Invent math (2012) 188:429–463
DOI 10.1007/s00222-011-0350-7

Self-intersections in combinatorial topology: statistical structure

Moira Chas · Steven P. Lalley

Received: 6 December 2010 / Accepted: 21 July 2011 / Published online: 16 August 2011

Abstract Oriented closed curves on an orientable surface with boundary are described up to continuous deformation by reduced cyclic words in the generators of the fundamental group and their inverses. By self-intersection number one means the minimum number of transversal self-intersection points of representatives of the class. We prove that if a class is chosen at random from among all classes of m letters, then for large m the distribution of the self-intersection number approaches the Gaussian distribution. The theorem was strongly suggested by a computer experiment with four million curves producing a very nearly Gaussian distribution.

Mathematics Subject Classification Primary 57M05 · Secondary 53C22 · 37D40

Contents

Supported by NSF grants DMS-0805755 and DMS-0757277.

M. Chas
Department of Mathematics, Stony Brook University, Stony Brook, NY 11794, USA
e-mail: moira@math.sunysb.edu

S.P. Lalley (✉)
Department of Statistics, University of Chicago, 5734 University Avenue, Chicago, IL 60637, USA
e-mail: lalley@galton.uchicago.edu

1 Introduction

Oriented closed curves in a surface with boundary are, up to continuous deformation, described by reduced cyclic words in a set of free generators of the fundamental group and their inverses. (Recall that such words represent the conjugacy classes of the fundamental group.) Given a reduced cyclic word α, define the *self-intersection number* $N(\alpha)$ to be the minimum number of transversal double points among all closed curves represented by α. (See Fig. 1.) Fix a positive integer n and consider how the self-intersection number $N(\alpha)$ varies over the population $\mathcal{F}_n$ of all reduced cyclic words of length n. The value of $N(\alpha)$ can be as small as 0, but no larger than $O(n^2)$. See [6, 7] for precise results concerning the maximum of $N(\alpha)$ for $\alpha \in \mathcal{F}_n$, and [14] for sharp results on the related problem of determining the growth of the number of non self-intersecting closed geodesics up to a given length relative to a hyperbolic metric.

For small values of n, using algorithms in [8], and [5], we computed the self-intersection counts $N(\alpha)$ for all words $\alpha \in \mathcal{F}_n$ (see [4]). Such computations show that, even for relatively small n, the distribution of $N(\alpha)$ over $\mathcal{F}_n$

Fig. 1 Two representatives of $aa\bar{b}\bar{b}$ in the doubly punctured plane. The *second curve* has fewest self-intersections in its free homotopy class

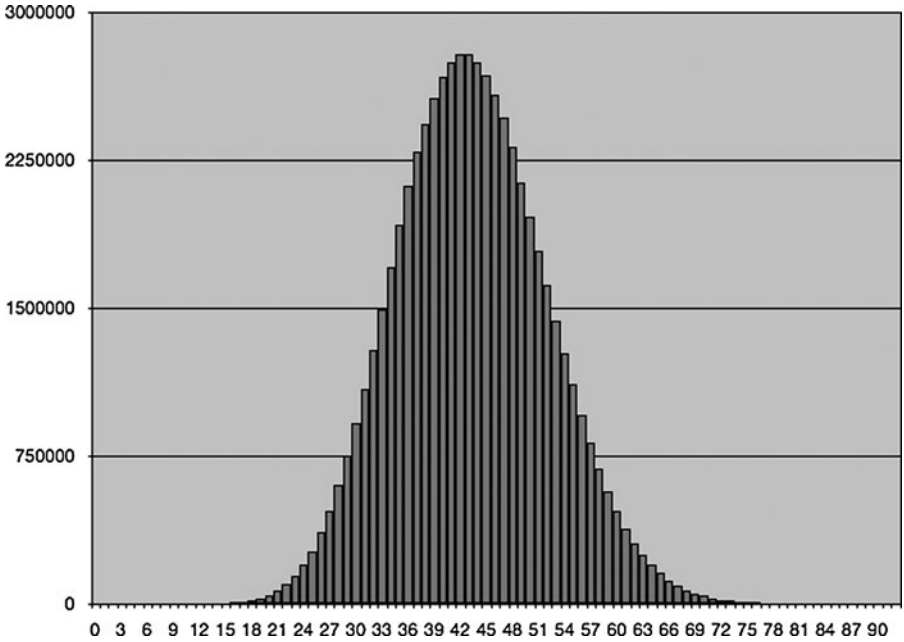

Fig. 2 A *histogram* showing the distribution of self-intersection numbers over all reduced cyclic words of length 19 in the doubly punctured plane. The *horizontal coordinate* shows the self-intersection count k; the *vertical coordinate* shows the number of cyclic reduced words for which the self-intersection number is k

is very nearly Gaussian. (See Fig. 2.) The purpose of this paper is to prove that as $n \to \infty$ the distribution of $N(\alpha)$ over the population $\mathcal{F}_n$, suitably scaled, does indeed approach a Gaussian distribution:

Main Theorem *Let Σ be an orientable, compact surface with boundary and negative Euler characteristic χ, and set*

$$\kappa = \kappa_\Sigma = \frac{\chi}{3(2\chi - 1)} \quad \textit{and} \quad \sigma^2 = \sigma_\Sigma^2 = \frac{2\chi(2\chi^2 - 2\chi + 1)}{45(2\chi - 1)^2(\chi - 1)}. \tag{1}$$

Then for any $a < b$ the proportion of words $\alpha \in \mathcal{F}_n$ such that

$$a < \frac{N(\alpha) - \kappa n^2}{n^{3/2}} < b$$

converges, as $n \to \infty$, to

$$\frac{1}{\sqrt{2\pi}\sigma} \int_a^b \exp\{-x^2/2\sigma^2\}\, dx.$$

Observe that the limiting variance σ^2 is *positive* if the Euler characteristic is negative. Consequently, the theorem implies that (i) for most words $\alpha \in \mathcal{F}_n$ the self-intersection number $N(\alpha)$ is to first order well-approximated by κn^2; and (ii) typical variations of $N(\alpha)$ from this first-order approximation ("fluctuations") are of size $n^{3/2}$.

It is relatively easy to understand (if not to prove) why the number of self-intersections of typical elements of $\mathcal{F}_n$ should grow like n^2. Here follows a short heuristic argument: consider the lift of a closed curve with minimal self-intersection number in its class to the universal cover of the surface Σ. This lift will cross n images of the fundamental polygon, where n is the corresponding word length, and these crossings can be used to partition the curve into n nonoverlapping segments in such a way that each segment makes one crossing of an image of the fundamental polygon. The self-intersection count for the curve is then the number of *pairs* of these segments whose images in the fundamental polygon cross. It is reasonable to guess that for typical classes $\alpha \in \mathcal{F}_n$ (at least when n is large) these segments look like a *random* sample from the set of all such segments, and so the law of large numbers then implies that the number of self-intersections should grow like $n^2\kappa'/2$ where κ' is the probability that two randomly chosen segments across the fundamental polygon will cross. The difficulty in making this argument precise, of course, is in quantifying the sense in which the segments of a typical closed curve look like a random sample of segments. The arguments below (see Sect. 4) will make this clear.

The mystery, then, is not why the mean number of self-intersections grows like n^2, but rather why the size of typical fluctuations is of order $n^{3/2}$ and why the limit distribution is Gaussian. This seems to be connected to geometry. If the surface Σ is equipped with a finite-area Riemannian metric of negative curvature, and if the boundary components are (closed) geodesics then each free homotopy class contains a unique closed geodesic (except for the free homotopy classes corresponding to the punctures). It is therefore possible to order the free homotopy classes by the length of the geodesic representative. Fix L, and let $\mathbb{G}_L$ be the set of all free homotopy classes whose closed geodesics are of length $\leq L$. The main result of [12] (see also [13]) describes the variation of the self-intersection count $N(\alpha)$ as α ranges over the population $\mathbb{G}_L$:

Geometric Sampling Theorem *If the Riemannian metric on Σ is hyperbolic (i.e., constant-curvature -1) then there exists a possibly degenerate probability distribution G on $\mathbb{R}$ such that for all $a < b$ the proportion of words $\alpha \in \mathbb{G}_L$ such that*

$$a < \frac{N(\alpha) + L^2/(\pi^2\chi)}{L} < b$$

converges, as $L \to \infty$, to $G(b) - G(a)$.

The limit distribution is not known, but is likely not Gaussian. The result leaves open the possibility that the limit distribution is degenerate (that is, concentrated at a single point); if this were the case, then the true order of magnitude of the fluctuations might be a fractional power of L. The Geometric Sampling Theorem implies that the typical variation in self-intersection count for a closed geodesic chosen randomly according to hyperbolic length is of order L. Together with the Main Theorem, this suggests that the much larger variations that occur when sampling by word length are (in some sense) due to $\sqrt{n}$-variations in hyperbolic length over the population $\mathcal{F}_n$.

The Main Theorem can be reformulated in probabilistic language as follows (see Appendix B for definitions):

Main Theorem* *Let Σ be an orientable, compact surface with boundary and negative Euler characteristic χ, and let κ and σ be defined by (1). Let N_n be the random variable obtained by evaluating the self-intersection function N at a randomly chosen $\alpha \in \mathcal{F}_n$. Then as $n \to \infty$,*

$$\frac{N_n - n^2\kappa}{\sigma n^{3/2}} \quad \Longrightarrow \quad Normal(0,1) \tag{2}$$

where $Normal(0,1)$ is the standard Gaussian distribution on $\mathbb{R}$ and $\Rightarrow$ denotes convergence in distribution.

2 Combinatorics of self-intersection counts

Our analysis is grounded on a purely combinatorial description of the self-intersection counts $N(\alpha)$, due to [3, 5, 8]. For an example of this analysis, see Appendix A.

Since Σ has non-empty boundary, its fundamental group $\pi_1(\Sigma)$ is free. We will work with a generating set of $\pi_1(\Sigma)$ such that each element has a non-self-intersecting representative. (Such a basis is a natural choice to describe self-intersections of free homotopy classes.) Denote by $\mathcal{G}$ the set containing the elements of the generating set and their inverses and by g the cardinality of $\mathcal{G}$. Thus, $g = 2 - 2\chi$, where χ denotes the Euler characteristic of Σ. We shall assume throughout the paper that $\chi \leq -1$, and so $g \geq 4$. It is not hard to see that there exists a (non-unique and possibly non-reduced) cyclic word $\mathcal{O}$ of length g such that

(1) $\mathcal{O}$ contains each element of $\mathcal{G}$ exactly once.
(2) The surface Σ can be obtained as follows: label the edges of a polygon with $2g$ sides, alternately (so every other edge is not labelled) with the letters of $\mathcal{O}$ and glue edges labeled with the same letter without creating Moebius bands.

This cyclic word $\mathcal{O}$ encodes the intersection and self-intersection structure of free homotopy classes of curves on Σ.

Since $\pi_1(\Sigma)$ is a free group, the elements of $\pi_1(\Sigma)$ can be identified with the *reduced words* (which we will also call *strings*) in the generators and their inverses. A string is *joinable* if each cyclic permutation of its letters is also a string, that is, if its last letter is not the inverse of its first. A *reduced cyclic word* (also called a *necklace*) is an equivalence class of joinable strings, where two such strings are considered equivalent if each is a cyclic permutation of the other. Denote by $\mathcal{S}_n$, $\mathcal{J}_n$, and $\mathcal{F}_n$ the sets of strings, joinable strings, and necklaces, respectively, of length n. Since necklaces correspond bijectively with the conjugacy classes of the fundamental group, the self-intersection count $\alpha \mapsto N(\alpha)$ can be regarded as a function on the set $\mathcal{F}_n$ of necklaces. This function pulls back to a function on the set $\mathcal{J}_n$ of joinable strings, which we again denote by $N(\alpha)$, that is constant on equivalence classes. By [5] this function has the form

$$N(\alpha) = \sum_{1 \le i < j \le n} H(\sigma^i \alpha, \sigma^j \alpha), \tag{3}$$

where $H = H(\mathcal{O})$ is a symmetric function with values in $\{0, 1\}$ on $\mathcal{J}_n \times \mathcal{J}_n$ and $\sigma^i \alpha$ denotes the ith cyclic permutation of α. (Note: σ^2 also denotes the limiting variance in (1), but it will be clear from the context which of the two meanings is in force.)

To describe the function H in the representation (3), we must explain the *cyclic ordering* of letters. For a cyclic word α (not necessarily reduced), set $o(\alpha) = 1$ if the letters of α occur in cyclic (clockwise) order in $\mathcal{O}$, set $o(\alpha) = -1$ if the letters of α occur in *reverse* cyclic (anti-clockwise) order, and set $o(\alpha) = 0$ otherwise. Consider two (finite or infinite) strings, $\omega = c_1 c_2 \dots$ and $\omega' = d_1 d_2 \dots$. For each integer $k \ge 2$ define functions u_k and v_k of such pairs (ω, ω') as follows: First, set $u_k(\omega, \omega') = 0$ unless

(a) both ω and ω' are of length at least k; and
(b) $c_1 \ne d_1$, $c_k \ne d_k$, and $c_j = d_j$ for all $1 < j < k$.

For any pair (ω, ω') such that both (a) and (b) hold, define

$$u_k(\omega, \omega') = \begin{cases} 1 & \text{if } k = 2, \text{ and } o(\bar{c}_1 \bar{d}_1 c_2 d_2) \ne 0; \\ 1 & \text{if } k \ge 3, \text{ and } o(\bar{c}_1 \bar{d}_1 c_2) = o(c_k d_k \bar{c}_{k-1}); \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

Finally, define $v_2(\omega, \omega') = 0$ for all strings ω, ω', and for $k \ge 3$ define $v_k(\omega, \omega') = 0$ unless both ω and ω' are of length at least k, in which case

$$v_k(\omega, \omega') = u_k(c_1 c_2 \dots c_k, \bar{d}_k \bar{d}_{k-1} \dots \bar{d}_1).$$

(Note: The only reason for defining v_2 is to avoid having to write separate sums for the functions v_j and u_j in formula (4) and the arguments to follow.) Observe that both u_k and v_k depend only on the first k letters of their arguments. Furthermore, u_k and v_k are defined for arbitrary pairs of strings, finite or infinite; for *doubly* infinite sequences $\mathbf{x} = \ldots x_{-1}x_0x_1\ldots$ and $\mathbf{y} = \ldots y_{-1}y_0y_1\ldots$ we adopt the convention that

$$u_k(\mathbf{x}, \mathbf{y}) = u_k(x_1x_2\ldots x_k, y_1y_2\ldots y_k) \quad \text{and}$$
$$v_k(\mathbf{x}, \mathbf{y}) = v_k(x_1x_2\ldots x_k, , y_1y_2\ldots y_k).$$

Proposition 2.1 (Chas [5]) *Let α be a primitive necklace of length $n \geq 2$. Unhook α at an arbitrary location to obtain a string $\alpha^* = a_1a_2\ldots a_n$, and let $\sigma^j\alpha^*$ be the jth cyclic permutation of α^*. Then*

$$N(\alpha) = \sum_{i=1}^{n}\sum_{j=i+1}^{n}\sum_{k=2}^{n}(u_k(\sigma^i\alpha^*, \sigma^j\alpha^*) + v_k(\sigma^i\alpha^*, \sigma^j\alpha^*)). \tag{4}$$

3 Proof of the Main Theorem: strategy

Except for the exact values (1) of the limiting constants κ and σ^2, which of course depend on the specific form of the functions u_k and v_k, the conclusions of the Main Theorem hold more generally for random variables defined by sums of the form

$$N(\alpha^*) = \sum_{i=1}^{n}\sum_{j=i+1}^{n}\sum_{k=2}^{n}h_k(\sigma^i\alpha^*, \sigma^j\alpha^*) \tag{5}$$

where h_k are real-valued functions on the space of reduced sequences α^* with entries in $\mathcal{G}$ satisfying the hypotheses (H0)–(H3) below. The function N extends to *necklaces* in an obvious way: for any necklace α of length n, unhook α at an arbitrary place to obtain a joinable string α^*, then define $N(\alpha) = N(\alpha^*)$. Denote by λ_n, μ_n, and ν_n the uniform probability distributions on the sets $\mathcal{J}_n$, $\mathcal{F}_n$, and $\mathcal{S}_n$, respectively.

(H0) Each function h_k is symmetric.
(H1) There exists $C < \infty$ such that $|h_k| \leq C$ for all $k \geq 1$.
(H2) For each $k \geq 1$ the function h_k depends only on the first k entries of its arguments.
(H3) There exist constants $C < \infty$ and $0 < \beta < 1$ such that for all $n \geq k \geq 1$ and $1 \leq i < n$,

$$E_{\lambda_n}|h_k(\alpha, \sigma^i\alpha)| \leq C\beta^k$$

In view of (H2), the function h_k is well-defined for any pair of sequences, finite or infinite, provided their lengths are at least k. Hypotheses (H0)–(H2) are clearly satisfied for $h_k = u_k + v_k$, where u_k and v_k are as in formula (4) and $u_1 = v_1 = 0$; see Lemma 4.8 of Sect. 4.5 for hypothesis (H3).

Theorem 3.1 *Assume that the functions h_k satisfy hypotheses* (H0)–(H3), *and let $N(\alpha)$ be defined by* (5) *for all necklaces α of length n. There exist constants κ and σ^2 (given by* (22) *below) such that if F_n is the distribution of the random variable $(N(\alpha) - n^2\kappa)/n^{3/2}$ under the probability measure μ_n, then as $n \to \infty$,*

$$F_n \quad \Longrightarrow \quad \mathit{Normal}(0, \sigma^2). \tag{6}$$

Formulas for the limiting constants κ, σ are given (in more general form) in Theorem 5.1 below. In Sect. 6 we will show that in the case of particular interest, namely $h_k = u_k + v_k$ where u_k, v_k are as in Proposition 2.1, the constants κ and σ defined in Theorem 5.1 assume the values (1) given in the statement of the Main Theorem.

Modulo the proof of Lemma 4.8 and the calculation of the constants σ and κ, the Main Theorem follows directly from Theorem 3.1. The proof of Theorem 3.1 will proceed roughly as follows. First we will prove (Lemma 4.2) that there is a shift-invariant, Markov probability measure ν on the space $\mathcal{S}_\infty$ of infinite sequences $\mathbf{x} = x_1 x_2 \dots$ whose marginals (that is, the push-forwards under the projection mappings to $\mathcal{S}_n$) are the uniform distributions ν_n. Using this representation we will prove, in Sect. 4.4, that when n is large the distribution of $N(\alpha)$ under μ_n differs negligibly from the distribution of a related random variable defined on the Markov chain with distribution ν. See Proposition 4.7 for a precise statement. Theorem 3.1 will then follow from a general limit theorem for certain *U-statistics* of Markov chains (see Theorem 5.1).

4 The associated Markov chain

4.1 Necklaces, strings, and joinable strings

Recall that a *string* is a sequence with entries in the set $\mathcal{G}$ of generators and their inverses such that no two adjacent entries are inverses. A finite string is *joinable* if its first and last entries are not inverses. The sets of length-n strings, joinable strings, and necklaces are denoted by $\mathcal{S}_n$, $\mathcal{J}_n$, and $\mathcal{F}_n$, respectively, and the uniform distributions on these sets are denoted by ν_n, λ_n, and μ_n. Let A be the involutive permutation matrix with rows and columns indexed by $\mathcal{G}$ whose entries $a(x, y)$ are 1 if x and y are inverses and 0 otherwise. Let B be

the matrix with all entries 1. Then for any $n \geq 1$,

$$|\mathcal{S}_n| = \mathbf{1}^T (B - A)^{n-1} \mathbf{1} \quad \text{and} \quad |\mathcal{J}_n| = \text{trace}(B - A)^{n-1},$$

where $\mathbf{1}$ denotes the (column) vector all of where entries are 1. Similar formulas can be written for the number of strings (or joinable strings) with specified first and/or last entry. The matrix $B - A$ is a Perron–Frobenius matrix with lead eigenvalue $(g - 1)$; this eigenvalue is simple, so both $|\mathcal{S}_n|$ and $|\mathcal{J}_n|$ grow at the precise exponential rate $(g - 1)$, that is, there exist positive constants $C_{\mathcal{S}} = g/(g - 1)$ and $C_{\mathcal{J}}$ such that

$$|\mathcal{S}_n| \sim C_{\mathcal{S}}(g - 1)^n \quad \text{and} \quad |\mathcal{J}_n| \sim C_{\mathcal{J}}(g - 1)^n.$$

Every necklace of length n can be obtained by joining the ends of a joinable string, so there is a natural surjective mapping $p_n : \mathcal{J}_n \to \mathcal{F}_n$. This mapping is nearly n to 1: In particular, no necklace has more than n pre-images, and the only necklaces that do not have exactly n pre-images are those which are periodic with some period $d|n$ smaller than n. The number of these exceptional necklaces is vanishingly small compared to the total number of necklaces. To see this, observe that the total number of *strings* of length $n \geq 2$ is $g(g - 1)^{n-1}$; hence, the number of *joinable* strings is between $g(g - 1)^{n-2}$ and $g(g - 1)^{n-1}$. The number of length-n strings with period $< n$ is bounded above by

$$\sum_{d|n} g(g - 1)^{d-1} \leq \text{constant} \times (g - 1)^{n/2}.$$

This is of smaller exponential order of magnitude than $|\mathcal{J}_n|$, so for large n most necklaces of length n will have exactly n pre-images under the projection p_n. Consequently, as $n \to \infty$

$$|\mathcal{F}_n| \sim C_{\mathcal{J}}(g - 1)^n / n.$$

More important, this implies the following.

Lemma 4.1 *Let λ_n be the uniform probability distribution on the set $\mathcal{J}_n$, and let $\mu_n \circ p_n^{-1}$ be the push-forward to $\mathcal{J}_n$ of the uniform distribution on $\mathcal{F}_n$. Then*

$$\lim_{n \to \infty} \|\lambda_n - \mu_n \circ p_{n^{-1}}\|_{TV} = 0. \tag{7}$$

Here $\| \cdot \|_{TV}$ denotes the total variation norm on measures—see the Appendix. By Lemma B.3 of the Appendix, it follows that the distributions of the random variable $N(\alpha)$ under the probability measures λ_n and μ_n are asymptotically indistinguishable.

4.2 The associated Markov measure

The matrix $(B - A)$ has the convenient feature that its row sums and column sums are all $g - 1$. Therefore, the matrix $\mathbb{P} := (B - A)/(g - 1)$ is a stochastic matrix, with entries

$$p(a, b) = \begin{cases} \theta & \text{if } b \neq a^{-1}, \text{ and} \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

where

$$\theta = (g - 1)^{-1}. \tag{9}$$

In fact, $\mathbb{P}$ is *doubly stochastic*, that is, both its rows and columns sum to 1. Moreover, $\mathbb{P}$ is aperiodic and irreducible, that is, for some $k \geq 1$ (in this case $k = 2$) the entries of $\mathbb{P}^k$ are strictly positive. It is an elementary result of probability theory that for any aperiodic, irreducible, doubly stochastic matrix $\mathbb{P}$ on a finite set $\mathcal{G}$ there exists a shift-invariant probability measure ν on sequence space $\mathcal{S}_\infty$, called a *Markov measure*, whose value on the cylinder set $C(x_1 x_2 \dots x_n)$ consisting of all sequences whose first n entries are $x_1 x_2 \dots x_n$ is

$$\nu(C(x_1 x_2 \dots x_n)) = \frac{1}{|\mathcal{G}|} \prod_{i=1}^{n-1} p(x_i, x_{i+1}). \tag{10}$$

Any random sequence $\mathbf{X} = (X_1 X_2 \dots)$ valued in $\mathcal{S}_\infty$, defined on any probability space (Ω, P), whose distribution is ν is called a *stationary Markov chain with transition probability matrix* $\mathbb{P}$. In particular, the coordinate process on $(\mathcal{S}_\infty, \nu)$ is a Markov chain with t.p.m. $\mathbb{P}$.

Lemma 4.2 *Let* $\mathbf{X} = (X_1 X_2 \dots)$ *be a stationary Markov chain with transition probability matrix* $\mathbb{P}$ *defined by* (8). *Then for any* $n \geq 1$ *the distribution of the random string* $X_1 X_2 \dots X_n$ *is the uniform distribution* ν_n *on the set* $\mathcal{S}_n$.

Proof The transition probabilities $p(a, b)$ take only two values, 0 and θ, so for any n the nonzero cylinder probabilities (10) are all the same. Hence, the distribution of $X_1 X_2 \dots X_n$ is the uniform distribution on the set of all strings $\xi = x_1 x_2 \dots x_n$ such that the cylinder probability $\nu(C(\xi))$ is positive. These are precisely the *strings* of length n. □

4.3 Mixing properties of the Markov chain

Because the transition probability matrix $\mathbb{P}$ defined by (8) is aperiodic and irreducible, the m-step transition probabilities (the entries of the mth power $\mathbb{P}^m$ of $\mathbb{P}$) approach the stationary (uniform) distribution exponentially fast. The

one-step transition probabilities (8) are simple enough that precise bounds can be given:

Lemma 4.3 *The m-step transition probabilities* $p_m(a,b)$ *of the Markov chain with* 1*-step transition probabilities* (8) *satisfy*

$$\left| p_m(a,b) - \frac{1}{g} \right| \le \theta^m \tag{11}$$

where $\theta = 1/(g-1)$.

Proof Recall that $\mathbb{P} = \theta(B - A)$ where B is the matrix with all entries 1 and A is an involutive permutation matrix. Hence, $BA = AB = B$ and $B^2 = gB = ((\theta+1)/\theta)B$. This implies, by a routine induction argument, that for every integer $m \ge 1$,

$$(B-A)^m = \left(\frac{\theta^{-m}+1}{g}\right)B - A \quad \text{if } m \text{ is odd, and}$$

$$](B-A)^m = \left(\frac{\theta^{-m}-1}{g}\right)B + I \quad \text{if } m \text{ is even.}$$

The inequality (11) follows directly. □

The next lemma is a reformulation of the exponential convergence (11). Let $\mathbf{X} = (X_j)_{j\in\mathbb{Z}}$ be a stationary Markov chain with transition probabilities (8). For any finite subset $J \subset \mathbb{N}$, let X_J denote the restriction of $\mathbf{X}$ to the index set J, that is,

$$X_J = (X_j)_{j\in J};$$

for example, if J is the interval $[1,n] := \{1,2,\dots,n\}$ then X_J is just the random string $X_1X_2\dots X_n$. Denote by ν_J the distribution of X_J, viewed as a probability measure on the set $\mathcal{G}^J$; thus, for any subset $F \subset \mathcal{G}^J$,

$$\nu_J(F) = P\{X_J \in F\}. \tag{12}$$

If J, K are non-overlapping subsets of $\mathbb{N}$, then $\nu_{J\cup K}$ and $\nu_J \times \nu_K$ are both probability measures on $\mathcal{G}^{J\cup K}$, both with support set equal to the set of all restrictions of infinite strings.

Lemma 4.4 *Let* $J, K \subset \mathbb{N}$ *be two finite subsets such that* $\max(J) + m \le \min(K)$ *for some* $m \ge 1$. *Then on the support of the measure* $\nu_J \times \nu_K$,

$$1 - g\theta^m \le \frac{d\nu_{J\cup K}}{d\nu_J \times \nu_K} \le 1 + g\theta^m \tag{13}$$

where $\theta = 1/(g-1)$ and $d\alpha/d\beta$ denotes the Radon–Nikodym derivative ("likelihood ratio") of the probability measure α and β.

Proof It suffices to consider the special case where J and K are intervals, because the general case can be deduced by summing over excluded variables. Furthermore, because the Markov chain is stationary, the measures ν_J are invariant by translations (that is, $\nu_{J+n} = \nu_J$ for any $n \geq 1$), so we may assume that $J = [1, n]$ and $K = [n+m, n+q]$. Let $x_{J\cup K}$ be the restriction of some infinite string to $J \cup K$; then

$$\nu_J \times \nu_K(x_{J\cup K})\pi(x_1)\left(\prod_{j=1}^{n-1} p(x_j, x_{j+1})\right)\pi(x_{n+m})\left(\prod_{j=n+m}^{m+q-1} p(x_j, x_{j+1})\right)$$

and

$$\begin{aligned}&\nu_{J\cup K}(x_{J\cup K})\\ &\quad = \pi(x_1)\left(\prod_{j=1}^{n-1} p(x_j, x_{j+1})\right)p_m(x_n, x_{n+m})\left(\prod_{j=n+m}^{m+q-1} p(x_j, x_{j+1})\right).\end{aligned}$$

The result now follows directly from the double inequality (11), as this implies that for any two letters a, b,

$$\left|\frac{p_m(a,b)}{\pi(b)} - 1\right| \leq g\theta^m. \qquad \square$$

4.4 From random joinable strings to random strings

Since $\mathcal{J}_n \subset \mathcal{S}_n$, the uniform distribution λ_n on $\mathcal{J}_n$ is gotten by restricting the uniform distribution ν_n on $\mathcal{S}_n$ to $\mathcal{J}_n$ and then renormalizing:

$$\lambda_n(F) = \frac{\nu_n(F \cap \mathcal{J}_n)}{\nu_n(\mathcal{J}_n)}.$$

Equivalently, the distribution of a random *joinable* string is the conditional distribution of a random string given that its first and last entries are not inverses. Our goal here is to show that the distributions of the random variable $N(\alpha)$ defined by (5) under the probability measures λ_n and ν_n differ negligibly when n is large. For this we will show first that the distributions under λ_n and ν_n, respectively, of the *substring* gotten by deleting the last $n^{1/2-\varepsilon}$ letters are close in total variation distance; then we will show that changing the last $n^{1/2-\varepsilon}$ letters has only a small effect on the value of $N(\alpha)$.

Lemma 4.5 *Let $X_1X_2\dots X_n$ be a random string of length n, and $Y_1Y_2\dots Y_n$ a random* joinable *string. For any integer $m \in [1, n-1]$ let $\nu_{n,m}$ and $\lambda_{n,m}$ denote the distributions of the random substrings $X_1X_2\dots X_{n-m}$ and $Y_1Y_2\dots Y_{n-m}$. Then the measure $\lambda_{n,m}$ is absolutely continuous with respect to $\nu_{n,m}$, and the Radon–Nikodym derivative satisfies*

$$\frac{1-g\theta^m}{1+g\theta^m} \le \frac{d\lambda_{n,m}}{d\nu_{n,m}} \le \frac{1+g\theta^m}{1-g\theta^m} \tag{14}$$

where $\theta = 1/(g-1)$. Consequently, the total variation distance between the two measures satisfies

$$\|\nu_{n,m} - \lambda_{n,m}\|_{TV} \le 2\left(\frac{1+g\theta^m}{1-g\theta^m} - 1\right). \tag{15}$$

Proof The cases $m = 0$ and $m = 1$ are trivial, because in these cases the lower bound is non-positive and the upper bound is at least 2. The general case $m \ge 2$ follows from the exponential ergodicity estimates (11) by an argument much like that used to prove Lemma 4.4. For any string $x_1x_2\dots x_{n-m}$ with initial letter $x_1 = a$,

$$\nu_{n,m}(x_1x_2\dots x_{n-m}) = \frac{1}{g}\prod_{i=1}^{n-m-1} p(x_i, x_{i+1}).$$

Similarly, by Lemma 4.2,

$$\lambda_{n,m}(x_1x_2\dots x_{n-m}) = \frac{1}{g}\left(\prod_{i=1}^{n-m-1} p(x_i, x_{i+1})\right)\frac{\sum_{b\ne x_1^{-1}} p_m(x_{n-m}, b)}{g^{-1}\sum_a \sum_{b\ne a^{-1}} p_n(a,b)}.$$

Inequality (11) implies that the last fraction in this expression is between the bounding fractions in (14). The bound on the total variation distance between the two measures follows routinely. □

Corollary 4.6 *Let $\mathbf{X}$ be a stationary Markov chain with transition probability matrix $\mathbb{P}$. Assume that the functions h_k satisfy hypotheses* (H0)–(H3) *of Sect. 3. Then for all $k, i \ge 1$,*

$$E|h_k(\mathbf{X}, \tau^i\mathbf{X})| \le C\beta^k. \tag{16}$$

Proof The function $h_k(\mathbf{x}, \tau^i\mathbf{x})$ is a function only of the coordinates $x_1x_2\dots x_{i+k}$, and so for any joinable string $\mathbf{X}$ of length $> i+k$,

$$h_k(\mathbf{x}, \tau^i\mathbf{x}) = h_k(\mathbf{x}, \sigma^i\mathbf{x}).$$

By Lemma 4.5, the difference in total variation norm between the distributions of the substring $x_1x_2\dots x_{i+k}$ under the measures λ_n and ν_n converges to 0 as $n\to\infty$. Therefore,

$$E|h_k(\mathbf{X},\tau^i\mathbf{X})| = \lim_{n\to\infty} E_{\lambda_n}|h_k(\alpha,\sigma^i\alpha)| \le C\beta^k. \qquad \square$$

Now we are in a position to compare the distribution of the random variable $N(\alpha)$ under μ_n with the distribution of a corresponding random variable N_n^S on the sequence space $\mathcal{S}_\infty$ under the measure ν. (Recall that the *distribution* of a random variable Z defined on a probability space $(\Omega,\mathcal{F},P)$ is the pushforward measure $P\circ Z^{-1}$. See the Appendix for a resume of common terminology from the theory of probability and basic results concerning convergence in distribution.) The function N_n^S is defined by

$$N_n^S(\mathbf{x}) = \sum_{i=1}^{n}\sum_{j=i+1}^{n}\sum_{k=1}^{\infty} h_k(\tau^i\mathbf{x},\tau^j\mathbf{x}). \tag{17}$$

Proposition 4.7 *Assume that the functions h_k satisfy hypotheses* (H0)–(H3), *and let $\kappa = \sum_{k=1}^{\infty} E\,H_k(\mathbf{X})$. Let F_n be the distribution of the random variable $(N(\alpha)-n^2\kappa)/n^{3/2}$ under the uniform probability measure μ_n on $\mathcal{F}_n$, and G_n the distribution of $(N_n^S(\mathbf{x})-\kappa n^2)/n^{3/2}$ under ν. Then for any metric ϱ that induces the topology of weak convergence on probability measures,*

$$\lim_{n\to\infty}\varrho(F_n,G_n) = 0. \tag{18}$$

Consequently, $F_n\Rightarrow\Phi_\sigma$ if and only if $G_n\Rightarrow\Phi_\sigma$.

Proof Let F_n' be the distribution of the random variable $(N(\alpha)-n^2\kappa)/n^{3/2}$ under the uniform probability measure λ_n on $\mathcal{J}_n$. By Lemma 4.1, the total variation distance between λ_n and $\mu_n\circ p_n^{-1}$ is vanishingly small for large n. Hence, by Lemma B.3 and the fact that total variable distance is never increased by mapping (cf. inequality (47) of the Appendix),

$$\lim_{n\to\infty}\varrho(F_n,F_n') = 0.$$

Therefore, it suffices to prove (18) with F_n replaced by F_n'.

Partition the sums (5) and (17) as follows. Fix $0<\delta<1/2$ and set $m=m(n)=[n^\delta]$. By hypothesis (H3) and Corollary (4.6),

$$E_\mu\sum_{i=1}^{n}\sum_{j=i+1}^{n}\sum_{k>m(n)}|h_k(\tau^i\mathbf{x},\tau^j\mathbf{x})| \le Cn^2\beta^{m(n)} \quad \text{and}$$

$$E_{\lambda_n} \sum_{i=1}^{n} \sum_{j=i+1}^{n} \sum_{k>m(n)} |h_k(\sigma^i\alpha, \sigma^j\alpha)| \le Cn^2\beta^{m(n)}.$$

These upper bounds are rapidly decreasing in n. Hence, by Markov's inequality (i.e., the crude bound $P\{|Y| > \varepsilon\} \le E|Y|/\varepsilon$), the distributions of both of the sums converge weakly to 0 as $n \to \infty$. Thus, by Lemma B.3, to prove the proposition it suffices to prove that

$$\lim_{n\to\infty} \varrho(F_n^A, G_n^A) = 0$$

where F_n^A and G_n^A are the distributions of the truncated sums obtained by replacing the inner sums in (5) and (17) by the sums over $1 \le k \le m(n)$.

The outer sums in (5) and (17) are over pairs of indices $1 \le i < j \le n$. Consider those pairs for which $j > n - 2m(n)$: there are only $2nm(n)$ of these. Since $nm(n) = O(n^{1+\delta})$ and $\delta < 1/2$, and since each term in (5) and (17) is bounded in absolute value by a constant C (by Hypothesis (H1)), the sum over those index pairs $i < j$ with $n - 2m(n) < j \le n$ is $o(n^{3/2})$. Hence, by Lemma B.3, it suffices to prove that

$$\lim_{n\to\infty} \varrho(F_n^B, G_n^B) = 0$$

where F_n^B and G_n^B are the distributions under λ_n and ν of the sums (5) and (17) with the limits of summation changed to $i < j < n - 2m(n)$ and $k \le m(n)$. Now if $i < j < n - 2m(n)$ and $k \le m(n)$ then $h_k(\tau^i\mathbf{x}, \tau^j\mathbf{x})$ and $h_k(\sigma^i\alpha, \sigma^j\alpha)$ depend only on the first $n - n(m)$ entries of $\mathbf{x}$ and α. Consequently, the distributions F_n^B and G_n^B are the distributions of the sums

$$\sum_{i=1}^{n-2m(n)} \sum_{j=i+1}^{n-2m(n)} \sum_{k\le m(n)} h_k(\tau^i\mathbf{x}, \tau^j\mathbf{x})$$

under the probability measures $\lambda_{n,m}$ and $\nu_{n,m}$, respectively, where $\lambda_{n,m}$ and $\nu_{n,m}$ are as defined in Lemma 4.5. But the total variation distance between $\lambda_{n,m}$ and $\nu_{n,m}$ converges to zero, by Lemma 4.5. Therefore, by the mapping principle (47) and Lemma B.3,

$$\varrho(F_n^B, G_n^B) \longrightarrow 0. \qquad \square$$

4.5 Mean estimates

In this section we show that the hypothesis (H3) is satisfied by the functions $h_k = u_k + v_k$, where u_k and v_k are as in Proposition 2.1.

Lemma 4.8 *Let σ^i be the ith cyclic shift on the set $\mathcal{J}_n$ of joinable sequences α. There exists $C < \infty$ such that for all $2 \le k \le n$ and $0 \le i < j < n$,*

$$\begin{aligned} E_{\lambda_n} u_k(\sigma^i\beta, \sigma^j\beta) &\le C\theta^{k/2} \quad \text{and} \\ E_{\lambda_n} v_k(\sigma^i\beta, \sigma^j\beta) &\le C\theta^{k/2}. \end{aligned} \tag{19}$$

Proof Because the measure λ_n is invariant under both cyclic shifts and the reversal function, it suffices to prove the estimates only for the case where one of the indices i, j is 0. If the proper choice is made ($i = 0$ and $j \le n/2$), then a necessary condition for $u_k(\alpha, \sigma^j\alpha) \ne 0$ is that the strings α and $\sigma^j\alpha$ agree in their second through their $(k-1)/2$th slots. By routine counting arguments (as in Sect. 4.1) it can be shown that the number of joinable strings of length n with this property is bounded above by $C(g-1)^{n-k/2}$, where $C < \infty$ is a constant independent of both n and $k \le n$. This proves the first inequality. A similar argument proves the second. □

5 U-statistics of Markov chains

Proposition 4.7 implies that for large n the distribution F_n considered in Theorem 3.1 is close in the weak topology to the distribution G_n of the random variable N_n^S defined by (17) under the Markov measure ν. Consequently, if it can be shown that $G_n \Rightarrow \Phi_\sigma$ then the conclusion $F_n \Longrightarrow \Phi_\sigma$ will follow, by Lemma B.3 of the Appendix. This will prove Theorem 3.1.

Random variables of the form (17) are known generically in probability theory as *U-statistics* (see [10]). Second order U-statistics of Markov chains are defined as follows. Let $\mathbf{Z} = Z_1 Z_2 \ldots$ be a stationary, aperiodic, irreducible Markov chain on a finite state space $\mathcal{A}$ with transition probability matrix $\mathbb{Q}$ and stationary distribution π. Let τ be the forward shift on the sequence space $\mathcal{A}^{\mathbb{N}}$. The U-statistics of order 2 with kernel $h : \mathcal{A}^{\mathbb{N}} \times \mathcal{A}^{\mathbb{N}} \to \mathbb{R}$ are the random variables

$$W_n = \sum_{i=1}^{n} \sum_{j=i+1}^{n} h(\tau^i\mathbf{Z}, \tau^j\mathbf{Z}).$$

The Hoeffding projection of a kernel h is the function $H : \mathcal{A}^{\mathbb{N}} \to \mathbb{R}$ defined by

$$H(\mathbf{z}) = Eh(\mathbf{z}, \mathbf{Z}).$$

Theorem 5.1 *Suppose that $h = \sum_{k=1}^{\infty} h_k$ where $\{h_k\}_{k\ge1}$ is a sequence of kernels satisfying hypotheses* (H0)–(H2) *and the following*: *There exist constants $C < \infty$ and $0 < \beta < 1$ such that for all $k, i \ge 1$,*

$$E|h_k(\mathbf{Z}, \tau^i\mathbf{Z})| \le C\beta^k. \tag{20}$$

Then as $n \to \infty$,

$$\frac{W_n - n^2\kappa}{n^{3/2}} \quad \Longrightarrow \quad \Phi_\sigma \tag{21}$$

where the constants κ *and* σ^2 *are*

$$\kappa = \sum_{k=1}^{\infty} E\,H_k(\mathbf{Z}) \quad \textit{and} \quad \sigma^2 = \lim_{n\to\infty} \frac{1}{n} E\left(\sum_{i=1}^{n}\sum_{k=1}^{\infty} H_k(\tau^i \mathbf{Z}) - n\kappa\right)^2. \tag{22}$$

There are similar theorems in the literature, but all require some degree of additional continuity of the kernel h. In the special case where all but finitely many of the functions h_k are identically 0 the result is a special case of Theorem 1 of [9] or Theorem 2 of [11]. If the functions h_k satisfy the stronger hypothesis that $|h_k| \le C\beta^k$ pointwise then the result follows (with some work) from Theorem 2 of [11]. Unfortunately, the special case of interest to us, where $h_k = u_k + v_k$ and u_k, v_k are the functions defined in Sect. 2, does not satisfy this hypothesis.

The rest of Sect. 5 is devoted to the proof. The main step is to reduce the problem to the special case where all but finitely many of the functions h_k are identically 0 by approximation; this is where the hypothesis (20) will be used. The special case, as already noted, can be deduced from the results of [9] or [11], but instead we shall give a short and elementary argument.

In proving Theorem 5.1 we can assume that all of the Hoeffding projections H_k have mean

$$E\,H_k(\mathbf{Z}) = 0,$$

because subtracting a constant from both h and κ has no effect on the validity of the theorem. Note that this does *not* imply that $Eh_k(\tau^i\mathbf{Z}, \tau^j\mathbf{Z}) = 0$, but it does imply (by Fubini's theorem) that if $\mathbf{Z}$ and $\mathbf{Z}'$ are independent copies of the Markov chain then

$$Eh_k(\mathbf{Z}, \mathbf{Z}') = 0.$$

5.1 Proof in the special case

If all but finitely many of the functions h_k are 0 then for some finite value of K the kernel h depends only on the first K entries of its arguments.

Lemma 5.2 *Without loss of generality, we can assume that* $K = 1$.

Proof If $Z_1 Z_2 \ldots$ is a stationary Markov chain, then so is the sequence $Z_1^K Z_2^K \ldots$ where

$$Z_i^K = Z_i Z_{i+1} \ldots Z_{i+K}$$

is the length-$(K+1)$ word obtained by concatenating the $K+1$ states of the original Markov chain following Z_i. Hence, the U-statistics W_n can be represented as U-statistics on a different Markov chain with kernel depending only on the first entries of its arguments. It is routine to check that the constants κ and σ^2 defined by (22) for the chain Z_n^K equal those defined by (22) for the original chain. □

Assume now that h depends only on the first entries of its arguments. Then the Hoeffding projection H also depends only on the first entry of its argument, and can be written as

$$H(z) = Eh(z, Z_1) = \sum_{z' \in \mathcal{A}} h(z, z')\pi(z').$$

Since the Markov chain Z_n is stationary and ergodic, the covariances $EH(Z_i)H(Z_{i+n}) = EH(Z_1)H(Z_{1+n})$ decay exponentially in n, so the limit

$$\sigma^2 := \lim_{n\to\infty} \frac{1}{n} E\left(\sum_{j=1}^{n} H(Z_j)\right)^2 \tag{23}$$

exists and is nonnegative. It is an elementary fact that $\sigma^2 > 0$ unless $H \equiv 0$. Say that the kernel h is *centered* if this is the case. If h is not centered then the adjusted kernel

$$h^*(z, z') := h(z, z') - H(z) - H(z') \tag{24}$$

is centered, because its Hoeffding projection satisfies

$$\begin{aligned} H^*(z) :&= Eh^*(z, Z_1) \\ &= Eh(z, Z_1) - EH(z) - EH(Z_1) \\ &= H(z) - H(z) - 0. \end{aligned}$$

Define

$$T_n = \sum_{i=1}^{n}\sum_{j=1}^{n} h(Z_i, Z_j) \quad \text{and} \quad D_n = \sum_{i=1}^{n} h(Z_i, Z_i);$$

then since the kernel h is symmetric,

$$W_n = \frac{1}{2}(T_n - D_n). \tag{25}$$

Proposition 5.3 *If h is centered, then*

$$T_n/n \quad \Longrightarrow \quad Q \tag{26}$$

where Q is a quadratic form in no more than $m = |\mathcal{A}|$ independent, standard normal random variables.

Proof Consider the linear operator L_h on $\ell^2(\mathcal{A}, \pi)$ defined by

$$L_h f(z) := \sum_{z' \in \mathcal{A}} h(z, z') f(z') \pi(z').$$

This operator is symmetric (real Hermitean), and consequently has a complete set of orthonormal real eigenvectors $\varphi_j(z)$ with real eigenvalues λ_j. Since h is centered, the constant function $\varphi_1 := 1/\sqrt{m}$ is an eigenvector with eigenvalue $\lambda_1 = 0$; therefore, all of the other eigenvectors φ_j, being orthogonal to φ_1, must have mean zero. Hence, since $\lambda_1 = 0$,

$$h(z, z') = \sum_{j=2}^{m} \lambda_j \varphi_j(z) \varphi_j(z'),$$

and so

$$\begin{aligned} T_n &= \sum_{k=2}^{m} \lambda_k \sum_{i=1}^{n} \sum_{j=1}^{n} \varphi_k(Z_i) \varphi_k(Z_j) \\ &= \sum_{k=2}^{m} \lambda_k \left(\sum_{i=1}^{n} \varphi_k(Z_i) \right)^2 . \end{aligned} \tag{27}$$

Since each φ_k has mean zero and variance 1 relative to π, the central limit theorem for Markov chains implies that as $n \to \infty$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi_k(Z_i) \quad \Longrightarrow \quad \mathit{Normal}(0, \sigma_k^2), \tag{28}$$

with limiting variances $\sigma_k^2 \geq 0$. In fact, these normalized sums converge *jointly*[1] (for $k = 2, 3, \ldots, m$) to a multivariate normal distribution with

[1] Note, however, that the normalized sums in (28) need not be asymptotically independent for different k, despite the fact that the different functions φ_k are uncorrelated relative to π. This is because the arguments Z_i are serially correlated: in particular, even though $\varphi_k(Z_i)$ and $\varphi_l(Z_i)$ are uncorrelated, the random variables $\varphi_k(Z_i)$ and $\varphi_l(Z_{i+1})$ might well be correlated.

marginal variances $\sigma_k^2 \geq 0$. The result therefore follows from the spectral representation (27). □

Corollary 5.4 *If h is not centered, then with $\sigma^2 > 0$ as defined in* (23),

$$W_n/n^{3/2} \implies \mathit{Normal}(0, \sigma^2). \tag{29}$$

Proof Recall that $W_n = (T_n - D_n)/2$. By the ergodic theorem, $\lim_{n\to\infty} D_n/n = Eh(Z_1, Z_1)$ almost surely, so $D_n/n^{3/2} \Rightarrow 0$. Hence, by Lemma B.3 of the Appendix, it suffices to prove that if h is not centered then

$$T_n/n^{3/2} \implies \mathit{Normal}(0, 4\sigma^2). \tag{30}$$

Define the centered kernel h^* as in (24). Since the Hoeffding projection of H^* is identically 0,

$$T_n = T_n^* + 2n\sum_{i=1}^{n} H(Z_i) \quad \text{where}$$

$$T_n^* = \sum_{i=1}^{n}\sum_{j=1}^{n} h^*(Z_i, Z_j).$$

Proposition 5.3 implies that T_n^*/n converges in distribution, and it follows that $T_n^*/n^{3/2}$ converges to 0 in distribution. On the other hand, the central limit theorem for Markov chains implies that

$$n^{-3/2}\left(2n\sum_{i=1}^{n} H(Z_i)\right) \implies \mathit{Normal}(0, 4\sigma^2),$$

with $\sigma^2 > 0$, since by hypothesis the kernel h is not centered. The weak convergence (30) now follows by Lemma B.3. □

5.2 Variance/covariance bounds

To prove Theorem 5.1 in the general case we will show that truncation of the kernel h, that is, replacing $h = \sum_{k=1}^{\infty} h_k$ by $h^K = \sum_{k=1}^{K} h_k$, has only a small effect on the distributions of the normalized random variables $W_n/n^{3/2}$ when K is large. For this we will use second moment bounds. To deduce these from the first-moment hypothesis (20) we shall appeal to the fact that any aperiodic, irreducible, finite-state Markov chain is exponentially mixing. Exponential mixing is expressed in the same manner as for the Markov chain considered in Sect. 4.3. For any finite subset $J \subset \mathbb{N}$, let $Z_J = (Z_j)_{j\in J}$ denote

the restriction of $\mathbf{Z}$ to the index set J, and denote by μ_J the distribution of Z_J. If I, J are nonoverlapping subsets of $\mathbb{N}$ then both $\mu_{I\cup J}$ and $\mu_I \times \mu_J$ are probability measures supported by $\mathcal{A}^{I\cup J}$. If the distance between the sets I and J is at least m_*, where m_* is the smallest integer such that all entries of $\mathbb{Q}^{m_*}$ are positive, then $\mu_{I\cup J}$ and $\mu_I \times \mu_J$ are mutually absolutely continuous.

Lemma 5.5 *There exist constants $C < \infty$ and $0 < \delta < 1$ such that for any two subsets $I, J \subset \mathbb{N}$ satisfying $\min(J) - \max(I) = m \geq m_*$,*

$$1 - C\delta^m \leq \frac{d\mu_{I\cup J}}{d\mu_I \times \mu_J} \leq 1 + C\delta^m.$$

The constant C need not be the same as the constant in the hypothesis (20); however, the exact values of these constants are irrelevant to our purposes, and so we shall minimize notational clutter by using the letter C generically for such constants. The proof of the lemma is nearly identical to that of Lemma 4.4, except that the exponential convergence bounds of Lemma 4.3 must be replaced by corresponding bounds for the transition probabilities of $\mathbf{Z}$. The corresponding bounds are gotten from the Perron–Frobenius theorem.

For any two random variables U, V denote by $\operatorname{cov}(U, V) = E(UV) - EUEV$ their covariance. (When $U = V$ the covariance $\operatorname{cov}(U, V) = \operatorname{Var}(U)$.)

Lemma 5.6 *For any two pairs $i < j$ and $i' < j'$ of indices, let $\Delta = \Delta(i, i', j, j')$ be the distance between the sets $\{i, j\}$ and $\{i', j'\}$ (that is, the minimum distance between one of i, j and one of i', j'). Then for suitable constants $0 < C, C' < \infty$, for all $\Delta \geq \max(k, k') + m_*$,*

$$|\operatorname{cov}(h_k(\tau^i\mathbf{Z}, \tau^j\mathbf{Z}), h_{k'}(\tau^{i'}\mathbf{Z}, \tau^{j'}\mathbf{Z}))| \leq C'\beta^{k+k'-4}\varrho_{\Delta-\max(k,k')} \tag{31}$$

where

$$\varrho_m = (1 + C\delta^m)^5 \quad \text{for } m \geq m_*$$

and β is the exponential decay rate in (20).

Remark 5.7 What is important is that the covariances decay exponentially in both $k + k'$ and Δ; the rates will not matter. When $\Delta \leq \max(k, k') + m_*$ the bounds (31) do not apply. However, in this case, since the functions h_k are bounded above in absolute value by a constant $C < \infty$ independent of k (hypothesis (H1)), the Cauchy–Schwartz inequality implies

$$\begin{aligned}&|\operatorname{cov}(h_k(\tau^i\mathbf{Z}, \tau^j\mathbf{Z}), h_{k'}(\tau^{i'}\mathbf{Z}, \tau^{j'}\mathbf{Z}))|^2\\ &\quad = (Eh_k(\tau^i\mathbf{Z}, \tau^j\mathbf{Z})h_{k'}(\tau^{i'}\mathbf{Z}, \tau^{j'}\mathbf{Z}))^2\end{aligned}$$

$$\begin{aligned}
&\le (Eh_k(\tau^i\mathbf{Z}, \tau^j\mathbf{Z})Eh_{k'}(\tau^{i'}\mathbf{Z}, \tau^{j'}\mathbf{Z}))^2\\
&\le C^2 Eh_k(\tau^i\mathbf{Z}, \tau^j\mathbf{Z})Eh_{k'}(\tau^{i'}\mathbf{Z}, \tau^{j'}\mathbf{Z})\\
&\le C_*\beta^{k+k'},
\end{aligned}$$

the last by the first moment hypothesis (20).

Proof of Lemma 5.6 Since the random variables h_k are functions only of the first k letters of their arguments, the covariances can be calculated by averaging against the measures $\mu_{J\cup K}$, where

$$J = [i, i+k] \cup [j, j+k] \quad \text{and} \quad K = [i', i'+k'] \cup [j', j+k'].$$

The simplest case is where $j + k < i'$; in this case the result of Lemma 5.5 applies directly, because the sets J and K are separated by $m = \Delta - k$. Since the functions h_k are uniformly bounded, Lemma 5.5 implies

$$1 - C\delta^m \le \frac{Eh_k(\tau^i\mathbf{Z}, \tau^j\mathbf{Z})h_{k'}(\tau^{i'}\mathbf{Z}, \tau^{j'}\mathbf{Z})}{Eh_k(\tau^i\mathbf{Z}, \tau^j\mathbf{Z})Eh_{k'}(\tau^{i'}\mathbf{Z}, \tau^{j'}\mathbf{Z})} \le 1 + C\delta^m.$$

The inequalities in (31) now follow, by the assumption (20). (In this special case the bounds obtained are tighter that those in (31).)

The other cases are similar, but the exponential ergodicity estimate (13) must be used indirectly, since the index sets J and K need not be ordered as required by Lemma 4.4. Consider, for definiteness, the case where

$$i + k \le i' \le i' + k' \le j \le j + k \le j' \le j' + k'.$$

To bound the relevant likelihood ratio in this case, use the factorization

$$\begin{aligned}
\frac{d\mu_{J\cup K}}{d\mu_J \times \mu_K} = &\frac{d\mu_{J\cup K}}{d\mu_{J^-\cup K^-} \times \mu_{J^+\cup K^+}} \times \frac{d\mu_{J^-\cup K^-} \times \mu_{J^+\cup K^+}}{d\mu_{J^-} \times \mu_{K^-} \times \mu_{J^+} \times \mu_{K^+}}\\
&\times \frac{d\mu_{J^-} \times \mu_{K^-} \times \mu_{J^+} \times \mu_{K^+}}{d\mu_J \times \mu_K}
\end{aligned}$$

where $J^- = [i, i+k]$, $J^+ = [j, j+k]$, $K^- = [i', i'+k']$, and $K^+ = [j', j'+k']$. For the second and third factors, use the fact that Radon–Nikodym derivatives of product measures factor, e.g.,

$$\begin{aligned}
&\frac{d\mu_{J^-} \times \mu_{K^-} \times \mu_{J^+} \times \mu_{K^+}}{d\mu_J \times \mu_K}(x_J, x_K)\\
&= \frac{d\mu_{J^-} \times \mu_{J^+}}{d\mu_J}(x_J) \times \frac{d\mu_{K^-} \times \mu_{K^+}}{d\mu_K}(x_K)
\end{aligned}$$

Now Lemma 5.5 can be used to bound each of the resulting five factors. This yields the following inequalities:

$$(1 - C\delta^m)^5 \le \frac{d\mu_{J\cup K}}{d\mu_J \times \mu_K} \le (1 + C\delta^m)^5,$$

and so by the same reasoning as used earlier,

$$(1 - C\delta^m)^5 \le \frac{Eh_k(\tau^i\mathbf{Z}, \tau^j\mathbf{Z})h_{k'}(\tau^{i'}\mathbf{Z}, \tau^{j'}\mathbf{Z})}{Eh_k(\tau^i\mathbf{Z}, \tau^j\mathbf{Z})Eh_{k'}(\tau^{i'}\mathbf{Z}, \tau^{j'}\mathbf{Z})} \le (1 + C\delta^m)^5.$$

The remaining cases can be handled in the same manner. □

Corollary 5.8 *There exist $C, C' < \infty$ such that for all $n \ge 1$ and all $1 \le K \le L \le \infty$,*

$$\begin{aligned}&\operatorname{Var}\left(\sum_{i=1}^{n}\sum_{j=i+1}^{n}\sum_{k=K}^{L} h_k(\tau^i\mathbf{X}, \tau^j\mathbf{Z})\right)\\ &\quad\le Cn^3\sum_{k=K}^{\infty}\sum_{k'=K}^{\infty}\left\{(k'+k+C')\beta^{k+k'}\right\}. \end{aligned} \tag{32}$$

Consequently, for any $\varepsilon > 0$ there exists $K < \infty$ such that for all $n \ge 1$,

$$\operatorname{Var}\left(\sum_{i=1}^{n}\sum_{j=i+1}^{n}\sum_{k=K+1}^{\infty} h_k(\tau^i\mathbf{Z}, \tau^j\mathbf{Z})\right) \le \varepsilon n^3. \tag{33}$$

Proof The variance is gotten by summing the covariances of all possible pairs of terms in the sum. Group these by size, according to the value of $\Delta(i, i', j, j')$: for any given value of $\Delta \ge 2$, the number of quadruples i, i', j, j' in the range $[1, n]$ with $\Delta(i, i', j, j') = \Delta$ is no greater than $24n^3$. For each such quadruple and any pair k, k' such that $K < k \le k'$ Lemma 5.6 implies that if $\Delta \ge k + m_*$ then

$$|\operatorname{cov}(h_k(\tau^i\mathbf{Z}, \tau^j\mathbf{Z}), h_{k'}(\tau^{i'}\mathbf{Z}, \tau^{j'}\mathbf{Z}))| \le C\beta^{k+k'}\varrho_{\Delta-k'}.$$

If $\Delta \le m_* + k'$ then the crude Cauchy–Schwartz bounds of Remark 5.7 imply that

$$|\operatorname{cov}(h_k(\tau^i\mathbf{Z}, \tau^j\mathbf{Z}), h_{k'}(\tau^{i'}\mathbf{Z}, \tau^{j'}\mathbf{Z}))| \le C\beta^{k+k'}$$

where $C < \infty$ is a constant independent of i, i', j, j', k, k'. Summing these bounds we find that the variance on the left side of (32) is bounded by

$$Cn^3 \sum_{k=K}^{L} \sum_{k=K}^{L} \beta^{k+k'} \left(m_* + k + k' + \sum_{\Delta=m_*}^{\infty} \varrho_\Delta \right).$$

Since ϱ_j is exponentially decaying in j, the inner sum is finite. This proves the inequality (32). The second assertion now follows. □

5.3 Proof of Theorem 5.1

Given Corollary 5.8—in particular, the assertion (33)—Theorem 5.1 follows from the special case where all but finitely many of the functions h_k are identically zero, by Lemma B.4 of the Appendix. To see this, observe that under the hypotheses of Theorem 5.1, the random variable W_n can be partitioned as

$$W_n = W_n^K + R_n^K$$

where

$$W_n^K = \sum_{i=1}^{n} \sum_{j=i+1}^{n} \sum_{k=1}^{K} h_k(\tau^i \mathbf{Z}, \tau^j \mathbf{Z}) \quad \text{and}$$

$$R_n^K = \sum_{i=1}^{n} \sum_{j=i+1}^{n} \sum_{k=K+1}^{\infty} h_k(\tau^i \mathbf{Z}, \tau^j \mathbf{Z}).$$

By Proposition 5.3 and Corollary 5.4, for any finite K the sequence $W_n^K/n^{3/2}$ converges to a normal distribution with mean 0 and finite (but possibly zero) variance σ_K^2. By (33), for any $\varepsilon > 0$ there exists $K < \infty$ such that $E|R_n^K|^2/n^3 < \varepsilon$ for all $n \geq 1$. Consequently, by Lemma B.4, $\sigma^2 = \lim_{K\to\infty} \sigma_K^2$ exists and is finite, and

$$W_n/n^{3/2} \quad \Longrightarrow \quad \mathit{Normal}(0, \sigma^2).$$

6 Mean/variance calculations

In this section we verify that in the special case $h_k = u_k + v_k$, where u_k and v_k are the functions defined in Sect. 2 and $h_1 \equiv 0$, the constants κ and σ^2 defined by (22) coincide with the values (1).

Assume throughout this section that $\mathbf{X} = X_1 X_2 \ldots$ and $\mathbf{X}' = X_1' X_2' \ldots$ are two *independent* stationary Markov chains with transition probabilities (8),

both defined on a probability space (Ω, P) with corresponding expectation operator E. Set $h_k = u_k + v_k$. For each fixed (nonrandom) string $x_1x_2\ldots$ of length $\geq k$ define

$$H_k = U_k + V_k \quad \text{where} \tag{34}$$

$$\begin{aligned} U_k(x_1x_2\ldots) &= Eu_k(x_1x_2\ldots x_k, X_1'X_2'\ldots) \quad \text{and} \\ V_k(x_1x_2\ldots) &= Ev_k(x_1x_2\ldots x_k, X_1'X_2'\ldots), \end{aligned} \tag{35}$$

and set

$$S_k(x_1x_2\ldots) = U_2(x_1x_2) + \sum_{l=3}^{k}(U_l + V_l)(x_1x_2\ldots x_l). \tag{36}$$

Since the summands are all nonnegative and satisfy hypotheses (H0)–(H3), the last sum is well-defined and finite even for $k = \infty$. The restrictions of U_k and V_k to the space of infinite sequences are the *Hoeffding projections* of the functions u_k and v_k (see Sect. 3). Note that each of the functions U_k, V_k, H_k depends only on the first k letters of the string $x_1x_2\ldots$. By (22) of Theorem 5.1, the limit constants κ and σ^2 are

$$\kappa = \sum_{k=2}^{\infty} EH_k(\mathbf{X}) \quad \text{and} \quad \sigma^2 = \lim_{n\to\infty} \frac{1}{n} E\left(\sum_{i=1}^{n}\sum_{k=2}^{\infty} H_k(\tau^i\mathbf{X}) - n\kappa\right)^2.$$

We will prove (Corollary 6.4) that in the particular case of interest here, where $h_k = u_k + v_k$, the random variables $H_k(\tau^i\mathbf{X})$ and $H_{k'}(\tau^{i'}\mathbf{X})$ are uncorrelated unless $i = i'$ and $k = k'$. It then follows that the terms of the sequence defining σ^2 are all equal, and so

$$\sigma^2 = \sum_{k=2}^{\infty} \mathrm{Var}(H_k(\mathbf{X})).$$

Lemma 6.1 *For each string $x_1x_2\ldots x_k$ of length $k \geq 2$ and each index $i \leq k-1$, define $j_i = j_i(x_1x_2\ldots x_k)$ to be the number of letters between $\bar{x}_i$ and x_{i+1} in the reference word $\mathcal{O}$ in the clockwise direction (see Fig. 3). Then*

$$V_2 = 0 \quad \textit{and} \quad U_k = V_k \quad \textit{for all } k \geq 3, \tag{37}$$

and

$$U_k(x_1x_2\ldots) = \frac{t(j_1, j_{k-1})}{g(g-1)^{k-1}} \tag{38}$$

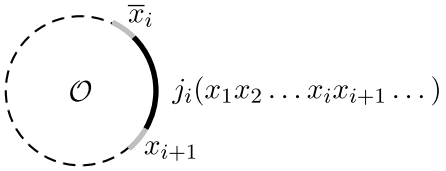

Fig. 3 The interval of length j_i in $\mathcal{O}$

where $t(a,b) = a(g-2-b) + b(g-2-a)$.

Therefore,

$$S_K(x_1x_2\dots) = \frac{t(j_1,j_1)}{g(g-1)} + 2\sum_{k=3}^{K} \frac{t(j_1,j_{k-1})}{g(g-1)^{k-1}}. \tag{39}$$

Proof The Markov chain with transition probabilities (8) is reversible (the transition probability matrix (8) is symmetric), and the transition probabilities are unchanged by inversion $a \mapsto \bar{a}$ and $a' \mapsto \bar{a}'$. Hence, the random strings $X_1'X_2'\dots X_k'$ and $\bar{X}_k'\bar{X}_{k-1}'\dots\bar{X}_1'$ have the same distribution. It follows that for each $k \geq 2$,

$$U_k(x_1x_2\dots) = V_k(x_1x_2\dots).$$

Consider the case $k = 2$. In order that $u_2(x_1x_2, X_1'X_2') \neq 0$ it is necessary and sufficient that the letters $\bar{x}_1\bar{X}_1'x_2X_2'$ occur in cyclic order (either clockwise or counterclockwise) in the reference word $\mathcal{O}$. For *clockwise* cyclic ordering, the letter $\bar{X}_1'$ must be one of the j_1 letters between $\bar{x}_1$ and x_2, and X_2' must be one of the $g-2-j_1$ letters between x_2 and $\bar{x}_1$. Similarly, for *counterclockwise* cyclic ordering, $\bar{X}_1'$ must be one of the $g-2-j_1$ letters between x_2 and $\bar{x}_1$, and X_2 one of the j_1 letters between $\bar{x}_1$ and x_2. But X_1', and hence also its inverse $\bar{X}_1'$, is uniformly distributed on the g letters, and given the value of $\bar{X}_1'$ the random variable X_2' is uniformly distributed on the remaining $(g-1)$ letters. Therefore,

$$U_2(x_1x_2) = \frac{t(j_1,j_1)}{g(g-1)}.$$

The case $k \geq 3$ is similar. In order that $u_k(x_1x_2\dots, X_1'X_2'\dots)$ be nonzero it is necessary and sufficient that the strings $x_1x_2\dots x_k$ and $X_1'X_2'\dots X_k'$ differ precisely in the first and kth entries, and that the letters $\bar{x}_1\bar{X}_1'x_2$ occur in the same cyclic order as the letters $x_kX_k'\bar{x}_{k-1}$. This order will be *clockwise* if and only if $\bar{X}_1'$ is one of the j_1 letters between $\bar{x}_1$ and x_2 *and* X_k' is one of the $g-2-j_{k-1}$ letters between x_k and $\bar{x}_{k-1}$. The order will be *counterclockwise* if and only if $\bar{X}_1'$ is one of the $g-2-j_1$ letters between x_2 and $\bar{x}_2$ *and* X_k' is one of the j_{k-1} letters between $\bar{x}_{k-1}$ and x_k. Observe that all of these possible

choices will lead to reduced words $X_1' x_2 x_3 \dots x_{k-1} X_k'$. By (8), the probability of one of these events occurring is

$$U_k(x_1 x_2 \dots x_k) = \frac{t(j_1, j_{k-1})}{g(g-1)^{k-1}}. \qquad \square$$

For $i = 1, 2, \dots$, define $J_i = j_i(X_1 X_2 \dots)$ to be the random variable obtained by evaluating the function j_i at a random string generated by the Markov chain, that is, J_i is the number of letters between $\bar{X}_i$ and X_{i+1} in the reference word $\mathcal{O}$ in the clockwise direction. Because X_{i+1} is obtained by randomly choosing one of the letters of $\mathcal{G}$ other than $\bar{X}_i$, the random variable J_i is independent of X_i. Since these random choices are all made independently, the following is true:

Lemma 6.2 *The random variables $X_1, J_1, J_2, \dots$ are mutually independent, and each J_i has the uniform distribution on the set $\{0, 1, 2, \dots, g-2\}$. Consequently,*

$$\begin{aligned} EJ_i &= (g-2)/2, \\ EJ_i^2 &= (g-2)(2g-3)/6, \\ EJ_i^3 &= (g-2)^2(g-1)/4, \\ EJ_i^4 &= (g-2)(2g-3)(3g^2-9g+5)/30 \quad \textit{and} \\ EJ_i J_{i'} &= EJ_i EJ_{i'} = (g-2)^2/4 \quad \textit{for } i \neq i'. \end{aligned} \tag{40}$$

By Lemma 6.1, the conditional expectations U_k, V_k are quadratic functions of the cycle gaps $J_1, J_2, \dots$. Consequently, the unconditional expectations

$$Eu_k(\mathbf{X}, \mathbf{X}') = EU_k(\mathbf{X})$$

can be deduced from the elementary formulas of Lemma 6.2 by linearity of expectation. Consider first the case $k \geq 3$:

$$\begin{aligned} g(g-1)^{k-1} EU_k(\mathbf{X}) &= Et(J_1, J_{k-1}) \\ &= 2EJ_1(g-2-J_{k-1}) \\ &= 2(g-2)EJ_1 - 2EJ_1 J_{k-1} \\ &= (g-2)^2 - (g-2)^2/2 \\ &= (g-2)^2/2. \end{aligned}$$

For $k = 2$:

$$\begin{aligned} g(g-1)EU_2(\mathbf{X}) &= Et(J_1, J_1) \\ &= 2EJ_1(g-2-J_1) \end{aligned}$$

$$= 2(g-2)EJ_1 - 2EJ_1^2$$
$$= (g-2)^2 - (g-2)(2g-3)/3$$
$$= (g-2)(g-3)/3.$$

Corollary 6.3 *If* $\mathbf{X} = X_1X_2\dots$ *and* $\mathbf{X}' = X_1'X_2'\dots$ *are* independent *realizations of the stationary Markov chain with transition probabilities* (8), *then*

$$Eu_2(\mathbf{X}, \mathbf{X}') = EU_2(\mathbf{X}) = \frac{(g-2)(g-3)}{3g(g-1)}, \tag{41}$$

$$Eu_k(\mathbf{X}, \mathbf{X}') = EU_k(\mathbf{X}) = \frac{(g-2)^2}{2g(g-1)^{k-1}} \quad \textit{for } k \geq 3, \quad \textit{and} \tag{42}$$

$$ES_\infty(\mathbf{X}) = Eu_2(\mathbf{X}, \mathbf{X}') + \sum_{k=3}^{\infty} E(u_k + v_k)(\mathbf{X}, \mathbf{X}') = \frac{(g-2)}{3(g-1)}. \tag{43}$$

The variances and covariances of the random variables $U_k(\mathbf{X})$ can be calculated in similar fashion, using the independence of the cycle gaps J_k and the moment formulas in Lemma 6.2. It is easier to work with the scaled variables $t(J_1, J_k) = g(g-1)^k U_{k+1}$ rather than with the variables U_k, and for convenience we will write $J_i^R = g - 2 - J_i$. Note that by definition and Lemma 6.2 the random variables J_i and J_i^R both have the same distribution (uniform on the set $\{0, 1, \dots, g-2\}$), and therefore also the same moments.

Case 0: If i, j, k, m are distinct, or if $i = j$ and i, k, m are distinct, then

$$Et(J_i, J_j)t(J_k, J_m) = Et(J_i, J_j)Et(J_k, J_m),$$

since the random variables J_i, J_j, J_k, J_m (or in the second case J_i, J_k, J_m) are independent. It follows that for any indices i, j, k, m such that and $i + k \neq j + m$, the random variables $U_k(\tau^i\mathbf{X})$ and $U_m(\tau^j\mathbf{X})$ are uncorrelated. (Here, as usual, τ is the forward shift operator.)

Case 1: If $i, k, m \geq 1$ are distinct then

$$\begin{aligned} Et(J_i, J_k)t(J_i, J_m) &= E(J_iJ_k^R + J_kJ_i^R)(J_iJ_m^R + J_mJ_i^R) \\ &= EJ_iJ_k^RJ_iJ_m^R + EJ_iJ_k^RJ_i^RJ_m + EJ_i^RJ_kJ_iJ_m^R \\ &\quad + EJ_i^RJ_kJ_i^RJ_m \\ &= ((g-2)^2/4)(EJ_i^2 + EJ_iJ_i^R + EJ_i^RJ_i + EJ_i^RJ_i^R) \\ &= ((g-2)^2/4)E(J_i + J_i^R)^2 \end{aligned}$$

$$
\begin{aligned}
&= (g-2)^4/4 \\
&= Et(J_i, J_k)Et(J_i, J_m).
\end{aligned}
$$

Thus, the random variables $t(J_i, J_k)$ and $t(J_i, J_m)$ are uncorrelated. Consequently, for all choices of $i, j, k \geq 1$ such that $j \neq k$, the random variables $U_j(\tau^i \mathbf{X})$ and $U_m(\tau^i \mathbf{X})$ are uncorrelated.

Case 2: If $i \neq k$ then

$$
\begin{aligned}
Et(J_i, J_i)t(J_i, J_k) &= EJ_iJ_i^RJ_iJ_k^RJ_k + EJ_iJ_i^RJ_i^RJ_k \\
&\quad + EJ_i^RJ_iJ_iJ_k^RJ_k + EJ_i^RJ_iJ_i^RJ_k \\
&= ((g-2)/2)(2EJ_iJ_iJ_i^RJ_k + 2EJ_iJ_i^RJ_i^RJ_k) \\
&= 2(g-2)(EJ_iJ_i(g-2-J_i)) \\
&= 2(g-2)((g-2)EJ_i^2 - EJ_i^3) \\
&= (g-2)^3(g-3)/6 \\
&= Et(J_i, J_k)Et(J_i, J_i)
\end{aligned}
$$

Once again, the two random variables are uncorrelated. It follows that for all $i \geq 1$ and $m \geq 3$ the random variables $U_2(\tau^i \mathbf{X})$ and $U_m(\tau^i \mathbf{X})$ are uncorrelated.

Case 3: If $k \geq 2$ then

$$
\begin{aligned}
Et(J_1, J_k)^2 &= EJ_1J_1J_k^RJ_k^R + EJ_1^RJ_1^RJ_kJ_k + 2EJ_1^RJ_1J_k^RJ_k \\
&= 2(EJ_1^2)^2 + 2(EJ_1J_1^R)^2 \\
&= (g-2)^2(2g-3)^2/18 + (g-2)^2(g-3)^2/18
\end{aligned}
$$

and so

$$
\begin{aligned}
\operatorname{var}(t(J_1, J_k)) &= Et(J_1, J_k)^2 - (Et(J_1, J_k))^2 \\
&= (g-2)^2(2g-3)^2/18 + (g-2)^2(g-3)^2/18 - (g-2)^4/4 \\
&= g^2(g-2)^2/36.
\end{aligned}
$$

Case 4: When $k = 1$:

$$
\begin{aligned}
Et(J_1, J_1)^2 &= 4EJ_1J_1J_1^RJ_1^R \\
&= 4((g-2)^2EJ_1^2 - 2(g-2)EJ_1^3 + EJ_1^4) \\
&= 2(g-2)(g-3)(g^2-4g+5)/15,
\end{aligned}
$$

so

$$\begin{aligned}\operatorname{var}(t(J_1, J_1)) &= Et(J_1, J_1)^2 - (Et(J_1, J_1))^2 \\ &= 2(g-2)(g-3)(g^2-4g+5)/15 - (g-2)^2(g-3)^2/9 \\ &= g(g-2)(g-3)(g+1)/45.\end{aligned}$$

This proves:

Corollary 6.4 *The random variables $U_k(\tau^i \mathbf{X})$, where $i \geq 0$ and $k \geq 2$, are uncorrelated, and have variances*

$$\operatorname{Var}(U_k(\tau^i \mathbf{X})) = \frac{(g-2)^2}{36(g-1)^{2k-2}} \quad \text{for } k \geq 3, \tag{44}$$

$$\operatorname{Var}(U_2(\tau^i \mathbf{X})) = \frac{(g-2)(g-3)(g+1)}{45g(g-1)^2}.$$

Consequently,

$$\begin{aligned}\operatorname{Var}(S_\infty(\tau^i \mathbf{X})) &= \operatorname{Var}(U_2(\mathbf{X})) + \sum_{k=3}^{\infty} \operatorname{Var}(2U_k(\mathbf{X})) \\ &= \operatorname{Var}(U_2(\mathbf{X})) + \lim_{K\to\infty} \sum_{k=3}^{K} \operatorname{Var}(2U_k(\mathbf{X})) \\ &= \frac{(g-2)(g-3)(g+1)}{45g(g-1)^2} + \frac{g-2}{9g(g-1)^2} \\ &= \frac{(g-2)(g^2-2g+2)}{45g(g-1)^2} \\ &= \frac{2\chi(2\chi^2-2\chi+1)}{45(2\chi-1)^2(\chi-1)}.\end{aligned} \tag{45}$$

Appendix A: An example of the combinatorics of self-intersection counts

The counting of self-intersection numbers is based on the following idea: Two strands on a surface come close, stay together for some time and then separate. If one strand enters the strip from "above" and exits "below" and the other vice versa there must be an intersection. This intersection is measured by the functions u_k and v_k where k gives the "length of the time" that the

strands stay together. (See Fig. 1 showing pairs of subwords for which $u_2 \neq 0$ and $u_3 \neq 0$.)

Example A.1 Let $\mathcal{O}$ denote the cyclic word of Fig. 4. Consider a 16-gon with alternate sides labeled with the letters of $\mathcal{O}$ as in Fig. 1(I). By "glueing" the sides of this polygon labeled by the same letter, one obtains a surface Σ of genus two and one boundary component.

Let α be a necklace which can be unhooked to $\alpha^* = ab\overline{c}aac\overline{b}a$. There is a one to one correspondance between the self-intersection points of a representative of α with minimal self-intersection and the pairs of subwords of α listed in (a), (b) and (c).

(a) $(b\overline{c}, \overline{c}a)$, $(b\overline{c}, ac)$, $(\overline{c}a, c\overline{b})$ and $(ac, c\overline{b})$. (These are the all the pairs of the form (c_1c_2, d_1d_2) such that if w and w' are words with finite or infinite letters, $w = c_1c_2\ldots$ and $w' = d_1d_2\ldots$ then $u_2(w, w') = 1$ and $u_2(w', w) = 1$.)

(b) $(\overline{c}aa, aac)$ and $(\overline{b}aa, aab)$. (These are all the pairs $(c_1c_2\ldots c_k, d_1d_2\ldots d_k)$ of subwords of α, with $k \geq 3$ such that if $w = c_1c_2\ldots c_k\ldots$ and $w' = d_1d_2\ldots d_k\ldots$ then $u_k(w, w') = 1$ and $u_k(w', w) = 1$.)

(c) $(ab\overline{c}a, ac\overline{b}a)$. (This is the only pair of subwords $(c_1c_2\ldots c_k\ldots, d_1d_2\ldots d_k\ldots)$ of α of more than two letters such that if $w = c_1c_2\ldots c_k\ldots$ and $w' = d_1d_2\ldots d_k\ldots$ $v_k(w, w') = 1$ and $v_k(w', w) = 1$.)

Since there are seven pairs listed in (a), (b) and (c), the self-intersection number of α equals to seven.

Clearly the arcs corresponding to the subwords of α, $b\overline{c}$ and $\overline{c}a$ intersect in the polygon (see Fig. 5(I)). This suggests that the occurrence of $b\overline{c}$ and $\overline{c}a$ as subwords of a cyclic word will imply a self-intersection point in every representative of the cycled word.

Now, consider the pair of subwords of α, aa and $\overline{c}a$ (see Fig. 5(II)). Since both of the corresponding arcs land in the edge a of the polygon, the occurrence of these two subwords does not provide enough information to deduce the existence of a self-intersection point. In order to understand better this configuration of segments, we prolong the subwords starting with aa and $\overline{c}a$ until they have different letters at the beginning and at the end. Then we study how the arcs corresponding to these subwords intersect. So in our example we get $\overline{c}aa$ and aac, implying a self-intersection point (Fig. 5(II)).

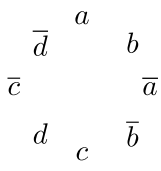

Fig. 4 An example of a word $\mathcal{O}$

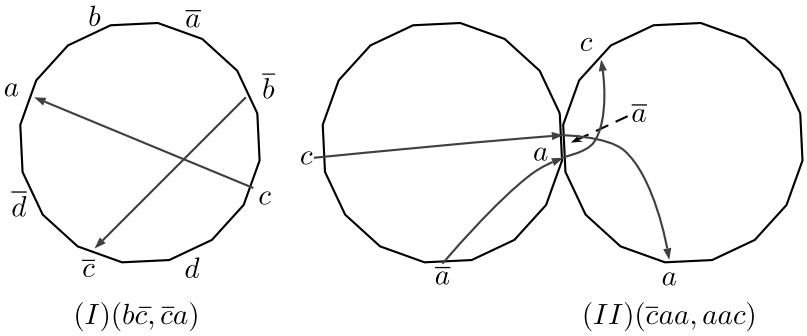

Fig. 5 Example

Appendix B: Background: probability, Markov chains, weak convergence

For the convenience of the reader we shall review some of the terminology of the subject here. (All of this is standard, and can be found in most introductory textbooks, for instance, [1, 2].)

A *probability space* is a measure space $(\Omega, \mathcal{B}, P)$ with total mass 1. Integrals with respect to P are called *expectations* and denoted by the letter E, or by E_P if the dependence on P must be emphasized. A *random variable* is a measurable, real-valued function on Ω; similarly, a *random vector* or a *random sequence* is a measurable function taking values in a vector space or sequence space. The *distribution* of a random variable, vector, or sequence X is the induced probability measure $P \circ X^{-1}$ on the range of X. Most questions of interest in the subject concern the distributions of various random objects, so the particular probability space on which these objects are defined is usually not important; however, it is sometimes necessary to move to a "larger" probability space (e.g., a product space) to ensure that auxiliary random variables can be defined. This is the case, for instance, in sec. 6, where independent copies of a Markov chain are needed.

Definition B.1 A sequence $\ldots, X_{-1}, X_0, X_1, \ldots$ of $\mathcal{G}$-valued random variables defined on some probability space $(\mathcal{X}, \mathcal{B}, P)$ is said to be a *stationary Markov chain* with *stationary distribution* π and transition probabilities $p(a, a')$ if for every finite sequence $w = w_0 w_1 \ldots w_k$ of elements of $\mathcal{G}$ and every integer m,

$$P\{X_{m+j} = w_j \text{ for each } 0 \le j \le k\} = \pi(w_0) \prod_{j=0}^{k-1} p(w_j, w_{j+1}). \tag{46}$$

If $p(a, a')$ is a stochastic matrix on set $\mathcal{G}$ and π satisfies the stationarity condition $\pi(a) = \sum_{a'} \pi(a')p(a', a)$ then there is a probability measure on the sequence space $\mathcal{G}^{\mathbb{Z}}$ under which the coordinate variables form a Markov chain with transition probabilities $p(a, a')$ and stationary distribution π. This follows from standard measure extension theorems—see, e.g., [1], Sect. 1.8.

Definition B.2 A sequence of random variables X_n (not necessarily all defined on the same probability space) is said to converge *weakly* or *in distribution* to a limit distribution F on $\mathbb{R}$ (denoted by $X_n \Rightarrow F$) if the distributions F_n of X_n converge to F in the weak topology on measures, that is, if for every bounded, continuous function $\varphi : \mathbb{R} \to \mathbb{R}$ (or equivalently, for every continuous function φ with compact support),

$$\lim_{n\to\infty} \int \varphi \, dF_n = \int \varphi \, dF$$

as $n \to \infty$.

It is also customary to write $F_n \Longrightarrow F$ for this convergence, since it is really a property of the distributions. When the limit distribution F is the point mass δ_0 at 0 we may sometimes write $X_n \Rightarrow 0$ instead of $X_n \Rightarrow \delta_0$. The weak topology on probability measures is metrizable; when necessary we will denote by ϱ a suitable metric. It is an elementary fact that weak convergence of probability distributions on $\mathbb{R}$ is equivalent to the pointwise convergence of the cumulative distribution functions at all points of continuity of the limit cumulative distribution function. Thus, Theorem 3.1 is equivalent to the assertion that the random variables $(N(\alpha) - n^2\kappa)/n^{3/2}$ on the probability spaces $(\mathcal{F}_n, \mu_n)$ converge in distribution to Φ_σ.

We conclude with several elementary tools of weak convergence that will be used repeatedly throughout the paper. First, given any countable family X_n of random variables, possibly defined on different probability spaces, there exist on the Lebesgue space ([0, 1], Lebesgue) random variables Y_n such that for each n the random variables X_n and Y_n have the same distribution. Furthermore, the random variables Y_n can be constructed in such a way that if the random variables X_n converge in distribution then the random variables Y_n converge pointwise on [0, 1] (the converse is trivial). Next, define the *total variation distance* between two probability measures μ and ν defined on a common measurable space $(\Omega, \mathcal{B})$ by

$$\|\mu - \nu\|_{TV} = \max(\mu(A) - \nu(A))$$

where A ranges over all measurable subsets (events) of Ω. Total variation distance is never increased by mapping, that is, if $T : \Omega \to \Omega'$ is a measurable

transformation then

$$\|\mu \circ T^{-1} - \nu \circ T^{-1}\|_{TV} \leq \|\mu - \nu\|_{TV}. \tag{47}$$

Also, if μ and ν are mutually absolutely continuous, with Radon–Nikodym derivative $d\mu/d\nu$, then

$$\|\mu - \nu\|_{TV} = \frac{1}{2} E_\nu \left| \frac{d\mu}{d\nu} - 1 \right|. \tag{48}$$

It is easily seen that if a sequence of probability measures $\{\mu_n\}_{n\geq 1}$ on $\mathbb{R}$ is Cauchy in total variation distance then the sequence converges in distribution. The following lemma is elementary:

Lemma B.3 *Let X_n and Y_n be two sequences of random variables, all defined on a common probability space, let a_n be a sequence of scalars, and fix $r > 0$. Denote by F_n and G_n the distributions of X_n and Y_n, respectively. Then the equivalence*

$$\frac{Y_n - a_n}{n^r} \implies F \quad \text{if and only if} \quad \frac{X_n - a_n}{n^r} \implies F \tag{49}$$

holds if either

$$(X_n - Y_n)/n^r \implies 0 \quad \text{or} \tag{50}$$

$$\|F_n - G_n\|_{TV} \longrightarrow 0 \tag{51}$$

as $n \to \infty$. Furthermore, (51) implies (50).

The following lemma is an elementary consequence of Chebyshev's inequality and the definition of weak convergence.

Lemma B.4 *Let X_n be a sequence of random variables. Suppose that for every $\varepsilon > 0$ there exist random variables X_n^ε and R_n^ε such that*

$$X_n = X_n^\varepsilon + R_n^\varepsilon, \tag{52}$$

$$X_n^\varepsilon \implies \text{Normal}(0, \sigma_\varepsilon^2), \quad \text{and}$$

$$E|R_n^\varepsilon|^2 \leq \varepsilon.$$

Then $\lim_{\varepsilon\to 0} \sigma_\varepsilon^2 := \sigma^2 \geq 0$ exists and is finite, and

$$X_n \implies \text{Normal}(0, \sigma^2). \tag{53}$$

References

1. Billingsley, P.: Probability and Measure, 3rd edn. Wiley Series in Probability and Mathematical Statistics. Wiley, New York (1995). A Wiley-Interscience Publication
2. Billingsley, P.: Convergence of Probability Measures, 2nd edn. Wiley Series in Probability and Statistics: Probability and Statistics. Wiley, New York (1999). A Wiley-Interscience Publication
3. Birman, J.S., Series, C.: An algorithm for simple curves on surfaces. J. Lond. Math. Soc. (2) **29**(2), 331–342 (1984)
4. Chas, M.: Experimental results in combinatorial topology. Manuscript
5. Chas, M.: Combinatorial Lie bialgebras of curves on surfaces. Topology **43**(3), 543–568 (2004)
6. Chas, M., Phillips, A.: Self-intersection numbers of curves in the doubly punctured plane. arXiv:1001.4568
7. Chas, M., Phillips, A.: Self-intersection numbers of curves on the punctured torus. Exp. Math. **19**(2), 129–148 (2010)
8. Cohen, M., Lustig, M.: Paths of geodesics and geometric intersection numbers. I. In: Combinatorial Group Theory and Topology, Alta, Utah, 1984. Annals of Mathematics Studies, vol. 111, pp. 479–500. Princeton University Press, Princeton (1987)
9. Denker, M., Keller, G.: On U-statistics and v. Mises' statistics for weakly dependent processes. Z. Wahrscheinlichkeitstheor. Verw. Geb. **64**(4), 505–522 (1983)
10. Hoeffding, W.: A class of statistics with asymptotically normal distribution. Ann. Math. Stat. **19**, 293–325 (1948)
11. Iosifescu, M.: On U-statistics and von Mises statistics for a special class of Markov chains. J. Stat. Plan. Inference **30**(3), 395–400 (1992)
12. Lalley, S.P.: Self-intersections of random geodesics on negatively curved surfaces. arXiv:0907.0259
13. Lalley, S.P.: Self-intersections of closed geodesics on a negatively curved surface: statistical regularities. In: Convergence in Ergodic Theory and Probability, Columbus, OH, 1993. Ohio State University Mathematical Research Institute Publications, vol. 5, pp. 263–272. de Gruyter, Berlin (1996)
14. Mirzakhani, M.: Growth of the number of simple closed geodesics on hyperbolic surfaces. Ann. Math. (2) **168**(1), 97–125 (2008)