound shows that the order of convergence cannot be improved, so that $E|L_n - EL_n|^r \asymp n^{\frac{r}{2}}$.

2. GENERAL LOWER BOUND IN TWO-STEP APPROACH

In this paper, we follow so-called *two-step* approach. This approach has actually been used in the most of the papers for obtaining the lower bound of variance, but formalized in [31, 25].

Generally speaking, the *first step* is to find a random mapping independent of Z , usually called by us as *random transformation*,

$$\mathcal{R} : \mathcal{Z}_n \rightarrow \mathcal{Z}_n$$

such that for an universal constant $\epsilon_o > 0$, the following convergence holds:

$$P(E[L(\mathcal{R}(Z)) - L(Z)|Z] \geq \epsilon_o) \rightarrow 1 \tag{2.1}$$

Here, abusing a bit of the notation, E denotes the expectation over the randomness involved in \mathcal{R} and P denotes the law of Z .

The *second step* is to show that equation (2.1) implies the optimal rate of convergence of absolute moments

$$\Phi(|L_n - EL_n|),$$

where $\Phi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is a convex non-decreasing function. To do so, we look for $U := \mathbf{u}(Z)$ and $V := \mathbf{v}(Z)$ two new random vectors (functions of Z), such that $\Phi(|L_n - EL_n|)$ can be somehow estimated from below by $\Phi(U)$. So, the control of fluctuations of L_n is reduced to the (easier) control of fluctuations of U and V . The current paper deals with the second step - the goal is to provide a general theorem and to apply it for pairwise Markov chains.

2.1. The main theorem. Consider two given functions

$$\mathbf{u} : \mathcal{Z}_n \rightarrow \mathbb{Z}, \quad \mathbf{v} : \mathcal{Z}_n \rightarrow \mathbb{Z}^d$$

and define $U := \mathbf{u}(Z)$ (resp. $V := \mathbf{v}(Z)$) an integer value random variable (resp. vector). Denote by \mathcal{S}_n , \mathcal{S}_n^U and \mathcal{S}_n^V the support of distributions of (U, V) , U and V , respectively. Hence $\mathcal{S}_n \subset \mathbb{Z}^{d+1}$, $\mathcal{S}_n^U \subset \mathbb{Z}$ and $\mathcal{S}_n^V \subset \mathbb{Z}^d$. For every $v \in \mathcal{S}_n^V$, we define the fiber of \mathcal{S}_n^U as follows

$$\mathcal{S}_n(v) := \{u \in \mathcal{S}_n^U : (u, v) \in \mathcal{S}_n\}.$$

For any $(u, v) \in \mathcal{S}_n$, let

$$l(u, v) := E[L(Z)|U = u, V = v].$$

For any $(u, v) \in \mathcal{S}_n$, let $P_{(u,v)}$ denote the law of $Z = (X, Y)$ given $U = u$ and $V = v$, namely

$$P_{(u,v)}(z) = P(Z = z|U = u, V = v).$$

Assumptions. The choice of the random transformation \mathcal{R} and U, V are linked together through the following assumptions :

A1: There exist universal constant $\epsilon_o > 0$ and a sequence $\Delta_n \rightarrow 0$ such that

$$P(E[L(\mathcal{R}(Z)) - L(Z)|Z] \geq \epsilon_o) \geq 1 - \Delta_n.$$

A2: There exists an universal constant (independent of n) $A < \infty$ such that $L(\mathcal{R}(Z)) - L(Z) \geq -A$.

A3: There exists sets $\mathcal{V}_n \subset \mathcal{S}_n^V$ and

$$\mathcal{U}_n(v) := \{u_n(v) + 1, u_n(v) + 2, \dots, u_n(v) + m_n(v)\} \subset \mathcal{S}_n(v)$$

such that for any (u, v) such that $v \in \mathcal{V}_n$ and $u \in \mathcal{U}_n(v)$, the following implication holds:

$$\text{If } Z \sim P_{(u,v)}, \text{ then } \mathcal{R}(Z) \sim P_{(u+1,v)}.$$

A4: There exists $n_1 > 0$ and a function $c(v) > 0$ (independent of n) such that for every $n \geq n_1$ and for every $v \in \mathcal{V}_n$,

$$m_n(v) \geq c(v)\varphi_v(n)^{-1},$$

where $\varphi_v(n) > 0$ satisfies

$$\min_{u \in \mathcal{U}_n(v)} P(U = u|V = v) \geq \varphi_v(n). \quad (2.2)$$

We note that **A4** is equivalent to the existence of $n_1 > 0$ and a function $c(v) > 0$ (independent of n) such that for every $n \geq n_1$ and for every $v \in \mathcal{V}_n$,

$$m_n(v) \cdot \min_{u \in \mathcal{U}_n(v)} P(U = u|V = v) \geq c(v)$$

and φ_v can be chosen as any function satisfying

$$\varphi_v(n) \in [c(v)/m_n(v), \min_{u \in \mathcal{U}_n(v)} P(U = u|V = v)]. \quad (2.3)$$

In what follows, we are interested in taking $\varphi_v(n)$ as small as possible, because the smaller $\varphi_v(n)$, the bigger the lower bound of $E\Phi(|L(Z) - \mu_n|)$ (see equation (2.4) in the theorem below). By equation (2.3), small $\varphi_v(n)$ means big $m_n(v)$, but too big a set $\mathcal{U}_n(v)$ typically means an exponential decay in probability and then the constant $c(v)$ might not exist. Therefore **A4** ties the lower bound of $E\Phi(|L(Z) - \mu_n|)$ with the size of $\mathcal{U}_n(v)$. Clearly $\varphi_v(n)$ can be chosen in such a way that $\varphi_v(n) \rightarrow 0$ as $n \rightarrow \infty$ if and only if $m_n(v) \rightarrow \infty$ as $n \rightarrow \infty$.

Theorem 2.1. *Let $\Phi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ be convex non-decreasing function and let μ_n be a sequence of reals. Assume **A1**, **A2**, **A3**, **A4**. Suppose that $c := c(v)$ is independent of $v \in \mathcal{V}_n$ and that $\varphi(n) := \sup_{v \in \mathcal{V}_n} (\varphi_v(n)) \rightarrow 0$ as $n \rightarrow \infty$. If, in addition, there exists $b_o > 0$ such that $P(V \in \mathcal{V}_n) \geq b_o$ for any n big enough, then given any constant $c_o \in (0, b_o c/8)$ for every sufficiently large n*

$$E\Phi(|L(Z) - \mu_n|) \geq \Phi\left(\frac{\epsilon_o c}{16\varphi(n)}\right)c_o. \quad (2.4)$$

3. PAIRWISE MARKOV CHAINS

In this paper, we consider a rather general model. Let X_1, X_2, \dots and Y_1, Y_2, \dots be two random processes on common state-space $\mathbb{A} = \{a_1, \dots, a_k\}$ (i.e. r.-variables X_i and Y_i take values on \mathbb{A}) such that the 2D process Z_1, Z_2, \dots , where $Z_i = (X_i, Y_i)$ is an aperiodic stationary MC with state space $\mathbb{A} \times \mathbb{A}$. Now the words $X = (X_1, \dots, X_n)$ and $Y = (Y_1, \dots, Y_n)$ are taken as the first n elements from these sequences. In what follows, we shall denote the elements of $\mathbb{A} \times \mathbb{A}$ by capital letters and we denote by $\mathbb{P} = (p_{\mathbf{A}\mathbf{B}})_{\mathbf{A}, \mathbf{B} \in \mathbb{A} \times \mathbb{A}}$ the transition matrix of Z . By aperiodicity assumption, there exists an integer m such that \mathbb{P}^m is primitive, i.e. has all strictly positive entries.

Random variable V . To construct V , we fix pairs $\mathbf{A}, \mathbf{B} \in \mathbb{A} \times \mathbb{A}$ such that $P(Z_1 = \mathbf{A}, Z_3 = \mathbf{B}) > 0$ and define $f := I_G$, where I_G stands for an indicator function and

$$G := \{\mathbf{A}\} \times \mathbb{A} \times \mathbb{A} \times \{\mathbf{B}\}.$$

Let us now define a Markov chain $\xi := \xi_1, \xi_2, \dots$ as follows

$$\xi_1 := (Z_1, Z_2, Z_3), \quad \xi_2 := (Z_4, Z_5, Z_6), \dots$$

Thus, the state space of ξ -chain is a subset $\mathcal{X} \subset \mathbb{A}^6$ (not necessarily \mathbb{A}^6 , because given the zeros in \mathbb{P} , it might be that some triplets have zero probability). Since Z is stationary, so is ξ ; moreover the aperiodicity of Z implies that of ξ . The random variable V is defined now as follows

$$V :=: \mathbf{v}(Z) := \sum_{i=1}^{\lfloor \frac{n}{3} \rfloor} f(\xi_i) = \sum_{i=1}^{\lfloor \frac{n}{3} \rfloor} I_G(\xi_i).$$

Therefore,

$$EV = \sum_{i=1}^{\lfloor \frac{n}{3} \rfloor} P(\xi \in G) = \lfloor \frac{n}{3} \rfloor P(Z_1 = \mathbf{A}, Z_3 = \mathbf{B}),$$

where the last equality follows from the stationarity. Let $\alpha := \frac{1}{3}P(Z_1 = \mathbf{A}, Z_3 = \mathbf{B})$. Then

$$EV = \lfloor \frac{n}{3} \rfloor 3\alpha.$$

When $n = 3m$, for some integer m , then $EV = \alpha n$, otherwise

$$EV := \alpha_n n, \quad \text{where} \quad \alpha - \frac{3\alpha}{n} < \alpha_n \leq \alpha.$$

Let us define

$$\mathcal{V}_n = [\alpha_n n - K\sqrt{n}, \alpha_n n + K\sqrt{n}] \cap \mathcal{S}_n^V,$$

where K is a constant specified later.

Proving $P(V \in \mathcal{V}_n) > b_o$ with b_o arbitrary close to 1. Let us now show that $P(V \in \mathcal{V}_n)$ is bounded away from zero for any sufficiently large n . Let $b_o \in (0, 1)$ be fixed. For that we use Hoeffding inequality for Markov chain proven in [19]. The theorem assumes that ξ_1, ξ_2, \dots satisfies the following condition: there exists probability measure Q on \mathcal{X} , $\lambda > 0$ and integer $r \geq 1$ such that for any state $x \in \mathcal{X}$

$$P_x(\xi_{r+1} \in \cdot) \geq \lambda Q(\cdot) \tag{3.1}$$

where $P_x(\cdot) := P(\cdot | \xi_1 = x)$. Recall that Z is aperiodic and so is ξ , hence there is a r such that for all states $x, y \in \mathcal{X}$, it holds $P_x(\xi_r = y) > 0$. That implies that equation (3.1) holds with Q being uniform over \mathcal{X} and

$$\lambda = \min_{x, y} P(\xi_{1+r} = y | \xi_1 = x) |\mathcal{X}|.$$

Then, according to the theorem, given a function $f : \mathcal{X} \rightarrow \mathbb{R}$, $S_m := \sum_{i=1}^m f(\xi_i)$, Hoeffding inequality is as follows: for any $x \in \mathcal{X}$

$$P_x(S_m - ES_m > m\epsilon) \leq \exp\left[-\frac{\lambda^2(m\epsilon - \frac{2r}{\lambda}\|f\|_\infty)^2}{2m\|f\|_\infty^2 r^2}\right], \quad \text{if } m > 2r(\lambda\epsilon)^{-1}\|f\|_\infty \tag{3.2}$$

where $\|f\|_\infty := \sup\{|f(x)|: x \in \mathcal{X}\}$. We take $m = \lfloor \frac{n}{3} \rfloor$ (remember that $V := S_{\lfloor \frac{n}{3} \rfloor}$) and $f = I_G$ so that $\|f\|_\infty = 1$. The inequality (3.2) is now: for every $\epsilon > 0$ and $x \in \mathcal{X}$

$$P_x(V - EV > \lfloor \frac{n}{3} \rfloor \epsilon) \leq \exp\left[-\frac{\lambda^2(\lfloor \frac{n}{3} \rfloor \epsilon - \frac{2r}{\lambda})^2}{2\frac{n}{3}r^2}\right], \quad \text{if } n > 6r(\lambda\epsilon)^{-1} + 3. \quad (3.3)$$

Take now K so big that

$$\exp\left[-\frac{3}{8}\left(\frac{\lambda}{r}\right)^2 K^2\right] < \frac{1 - b_o}{2}.$$

Take now $\epsilon = K\frac{3}{\sqrt{n}}$; then

$$K\sqrt{n} - \frac{3K}{\sqrt{n}} \leq \lfloor \frac{n}{3} \rfloor \epsilon \leq K\sqrt{n}.$$

If n is so big that

$$\frac{K\sqrt{n}}{2} > \frac{3K}{\sqrt{n}} + \frac{2r}{\lambda},$$

then $n > 6r(\lambda\epsilon)^{-1} + 3$ and inequality (3.3) implies

$$P_x(V - EV > K\sqrt{n}) \leq \exp\left[-\frac{3\lambda^2(K\sqrt{n} - \frac{3K}{\sqrt{n}} - \frac{2r}{\lambda})^2}{2r^2n}\right] \leq \exp\left[-\frac{3\lambda^2(\frac{1}{2}K\sqrt{n})^2}{2r^2n}\right] = \exp\left[-\frac{3}{8}\left(\frac{\lambda K}{r}\right)^2\right] \leq \frac{1 - b_o}{2}.$$

Since the left hand side holds for any initial probability distribution of ξ and, then also, for any initial probability distribution of Z . Thus, we have shown that there exists n_1 so that for every $n > n_1$

$$P(V - EV \leq K\sqrt{n}) = P(V \leq \alpha n + K\sqrt{n}) > \frac{1}{2} + \frac{b_o}{2}.$$

Applying the same argument for $f = -I_G$, we obtain that

$$P(V + EV \geq -K\sqrt{n}) = P(V \geq \alpha n - K\sqrt{n}) > \frac{1}{2} + \frac{b_o}{2}.$$

These two inequalities together give

$$P(\alpha_n n - K\sqrt{n} \leq V \leq \alpha_n n + K\sqrt{n}) = P(V \in \mathcal{V}_n) > b_o, \quad \forall n > n_1.$$

Random variable U . Let us now define U . To this aim, fix a letter in $\mathbb{A} \times \mathbb{A}$ and let us call it \mathbf{D} . The random variable U is the number of states \mathbf{D} in the middle of the $(\mathbf{A} \cdot \mathbf{B})$ -triplets. For the formal definition let, for any $z \in (\mathbb{A} \times \mathbb{A})^n$,

$$\mathbf{n}_i(z) := f(z_{3(i-1)+1}, z_{3(i-1)+2}, z_{3(i-1)+3}), \quad i = 1, \dots, \lfloor \frac{n}{3} \rfloor$$

and let us denote $(\mathbf{n}_1(z), \dots, \mathbf{n}_{\lfloor \frac{n}{3} \rfloor}(z))$ by $\mathbf{n}(z)$. Hence

$$\mathbf{v}(z) = \sum_{i=1}^{\lfloor \frac{n}{3} \rfloor} \mathbf{n}_i(z).$$

The function $\mathbf{u}(z)$ and random variable U are defined as follows

$$\mathbf{u}(z) = \sum_{i=1}^{\lfloor \frac{n}{3} \rfloor} \mathbf{n}_i(z) I_{\mathbf{D}}(z_{3(i-1)+2}), \quad U = \mathbf{u}(Z).$$

Clearly $\mathcal{S}_n(v) = \{0, 1, \dots, v\}$. Moreover, by the Markov property given $V = v$, $U \sim B(v, q)$, i.e. the random variable U has binomial distribution with parameters v and

$$q := \frac{p_{\mathbf{A}\mathbf{D}}p_{\mathbf{D}\mathbf{B}}}{\sum_{\mathbf{D}' \in \mathbb{A} \times \mathbb{A}} p_{\mathbf{A}\mathbf{D}'}p_{\mathbf{D}'\mathbf{B}}}.$$

The letter \mathbf{D} is chosen in such a way that $q > 0$. Take now, for any $v \in \mathcal{S}_n^V$

$$\mathcal{U}_n(v) := [vq - \sqrt{v}, vq + \sqrt{v}] \cap \mathcal{S}_n(v).$$

When v is big enough, then

$$\mathcal{U}_n(v) = [vq - \sqrt{v}, vq + \sqrt{v}] \cap \mathbb{Z}.$$

In this case the interval contains at most $\lfloor 2\sqrt{v} + 1 \rfloor$ integers.

Proving A4 for $c(v)$ and $\varphi_v(n)$ independent of v .

Lemma 3.1. *Let $X \sim B(m, p)$ be a binomial random variable with parameters m and p . Then, for any constant $\beta > 0$, there exists $b(\beta, p)$ and $m_o(\beta, p)$ such that for every $b \geq b(\beta, p)$, $m > m_o$ and*

$$i \in [mp - \beta\sqrt{m}, mp + \beta\sqrt{m}],$$

we have

$$P(X = i) = \binom{m}{i} p^i (1-p)^{m-i} \geq \frac{1}{b\sqrt{m}}. \quad (3.4)$$

It can be shown (see [25, equation (4.11)]) that the constant $b(\beta, p)$ can be taken as

$$b(\beta, p) := \sqrt{2\pi p(1-p)} \exp \left[\frac{\beta^2}{2p(1-p)} \right].$$

From this lemma, it follows that there exists universal constant $b(q) > 0$ and integer v_o so big that for every $u \in \mathcal{U}_n(v)$,

$$P(U = u | V = v) \geq \frac{1}{b\sqrt{v}}, \quad v > v_o \quad (3.5)$$

given any constant b satisfying

$$b > b(q) := \sqrt{2\pi q(1-q)} \exp \left[\frac{1}{2q(1-q)} \right]. \quad (3.6)$$

Recall the definition of \mathcal{V}_n . There exists $n_2 > n_1$ large enough such that if $n > n_2$, then $\alpha_n n - K\sqrt{n} > v_o$ and $\alpha_n n + K\sqrt{n} \leq \frac{n}{3} < n$. Therefore, if $n > n_2$ then

$$\mathcal{V}_n = [\alpha_n n - K\sqrt{n}, \alpha_n n + K\sqrt{n}] \cap \mathbb{Z},$$

every $v \in \mathcal{V}_n$ is smaller than n and equation (3.5) holds. Hence, when $n > n_2$, we have

$$P(U = u | V = v) \geq \frac{1}{b\sqrt{v}} \geq \frac{1}{b\sqrt{n}}, \quad \forall v \in \mathcal{V}_n \quad \forall n > n_2.$$

Thus equation (2.2) holds with

$$\varphi_v(n) :=: \varphi(n) := \frac{1}{b\sqrt{n}}.$$

We can find $n_3 > n_2$ large enough such that

$$\sqrt{\alpha_n n - K\sqrt{n}} \geq \sqrt{\frac{\alpha}{2}n} + \frac{1}{2}.$$

Therefore, if $n \geq n_3$, then every $v \in \mathcal{V}_n$ satisfies $\sqrt{v} \geq \sqrt{\alpha n/2} + 1/2$. Since the minimum number of integers in the interval $\mathcal{U}_n(v)$ is $\lfloor 2\sqrt{v} \rfloor$, we obtain the following inequality

$$m_n(v) > 2\sqrt{v} - 1 \geq 2\sqrt{\frac{\alpha}{2}}\sqrt{n} = b^{-1}\sqrt{2\alpha}\varphi(n)^{-1}, \quad \forall v \in \mathcal{V}_n.$$

Thus A4 with $c = b^{-1}\sqrt{2\alpha}$ holds. Therefore,

$$\frac{\epsilon_o c}{16\varphi(n)} = \frac{\epsilon_o \sqrt{2\alpha n}}{16}.$$

Thus the right hand side of equation (2.4) is

$$c_o \Phi \left(\frac{\epsilon_o \sqrt{2\alpha n}}{16} \right),$$

where for n big enough

$$c = \frac{\sqrt{2\alpha}}{b}, \quad c_o \leq \frac{c}{8}(b_o - \sqrt{\Delta_n})$$

Since b_o could be taken arbitrary close to one and $\Delta_n \rightarrow 0$, we can take any $c_o < \frac{c}{8}$.

The random transformation \mathcal{R} and the assumption A3. The random transformation \mathcal{R} picks a random $(\mathbf{A} \cdot \mathbf{B})$ -triplet which does not have a letter \mathbf{D} in-between (with uniform distribution) and changes the letter in the middle of the triplet into a \mathbf{D} -letter. Let $\{i_1(z), \dots, i_{\mathbf{v}(z)}(z)\}$ be the set of indexes corresponding to 1s in $\mathbf{n}(z)$ and define $\mathbf{b}(z) := (\mathbf{b}_1(z), \dots, \mathbf{b}_{\mathbf{v}(z)}(z))$ where

$$\mathbf{b}_j(z) := I_{\mathbf{D}}(z_{3(i_j(z)-1)+2}), \quad j = 1, \dots, \mathbf{v}(z).$$

With this notation, the number of \mathbf{D} -letters in-between the triplets is

$$\mathbf{u}(z) = \sum_{j=1}^{\mathbf{v}(z)} \mathbf{b}_j(z).$$

The random transformation \mathcal{R} acts on the set of sequences z satisfying the following condition: $\mathbf{u}(z) < \mathbf{v}(z)$. Given such a sequence, \mathcal{R} picks a random zero out of $\mathbf{v}(z) - \mathbf{u}(z)$ zeros in the vector $\mathbf{b}(z)$ (uniform distribution). Suppose that the chosen zero is the k -th element of $\mathbf{b}(z)$. Then $z_{3(i_k(z)-1)+2} \neq \mathbf{D}$ and \mathcal{R} changes that letter into \mathbf{D} . Thus $\mathcal{R}(z)$ is a sequence such that $\mathbf{n}_i(\mathcal{R}(z)) = \mathbf{n}_i(z)$ for every $i = 1, \dots, \lfloor \frac{n}{3} \rfloor$, thus $\mathbf{v}(\mathcal{R}(z)) = \mathbf{v}(z)$; but $\mathbf{u}(\mathcal{R}(z)) = \mathbf{u}(z) + 1$.

The following is an auxiliary and almost trivial result which we prove for the sake of completeness.

Proposition 3.1. *Let $Z := (Z_1, \dots, Z_m)$ be a vector of iid Bernoulli random variables of parameter p and let $P_{(u)}$ be the law of Z given $U := \sum_{i=1}^m Z_i = u$. Let $W \sim P_{(u)}$, where $u < m$. Then choose a random 0 in W with uniform distribution and change it into one. Let \widetilde{W} be the resulting random variable. Then $\widetilde{W} \sim P_{(u+1)}$.*

Proof. For any $u = 0, \dots, m$, let $\mathcal{A}(u) \subseteq \{0, 1\}^m$ consist of all binary sequences containing exactly u ones. For any $z \in \mathcal{A}(u)$,

$$P(Z = z | U = u) = \frac{P(Z = z, U = u)}{P(U = u)} = \frac{P(Z = z)}{P(U = u)} = \frac{p^u(1-p)^{m-u}}{\binom{m}{u} p^u(1-p)^{m-u}} = \binom{m}{u}^{-1}.$$

In other words, $P_{(u)}$ is the uniform distribution on $\mathcal{A}(u)$. Now for any $u = 0, \dots, m-1$, let W be any random vector such that $W \sim P_{(u)}$, then \widetilde{W} is supported on $\mathcal{A}(u+1)$. For any $z \in \mathcal{A}(u+1)$, let $0 \leq i_1 < \dots < i_{u+1} \leq m$ be the positions of ones in z , and let \hat{z}_{i_j} , $j = 1, \dots, u+1$, be the sequence in $\mathcal{A}(u)$ obtained from z by replacing the symbol 1 at position i_j with 0. We have

$$P(\widetilde{W} = z) = \sum_{j=1}^{u+1} P(\widetilde{W} = z | W = \hat{z}_{i_j}) P(W = \hat{z}_{i_j}) = (u+1) \frac{1}{m-u} \binom{m}{u}^{-1} = \binom{m}{u+1}^{-1} = P_{(u+1)}(z).$$

□

Let us consider the sequence $Z = Z_1, \dots, Z_n$ and let

$$m := \lfloor \frac{n}{3} \rfloor.$$

Recall that

$$V = \sum_{j=1}^m f(\xi_j) = \sum_{i=1}^m \eta_i,$$

where

$$\eta_i := f(\xi_i) = \mathbf{n}_i(Z) \in \{0, 1\}.$$

Since Z is stationary, we have that the sequence $\eta := (\eta_1, \dots, \eta_m)$ is a stationary binary sequence. As in the proof of Proposition 3.1, let $\mathcal{A}(v)$ be the set of binary sequences of length m having v ones. It is easy to see that additional conditioning on U will not change the conditional probability of η (given $u \leq v$), because for any vector $a \in \mathcal{A}(v)$ we have $\{\eta = a\} \subseteq \{V = v\}$ and

$$\begin{aligned} P(\eta = a | V = v, U = u) &= \frac{P(U = u, V = v, \eta = a)}{P(U = u, V = v)} = \frac{P(U = u | \eta = a) P(\eta = a)}{P(U = u, V = v)} \\ &= \frac{\binom{v}{u} q^u (1-q)^{v-u} P(\eta = a)}{P(U = u, V = v)}. \end{aligned}$$

Since

$$P(U = u, V = v) = \sum_{a \in \mathcal{A}(v)} P(U = u | \eta = a) P(\eta = a) = \binom{v}{u} q^u (1-q)^{v-u} P(V = v),$$

we have

$$P(\eta = a | V = v, U = u) = \frac{P(\eta = a, V = v)}{P(V = v)} = P(\eta = a | V = v). \quad (3.7)$$

For any $u \leq v \leq m$, let $\mathcal{B}(u, v)$ be the set of sequences such that the value of \mathbf{u} and \mathbf{v} are u and v respectively, that is

$$\mathcal{B}(u, v) = \{z \in (\mathbb{A} \times \mathbb{A})^n, \quad \mathbf{u}(z) = u, \quad \mathbf{v}(z) = v\}.$$

Fix $u \leq v \leq m$ and $Z_{(u,v)} \sim P_{(u,v)}$ (i.e. $P(Z_{(u,v)} = z) = P(Z = z | U = u, V = v)$). Let us compute $P_{(u,v)}$. To this aim define $B := (B_1, \dots, B_V) \equiv \mathbf{b}(Z)$. Now for any $z \in \mathcal{B}(u, v)$, by definition of $P_{(u,v)}$, since $\{Z = z\} \subseteq \{\eta = \mathbf{n}(z), B = \mathbf{b}(z)\} \subseteq \{U = u, V = v\}$, we have

$$\begin{aligned} P(Z_{(u,v)} = z) &= P(Z = z | U = u, V = v) = P(Z = z, \eta = \mathbf{n}(z), B = \mathbf{b}(z) | U = u, V = v) \\ &= P(Z = z | \eta = \mathbf{n}(z), B = \mathbf{b}(z)) P(\eta = \mathbf{n}(z), B = \mathbf{b}(z) | U = u, V = v). \end{aligned}$$

Given η , let Z' be the random vector obtained by collecting all random variables from (Z_1, \dots, Z_n) corresponding to the triplets where $\eta_i = 0$. And, analogously, let z' be the vector obtained by z by collecting the triplets where $\mathbf{n}_i(z) = 0$. From the Markov property we have

$$P(Z' = z' | \eta = \mathbf{n}(z), B = \mathbf{b}(z)) = P(Z' = z' | \eta = \mathbf{n}(z)).$$

Let $1 \leq i_1 < \dots < i_v \leq m$ be the indexes of corresponding ones in $\mathbf{n}(z)$. Then, from $\mathbf{b}(z)$ we know for every $j = 1, \dots, v$, whether $z_{3(i_j-1)+2}$ equals \mathbf{D} or not. But this does not fully determine the values of $z_{3(i_j-1)+2}$. Hence

$$\begin{aligned} P(Z = z | \eta = \mathbf{n}(z), B = \mathbf{b}(z)) &= \\ P(Z' = z' | \eta = \mathbf{n}(z)) \prod_{j=1}^v P(Z_{3(i_j-1)+2} = z_{3(i_j-1)+2} | Z_{3(i_j-1)+1} = \mathbf{A}, Z_{3(i_j-1)+3} = \mathbf{B}, B_j = \mathbf{b}_j(z)). \end{aligned}$$

If, in the product above, $\mathbf{b}_j(z) = 1$, then $z_{3(i_j-1)+2} = \mathbf{D}$ and

$$P(Z_{3(i_j-1)+2} = z_{3(i_j-1)+2} | Z_{3(i_j-1)+1} = \mathbf{A}, Z_{3(i_j-1)+3} = \mathbf{B}, B_j = \mathbf{b}_j(z)) = 1,$$

otherwise

$$\begin{aligned} P(Z_{3(i_j-1)+2} = z_{3(i_j-1)+2} | Z_{3(i_j-1)+1} = \mathbf{A}, Z_{3(i_j-1)+3} = \mathbf{B}, B_j = 0) &= \\ P(Z_{3(i_j-1)+2} = z_{3(i_j-1)+2} | Z_{3(i_j-1)+1} = \mathbf{A}, Z_{3(i_j-1)+3} = \mathbf{B}, Z_{3(i_j-1)+2} \neq \mathbf{D}) &=: \rho(3(i_j - 1) + 2, z); \end{aligned}$$

note that

$$\sum_{F \in \mathbb{A} \times \mathbb{A}: F \neq \mathbf{D}} P(Z_{3(i_j-1)+2} = F | Z_{3(i_j-1)+1} = \mathbf{A}, Z_{3(i_j-1)+3} = \mathbf{B}, Z_{3(i_j-1)+2} \neq \mathbf{D}) = 1. \quad (3.8)$$

Thus

$$\begin{aligned}
& \prod_{j=1}^v P(Z_{3(i_j-1)+2} = z_{3(i_j-1)+2} | Z_{3(i_j-1)+1} = \mathbf{A}, Z_{3(i_j-1)+3} = \mathbf{B}, B_j = \mathbf{b}_j(z)) \\
&= \prod_{\substack{j=1, \dots, v \\ \mathbf{b}_j(z)=0}} P(Z_{3(i_j-1)+2} = z_{3(i_j-1)+2} | Z_{3(i_j-1)+1} = \mathbf{A}, Z_{3(i_j-1)+3} = \mathbf{B}, Z_{3(i_j-1)+2} \neq \mathbf{D}) \\
&= \prod_{\substack{j=1, \dots, v \\ \mathbf{b}_j(z)=0}} \rho(3(i_j-1) + 2, z) =: \rho_z.
\end{aligned}$$

From equation (3.7), we know

$$P(\eta = \mathbf{n}(z) | U = u, V = v) = P(\eta = \mathbf{n}(z) | V = v).$$

By the Markov property

$$P(B = \mathbf{b}(z) | \eta = \mathbf{n}(z), U = u, V = v) = P(B = \mathbf{b}(z) | \eta = \mathbf{n}(z), U = u)$$

and this probability is equal to the probability that v i.i.d Bernoulli random variables take values $\mathbf{b}(z)$ given their sum is equal to u . This probability is $\binom{v}{u}^{-1}$. Thus,

$$P(B = \mathbf{b}(z) | \eta = \mathbf{n}(z), U = u, V = v) = \binom{v}{u}^{-1}.$$

Therefore, for any $z \in \mathcal{B}(u, v)$, we have

$$P(Z_{(u,v)} = z) = P(Z' = z' | \eta = \mathbf{n}(z)) \rho_z P(\eta = \mathbf{n}(z) | V = v) \binom{v}{u}^{-1}. \quad (3.9)$$

We apply now the random transformation and we compute $P(\mathcal{R}(Z_{(u,v)}) = z)$. Clearly, given $z \in \mathcal{B}(u+1, v)$,

$$P(\mathcal{R}(Z_{(u,v)}) = z) = \sum_{\tilde{z} \in \mathcal{B}(u,v)} P(\mathcal{R}(Z_{(u,v)}) = z | Z_{(u,v)} = \tilde{z}) P(Z_{(u,v)} = \tilde{z}) = (*)$$

and

$$P(\mathcal{R}(Z_{(u,v)}) = z | Z_{(u,v)} = \tilde{z}) = \begin{cases} 0 & \text{if } \tilde{z} \notin H_{\mathcal{R}}(z) \\ 1/(v-u) & \text{if } \tilde{z} \in H_{\mathcal{R}}(z) \end{cases}$$

where

$$H_{\mathcal{R}}(z) := \{\tilde{z} : P(\mathcal{R}(\tilde{z}) = z) > 0\} = \bigcup_{\substack{j=1, \dots, v \\ \mathbf{b}_j(z)=1}} \{\tilde{z} : P(\mathcal{R}(\tilde{z}) = z) > 0, \tilde{z}_{3(i_j-1)+2} \neq \mathbf{D}\}$$

the latter being the union of $u+1$ pairwise disjoint sets. Define $\tilde{\eta} := \mathbf{n}(\mathcal{R}(Z_{(u,v)}))$ and observe that if $\tilde{z} \in H_{\mathcal{R}}(z)$ then $P(\mathcal{R}(Z_{(u,v)})' = z' | \tilde{\eta} = \mathbf{n}(z)) = P(Z' = z' | \eta = \mathbf{n}(\tilde{z}))$ and $P(\tilde{\eta} = \mathbf{n}(z) | V = v) = P(\eta = \mathbf{n}(\tilde{z}) | V = v)$. Moreover $\sum_{\tilde{z} \in H_{\mathcal{R}}(z)} \rho_{\tilde{z}} = (u+1)\rho_z$ (decompose the sum using the above partition of $H_{\mathcal{R}}(z)$ into $u+1$ subsets and use equation (3.8)). Thus, by computing $P(Z_{(u,v)} = \tilde{z})$ by means of equation (3.9), we obtain

$$\begin{aligned}
(*) &= \sum_{\tilde{z} \in H_{\mathcal{R}}(z)} P(Z' = z' | \eta = \mathbf{n}(\tilde{z})) \rho_{\tilde{z}} P(\eta = \mathbf{n}(\tilde{z}) | V = v) \binom{v}{u}^{-1} \frac{1}{v-u} \\
&= \sum_{\tilde{z} \in H_{\mathcal{R}}(z)} P(\mathcal{R}(Z_{(u,v)})' = z' | \tilde{\eta} = \mathbf{n}(z)) \rho_{\tilde{z}} P(\tilde{\eta} = \mathbf{n}(z) | V = v) \binom{v}{u}^{-1} \frac{1}{v-u} \\
&= P(\mathcal{R}(Z_{(u,v)})' = z' | \tilde{\eta} = \mathbf{n}(z)) \rho_z P(\tilde{\eta} = \mathbf{n}(z) | V = v) \binom{v}{u+1}^{-1}
\end{aligned}$$

which, according to equation (3.9), implies that $\mathcal{R}(Z_{(u,v)}) \sim P_{(u+1,v)}$ and the proof is complete.

Main result. We have defined the random transformation \mathcal{R} , random variables U, V and sets $\mathcal{U}_n(v)$ and \mathcal{V}_n such that assumptions **A3** and **A4** with $\varphi(n)$ hold. Since \mathcal{R} changes at most two letters at time, by equation (1.2), the assumption **A2** holds for $A = 2\Delta$. Thus, recalling that b_o can be chosen arbitrarily close to 1, from Theorem 2.1 we have the following result.

Theorem 3.1. *Let $\Phi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ be convex non-decreasing function and let μ_n be a sequence of reals. If there exists $\epsilon_o > 0$ such that the random transformation \mathcal{R} satisfies **A1**, then for every n sufficiently large, the following inequality holds*

$$E\Phi(|L(Z) - \mu_n|) \geq c_o \Phi\left(\frac{\epsilon_o \sqrt{2\alpha n}}{16}\right), \quad (3.10)$$

where $\alpha = \frac{1}{3}P(Z_1 = \mathbf{A}, Z_3 = \mathbf{B})$ and $0 < c_o < b(q)^{-1}\sqrt{2\alpha}/8$ ($b(q)$ defined as in equation (3.6)).

In particular, when $\Phi(x) = x^r$, for $r \geq 1$ and $\mu_n = EL(Z)$, then equation (3.10) is

$$E |L(Z) - EL(Z)|^r \geq c_o \left(\frac{\epsilon_o \sqrt{2\alpha}}{16}\right)^r n^{\frac{r}{2}},$$

where

$$c_o < \frac{\sqrt{2\alpha}}{8b(q)}.$$

Taking $r = 2$, we obtain the lower bound for variance

$$\text{Var}(L(Z)) \geq a_o n, \quad a_o := \frac{2c_o \alpha}{16^2} \epsilon_o^2.$$

3.1. Combining random transformations. Suppose $\mathbf{A}_i, \mathbf{B}_i, \mathbf{D}_i$, $i = 1, 2$ are pairs of letters and let us briefly consider a random transformation \mathcal{R} that picks either a random $(\mathbf{A}_1 \cdot \mathbf{B}_1)$ -triplet which does not have a letter \mathbf{D}_1 in-between or a random $(\mathbf{A}_2 \cdot \mathbf{B}_2)$ -triplet which does not have a letter \mathbf{D}_2 in-between (with uniform distribution over both kind of triplets) and changes the letter in the middle of the triplet either into \mathbf{D}_1 -letter (if the chosen triplet was $(\mathbf{A}_1 \cdot \mathbf{B}_1)$) or into \mathbf{D}_2 -letter (if the chosen triplet was $(\mathbf{A}_2 \cdot \mathbf{B}_2)$). Such a transformation \mathcal{R} can be considered as a combination of two random transformations: \mathcal{R}_1 that acts on $(\mathbf{A}_1 \cdot \mathbf{B}_1)$ -triplets and \mathcal{R}_2 that acts on $(\mathbf{A}_2 \cdot \mathbf{B}_2)$. We suppose that $(\mathbf{A}_1, \mathbf{B}_1) \neq (\mathbf{A}_2, \mathbf{B}_2)$. Thus, for $i = 1, 2$, we now have the random variables V_i that count $(\mathbf{A}_i \cdot \mathbf{B}_i)$ -triplets (and are dependent on each other) and random variables U_i that counts number of states \mathbf{D}_i in-between the triplets. Let q_i be the probability of finding a \mathbf{D}_i -letter inside $(\mathbf{A}_i \cdot \mathbf{B}_i)$ -triplet. Thus given $V_i = v_i$, $U_i \sim B(v_i, q_i)$, $i = 1, 2$. Given V_1 and V_2 , the random variables U_1 and U_2 are independent.

We are now going to define the combined random transformation \mathcal{R} . In what follows, let $V = (V_1, V_2)$ and $U = U_1 + U_2$. Given $V = v := (v_1, v_2)$, the random variable U takes values from $0, 1, \dots, v_1 + v_2$. Now define the probabilities

$$p(l|u, v) := P(U_1 = l | U = u, V = v), \quad l = l_1, l_1 + 1, \dots, l_2,$$

where $l_1 = l_1(u, v_2) := (u - v_2) \vee 0$ and $l_2 = l_2(u, v_1) := u \wedge v_1$. Thus $p(l|u, v)$ is the probability that there are l \mathbf{D}_1 -letters (inside the triplets) given the sum of \mathbf{D}_1 and \mathbf{D}_2 letters (inside the corresponding triplets) is u . The random transformation \mathcal{R} picks the side i with certain probability r_i and then applies the transformation \mathcal{R}_i . In order for the composed random transformation \mathcal{R} to satisfy **A3**, the probabilities r_i should be chosen carefully. To this aim, given z , define $\mathbf{u}_i(z)$ and $\mathbf{v}_i(z)$, $i = 1, 2$ as usual and let $\mathbf{w}(z) := (\mathbf{u}_1(z), \mathbf{u}_2(z), \mathbf{v}_1(z), \mathbf{v}_2(z))$. We now define the probabilities $r_i(z) = r_i(\mathbf{w}(z)) = r_i(u_1, u_2, v)$ such that $r_1(z) + r_2(z) = 1$, $r_1(v_1, u_2, v) = r_2(u_1, v_2, v) = 0$ and the following conditions hold:

$$r_1(l-1, u-l+1, v)p(l-1|u, v) + r_2(l, u-l, v)p(l|u, v) = p(l|u+1, v), \quad l_2 \geq l > l_1 \quad (3.11)$$

$$r_2(0, u, v)p(0|u, v) = p(0|u+1, v), \text{ when } u < v_2 \quad (3.12)$$

$$r_1(u, 0, v)p(u|u, v) = p(u+1|u+1, v), \text{ when } u < v_1 \quad (3.13)$$

for all $u = 0, \dots, v_1 + v_2 - 1$. Now for any $w := (u_1, v_1, u_2, v_2)$, such that $v_i \geq u_i \geq 0$ and $v_1 + v_2 \leq m$, we define a random variable T_w such that $P(T_w = i) = r_i(w)$, $i = 1, 2$ and given the random variables

$U_i = u_i, V_i = v_i, T_w$ is independent of Z . The transformation \mathcal{R} is now formally defined as follows:

$$\mathcal{R} = \begin{cases} \mathcal{R}_1(z), & \text{if } T_w(z) = 1; \\ \mathcal{R}_2(z), & \text{if } T_w(z) = 2. \end{cases}$$

In general, the probabilities r_i depend on the probabilities q_i . When $q_1 = q_2$, then

$$r_i(u_1, u_2, v) := \frac{v_i - u_i}{(v_1 - u_1) + (v_2 - u_2)}, \quad i = 1, 2$$

satisfy the requirements. Thus, in that case \mathcal{R} just picks one $(\mathbf{A}_i \cdot \mathbf{B}_i)$ -triplet over all such triplets with no \mathbf{D}_i -letter inside with *uniform distribution*, whilst in the case $q_1 \neq q_2$, the distribution is not uniform. It is easy to see that such r_i satisfy conditions (3.11) (3.13) and (3.12). Indeed, the reader can easily prove the following statement.

Proposition 3.2. *Let $X \sim B(v_1, q)$ and $Y \sim B(v_2, q)$ two independent binomially distributed random variables. Then for any integers l and u such that $u < v_1 + v_2$ and $u \wedge v_1 \geq l > (u - v_2) \vee 0$ we have*

$$\frac{v_1 - l + 1}{v_1 + v_2 - u} P(X = l - 1 | X + Y = u) + \frac{v_2 - u + l}{v_1 + v_2 - u} P(X = l | X + Y = u) = P(X = l | X + Y = u + 1).$$

Moreover, when $u < v_2$, then

$$\frac{v_2 - u}{v_1 + v_2 - u} P(X = 0 | X + Y = u) = P(X = 0 | X + Y = u + 1).$$

Clearly \mathcal{R} satisfies **A2**. We now show that it also satisfies **A3**. Fix $v = (v_1, v_2)$ such that $v_1 + v_2 \leq m$. Now, we can decompose the measure $P_{(u,v)}$ as follows

$$P_{(u,v)} = \sum_{l=l_1}^{l_2} P_{(l,u-l,v)} p(l|u,v), \quad (3.14)$$

where $P_{(l,u-l,v)}$ is the distribution of Z given $U_1 = l, U = u, V = v$. We know that \mathcal{R}_i satisfies **A3** for any $u = \{0, 1, \dots, v_i - 1\}$, thus the following holds: when $Z \sim P_{(l,u-l,v)}$ and $l < v_1, u - l < v_2$, then $\mathcal{R}_1(Z) \sim P_{(l+1,u-l,v)}$ and $\mathcal{R}_2(Z) \sim P_{(l,u-l+1,v)}$. Therefore, if $Z \sim P_{(u,v)}$, then

$$\mathcal{R}(Z) \sim \sum_{l=l_1}^{l_2} (P_{(l+1,u-l,v)} r_1(l, u-l, v) + P_{(l,u-l+1,v)} r_2(l, u-l, v)) p(l|u, v).$$

Thus, by equation (3.11)

$$\begin{aligned} \mathcal{R}(Z) &\sim P_{(l_1, u-l_1+1, v)} r_2(l_1, u-l_1, v) p(l_1|u, v) \\ &+ \sum_{l=l_1+1}^{l_2} P_{(l, u-l+1, v)} (r_1(l-1, u-l+1, v) p(l-1|u, v) + r_2(l, u-l, v) p(l|u, v)) \\ &+ P_{(l_2+1, u-l_2, v)} r_1(l_2, u-l_2, v) p(l_2|u, v) = P_{(l_1, u-l_1+1, v)} r_2(l_1, u-l_1, v) p(l_1|u, v) \\ &+ \sum_{l=l_1+1}^{l_2} P_{(l, u+1-l, v)} p(l|u+1, v) + P_{(l_2+1, u-l_2, v)} r_1(l_2, u-l_2, v) p(l_2|u, v) = (*). \end{aligned}$$

If $u < v_1$ and $u < v_2$, then $l_1(u, v_2) = l_1(u+1, v_2) = 0$ and $l_2(u+1, v_1) = u+1 = l_2(u, v_1) + 1$, thus by equations (3.12) and (3.13) we obtain that (*) equals

$$\begin{aligned} P_{(0, u+1, v)} p(0|u+1, v) + \sum_{l=1}^u P_{(l, u+1-l, v)} p(l|u+1, v) + P_{(u+1, 0, v)} p(u+1|u+1, v) = \\ \sum_{l=l_1(u+1, v_2)}^{l_2(u+1, v_1)} P_{(l, u+1-l, v)} p(l|u+1, v) = P_{(u+1, v)}. \end{aligned}$$

If $u \geq v_1$ and $u < v_2$, then $l_1(u, v_2) = l_1(u + 1, v_2) = 0$, $l_2(u, v_1) = l_2(u + 1, v_1) = v_1$ and then by equation (3.12) and since $r_1(v_1, u - v_1, v) = 0$, we have that (*) equals

$$P_{(0, u+1, v)} p(0|u+1, v) + \sum_{l=1}^{v_1} P_{(l, u+1-l, v)} p(l|u+1, v) = \sum_{l=l_1(u+1, v_2)}^{l_2(u+1, v_1)} P_{(l, u+1-l, v)} p(l|u+1, v) = P_{(u+1, v)}.$$

If $u < v_1$ and $u \geq v_2$, then $l_1(u + 1, v_2) = u + 1 - v_2 = l_1(u, v_2) + 1$ and $l_2(u + 1, v_1) = u + 1 = l_2(u, v_1) + 1$, thus by equation (3.13) and since $r_2(u - v_2, v_2, v) = 0$ we obtain that (*) equals

$$\sum_{l=u-v_2+1}^u P_{(l, u+1-l, v)} p(l|u+1, v) + P_{(u+1, 0, v)} p(u+1|u+1, v) = \sum_{l=l_1(u+1, v_2)}^{l_2(u+1, v_1)} P_{(l, u+1-l, v)} p(l|u+1, v) = P_{(u+1, v)}.$$

Finally, if $u \geq v_1$ and $u \geq v_2$, then $l_1(u + 1, v_2) = u + 1 - v_2 = l_1(u, v_2) + 1$, $l_2(u, v_1) = l_2(u + 1, v_1) = v_1$. Since $r_2(u - v_2, v_2, v) = r_1(v_1, u - v_1, v) = 0$ and (*) equals

$$\sum_{l=u-v_2+1}^{v_1} P_{(l, u+1-l, v)} p(l|u+1, v) = \sum_{l=l_1(u+1, v_2)}^{l_2(u+1, v_1)} P_{(l, u+1-l, v)} p(l|u+1, v) = P_{(u+1, v)}.$$

Thus, we have shown that $\mathcal{R}(Z) \sim P_{(u+1, v)}$ and **A3** is fulfilled for any $u \in \{0, 1, \dots, v_1 + v_2 - 1\}$.

To the end of the paragraph, let us skip n from the notation and let $\mathcal{V} := \mathcal{V}_1 \times \mathcal{V}_2$. For $(v_1, v_2) \in \mathcal{V}$, let $\mathcal{U}_i(v_i) := [v_i q_i - \sqrt{v_i}, v_i q_i + \sqrt{v_i}] \cap \mathbb{Z}$ and

$$\mathcal{U}(v) := [(v_1 q_1 + v_2 q_2) - \sqrt{v_1} \wedge \sqrt{v_2}, (v_1 q_1 + v_2 q_2) + \sqrt{v_1} \wedge \sqrt{v_2}] \cap \mathbb{Z}.$$

It is not difficult to show that for every $u \in \mathcal{U}(v)$ the cardinality of $\{(u_1, u_2) \in \mathcal{U}_1(v_1) \times \mathcal{U}_2(v_2) : u_1 + u_2 = u\}$ is at least $\lfloor \sqrt{v_1} \wedge \sqrt{v_2} \rfloor$. In order to show that \mathcal{R} satisfies **A4**, we assume without loss of generality that $v_1 \leq v_2$, whence $\sqrt{v_1} \wedge \sqrt{v_2} = \sqrt{v_1}$. We know that \mathcal{R}_i satisfies **A4**, so for $i = 1, 2$ there exists a constant b_i such that for every $u_i \in \mathcal{U}_i(v_i)$ and n big enough, we have

$$P(U_i = u_i | V_i = v_i) \geq \frac{1}{b_i \sqrt{v_i}}$$

where b_i depends only on q_i (see Lemma 3.1). Now observe that $\lfloor x \rfloor / x \geq 1/2$ for all $x \geq 1$; thus, for every $u \in \mathcal{U}(v)$ and every sufficiently large n ,

$$\begin{aligned} P(U = u | V = v) &\geq \sum_{u_i \in \mathcal{U}_i(v_i) : u_1 + u_2 = u} P(U_1 = u_1 | V_1 = v_1) P(U_2 = u_2 | V_2 = v_2) \\ &\geq \sum_{u_i \in \mathcal{U}_i(v_i) : u_1 + u_2 = u} \frac{1}{b_1 b_2 \sqrt{v_1 v_2}} \geq \frac{1}{2 b_1 b_2 \sqrt{v_2}} \geq \frac{1}{2 b_1 b_2 \sqrt{n}}. \end{aligned}$$

The number of elements in $\mathcal{U}(v)$ is bigger than $2\sqrt{v_1} - 1$ and since there exists a constant $c > 0$ such that $\sqrt{v_1} \geq c\sqrt{n}$, we see that **A4** holds with $\varphi_v(n)$ independent of v .

Finally, we note that from $P(V_i \in \mathcal{V}_i) \geq b_o$ (where b_o is close to 1), it follows that $P(V \in \mathcal{V}) \geq 1 - 2(1 - b_o)$.

3.2. About assumption A1 for the longest common subsequence. The assumption **A1** depends very much on concrete model and the scoring function S . Even when **A1** it is intuitively understandable, it is, in general, very difficult to prove. Let us briefly explain the intuition behind **A1** in the case of the longest common subsequence. Thus $L(Z) = \ell(Z)$ is the length of the longest common subsequence.

Suppose that there is a letter in \mathbb{A} , say a so that the pair $\mathbf{A}^* := (a, a)$ has high probability. Such a situation might occur in many cases in practice, for example when X and Y are independent stationary Markov chains having the same distribution and the probability $P(X_1 = a)$ is very high. Since the pair \mathbf{A}^* has high probability, typically the sequence Z_1, Z_2, \dots, Z_n has many \mathbf{A}^* s. Then, in the construction of V and U , take $\mathbf{A} = \mathbf{B} = \mathbf{D} = \mathbf{A}^*$. In this case the random variable V counts the number of $(\mathbf{A}^* \cdot \mathbf{A}^*)$ triplets in certain positions (i.e first three letters, then letters 4, 5, 6 etc) and U counts the number of \mathbf{A}^* s between these triplets. The random variable \mathcal{R} now picks any non- \mathbf{A}^* in-between the triplet (with uniform distribution) and changes it into \mathbf{A}^* . Clearly \mathcal{R} then changes at least one non- a -letter into a -letter. As a result, the number of \mathbf{A}^* s increases and the number of a s in X and Y increases as well. Since there are many

\mathbf{A}^* s in Z and, therefore, many as in X and Y , any longest common subsequence has to connect many as on the X -side with as on the Y -side. If the probability of a in X is very high, then any longest common subsequence consists of mostly a -pairs. It does not necessarily mean that two as in the same position (thus a \mathbf{A}^* -pair) would be necessarily connected by LCS, but it is very likely that both as in a \mathbf{A}^* are connected. In fact, as the simulations in [3] showed, with highly asymmetrical distribution of X_i (i.e. having a letter a with high probability) the subsequence that aligns as many as as possible is very close to being the longest. Hence, if X and Y sequences both have many a -letters then any LCS connects mostly a -letters. That implies that non- a -letters have bigger likelihood to remain unconnected, because connecting a pair of non- a -letters will typically destroy many connected a -letter pairs. Thus changing at least one non- a -letter into an a -letter, has tendency to increase LCS.

The above-described approach has been formalized in [20, 29]. In those articles, X and Y are considered independent i.i.d. sequences, where X_1 and Y_1 have the common asymmetric distribution over \mathbb{A} (in [29] a two letter alphabet is considered; in [20] the result is generalized for many letter alphabet). The asymmetry means that one letter, say a , has the probability close to one. Thus both sequences consist mostly of as . In these papers, the random transformation picks any non- a letter from these letters in X and Y letters and then changes it into a . In this case, the random variable U counts as in X and Y sequence, and there is no need for V -variable, formally take $V \equiv 2n$.

Formally the described random transformation used in these two papers differs from the one in the present article by several aspects:

- (1) The sequences X and Y are considered separately, not pairwise. This is due to the independence of X and Y . If X and Y are independent Markov chains, then we could define \mathcal{R} also as follows: consider all (non-overlapping) triplets in X and Y sequences separately and let V count the $a \cdot a$ -ones. The maximal number of such triplets would be $2\lfloor \frac{n}{3} \rfloor$, not $\lfloor \frac{n}{3} \rfloor$ as in our case. Then pick any triplet with non- a -letter in between (either in X or Y sequence) and change the middle letter into a . The random variable U counts the as in the middle of the triplets. Surely, due to the independence of X and Y , the conditional distribution of U given $V = v$ is still Binomial and it is straightforward to verify that everything else holds as well. When X and Y are dependent, one needs them to consider them pairwise in order to obtain the conditional independence of B_1, \dots, B_v given $V = v$.
- (2) There are no fixed $a \cdot a$ -neighborhoods and hence also no V -variable. The fixed neighborhood is not needed, because X and Y already consists of independent random variables. And the number of as is Binomially distributed. In the case on Markov chains, the fixed neighborhoods are needed, again, to obtain the conditional independence of B_1, \dots, B_v given $V = v$. Without neighborhoods, there is obviously no need for prescribed triplet-locations.

Thus, although formally different, the random transformation in the present article is of the same nature as the ones used in [29, 20], where it is proven that when the probability of a is close to one then assumption **A1** holds (see [20, Theorem 2.1], [29, Theorem 2.2]). Therefore, it is reasonable to believe, that in the case where an $\mathbf{A}^* = (a, a)$ pair has high enough probability, then \mathcal{R} that replaces a random non- \mathbf{A}^* pair by \mathbf{A}^* satisfies **A1**. To prove that, however, is beyond the scope of the current paper and needs a separate article.

Suppose now that there is a pair of different letters (a, b) such that $P(Z_1 = (a, b))$ is close to one. Then take $\mathbf{A} = \mathbf{B} = \mathbf{D} = (a, b)$ and let the random transformation to change a non (a, b) -pair into (a, b) -pair. Clearly such a random transformation tends to decrease the length of LCS. But when such a transformation decreases the length of LCS by a fixed ϵ_o , then defining $L(Z) = n - \ell(Z)$, we see that **A1** still holds. In other words, it is not important whether \mathcal{R} actually increases or decreases the score, important is that it influences it. Hence, if there is a pair in $\mathbb{A} \times \mathbb{A}$ occurring with sufficiently large probability, then the approach in [20, 29] applies.

3.3. Simulations. The goal of the present subsection is to check the assumption **A1** by simulations. Given random transformation \mathcal{R} and a sequence $Z = Z_1, \dots, Z_n$, let us denote

$$E_n := E[L(\mathcal{R}(Z))|Z] - L(Z),$$

where the expectation is taken over the random transformation. Under **A1**, there exists $\epsilon_o > 0$ such that

$$P(E_n \geq \epsilon_o) \rightarrow 1.$$

If the convergence above is fast enough, then $P(E_n \geq \epsilon_o, \text{ev}) = 1$ implying that $\liminf_n E_n \geq \epsilon_o$, a.s.. Our objective now is to study the asymptotic behavior of E_n for several PMC-models. Throughout the subsection the score is the length of LCS, i.e. $L(Z) = \ell(Z)$. Let us start with the model.

The model. Before introducing our specific model we state the following lemma whose proof is included for the sake of completeness.

Lemma 3.2. *Let Z_1, Z_2, \dots be a Markov chain on X with transition matrix $\mathbb{P} = (p_{xy})_{x,y \in X}$. Suppose that $\{A_i\}_{i \in I}$ is a partition of X and define $\pi : X \rightarrow I$ as $\pi(x) = i$ if and only if $x \in A_i$. Then the following assertions are equivalent:*

- (1) *for every initial distribution of Z_0 , $\pi(Z_1), \pi(Z_2), \dots$ is a Markov chain on I with transition matrix $\mathbb{Q} := (q_{ij})_{i,j \in I}$;*
- (2) *for all $i, j \in I$, $x \in A_i$*

$$\sum_{y \in A_j} p_{xy} = q_{ij}. \quad (3.15)$$

Proof. Let us denote by μ the initial distribution of Z_0 ; hence, $P(\pi(Z_0) = i) = \mu(A_i)$.

(1) \implies (2). From the hypotheses

$$q_{ij} = P(\pi(Z_1) = j | \pi(Z_0) = i) = \frac{\sum_{x \in A_i} P(\pi(Z_1) = j | Z_0 = x) \mu(x)}{P(\pi(Z_0) = i)}$$

and this holds for every distribution μ (s.t. $\mu(A_i) > 0$) if and only if $q_{ij} = P(\pi(Z_1) = j | Z_0 = x)$ for every $x \in A_i$, that is, $\sum_{y \in A_j} p_{xy} = q_{ij}$ for all $x \in A_i$.

(2) \implies (1). If we compute $P(\pi(Z_n) = i_n | \pi(Z_{n-1}) = i_{n-1}, \dots, \pi(Z_0) = i_0)$ by means of the decomposition

$$\{\pi(Z_n) = i_n, \pi(Z_{n-1}) = i_{n-1}, \dots, \pi(Z_0) = i_0\} = \bigcup_{z \in A_{i_0} \times A_{i_1} \times \dots \times A_{i_n}} \{Z_n = z_{n+1}, Z_{n-1} = z_n \dots, Z_0 = z_1\}$$

and by using the Markov property of Z_1, Z_2, \dots and equation (3.15)

$$P(\pi(Z_n) = i_n | \pi(Z_{n-1}) = i_{n-1}, \dots, \pi(Z_0) = i_0) = P(\pi(Z_n) = i_n | \pi(Z_{n-1}) = i_{n-1}) = q_{i_{n-1} i_n}$$

follows easily. \square

From this result we can easily derive the most general transition matrix of a 2-dimensional random walk $Z_n = (X_n, Y_n)$ with state space $\{(1, 1), (1, 0), (0, 1), (0, 0)\}$ whose marginals are Markov chains. More precisely, given the marginals of X and Y with state space $\mathbb{A} = \{0, 1\}$

$$\begin{pmatrix} p & 1-p \\ q & 1-q \end{pmatrix} \quad \begin{pmatrix} p' & 1-p' \\ q' & 1-q' \end{pmatrix}$$

the most general joint transition matrix can be easily obtained by applying Lemma 3.2 twice: first with $A_1 := \{(1, 1), (1, 0)\}$, $A_2 := \{(0, 1), (0, 0)\}$ (to ensure that X_n is a Markov chain) and then with $A_1 := \{(1, 1), (0, 1)\}$, $A_2 := \{(1, 0), (0, 0)\}$ (to ensure that Y_n is a Markov chain). The final result is

$$\begin{pmatrix} p\lambda_1 & p(1-\lambda_1) & p' - p\lambda_1 & 1 + p\lambda_1 - p' - p \\ p\lambda_2 & p(1-\lambda_2) & q' - p\lambda_2 & 1 + p\lambda_2 - q' - p \\ q\mu_1 & q(1-\mu_1) & p' - q\mu_1 & 1 + q\mu_1 - p' - q \\ q\mu_2 & q(1-\mu_2) & q' - q\mu_2 & 1 + q\mu_2 - q' - q \end{pmatrix}$$

with the constraints

$$\lambda_1 \in \left[\frac{p' + p - 1}{p} \vee 0, \frac{p'}{p} \wedge 1 \right], \quad \lambda_2 \in \left[\frac{q' + p - 1}{p} \vee 0, \frac{q'}{p} \wedge 1 \right],$$

$$\mu_1 \in \left[\frac{p' + q - 1}{q} \vee 0, \frac{p'}{q} \wedge 1 \right], \quad \mu_2 \in \left[\frac{q' + q - 1}{q} \vee 0, \frac{q'}{q} \wedge 1 \right].$$

This 4-parameter model is sufficiently flexible and general to cover a large variety of cases. When $p = p'$ and $q = q'$, i.e. X and Y have the same distribution, then the transition matrix is simply

$$\begin{pmatrix} p\lambda_1 & p(1-\lambda_1) & p(1-\lambda_1) & 1+p(\lambda_1-2) \\ p\lambda_2 & p(1-\lambda_2) & q-p\lambda_2 & 1+p\lambda_2-q-p \\ q\mu_1 & q(1-\mu_1) & p-q\mu_1 & 1+q\mu_1-p-q \\ q\mu_2 & q(1-\mu_2) & q(1-\mu_2) & 1+q(\mu_2-2) \end{pmatrix}.$$

This is the case we are considering in the present subsection. In what follows, without loss of generality, we shall assume that $p \geq q$. The parameters λ_i and μ_i regulate the dependence between marginal sequences X and Y . Clearly X and Y are independent if and only if $\lambda_1 = \mu_1 = p$ and $\lambda_2 = \mu_2 = q$. The transition matrix corresponding to that particular choice of parameters will be denoted by \mathbb{P}_{ind} . If λ_i and μ_i are maximal i.e.

$$\lambda_1 = 1, \quad \lambda_2 = q/p, \quad \mu_1 = \mu_2 = 1,$$

then X and Y are (in a sense) maximally positive-dependent and the corresponding transition matrix is

$$\begin{pmatrix} p & 0 & 0 & 1-p \\ q & p-q & 0 & 1-p \\ q & 0 & p-q & 1-p \\ q & 0 & 0 & 1-q \end{pmatrix}.$$

We shall call this case *maximal dependence*, and the "maximal" here means the maximal number of similar pairs $(1, 1)$ or $(0, 0)$. The matrix above is not irreducible and therefore in the simulations below, we shall use the following "nearly" maximal dependence matrix

$$\mathbb{P}_{\text{max}}(p, q) = \begin{pmatrix} p-\epsilon & \epsilon & \epsilon & 1-p-\epsilon \\ q & p-q & 0 & 1-p \\ q & 0 & p-q & 1-p \\ q-\epsilon & \epsilon & \epsilon & 1-q-\epsilon \end{pmatrix}, \quad (3.16)$$

where to the end of the subsection $\epsilon = 0.05$. Clearly the distribution of X and Y is not affected by adding ϵ . The (nearly) maximal dependent Z favors pairs $(0, 0)$ and $(1, 1)$. When p and q are relatively high, then typical outcome of Z will have many pairs $(1, 1)$ and then changing a non $(1, 1)$ -pair into a $(1, 1)$ -pair has a tendency to increase the score.

We shall also consider the "minimal dependence" matrix that corresponds to the small λ_i and μ_i . Such model favors dissimilar pairs $(0, 1)$ and $(1, 0)$. Due to the fact that X and Y sequences have the same transition matrix, unlike in the case of maximal dependence, the minimal dependence (corresponding to the correlation -1) is not always totally achieved and the structure of minimal dependence matrix depends more on p and q . In the simulations we shall use the following minimal dependence matrices (again ϵ is added to have an irreducible chain):

$$\mathbb{P}_{\text{min}}(p, q) = \begin{cases} \begin{pmatrix} 2p-1+\epsilon & 1-p-\epsilon & 1-p-\epsilon & \epsilon \\ p+q-1 & 1-q & 1-p & 0 \\ p+q-1 & 1-p & 1-q & 0 \\ 2q-1+\epsilon & 1-q-\epsilon & 1-q-\epsilon & \epsilon \end{pmatrix}, & \text{if } p+q > 1 \text{ and } q \geq \frac{1}{2}; \\ \begin{pmatrix} 2p-1+\epsilon & 1-p-\epsilon & 1-p-\epsilon & \epsilon \\ p+q-1 & 1-q & 1-p & 0 \\ p+q-1 & 1-p & 1-q & 0 \\ \epsilon & q-\epsilon & q-\epsilon & 1-2q+\epsilon \end{pmatrix}, & \text{if } p+q > 1 \text{ and } q < \frac{1}{2}. \end{cases} \quad (3.17)$$

The simulations. Let us briefly describe the simulations for a fixed transition matrix \mathbb{P} . First, let us fix $\mathbf{A} \in \{0, 1\}$. Then we fix a transition matrix \mathbb{P} and generate a Markov sequence $Z_1, \dots, Z_{3 \cdot 7500}$ according to the stationary distribution corresponding to \mathbb{P} . Denote

$$J_m := \{j : j \leq m, Z_{3j-2} = Z_{3j} = \mathbf{A}, Z_{3j-1} \neq \mathbf{A}\}.$$

For each $m = 100, 200, \dots, 7500$ we do the following procedure. If $J_m = \emptyset$ we don't do anything and just pick the next m . Suppose now that $J_m \neq \emptyset$ (obviously then also $J_{m+100} \neq \emptyset$). We compute $l(m)$, the length of LCS of (Z_1, \dots, Z_{3m}) . Next, for each $j = 1, \dots, |J_m|$ we do the following subprocedure. We compute $l(m, j)$, the length of LCS of the sequence

$$(Z_1, \dots, Z_{3j-2}, \mathbf{A}, Z_{3j}, \dots, Z_{3m}).$$

Next we compute the difference

$$r(m, j) := l(m, j) - l(m).$$

Note that $r(m, j) \in \{-2, -1, 0, 1, 2\}$. By the end of this subprocedure we have $|J_m|$ values $r(m, 1), \dots, r(m, |J_m|)$ and we compute

$$E(m) := \frac{1}{|J_m|} \sum_{i=1}^{|J_m|} r(m, i).$$

Recall that $E_n = E[L(\mathcal{R}(Z_1, \dots, Z_n))|Z] - L(Z_1, \dots, Z_n)$. Note that $E_{3m} \stackrel{d}{=} E(m)$, where \mathcal{R} is the random transformation used in the proof of **A3**, with $\mathbf{B} = \mathbf{D} = \mathbf{A}$. The final goal is to see whether there are indications of the existence of a positive ϵ_o such that $|E(m)| \geq \epsilon_o$ eventually.

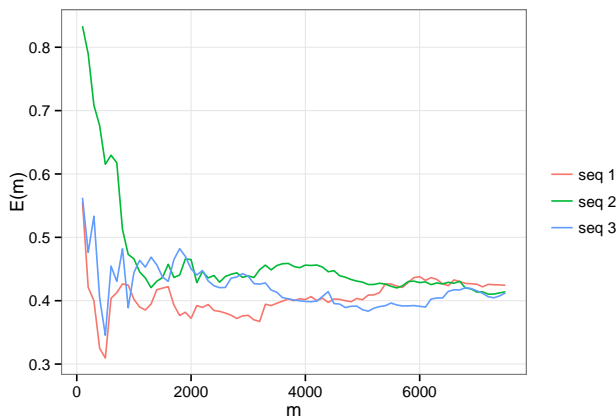
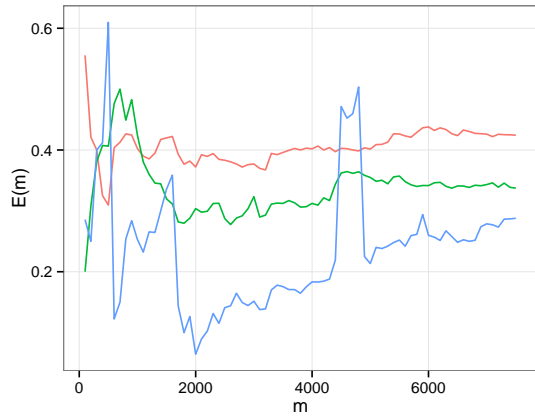


FIGURE 1. The behaviour of $E(m)$ with $\mathbb{P} = \mathbb{P}_{\max}$, $p = 0.9$, $q = 0.7$, and $\mathbf{A} = (1, 1)$. Note that for all three chains, $E(m)$ seems to converge to same positive and constant limit.

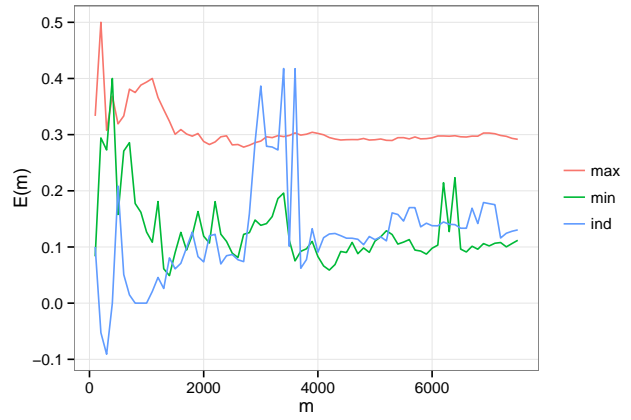
We start our simulations with $\mathbf{A} = (1, 1)$. In Figure 1, three different sequences $Z_1, \dots, Z_{3 \cdot 7500}$ are generated with the same distribution corresponding to matrix $\mathbb{P}_{\max}(0.9, 0.7)$. In this case, for every t , $P(Z_t = (1, 1)) = 0.819$ and turning a non-(1,1)-pair into a (1,1)-pair clearly has positive effect to the score. From Figure 1, it is evident that $E(m)$ not only is bounded away from zero, but also converges to a strictly positive constant limit (which we estimate to be around 0.4). The convergence is not needed for **A1** to hold, but based on that picture, we conjecture that (at least for some models) E_n a.s. tends to a constant limit.

Figure 1 also indicates that our choice of m is big enough to in the sense that all different sequences behave similarly. Thus, in what follows, we shall generate only one sequence for every \mathbb{P} . In Figure 2, for several choices of (p, q) three different models, independent, maximal dependent (3.16) and minimal dependent (3.17), are considered. From Figure 2, we see that in the case of \mathbb{P}_{\max} (where the probability of (1,1)-pair is the highest, corresponding to the red line) $E(m)$ clearly is bounded away from zero for every (p, q) . For independent sequences and the sequences corresponding to \mathbb{P}_{\max} , the desired boundedness is evident for models with relatively big p and q (upper row), whilst for smaller p and q , it might not be so. This is due to relatively low number of (1,1)-pairs. Indeed, for independent sequences the probability of (1,1)-pair is pq and so if $p = 0.7$ and $q = 0.4$ (D), the proportion of (1,1)-pairs is too small for our random transformation to have positive effect to the score.

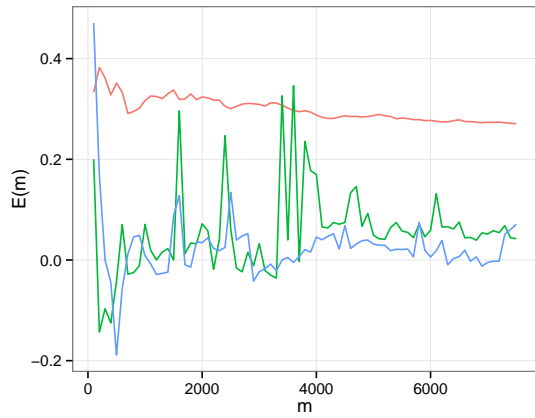
The random transformation considered so far is designed to increase the score and for most of the models



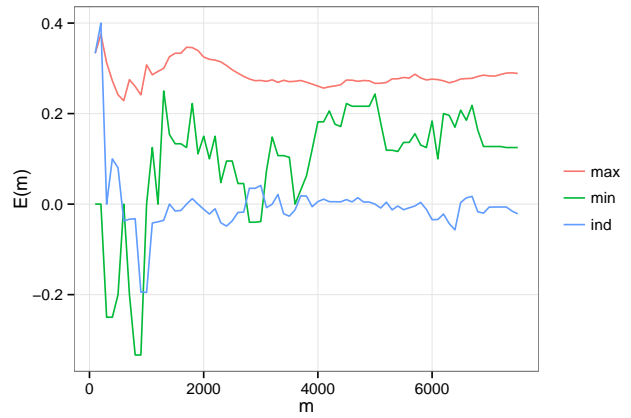
(A) $p = 0.9, q = 0.7$



(B) $p = 0.8, q = 0.6$



(C) $p = 0.7, q = 0.7$

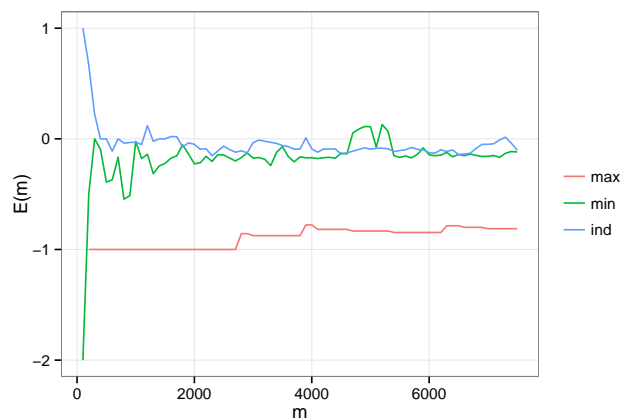


(D) $p = 0.7, q = 0.4$

FIGURE 2. The behaviour of $E(m)$ with transition matrices \mathbb{P}_{\max} , \mathbb{P}_{\min} and \mathbb{P}_{ind} , with $\mathbf{A} = (1, 1)$.



(A) $p = 0.7, q = 0.7$



(B) $p = 0.7, q = 0.4$

FIGURE 3. The behavior of $E(m)$ with transition matrices \mathbb{P}_{\max} , \mathbb{P}_{\min} and \mathbb{P}_{ind} , with $\mathbf{A} = (0, 1)$.

in Figure 2, it indeed does so. We next consider a new \mathcal{R} that tends to decrease the score. For that, we just take $\mathbf{A} = (0, 1)$. In Figure 3, we repeat, with this new \mathcal{R} , the same simulations of cases (C) and (D) of Figure 2. The choice of these cases is due to the fact that, for \mathbb{P}_{\min} and \mathbb{P}_{ind} , the former transformation \mathcal{R} (with $\mathbf{A} = (1, 1)$) did not convincingly show the existence of the positive lower bound ϵ_0 . For the independent marginals case ($p = q = 0.7$), the behavior of $E(m)$ is much better now and we can conclude that $E(m)$ converges a.s. to a constant limit that for cases \mathbb{P}_{\min} and \mathbb{P}_{ind} are in $(-0.25, -0.5)$. Recall that the negative limit also ensures **A1**, we just formally have to consider a different score function. In the other case, namely the case (B) of Figure 3, we see indications of the convergence of $E(m)$, but for \mathbb{P}_{\min} and \mathbb{P}_{ind} , it is difficult to conclude whether the limit is different from zero or not.

In the case (A) of Figure 3, the probability $P(Z_t = (0, 1))$ is 0.045 (max), 0.28 (min) and 0.49 (ind). The same probabilities in the case (B) are 0.063, 0.418 and 0.245 respectively. We see that, especially for \mathbb{P}_{\max} , the number of $(0, 1)$ -pairs in the sequence is very small and that jeopardies the simulations in this case. The small number of $(0, 1)$ pairs is evident from the pictures, where the red line is not varying much. Therefore, we combine the transformations by taking $\mathbf{A}_1 = \mathbf{B}_1 = \mathbf{D}_1 = (1, 0)$ and $\mathbf{A}_2 = \mathbf{B}_2 = \mathbf{D}_2 = (0, 1)$. As before, we generate a Markov sequence $Z_1, \dots, Z_{3 \cdot 7500}$ according to the stationary distribution. We then apply the procedure described above twice: first with $\mathbf{A} = (1, 0)$, and then with $\mathbf{A} = (0, 1)$. In this way we obtain the sets J_m^1 and the LCS-differences $r_1(m, i)$ (corresponding to the pair $(1, 0)$), and the sets J_m^2 and the LCS-differences $r_2(m, i)$ (corresponding to the pair $(0, 1)$). Finally we define

$$E(m) := \frac{1}{|J_m^1| + |J_m^2|} \left(\sum_{i=1}^{|J_m^1|} r_1(m, i) + \sum_{i=1}^{|J_m^2|} r_2(m, i) \right).$$

Again, note that $R_{3m} \stackrel{d}{=} E(m)$, where \mathcal{R} is now the combined random transformation with $\mathbf{A}_1 = \mathbf{B}_1 = \mathbf{D}_1 = (1, 0)$ and $\mathbf{A}_2 = \mathbf{B}_2 = \mathbf{D}_2 = (0, 1)$. When we described the combined transformation in Section 3.1, we mainly considered the case $q_1 = q_2$: this is true for our transition matrices \mathbb{P}_{\max} , \mathbb{P}_{\min} , \mathbb{P}_{ind} , so the use of combined random transformations in the simulations is justified¹. The results of these new simulations are presented in Figure 4. Since in all cases $P(Z_t = (0, 1)) = P(Z_t = (1, 0))$, including $(1, 0)$ into \mathcal{R} has the same effect as doubling the number of simulations in Figure 3. We see that the red line now varies more and we can believe that there is a convergence. In the case $p = q = 0.7$, the convergence of green and blue lines to the limits around -0.4 is now even more evident, and for the most difficult case $p = 0.7, q = 0.4$, we now can deduce that $\limsup_m E(m) < 0$, i.e. **A1** also holds in this case.

4. THE UPPER BOUND

In order to judge the sharpness of the lower bound, we briefly calculate the upper bounds of $\Phi(|L_n - EL_n|)$ in the case $\phi(x) = x^r$. In the case of independent random variables, there are many ways of finding upper bound starting from Efron-Stein type of inequalities when $r = 2$. For an overview of several methods for obtaining the upper bound, see [25]. However, most of the methods assume independence of random letters. In the case of PMC-model, probably the easiest way to get an upper bound of the correct order seems to be via the following McDiarmid's-type of inequality for Markov chains (see [39, Corollary 2.9])

Theorem 4.1. *Let $Z := Z_1, \dots, Z_n$ be a homogeneous Markov chain with state space \mathcal{Z} and mixing time $t(\epsilon)$. Let $f : \mathcal{Z}^n \rightarrow \mathbb{R}$ be function satisfying the bounded difference inequality; for every $z, z' \in \mathcal{Z}^n$*

$$|f(z) - f(z')| \leq \sum_{i=1}^n c_i I_{\{z_i \neq z'_i\}},$$

¹More specifically, note that when $|\mathbb{A}| = 2$, then, as it is easy to see, the following conditions are sufficient for $q_1 = q_2$ to hold: $\mathbb{P}_{22} = \mathbb{P}_{33}, \mathbb{P}_{23} = \mathbb{P}_{32}, \mathbb{P}_{21} = \mathbb{P}_{31}, \mathbb{P}_{12} = \mathbb{P}_{13}, \mathbb{P}_{24} = \mathbb{P}_{34}, \mathbb{P}_{42} = \mathbb{P}_{43}$. The transition matrices $\mathbb{P}_{\max}, \mathbb{P}_{\min}, \mathbb{P}_{\text{ind}}$ satisfy those equalities.

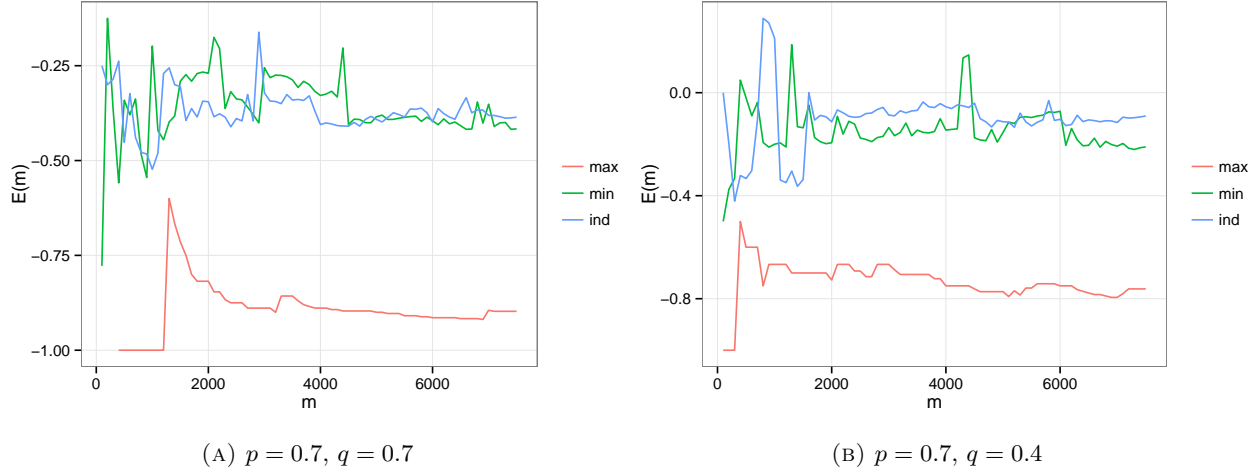


FIGURE 4. The behaviour of $E(m)$ with transition matrices \mathbb{P}_{\max} , \mathbb{P}_{\min} and \mathbb{P}_{ind} using combined random transformations: $\mathbf{A}_1 = \mathbf{B}_1 = \mathbf{D}_1 = (1, 0)$ and $\mathbf{A}_2 = \mathbf{B}_2 = \mathbf{D}_2 = (0, 1)$.

where $c := (c_1, \dots, c_n)$ are some nonnegative constants. Then, for any $s > 0$

$$P\left(|f(Z) - Ef(Z)| > s\right) \leq 2 \exp\left[-\frac{s^2}{8\|c\|^2 t_{\text{mix}}}\right], \quad (4.1)$$

where $\|c\|^2 = \sum_i c_i^2$ and $t_{\text{mix}} = t(1/4)$.

We are going to apply this theorem for $f = L$. Since the change of a value of Z_i changes the score by at most 2Δ , we have the bounded difference property with $c_i = 2\Delta$ and $\|c\|^2 = n4\Delta^2$. Since, by assumption Z is aperiodic, there exists $m \geq 1$ such that

$$\min_{\mathbf{A}, \mathbf{B} \in \mathbb{A} \times \mathbb{A}} P(Z_{1+m} = \mathbf{B} | Z_1 = \mathbf{A}) =: p_o > 0.$$

Then, as it is well-known,

$$\max_{\mathbf{A} \in \mathbb{A} \times \mathbb{A}} \|\pi(\cdot) - P(Z_{1+n} \in \cdot | Z_1 = \mathbf{A})\| \leq C\rho^n,$$

where π is stationary distribution of Z , $\|\cdot\|$ is total variation distance, $\rho := (1 - |\mathbb{A}|^2 p_o)^{\frac{1}{m}}$ and $C = 1$ if $r = 1$, and $C := (1 - |\mathbb{A}|^2 p_o)^{-1}$, otherwise. Therefore

$$t(\epsilon) \leq \frac{\ln \epsilon - \ln C}{\ln(\rho)} < \infty, \quad t_{\text{mix}} \leq \frac{-(\ln 4 + \ln C)}{\ln(\rho)} < \infty.$$

Applying now equation (4.1), we get

$$P(|L(Z) - E(L(Z))| \geq s) \leq 2 \exp\left(-\frac{s^2}{nF}\right), \quad (4.2)$$

where $F := 32\Delta^2 t_{\text{mix}}$. From that it is trivial to get the upper bound. Take $W_n = |L(Z) - E(L(Z))|$

$$E(W_n^r) = \int_0^\infty P(W_n \geq t^{\frac{1}{r}}) dt \leq x + 2 \int_x^\infty \exp\left(-\frac{t^{\frac{2}{r}}}{nF}\right) dt.$$

Minimizing in x , i.e., taking $x = (F(\ln 2)n)^{r/2}$, and changing variables $u = t^{2/r}/(Fn)$, lead to:

$$E(W_n^r) \leq (F(\ln 2)n)^{\frac{r}{2}} + rF^{r/2}n^{\frac{r}{2}} \int_{\ln 2}^\infty e^{-u} u^{\frac{r}{2}-1} du = n^{\frac{r}{2}} F^{r/2} \left[(\ln 2)^{\frac{r}{2}} + r \int_{\ln 2}^\infty e^{-u} u^{\frac{r}{2}-1} du \right],$$

an upper bound of the form $C(r)n^{r/2}$, where

$$C(r) := F^{r/2} \left[(\ln 2)^{\frac{r}{2}} + r \int_{\ln 2}^{\infty} e^{-u} u^{\frac{r}{2}-1} du \right].$$

When $x = 0$, the corresponding constant is slightly bigger than $C(r)$, and is given by:

$$D(r) := rF^{r/2} \int_0^{\infty} e^{-u} u^{\frac{r}{2}-1} du = rF^{r/2} \Gamma\left(\frac{r}{2}\right).$$

4.1. Appendix: proof of Theorem 2.1. Let $B_n \subset \mathcal{Z}_n$ be the set of outcomes of Z such that

$$\{E[L(\mathcal{R}(Z)) - L(Z)|Z] \geq \epsilon_o\} = \{Z \in B_n\}.$$

Let the set $\mathcal{V}_n^o \subset \mathcal{S}_n^V$ be defined as follows:

$$v \in \mathcal{V}_n^o \Leftrightarrow P(Z \notin B_n | V = v) \leq \sqrt{\Delta_n}. \quad (4.3)$$

Now

$$\Delta_n \geq P(Z \notin B_n) \geq \sum_{v \notin \mathcal{V}_n^o} P(Z \notin B_n | V = v) P(V = v) > \sqrt{\Delta_n} P(V \notin \mathcal{V}_n^o), \Rightarrow P(V \notin \mathcal{V}_n^o) \leq \Delta_n^{\frac{1}{2}}.$$

Furthermore, for every $v \in \mathcal{V}_n^o$, let $\mathcal{U}_n^o(v) \subset \mathcal{S}_n(v)$ be defined as follows

$$u \in \mathcal{U}_n^o(v) \Leftrightarrow P(Z \notin B_n | V = v, U = u) \leq \Delta_n^{\frac{1}{4}}. \quad (4.4)$$

Again,

$$\begin{aligned} \sqrt{\Delta_n} &\geq P(Z \notin B_n | V = v) \geq \sum_{u \notin \mathcal{U}_n^o(v)} P(Z \notin B_n | V = v, U = u) P(U = u | V = v) \\ &> \Delta_n^{\frac{1}{4}} P(U \notin \mathcal{U}_n^o(v) | V = v), \Rightarrow P(U \notin \mathcal{U}_n^o(v) | V = v) \leq \Delta_n^{\frac{1}{4}}. \end{aligned}$$

We now show that there exists n_o so big that when $v \in \mathcal{V}_n^o \cap \mathcal{V}_n$ and $u \in \mathcal{U}_n(v) \cap \mathcal{U}_n^o(v)$, then

$$l(u+1, v) - l(u, v) \geq \frac{\epsilon_o}{2}. \quad (4.5)$$

Let $Z_{(u,v)}$ be a random vector having the distribution $P_{(u,v)}$. By **A3**, thus,

$$l(u+1, v) = E[L(\mathcal{R}(Z_{(u,v)}))].$$

Hence

$$\begin{aligned} l(u+1, v) - l(u, v) &= E[L(\mathcal{R}(Z_{(u,v)}))] - E[L(Z_{(u,v)})] = E[L(\mathcal{R}(Z_{(u,v)})) - L(Z_{(u,v)})] \\ &= E(E[L(\mathcal{R}(Z_{(u,v)})) - L(Z_{(u,v)}) | Z_{(u,v)}]). \end{aligned}$$

By assumption **A2**, for any pair of sequences z , the worst decrease of the score, when applying the random transformation is $-A$. Hence,

$$E(E[L(\mathcal{R}(Z_{(u,v)})) - L(Z_{(u,v)}) | Z_{(u,v)}]) \geq \epsilon_o P(Z_{(u,v)} \in B_n) - AP(Z_{(u,v)} \notin B_n) \geq \epsilon_o(1 - \Delta_n^{\frac{1}{4}}) - A\Delta_n^{\frac{1}{4}}.$$

The last inequality follows from the fact that by definition of $\mathcal{U}_n^o(v)$, when $v \in \mathcal{V}_n^o$ and $u \in \mathcal{U}_n^o(v)$, it holds

$$P(Z_{(u,v)} \in B_n) = P(Z \in B_n | V = v, U = u) \geq 1 - \Delta_n^{\frac{1}{4}}.$$

Since $\Delta_n \rightarrow 0$, there exists n_o so big that $\epsilon_o(1 - \Delta_n^{\frac{1}{4}}) - A\Delta_n^{\frac{1}{4}} \geq \frac{\epsilon_o}{2}$, provided $n > n_o$. In what follows, we assume $n > n_o$.

Fix $v \in \mathcal{V}_n^o \cap \mathcal{V}_n$ and consider the set $\mathcal{U}_n(v) \cap \mathcal{U}_n^o(v)$ and $n > n_o$. When $u \in \mathcal{U}_n(v) \cap \mathcal{U}_n^o(v)$, then by equation (4.5) $l(u+1, v) - l(u, v) \geq \frac{\epsilon_o}{2}$. When $u \notin \mathcal{U}_n(v) \cap \mathcal{U}_n^o(v)$, then $l(u+1, v) - l(u, v) \geq -A$.

Recall $\mathcal{U}_n(v) = \{u_n(v) + 1, \dots, u_n(v) + m_n(v)\}$. The set $\mathcal{U}_n(v) \cap \mathcal{U}_n^o(v)$ can be represented as the union of disjoint intervals of $\mathcal{U}_n(v)$, i.e.

$$\mathcal{U}_n(v) \cap \mathcal{U}_n^o(v) = \bigcup_{j=1}^{k(v)} I_j(v),$$

where

$$I_j(v) = \{u_n(j, v) + 1, \dots, u_n(j, v) + m_n(j, v)\}$$

is a subinterval of $\mathcal{U}_n(v)$. Obviously the number of intervals $k(v)$ as well as the intervals $I_j(v)$ depend on n . On every interval $I_j(v)$, the function $l(\cdot, v)$ increases with the slope at least $\frac{\epsilon_o}{2}$ i.e.

$$\text{if } u \in I_j(v), \text{ then } l(u+1, v) - l(u, v) \geq \frac{\epsilon_o}{2}. \quad (4.6)$$

Let us consider the sets

$$J_j(v) := \{l(u_n(j, v) + 1, v), \dots, l(u_n(j, v) + m_n(j, v), v)\} \quad j = 1, \dots, k(v).$$

Thus $J_j(v)$ is the image of the set $I_j(v)$ when applying $l(\cdot, v)$. Note that if $u = u_n(j, v) + m_n(j, v)$, i.e. u is the last element in the interval, then $l(u+1, v)$ is outside of the interval $J_j(v)$. We know that all elements of $J_j(v)$ are at least $\frac{\epsilon_o}{2}$ -apart from each other. However, the intervals $J_j(v)$ might overlap (even though we know that the intervals $I_j(v)$ do not). Since for any $u \in \mathcal{U}_n(v) \setminus \mathcal{U}_n^o(v)$, it holds that $l(u+1, v) - l(u, v) \geq -A$, we have

$$\sum_{u \in \mathcal{U}_n(v) \setminus \mathcal{U}_n^o(v)} (l(u+1, v) - l(u, v)) \geq -A|\mathcal{U}_n(v) \setminus \mathcal{U}_n^o(v)|. \quad (4.7)$$

The inequality equation (4.7) together with equation (4.6) implies that the sum of the lengths of (integer) intervals $J_j(v)$ differs from the length of the set $J(v) := \bigcup_{j=1}^{k(v)} J_j(v)$ at most by $A|\mathcal{U}_n(v) \setminus \mathcal{U}_n^o(v)|$. Formally, defining for any finite set of real numbers T the length $\ell(T)$ as the difference between maximum and minimum element of T i.e.

$$\ell(J_j(v)) := l(u_n(j, v) + m_n(j, v), v) - l(u_n(j, v) + 1, v),$$

we obtain

$$\sum_{j=1}^{k(v)} \ell(J_j(v)) - \ell(J(v)) \leq \sum_{j=1}^{k(v)} \ell(J_j(v)) - \left(l(u_n(k, v) + m_n(k, v), v) - l(u_n(1, v) + 1, v) \right) \leq A|\mathcal{U}_n(v) \setminus \mathcal{U}_n^o(v)|. \quad (4.8)$$

The first inequality follows from the fact that

$$l(u_n(k, v) + m_n(k, v), v) - l(u_n(1, v) + 1, v) \leq \ell(J(v))$$

and the second from equations (4.7) and (4.6).

The number of $\frac{\epsilon_o}{2}$ -apart points needed for covering an (real) interval with length $A|\mathcal{U}_n(v) \setminus \mathcal{U}_n^o(v)|$ is at most

$$\frac{2A|\mathcal{U}_n(v) \setminus \mathcal{U}_n^o(v)|}{\epsilon_o} + 1.$$

This means that due to the overlapping at most $\frac{2A|\mathcal{U}_n(v) \setminus \mathcal{U}_n^o(v)|}{\epsilon_o} + 1$ points that are $\frac{\epsilon_o}{2}$ -apart will be lost implying that in the set $J(v)$ there are at least

$$|\mathcal{U}_n(v)| - \frac{2A|\mathcal{U}_n(v) \setminus \mathcal{U}_n^o(v)|}{\epsilon_o} - 1 = m_n(v) - \frac{2A|\mathcal{U}_n(v) \setminus \mathcal{U}_n^o(v)|}{\epsilon_o} - 1$$

points that are (at least) $\frac{\epsilon_o}{2}$ -apart from each other.

Using the inequality (recall $v \in \mathcal{V}_n^o$)

$$P(U \notin \mathcal{U}_n^o(v) | V = v) \leq \Delta_n^{\frac{1}{4}}$$

and equation (2.2) we obtain

$$\Delta_n^{\frac{1}{4}} \geq P(U \in \mathcal{U}_n(v) \setminus \mathcal{U}_n^o(v) | V = v) = \sum_{u \in \mathcal{U}_n(v) \setminus \mathcal{U}_n^o(v)} P(U = u | V = v) \geq |\mathcal{U}_n(v) \setminus \mathcal{U}_n^o(v)| \varphi_v(n)$$

implying that

$$|\mathcal{U}_n(v) \setminus \mathcal{U}_n^o(v)| \leq \Delta_n^{\frac{1}{4}} \varphi_v(n)^{-1}.$$

Thus by **A4**, there exists n_1 such that

$$m_n(v) - \frac{2A|\mathcal{U}_n(v) \setminus \mathcal{U}_n^o(v)|}{\epsilon_o} - 1 \geq m_n(v) - 2A \frac{\Delta_n^{\frac{1}{4}}}{\epsilon_o \varphi_v(n)} - 1 \geq \frac{(c(v)\epsilon_o - 2A\Delta_n^{\frac{1}{4}} - \epsilon_o \varphi_v(n))}{\epsilon_o \varphi_v(n)} \geq \frac{r_v(n)}{\varphi_v(n)}, \quad \forall n > n_1$$

where

$$r_v(n) := c(v) - \frac{2A\Delta_n^{\frac{1}{4}}}{\epsilon_o} - \varphi_v(n) \rightarrow c(v)$$

uniformly with respect to $v \in \mathcal{V}_n$ (for the definition of uniform convergence with respect to a variable in a sequence of sets, see for instance [4, Definition 2.2]). To summarize: the set

$$J(v) \subseteq \{l(u_n(v) + 1, v), \dots, l(u_n(v) + m_n(v), v)\}$$

contains at least $\frac{r_v(n)}{\varphi_v(n)}$ elements being $\frac{\epsilon_o}{2}$ -apart from each other.

Finally, define the set

$$\mathcal{A}_n(v) := \left\{ u \in \mathcal{U}_n(v) : |l(u, v) - \mu_n| \geq \frac{\epsilon_o r_v(n)}{8\varphi_v(n)} \right\}.$$

Since the interval

$$\left[\mu_n - \frac{\epsilon_o r_v(n)}{8\varphi_v(n)}, \mu_n + \frac{\epsilon_o r_v(n)}{8\varphi_v(n)} \right]$$

contains at most $\frac{r_v(n)}{2\varphi_v(n)} + 1$ elements that are $\frac{\epsilon_o}{2}$ -apart from each other and $J(v)$ contains at least $\frac{r_v(n)}{\varphi_v(n)}$ of such elements, it follows that that the set

$$\mathcal{B}_n(v) := \{l(u, v) : u \in \mathcal{A}_n(v)\}$$

contains at least $\frac{r_v(n)}{2\varphi_v(n)} - 1$ points being $\frac{\epsilon_o}{2}$ -apart from each other, and in particular, the set $\mathcal{A}_n(v)$ contains at least $\frac{r_v(n)}{2\varphi_v(n)} - 1$ points i.e $|\mathcal{A}_n(v)| \geq \frac{r_v(n)}{2\varphi_v(n)} - 1$.

By conditional Jensen (recall Φ is convex), we get

$$E[\Phi(|L(Z) - \mu_n|) | V, U] \geq \Phi(|E[L(Z) | V, U] - \mu_n|) = \Phi(|l(U, V) - \mu_n|).$$

Therefore (recall also that Φ is increasing)

$$\begin{aligned}
E\Phi(|L(Z) - \mu_n|) &= E\left(E[\Phi(L(Z) - \mu_n)|V, U]\right) \geq E\Phi(|l(U, V) - \mu_n|) \\
&\geq \sum_{v \in \mathcal{V}_n \cap \mathcal{V}_n^o} \sum_{u \in \mathcal{U}_n(v)} \Phi(|l(u, v) - \mu_n|) P(U = u|V = v) P(V = v) \\
&\geq \sum_{v \in \mathcal{V}_n \cap \mathcal{V}_n^o} \sum_{u \in \mathcal{A}_n(v)} \Phi(|l(u, v) - \mu_n|) P(U = u|V = v) P(V = v) \\
&\geq \sum_{v \in \mathcal{V}_n \cap \mathcal{V}_n^o} \sum_{u \in \mathcal{A}_n(v)} \Phi(|l(u, v) - \mu_n|) \varphi_v(n) P(V = v) \\
&\geq \sum_{v \in \mathcal{V}_n \cap \mathcal{V}_n^o} \Phi\left(\frac{\epsilon_o r_v(n)}{8\varphi_v(n)}\right) |\mathcal{A}_n(v)| \varphi_v(n) P(V = v) \\
&\geq \sum_{v \in \mathcal{V}_n \cap \mathcal{V}_n^o} \Phi\left(\frac{\epsilon_o r_v(n)}{8\varphi_v(n)}\right) \left(\frac{r_v(n)}{2\varphi_v(n)} - 1\right) \varphi_v(n) P(V = v) \\
&= \sum_{v \in \mathcal{V}_n \cap \mathcal{V}_n^o} \Phi\left(\frac{\epsilon_o r_v(n)}{8\varphi_v(n)}\right) \left(\frac{r_v(n)}{2} - \varphi_v(n)\right) P(V = v).
\end{aligned}$$

In particular, if $\varphi(n) = \sup_{v \in \mathcal{V}_n} \varphi_v(n) \rightarrow 0$ as $n \rightarrow \infty$ (that is, $\varphi_v(n)$ converges to 0 uniformly with respect to $v \in \mathcal{V}_n$) then there exists n_2 such that $r_v(n) > \frac{c}{2}$ and $\varphi_v(n) \leq c/8$ for all $n \geq n_2, v \in \mathcal{V}_n$. Thus, if $n \geq n_2$, then

$$E\Phi(|L(Z) - \mu_n|) \geq \Phi\left(\frac{\epsilon_o c}{16\varphi(n)}\right) \left(\frac{c}{4} - \varphi(n)\right) P(V \in \mathcal{V}_n \cap \mathcal{V}_n^o) \geq \Phi\left(\frac{\epsilon_o c}{16\varphi(n)}\right) \frac{c}{8} (P(V \in \mathcal{V}_n) - \Delta_n^{\frac{1}{2}}).$$

If, in addition, $P(V \in \mathcal{V}_n)$ is bounded away from zero, then for any constant c_o satisfying $b_o c/8 > c_o > 0$ we can choose $n_3 \geq n_2$ such that for all $n \geq n_3$

$$E\Phi(|L(Z) - \mu_n|) \geq \Phi\left(\frac{\epsilon_o c}{16\varphi(n)}\right) c_o. \tag{4.9}$$

ACKNOWLEDGEMENTS

The authors acknowledge financial support from INDAM (Istituto Nazionale di Alta Matematica), from Estonian Science foundation Grant no. 5822 and from Estonian institutional research funding IUT34-5.

REFERENCES

- [1] K. S. ALEXANDER, The rate of convergence of the mean length of the longest common subsequence. *Ann. Appl. Probab.*, **4** n. 4 (1994), 1074–1082.
- [2] R. A. BAEZA-YATES, R. GAVALDÀ, G. NAVARRO, AND R. SCHEIHING, Bounding the expected length of longest common subsequences and forests, *Theory Comput. Syst.*, **32** n. 4 (1999), 435–452.
- [3] S. BARDER, J. LEMBER, H. MATZINGER, M. TOOTS, On suboptimal LCS-alignments for independent Bernoulli sequences with asymmetric distributions, *Meth. and Comp. in Appl. Prob.*, **2** (2012), 357 – 382.
- [4] D. BERTACCHI, F. ZUCCA Uniform asymptotic estimates of transition probabilities on combs, *Journal of the Australian Mathematical Society (Ser. A)*, **75** (2003), 325 – 353.
- [5] F. BONETTO, H. MATZINGER, Fluctuations of the Longest Common Subsequence in the Asymmetric Case of 2- and 3-Letter Alphabets, *Latin American Journal of Probability and Mathematical Statistics*, **2** (2006), 195-216.
- [6] H. S. BOOTH, S.F. MACNAMARA, O.M. NIELSEN, S.R. WILSON, An iterative approach to determining the length of the longest common subsequence of two strings, *Methodology Comput. Appl. Probab.*, **6** n. 4 (2004), 401–421.
- [7] S. BOUCHERON, G. LUGOSI, P. MASSART, *Concentration Inequalities: A Nonasymptotic Theory of Independence*, Oxford University Press, Oxford, United Kingdom, 2013.
- [8] J. BOUTET DE MONVEL, Extensive simulations for longest common subsequences, *Eur. Phys. J. B* **7** (1999), 293–308.
- [9] K. M. CHAO, L. ZHANG, *Sequence comparison: theory and methods*, Springer, 2009.
- [10] N. CHRISTIANINI, M. W. HAHN, *Introduction to Computational Genomics*, Cambridge University Press, 2007.
- [11] V. CHVATAL, D. SANKOFF, Longest common subsequences of two random sequences, *J. Appl. Probab.* **12** (1975), 306–315.
- [12] V. DANCİK, M. PATERSON, Upper bounds for the expected length of a longest common subsequence of two binary sequences, *Random Structures Algorithms*, **6** n. 4 (1995), 449–458.
- [13] J. G. DEKEN, Some limit results for longest common subsequences, *Discrete Math.*, **26** n. 1, 17–31.

- [14] S. DERRODE, W. PIECZYNSKI, Signal and image segmentation using Pairwise Markov chains, *IEEE Trans. on Signal Processing* **52** n. 9 (2004), 2477–2489.
- [15] R. DURBIN, S. EDDY, A. KROGH, G. MITCHISON, *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*, Cambridge University Press, 1998.
- [16] S. DERRODE, W. PIECZYNSKI, Unsupervised data classification using pairwise Markov chains with automatic copulas selection, *Computational Statistics and Data Analysis* **63** (2013), 81–98.
- [17] R. DURRETT, *Probability: Theory and Examples*, Thomson, 2005. Third Edition.
- [18] C. DURRINGER, J. LEMBER, H. MATZINGER, Deviation from the Mean in Sequence Comparison with a Periodic Sequence, *Latin American Journal of Probability and Mathematical Statistics*, **3** (2007), 1–29.
- [19] P. W. GLYNN, D. ORMONEIT, Hoeffding inequality for uniformly ergodic Markov chains, *Stat. and Prob. Letters.*, **56** n. 1 (2002), 143–146.
- [20] C. HOUDRÉ, J. MA, On the Order of the Central Moments of the Length of the Longest Common Subsequence, *Preprint*, 2014. Available at arXiv:1212.3265v3.
- [21] R. HAUSER, S. MARTINEZ, H. MATZINGER, Large deviation based upper bounds for the LCS-problem, *Advances in Applied Probability*. **38** (2006), 827–852.
- [22] C. HOUDRÉ, H. MATZINGER, On the Variance of the Optimal Alignment Score for an Asymmetric Scoring Function, *Preprint*, 2007. Available at arXiv:math/0702036.
- [23] R. HAUSER, H. MATZINGER, C. DURRINGER, Approximation to the mean curve in the lcs-problem, *Stochastic Proc. and Appl.*, **118** n. 4 (2008), 629–648.
- [24] C. HOUDRE, J. LEMBER, HEINRICH MATZINGER, On the longest common increasing binary subsequence, *C.R. Acad. Sci. Paris, Ser. I* **343** (2006), 589–594.
- [25] C. HOUDRE, J. LEMBER, Lower bounds on the generalized central moments of the optimal alignments score of random sequences, *ArXiv*, 2015.
- [26] C. HOUDRE, H. MATZINGER, Fluctuations of the Optimal Alignment Score with and Asymmetric Scoring Function, [arXiv:math/0702036]
- [27] J. D. KEVARČIĆ, P. M. VASIĆ, Some Inequalities for the Gamma Function, *Publications de L’institut Mathématique, Nouvelle Série*, tome 11, Vol. 25 (1971), 107–114.
- [28] M. A. KIWI, M. LOEBL, J. MATOUSEK, Expected length of the longest common subsequence for large alphabets, *Advances in Mathematics*, **197** n. 2 (2005), 480–498.
- [29] J. LEMBER, H. MATZINGER, Standard Deviation of the Longest Common Subsequence, *The Annals of Probability*, **37** (2009), No. 3, 1192–1235.
- [30] J. LEMBER, H. MATZINGER, F. TORRES, The Rate of the Convergence of the Mean Score in Random Sequence Comparison, *The Annals of Applied Probability*, **22** n. 3 (2012), 1046–1058.
- [31] J. LEMBER, H. MATZINGER, F. TORRES, General Approach to the Fluctuations Problem in Random Sequence Comparison, *Preprint*, 2012. Available at arXiv:1211.5072v1.
- [32] J. LEMBER, H. MATZINGER, F. TORRES, Proportion of Gaps and Fluctuations of the Optimal Score in Random Sequence Comparison, In *Limit Theorems in Probability, Statistics and Number Theory (In Honor of Friedrich Götze)*, Springer, Vol. 42, 207–234, 2013.
- [33] K.W. LI AND C. C. YANG, Automatic construction of english/chinese parallel corpora, *Journal of the American Society for Information Science and Technology* **54** (2003), 730–742.
- [34] C. Y. LIN, F. J. OCH, Automatic evaluation of machine translation quality using longest common subsequence and skip-bigram statistics, In *ACL '04: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, page 605, 2004.
- [35] G. S. LUEKER, Improved bounds on the average length of longest common subsequences, *J. ACM*, **56** n. 3 (2009), 1–38.
- [36] I. D. MELAMED, Automatic evaluation and uniform filter cascades for inducing n-best translation lexicons, In *Proceedings of the Third Workshop on Very Large Corpora*, 1995.
- [37] I. D. MELAMED, Bixtext maps and alignment via pattern recognition, *Comput. Linguist.*, **25** n. 1 (1999), 107–130.
- [38] M. PATERSON, V. DANCİK, Longest common subsequences, In *Mathematical foundations of computer science 1994 (Kosice, 1994)*, volume 841 of *Lecture Notes in Comput. Sci.*, Springer, Berlin, 1994, pages 127–142.
- [39] D. PAULIN Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electronic Journal of Prob.* **20** (2015), 1–32.
- [40] P. PEVZNER, *Computational Molecular Biology*, MIT Press. An algorithmic approach, A Bradford Book, Cambridge, 2000.
- [41] W. PIECZYNSKI, Pairwise Markov chains, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **25** n. 5 (2003), 634–639.
- [42] T. F. SMITH, M. S. WATERMAN, Identification of common molecular subsequences, *J. Mol. Bio.* **147** (1981), 195–197.
- [43] M. J. STEELE, An Efron-Stein inequality for non-symmetric statistics, *Annals of Statistics*, **14** (1986), 753–758.
- [44] F. TORRES, On the probabilistic longest common subsequence problem for sequences of independent blocks, PhD thesis, Bielefeld University, 2009, Online at <http://bieson.ub.uni-bielefeld.de/volltexte/2009/1473/>
- [45] M. S. WATERMAN, Estimating Statistical Significance of Sequence Alignments, *Philosophical Transactions of the Royal Society: Biological Sciences*, Vol. 344, Issue 1310, 383–390, 1994.
- [46] M. S. WATERMAN, *Introduction to Computational Biology*, Chapman & Hall, 1995.

J. LEMBER, INSTITUTE OF MATHEMATICS AND STATISTICS, UNIVERSITY OF TARTU, J. LIIV 2, 50409 TARTU, ESTONIA.
E-mail address: `juri.lember@ut.ee`

H. MATZINGER, SCHOOL OF MATHEMATICS, GEORGIA TECH, ATLANTA (GA), 30332 USA.
E-mail address: `matzi@math.getech.edu`

J. SOVA, INSTITUTE OF MATHEMATICS AND STATISTICS, UNIVERSITY OF TARTU, J. LIIVI 2, 50409 TARTU, ESTONIA.
E-mail address: `joonas.sova@ut.ee`

F. ZUCCA, DIPARTIMENTO DI MATEMATICA, POLITECNICO DI MILANO, PIAZZA LEONARDO DA VINCI 32, 20133 MILANO, ITALY.
E-mail address: `fabio.zucca@polimi.it`