THE GENERALIZED ALICE HH VS BOB HT PROBLEM

SVANTE JANSON, MIHAI NICA, AND SIMON SEGERT

Abstract. In 2024, Daniel Litt posed a simple coinflip game pitting Alice's "Heads-Heads" vs Bob's "Heads-Tails": who is more likely to win if they score 1 point per occurrence of their substring in a sequence of n fair coinflips? This attracted over 1 million views on X and quickly spawned several articles explaining the counterintuitive solution. We study the generalized game, where the set of coin outcomes, $\{\text{Heads}, \text{Tails}\}$, is generalized to an arbitrary finite alphabet $\mathcal{A}$, and where Alice's and Bob's substrings are any finite $\mathcal{A}$-strings of the same length. We find that the winner of Litt's game can be determined by a single quantity which measures the amount of prefix/suffix self-overlaps in each string; whoever's string has more overlaps *loses*. For example, "Heads-Tails" beats "Heads-Heads" in the original problem because "Heads-Heads" has a prefix/suffix overlap of length 1 while "Heads-Tails" has none. The method of proof is to develop a precise Edgeworth expansion for discrete Markov chains, and apply this to calculate Alice's and Bob's probability to win the game correct to order $O(1/n)$.

1. Introduction

On March 16, 2024 Daniel Litt posted the following brainteaser on X [17]:

> Flip a fair coin 100 times—it gives a sequence of heads (H) and tails (T). For each HH in the sequence of flips, Alice gets a point; for each HT, Bob does, so e.g. for the sequence THHHT Alice gets 2 points and Bob gets 1 point. Who is most likely to win?

The post gained popularity due to the deceptively unintuitive nature of the problem, with plurality of respondents in an attached poll believing (incorrectly) that the game is fair, and the correct answer being chosen the least often.

This problem turned out to have a surprisingly rich mathematical structure. In short order, a succession of papers appeared: Ekhad and Zeilberger [5], Segert [26], and Grimmett [8], each of which rigorously analyzed the problem via different techniques (respectively, symbolic computation, asymptotic analysis, and probability theory). We highlight here two key results. Firstly, Litt's qualitative question "Who is more likely to win" was answered for an arbitrary number n of coin flips: Bob is strictly favored as long as n is at least 3 [26; 8]. Secondly, and more immediately relevant to the present work, for the quantitative question of finding the winning probabilities, the following asymptotic as $n \to \infty$ was found by [5; 26; 8] (by different methods):

$$\mathbb{P}(\text{Bob wins}) - \mathbb{P}(\text{Alice wins}) \sim \frac{1}{2\sqrt{\pi n}} \tag{1.1}$$

Date: March 24, 2025.

SJ is supported by the Knut and Alice Wallenberg Foundation and the Swedish Research Council.
MN is supported by the National Sciences and Engineering Research Council of Canada.

and, more precisely,

$$\frac{1}{2} - \mathbb{P}(\text{Alice wins}) \sim \frac{3}{4\sqrt{\pi n}}, \qquad \frac{1}{2} - \mathbb{P}(\text{Bob wins}) \sim \frac{1}{4\sqrt{\pi n}}. \tag{1.2}$$

A natural generalization of Litt's problem, consider for example by [5], is to let Alice's and Bob's strings be arbitrary finite sequences of the same length ℓ in an arbitrary finite alphabet $\mathcal{A}$; we then assume that Alice and Bob do a generalized coin tossing where a sequence of random letters are generated, with the letters independent and uniformly distributed. (See Section 4 for details. We leave extensions to non-uniform letter distributions to the reader.) We call also this generalization *Litt's game.*

This generalization to arbitrary strings was considered by Basdevant et al. [1] who established, by a combinatorial argument, a condition under which such a game is exactly fair (see Remark 1.2 below). The main purpose of the present paper is to establish asymptotics extending (1.1)–(1.2) for the generalized problem.

Somewhat surprisingly, with the exception of a small set of pathological strings, we find that who has the advantage in the generalized Litt's game depends *only* on how Alice's/Bob's string overlap themselves in their prefixs/suffixes. More precisely, for a string $A = a_1a_2\dots a_\ell$ of length ℓ from an alphabet $\mathcal{A}$ with q letters, one can calculate a single quantity $\theta_{AA} \in \mathbb{R}$ that measures the size of the prefix/suffix overlap of A defined as follows:

$$\Theta(A, A) := \{1 \leqslant k \leqslant \ell - 1 : a_{\ell-k+1} \cdots a_\ell = a_1 \cdots a_k\}, \tag{1.3}$$

$$\theta_{AA} := \sum_{k\in\Theta(A,A)} q^{k-\ell} = q^{-\ell} \sum_{k\in\Theta(A,A)} q^k. \tag{1.4}$$

(This is a special case of θ_{UV} between two strings U, V defined in (4.10).) We will prove the following result, which shows that the asymptotic winner of Litt's game is determined by comparing the value θ_{AA} of Alice's string to θ_{BB} of Bob's string; whoever's θ overlap value is larger *loses* Litt's game asymptotically.

Theorem 1.1. *Let Alice and Bob play Litt's game with distinct words A and B of the same length ℓ in an alphabet $\mathcal{A}$ with q letters, and assume that n letters are chosen at random, uniformly and independently.*

Exclude the two cases, both with $q = 2$:

(i) *$A = \mathsf{H}\mathsf{T}^{\ell-1}$ and $B = \mathsf{T}^{\ell-1}\mathsf{H}$ for some $\ell \geqslant 2$ (see Example 6.2),*

(ii) *$A = \mathsf{H}$ and $B = \mathsf{T}$ (see Example 6.3),*

and their variants obtained by interchanging Alice and Bob or H *and* T *(or both).*

Then, with θ_{UU} given by (1.4) *(or* (4.10)*) and $\sigma^2 = \sigma^2(A, B)$ given by* (4.19)*, we have $\sigma^2 > 0$, and*

$$\mathbb{P}(\text{Alice wins}) = \frac{1}{2} + \frac{\theta_{BB} - \theta_{AA} - 1}{2\sqrt{2\pi\sigma^2}} n^{-1/2} + O(n^{-1}), \tag{1.5}$$

$$\mathbb{P}(\text{Bob wins}) = \frac{1}{2} + \frac{\theta_{AA} - \theta_{BB} - 1}{2\sqrt{2\pi\sigma^2}} n^{-1/2} + O(n^{-1}), \tag{1.6}$$

$$\mathbb{P}(\text{Tie}) = \frac{1}{\sqrt{2\pi\sigma^2}} n^{-1/2} + O(n^{-1}), \tag{1.7}$$

and thus

$$\mathbb{P}(\text{Alice wins}) - \mathbb{P}(\text{Bob wins}) = \frac{\theta_{BB} - \theta_{AA}}{\sqrt{2\pi\sigma^2}} n^{-1/2} + O(n^{-1}). \tag{1.8}$$

In the two excluded cases, the conclusions (1.5)–(1.7) fail (in somewhat different ways), see Examples 6.2 and 6.3.

Remark 1.2. In particular, Theorem 1.1 shows that the game is fair up to order n^{-1} if and only if $\theta_{AA} = \theta_{BB}$. This is (in different notation) the condition by Basdevant et al. [1], who showed (by completely different methods) that in this case the game is perfectly fair for every n: $\mathbb{P}(\text{Alice wins}) = \mathbb{P}(\text{Bob wins})$. (The result in [1] is stated for $q = 2$, but the proof holds for an arbitrary finite alphabet.) Our result thus shows that if the condition in [1] is not satisfied, then the game is for all large n not fair; thus their condition for fairness is both necessary and sufficient. △

Our method to prove Theorem 1.1 is to recognize (1.5) and (1.6) as examples of (first order) Edgeworth expansions. Edgeworth expansions are useful approximations in many situations in probability and statistics, and they have been rigorously established in many situations, see Section 2.3 for a background. We use a general result giving an Edgeworth expansion for partial sums of an integer-valued function of a finite-state Markov chain. There is a large literature on Edgeworth expansions in various situations, including Markov chains, see Section 2.3. However, we have failed to find a general theorem that is directly applicable here. We therefore state such a theorem for finite-state Markov chains here (in two versions, Theorems 3.2 and 3.4); it is perhaps not new, but since we do not know a reference we give for completeness a complete proof.

2. Preliminaries

2.1. **Notation.**

The *characteristic function* of a random variable X is defined by

$$\varphi(t) = \varphi_X(t) := \mathbb{E}\, e^{\mathrm{i}tX}, \qquad t \in \mathbb{R}. \tag{2.1}$$

Vectors are generally regarded as column vectors in formulas using matrix notation; the row vector that is the transpose of a column vector v is denoted v^{t}.

We let $\mathbf{1} = (1, \dots, 1)^{\mathsf{t}}$, the (column) vector with all entries 1, with the dimension of the vector determined by context.

For vectors v (in $\mathbb{C}^m$ for some $m \in \mathbb{N}$) we let $\|v\|$ be the usual Euclidean norm. For a square matrix A, let $\|A\|$ denote its *operator norm*

$$\|A\| := \sup\bigl\{\|Av\| : \|v\| = 1\bigr\} \tag{2.2}$$

and let $\rho(A)$ denote its *spectral norm*

$$\rho(A) := \max\{|\lambda| : \lambda \text{ is an eigenvalue of } A\}. \tag{2.3}$$

Recall the *spectral radius formula*

$$\rho(A) = \lim_{n\to\infty} \|A^n\|^{1/n}. \tag{2.4}$$

A square matrix A with non-negative entries is *irreducible* if for every pair of indices i, j we have $(A^n)_{ij} > 0$ for some $n \geqslant 1$; the matrix is *primitive* if we further can choose the same n for all pairs i, j; see e.g. [27, Chapter 1]. Recall that a matrix is primitive if and only if it is irreducible and *aperiodic* [27, Theorem 1.4]. A *stochastic* matrix is a square matrix with non-negative entries where all row sums are 1. It follows from the Perron–Frobenius theorem [27, Theorem 1.5] that an irreducible stochastic matrix has an eigenvalue 1 which is (algebraically) simple, and that the spectral radius is 1; corresponding right and left eigenvectors are $\mathbf{1}$ and π^{t}, the stationary distribution of the corresponding Markov chain.

For $z_0 \in \mathbb{C}$ and $r > 0$, let $D(z_0, r) := \{z \in \mathbb{C} : |z - z_0| < r\}$, the open disc with centre z_0 and radius r.

We let $\mathbf{1}_{\mathcal{E}}$ denote the indicator of an event $\mathcal{E}$; this is thus 1 if $\mathcal{E}$ occurs and 0 otherwise.

C denotes unspecified constants that may vary from one occurrence to the next. They may depend on parameters, and we may write e.g. $C(\delta)$ to stress this.

2.2. **Cumulants.** If X is a random variable, with (for simplicity) all moments finite, then its *cumulants* (also called *semi-invariants*) are defined by, for integers $m \geqslant 1$, recalling (2.1),

$$\kappa_m = \kappa_m(X) := \mathrm{i}^{-m} \frac{\mathrm{d}^m}{\mathrm{d}t^m} \log \varphi_X(t)\big|_{t=0}; \tag{2.5}$$

in other words, $\log \varphi_X(t)$ has the Taylor expansion, for $M \geqslant 1$ and small t,

$$\log \varphi_X(t) = \sum_{m=1}^{M} \kappa_m \frac{(\mathrm{i}t)^m}{m!} + O(|t|^{M+1}). \tag{2.6}$$

See for example [9, Section 4.6]. The cumulant κ_m can be expressed as a polynomial in moments of order at most m. In particular [9, Theorem 4.6.4],

$$\kappa_1 = \mathbb{E}\, X, \tag{2.7}$$

$$\kappa_2 = \mathbb{E}\,[X^2] - (\mathbb{E}\, X)^2 = \operatorname{Var}(X), \tag{2.8}$$

$$\kappa_3 = \mathbb{E}\,[(X - \mathbb{E}\, X)^3]. \tag{2.9}$$

Thus κ_2 and κ_3 are simply the corresponding central moments. (Higher cumulants have more complicated formulas.)

2.3. **Background on Edgeworth expansions.** The idea of an Edgeworth expansion (or Edgeworth approximation) [4] is that many random variables in probability theory and statistics have distributions that are approximatively normal, and that the approximation often can be improved by adding extra terms to the normal distribution. It turns out that the natural terms to add are derivatives of the normal density function, with coefficients that are given by some (explicit but a little complicated) polynomials of cumulants κ_m, $m \geqslant 3$, divided by powers of the variance. (Part of the motivation is that for a normal random variable, all cumulants $\kappa_m = 0$ for $m \geqslant 3$; thus the cumulants measure in some way deviations from normality. For detailed motivations, see [4] and [2, Chapter 17.6-7].) For a continuous random variable Z, for simplicity normalized to have $\mathbb{E}\, Z = 0$ and $\operatorname{Var} Z = 1$, the one-term Edgeworth appoximation is

$$\mathbb{P}(Z \leqslant x) \approx \Phi(x) + \frac{\kappa_3(Z)}{6\sqrt{2\pi}}(1 - x^2)e^{-x^2/2}, \qquad -\infty < x < \infty, \tag{2.10}$$

where

$$\Phi(x) := \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, \mathrm{d}x \tag{2.11}$$

is the standard normal distribution function. This approximation can be made rigorous, with error bounds, in many situations. The most important, and archetypical, case is when the random variable Z is a normalized sum $\widetilde{S}_n := \frac{1}{\sigma\sqrt{n}} \sum_1^n X_i$ of n independent and identially distributed (i.i.d.) random variables X_i with $\mathbb{E}\, X_i = 0$, $\operatorname{Var} X_i = \sigma^2$, and a finite third moment $\mathbb{E}\, |X_i|^3$. Then $\kappa_3(\widetilde{S}_n) = n^{-1/2}\kappa_3(X_i)/\sigma^3$, and, as shown by [6, Theorem IV.2, p. 49], if the distribution of X_i is non-lattice, then (2.10) holds with an error $o(n^{-1/2})$, i.e., with $S_n := \sum_1^n X_i$ the unnormalized sum,

$$\mathbb{P}(S_n \leqslant x\sigma\sqrt{n}) = \mathbb{P}(\widetilde{S}_n \leqslant x) = \Phi(x) + \frac{\kappa_3(X_1)}{6\sigma^3\sqrt{2\pi n}}(1-x^2)e^{-x^2/2} + o\bigl(n^{-1/2}\bigr), \tag{2.12}$$

uniformly in x. (Note that the right-hand side is $\Phi(x) + O(n^{-1/2})$ uniformly in x, so (2.12) implies, and can be seen as a precise version of, the Berry–Esseen theorem [9, Theorem 7.6.1].) Furthermore, under somewhat stronger conditions, the expansion (2.12) can be continued to any number of terms, where the mth term is of the order $n^{-m/2}$; see [6, Theorem IV.1, p. 48] or [23, Theorem VI.4 and (VI.1.13)] for a precise statement and the general form of the terms.

However, in this paper we are interested in integer-valued variables, and then (2.12) is not appropriate, since the right-hand side ignores the jumps in the left-hand side when $x\sigma\sqrt{n}$ is an integer. The correct version of (2.12) for integer-valued X_i with $\mathbb{E}\, X_i = 0$ is

$$\begin{aligned}\mathbb{P}(S_n \leqslant x\sigma\sqrt{n}) = \mathbb{P}(\widetilde{S}_n \leqslant x) = \Phi(x) &+ \frac{\kappa_3(X_1)}{6\sigma^3\sqrt{2\pi n}}(1-x^2)e^{-x^2/2} \\ &+ \frac{1}{\sigma\sqrt{2\pi n}}\vartheta(x\sigma\sqrt{n})e^{-x^2/2} + o\bigl(n^{-1/2}\bigr),\end{aligned} \tag{2.13}$$

where we have added a correction term with

$$\vartheta(x) := \tfrac{1}{2} - (x - \lfloor x \rfloor), \tag{2.14}$$

see [6, Theorem IV.3, p. 56] (which also includes the case $\mathbb{E}\, X_i \neq 0$, for simplicity ignored here). See also [6, Theorem IV.4, p. 61] and [23, Theorem VI.6] for the corresponding result with further terms (similar, but more complicated), and [16] for an interpretation of this expansion as a Sheppard's correction of the version for the continuous (or, more generally, non-lattice) case.

Note that $\vartheta(x) = 0$ when x is the midpoint between two consecutive integers. In fact, (2.13) can be regarded as an approximation as in (2.12) of the distribution function of S_n interpolated linearly between such half-integer points, see [7, Theorem XVI.4.2].

In the present paper we are interested in the distribution function at or close to the mean, i.e., the case $x \approx 0$ above. Note that if, say, $x\sigma\sqrt{n} = O(1)$, so $x = O(n^{-1/2})$, then (2.13) simplifies to

$$\mathbb{P}(S_n \leqslant x\sigma\sqrt{n}) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}}x + \frac{\kappa_3(X_1)}{6\sigma^3\sqrt{2\pi n}} + \frac{1}{\sigma\sqrt{2\pi n}}\vartheta(x\sigma\sqrt{n}) + O\bigl(n^{-1}\bigr). \tag{2.15}$$

We remark that under suitable conditions there are also corresponding Edgeworth expansions of the density function (in the absolutely continuous case) or point probabilities (in the discrete case), which refine the local limit theorem in a similar way, see for example [6, Theorem IV.5, p. 63] or [23, Theorem VII.13]; we will not use these expansions here.

In the present paper, we cannot use the basic results above, since the random variable we are interested in is *not* a sum of i.i.d. variables. Nevertheless, the Edgeworth expansions above have been extended to various other situations, and there is a large literature on Edgeworth expansions under various situations with dependency; we mention only a few references that are closely related to our case, although none of them is directly applicable to it, see further the discussion before Theorem 3.2. We consider below an irreducible finite-state Markov chain, and random variables defined by (3.7). For this case, Siraždinov [29] proved (under some conditions) an Edgeworth expansion for point probabilities (and a Berry–Esseen type estimate for the distribution function). Nagaev [21] considered more general Markov chains, allowing an infinite number of states, and proved Edgeworth expansions for both continuous and discrete cases under some conditions. (For the continuous case, see also [3] with some corrections and improvements.) For the integer-valued case that we are interested in, Hipp [12, Theorem (3.1)] proved an Edgeworth expansion for point probabilities under more general conditions. The variables in our application to Litt's game are m-dependent, so an alternative approach would be to use results on Edgeworth expansions for sums of m-dependent variables. However, we have not found a suitable such theorem for integer-valued variables; for the non-lattice case, see e.g. Heinrich [11], Rhee [24], and [18]; see also Rinott and Rotar [25] for a more general result, and the further references there.

2.4. **Group Inverse.** If A is a square (possibly singular) matrix, then the *group inverse* A^{g} is defined to be the matrix satisfying:

$$AA^{\mathsf{g}} = A^{\mathsf{g}}A, \tag{2.16}$$

$$AA^{\mathsf{g}}A = A, \tag{2.17}$$

$$A^{\mathsf{g}}AA^{\mathsf{g}} = A^{\mathsf{g}}, \tag{2.18}$$

The matrix A^{g} is unique if it exists, however it may not exist in the first place.

If we regard A as a linear operator in some or $\mathbb{R}^q$ (or $\mathbb{C}^q$), with kernel $\ker(A)$ and range $\operatorname{ran}(A)$, then A^{g} exists if and only if $\ker(A)\cap\operatorname{ran}(A)=\{0\}$ and thus $\mathbb{R}^q$ (resp. $\mathbb{C}^q$) is the direct sum $\ker(A)\oplus\operatorname{ran}(A)$; in this case $A^{\mathsf{g}}|_{\ker(A)}=0$ and $A^{\mathsf{g}}|_{\operatorname{ran}(A)}$ is the inverse of $A:\operatorname{ran}(A)\to\operatorname{ran}(A)$.

It can be shown that A^{g} does exist whenever $A=I-P$ where P is a stochastic matrix, see [19]. In this case, there is actually more that we can say. One useful identity that holds if P is irreducible is

$$(I-P)^{\mathsf{g}}(I-P) = I - \mathbf{1}\pi^{\mathsf{t}} \tag{2.19}$$

where π is the stationary distribution, see Theorem 2.2 in [19].

Most of the time when we want to actually compute $(I-P)^{\mathsf{g}}$, we will use the following well-known representation.

Proposition 2.1. *If P is an irreducible stochastic matrix, then*

$$\begin{aligned}(I-P)^{\mathsf{g}} &= (I-P+\mathbf{1}\pi^{\mathsf{t}})^{-1}-\mathbf{1}\pi^{\mathsf{t}}\\ &= \lim_{t\nearrow 1}\big((I-tP)^{-1}-\mathbf{1}\pi^{\mathsf{t}}/(1-t)\big).\end{aligned} \tag{2.20}$$

Proof. The first equality is proved in [19]. For the second equality, write

$$(I-P+\mathbf{1}\pi^{\mathsf{t}})^{-1} = \lim_{t\nearrow 1}(I-t(P-\mathbf{1}\pi^{\mathsf{t}}))^{-1} \tag{2.21}$$

Since for any $t<1$, the spectral radius of $t(P-\mathbf{1}\pi^{\mathsf{t}})$ is <1, the inverse on the right side can be expanded as a geometric series. Moreover, it is easily seen that $(P-\mathbf{1}\pi^{\mathsf{t}})^n = P^n-\mathbf{1}\pi^{\mathsf{t}}$ for $n\geqslant 1$, which implies the second equality after a simple calculation. ∎

Remark 2.2. Note that the group inverse does not in general coincide with the more well-known Moore–Penrose pseudoinverse A^+. In fact, it can be shown that if P is an irreducible stochastic matrix, then $(I-P)^{\mathsf{g}} = (I-P)^+$ if and only if the stationary distribution of P is uniform [19, Theorem 6.1]. △

3. An Edgeworth expansion for finite-state Markov chains

3.1. **Markov chain preliminaries.** Let $(W_k)_1^\infty$ be a homogeneous Markov chain on a finite state space $\mathcal{W}$, with transition probabilities given by the matrix $P = (P_{ij})_{i,j\in\mathcal{W}}$. (All matrices and vectors in this section will be indexed by $\mathcal{W}$. The reader that so prefers may without loss of generality assume that $\mathcal{W} = [m]$ for some integer m in this section.) For basic facts on Markov chains used below, see e.g. [22, Chapter 1].

We denote the distribution of W_k by

$$\pi_k = (\pi_{k;i})_{i\in\mathcal{W}} \quad \text{where} \quad \pi_{k;i} := \mathbb{P}(W_k = i), \quad i \in \mathcal{W}. \tag{3.1}$$

In particular, π_1 is the initial distribution of the Markov chain. We regard π_k as a column vector; recall that its transpose (a row vector) is denoted by π_k^{t}. It is well-known that

$$\pi_k^{\mathsf{t}} = \pi_1^{\mathsf{t}} P^{k-1}, \qquad k \geqslant 1. \tag{3.2}$$

We say that a sequence $i_0, \dots, i_\ell$ of elements of $\mathcal{W}$ is a *path* if it can appear with positive probability in the Markov chain, i.e., if $P_{i_k i_{k+1}} > 0$ for $0 \leqslant k < \ell$. We say that ℓ is the *length* of the path; we denote the length of a path $\mathcal{Q}$ by $\ell(\mathcal{Q})$. A *closed path* is a path $i_0, \dots, i_\ell$ such that $i_\ell = i_0$. We assume throughout this section that the Markov chain is *irreducible*, i.e., that

$$\text{for every pair } i, j \in \mathcal{W}, \text{ there exists a path } i = i_0, \dots, i_\ell = j \text{ from } i \text{ to } j. \tag{3.3}$$

We assume also that the Markov chain is *aperiodic*, i.e.,

$$\gcd\bigl\{\ell(\mathcal{Q}) : \mathcal{Q} \text{ is a closed path}\bigr\} = 1. \tag{3.4}$$

It is well-known that our assumptions (3.3)–(3.4) that the Markov chain is irreducible and aperiodic imply that there is a unique stationary distribution π, i.e., a distribution satisfying

$$\pi^{\mathsf{t}} = \pi^{\mathsf{t}} P. \tag{3.5}$$

Moreover, for any initial distribution π_1, we have

$$\pi_n \to \pi \qquad \text{as } n \to \infty. \tag{3.6}$$

3.2. **The main result.** Let $g : \mathcal{W} \to \mathbb{Z}$ be an integer-valued function, and define the integer-valued random variables $X_k := g(W_k)$ and

$$S_n := \sum_{k=1}^n X_k = \sum_{k=1}^n g(W_k). \tag{3.7}$$

Our goal is to prove an Edgeworth expansion for S_n.

We will, besides the aperiodicity condition (3.4), also assume a similar aperiodicity condition for the function g. For a path $\mathcal{Q}$ given by $i_0, \dots, i_\ell$, define the *value* of $\mathcal{Q}$ as

$$g(\mathcal{Q}) := \sum_{k=1}^{\ell} g(i_k) \in \mathbb{Z}. \tag{3.8}$$

We then assume:

$$\text{The set } \big\{(g(\mathcal{Q}),\ell(\mathcal{Q})) : \mathcal{Q} \text{ is a closed path}\big\} \text{ generates } \mathbb{Z}^2 \text{ as a group.} \tag{3.9}$$

Remark 3.1. It is easy to see, arguing similarly as in the proof of Proposition 6.8 below, that in (3.9), it is equivalent to consider only closed paths starting at any given state i_0. This implies the following probabilistic formulation of the condition: Start the Markov chain in state i_0 at time 0, and let $T := \min\{k \geqslant 1 : W_k = i_0\}$ be the time of the first return there. Then (3.9) is equivalent to:

$$\text{The random vector } (S_T, T) \text{ is not supported on a proper sublattice of } \mathbb{Z}^2. \tag{3.10}$$

This formulation is used e.g. by [12]. △

The following results will be proved in Section 5 . As noted in Section 2.3, there is a large literature on Edgeworth expansions, and in particular Nagaev [21] has proven several similar results for more general Markov chains (allowing also infinite state spaces); however, his theorems for the integer-valued (or, equivalently, lattice-valued) case assume instead of (3.9) a strong aperiodicity condition ([21, Condition C], see Example 6.7) which is not satisfied in our application. Hipp [12, Theorem (3.1)] has, by another method, proven a general result on approximation of the point probabilities using weaker assumptions than [21] and in particular under the same aperiodicity condition (3.9) as we. However, his result is not in a form directly applicable to our probabilities; it seems possible to derive our result from his with some extra work but we have not attempted this; we give instead a direct proof, which also leads to the explicit formulas (3.15)–(3.19) for the parameters (cumulants) in the approximation (3.14) (which are not explicit in [12]). Our proof is similar to the proofs in [21], but much simpler since we consider only finite state spaces; our argument is thus structurally closer to the classical proof of the i.i.d. result using characteristic functions than the arguments by [12].

Theorem 3.2. *Let $(W_k)_1^\infty$ be a stationary, irreducible and aperiodic homogeneous Markov chain on a finite state space $\mathcal{W}$, and let S_n be defined by* (3.7) *for some function $g : \mathcal{W} \to \mathbb{Z}$ such that* (3.9) *holds. Let*

$$\mu := \mathbb{E}\, X_1 = \mathbb{E}\, g(W_1), \tag{3.11}$$

$$\sigma^2 := \lim_{n\to\infty} n^{-1} \operatorname{Var}(S_n), \tag{3.12}$$

$$\varkappa_3 := \lim_{n\to\infty} n^{-1} \kappa_3(S_n). \tag{3.13}$$

(The limits exist and are finite under our assumptions.) Then $\sigma^2 > 0$, $\mathbb{E}\, S_n = n\mu$, and with $\Phi(x)$ and $\vartheta(x)$ as in (2.11) *and* (2.14),

$$\mathbb{P}(S_n - n\mu \leqslant x\sigma\sqrt{n}) = \Phi(x) + \frac{\varkappa_3}{6\sigma^3\sqrt{2\pi n}}(1-x^2)e^{-x^2/2} + \frac{1}{\sigma\sqrt{2\pi n}}\vartheta(x\sigma\sqrt{n}+n\mu)e^{-x^2/2} + O\big(n^{-1}\big), \tag{3.14}$$

uniformly in $x \in \mathbb{R}$ and $n \geqslant 1$. Moreover, μ, σ^2, and $\varkappa_3$ have the following explicit expressions:

$$\mu = \pi^{\mathsf{t}} G \mathbf{1} \tag{3.15}$$

$$\sigma^2 = \pi^{\mathsf{t}}\left(G^2 + 2GQPG\right)\mathbf{1} - \mu^2 \tag{3.16}$$

$$\varkappa_3 = \pi^{\mathsf{t}}\left(G^3 + 3GQPG^2 + 3G^2QPG + 6GQPGQPG\right)\mathbf{1}$$

$$- \mu \left(6\pi^{\mathsf{t}} G Q^2 P G \mathbf{1} + 3\sigma^2 + \mu^2\right), \tag{3.17}$$

where π is the stationary distribution of P and $G = \operatorname{Diag}(g(1), \dots, g(m))$ and $Q := (I-P)^{\mathsf{g}}$. Equivalently, with $Q' := Q - I = (I-P)^{\mathsf{g}} - I$, we have

$$\sigma^2 = \pi^{\mathsf{t}}(G^2 + 2GQ'G)\mathbf{1} + \mu^2 \tag{3.18}$$

$$\begin{aligned} \varkappa_3 &= \pi^{\mathsf{t}}(G^3 + 3GQ'G^2 + 3G^2Q'G + 6GQ'GQ'G)\mathbf{1} \\ &\quad + \mu\left(3\pi^{\mathsf{t}}G^2\mathbf{1} - 6\pi^{\mathsf{t}}GQ'^2G\mathbf{1} + 2\mu^2\right). \end{aligned} \tag{3.19}$$

The conditions (3.9) and $\sigma^2 > 0$ are discussed further in Section 6.2 and 6.3.
In particular, for the case $\mu = 0$ and $x = 0$ we obtain the following.

Corollary 3.3. *Under the assumptions of Theorem 3.2, if furthermore $\mu = 0$, then*

$$\mathbb{P}(S_n \leqslant 0) = \frac{1}{2} + \frac{\varkappa_3}{6\sigma^3\sqrt{2\pi n}} + \frac{1}{2\sigma\sqrt{2\pi n}} + O\big(n^{-1}\big) \tag{3.20}$$

and

$$\mathbb{P}(S_n < 0) = \frac{1}{2} + \frac{\varkappa_3}{6\sigma^3\sqrt{2\pi n}} - \frac{1}{2\sigma\sqrt{2\pi n}} + O\big(n^{-1}\big). \tag{3.21}$$

Consequently,

$$\mathbb{P}(S_n < 0) - \mathbb{P}(S_n > 0) = \frac{\varkappa_3}{3\sigma^3\sqrt{2\pi n}} + O\big(n^{-1}\big) \tag{3.22}$$

and

$$\mathbb{P}(S_n = 0) = \frac{1}{\sigma\sqrt{2\pi n}} + O\big(n^{-1}\big). \tag{3.23}$$

The local limit theorem (3.23) is here shown as a consequence of (3.20)–(3.21), and thus of (3.14). It can also easily be proved directly from the estimates in (5.37) and (5.52) below using Fourier inversion; furthermore, further terms can be obtained as in [6, Theorem IV.5], see [29, Theorem 2].

We can generalize the set-up above and consider a function $g : \mathcal{W} \times \mathcal{W} \to \mathbb{Z}$ of two variables; we then define

$$S_n := \sum_{k=1}^{n} g(W_{k-1}, W_k), \tag{3.24}$$

where we for convenience index the Markov chain as $(W_k)_0^\infty$. We then similarly define, if $\mathcal{Q}$ is the path $i_0, \dots, i_\ell$,

$$g(\mathcal{Q}) := \sum_{k=1}^{\ell} g(i_{k-1}, i_k) \in \mathbb{Z}. \tag{3.25}$$

Theorem 3.4. *Let $(W_k)_1^\infty$ be a stationary, irreducible and aperiodic homogeneous Markov chain on a finite state space $\mathcal{W}$, and let S_n be defined by (3.24) for some function $g : \mathcal{W}^2 \to \mathbb{Z}$ such that (3.9) holds. Then the conclusions of Theorem 3.2 and Corollary 3.3 hold. The explicit expressions for the moments are modified accordingly, see Section 8 for details.*

For further extensions, see Section 9.

4. Coin tossing

We consider in this section Litt's game, in the general version described in the introduction which is defined as follows.

Let $\mathcal{A}$ be a finite alphabet, with $q := |\mathcal{A}| \geqslant 2$ letters. (We are mainly interested in the case $q = 2$, and then we let $\mathcal{A} = \{\mathsf{H}, \mathsf{T}\}$ as in Litt's original formulation with coin tossing.) Let $\Xi_n = \xi_1 \cdots \xi_n$ be a random word with n letters $\xi_i \in \mathcal{A}$, chosen independently and uniformly at random. Hence, for any given word $a_1 \cdots a_n$, the probability

$$\mathbb{P}(\xi_1 \cdots \xi_n = a_1 \cdots a_n) = q^{-n}. \tag{4.1}$$

Fix two distinct words $A = a_1 \cdots a_\ell$ and $B := b_1 \cdots b_\ell$ of the same length $\ell \geqslant 1$. The letters in Ξ_n are drawn one by one; Alice scores a point when the last ℓ letters form A, and Bob scores a point when they form B. Denote their total scores by $S_{A,n}$ and $S_{B,n}$; we are interested in the difference $\widehat{S}_n := S_{A;n} - S_{B,n}$.

To put this into our framework, define for $1 \leqslant k \leqslant n - k + 1$ the subword

$$W_k := \xi_k \cdots \xi_{k+\ell-1} \in \mathcal{A}^\ell, \tag{4.2}$$

and for any word $U \in \mathcal{A}^\ell$, define the indicator

$$I_U(V) := \mathbf{1}_{U=V}, \qquad V \in \mathcal{A}^\ell. \tag{4.3}$$

Then the net score of the $(k + \ell - 1)$th draw (i.e., the kth draw that may score) is

$$X_k := I_A(W_k) - I_B(W_k), \tag{4.4}$$

and thus the final net score is

$$\widehat{S}_n := \sum_{k=1}^{n-\ell+1} X_k. \tag{4.5}$$

We may as well assume that the random letter ξ_i is defined for every $i \geqslant 1$; then also W_k, X_k, and S_n are defined for all $k \geqslant 1$ and $n \geqslant \ell$. It is obvious that the sequence of random words $(W_k)_1^\infty$ forms a stationary, homogeneous Markov chain with state space $\mathcal{A}^\ell$; furthermore, it is easy to see that this chain is irreducible and aperiodic. By comparing (4.5) and (3.7) we see that

$$\widehat{S}_n = S_{n-\ell+1} \tag{4.6}$$

for the function $g : \mathcal{A}^\ell \to \mathbb{Z}$ given by $g := I_A - I_B$. Note that for any fixed word $U \in \mathcal{A}^\ell$ and any $k \geqslant 1$ we have

$$\mathbb{E}\,[I_U(W_k)] = \mathbb{P}(W_k = U) = q^{-\ell}, \tag{4.7}$$

and thus

$$\mu := \mathbb{E}\, g(W_1) = \mathbb{E}\,[I_A(W_1)] - \mathbb{E}\,[I_B(W_1)] = q^{-\ell} - q^{-\ell} = 0. \tag{4.8}$$

We are thus in the setting of Theorem 3.2 and Corollary 3.3, provided that (3.9) holds. All that remains is thus to calculate σ^2 and $\varkappa_3$ and to verify (3.9).

We provide two methods, one a direct method that is specific to this application, and a second method that involves the explicit formulas (3.15)–(3.17) for the moments.

We introduce some more notation. For a pair of two words $U = u_1 \cdots u_\ell$ and $V = v_1 \cdots v_\ell$ of length ℓ, let

$$\Theta(U, V) := \{1 \leqslant k \leqslant \ell - 1 : u_{\ell-k+1} \cdots u_\ell = v_1 \cdots v_k\}, \tag{4.9}$$

i.e., the set of integers $k < n$ such that the words consisiting of the last k letters of U and the first k letters of V are the same; in other words, U and V may be concatenated with an overlap of k letters. (Note that this is not symmetric: in general $\Theta(U,V) \neq \Theta(V,U)$.) Note that we do not include ℓ in $\Theta(U,V)$, even if $U = V$. We define also the rational number, which may be called *overlap index*,

$$\theta_{UV} := \sum_{k\in\Theta(U,V)} q^{k-\ell} = q^{-\ell} \sum_{k\in\Theta(U,V)} q^k. \tag{4.10}$$

This is equivalent to the quantity $[U|V]$ defined by Basdevant et al. [1], which plays a key role in their analysis. More precisely, $\theta_{UV} = q^{-\ell}[U|V]$.

For two fixed words $U, V \in \mathcal{A}^\ell$, and $1 \leqslant j \leqslant \ell - 1$, we have $I_U(W_i)I_V(W_{i+j}) = 0$ unless $\ell - j \in \Theta(U,V)$, and it follows that, since W_i and W_{i+j} together contain $\ell + j$ random letters,

$$\mathbb{E}\left[I_U(W_i)I_V(W_{i+j})\right] = \mathbf{1}_{\ell-j\in\Theta(U,V)}\, q^{-(\ell+j)}. \tag{4.11}$$

Hence

$$\begin{aligned}\sum_{j=1}^{\ell-1} \mathbb{E}\left[I_U(W_i)I_V(W_{i+j})\right] &= \sum_{j=1}^{\ell-1} \mathbf{1}_{\ell-j\in\Theta(U,V)}\, q^{-(\ell+j)} \\ &= \sum_{i=1}^{\ell-1} \mathbf{1}_{i\in\Theta(U,V)}\, q^{i-2\ell} = q^{-\ell}\theta_{UV}. \end{aligned} \tag{4.12}$$

We return to S_n. We have $\mathbb{E}\, S_n = n\mu = 0$ by (4.8) and thus $\operatorname{Var} S_n = \mathbb{E}\,[S_n^2]$. Furthermore, by the definition (3.7) (or (4.5)–(4.6)) and expanding,

$$\mathbb{E}\,[S_n^2] = \sum_{i,j=1}^{n} \mathbb{E}\,[X_iX_j] = \sum_{i=1}^{n} \mathbb{E}\,[X_i^2] + 2\sum_{i=1}^{n}\sum_{k=1}^{n-i} \mathbb{E}\,[X_iX_{i+k}]. \tag{4.13}$$

If $k \geqslant \ell$, then W_i and W_{i+k} consist of different letters from Ξ, and thus they are independent, which implies $\mathbb{E}\,[X_iX_{i+k}] = \mathbb{E}\,[X_i]\,\mathbb{E}\,[X_{i+k}] = 0$; hence it suffices to take $k < \ell$ in the inner sum in (4.13). Furthermore, it is clear that for any fixed $k \geqslant 0$, the expectation $\mathbb{E}\,[X_iX_{i+k}]$ does not depend on i. Hence, (4.13) yields

$$\mathbb{E}\,[S_n^2] = n\,\mathbb{E}\,[X_1^2] + 2n\sum_{k=1}^{\ell-1} \mathbb{E}\,[X_1X_{1+k}] + O(1), \tag{4.14}$$

where the $O(1)$ comes from the missing terms with $0 \leqslant n - i < k < \ell$. Consequently, (3.12) yields

$$\sigma^2 = \mathbb{E}\,[X_1^2] + 2\sum_{k=1}^{\ell-1} \mathbb{E}\,[X_1X_{1+k}]. \tag{4.15}$$

Recalling (4.4), we see that

$$X_k^2 = I_A(W_k) + I_B(W_k). \tag{4.16}$$

Hence, (4.7) yields

$$\mathbb{E}\,[X_1^2] = \mathbb{E}\,[I_A(W_k)] + \mathbb{E}\,[I_B(W_k)] = 2q^{-\ell}. \tag{4.17}$$

Furthermore, (4.4) and (4.12) yield

$$\sum_{k=1}^{\ell-1} \mathbb{E}\,[X_1 X_{1+k}] = q^{-\ell}\bigl(\theta_{AA} - \theta_{AB} - \theta_{BA} + \theta_{BB}\bigr). \tag{4.18}$$

We conclude from (4.15) and (4.17)–(4.18) that

$$\sigma^2 = 2q^{-\ell}\bigl(1 + \theta_{AA} - \theta_{AB} - \theta_{BA} + \theta_{BB}\bigr). \tag{4.19}$$

For the third cumulant $\varkappa_3$ we argue similarly. First, since $\mathbb{E}\,S_n = 0$, we have $\kappa_3(S_n) = \mathbb{E}\,[S_n^3]$, and by expanding followed by combining equal terms,

$$\begin{aligned}\mathbb{E}\,[S_n^3] = \sum_{i,j,k=1}^{n} \mathbb{E}\,[X_i X_j X_k] = \sum_{1\leqslant i\leqslant n} \mathbb{E}\,[X_i^3] + 3 \sum_{1\leqslant i<j\leqslant n} \mathbb{E}\,[X_i^2 X_j] \\ + 3 \sum_{1\leqslant i<j\leqslant n} \mathbb{E}\,[X_i X_j^2] + 6 \sum_{1\leqslant i<j<k\leqslant n} \mathbb{E}\,[X_i X_j X_k]. \end{aligned} \tag{4.20}$$

Furthermore, all terms in the sums on the right-hand side with $j \geqslant i+\ell$ or $k \geqslant j+\ell$ vanish by independence, and the terms are invariant under a simultaneous shift of the indices. Hence (3.13) yields

$$\varkappa_3 = \mathbb{E}\,[X_1^3] + 3\sum_{s=1}^{\ell-1} \mathbb{E}\,[X_1^2 X_{1+s}] + 3\sum_{s=1}^{\ell-1} \mathbb{E}\,[X_1 X_{1+s}^2] + 6\sum_{s,t=1}^{\ell-1} \mathbb{E}\,[X_1 X_{1+s} X_{1+s+t}]. \tag{4.21}$$

We have $X_1^3 = X_1$, and thus $\mathbb{E}\,[X_1^3] = 0$. Furthermore, (4.4) and (4.16) together with (4.12) show that

$$\sum_{s=1}^{\ell-1} \mathbb{E}\,[X_1^2 X_{1+s}] = q^{-\ell}\bigl(\theta_{AA} - \theta_{AB} + \theta_{BA} - \theta_{BB}\bigr), \tag{4.22}$$

$$\sum_{s=1}^{\ell-1} \mathbb{E}\,[X_1 X_{1+s}^2] = q^{-\ell}\bigl(\theta_{AA} + \theta_{AB} - \theta_{BA} - \theta_{BB}\bigr). \tag{4.23}$$

Arguing as in (4.12) yields, for three fixed words $U, V, T \in \mathcal{A}^\ell$,

$$\begin{aligned}\sum_{j,k=1}^{\ell-1} \mathbb{E}\,\bigl[I_U(W_i) I_V(W_{i+j}) I_T(W_{i+j+k})\bigr] \\ = \sum_{j,k=1}^{\ell-1} \mathbf{1}_{\ell-j\in\Theta(U,V)} \mathbf{1}_{\ell-k\in\Theta(U,V)}\, q^{-(\ell+j+k)} \\ = q^{-\ell}\theta_{UV}\theta_{VT}. \end{aligned} \tag{4.24}$$

Hence, (4.4) yields

$$\begin{aligned}\sum_{s,t=1}^{\ell-1} \mathbb{E}\,[X_1 X_{1+s} X_{1+s+t}] = q^{-\ell}\bigl(\theta_{AA}\theta_{AA} - \theta_{AA}\theta_{AB} - \theta_{AB}\theta_{BA} + \theta_{AB}\theta_{BB} \\ - \theta_{BA}\theta_{AA} + \theta_{BA}\theta_{AB} + \theta_{BB}\theta_{BA} - \theta_{BB}\theta_{BB}\bigr) \\ = q^{-\ell}(\theta_{AA} - \theta_{BB})(\theta_{AA} + \theta_{BB} - \theta_{AB} - \theta_{BA}). \end{aligned} \tag{4.25}$$

Finally we obtain from (4.21) and (4.22)–(4.25), after some cancellations,

$$\varkappa_3 = 6q^{-\ell}(\theta_{AA} - \theta_{BB})(1 + \theta_{AA} + \theta_{BB} - \theta_{AB} - \theta_{BA}) = 3\sigma^2(\theta_{AA} - \theta_{BB}). \quad (4.26)$$

It is also possible to obtain this result from the explicit moment formulas. Denote by P_ℓ the transition matrix, indexed by all q^ℓ strings of length ℓ. Here, a transition $U \to V$ exists if and only if the last $\ell - 1$ characters of U coincide with the first $\ell - 1$ characters of V. It is easy to see that P_ℓ is primitive and the stationary distribution is uniform: $\pi = q^{-\ell}\mathbf{1}$; see (4.29) below. The only remaining difficulty in applying the formulas (3.15)–(3.17) is to calculate the group inverse $(I - P_\ell)^{\mathrm{g}}$. It turns out that the entries of this matrix coincide with the overlap indices θ_{UV}, up to a normalization.

Proposition 4.1. *The entries of the group inverse $Q := (I - P_\ell)^{\mathrm{g}}$ are given by*

$$(I - P_\ell)^{\mathrm{g}}_{U,V} = \mathbf{1}_{U=V} + \theta_{UV} - \ell q^{-\ell} \quad (4.27)$$

Proof. Introduce the notation $tail_k(U)$ for the string obtained by deleting the first k characters of U. Analogously for $head_k(U)$. It is straightforward to see that for any U, V we have

$$(P^j)_{U,V} = q^{-j}\mathbf{1}_{tail_j(U)=head_j(V)}, \qquad 0 \leqslant j < \ell, \quad (4.28)$$

$$(P^j)_{U,V} = q^{-\ell}, \qquad j \geqslant \ell. \quad (4.29)$$

Further,

$$\theta_{UV} = \sum_{k=1}^{l-1} q^{k-l}\mathbf{1}_{tail_{l-k}(U)=head_{l-k}(V)} = \sum_{j=1}^{l-1} q^{-j}\mathbf{1}_{tail_j(U)=head_j(V)} = \left(\sum_{j=1}^{\ell-1} P^j\right)_{U,V}. \quad (4.30)$$

At this point, the result follows by a simple calculation from the formula $(I - P)^{\mathrm{g}} = \lim_{t \nearrow 1}\bigl(\sum_{k\geqslant 0} t^k P^k - \mathbf{1}\pi^{\mathrm{t}}/(1-t)\bigr)$ in Proposition 2.1. ☐

In our case, the diagonal matrix G is given by $G_{AA} = 1$, $G_{BB} = -1$, and G is otherwise zero.

Clearly $\mu = 0$, by (3.11) or (3.15). For σ^2 and $\varkappa_3$ we use (3.18)–(3.19), noting that by Proposition 4.1, $Q'_{U,V} = \theta_{UV} - \ell q^{-\ell}$. Hence,

$$\pi^{\mathrm{t}}GQ'G\mathbf{1} = q^{-\ell}\mathbf{1}^{\mathrm{t}}GQ'G\mathbf{1} = \theta_{AA} + \theta_{BB} - \theta_{AB} - \theta_{BA}. \quad (4.31)$$

Furthermore, evidently $\pi^{\mathrm{t}}G^2\mathbf{1} = 2q^{-\ell}$. Hence, (3.18) yields

$$q^\ell\sigma^2 = \mathbf{1}^{\mathrm{t}}G^2\mathbf{1} + 2\mathbf{1}^{\mathrm{t}}GQ'G\mathbf{1} + 0 \quad (4.32)$$

$$= 2 + 2(\theta_{AA} + \theta_{BB} - \theta_{AB} - \theta_{BA}) \quad (4.33)$$

in agreement with (4.19). The calculation of the third cumulant (4.26) by this method is similar, and the details are omitted.

We now obtain a preliminary result for Litt's game. This will be improved to Theorem 1.1 in Section 7 where we identify the cases where the condition (3.9) does not hold.

Theorem 4.2. *Let Alice and Bob play Litt's game above with distinct words A and B of the same length ℓ in an alphabet $\mathcal{A}$ with q letters, and assume that n letters* *are chosen at random, uniformly and independently. Assume also that* (3.9) *holds. Then, with* θ_{UV} *and* σ^2 *given by* (4.10) *and* (4.19), $\sigma^2 > 0$ *and*

$$\mathbb{P}(\text{Alice wins}) = \frac{1}{2} + \frac{\theta_{BB} - \theta_{AA} - 1}{2\sqrt{2\pi\sigma^2}} n^{-1/2} + O(n^{-1}), \tag{4.34}$$

$$\mathbb{P}(\text{Bob wins}) = \frac{1}{2} + \frac{\theta_{AA} - \theta_{BB} - 1}{2\sqrt{2\pi\sigma^2}} n^{-1/2} + O(n^{-1}), \tag{4.35}$$

$$\mathbb{P}(\text{Tie}) = \frac{1}{\sqrt{2\pi\sigma^2}} n^{-1/2} + O(n^{-1}), \tag{4.36}$$

and thus

$$\mathbb{P}(\text{Alice wins}) - \mathbb{P}(\text{Bob wins}) = \frac{\theta_{BB} - \theta_{AA}}{\sqrt{2\pi\sigma^2}} n^{-1/2} + O(n^{-1}). \tag{4.37}$$

Proof. A consequence of Corollary 3.3, recalling (4.5) and using (4.19) and (4.26) above. □

Example 4.3 (HH vs HT)**.** Litt's original game is the case $\mathcal{A} = \{\mathsf{H}, \mathsf{T}\}$, $q = 2$, $\ell = 2$, $A = \mathsf{HH}$, $B = \mathsf{HT}$ of Theorem 1.1. We have $\Theta(A, A) = \Theta(A, B) = \{1\}$ and $\Theta(B, A) = \Theta(B, B) = \emptyset$, and thus

$$\theta_{AA} = \theta_{AB} = \tfrac{1}{2}, \quad \theta_{BA} = \theta_{BB} = 0. \tag{4.38}$$

Hence, (4.19) yields $\sigma^2 = \frac{1}{2}$, and (1.5)–(1.8) yield (1.1)–(1.2), with error terms $O(n^{-1})$. △

Remark 4.4. We have here represented the score $\widehat{S}_n$ using a (finite-state) Markov chain. We may also note that the sequence X_k is $(\ell - 1)$-dependent, which means that general results for sums of m-dependent variables can be applied to $\widehat{S}_n$. See e.g. [11; 24; 18] for some related results; however, we have not been able to find a general result that applies to our situation. △

5. Proof of Theorems 3.2–3.4

5.1. **A lemma.** We will use the following simple uniform version of the spectral radius formula (2.4). We do not know a reference so we give a proof for completeness.

Lemma 5.1. *Let* $z \mapsto A(z)$ *be a continuous square-matrix-valued function defined on some compact set* K*. Suppose that for some* $r > 0$*, we have* $\rho(A(z)) < r$ *for every* $z \in K$*. Then there exists* $\tilde{r} < r$ *such that*

$$\|A(z)^n\| \leqslant C\tilde{r}^n \tag{5.1}$$

for some constant C*, uniformly for all* $z \in K$ *and* $n \geqslant 1$*.*

Proof. Let $z \in K$, and choose r_z with $\rho(A(z)) < r_z < r$. Then the assumption and (2.4) show that there exists N such that $\|A(z)^N\| < r_z^N$. Since $z \mapsto A(z)$ is continuous, and the operator norm is a continuous functional, it follows that there exists an open neighbourhood U_z of z such that for all $w \in U_z$,

$$\|A(w)^N\| < r_z^N. \tag{5.2}$$

Since $\|AB\| \leqslant \|A\| \cdot \|B\|$ for two matrices A and B, it follows that for any $w \in U_z$ and all $n \geqslant 1$, by writing $n = kN + \ell$ with $0 \leqslant \ell < N$,

$$\|A(w)^n\| \leqslant \|A(w)^N\|^k \|A(w)^\ell\| \leqslant r_z^{kN} \|A(w)^\ell\| \leqslant r_z^n \max_{0 \leqslant \ell < N} \bigl(r_z^{-\ell} \|A(w)^\ell\|\bigr). \tag{5.3}$$

For fixed ℓ, $\|A(w)^\ell\|$ is a continuous function of z, and is thus bounded in K. Consequently (5.3) implies that

$$\|A(w)^n\| \leqslant C_z r_z^n \tag{5.4}$$

holds for some C_z, uniformly for $w \in U_z$ and $n \geqslant 1$. We may cover the compact set K by a finite number of such open sets U_{z_i}, $i = 1, \dots, M$; thus (5.1) follows for all $z \in K$ with $C := \max_i C_{z_i}$ and $\tilde{r} := \max_i r_{z_i} < r$. □

5.2. **Proof of Theorem 3.2.** We assume in this subsection the assumptions of Theorem 3.2.

For complex $z \neq 0$, define the matrix

$$P(z) = (P(z)_{ij})_{i,j\in\mathcal{W}} \quad \text{where} \quad P(z)_{ij} := P_{ij} z^{g(j)} \tag{5.5}$$

and the vector

$$\pi_1(z) = (\pi_1(z)_i)_{i\in\mathcal{W}} \quad \text{where} \quad \pi_1(z)_i := \pi_{1;i} z^{g(i)}. \tag{5.6}$$

Note that $P(1) = P$ and $\pi_1(1) = \pi_1$. Furthermore, the matrix-valued function $P(z)$ and the vector-valued function $\pi_1(z)$ are analytic functions of $z \neq 0$.

Let

$$G_n(z) := \mathbb{E}\, z^{S_n}, \tag{5.7}$$

i.e., the probability generating function of S_n (in a generalized sense since S_n may take both positive and negative integer values). Then $G_n(z)$ is a well-defined analytic (in fact, rational) function for complex $z \neq 0$, since S_n takes only a finite number of values for each n. Note that for real t,

$$G_n(e^{\mathrm{i}t}) = \mathbb{E}\, e^{\mathrm{i}tS_n} = \varphi_{S_n}(t), \tag{5.8}$$

the characteristic function of S_n.

We will use the following well-known representation. Recall that $\mathbf{1} = (1, \dots, 1)$, the (column) vector with all entries 1.

Lemma 5.2. *For any $z \neq 0$ and $n \geqslant 1$,*

$$G_n(z) = \pi_1(z)^{\mathrm{t}} P(z)^{n-1} \mathbf{1}. \tag{5.9}$$

Proof. Since $\mathbb{P}(W_1 = i_1, \dots, W_n = i_n) = \pi_{1;i_1} P_{i_1 i_2} P_{i_2 i_3} \cdots P_{i_{n-1} i_n}$, we have

$$G_n(z) = \mathbb{E} \prod_{k=1}^n z^{g(W_k)} = \sum_{i_1,\dots,i_n\in\mathcal{W}} \pi_{1;i_1} z^{g(i_1)} P_{i_1 i_2} z^{g(i_2)} \cdots P_{i_{n-1} i_n} z^{g(i_n)}, \tag{5.10}$$

which equals the right-hand side of (5.9). □

We aim at finding good estimates of the characteristic function $\varphi_{S_n}(t) = G_n(e^{\mathrm{i}t})$. We consider first small t. The following asymptotic formula is central in our arguments.

Lemma 5.3. *If $\delta \in (0, 1)$ is small enough, then there exist analytic functions $\eta(z)$ and $\lambda(z)$ in $D_\delta := D(1, \delta)$ and a constant $c \in (0, 1)$ such that for all $z \in D_\delta$ and $n \geqslant 1$*

$$G_n(z) = \eta(z)\lambda(z)^n \bigl(1 + O(c^n)\bigr) \tag{5.11}$$

and, somewhat more precisely,

$$G_n(z) = \eta(z)\lambda(z)^n \bigl(1 + O(c^n|z - 1|)\bigr). \tag{5.12}$$

Furthermore, $\lambda(z)$ is an eigenvalue of $P(z)$, and $\eta(1) = \lambda(1) = 1$.

Proof. Denote the set of eigenvalues of $P(z)$ by $\Lambda(z)$.

Consider first $z = 1$. The matrix $P(1) = P$ is stochastic, which means that $P\mathbf{1} = \mathbf{1}$, i.e., $\mathbf{1}$ is a right eigenvector with eigenvalue 1. Moreover, by the Perron–Frobenius theorem and our assumptions (3.3)–(3.4), this eigenvalue is simple and all other eigenvalues λ_i of P satisfy $|\lambda_i| < 1$. Let

$$\rho' := \max\{|\lambda| : \lambda \in \Lambda(1) \setminus 1\} < 1. \tag{5.13}$$

Let also $\rho_0 = (2\rho' + 1)/3$ and $\rho_1 := (1 - \rho')/3$. Thus $0 < \rho_0 < 1 - \rho_1 < 1$. Furthermore, $P(1)$ has exactly one eigenvalue (viz. 1) in the open disc $D(1, \rho_1)$, and all other eigenvalues in the (disjoint) open disc $D(0, \rho_0)$. The eigenvalues $\Lambda(z)$ are the roots of the characteristic polynomial of $P(z)$, and the coefficients in this polynomial are continuous (in fact, analytic) functions of z. Hence it follows that there exists a small $\delta \in (0, 1)$ such that if $z \in D_\delta$ (i.e., $|z - 1| < \delta$), then $P(z)$ has exactly 1 simple eigenvalue in $D(1, \rho_1)$, and all other eigenvalues in $D(0, \rho_0)$. For $z \in D_\delta$, denote the eigenvalue in $D(1, \rho_1)$ by $\lambda(z)$. Thus, for $z \in D_\delta$,

$$|\lambda(z) - 1| < \rho_1 \quad \text{and} \quad |\lambda(z)| > 1 - \rho_1 > \tfrac{2}{3}. \tag{5.14}$$

Since $\lambda(z)$ is a simple root of the characteristic polynomial of $P(z)$, it follows from the implicit function theorem that $\lambda(z)$ is an analytic function of $z \in D_\delta$. Moreover, provided δ is chosen small enough, the corresponding left and right eigenvectors $u(z)^{\mathsf{t}}$ and $v(z)$ can (and will) be normalized by

$$u(z)^{\mathsf{t}}\mathbf{1} = 1 = u(z)^{\mathsf{t}}v(z), \tag{5.15}$$

for every $z \in D_\delta$, and then $u(z)^{\mathsf{t}}$ and $v(z)$ are analytic functions of $z \in D_\delta$. Note that, since $P(1) = P$, it follows from (3.5) and $P\mathbf{1} = \mathbf{1}$ that $\lambda(1) = 1$ with normalized eigenvectors

$$u(1)^{\mathsf{t}} = \pi^{\mathsf{t}} \quad \text{and} \quad v(1) = \mathbf{1}. \tag{5.16}$$

Let $z \in D_\delta$ and let

$$\Pi(z) := v(z)u(z)^{\mathsf{t}}; \tag{5.17}$$

in other words, the matrix $\Pi(z)$ defines the operator $v \mapsto (u(z)^{\mathsf{t}}v)v(z)$, which is a projection onto the eigenspace spanned by $v(z)$; in particular,

$$\Pi(z)^2 = \Pi(z). \tag{5.18}$$

Moreover, $\Pi(z)$ commutes with $P(z)$ and

$$P(z)\Pi(z) = \Pi(z)P(z) = \lambda(z)\Pi(z). \tag{5.19}$$

Consequently, by elementary spectral theory,

$$\widetilde{P}(z) := P(z) - \lambda(z)\Pi(z) = (I - \Pi(z))P(z) \tag{5.20}$$

has the set of eigenvalues $\Lambda(z) \setminus \{\lambda(z)\} \cup \{0\}$. This set is contained in $D(0, \rho_0)$ and thus by the definition (2.3) the spectral radius

$$\rho(\widetilde{P}(z)) < \rho_0. \tag{5.21}$$

Consequently, the spectral radius formula (2.4) implies that for some constant $C = C(z)$

$$\|\widetilde{P}(z)^n\| \leqslant C\rho_0^n. \qquad n \geqslant 1, \tag{5.22}$$

By decreasing δ a little, we may assume that (5.22) holds on $\overline{D_\delta}$, and then, by Lemma 5.1, (5.22) holds uniformly for all $z \in D_\delta$ with some $C = C(\delta)$.

By (5.19)–(5.20), $\widetilde{P}(z)$ and $\Pi(z)$ commute, and $\widetilde{P}(z)\Pi(z) = \Pi(z)\widetilde{P}(z) = 0$. Hence, (5.20) and (5.18) imply

$$P(z)^n = \bigl(\lambda(z)\Pi(z) + \widetilde{P}(z)\bigr)^n = \lambda(z)^n\Pi(z) + \widetilde{P}(z)^n. \tag{5.23}$$

Consequently, Lemma 5.2 yields, recalling (5.14) and (5.15), for $z \in D_\delta$,

$$\begin{aligned} G_n(z) &= \pi_1(z)^{\mathsf{t}}\bigl(\lambda(z)^{n-1}\Pi(z) + \widetilde{P}(z)^{n-1}\bigr)\mathbf{1} \\ &= \lambda(z)^{n-1}\pi_1(z)^{\mathsf{t}}\Pi(z)\mathbf{1} + O(\rho_0^{n-1}) \\ &= \lambda(z)^{n-1}(\pi_1(z)^{\mathsf{t}}v(z))(u(z)^{\mathsf{t}}\mathbf{1}) + O(\rho_0^n), \\ &= \lambda(z)^{n-1}\Bigl(\pi_1(z)^{\mathsf{t}}v(z) + O\bigl(c^n\bigr)\Bigr) \end{aligned} \tag{5.24}$$

with $c := \rho_0/(1-\rho_1) < 1$; note that the O terms are uniform for $z \in D_\delta$ since (5.22) is and $\pi_1(z)$ is bounded for $z \in D_\delta$. We may, by decreasing δ if necessary, assume that $|\pi_1(z)^{\mathsf{t}}v(z)| > 1/2$ for $z \in D_\delta$, and then (5.24) yields (5.11) with

$$\eta(z) := \lambda(z)^{-1}\pi_1(z)^{\mathsf{t}}v(z). \tag{5.25}$$

We have noted $\lambda(1) = 1$, and thus (5.25) and (5.16) yield $\eta(1) = \pi_1^{\mathsf{t}}\mathbf{1} = 1$.

Finally, (5.11) can be written

$$\left|\frac{G_n(z)}{\eta(z)\lambda(z)^n} - 1\right| \leqslant Cc^n, \qquad z \in D_\delta,\ n \geqslant 1. \tag{5.26}$$

The function $G_n(z)/\bigl(\eta(z)\lambda(z)^n\bigr) - 1$ on the left-hand side is analytic and vanishes at $z = 1$; hence we can divide it by $z - 1$ and obtain, by the maximum principle,

$$\left|\frac{G_n(z)/(\eta(z)\lambda(z)^n) - 1}{z-1}\right| \leqslant Cc^n, \qquad z \in D_\delta,\ n \geqslant 1, \tag{5.27}$$

which is (5.12). □

We may assume (and actually already have assumed in the proof) that δ is so small that $\eta(z) \neq 0$ and $\lambda(z) \neq 0$ in D_δ, and thus $\log\eta(z)$ and $\log\lambda(z)$ are defined there. Let δ_0 be so small that if z is a complex number with $|z| < \delta_0$, then $e^{\mathrm{i}z} \in D_\delta$. We then can define the analytic functions, for $|z| < \delta_0$,

$$\psi(z) := \log\lambda\bigl(e^{\mathrm{i}z}\bigr) \quad\text{and}\quad \gamma(z) := \log\eta\bigl(e^{\mathrm{i}z}\bigr). \tag{5.28}$$

We obtain from (5.12) the estimate

$$\log G_n(e^{\mathrm{i}z}) = n\psi(z) + \gamma(z) + O\bigl(|z|c^n\bigr), \qquad |z| < \delta_0. \tag{5.29}$$

Note that, recalling (5.8), (5.29) resembles the elementary decomposition of the characteristic function of the sum of i.i.d. variables as a power of the characteristic function of an individual variable, but we have here also two “error terms”.

Lemma 5.4. *The cumulants of S_n are given by*

$$\kappa_m(S_n) = \mathrm{i}^{-m}\psi^{(m)}(0)n + O(1) = \varkappa_m n + O(1) \tag{5.30}$$

for every $m \geqslant 1$, where

$$\varkappa_m := \mathrm{i}^{-m}\psi^{(m)}(0). \tag{5.31}$$

(The implicit constant may depend on m, but not on n.) In particular,

$$\frac{\kappa_m(S_n)}{n} \to \varkappa_m \quad \text{as } n \to \infty. \tag{5.32}$$

Proof. Since the functions in (5.29) are analytic, we may differentiate an arbitrary number of times, and obtain by (2.5) and Cauchy's estimate, for every $m \geqslant 1$,

$$\mathrm{i}^m \kappa_m(S_n) = \frac{\mathrm{d}^m}{\mathrm{d}t^m} \log G_n(e^{\mathrm{i}t})\big|_{t=0} = n\psi^{(m)}(0) + \gamma^{(m)}(0) + O\big(c^n\big), \tag{5.33}$$

which is a more precise form of (5.30)–(5.31). □

We have so far allowed any initial distribution π_1, but we now, for simplicity, assume that the Markov chain $(W_n)_1^\infty$ is stationary, i.e., that the initial distribution $\pi_1 = \pi$, and thus $\pi_k = \pi$ for every k by (3.2) and (3.5). Then the random variables X_k have the same distribution. As in (3.11), denote their mean by

$$\mu := \mathbb{E}\, X_k. \tag{5.34}$$

Corollary 5.5. *Suppose that the Markov chain $(W_n)_1^\infty$ is stationary. Then*

$$\varkappa_1 = \mu, \qquad \psi'(0) = \mathrm{i}\mu, \qquad \gamma'(0) = 0. \tag{5.35}$$

Proof. Since the random variables X_k have the same distribution, (2.7) yields

$$\kappa_1(S_n) = \mathbb{E}\, S_n = \sum_{k=1}^n \mathbb{E}\, X_k = n\mu. \tag{5.36}$$

The result follows from the case $m = 1$ of (5.32) and (5.33). □

We are now prepared to prove the estimate of $\varphi_{S_n}(t)$ that we need for small t.

Lemma 5.6. *Suppose that the Markov chain $(W_n)_1^\infty$ is stationary. Then there exists $\delta > 0$ such that, if $|t| \leqslant \delta$, then*

$$\varphi_{S_n}(t) = e^{\mathrm{i}n\mu t - n\varkappa_2 t^2/2}\Big(1 - \mathrm{i}n\frac{\varkappa_3}{6}t^3\Big) + O\big((t^2 + n^2t^6)e^{-n\varkappa_2 t^2/4}\big). \tag{5.37}$$

Proof. Let $\delta < \delta_0$. Then (5.8), (5.29) and Cauchy's estimate yield, for $|t| \leqslant \delta$,

$$\frac{\mathrm{d}^4}{\mathrm{d}t^4} \log \varphi_{S_n}(t) = \frac{\mathrm{d}^4}{\mathrm{d}t^4} \log G_n(e^{\mathrm{i}t}) = n\psi^{(4)}(t) + \gamma^{(4)}(t) + O\big(c^n\big) = O(n). \tag{5.38}$$

Consequently, a Taylor expansion as in (2.6) yields, using $\varphi_{S_n}(1) = 1$, (5.36) and (5.30), for $|t| \leqslant \delta$ and $n \geqslant 1$,

$$\begin{aligned}\log \varphi_{S_n}(t) &= \mathrm{i}\kappa_1(S_n)t - \frac{\kappa_2(S_n)}{2}t^2 - \mathrm{i}\frac{\kappa_3(S_n)}{6}t^3 + O(nt^4) \\ &= \mathrm{i}n\mu t - n\frac{\varkappa_2}{2}t^2 - \mathrm{i}n\frac{\varkappa_3}{6}t^3 + O(t^2 + nt^4).\end{aligned} \tag{5.39}$$

Hence,

$$e^{-\mathrm{i}n\mu t + n\varkappa_2 t^2/2}\varphi_{S_n}(t) = \exp\Big(-\mathrm{i}n\frac{\varkappa_3}{6}t^3 + O(t^2 + nt^4)\Big). \tag{5.40}$$

If δ is small enough, then the real part of the argument of the exponential function in (5.40) is less than $n\varkappa_2 t^2/4 + C$, and consequently, by a Taylor expansion,

$$e^{-\mathrm{i}n\mu t + n\varkappa_2 t^2/2}\varphi_{S_n}(t) = 1 - \mathrm{i}n\frac{\varkappa_3}{6}t^3 + O(t^2 + nt^4) + O\Big(\big(n|t|^3 + t^2\big)^2 e^{n\varkappa_2 t^2/4}\Big), \tag{5.41}$$

which yields (5.37), recalling that $|t| = O(1)$ and noting $nt^4 \leqslant (t^2 + n^2t^6)/2$. □

We turn to estimating $\varphi_{S_n}(t)$ for larger t, and again begin by studying the matrix $P(z)$. This is where we use our assumption (3.9).

Lemma 5.7. *If* $0 < |t| \leqslant \pi$, *then*

$$\rho\big(P(e^{\mathrm{i}t})\big) < 1. \tag{5.42}$$

Proof. Let λ be an eigenvalue of $P(e^{\mathrm{i}t})$, and let $u^{\mathrm{t}} = (u_j)_1^n$ be a corresponding left eigenvector. Then

$$\lambda u_k = \sum_{j\in\mathcal{W}} u_j P(e^{\mathrm{i}t})_{jk} = \sum_{j\in\mathcal{W}} u_j P_{jk} e^{\mathrm{i}g(k)t}, \qquad k \in \mathcal{W}, \tag{5.43}$$

and thus, by the triangular inequality,

$$|\lambda|\,|u_k| \leqslant \sum_{j\in\mathcal{W}} |u_j| P_{jk}, \qquad k \in \mathcal{W}. \tag{5.44}$$

Summing over all $k \in \mathcal{W}$ yields, since (P_{jk}) is a stochastic matrix,

$$|\lambda| \sum_{k\in\mathcal{W}} |u_k| \leqslant \sum_{j\in\mathcal{W}} |u_j| \sum_{k\in\mathcal{W}} P_{jk} = \sum_{j\in\mathcal{W}} |u_j|. \tag{5.45}$$

Consequently, $|\lambda| \leqslant 1$.

Suppose now, to obtain a contradiction, that $|\lambda| = 1$. We then have equality in (5.45), and thus in (5.44) for every k. Note first that since (P_{jk}) is irreducible, it follows easily from equality in (5.44) that we have $|u_k| > 0$ for every $k \in \mathcal{W}$. Furthermore, equality when applying the triangle inequality to (5.43) implies

$$\frac{u_j e^{\mathrm{i}g(k)t}}{\lambda u_k} > 0 \tag{5.46}$$

for all $j, k \in \mathcal{W}$ such that $P_{jk} > 0$. It follows that for any path $\mathcal{Q}$ given by $i_0, \dots, i_\ell$,

$$\frac{u_{i_0} e^{\mathrm{i}g(\mathcal{Q})t}}{\lambda^{\ell(\mathcal{Q})} u_{i_\ell}} = \prod_{k=1}^{\ell} \frac{u_{i_{k-1}} e^{\mathrm{i}g(i_k)t}}{\lambda u_{i_k}} > 0 \tag{5.47}$$

and in particular, for a closed path $\mathcal{Q}$

$$e^{\mathrm{i}g(\mathcal{Q})t}\lambda^{-\ell(\mathcal{Q})} = \frac{e^{\mathrm{i}g(\mathcal{Q})t}}{\lambda^{\ell}(\mathcal{Q})} > 0. \tag{5.48}$$

Furthermore, $|e^{\mathrm{i}g(\mathcal{Q})t}| = |\lambda^{\ell}(\mathcal{Q})| = 1$, and thus, for every closed path $\mathcal{Q}$

$$e^{\mathrm{i}g(\mathcal{Q})t}\lambda^{-\ell(\mathcal{Q})} = 1. \tag{5.49}$$

The set of all $(k, \ell) \in \mathbb{Z}^2$ such that

$$e^{\mathrm{i}kt}\lambda^{-\ell} = 1 \tag{5.50}$$

is a subgroup, and thus it follows from (3.9) and (5.49) that (5.50) holds for all $(k, \ell) \in \mathbb{Z}^2$. Taking $(k, \ell) = (1, 0)$, this shows $e^{\mathrm{i}t} = 1$, which contradicts the assumption on t.

This contradiction shows that $|\lambda| < 1$ for every eigenvalue λ of $P(e^{\mathrm{i}t})$, which is the same as (5.42) by (2.3). ◻

Lemma 5.8. *For every* $\delta > 0$, *there exist* $r = r(\delta) < 1$ *and* C *such that if* $\delta \leqslant |t| \leqslant \pi$ *and* $n \geqslant 1$, *then*

$$\|P(e^{\mathrm{i}t})^n\| \leqslant Cr^n. \tag{5.51}$$

Proof. A consequence of Lemmas 5.7 and 5.1, taking $K := \{t \in \mathbb{R} : \delta \leqslant |t| \leqslant \pi\}$. ◻

Lemma 5.9. *For every $\delta > 0$, there exist $r = r(\delta) < 1$ and C such that if $\delta \leqslant |t| \leqslant \pi$ and $n \geqslant 1$, then*

$$\bigl|\varphi_{S_n}(t)\bigr| \leqslant Cr^n. \tag{5.52}$$

Proof. A consequence of Lemmas 5.8 and 5.2 together with (5.8). ☐

Proof of Theorem 3.2. The limits (3.12) and (3.13) exist by (5.32); note that $\sigma^2 = \varkappa_2$. We postpone the proof that $\sigma^2 > 0$ to Corollary 6.10 in Section 6. Define the normalized variable $\widetilde{S}_n := (S_n - n\mu)/(\sigma\sqrt{n})$ which has mean $\mathbb{E}\,\widetilde{S}_n = 0$ and variance $\operatorname{Var}(\widetilde{S}_n) = 1 + O(1/n)$ by (5.36) and (5.30).

Using the estimates of $\varphi_{S_n}(t)$ in Lemmas 5.6 and 5.9, the result (3.14) follows as the classical corresponding result for sums of i.i.d. random variables [6, Theorem IV.3]. There are actually two proofs of this theorem in [6]. We use the second proof in [6, pp. 64–65], taking there

$$f_n(t) := \varphi_{\widetilde{S}_n}(t) = e^{-\mathrm{i}\mu\sqrt{n}t/\sigma}\varphi_{S_n}\Bigl(\frac{t}{\sigma\sqrt{n}}\Bigr). \tag{5.53}$$

We further take $T_{3\,n}$ there as $\delta\sqrt{n}$ with the same δ as in Lemma 5.6, and we have $d = 1$ and $t_0 = 2\pi$. A crucial part of the estimate is that Lemma 5.6 yields

$$\varphi_{\widetilde{S}_n}(t) = e^{-t^2/2}\Bigl(1 - \mathrm{i}\frac{\varkappa_3}{6\sigma^3\sqrt{n}}t^3\Bigr) + O\bigl(n^{-1}(t^2 + t^6)e^{-t^2/4}\bigr), \qquad |t| \leqslant \delta\sigma\sqrt{n}, \tag{5.54}$$

and thus, with $g(t) := e^{-t^2/2}\bigl(1 - \mathrm{i}\frac{\varkappa_3}{6\sigma^3\sqrt{n}}t^3\bigr)$,

$$\int_{-\delta\sigma\sqrt{n}}^{\delta\sigma\sqrt{n}} \left|\frac{\varphi_{\widetilde{S}_n}(t) - g(t)}{t}\right| \mathrm{d}t = O\bigl(n^{-1}\bigr), \tag{5.55}$$

which implies that $I_2' = O(n^{-1})$ on [6, p. 65]. The rest of the estimates in [6, pp. 64–65] hold without changes, and we obtain (3.14); we omit the details. We may also use the first proof in [6, pp. 57–59], again using (5.55); however, this yields (3.14) with the slightly weaker error term $O\bigl(\frac{\log n}{n}\bigr)$ from the estimates of I_k' and ε_1 on [6, p. 59].

It remains to verify the matrix-based cumulant formulas (3.15)–(3.19). By Lemma 5.4 and (5.28), the cumulants are given in terms of the first three derivatives of $\psi(z) = \log\lambda(e^{iz})$ at $z = 0$. By analyticity of $\lambda(e^{iz})$, it is enough to calculate the derivatives of $\log\lambda(e^t)$ where t is real. This can be done by repeated differentiation of the eigenvalue equation and some algebra, see Appendix A, but we will here instead make use of [10], which studies this problem in great generality; in particular, [10, Theorem 4.1] (with a correction of the sign of ρ_3) provides for P as above and any matrix E and small t, the Taylor series

$$\lambda(P + tE) = 1 + t\rho_1(E) + t^2\rho_2(E) + t^3\rho_3(E) + O(t^4), \tag{5.56}$$

where $\lambda(P+tE)$ denotes the largest eigenvalue, and ρ_i are the following expressions:

$$\rho_1(E) = \pi^{\mathsf{t}}E\mathbf{1}, \tag{5.57}$$

$$\rho_2(E) = \pi^{\mathsf{t}}EQE\mathbf{1}, \tag{5.58}$$

$$\rho_3(E) = \pi^{\mathsf{t}}EQEQE\mathbf{1} - \rho_1(E)\cdot\pi^{\mathsf{t}}EQ^2E\mathbf{1}. \tag{5.59}$$

Here, as before, we set $Q := (I - P)^{\mathsf{g}}$. (Recall that since P is irreducible, it has a simple largest eigenvalue; by continuity, this holds also for all small perturbations.) We will use this result in a different, more general form. The expansion (5.56) holds uniformly for all matrices E in a bounded set. (This follows e.g. since the implicit function theorem implies that the eigenvalue $\lambda(P+tE)$ is (for small t) an analytic function of the entries in tE.) Consequently, (5.56) holds also for a continuous matrix-valued function $E(t)$. Let $F(t)$ be a smooth one-parameter family of matrices with $F(0)=P$ and take $E(t):=(F(t)-F(0))/t$ (with $E(0):=F'(0)$); then (5.56) gives an expansion of $\lambda(F(t))$. We want to extract the first three derivatives at the origin. For this purpose, it is sufficient to replace $F(t)$ with its third-order Taylor series $P+tF'(0)+\frac{1}{2}t^2F''(0)+\frac{1}{6}t^3F'''(0)$, which gives

$$E(t):=F'(0)+\frac{1}{2}tF''(0)+\frac{1}{6}t^2F'''(0). \tag{5.60}$$

By substituting (5.60) in (5.56)–(5.59), we obtain

$$\begin{aligned}\lambda(F(t))&=1+t\rho_1(E(t))+t^2\rho_2(E(t))+t^3\rho_3(E(t))+O(t^4)\\&=1+m_1t+\tfrac{1}{2}m_2t^2+\tfrac{1}{6}m_3t^3+O(t^4),\end{aligned} \tag{5.61}$$

where, collecting terms,

$$m_1:=[t^1]\lambda(F(t))=[t^0]\rho_1(E(t))=\pi^{\mathsf{t}}F'(0)\mathbf{1}, \tag{5.62}$$

$$\begin{aligned}\tfrac{1}{2}m_2&:=[t^2]\lambda(F(t))=[t^1]\rho_1(E(t))+[t^0]\rho_2(E(t))\\&=\tfrac{1}{2}\pi^{\mathsf{t}}F''(0)\mathbf{1}+\pi^{\mathsf{t}}F'(0)QF'(0)\mathbf{1},\end{aligned} \tag{5.63}$$

$$\begin{aligned}\tfrac{1}{6}m_3&:=\lambda(F(t))[t^3]=[t^2]\rho_1(E(t))+[t^1]\rho_2(E(t))+[t^0]\rho_3(E(t))\\&=\tfrac{1}{6}\pi^{\mathsf{t}}F'''(0)\mathbf{1}+\tfrac{1}{2}\pi^{\mathsf{t}}F'(0)QF''(0)\mathbf{1}+\tfrac{1}{2}\pi^{\mathsf{t}}F''(0)QF'(0)\mathbf{1}\\&\qquad+\pi^{\mathsf{t}}F'(0)QF'(0)QF'(0)\mathbf{1}-\pi^{\mathsf{t}}F'(0)\mathbf{1}\cdot\pi^{\mathsf{t}}F'(0)Q^2F'(0)\mathbf{1}.\end{aligned} \tag{5.64}$$

In our case, we set $F(t)=P(e^t)$, and it is easy to see from (5.5) that the derivatives are given by $F^{(k)}(0)=PG^k$ where $G=\mathrm{Diag}(g(1),\dots,g(m))$. Furthermore, the relation between the raw moments $m_i=\frac{d^i}{dt^i}\big|_{t=0}\lambda(F(t))$ and cumulants $\varkappa_i=\frac{d^i}{dt^i}\big|_{t=0}\log\lambda(F(t))$ is as usual:

$$\mu=\varkappa_1=m_1, \tag{5.65}$$

$$\sigma^2=\varkappa_2=m_2-m_1^2, \tag{5.66}$$

$$\varkappa_3=m_3-3m_2m_1+2m_1^3. \tag{5.67}$$

Plugging in (5.62)–(5.64), and recalling $\pi^{\mathsf{t}}P=\pi^{\mathsf{t}}$, we get the indicated expressions in equations (3.15)–(3.17).

Finally, (2.19) yields

$$QP=Q-I+\mathbf{1}\pi^{\mathsf{t}}=Q'+\mathbf{1}\pi^{\mathsf{t}}, \tag{5.68}$$

and as a consequence, since $Q\mathbf{1}=0$,

$$Q^2P=Q(QP)=Q(Q'+\mathbf{1}\pi^{\mathsf{t}})=QQ'=Q'^2+Q'. \tag{5.69}$$

Substituting these in (3.16)–(3.17) yields (3.18)–(3.19). ☐

Proof of Corollary 3.3. We obtain (3.20) directly from (3.14) by taking $\mu=0$ and $x=0$, noting that $\vartheta(0)=1/2$ by (2.14). To obtain (3.21) we instead take $x<0$ and let $x\nearrow 0$, noting that $\vartheta(0-)=-1/2$.

The formula (3.22) then follows from

$$\mathbb{P}(S_n<0)-\mathbb{P}(S_n>0)=\mathbb{P}_n(S_n<0)+\mathbb{P}(S_n\leqslant 0)-1, \tag{5.70}$$

and (3.23) follows from $\mathbb{P}(S_n = 0) = \mathbb{P}_n(S_n \leqslant 0) - \mathbb{P}(S_n < 0)$. □

5.3. **Proof of Theorem 3.4.**

Proof of Theorem 3.4. Let $\widehat{W}_k := (W_{k-1}, W_k)$; this is easily seen to be a Markov chain in the state space

$$\widehat{\mathcal{W}} := \big\{(i,j) \in \mathcal{W}^2 : P_{ij} > 0\big\}. \tag{5.71}$$

We can regard g as a function $\widehat{\mathcal{W}} \to \mathbb{Z}$, and then the definition (3.24) yields the same S_n as (3.7) for the chain $(\widehat{W}_k)_1^\infty$. It is easily seen that the chain $(\widehat{W}_k)_1^\infty$ is stationary, irreducible and aperiodic when $(W_k)_0^\infty$ is, and the condition (3.9) is the same for both chains. Hence the result follows from Theorem 3.2 applied to $(\widehat{W}_k)_1^\infty$. □

6. The aperiodicity condition and non-degeneracy

The general results above use the aperiodicity condition (3.9). In this section we give several equivalent forms of it, and also of the condition $\sigma^2 > 0$ which, as stated in Theorem 3.2 is a consequence. We consider irreducible and aperiodic finite-state Markov chains as in Section 3. In particular, we specialize to Litt's game, and show that these conditions are always satisfied except for the rather trivial cases in Examples 6.2, 6.3, and 6.4 below.

6.1. **Bad examples.**

We begin with a few examples where (3.9) fails, illustrating the reasons for it and showing the consequences. The first two examples are degenerate with also $\sigma^2 = 0$.

Example 6.1. Consider Litt's game with $A = \mathsf{HT}$ and $B = \mathsf{TH}$. It is then obvious by induction that

$$\widehat{S}_n = \mathbf{1}_{\xi_n = \mathsf{T}} - \mathbf{1}_{\xi_1 = \mathsf{T}} \in \{-1, 0, 1\}. \tag{6.1}$$

Thus (3.12) yields $\sigma^2 = 0$. Furthermore, it is easily seen that $g(\mathcal{Q}) = 0$ for every closed path $\mathcal{Q}$, and thus (3.9) fails. It is easy to see that the matrix $P(e^{\mathrm{i}t})$ in Section 5 has an eigenvalue $\lambda(e^{\mathrm{i}t}) = 1$ for every real t, and thus Lemmas 5.7–5.9 fail. It follows trivially from (6.1) that for every $n \geqslant 2$,

$$\mathbb{P}(\text{Alice wins}) = \mathbb{P}(\text{Bob wins}) = \tfrac14, \qquad \mathbb{P}(\text{tie}) = \tfrac12. \tag{6.2}$$

Hence, (1.5)–(1.7) utterly fail, while (1.8) holds trivially. △

Example 6.2. More generally, let $\ell \geqslant 2$ and consider Litt's game with $A = \mathsf{HT}^{\ell-1}$ and $B = \mathsf{T}^{\ell-1}\mathsf{H}$. Then Alice scores when a run of at least $\ell - 1$ tails has begun, and Bob scores when such a run ends, and again it is obvious that $\widehat{S}_n \in \{-1, 0, 1\}$. Thus $\sigma^2 = 0$. Also, again $g(\mathcal{Q}) = 0$ for every closed path $\mathcal{Q}$, and (3.9) fails. As in Example 6.1, it is easy to see that (1.5)–(1.7) fail, while (1.8) holds trivially by symmetry (reversing the order of the coin tosses, cf. [8]).

We will see below (Theorem 7.1) that, as stated without proof in [1], this example and its obvious equivalent variants obtained by interchanging Alice and Bob or H and T (or both) are the only cases in Litt's game where $\sigma^2 = 0$. △

In the following examples, (3.9) fails but $\sigma^2 > 0$. We will see that Lemma 5.7 fails for them, and as a consequence also Lemma 5.8; not surprisingly also Lemma 5.9 fails.

Example 6.3. Consider the trivial case of Litt's game with $\ell = 1$, $A = \mathsf{H}$ and $B = \mathsf{T}$. In this case, S_n is just the number of heads minus the number of tails. Thus,

$$S_n = 2S_n' - n \quad \text{with} \quad S_n' \in \mathrm{Bin}(n, \tfrac{1}{2}). \tag{6.3}$$

In particular $S_n \equiv n \pmod 2$ and thus $\mathbb{P}(S_n = 0) = 0$ when n is odd, so (1.7) fails, and consequently (by symmetry) also (1.5)–(1.6) fail. (Asymptotic expansions of the probabilities are easily obtained from (6.3), treating n even and n odd separately; we leave this to the reader.)

Similarly, $g(\mathcal{Q}) \equiv \ell(\mathcal{Q}) \pmod 2$ for every path $\mathcal{Q}$ (closed or not). As a consequence, (3.9) does not hold, which is the reason why Theorem 3.2 and Corollary 3.3 do not apply. We have $\sigma^2 = 1 > 0$ by (4.19) (or as a consequence of (6.3)). In the arguments in Section 5, we have by (5.5)

$$P(z) = \frac{1}{2}\begin{pmatrix} z & z^{-1} \\ z & z^{-1} \end{pmatrix}. \tag{6.4}$$

This matrix has rank 1, with eigenvalues $(z + z^{-1})/2$ and 0, and it follows easily from (5.8) and (5.14) that

$$\varphi_{S_n}(t) = G_n(e^{\mathrm{i}t}) = \pi_1(e^{\mathrm{i}t})^{\mathsf{t}} P(e^{\mathrm{i}t})^{n-1}\mathbf{1} = \cos^n t, \tag{6.5}$$

which of course also follows directly from (6.3). Note that $P(-1)$ has an eigenvalue -1, which means that Lemmas 5.7 and 5.8 do not hold for $t = \pi$; similarly, (6.5) shows that Lemma 5.9 fails for $t = \pi$. △

Example 6.4. Consider Litt's game with $A = \mathsf{HH}$ and $B = \mathsf{TT}$. This game is obviously fair by symmetry, so (1.8) holds trivially (with $\theta_{AA} = \theta_{BB} = \frac{1}{2}$), and it is checked below that (1.5)–(1.7) also hold. Thus the conclusions of Theorem 1.1 hold; nevertheless, Theorem 3.2 and Corollary 3.3 do not apply because (3.9) does not hold. In fact, if we instead use the alphabet $\mathcal{A} = \{0, 1\}$, we have, recalling (4.2),

$$g(W_k) = \pm \mathbf{1}_{\xi_k = \xi_{k+1}} \equiv \xi_k - \xi_{k+1} + 1 \pmod 2, \tag{6.6}$$

and consequently, $g(\mathcal{Q}) \equiv \ell(\mathcal{Q}) \pmod 2$ for every closed path $\mathcal{Q}$, which shows that (3.9) does not hold.

In this case, it is easy to calculate the distribution of $\widehat{S}_{n+1} = S_n$ exactly using the arguments in Section 5. We may simplify the calculations by letting (W_k) be a sequence of i.i.d. random bits with $\mathbb{P}(W_k = 0) = \mathbb{P}(W_k = 1) = \frac{1}{2}$, which is a trivial Markov chain, and use the version (3.24) with $g(i, j) := \mathbf{1}_{i=j=0} - \mathbf{1}_{i=j=1}$. It is easily seen that for this version, (5.9) is modified to

$$G_n(z) = \pi_0^{\mathsf{t}} P(z)^n \mathbf{1} \tag{6.7}$$

with

$$P(z) := \bigl(P_{ij} z^{g(i,j)}\bigr)_{i,j \in \mathcal{W}} = \frac{1}{2}\begin{pmatrix} z & 1 \\ 1 & z^{-1} \end{pmatrix}. \tag{6.8}$$

This matrix has rank 1, with eigenvalues $(z + z^{-1})/2$ and 0 (just as (6.4)), and it follows that, for $n \geqslant 1$,

$$P(e^{\mathrm{i}t})^n = (\cos t)^{n-1} P(e^{\mathrm{i}t}). \tag{6.9}$$

Hence, (6.7) yields

$$\varphi_{S_n}(t) = G_n(e^{\mathrm{i}t}) = \frac{1 + \cos t}{2} \cos^{n-1} t. \tag{6.10}$$

(This factorization has the probabilistic interpretation that S_n has the same distribution as a sum $\sum_0^{n-1} Y_j$ of independent random variables, where Y_0+1 has the binomial distribution $\operatorname{Bin}(2,\frac12)$ and $Y_j=\pm1$ with probability $\frac12$ each for $j\geqslant 1$.) It follows easily that, for example, with $m:=\lfloor n/2\rfloor$,

$$\mathbb{P}(S_n=0)=2^{-2m-1}\binom{2m}{m}=\frac{1}{\sqrt{2\pi n}}+O(n^{-3/2}). \tag{6.11}$$

We have $\sigma^2=1$, by (4.19) or directly from (6.10) which implies that $\operatorname{Var} S_n=n-\frac12$. Hence (1.7) holds, and thus (by symmetry) also (1.5)–(1.6).

This is as expected, but note that the next term in the expansion (6.11), of order $n^{-3/2}$, will depend on the parity of n. The reason is that the matrix $P(e^{\mathrm{i}\pi})=P(-1)$ has an eigenvalue -1 on the unit circle; thus, although the characteristic function (6.10) vanishes at $t=\pi$, it is not exponentially small for t close to π, as it is in Section 5 when we assume (3.9) and as a consequence Lemma 5.7 holds. △

We end with an (artificial) example of a different type of Markov chain where (3.9) fails.

Example 6.5. Consider a stationary Markov chain (W_k) with 4 states $\mathcal{A}=\{a,b,c,d\}$ and the transition matrix

$$P=\begin{pmatrix}0.45 & 0.45 & 0.1 & 0\\ 0.45 & 0.45 & 0.1 & 0\\ 0 & 0 & 0 & 1\\ 0.5 & 0.5 & 0 & 0\end{pmatrix}. \tag{6.12}$$

This is irreducible and aperiodic, with stationary distribution $(\frac{5}{12},\frac{5}{12},\frac{1}{12},\frac{1}{12})$. Let $g(a)=1$, $g(b)=-1$, $g(c)=g(d)=0$. If we partition $\mathcal{W}$ as $\{a,b,d\}\cup\{c\}$, then $g(W_k)=0$ if W_{k-1} and W_k are in different parts, but $g(W_k)=\pm1$ if W_{k-1} and W_k are in the same part, and it follows that

$$g(W_k)\equiv \mathbf{1}_{W_k=c}-\mathbf{1}_{W_{k-1}=c}+1 \pmod 2. \tag{6.13}$$

This implies that for any closed path $\mathcal{Q}$,

$$g(\mathcal{Q})\equiv \ell(\mathcal{Q}) \pmod 2, \tag{6.14}$$

and thus (3.9) does not hold. It is easy to see, e.g. by the calculations below or by Proposition 6.8, that $\sigma^2>0$.

It follows also from (6.13) and induction that

$$S_n=\sum_{k=1}^n g(W_k)\equiv \mathbf{1}_{W_1=d}+\mathbf{1}_{W_n=c}+n \pmod 2. \tag{6.15}$$

As $n\to\infty$, W_1 and W_n are asymptotically independent, and thus $\mathbb{P}(W_1=d,W_n=c)\to\frac{1}{12^2}$ and $\mathbb{P}(W_1\neq d,W_n\neq c)\to\frac{11^2}{12^2}$. Hence, if n is even, then S_n is even with probability $\approx\frac{122}{144}=\frac{61}{72}$, while if n is odd then S_n is even with probability $\approx\frac{11}{72}$. Hence S_n exhibits a strong periodicity, although the Markov chain itself is aperiodic.

In the arguments of Section 5, this is reflected in the easily verified fact that $P(-1)$ has an eigenvalue -1, and thus $\rho(P(-1))=1$ so Lemma 5.7 fails for $t=\pi$. Lemma 5.7 holds for $0<t<\pi$, and thus Lemmas 5.8 and 5.9 hold for $\delta\leqslant t\leqslant\pi-\delta$. Furthermore, for z close to -1, we have in analogy with (5.11) an approximation

$$G_n(z)=\eta_-(z)\lambda(z)^n\bigl(1+O(c^n)\bigr) \tag{6.16}$$

with $c < 1$, where $\eta_-(z)$ and $\lambda(z)$ are analytic and $\lambda(-1) = -1$; furthermore, $\lambda(z)$ is an eigenvalue of $P(z)$, and we have $\lambda(z) = -\lambda(-z)$.

We can now argue as in Section 5, but we have to include terms coming from $z = -1$. For simplicity, consider

$$\mathbb{P}(S_n = 0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \varphi_{S_n}(t) = \frac{1}{2\pi}\int_{-\pi}^{\pi} G_n(e^{\mathrm{i}t}). \tag{6.17}$$

Calculations show that $\lambda'(1) = \lambda'(-1) = 0$, $\lambda''(1) = \frac{5}{6}$, $\lambda''(-1) = -\frac{5}{6}$, $\eta(1) = 1$ (as always), and $\eta(-1) = \frac{25}{36}$; (6.17) and standard arguments then yield

$$\mathbb{P}(S_n = 0) = \frac{3}{\sqrt{5\pi n}}\Bigl(1 + (-1)^n\frac{25}{36} + o(1)\Bigr). \tag{6.18}$$

This shows periodicity for $\mathbb{P}(S_n = 0)$, and in particular (3.14) does not hold. △

6.2. **The aperiodicity condition** (3.9)**.** Let L be the subgroup of $\mathbb{Z}^2$ generated by the set $\bigl\{(g(\mathcal{Q}), \ell(\mathcal{Q})) : \mathcal{Q} \text{ is a closed path}\bigr\}$ in (3.9), and let $G := \mathbb{Z}^2/L$ be the corresponding quotient. Thus the condition (3.9) says $G = \{1\}$.

The projection $\Pi_2 : (x, y) \mapsto y$ maps L onto the subgroup L_2 of $\mathbb{Z}$ generated by the set of $\ell(\mathcal{Q})$ for all closed paths $\mathcal{Q}$, and since we assume the aperiodicity (3.4), we have $L_2 = \mathbb{Z}$; in other words, Π_2 is onto $\mathbb{Z}$. Consequently, there exists some $b \in \mathbb{Z}$ such that $(b, 1) \in L$. Let

$$L_1 := \{x \in \mathbb{Z} : (x, 0) \in L\}, \tag{6.19}$$

i.e., the kernel of Π_2 regarded as a subgroup of $\mathbb{Z}$. Then

$$(x, y) \in L \iff (x, y) - y(b, 1) \in L \iff x - yb \in L_1 \tag{6.20}$$

and thus

$$L = \bigl\{(z + yb, y) : y \in \mathbb{Z}, z \in L_1\bigr\}. \tag{6.21}$$

Similarly, it follows from $(b, 1) \in L$ that every coset $\overline{(x, y)} \in G$ of L has a representative of the form $(z, 0)$. Consequently, the homomorphism $z \mapsto \overline{(z, 0)}$ maps $\mathbb{Z}$ onto G, and thus $G \cong \mathbb{Z}/L_1$. Hence, G is a cyclic group. Let $N := |G| \in \mathbb{N} \cup \{\infty\}$. In particular, (3.9) holds if and only if $N = 1$. There are two cases:

(i) $L_1 = \{0\}$, $N = \infty$, and $G \cong \mathbb{Z}$. Then (6.21) shows that $L = \{y(b, 1) : y \in \mathbb{Z}\}$. The definition of L shows that this holds if and only if

$$g(\mathcal{Q}) = b\ell(\mathcal{Q}) \tag{6.22}$$

for every closed path. This is a very degenerate case, see Section 6.3.

(ii) $L_1 = N\mathbb{Z}$, $1 \leqslant N < \infty$, and $G \cong \mathbb{Z}_N$. It follows from (6.21) that L is a two-dimensional lattice with basis $\{(N, 0), (b, 1)\}$.

In both cases, we identify G with $\mathbb{Z}_N$ (where $\mathbb{Z}_\infty := \mathbb{Z}$) in the natural way by the isomorphism $\mathbb{Z}/L_1 \to G$.

Every possible step ij with $P_{ij} > 0$ in the Markov chain generates a vector $v_{ij} := (g(j), 1)$ in $\mathbb{Z}^2$, and thus a corresponding element $\gamma_{ij} := \overline{(g(j), 1)} \in G$. If we sum v_{ij} along a closed path, we get (by definition) an element of L, and thus the sum of γ_{ij} along a closed path vanishes in G. Fix an (arbitrary) element $o \in \mathcal{W}$, and pick for every $k \in \mathcal{W}$ a path $\mathcal{Q}_k$ from o to k. Let $\gamma_k \in G$ be the sum of γ_{ij} along this path. It is easy to see that since the sum along any closed path is 0, the sum γ_k does not depend on the choice of $\mathcal{Q}_k$. Furthermore, if $P_{ij} > 0$, then we may choose $\mathcal{Q}_j$ as the path $\mathcal{Q}_i$ followed by j, and thus

$$\gamma_j = \gamma_i + \gamma_{ij}, \qquad \text{if } P_{ij} > 0. \tag{6.23}$$

Note also that since $(b, 1) \in L$,

$$\gamma_{ij} = \overline{(g(j), 1)} = \overline{(g(j) - b, 0)}. \tag{6.24}$$

Hence, by identifying G with $\mathbb{Z}_N$, we simply have $\gamma_{ij} = g(j) - b$ in $\mathbb{Z}_N$, and thus by (6.23)

$$\gamma_j = \gamma_i + g(j) - b \qquad \text{in } \mathbb{Z}_N \text{ when } P_{ij} > 0. \tag{6.25}$$

This leads to the following characterizations of (3.9).

Proposition 6.6. *Let (W_k) be an irreducible and aperiodic Markov chain on a finite state space $\mathcal{W}$, and let $g : \mathcal{W} \to \mathbb{Z}$. Then, with notation as in Sections 3 and 5, the following are equivalent:*

(i) *The condition* (3.9) *does* not *hold.*
(ii) *$P(e^{\mathrm{i}t})$ has an eigenvalue λ with $|\lambda| = 1$ for some $t \in (0, \pi]$.*
(iii) *The spectral radius $\rho\bigl(P(e^{\mathrm{i}t})\bigr) = 1$ for some $t \in (0, \pi]$.*
(iv) *There exist an integer $N \geqslant 2$, an integer b, and integers γ_i, $i \in \mathcal{W}$, such that for every pair $i, j \in \mathcal{W}$ with $P_{ij} > 0$,*

$$\gamma_j \equiv \gamma_i + g(j) - b \pmod{N}. \tag{6.26}$$

(v) *There exists an integer $N \geqslant 2$, an integer b, and a partition of $\mathcal{W}$ into (possibly empty) sets $\mathcal{W}_k$, $k = 1, \dots, N$, such that if $i \in \mathcal{W}_k$ and $P_{ij} > 0$, then $j \in \mathcal{W}_{k+g(j)-b}$, with the index regarded modulo N.*

Proof. First, the proof of Lemma 5.7 shows that every eigenvalue λ of $P(e^{\mathrm{i}t})$ satisfies $|\lambda| \leqslant 1$, and thus (ii) $\iff$ (iii) by the definition of spectral radius.

Furthermore, (iv) $\iff$ (v), since if (iv) holds, we may define $\mathcal{W}_k := \{i : \gamma_i \equiv k \pmod{N}\}$, and conversely we may define $\gamma_i = k$ for $i \in \mathcal{W}_k$.

Lemma 5.7 shows that if (3.9) holds, then (iii) does not hold; thus (iii) $\implies$ (i).

Now suppose that (i) holds. This means that in the discussion above, $N > 1$. If $N < \infty$, then (6.25) shows the existence of b and γ_i such that (6.26) holds. If $N = \infty$, then (6.26) holds in $\mathbb{Z}$, and it thus holds for any integer $N \geqslant 2$. This shows (i) $\implies$ (iv).

Finally, suppose that (iv) holds. Let $t = 2\pi/N$ and $\omega := e^{\mathrm{i}t} = e^{2\pi\mathrm{i}/N}$, and let $v_i := \omega^{-\gamma_i}$. Then, by (5.5) and (6.26) if $P_{ij} > 0$, and trivially if $P_{ij} = 0$,

$$P(\omega)_{ij} v_j = P_{ij} \omega^{g(j) - \gamma_j} = P_{ij} \omega^{b - g(i)} = P_{ij} \omega^b v_i, \tag{6.27}$$

and by summing over j we see that $v := (v_i)_i$ is an eigenvector of $P(\omega)$ with eigenvalue ω^b. Hence (ii) holds. $\square$

Example 6.7. Nagaev [21, Condition C] makes the (rather strong) assumption that for every pair $i, j \in \mathcal{W}$, there exists $k \in \mathcal{W}$ such that $P_{ik}, P_{jk} > 0$. This in combination with (6.26) implies that

$$\gamma_i \equiv \gamma_k - g(k) + b \equiv \gamma_j \pmod{N}. \tag{6.28}$$

Thus, for every j, by choosing i such that $P_{ij} > 0$ and applying (6.26) again,

$$g(j) \equiv b \pmod{N}, \qquad j \in \mathcal{W}. \tag{6.29}$$

Since [21, Theorem 3] also assumes (in our notation) that (6.29) does not hold for any $N \geqslant 2$, we see that the assumptions in [21] imply that Proposition 6.6(iv) cannot hold, and thus Proposition 6.6 shows that (3.9) holds. △

6.3. **The asymptotic variance $\sigma^2 = 0$.** Another exceptional case when Theorem 1.1 fails is when the asymptotic variance $\sigma^2 = 0$. This is a very degenerate case; as shown in Proposition 6.8 below, it happens only if, after subtracting a suitable constant b from g, the sums S_n are deterministically bounded. We will also see that it cannot happen when (3.9) holds.

Proposition 6.8. *Let (W_k) be an irreducible and aperiodic Markov chain on a finite state space $\mathcal{W}$, and let $g : \mathcal{W} \to \mathbb{Z}$. Then, with notation as in Sections 3 and 6.2, the following are equivalent:*

(i) $\sigma^2 = 0$.

(ii) *There exists an integer (or real number) b such that deterministically*

$$|S_n - bn| \leqslant C \tag{6.30}$$

for some constant C, uniformly in n.

(iii) *There exists an integer (or real number) b such that*

$$g(\mathcal{Q}) = b\ell(\mathcal{Q}) \tag{6.31}$$

for every closed path $\mathcal{Q}$.

(iv) *There exist an integer b and integers γ_i, $i \in \mathcal{W}$, such that for every pair $i, j \in \mathcal{W}$ with $P_{ij} > 0$,*

$$\gamma_j = \gamma_i + g(j) - b. \tag{6.32}$$

(v) *There exists an integer b such that*

$$L = \mathbb{Z}(b, 1) = \{(yb, y) : y \in \mathbb{Z}\}. \tag{6.33}$$

(vi) *The lattice L is one-dimensional.*

(vii) $L_1 = \{0\}$.

(viii) $N = \infty$.

(ix) $G = \mathbb{Z}$.

Furthermore, the constant b in (ii), (iii), (iv), *and* (v) *is the same integer.*

Proof. The equivalences (vi) $\iff$ (v) $\iff$ (vii) $\iff$ (viii) $\iff$ (ix) $\iff$ (iii) follow from the discussion at the beginning of Section 6.2, in particular (6.21) and the text leading to (6.22). Note that if (6.31) holds for some real number b, then $b\ell(\mathcal{Q}) \in \mathbb{Z}$ for every closed path $\mathcal{Q}$, and thus $bx \in \mathbb{Z}$ for every x in the group L_2 generated by the set of all such $\ell(\mathcal{Q})$. As noted above, the aperiodicity (3.4) implies $L_2 = \mathbb{Z}$, and thus $b \in \mathbb{Z}$, so b in (6.31) has to be an integer.

Furthermore, (viii) $\Longrightarrow$ (iv) follows by (6.25) (with $N = \infty$).

(iv) $\Longrightarrow$ (ii): It follows from (6.32) that

$$\begin{aligned} S_n - bn = \sum_{k=1}^{n} (g(W_k) - b) &= g(W_1) - b + \sum_{k=2}^{n} (\gamma_{W_k} - \gamma_{W_{k-1}}) \\ &= g(W_1) - b + \gamma_{W_n} - \gamma_{W_1} \end{aligned} \tag{6.34}$$

and thus (6.30) holds with $C = \max_{i \in \mathcal{W}} |g(i) - b| + 2 \max_{j \in \mathcal{W}} |\gamma_j|$. (The same argument yields (iv) $\Longrightarrow$ (iii) directly, but we do not need this.)

(ii) $\Longrightarrow$ (i): Obvious by the definition (3.12).

(i) $\Longrightarrow$ (iii): Suppose that (iii) does not hold. Then there exist two closed paths $\mathcal{Q}_1$ and $\mathcal{Q}_2$ such that

$$g(\mathcal{Q}_1)/\ell(\mathcal{Q}_1) \neq g(\mathcal{Q}_2)/\ell(\mathcal{Q}_2). \tag{6.35}$$

Let w_i be the starting point of $\mathcal{Q}_i$. If $w_2 \neq w_1$, choose two paths $\mathcal{Q}_{12}$ and $\mathcal{Q}_{21}$ from w_1 to w_2 and from w_2 to w_1, respectively. We may then construct the closed paths $\mathcal{Q}' := \mathcal{Q}_{12} + \mathcal{Q}_{21}$ and $\mathcal{Q}'' := \mathcal{Q}_{12} + \mathcal{Q}_2 + \mathcal{Q}_{21}$, both starting at w_1, by concatenation in the obvious way. Then

$$g(\mathcal{Q}'') = g(\mathcal{Q}_{12}) + g(\mathcal{Q}_2) + g(\mathcal{Q}_{21}) = g(\mathcal{Q}') + g(\mathcal{Q}_2) \tag{6.36}$$

and similarly $\ell(\mathcal{Q}'') = \ell(\mathcal{Q}')+\ell(\mathcal{Q}_2)$; hence it follows from (6.35) that we cannot have $g(\mathcal{Q}')/\ell(\mathcal{Q}') = g(\mathcal{Q}_1)/\ell(\mathcal{Q}_1) = g(\mathcal{Q}'')/\ell(\mathcal{Q}'')$. Consequently, by relabelling either $\mathcal{Q}'$ or $\mathcal{Q}''$ as $\mathcal{Q}_2$, we may assume that (6.35) holds with the same starting point w_1 for both closed paths.

Next, let $\ell_i := \ell(\mathcal{Q}_i) \geqslant 1$. Replace $\mathcal{Q}_i$ by the closed path $\ell_{3-i}\mathcal{Q}_i$, i.e., the closed path $\mathcal{Q}_i$ repeated ℓ_{3-i} times. This replacement does not change $g(\mathcal{Q}_i)/\ell(\mathcal{Q}_i)$, and thus (6.35) still holds, but now both closed paths have the same length $\ell_1\ell_2$.

We may thus assume that (6.35) holds for two closed paths $\mathcal{Q}_1$ and $\mathcal{Q}_2$ of the same length $\ell \geqslant 1$ and starting from the same point $w_1 \in \mathcal{W}$; note that this implies $g(\mathcal{Q}_1) \neq g(\mathcal{Q}_2)$. We now fix these paths, and thus ℓ, and define a sequence of stopping times for the Markov chain by

$$T_0 := \min\{n \geqslant 1 : W_n = w_1\}, \tag{6.37}$$

$$T_k := \min\{n \geqslant T_{k-1} + \ell : W_n = w_1\}, \qquad k \geqslant 1. \tag{6.38}$$

Thus, at T_k we return to w_1, but we only consider returns after a time of at least ℓ.

Since the Markov chain is finite and irreducible, almost surely all T_k are finite. The Markov property shows that the times T_k are renewal times in the sense that the process starts again at each T_k. Let $\tau_k := T_k - T_{k-1}$, the waiting time of the k:th "renewal"; then $\tau_1, \tau_2, \dots$ are i.i.d. random variables. Using the fact that there exists some t_0 such that from every point in $\mathcal{W}$ we may with positive probability reach w_1 within t_0 steps, it is easily seen that the tail probability $\mathbb{P}(\tau_1 > t)$ decreases exponentially as $t \to \infty$; in particular, the moment $\mathbb{E}\,\tau_1^m$ is finite for every $m \geqslant 1$, Similarly, $\mathbb{E}\,T_0^m < \infty$.

Consider the two-dimensional process (n, S_n) and its "increments" between the renewal times T_k defined by

$$\zeta_k := \bigl(T_k - T_{k-1}, \{S_{T_{k-1}+i} - S_{T_{k-1}} : 0 \leqslant i \leqslant T_k - T_{k-1}\}\bigr), \qquad k \geqslant 1, \tag{6.39}$$

where the second component thus is a random process defined on an integer interval of random length. The Markov property shows that these increments are i.i.d. We may thus apply the general result from renewal theory stated in Lemma 6.9 below and conclude that, with the notation in (6.42)–(6.44), (6.45) holds with convergence of all moments. In particular,

$$\frac{\mathrm{Var}[S_n]}{n} = \mathrm{Var}\Bigl(\frac{S_n - an}{\sqrt{n}}\Bigr) \to \widetilde{\sigma}^2/\widetilde{\mu} \tag{6.40}$$

and thus, recalling (3.12),

$$\sigma^2 = \widetilde{\sigma}^2/\widetilde{\mu}. \tag{6.41}$$

Now recall the closed paths $\mathcal{Q}_1$ and $\mathcal{Q}_2$ of the same length ℓ constructed above. For each $i = 1, 2$, with positive probability the Markov chain starts in w_1 and then follows $\mathcal{Q}_i$ for the next ℓ steps; in this case $T_0 = 1$ and $T_1 = \ell + 1$ so $\tau_1 = \ell$, and also $S_{T_1} - S_{T_0} = g(\mathcal{Q}_i)$. Consequently, $S_{T_1} - S_{T_0} - aT_1$ takes two different values $g(\mathcal{Q}_i) - a\ell$ with positive probability, and thus $\widetilde{\sigma}^2 > 0$ by (6.44) and consequently $\sigma^2 > 0$ by (6.41), so (i) does not hold. This completes the proof of (i) $\Longrightarrow$ (iii), which completes the chain of equivalences. □

We used above the following renewal theoretic lemma, stated here for our Markov chain and S_n. The lemma follows easily from known general results in renewal theory; we find it convenient to use a version from [15] that uses a formulation from [28]. The powerful idea to analyze Markov chains (also with an infinite state space) by a suitable renewal sequence of stopping times is old, and was for example used by [12].

Lemma 6.9. *Let $(T_k)_{k=0}^{\infty}$ be an increasing sequence of (a.s. finite) random times and let $\tau_k := T_k - T_{k-1}$, $k \geqslant 1$. Assume that the increments ζ_k defined in* (6.39) *for $k \geqslant 1$ are i.i.d. (Thus, in particular, the increments τ_k are i.i.d.) Assume also that for every $r \geqslant 1$, $\mathbb{E}\, T_0^r < \infty$ and $\mathbb{E}\, \tau_1^r < \infty$. Let*

$$a := \mathbb{E}\,[S_{T_1} - S_{T_0}]/\,\mathbb{E}\,[\tau_1], \tag{6.42}$$

$$\widetilde{\mu} := \mathbb{E}\,[\tau_1], \tag{6.43}$$

$$\widetilde{\sigma}^2 := \operatorname{Var}[S_{T_1} - S_{T_0} - a\tau_1]. \tag{6.44}$$

Then, as $n \to \infty$,

$$\frac{S_n - an}{\sqrt{n}} \xrightarrow{\text{d}} N\big(0, \widetilde{\sigma}^2/\widetilde{\mu}\big) \tag{6.45}$$

with convergence of all moments.

Proof. Consider first the process only from time T_0. Formally we define

$$S'_n := S_{T_0+n} - S_{T_0} \tag{6.46}$$

and note that $(S'_n)_{n=0}^{\infty}$ is a stochastic process that starts at time $n = 0$ with $S'_0 = 0$; define also $T'_n := T_n - T_0$. Then the increments ζ_k in (6.39) are the same for (S'_n) and (T'_k) as for (S_n) and (T_k). By assumption these are i.i.d.; this means that in the terminology of [28] and [15], the stochastic process S'_n has regenerative increments over the times T'_k.

Since the state space $\mathcal{W}$ is finite, the function g is bounded, and thus $|S'_{T_1}| \leqslant C(T_1 - T_0) = C\tau_1$; consequently the moment $\mathbb{E}\,|S'_{T_1}|^r$ is finite for every $r \geqslant 1$. Moreover, if $M_1 := \sup_{1\leqslant n\leqslant T'_1} |S'_n|$ then $\mathbb{E}\, M_1^r < \infty$ by the same argument.

We may now apply [15, Theorems 1.4 and 3.1] and conclude that

$$\frac{S'_n - an}{\sqrt{n}} \xrightarrow{\text{d}} N\big(0, \widetilde{\sigma}^2/\widetilde{\mu}\big) \tag{6.47}$$

with convergence of all moments.

Finally, again since g is bounded, $|S_{T_0+n} - S_n| \leqslant CT_0$ and $|S_{T_0}| \leqslant CT_0$, and consequently

$$|S_n - S'_n| \leqslant |S_{T_0+n} - S_n| + |S_{T_0}| \leqslant CT_0. \tag{6.48}$$

It follows that, as $n \to \infty$,

$$\frac{S_n - S'_n}{\sqrt{n}} \to 0 \tag{6.49}$$

in probability, which together with (6.47) implies (6.45). Moreover, since $\mathbb{E}\, T_0^p < \infty$ for every $p \geqslant 1$, it follows from (6.48) that (6.49) holds in L^p for every $p \geqslant 1$, which together with the moment convergence in (6.47) implies that all moments converge in (6.45) too. □

Corollary 6.10. *Let (W_k) be an irreducible and aperiodic Markov chain on a finite state space $\mathcal{W}$, and let $g : \mathcal{W} \to \mathbb{Z}$. Then, with notation as in Sections 3, if* (3.9) *holds, then $\sigma^2 > 0$.*

Proof. Proposition 6.8 shows that if $\sigma^2 = 0$, then (6.31) holds and thus (3.9) does not hold. □

This corollary completes the proof of Theorem 3.2.

7. Back to Litt's game

We consider again Litt's game. We begin by showing that the only cases where $\sigma^2 = 0$ are the rather trivial cases in Example 6.2 (including Example 6.1). This (in the form (7.1) below) is stated in Basdevant et al. [1] without proof; since we also do not know of a proof given elsewhere, we give a complete proof.

Theorem 7.1. *For Litt's game in Section 4, we have $\sigma^2 > 0$ except in the case $A = \mathsf{H}\mathsf{T}^{\ell-1}$ and $B = \mathsf{T}^{\ell-1}\mathsf{H}$ (with $q = 2$ and some $\ell \geqslant 2$) and its variants obtained by interchanging Alice and Bob or* H *and* T *(or both).*

Proof. Suppose that A and B are distinct words such that $\sigma^2 = 0$. By (4.19), this is equivalent to

$$1 + \theta_{AA} - \theta_{AB} - \theta_{BA} + \theta_{BB} = 0. \tag{7.1}$$

Note first that for any two words U and V of length ℓ, (4.10) yields

$$0 \leqslant \theta_{UV} \leqslant \sum_{k=1}^{\ell-1} q^{k-\ell} = \frac{1 - q^{-(\ell-1)}}{q-1} < \frac{1}{q-1}. \tag{7.2}$$

Hence, (7.1) is impossible if $q \geqslant 3$, and thus we in the rest of the proof assume $q = 2$ and take the alphabet to be $\{\mathsf{H}, \mathsf{T}\}$. Furthermore, we have to have $\ell \geqslant 2$, since $\theta_{UV} = 0$ when $\ell = 1$.

Moreover, if $\ell - 1 \notin \Theta(U, V)$, then similarly

$$\theta_{UV} \leqslant \sum_{k=1}^{\ell-2} 2^{k-\ell} < \frac{1}{2}. \tag{7.3}$$

Hence, (7.1) cannot hold unless $\ell - 1 \in \Theta(A, B)$ or $\ell - 1 \in \Theta(B, A)$. By symmetry, we may assume the first, i.e., that the last $\ell - 1$ letters in A are the same as the first $\ell - 1$ letters in B. By symmetry, we may also assume that A ends with T. This means that for some word C of length $\ell - 2$ and some letters $a, b \in \{\mathsf{H}, \mathsf{T}\}$, we have

$$A = aC\mathsf{T}, \qquad B = C\mathsf{T}b. \tag{7.4}$$

Recall from Section 4 that $\widehat{S}_n = S_{n-\ell+1} = \sum_{k=1}^{n-\ell-1} g(\xi_k \cdots \xi_{k+\ell-1})$ where $g : \mathcal{A} \to \mathbb{Z}$ is given by $g = I_A - I_B$; thus

$$g(A) = 1, \quad g(B) = -1, \quad \text{and otherwise } g = 0. \tag{7.5}$$

We use [14, Theorem 2] which shows that then the asymptotic variance $\sigma^2 = 0$ if and only if there exists a function $h : \mathcal{A}^{\ell-1} \to \mathbb{R}$ and a constant μ such that for any word $a_1 \cdots a_\ell$ we have

$$g(a_1 \cdots a_\ell) = h(a_2 \cdots a_\ell) - h(a_1 \cdots a_{\ell-1}) + \mu. \tag{7.6}$$

(In fact, it follows from (7.6) that $\mu = \mathbb{E}\, g(\xi_1 \cdots \xi_\ell) = \mathbb{E}\, g(W_1) = 0$, and thus $\mu = 0$ is the same as before, see (4.8).)

We have $A = aC\mathsf{T}$ by (7.4), and thus $aC\mathsf{H} \neq A$. Hence, (7.5) and (7.6) yield

$$0 < g(aC\mathsf{T}) - g(aC\mathsf{H}) = h(C\mathsf{T}) - h(C\mathsf{H}). \tag{7.7}$$

Let $\bar{a}$ be the letter in $\{\mathsf{H}, \mathsf{T}\}$ distinct from a. Then by (7.6) again and (7.7),

$$g(\bar{a}C\mathsf{T}) - g(\bar{a}C\mathsf{H}) = h(C\mathsf{T}) - h(C\mathsf{H}) > 0. \tag{7.8}$$

Since $\bar{a}C\mathsf{T} \neq A$, (7.8) is possible only if $\bar{a}C\mathsf{H} = B$. Thus, by (7.4),

$$\bar{a}C\mathsf{H} = B = C\mathsf{T}b. \tag{7.9}$$

Hence, $b = \mathsf{H}$ and

$$\bar{a}C = C\mathsf{T}. \tag{7.10}$$

By counting the number of Ts on both sides, we see that $\bar{a} = \mathsf{T}$, and thus $a = \mathsf{H}$. Finally, (7.10) shows that $C\mathsf{T}$ is invariant under cyclic permutations, and thus must contain only one letter; hence $C = \mathsf{T}^{\ell-2}$. The conclusion $A = \mathsf{H}\mathsf{T}^{\ell-1}$ and $B = \mathsf{T}^{\ell-1}\mathsf{H}$ now follows from (7.4). ☐

7.1. **The aperiodicity condition for Litt's game.** We next show that for Litt's game, also the aperiodicity condition (3.9) is satisfied except in the (more or less trivial) cases in Examples 6.2, 6.4 and 6.3, i.e., in the exceptional cases in Theorem 7.1 and also in the cases (H, T) and $(\mathsf{HH}, \mathsf{TT})$ (or vice versa).

Theorem 7.2. *For Litt's game in Section 4, the aperiodicity condition* (3.9) *holds except in the cases (all with $q = 2$)*

(i) $A = \mathsf{H}\mathsf{T}^{\ell-1}$ *and* $B = \mathsf{T}^{\ell-1}\mathsf{H}$ *for some* $\ell \geqslant 2$ *(see Example 6.2),*
(ii) $A = \mathsf{H}$ *and* $B = \mathsf{T}$ *(see Example 6.3),*
(iii) $A = \mathsf{HH}$ *and* $B = \mathsf{TT}$ *(see Example 6.4),*

and their variants obtained by interchanging Alice and Bob or H *and* T *(or both).*

Proof. Recall that a path $\mathcal{Q}$ is a sequence $i_0, \dots, i_m$ in $\mathcal{W} = \mathcal{A}^\ell$, with each transition having a positive probability. The length is defined as m and the value $g(\mathcal{Q})$ is $\sum_{k=1}^m g(i_k)$. Define the weight of the path as $w(\mathcal{Q}) := \prod_{k=1}^m P_{i_{k-1}, i_k}$; this is the probability that if the Markov chain starts in i_0, it will follow $\mathcal{Q}$ for the next m steps.

To satisfy the aperiodicity condition, it is sufficient to show:

(a) There exists a closed path $\mathcal{Q}_1$ such that $g(\mathcal{Q}_1) = 0$ and $\ell(\mathcal{Q}_1) = 1$.
(b) There exists a closed path $\mathcal{Q}_2$ such that $g(\mathcal{Q}_2) = 1$.

In this case, there exists some $N \geqslant 1$ such that $(0, 1)$ and $(1, N)$ are in the set $\big\{(g(\mathcal{Q}), \ell(\mathcal{Q})) : \mathcal{Q} \text{ is a closed path}\big\}$ in (3.9) and it is clear that these two vectors together generate $\mathbb{Z}^2$; thus (3.9) holds.

The first condition (a) is simple, since for every $a \in \mathcal{A}$, $a \cdots a \to a \cdots a$ is a closed path in $\mathcal{W}$ of length 1. If $q \geqslant 3$, or $q = 2$ and at least one of A or B is non-constant, then at least one of these paths also has value 0, thus satisfying (a). The remaining case $q = 2$ with $A = \mathsf{H}^\ell$ and $B = \mathsf{T}^\ell$ (or vice versa) is as noted an exception if $\ell \leqslant 2$; if $\ell \geqslant 3$, (a) does not hold but we note that the sequences of coin tosses obtained by repeating HT or HHT yields closed paths $\mathcal{Q}_2$ and $\mathcal{Q}_3$ with lengths 2 and 3 and value 0, which is just as good since $\gcd(2, 3) = 1$.

To show (b), it is sufficient to show there exists a closed path through A (of length $\geqslant 1$) that avoids B. Given such a path, we can truncate it to the time of first return to A to ensure that the value of the path is 1.

Let $W(m)$ be the total weight of all closed paths from A with length m, that avoid B, with $W(0) := 1$. It is easily seen that these numbers have the following generating function:

$$(I - tP_{\neq B})^{-1}_{A,A} = \sum_{m\geqslant 0} W(m)t^m \tag{7.11}$$

where $P_{\neq B}$ is given by zero-ing out the column of P corresponding to B.

We need to show that at least one $W(m)$ is non-zero for $m > 0$. We will do this by showing that $\sum_{m\geqslant 0} W(m) > W(0) = 1$. For this we use the formula (7.16) in Lemma 7.3 below, which together with (7.11) gives

$$\sum_{m\geqslant 0} W(m) = (I - P_{\neq B})^{-1}_{A,A} = Q_{AA} + Q_{BB} - Q_{AB} - Q_{BA}. \tag{7.12}$$

(An alternative proof of (7.12) is given in Appendix B.) The explicit formula (4.27) now yields

$$\sum_{m\geqslant 0} W(m) = 2 + \theta_{AA} + \theta_{BB} - \theta_{AB} - \theta_{BA}. \tag{7.13}$$

Hence, by (7.13) and (4.19),

$$\sum_{m\geqslant 0} W(m) > 1 \iff 1 + \theta_{AA} + \theta_{BB} - \theta_{AB} - \theta_{BA} > 0 \iff \sigma^2 > 0, \tag{7.14}$$

which holds by Theorem 7.1 except in the excluded case (i). ∎

The proof above used the following lemma. (See also Appendix B.)

Lemma 7.3. *Let P be an irreducible stochastic matrix on a finite state space $\mathcal{W}$, and let π be its stationary distribution. Then, for any $B \in \mathcal{W}$, the matrix $(I - P_{\neq B})$ is invertible, and the inverse has the entries, for $i, j \in \mathcal{W}$,*

$$(I - P_{\neq B})^{-1}_{ij} = Q_{ij} + \bigl(Q_{BB} + (I - Q)_{iB}\bigr)\frac{\pi_j}{\pi_B} - Q_{Bj}. \tag{7.15}$$

In particular, if the stationary distribution π is uniform, then

$$(I - P_{\neq B})^{-1}_{ij} = \mathbf{1}_{i=B} + Q_{ij} + Q_{BB} - Q_{iB} - Q_{Bj}. \tag{7.16}$$

Proof. Recall the identity $(I - P)Q = I - \mathbf{1}\pi^t$ (2.19). Hence, $PQ = Q - I + \mathbf{1}\pi^t$. Thus, using the Kronecker delta $\delta_{kj} := \mathbf{1}_{k=j}$,

$$\begin{aligned}(P_{\neq B}Q)_{kj} &= \sum_{i\neq B} P_{ki}Q_{ij} \\ &= (PQ)_{kj} - P_{kB}Q_{Bj} \\ &= Q_{kj} - \delta_{kj} - \pi_j - P_{kB}Q_{Bj}.\end{aligned} \tag{7.17}$$

Similarly

$$\sum_i (P_{\neq B})_{ki}(I - Q)_{iB} = (P(I - Q))_{kB} - P_{kB}(I - Q)_{BB}$$

$$\begin{aligned}
&= P_{kB} + (I - Q - \mathbf{1}\pi^t)_{kB} - P_{kB}(1 - Q_{BB}) \\
&= P_{kB} + \delta_{kB} - Q_{kB} - \pi_B - P_{kB} + Q_{BB}P_{kB} \\
&= \delta_{kB} - Q_{kB} - \pi_B + Q_{BB}P_{kB} \qquad (7.18)
\end{aligned}$$

So if we define M_{ij} to be the right hand side of (7.15), then, using $\sum_{i\neq B} P_{ki} = 1 - P_{kB}$,

$$\begin{aligned}
(P_{\neq B}M)_{kj} &= \sum_i (P_{\neq B})_{ki} \left(Q_{ij} + (Q_{BB} + (I-Q)_{iB}) \frac{\pi_j}{\pi_B} - Q_{Bj} \right) \\
&= Q_{kj} - \delta_{kj} + \pi_j - P_{kB}Q_{Bj} + \big((1 - P_{kB})Q_{BB} \\
&\qquad + \delta_{kB} - Q_{kB} - \pi_B + Q_{BB}P_{kB}\big) \frac{\pi_j}{\pi_B} - (1 - P_{kB})Q_{Bj} \\
&= -\delta_{kj} + Q_{kj} + \pi_j - P_{kB}Q_{Bj} + (Q_{BB} + \delta_{kB} - Q_{kB}) \frac{\pi_j}{\pi_B} \\
&\qquad - \pi_j - Q_{Bj} + P_{kB}Q_{Bj} \\
&= -\delta_{kj} + Q_{kj} + (Q_{BB} + (I-Q)_{kB}) \frac{\pi_j}{\pi_B} - Q_{Bj} \\
&= -\delta_{kj} + M_{kj}. \qquad (7.19)
\end{aligned}$$

Thus $(I - P_{\neq B})M = I$ and the claim follows. □

Remark 7.4. An alternative proof of this lemma proceeds by writing $I - P_{\neq B} = I - P + uv^t = (I - P + \mathbf{1}\pi^t) + (uv^t - \mathbf{1}\pi^t)$ for appropriate vectors u, v. Since the matrix $I - P + \mathbf{1}\pi^t$ is invertible (Proposition 2.1), we can now apply the standard Woodbury formula to write $(I - P_{\neq B})^{-1}$ as a rank-2 correction to $(I - P + \mathbf{1}\pi^t)^{-1} = Q + \mathbf{1}\pi^t$. A similar argument also works to derive a general formula for rank-k updates to $I - P$. The details are straightforward and left to the reader. △

Remark 7.5. The quantity $\sum_{m\geqslant 0} W(m) = (I - P_{\neq B})^{-1}_{AA}$ has the alternative probablistic interpretation the expected number of visits to A before visiting B for a Markov chain started at A (counting the initial state as a visit). This follows by considering N_i, the expected number of visits to A before visiting B for a random walk started at i, and observing the linear relation $N_i = \mathbf{1}_{i=A} + \sum_{j\neq B} P_{ij}N_j$. △

Proof of Theorem 1.1. By Theorem 7.2, except in the excluded cases and in the case $A = \mathsf{HH}$ and $B = \mathsf{TT}$ (or conversely), the condition (3.9) holds and then the result is given by Theorem 4.2.

In the remaining case $A = \mathsf{HH}$ and $B = \mathsf{TT}$, (3.9) fails but $\sigma^2 > 0$ and (1.5)–(1.8) still hold by simple direct calculations, see Example 6.4. □

8. Analysis of state space expansion

The proof of Theorem 3.4 involves passing to the expanded state space of pairs (i, j). In order to modify the explicit moment formulas for this case, it is necessary to know how the stationary distribution π and group inverse Q are transformed under this state-space-expansion operation.

The expanded transition matrix $\widehat{P}_{ij,i'j'}$ evidently satisfies

$$\widehat{P}_{ij,i'j'} = \mathbf{1}_{j=i'} P_{j,j'} \qquad (8.1)$$

We have

$$\begin{aligned}\sum_{ij}\pi_i P_{i,j}\widehat{P}_{ij,i'j'} &= \sum_{ij}\pi_i P_{i,j}\mathbf{1}_{j=i'}P_{j,j'}\\ &= \sum_{i}\pi_i P_{i,i'}P_{i',j'}\\ &= \pi_{i'}P_{i',j'} \qquad (8.2)\end{aligned}$$

Therefore the expanded stationary distribution is

$$\Pi_{ij} = \pi_i P_{ij} \qquad (8.3)$$

We also need to work out the expanded group inverse $\widehat{Q} := (I-\widehat{P})^{\mathsf{g}}$. A simple induction shows

$$(\widehat{P}^k)_{ij,i'j'} \quad = \quad (P^{k-1})_{j,i'}P_{i',j'} \qquad (8.4)$$

for $k \geqslant 1$. Another way to see this is to note that a length-k path in the extended state space from (i,j) to (i',j') can be decomposed into a length $k-1$ path in the original state spce between j and i', together with the transition $i' \to j$.

Therefore, for $0 \leqslant t < 1$:

$$\begin{aligned}(I-t\widehat{P})^{-1}_{ij,i'j'} &= I_{ij,i'j'} + \sum_{k\geqslant 1} t^k(\widehat{P}^k)_{ij,i'j'}\\ &= I_{ij,i'j'} + \sum_{k\geqslant 1} t^k(P^{k-1})_{j,i'}P_{i',j'}\\ &= I_{ij,i'j'} + P_{i',j'}t\sum_{k\geqslant 0} t^k(P^k)_{j,i'}\\ &= I_{ij,i'j'} + P_{i',j'}t(I-tP)^{-1}_{j,i'}. \qquad (8.5)\end{aligned}$$

Using the group inverse formula in Proposition 2.1 and (8.3) thus yields:

$$\begin{aligned}(I-\widehat{P})^{\mathsf{g}}_{ij,i'j'} &= \lim_{t\nearrow 1}\bigl((I-t\widehat{P})^{-1}_{ij,i'j'} - \pi_{i'}P_{i'j'}/(1-t)\bigr)\\ &= \lim_{t\nearrow 1}\bigl(\mathbf{1}_{i=i',j=j'} + tP_{i',j'}(I-tP)^{-1}_{j,i'} - \pi_{i'}P_{i',j'}/(1-t)\bigr)\\ &= \mathbf{1}_{i=i',j=j'} + P_{i',j'}\lim_{t\nearrow 1}\Bigl(t(I-tP)^{-1}_{j,i'} - \pi_{i'}/(1-t)\Bigr). \qquad (8.6)\end{aligned}$$

Now:

$$\begin{aligned}t(I-tP)^{-1} - \mathbf{1}\pi^{\mathsf{t}}/(1-t) \quad &= \quad t(I-tP)^{-1} - t\mathbf{1}\pi^{\mathsf{t}}/(1-t) + (-1+t)\mathbf{1}\pi^{\mathsf{t}}/(1-t)\\ &= \quad t\bigl((I-tP)^{-1} - \mathbf{1}\pi^{\mathsf{t}}/(1-t)\bigr) - \mathbf{1}\pi^{\mathsf{t}}. \qquad (8.7)\end{aligned}$$

Thus, using Proposition 2.1 again,

$$\lim_{t\nearrow 1}\bigl(t(I-tP)^{-1} - \mathbf{1}\pi^{\mathsf{t}}/(1-t)\bigr) = (I-P)^{\mathsf{g}} - \mathbf{1}\pi^{\mathsf{t}}. \qquad (8.8)$$

Returning to the previous calculation in (8.6), we thus find by (8.8)

$$\widehat{Q}_{ij,i'j'} = (I-\widehat{P})^{\mathsf{g}}_{ij,i'j'} = \mathbf{1}_{i=i',j=j'} + P_{i',j'}(Q_{j,i'} - \pi_{i'}), \qquad (8.9)$$

where $Q = (I-P)^{\mathsf{g}}$ as before. Consequently, $\widehat{Q}' := \widehat{Q} - I$ is given by

$$\widehat{Q}'_{ij,i'j'} = P_{i',j'}(Q_{j,i'} - \pi_{i'}), \qquad (8.10)$$

Now we can directly apply the explicit moment formulas (3.15)–(3.19) to the expanded state space.

9. Further remarks

9.1. **Some extensions.** We give here some rather brief comments on extensions of the basic Theorems 3.2 and 3.4 for finite-state Markov chains. We discuss three such extensions separately; they may without difficulty be combined.

9.1.1. *Higher order expansions.* We have in our results only considered one-term Edgeworth expansions, with an extra term of order $n^{-1/2}$ and error of order $O(n^{-1})$. It is by the methods above possible to obtain also expansions with further terms and errors of order $n^{-m/2}$ for, in principle, any integer m. This is achieved by taking the Taylor expansion in (5.39) to higher order, and then arguing as above and in [6]. Note, however, that then also $\gamma(t)$ has to be expanded explicitly in (5.39), and thus the higher order terms in the expansion will also depend on the first derivatives $\lambda^{(m)}(1)$ of the eigenvalue $\lambda(z)$ at $z = 1$. These derivatives can be found e.g. by repeatedly differentiating the characteristic equation

$$\det\bigl(P(z) - \lambda(z)I\bigr) = 0 \tag{9.1}$$

at $z = 1$. We leave the details to the reader. See also [5] for asymptotic expansions with higher order terms for the original HH vs HT problem.

9.1.2. *Arbitrary initial values.* One generalization of Theorem 3.2 is to add to S_n some initial value, say $h(X_1)$ for a given function $h : \mathcal{W} \to \mathbb{Z}$; thus now

$$S_n := h(X_1) + \sum_{k=1}^{n} g(W_k). \tag{9.2}$$

An example where this occurs is if Alice and Bob score words of different lengths. For a simple example, suppose that Alice scores a point when HH appears, while Bob scores a point when HHT or THT appears. From the third coinflip on, this is exactly the original problem in disguise, but here Alice may also gain a point when the second coin is tossed. If we as in Section 4 consider the Markov chain consisting of W_k given by (4.2), now taking $\ell = 3$, then the net score is S_{n-2} given by (9.2) with a suitable g and $h := I_{\mathsf{HHH}} + I_{\mathsf{HHT}}$.

In this version, Lemmas 5.2–5.4 still hold, provided we replace $\pi_1(z)$ defined in (5.6) by

$$\pi_1^h(z) = (\pi_1^h(z)_i)_{i\in\mathcal{W}} \quad \text{where} \quad \pi_1^h(z)_i := \pi_{1;i} z^{g(i)+h(i)}. \tag{9.3}$$

However, this modification changes also $\eta(z)$ by (5.25), and since $\gamma'(0) = \mathrm{i}\eta'(1)$ by (5.28), $\gamma'(0)$ may be modified and is in general no longer 0, so Corollary 5.5 does not hold. In fact, by taking derivatives in (5.25) and comparing with the original case with $h = 0$ (for which $\eta'(1) = \gamma'(0) = 0$ by the proof in Section 5), we see that

$$\eta'(1) = 0 + \sum_{j\in\mathcal{W}} \pi_{1;j} h(j) = \mathbb{E}\, h(W_1), \tag{9.4}$$

and hence,

$$\gamma'(0) = \mathrm{i}\eta'(1) = \mathrm{i}\,\mathbb{E}\, h(W_1). \tag{9.5}$$

If we still let $\mu := \mathbb{E}\, X_k$, and also

$$\Delta := -\mathrm{i}\gamma'(0) = \eta'(1) = \mathbb{E}\, h(W_1), \tag{9.6}$$

then obviously

$$\mathbb{E}\, S_n = n\mu + \Delta. \tag{9.7}$$

In particular, Lemma 5.4 implies $\varkappa_1 = \mu$.

It is now necessary to include a term $\gamma'(0)t = \mathrm{i}\Delta t$ in the Taylor expansion (5.39). The rest of the proof goes through if we replace $n\mu$ by $n\mu + \Delta$, and thus (3.14) holds with this change. We may then replace x by $x - \Delta/(\sigma\sqrt{n})$, which using Taylor expansions yields

$$\mathbb{P}(S_n - n\mu \leqslant x\sigma\sqrt{n}) = \Phi(x) + \frac{\varkappa_3}{6\sigma^3\sqrt{2\pi n}}(1-x^2)e^{-x^2/2} - \frac{\Delta}{\sigma\sqrt{2\pi n}}e^{-x^2/2} + \frac{1}{\sigma\sqrt{2\pi n}}\vartheta(x\sigma\sqrt{n} - n\mu)e^{-x^2/2} + O\bigl(n^{-1}\bigr). \tag{9.8}$$

In other words, (3.14) holds with an extra term $-\Delta/(\sigma\sqrt{2\pi n})e^{-x^2/2}$. We omit the details. In Corollary 3.3, this means an extra term $-\Delta/(\sigma\sqrt{2\pi n})$ in (3.20) and (3.21), and thus twice as much in (3.22).

This should not be surprising. Consider for simplicity the case $\mu = 0$ and $x = 0$ in Corollary 3.3. If $h(W) := 1$ deterministically, so we just add 1 to S_n, then $\mathbb{P}(S_n \leqslant 0)$ is decreased by $\mathbb{P}(S_n = 0)$, which up to $O(1/n)$ equals $1/(\sigma\sqrt{2\pi n})$ by (3.23). Hence the result just shown says that if (9.2) holds so the score on the average is increased by $\Delta = \mathbb{E}\, h(W_1)$ compared to the standard case, then the probability $\mathbb{P}(S_n \leqslant 0)$ shifts by Δ times as much as if we instead add 1 deterministically.

9.1.3. *Arbitrary initial distribution.* We have asssumed in Theorems 3.2 and 3.4 that the Markov chain is stationary. We may generalize to an arbitrary initial distribution π_1 (or, in Theorem 3.4, π_0). Then Lemmas 5.2–5.4 still hold, with $\lambda(z)$ still given by (5.25). However, as in Section 9.1.2, in general $\gamma'(0) \neq 0$, so Corollary 5.5 does not hold. (We still have $\gamma'(0) = \mathrm{i}\eta'(1)$ by (5.28), and $\eta(z)$ depends on π_1 by (5.24).) If we define $\mu := \varkappa_1$, then Lemma 5.4 shows that $\mathbb{E}\, S_n/n = \kappa_1(S_n)/n \to \mu$, but in general $\mathbb{E}\, S_n$ is not equal to $n\mu$; in fact, (5.33) with $m = 1$ implies that

$$\mathbb{E}\, S_n - n\mu = \sum_{k=1}^{n}(\mathbb{E}\, X_k - \mu) \to \Delta := -\mathrm{i}\gamma'(0). \tag{9.9}$$

As in Section 9.1.2, it is now necessary to include a term $\gamma'(0)t = \mathrm{i}\Delta t$ in the Taylor expansion (5.39). This has the same consequences as in Section 9.1.2, and again we have (9.8) and its consequences.

Again this is not surprising: a different initial distribution implies by (9.9) an average extra score Δ, and (up to order n^{-1}) the probabilities in Theorem 3.2 and Corollary 3.3 shift by an amount proportional to this average extra score, just as in Section 9.1.2.

9.2. **Litt's game with several words each.** We observe that our result Theorem 3.2 can be applied to an even more general formulation of Litt's game, in which Alice and Bob each have several words that yield positive points for them (this generalization is considered by [5]); moreover, we may let different words give different numbers of points. As before, the value of each word is encoded in the function g (with a $\pm$ sign depending on whether it gives points to Alice or Bob).

If we can verify (3.9) holds for the chain so defined, then Theorem 3.2 applies; combining (3.15)–(3.19) with Proposition 4.1 gives explicit formulas for the asymptotics directly analogous with the two-word case (although the formulas do not simplify as much in the general case). Verifying (3.9) in this case is however more difficult, since we do not have an analogue of Theorem 7.2. However we can often verify (3.9) on a case-by-case basis by an adaptation of the proof of Theorem 7.2. For example, if neither Alice nor Bob gets points for $\mathsf{H}\cdots\mathsf{H}$, and furthermore $[z]\operatorname{Tr}(I-P(z))^{-1}>0$, then (3.9) holds (and hence everything goes through as in the two-word case). In general, it is an interesting question to what extent we can characterize the conditions under which (3.9) holds in the multiple-word case.

Acknowledgements

We thank Daniel Litt for posing the problem, Doron Zeilberger for enthusiasm and bringing us together on the problem, and Geoffrey Grimmett for inspiring and interesting comments.

Appendix A. Taylor expansion of eigenvalues

We give here a direct proof of the Taylor expansion (5.61)–(5.64) of the eigenvalue $\lambda(F(t))$. Both the result and the method are old, but we do not know an explicit reference so we give the details here for completeness. It is obvious that the calculations can be continued to the derivative of an arbitrary order, but we need only the first three derivatives.

Proposition A.1. *Let $t\mapsto F(t)$ be a C^3 (three times continuously differentiable) map defined on some neighborhood of 0 and with values $F(t)$ in the space of $m\times m$ matrices for some $m\geqslant 1$. Suppose that $F(0)$ has an eigenvalue λ_0 that is (algebraically) simple. Let Q be the group inverse of $\lambda_0 I-F(0)$ and let u^{t} and v be left and right eigenvectors of $F(0)$ for the eigenvalue 0, with the normalization $u^{\mathsf{t}}v=1$. (Such vectors exist since the eigenvalue is simple.) Then, for t in a possibly smaller neighborhood of 0, $F(t)$ has an eigenvalue $\lambda(t)$ such that $\lambda(0)=\lambda_0$ and $t\mapsto\lambda(t)$ is C^3 with, writing F',F'',F''' for $F'(0),F''(0),F'''(0)$,*

$$\lambda'(0)=u^{\mathsf{t}}F'v, \tag{A.1}$$

$$\lambda''(0)=u^{\mathsf{t}}F''v+2u^{\mathsf{t}}F'QF'v. \tag{A.2}$$

$$\begin{aligned}\lambda'''(0)=&\,u^{\mathsf{t}}F'''v+3u^{\mathsf{t}}F''QF'v+3u^{\mathsf{t}}F'QF''v+6u^{\mathsf{t}}F'QF'QF'v\\&-6(u^{\mathsf{t}}F'v)(u^{\mathsf{t}}F'Q^2F'v).\end{aligned} \tag{A.3}$$

Proof. By replacing $F(t)$ by $F(t)-\lambda_0 I$ we may assume that $\lambda_0=0$. Then Q is the group inverse of $-F(0)$. (Sorry for the minus sign with our choice of notation!)

The characteristic polynomial $f(\lambda;t):=\det(\lambda I-F(t))$ has coefficients that are C^3, and $\lambda_0=0$ is a simple root of $f(\cdot;0)$; thus $f(0,0)=0$ and $\frac{\partial f}{\partial\lambda}(0,0)\neq 0$. Hence the implicit function theorem shows the existence of a C^3 function $\lambda(t)$ that is a simple root of $f(t)$, and thus a simple eigenvalue of $F(t)$, with $\lambda(0)=0$. It remains only to compute the derivatives at 0.

It follows, e.g. using Cramer's rule, that for small t there exists an eigenvector $v(t)$ of $F(t)$ such that $v(t)$ is a C^3 function of t with

$$v(0)=v, \tag{A.4}$$

$$F(t)v(t)=\lambda(t)v(t), \tag{A.5}$$

$$u^{\mathsf{t}}v(t)=1. \tag{A.6}$$

We now differentiate the eigenvalue equation (A.5), which yields

$$\lambda'(t)v(t) + \lambda(t)v'(t) = F'(t)v(t) + F(t)v'(t). \tag{A.7}$$

Multiply to the left by u^{t}, and note that repeated differentiating of (A.6) yields

$$u^{\mathsf{t}}v^{(k)}(t) = 0, \qquad k \geqslant 1. \tag{A.8}$$

Hence, (A.7) yields, using (A.6) and (A.8),

$$\lambda'(t) = u^{\mathsf{t}}\bigl(\lambda'(t)v(t) + \lambda(t)v'(t)\bigr) = u^{\mathsf{t}}F'(t)v(t) + u^{\mathsf{t}}F(t)v'(t). \tag{A.9}$$

Taking $t = 0$ and recalling $u^{\mathsf{t}}F(0) = 0$ we obtain (A.1).

Since Q is the group inverse of $-F(0)$, and $F(0)v = 0$, we have

$$Qv = 0. \tag{A.10}$$

Moreover, (A.8) shows that $v'(0)$ is orthogonal to u, which (since the eigenvalue 0 of $F(0)$ is simple) spans the null space of the transpose matrix $F'(0)$; hence $v'(0) \in \operatorname{ran}(F(0))$, which implies (by (2.16)–(2.17)) that $v'(0) = -QF(0)v'(0)$. Hence (A.7) implies, using also $\lambda(0) = 0$ and (A.10),

$$v'(0) = -QF(0)v'(0) = -Q\bigl(\lambda'(0)v - F'(0)v\bigr) = QF'(0)v. \tag{A.11}$$

Differentiate (A.7) again; this yields

$$\lambda''(t)v(t) + 2\lambda'(t)v'(t) + \lambda(t)v''(t) = F''(t)v(t) + 2F'(t)v'(t) + F(t)v''(t). \tag{A.12}$$

Multiplying to the left by u^{t} yields, using (A.8) and (A.11),

$$\lambda''(t) = u^{\mathsf{t}}F''(t)v(t) + 2u^{\mathsf{t}}F'(t)v'(t) + u^{\mathsf{t}}F(t)v''(t), \tag{A.13}$$

which yields (A.2) by taking $t = 0$ and using (A.11) and (A.8).

Furthermore, just as for $v'(0)$, (A.8) shows that $v''(0)$ is orthogonal to u, and thus $v''(0) = -QF(0)v''(0)$. Hence (A.12) implies, using $\lambda(0) = 0$, (A.10), (A.11), and (A.1),

$$\begin{aligned} v''(0) &= -QF(0)v''(0) = -2\lambda'(0)Qv'(0) + Q\bigl(F''(0)v(0) + 2F'(0)v'(0)\bigr) \\ &= -2(u^{\mathsf{t}}F'(0)v)Q^2F'(0)v + QF''(0)v + 2QF'(0)QF'(0)v. \end{aligned} \tag{A.14}$$

Finally, differentiating (A.13) and taking $t = 0$ yields, since $u^{\mathsf{t}}F(0) = 0$,

$$\lambda'''(0) = u^{\mathsf{t}}F'''(0)v + 3u^{\mathsf{t}}F''(0)v'(0) + 3u^{\mathsf{t}}F'(0)v''(0), \tag{A.15}$$

and (A.3) follows by substituting (A.11) and (A.14). □

Appendix B. Alternative proof of (7.12)

We give here an independent and more “probabilistic” proof of (7.12) in the proof of Theorem 7.2; this proof bypasses the matrix computations of Lemma 7.3.

Let V_A^B be the number of times that a random walk started from A hits A prior to hitting B (we count the initial state as a hit). By Remark 7.5, the sum $\sum_{m\geqslant 0} W(m)$ in the proof of Theorem 7.2 equals the expectation $\mathbb{E}\, V_A^B$. Recall that the crux of the proof of Theorem 7.2 was to show that this sum is greater than $W(0) = 1$, in other words, that $\mathbb{E}\, V_A^B > 1$. We showed this by showing the formula (7.12), which is a special case of Lemma 7.3, proved above by algebraic calculations.

We give here an alternative proof of the following special case of Lemma 7.3, which is sufficient for our use of it in (7.12); the proof uses standard methods and known results for Markov chains. (We guess that similar arguments can be used to prove Lemma 7.3 in general, but we have not pursued this.)

Lemma B.1. *Let P be an irreducible stochastic matrix on a finite state space $\mathcal{W}$, and let π be its stationary distribution. Then, for any distinct states $A, B \in \mathcal{W}$,*

$$(I - P_{\neq B})^{-1}_{AA} = \mathbb{E}\, V_A^B = Q_{AA} - Q_{BA} + (Q_{BB} - Q_{AB})\frac{\pi_A}{\pi_B}. \tag{B.1}$$

As a first step in the proof, we introduce the random variable

$$T_A^B := \text{time that a walk started from } A \text{ first visits } A \text{ after having visited } B. \tag{B.2}$$

Lemma B.2. *With the notation defined above,*

$$\mathbb{E}\, V_A^B = \pi_A \, \mathbb{E}\, T_A^B. \tag{B.3}$$

Proof. Let γ be a random walk starting from A. Define the sequence of random times $T_0, T_1, \ldots$ as $T_0 := 0$ and

$$T_i := \min\{j : j > T_{i-1}, \gamma_j = A, \gamma_{j'} = B \text{ for some } T_{i-1} < j' < j\}; \tag{B.4}$$

in other words, T_i is the time of the first return to A after having visited B after T_{i-1}. Since the chain is irreducible and finite, all T_i are finite, almost surely. Moreover, by the Markov property, all $T_i - T_{i-1}$ are i.i.d. By definition, $T_1 - T_0 = T_1 = T_A^B$.

We define associated "rewards" as

$$V_i = \#\{j : 0 \leqslant j < T_i : \gamma_j = A\} \tag{B.5}$$

i.e., the number of visits to A before T_i. Similarly to the above, the differences $V_i - V_{i-1}$ are i.i.d. by the Markov property. Slightly less obviously, we have $V_1 - V_0 = V_A^B$. This is because T_1 is the time of the first visit to A which occurs after the first visit to B. Therefore, any visits to A prior to T_1 must in fact occur prior to the first visit to B.

By the law of large numbers we thus have, as $n \to \infty$, a.s.,

$$\frac{T_n}{n} = \frac{\sum_{i=1}^n (T_i - T_{i-1})}{n} \to \mathbb{E}\,(T_1 - T_0) = \mathbb{E}\, T_A^B, \tag{B.6}$$

$$\frac{V_n}{n} = \frac{\sum_{i=1}^n (V_i - V_{i-1})}{n} \to \mathbb{E}\,(V_1 - V_0) = \mathbb{E}\, V_A^B, \tag{B.7}$$

and thus

$$\frac{V_n}{T_n} \to \frac{\mathbb{E}\, V_A^B}{\mathbb{E}\, T_A^B}. \tag{B.8}$$

On the other hand, by the Markov chain ergodic theorem, a.s.

$$\frac{V_n}{T_n} = \frac{\text{number of visits to } A \text{ before } T_n}{T_n} \to \pi_A. \tag{B.9}$$

Finally, (B.3) follows by comparing (B.8) and (B.9). □

Proof of Lemma B.1. The first equality in (B.1) was noted in Remark 7.5. For the second equality, we introduce the mean first passage times m_{ij}. For distinct i and j, m_{ij} is the average number of steps for a walk started at i to reach j for the first time. We define, for convenience, $m_{ii} := 0$.

We can express $\mathbb{E}\, T_A^B$ in terms of the mean first passage times. Indeed, waiting for the first visit to A after passing through B is equivalent to waiting for the first visit to B, and then subsequently waiting for the first visit to A thereafter. So by linearity of expectation,

$$\mathbb{E}\, T_A^B = m_{AB} + m_{BA}. \tag{B.10}$$

Combining with Lemma B.2, we thus have

$$\mathbb{E}\, V_A^B = \pi_A(m_{AB} + m_{BA}). \tag{B.11}$$

On the other hand, the mean first passage times are known to be closely related to the entries of the group inverse. More precisely, we have

$$Q_{ij} = \pi_j(\tau'_j - m_{ij}), \qquad i, j \in \mathcal{W}, \tag{B.12}$$

where $\tau'_j := \sum_{k\neq j} \pi_k m_{kj}$ [13, Corollary 11.7]. As a simple consequence,

$$m_{ij} = \frac{Q_{jj} - Q_{ij}}{\pi_j}. \tag{B.13}$$

Combining (B.13) with (B.11) yields the second equality in (B.1). □

References

[1] Anne-Laure Basdevant, Olivier Hénard, Edouard Maurel-Segala, and Arvind Singh. On cases where Litt's game is fair. Preprint, 2024. arXiv:2406.20049

[2] Harald Cramér. *Mathematical Methods of Statistics.* Almqvist & Wiksell, Uppsala, 1945.

[3] Somnath Datta and William P. McCormick. On the first-order Edgeworth expansion for a Markov chain. *J. Multivariate Anal.* **44**:2 (1993), 345–359.

[4] F. Y. Edgeworth. The law of error. *Trans. Cambridge Philos. Soc.* **20** (1905), 36–65, 113–141.

[5] Shalosh B. Ekhad and Doron Zeilberger. How to Answer Questions of the Type: If you toss a coin n times, how likely is HH to show up more than HT? Preprint, 2024. arXiv:2405.13561v1.

Accompanying web page, including Maple package, 2024. https://sites.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/litt.html

[6] Carl-Gustav Esseen. Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law. *Acta Math.* **77** (1945), 1–125.

[7] William Feller. *An Introduction to Probability Theory and its Applications, Volume II*, 2nd ed., Wiley, New York, 1971.

[8] Geoffrey R. Grimmett. Alice and Bob on $\mathbb{X}$: reversal, coupling, renewal. Preprint, 2024. arXiv:2409.00732

[9] Allan Gut. *Probability: A Graduate Course*, 2nd ed., Springer, New York, 2013.

[10] Moshe Haviv, Ya'acov Ritov, and Uriel G. Rothblum. Taylor expansions of eigenvalues of perturbed matrices with applications to spectral radii of nonnegative matrices. *Linear Algebra Appl.* **168** (1992), 159–188.

[11] Lothar Heinrich. Some remarks on asymptotic expansions in the central limit theorem for m-dependent random variables. *Math. Nachr.* **122** (1985), 151–155.

[12] Christian Hipp. Asymptotic expansions in the central limit theorem for compound and Markov processes. *Z. Wahrsch. Verw. Gebiete* **69**:3 (1985), 361–385.

[13] Jeffrey J. Hunter. Generalized inverses of Markovian kernels in terms of properties of the Markov chain. *Linear Algebra Appl.* **447** (2014), 38–55.

[14] Svante Janson. On degenerate sums of m-dependent variables. *J. Appl. Probab.* **52**:4 (2015), 1146–1155.

[15] Svante Janson. On a central limit theorem in renewal theory. *Statistics & Probability Letters* **204** (2024), 109948.

[16] John E. Kolassa and Peter McCullagh. Edgeworth series for lattice distributions. *Ann. Statist.* **18**:2 (1990), 981–985.

[17] Daniel Litt. Post on 𝕏, March 16, 2024. `https://x.com/littmath/status/1769044719034647001`

[18] Wei-Liem Loh. On m-dependence and Edgeworth expansions. *Ann. Inst. Statist. Math.* **46** (1994), 147–164.

[19] Carl D. Meyer. The role of the group generalized inverse in the theory of finite Markov chains. *SIAM Review*, **17**:3 (1975), 443–464.

[20] Sergey V. Nagaev. Some limit theorems for stationary Markov chains. (Russian.) *Teor. Veroyatnost. i Primenen.* **2**:4 (1957), 389–416. English transl.: *Theory Probab. Appl.* **2**:4 (1957), 378–406.

[21] Sergey V. Nagaev. More exact limit theorems for homogeneous Markov chains. (Russian.) *Teor. Verojatnost. i Primenen.* **6**:1 (1961), 67–86. English transl.: *Theory Probab. Appl.* **6**:1 (1961), 62–81.

[22] James R. Norris. *Markov Chains.* Cambridge University Press, Cambridge, 1998.

[23] Valentin V. Petrov. *Sums of Independent Random Variables.* Springer-Verlag, Berlin, 1975.

[24] Wan Soo Rhee. An Edgeworth expansion for a sum of m-dependent random variables. *Internat. J. Math. Math. Sci.* **8**:3 (1985), 563–569.

[25] Yosef Rinott and Vladimir Rotar. On Edgeworth expansions for dependency-neighborhoods chain structures and Stein's method. *Probab. Theory Related Fields* **126**:4 (2003), 528–570.

[26] Simon Segert. A proof that HT is more likely to outnumber HH than vice versa in a sequence of n coin flips. Preprint, 2024. `arXiv`:2405.16660

[27] E. Seneta. *Nonnegative Matrices and Markov Chains.* 2nd. ed. Springer-Verlag, New York, 1981.

[28] Richard Serfozo. *Basics of Applied Stochastic Processes.* Springer-Verlag, Berlin, 2009.

[29] S. H. Siraždinov. Refinement of limiting theorems for stationary Markov chains. (Russian) *Doklady Akad. Nauk SSSR (N.S.)* **84** (1952), 1143–1146.

SJ: Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden

Email address: `svante.janson@math.uu.se`

URL: `http://www2.math.uu.se/~svantejs/papers`

MN: Department of Mathematics and Statistics, University of Guelph, 50 Stone Road E, Guelph, ON N1G 2W1, Canada

Email address: `nicam@uoguelph.ca`

URL: `https://nicam.uoguelph.ca/`

SS: New York, NY, USA

Email address: `simonsegert@gmail.com`