Analysis of String Counting Problems

Simon Segert

April 2024

1 Introduction

Consider the following probability puzzle: A fair coin is flipped n times. For each occurrence of HT in the resulting sequence, Bob gets a point, and for each occurrence of HH Alice gets a point. At the end of the sequence, who is more likely to be ahead?

The puzzle is subtle because even though both players have the same expected score, it turns out that their win probabilities are not equal. Several heuristic arguments that Bob wins more often have been given; moreover, it is possible to algorithmically compute the win probabilities and thereby rigorously verify this for a specific value of n. However, an analytic proof that works for every n has not been given. **We provide a formal proof that Bob wins more often for every n**.

The proof uses generating function technology, which allows us to derive several interesting and unexpected results as a byproduct. To state these, let $P_n(Bob)$ denote the probability of the event that Bob has more points after n flips, and similarly for Alice. Let $\Delta_n = P_n(Bob) - P_n(Alice)$ be the difference in win probabilities. We show that **the asymptotic expansion of the differences is** $\Delta_n = \frac{1}{2\sqrt{n\pi}} + O(n^{-3/2})$ and moreover that Δ_n **can be computed in** $O(n)$ **time**.

There are two major steps of the proof: the first consists of establishing the following explicit formula for the generating function:

$$f(t) := \sum_{n \geq 0} \Delta_n (2t)^n = \frac{1}{2} \frac{1}{\sqrt{(1-t)(1-2t)(2t^2+t+1)}} - \frac{1}{2} \frac{1}{1-t},$$

, and the second consists of showing that the coefficients are positive. We actually provide two independent arguments for the positivity of the coefficients, so in a sense we provide 1.5 proofs that Bob wins more often. The first argument uses closure properties of the class of absolutely monotonic functions, and the second relates $f(t)$ to the generating function of a certain manifestly positive combinatorial sequence.

We then discuss computational aspects of the problem, and in particular the existence of efficient algorithms to compute Δ_n and related quantities. Finally, we show that the algorithms generalize to a much wider class of Markovian random walk counting problems and give applications.

2 Proof of Main result

We will prove that $\Delta_n > 0$ for $n \geq 3$. We will split up the argument into three subsections. In Section 2.1, we prove the claimed form of the generating function. We then give two independent arguments for the positivity of the coefficients, in Sections 2.2 and 2.3.

2.1 Derivation of generating function

The objective of this section is to prove the following theorem:

Theorem 2.1. *The generating function $f(t) = \sum_n \Delta_n (2t)^n$ is given by*

$$f(t) = \frac{1}{2} \frac{1}{\sqrt{(1-t)(1-2t)(2t^2+t+1)}} - \frac{1}{2} \frac{1}{1-t}$$

As a first step, we will derive an expression for Δ_n in terms of a complex integral [1].

Lemma 2.1. *For any $n > 0$, we have*

$$2^n \Delta_n = \int_C \boldsymbol{1}^t \frac{A(z)^{n-1} - A(1/z)^{n-1}}{1-z} \boldsymbol{1} dz$$

where $A(z) = \begin{pmatrix} 1 & 1/z \\ 1 & z \end{pmatrix}$, $\boldsymbol{1}$ is the vector of ones, and C is any simple contour in the complex plane that encloses the origin.

Proof. Define the random variable X_n to be the outcome of the nth flip (i.e. H or T), and define the random variable Y_n to be Alice's score minus Bob's score after n flips. Finally, consider the values

$$p_{n,k,T} = P(Y_n = k | X_n = T)$$

with $p_{n,k,H}$ defined analagously. By the chain rule of probability:

$$p_{n,k,T} = \frac{1}{2} P(Y_n = k | X_{n-1} = T, X_n = T) + \frac{1}{2} P(Y_n = k | X_{n-1} = H, X_n = T)$$

For the first term, we know that neither player can get a point for a subsequence that starts with T. Thus if the last two flips are TT and the cumulative score is k, then the cumulative score must already have been k after only $n-1$ flips. So the first term is $P(Y_n = k | X_{n-1} = T, X_n = T) = P(Y_{n-1} = k | n-1 = T, X_n = T) = p_{n-1,k,T}$ (since Y_{n-1} is independent of X_n). As for the second term, if the last two flips are HT, then this gives 1 point for Bob, and therefore

[1] Note that, by convention, we will assume that all complex integrals are multiplied by a factor of $\frac{1}{2\pi i}$, which we do not explicitly write out

the score difference after the first $n-1$ flips must have been $k+1$. So this is $p_{n-1,k+1,H}$. Thus we obtain the recurrence

$$p_{n,k,T} = \frac{1}{2}p_{n-1,k,T} + \frac{1}{2}p_{n-1,k+1,H}$$

Similarly, we can derive the recurrence

$$p_{n,k,H} = \frac{1}{2}p_{n-1,k,T} + \frac{1}{2}p_{n-1,k-1,H}$$

Next, consider the probability generating function $\phi_{n,T}(z) = \sum_n p_{n,k,T}z^k$, with $\phi_{n,H}$ defined analagously.

By multiplying each of the two recurrences above by z^k and summing over k, we obtain the matrix equation:

$$\begin{pmatrix} \phi_{n,T}(z) \\ \phi_{n,H}(z) \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & z^{-1} \\ 1 & z \end{pmatrix}\begin{pmatrix} \phi_{n-1,T}(z) \\ \phi_{n-1,H}(z) \end{pmatrix}$$

and by induction:

$$\begin{pmatrix} \phi_{n,T}(z) \\ \phi_{n,H}(z) \end{pmatrix} = \frac{1}{2^{n-1}}\begin{pmatrix} 1 & z^{-1} \\ 1 & z \end{pmatrix}^{n-1}\begin{pmatrix} \phi_{1,T}(z) \\ \phi_{1,H}(z) \end{pmatrix} = \frac{1}{2^{n-1}}\begin{pmatrix} 1 & z^{-1} \\ 1 & z \end{pmatrix}^{n-1}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Letting $\phi_n(z) = \sum_n P(Y_n = k)z^k$ be the probability generating function for Y_n, we evidently have the relation $\phi_n(z) = \frac{1}{2}\phi_{n,T}(z) + \frac{1}{2}\phi_{n,H}(z)$. Thus:

$$\phi_n(z) = 2^{-n}\begin{pmatrix} 1 & 1 \end{pmatrix}^T\begin{pmatrix} 1 & z^{-1} \\ 1 & z \end{pmatrix}^{n-1}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Now, note that Y_n can only take finitely many values (since neither player can get more than n-1 points after n flips). In particular, $\phi_n(z)$ is a meromorphic function, so by the residue formula:

$$P(Y_n = k) = \int_C z^{-1-k}\phi(z)dz$$

for any simple closed contour C which contains the origin. If we take C to lie completely within the (open) unit disc $|z| < 1$, then we have

$$\begin{aligned} P_n(\mathit{Bob\ wins}) &= \sum_{k<0} P(Y_n = k) \\ &= \sum_{k<0}\int_C z^{-1-k}\phi(z)dz \\ &= \sum_{k\geq 0}\int_C z^k\phi_n(z)dz \\ &= \int_C \frac{\phi_n(z)}{1-z}dz \end{aligned}$$

where the geometric sum converges by assumption on C. By a symmetric argument,

$$P_n(Alice\ wins) = \int_{C'} \frac{\phi_n(z)}{1-1/z}\frac{dz}{z^2}$$

where now C' lies in the *complement* of the (closed) unit disc $|z| \leq 1$. In particular, we can take C' to be the "inverse" of C: $C' = \{1/z : z \in C\}$. Making the change of variable $z \mapsto 1/z$ we see that

$$P_n(Alice\ wins) = \int_C \frac{\phi_n(1/z)}{1-z}dz$$

Note that there is a negative sign that arises from the fact that the function $1/z$ reverses the orientation along the curve. Therefore

$$\Delta_n = \int_C \frac{\phi_n(z) - \phi_n(1/z)}{1-z}dz$$

where C lies in the interior of the unit disc. Note that the integrand has a removable singularity at $z = 1$ and no other singularities except for $z = 0$; this implies that the above formula for Δ_n is valid for any C that encloses the origin (i.e. we can drop the condition that C lies in the unit disc). □

Proof. (of Theorem 2.1)

Starting with the complex integral formula, we see that, for t in a sufficiently small neighborhood of the origin, we have

$$f(t) = t\mathbf{1}^t \int_C \frac{(I - tA(z))^{-1} - (I - tA(1/z))^{-1}}{1-z}\mathbf{1}$$

Doing a bit of algebra, the integral becomes

$$f(t) = t^3 \int_C \frac{z^3 - z^2 - z + 1}{(t^2z^2 + (t-1)z + (t-t^2))((t^2-t)z^2 + (1-t)z - t^2)}dz$$

Interestingly, both quadratic equations in the denominator have the same discriminant $\Delta(t) = 4t^4 - 4t^3 + t^2 - 2t + 1$. By the quadratic formula, we see the integral has four poles located at:

$$\frac{-t+1 \pm \sqrt{\Delta(t)}}{2t^2}$$

$$\frac{t-1 \pm \sqrt{\Delta(t)}}{2t^2 - 2t}$$

Now, for small t, we have $\sqrt{\Delta(t)} = 1 - t + O(t^3)$. Thus as $t \to 0$ the pole $\frac{-t+1+\sqrt{\Delta(t)}}{2t^2}$ tends to infinity, while the complementary one tends to zero. Similarly for the other pair of poles, the "+" sign one tends to zero while the other one tends to infinity. Since C is fixed, we see in particular that for sufficiently small t that $\frac{-t+1-\sqrt{\Delta(t)}}{2t^2}$ and $\frac{t-1+\sqrt{\Delta(t)}}{2t^2-2t}$ will lie in C while the other two will lie outside of it. Therefore, the expression for $f(t)$ reduces to computing the residues at these two poles. Since all four poles are simple, we can directly read off the residue at one of the poles r_i as

$$Res_{r_i} = P(r_i)/Q'(r_i)$$

where P and Q are respectively the numerator and denominator of the expression in the integral. At this point, we can verify the claimed form of f through mechanical calculation.

□

2.2 First proof of positivity

Our goal is to show that $f^{(k)}(0) > 0$ for each $k \geq 3$. This is similar to the notion of *absolute monotonicity*, but differs in that we only care about the derivatives at a single point, and moreover we want a strict inequality. It will therefore be useful at this point to study this class of functions in more detail.

Definition 2.1.1. *Let g be a function with a convergent power series expansion in a neighborhood of 0. We say that g is* ***strictly absolutely monotonic at the origin*** *(or "**samo**" for short) if $g^{(i)}(0) > 0$ for every $i \geq 0$. More generally, we say that g is* ***k-samo*** *if $g^{(i)}(0) \geq 0$ for $i < k$ and $g^{(i)}(0) > 0$ for $i \geq k$* [2]*.*

We can rephrase our goal as showing that f is 3-samo. We now state a few simple but useful facts which mirror the corresponding results for absolutely monotonic functions:

Proposition 2.1.1. *The sum of two k-samo functions is k-samo. Moreover, a positive scalar multiple of a k-samo function is k-samo.*

Proof. Clear. □

Proposition 2.1.2. *Let f be k-samo. Then $f - f(0)$ is $max(k, 1)$-samo.*

Proof. Clear. □

Proposition 2.1.3. *Let f be k_1-samo and g be k_2-samo. Then fg is $k_1 + k_2$-samo.*

Proof. By the iterated product rule, the kth derivative of the product is a positive linear combination of terms of the form $f^{(i)}(0)g^{(k-i)}(0)$ for $i = 0, \ldots, k$. We consider two cases, the first being $k < k_1 + k_2$. In this case we must have either $i < k_1$ or $k - i < k_2$, and therefore one of the terms in the product $f^{(i)}(0)g^{(k-i)}(0)$ is equal to zero.

Now consider the second case $k \geq k_1 + k_2$. The derivative is clearly a sum of non-negative terms, so we just need to show that at least one of them is strictly

[2] Note that 0-samo is synonymous with samo

positive. Setting $i = k_1$ we have $f^{(k_1)}(0)g^{(k-k_1)}(0)$. The first factor is clearly positive by assumption on f, and the second term is similarly positive because $k - k_1 \geq k_2$. $\square$

Proposition 2.1.4. *If $log f$ is k-samo, then so is f.*

Proof. Apply Faa di Bruno's formula to $e^{log f}$. $\square$

Proposition 2.1.5. *If f is k-samo then $\int_0^x f(t)dt$ is $k+1$-samo.*

Proof. Clear. $\square$

We can use the above properties of samo functions to reduce our problem to the following simpler one:

Proposition 2.1.6. *Let $h(t)$ be defined by*

$$h(t) = \frac{t+2}{2t^2+t+1} + \frac{1}{1-2t}$$

If $h(t)$ is samo, then the generating function $f(t)$ is 3-samo.

Proof. We can see by direct computation that

$$\frac{2t^2}{1-t}h(t) = \left(\log \frac{\sqrt{1-t}}{\sqrt{(1-2t)(2t^2+t+1)}}\right)'$$

Assume h is samo. The function $\frac{2t^2}{1-t}$ is clearly 2-samo, so by assumption on h and Proposition 2.1.3, we conclude that RHS is 2-samo. By Proposition 2.1.5, we conclude that $\log \frac{\sqrt{1-t}}{\sqrt{(1-2t)(2t^2+t+1)}}$ is 3-samo, and by Proposition 2.1.4, we see that $\frac{\sqrt{1-t}}{\sqrt{(1-2t)(2t^2+t+1)}}$ is 3-samo. By Proposition 2.1.2, we have $\frac{\sqrt{1-t}}{\sqrt{(1-2t)(2t^2+t+1)}} - 1$ is 3-samo. Finally, because $\frac{1}{1-t}$ is samo, we conclude by Proposition 2.1.3 that

$$f(t) = \frac{1}{2(1-t)}\left(\frac{\sqrt{1-t}}{\sqrt{(1-2t)(2t^2+t+1)}} - 1\right)$$

is 3-samo as desired.

$\square$

At this point, the natural thing is to analyze h.

Proposition 2.1.7. *The coefficients of the function $h(t)$ defined in Proposition 2.1.6 are given by*

$$h(t)[t^n] = 2Re(\phi^n) + 2^n$$

where $\phi = \frac{-1+\sqrt{-7}}{2}$ and $n \geq 0$.

Proof. It suffices to show that $\frac{t+2}{2t^2+t+1} = \sum_n 2Re(\phi^n)t^n$. To wit, we have for sufficiently small t that

$$\begin{aligned}
\sum_n 2Re(\phi^n)t^n &= 2Re(\frac{1}{1-t\phi}) \\
&= \frac{1}{1-t\phi} + \frac{1}{1-t\overline{\phi}} \\
&= \frac{1-t\overline{\phi}+1-t\phi}{|1-t\phi|^2} \\
&= \frac{2+t}{(1+t/2)^2+7t^2/4} \\
&= \frac{2+t}{1+t+2t^2}
\end{aligned}$$

□

As a simple corollary:

Proposition 2.1.8. *The function $h(t)$ is samo.*

Proof. By Proposition 2.1.7, we need to show that $2^n + 2Re(\phi^n) > 0$ for every n. We have

$$Re(\phi^n) \geq -|\phi|^n = -2^{n/2}$$

and therefore the coefficients are strictly positive provided that $n \geq 3$. The cases $n = 0, 1, 2$ are easily verified directly. □

Combining Propositions 2.1.6, 2.1.7 and 2.1.8, we have proven the following:

Theorem 2.2. *The coefficients $f(t)[t^n]$ of the generating function are strictly positive for all $n \geq 3$. In particular, Bob is more likely to win the game than Alice for any $n \geq 3$.*

2.3 Second proof of positivity

We now provide an alternative argument for the postivity of Δ_n, by relating it to combinatorial sequence which is manifestly positive. This proof uses the explicit form of the generating function $f(t)$ but is otherwise completely independent from the argument given in Section 2.2.

Definition 2.2.1. *Let $S \subset \mathbb{Z}^2_{\geq 0} - \{(0,0)\}$ be a finite subset of the non-negative plane lattice. Assume that there is an associated positive integer $c_s \geq 1$ for each $s \in S$; we denote the collection of all by $C = \{c_s\}_{s \in S}$. A **colored lattice path** is a finite sequence $\{(s_i, x_i)\}_i$ such that $s_i \in S$ and $x_i \in \{1, \ldots, c_{s_i}\}$ for each i (the path may be empty). The **endpoint** of the path is defined to be $\sum_i s_i \in \mathbb{Z}^2$.*

Note that the values c_s have an obvious interpretation as a collection of possible colors for each type of edge.

Definition 2.2.2. *Let S, C be as above, and let $(a,b) \in \mathbb{Z}^2$. The number of colored lattice paths with endpoint (a,b) is denoted by $N_{S,C}(a,b)$.*

We can now state the main result of this section.

Theorem 2.3. *The sequence Δ_n can be expressed in terms of a count of colored lattice paths:*

$$\Delta_n = \frac{N_{S,C}(n,n) - 1}{2^{n+1}}$$

where $S = \{(6,5),(0,1),(1,1),(3,3)\}$, $c_{(3,3)} = 2$, and $c_{(i,j)} = 1$ for $(i,j) \in S - \{(3,3)\}$. In particular $\Delta_n > 0$ for $n \geq 3$.

The proof will follow easily from two simple lemmas:

Lemma 2.2. *The bivariate generating function $G_{S,C}(x,y) = \sum_{(a,b)\in\mathbb{Z}^2_{\geq 0}} N_{S,C}(a,b)x^a y^b$ is given by*

$$G_{S,C}(x,y) = \frac{1}{1 - \sum_{(i,j)\in S} c_{(i,j)} x^i y^j}$$

Lemma 2.3. *(Stanley) Let f, g, h be univariate polynomials with $h(0) = 0$. Consider the bivariate function $G(x,y) = \frac{1}{1 - xf(xy) - yg(xy) - h(xy)}$. Then the Diagonal $D(t) := \sum_{n\geq 0} G(x,y)[x^n y^n]t^n$ is given by*

$$D(t) = \frac{1}{\sqrt{(1-h(t))^2 - 4tf(t)g(t)}}$$

Lemma 2.2 is standard, and the proof is a simple exercise. Lemma 2.3 is Exercise 6.15 in [Sta23],and a solution is also provided therein.

We now prove Theorem 2.3 from the two lemmas. Taking the S and C defined in the statment of the Theorem, we see from Lemma 2.2 that

$$\begin{aligned}
G_{S,C}(x,y) &= \frac{1}{1 - x^6y^5 - y - xy - 2x^3y^3} \\
&= \frac{1}{1 - x(x^5y^5) - y(1) - (2x^3y^3 + xy)} \\
&= \frac{1}{1 - xf(xy) - yg(xy) - h(xy)} \\
f(t) &:= t^5 \\
g(t) &:= 1 \\
h(t) &:= 2t^3 + t
\end{aligned}$$

Therefore by Lemma 2.3 the diagonal $D_{S,C}(t) := \sum_n N_{S,C}(n,n)t^n$ satisfies

$$\begin{aligned}
D_{S,C}(t) &= \frac{1}{\sqrt{(1-h(t))^2 - 4tf(t)g(t)}} \\
&= \frac{1}{\sqrt{(1-t-2t^3)^2 - 4t^6}}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{4t^4 - 4t^3 + t^2 - 2t + 1}} \\
&= 2f(t) + \frac{1}{1-t}
\end{aligned}
$$

where $f(t) = \sum_n \Delta_n (2t)^n$ and we have used Theorem 2.1 in the last line. Theorem 2.3 follows immediately.

3 Asymptotic Analysis

It is easy to see that $\Delta_n \to 0$, but a natural and non-obvious follow up question is how fast is the convergence? Using the explicit form of the generating function, this can answered with a simple application of the famous Darboux formula (cf. [Wil93]).

Notation-wise, we define $d_n := 2^n \Delta_n + 1/2$, which are the coefficients of the function $\tilde{f}(t) = \frac{1}{2} \frac{1}{\sqrt{(1-t)(1-2t)(2t^2+t+1)}}$. Observe that $\tilde{f}(t)\sqrt{1/2 - t}$ can be extended to a homolomorphic function in the disc $|z| < 1/2 + \epsilon$. Darboux's formula then implies the following asymptotic form:

$$
2^{-n} d_n = \frac{1}{2} \binom{n - 1/2}{n} + O(n^{-3/2}) = \frac{1}{2\sqrt{\pi}} \frac{\Gamma(n + 1/2)}{\Gamma(n+1)} + O(n^{-3/2})
$$

By Stirling's formula, we in particular obtain:

$$
\Delta_n = \frac{1}{2\sqrt{n\pi}} + O(n^{-3/2})
$$

4 Algorithms

We discuss methods for practical computation of Δ_n and related quantitites.

4.1 Recurrence Relation Algorithm

Here we provide an algorithm to compute Δ_n in $O(n)$ time. To do so, we will use the reparametrization $d_n = 2^n \Delta_n + 1/2$ from above and compute d_n. The main observation is that the generating function $\tilde{f}(t)$ for d_n is an inverse square root of a rational function. In particular, it satisfies an algebraic relation of the form $p(t)\tilde{f}(t) + q(t)\tilde{f}'(t) = 0$ for some polynomials p and q. By equating the terms to zero, we see that the coefficients of $\tilde{f}$ satisfy a linear recurrence relation. Crucially, the number of terms in the recurrence relation does not depend on n (i.e. it is bounded by the degrees of p and q). Therefore, directly iterating the recurrence allows us to compute the nth term in $O(n)$ time [3].

[3] note that we cannot use the repeated squaring trick because the coefficients depend on n

To give the details, we obtain by direct computation:

$$(\log \tilde{f})'(t) = \frac{-8t^3 + 6t^2 - t + 1}{(1-x)(1-2t)(2t^2+t+1)}$$

In particular, we get an equation:

$$(4t^4 - 4t^3 + t^2 - 2t + 1)\tilde{f}'(t) + (8t^3 - 6t^2 + t - 1)\tilde{f}(t) = 0$$

Setting the coefficients of the LHS to zero gives the recurrence:

$$(-8/n+4)d_{n-4} - (-6/n+4)d_{n-3} + (-1/n+1)d_{n-2} - (-1/n+2)d_{n-1} + d_n = 0$$

We can also plug in $d_n = 2^n\Delta_n + 1/2$ to obtain a recurrence directly for Δ_n:

$$\Delta_n \quad = \quad \frac{1}{n2^n} + \left(\frac{1}{2n} - \frac{1}{4}\right)\Delta_{n-4} + \left(-\frac{3}{4n} + \frac{1}{2}\right)\Delta_{n-3} + \left(\frac{1}{4n} - \frac{1}{4}\right)\Delta_{n-2} + \left(-\frac{1}{2n} + 1\right)\Delta_{n-1}$$

An obvious consequence is that this algorithm not only computes Δ_n in $O(n)$ time, but can in fact generate the entire sequence $\Delta_1, \ldots, \Delta_n$ in $O(n)$ time.

4.2 Fourier Transform algorithm

The above algorithm is probably the fastest possible if we just want to compute Δ_n. But we can also efficiently (in $O(n \log n)$ time) compute the full distribution of the random variable $Y_n =$ *Alice's score* $-$ *Bob's score*. In addition to returning more information than just the single value Δ_n, this algorithm has the major virtue that it seamlessly generalizes to a very wide range of related problems, while the recurrence-relation-based algorithm does not generalize as readily, as we will see in Section 5.

Recalling that $|Y_n| \leq n-1$, we will consider the vector $v \in \mathbb{R}^{2n}$ defined by

$$v_i = P(Y_n = i - (n-1)), i = 0, \ldots, 2n-1$$

The discrete fourier transform $\hat{v}$ is then given by

$$\begin{aligned}
\hat{v}_k &= \sum_j e^{-\frac{\pi i}{n}kj} v_j \\
&= \sum_j e^{-\frac{\pi i}{n}kj} P(Y_n = j - (n-1)) \\
&= \sum_j e^{-\frac{\pi i}{n}k(j+n-1)} P(Y_n = j) \\
&= e^{-k\frac{\pi i(n-1)}{n}} \sum_j e^{-\frac{\pi i}{n}kj} P(Y_n = j) \\
&= e^{-k\frac{\pi i(n-1)}{n}} \phi_n(e^{-\frac{\pi i}{n}k})
\end{aligned}$$

Recall that $\phi_n(z) = \mathbf{1}^t A(z)^{n-1}\mathbf{1}$ where A is a certain 2x2 matrix. This means that, for any z, we can compute $\phi_n(z)$ in $O(\log n)$ time by using the repeated squaring trick. We can thus compute the DFT $\{\hat{v_k}\}_{k=0,\ldots,2n-1}$ in $O(n \log n)$. Finally, we can obtain the desired values v_i from $\hat{v}$ in $O(n \log n)$ time using the Fast Fourier Transform algorithm.

5 generalizations

Most of the above algorithmic analysis can be extended to the following more general setting. Consider some Markov chain with a finite number K of states. Let $T(k', k)$ denote the transition probability from $k' \to k$. Furthermore, each transition has an associated value, which is assumed to be an integer. Let $V(k', k)$ be the value associated with the transition $k' \to k$. Let $x_1, \ldots, x_n$ be a Markovian random walk; this walk has an associated value, Y_n, given by the sum of the value of each transition;

$$Y_n = \sum_{i=1}^{n-1} V(x_i, x_{i+1})$$

The question is to compute the distribution of Y_n. Note that the original game can be recovered by taking $K = 2$, $T = \frac{1}{2}\mathbf{1}\mathbf{1}^T$, $V = \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}$.

Proposition 5.0.1. *Let $\pi_0 \in \mathbb{R}^K$ be the distribution over the value of x_1. Then the probability generating function $\phi_n(z) = \sum_k P(Y_n = k)z^k$ is given by*

$$\phi_n(z) = \mathbf{1}^T A^{n-1}(z)(T^{n-1})^T \pi_0$$

where $A(z)_{k',k} := T(k', k)z^{V(k',k)}$ for $1 \le k, k' \le K$.

Given this, the Fourier-transform based algorithm described in Section 4.2 can be adapted *mutatis mutandis*, and we conclude:

Corollary 5.0.1. *The full distribution of Y_n can be computed in time $O(n \log n)$, where the implicit constant depends on K and $V_{max} := max_{k',k}|V(k', k)|$*

The proof of the proposition is likewise a straightforward generalization of the argument given in Section 2.1. That is, we consider the auxiliary random variables $Y_{n,k}$ which is the distribution of the total value, conditioned on the final state being k. We have

$$\begin{aligned}
P(Y_{n,k} = y) &= P(Y_n = y | X_n = k) \\
&= \sum_{k'} P(Y_n = y, X_{n-1} = k' | X_n = k) \\
&= \sum_{k'} P(Y_n = y | X_n = k, X_{n-1} = k')P(X_n = k | X_{n-1} = k')
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{k'} P(Y_{n-1} = y - V(k',k) | X_n = k, X_{n-1} = k') T(k',k) \\
&= \sum_{k'} P(Y_{n-1} = y - V(k',k) | X_{n-1} = k') T(k',k)
\end{aligned}
$$

where we used the fact that Y_{n-1} and X_n are independent in the last line. Passing to the generating function:

$$
\begin{aligned}
\sum_y P(Y_{n,k} = y) z^y &= \sum_{k',y} P(Y_{n-1,k'} = y - V(k',k)) T(k',k) z^y \\
&= \sum_{k'} z^{V(k',k)} T(k',k) \sum_y P(Y_{n-1,k'} = y - V(k',k)) z^{y - V(k',k)} \\
&= \sum_{k'} z^{V(k',k)} T(k',k) \phi_{n-1,k'}(z) \\
&= \sum_{k'} A(k',k) \phi_{n-1,k'}(z)
\end{aligned}
$$

If we let $\phi_{n,\cdot}(z)$ be the row vector containing each $\phi_{n,k}(z)$, we can write the above in matrix form $\phi_{n,\cdot}(z) = \phi_{n-1,\cdot}(z)^T A(z)$. By induction:

$$\phi_{n,\cdot}(z) = \phi_{1,\cdot}(z) A(z)^{n-1} = \mathbf{1}^T A(z)^{n-1}$$

To obtain $\phi_n(z)$, we need to marginalize over X_n, which is evidently

$$\phi_n(z) = \mathbf{1}^T A(z)^{n-1} \pi_n$$

where π_n is n-step distribution of the Markov chain regarded as a column vector; i.e. $\pi_n = (T^{n-1})^T \pi_0$. This proves the proposition.

Is there an analogue of the linear recurrence relation from Section 4.1? If we consider the differences $\Delta_n = P(Y_n > 0) - P(Y_n < 0)$, then the argument from the proof of Theorem 2.1 carries over exactly to give:

$$\Delta_n = \int_C \mathbf{1}^T \frac{A^{n-1}(z) - A^{n-1}(1/z)}{1-z} (T^{n-1})^T \pi_0 dz$$

for any simple closed contour enclosing the origin. Now, we can no longer use the geometric series trick to sum $\sum_n \Delta^n t^n$ because A and T might not commute. However, if we make the assumption that π_0 coincides with the stationary distribution of the Markov chain (i.e. $\pi_0^t T = \pi_0$), then the T^{n-1} drops out and we obtain

$$\Delta_n = \int_C \mathbf{1}^T \frac{A^{n-1}(z) - A^{n-1}(1/z)}{1-z} \pi_0 dz$$

Now we can obtain the generating funcion as before:

$$f(t) := \sum_n \Delta_n t^n = \int_C \mathbf{1}^T \frac{(1 - tA^{n-1}(z))^{-1} - (1 - tA^{n-1}(1/z))^{-1}}{1-z} \pi_0 dz$$

The integrand is clearly a rational function of t and z. The roots of the denominator are algebraic functions of t (Newton-Pusieux theorem), so we conclude

Proposition 5.0.2. *If π_0 coincides with the stationary distribution of T, then $f(t)$ is an algebraic function.*

This has a definite consequence with respect to the computational complexity of the values Δ_n:

Proposition 5.0.3. *If π_0 is the stationary distribution of T, then the differences $\Delta_n = P(Y_n > 0) - P(Y_n < 0)$ satisfy a recurrence relation of the form $\sum_{i=0}^{d} P_i(n)\Delta_{n-i} = 0$, for some polynomials P_i. In particular, Δ_n can be computed in $O(n)$ time .*

Proof. We know that $f(t)$ is an algebraic function, so by well-known results (e.g. Thm 6.4.6 in [Sta23]), there is a polynomial relation of the form

$$\sum_{i=0}^{d} Q_i(t) f^{(i)}(t) = 0$$

Extracting the coefficient of t^n from the left hand side gives the desired relation.

□

Note the subtle but important point that even though we known an $O(n)$ algorithm exists, it may be very difficult to explicitly describe the algorithm (since it depends on the roots of a certain polynomial, which may not have any simple expression in general). By contrast, the Fourier-transform based algorithm has no practical difficulties and is nearly as fast.

6 applications

The Fourier transform algorithm can trivially handle a wide class of follow ups to the original question. For example, what if Alice gets one point for HHH and three points for TTH and Bob gets two points for HHT or HTT?

A straightforward but interesting application is to simply replace HH and HT with two arbitrary subsequences (assumed to have the same length L). The algorithm is fast enough that filling out the full “win matrix” takes only a few seconds. These are shown below.

In this case, the states of the markov chain can be taken to be all subsequences of length $L - 1$, with the transition structure and associated values given in the obvious ways.

References

[Wil93] H Wilf. *generatingfunctionology.* Academic Press, 1993.

[Sta23] R Stanley. *Enumerative Combinatorics.* Vol. 2. Cambridge UP, 2023.

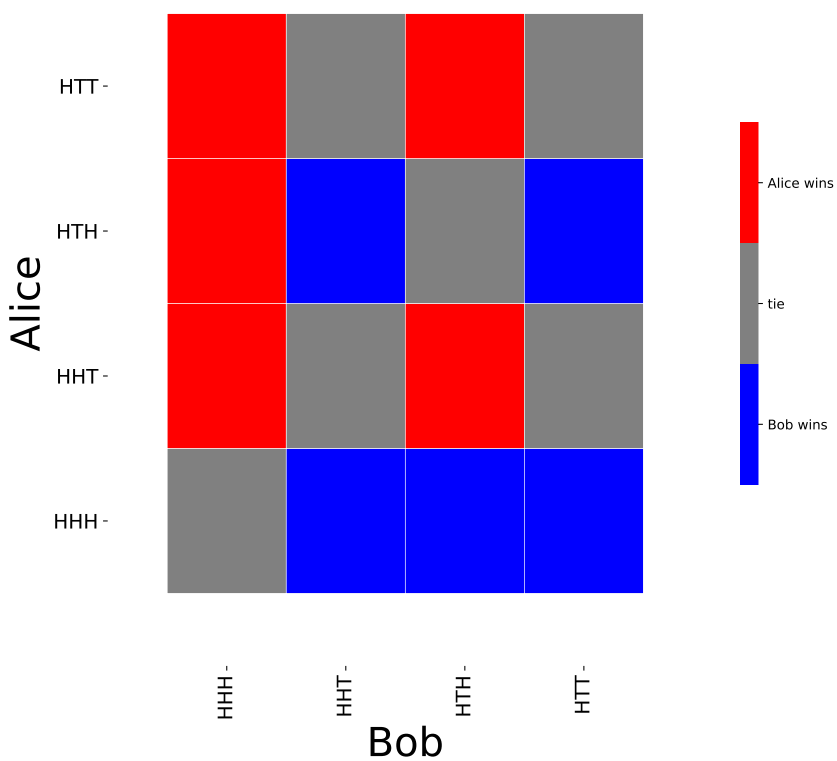

Figure 1: Win matrix for all pairs of subsequences of length 3 (modulo symmetry). The number of flips is 100 as in the original problem.

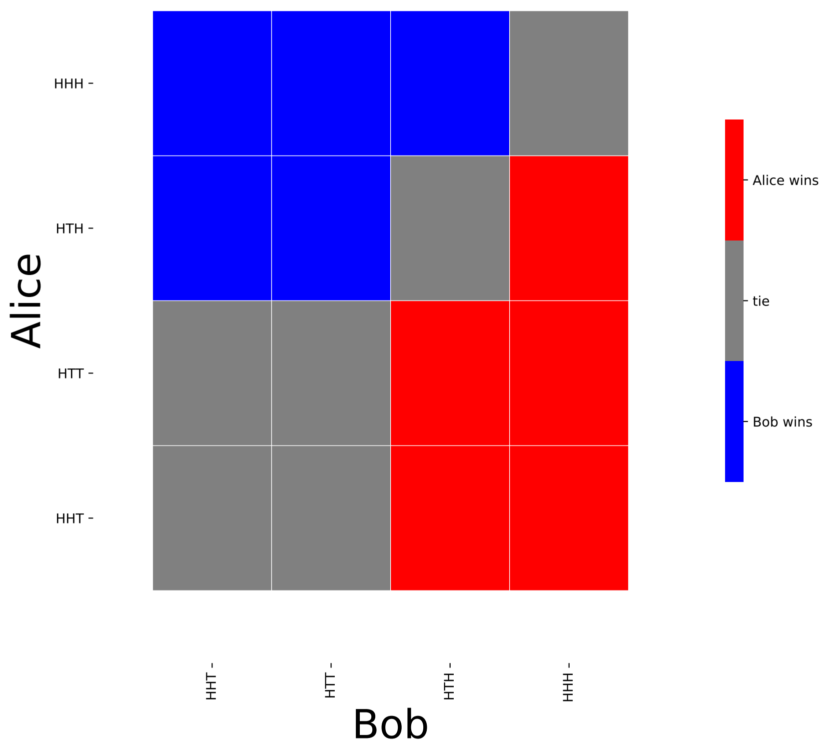

Figure 2: Same as Figure 1,except with sequences topologically ordered.

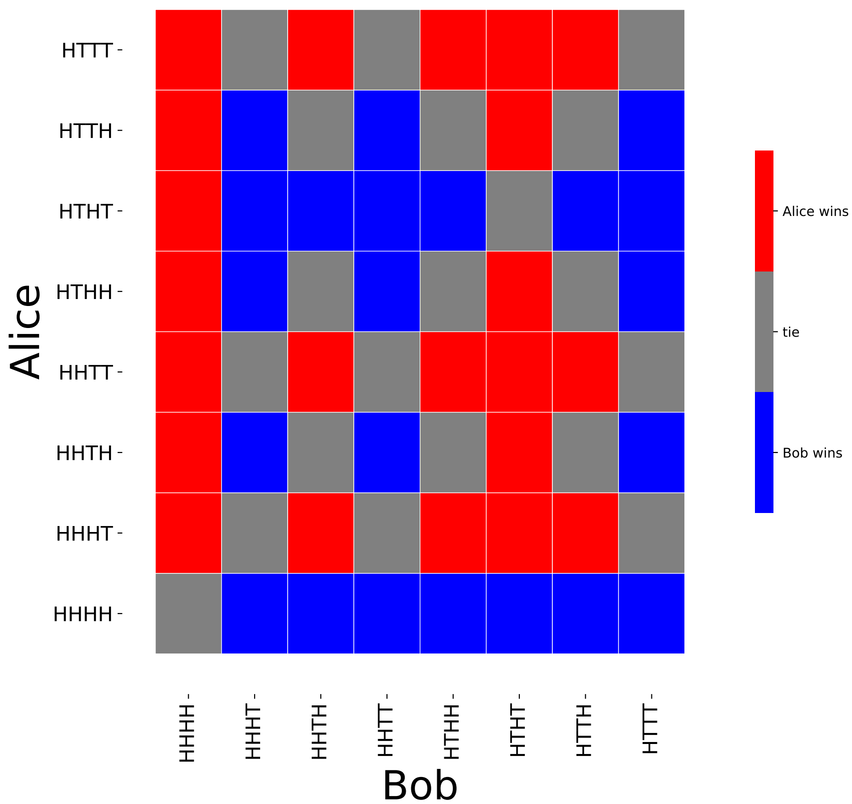

Figure 3: Analogous to Figure1.

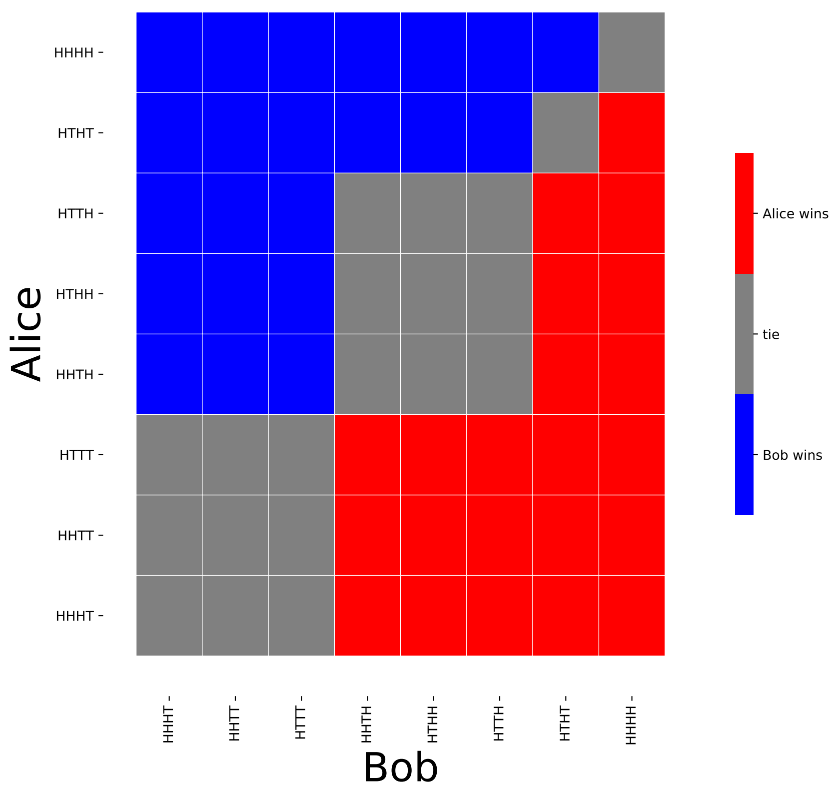

Figure 4: Analogous to Figure 2.

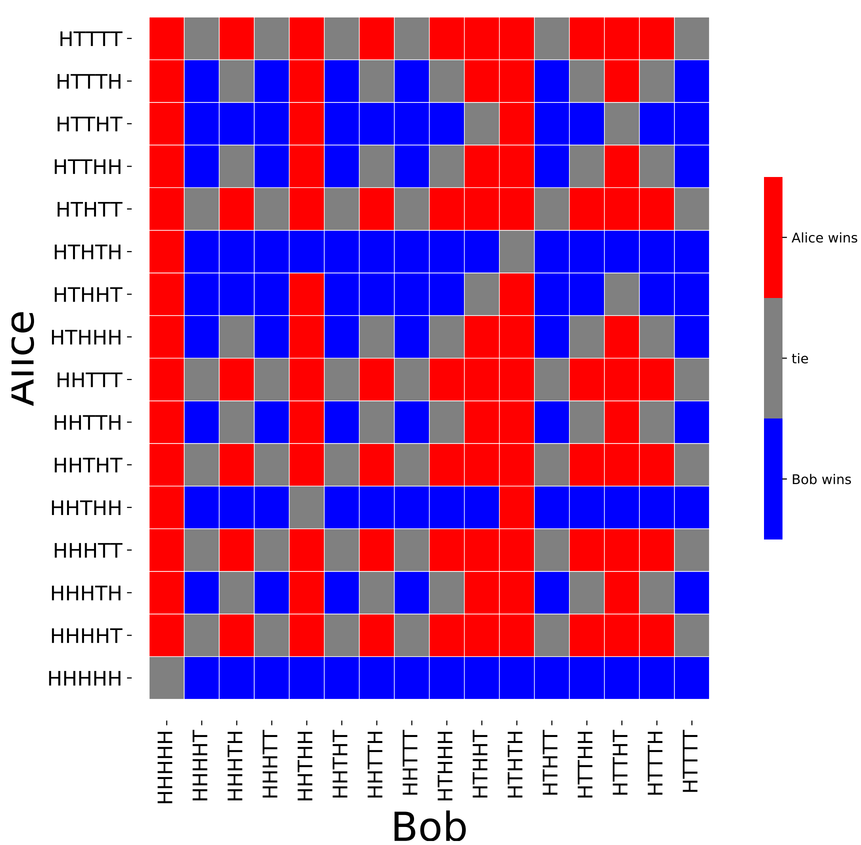

Figure 5: Analogous to Figure 1.

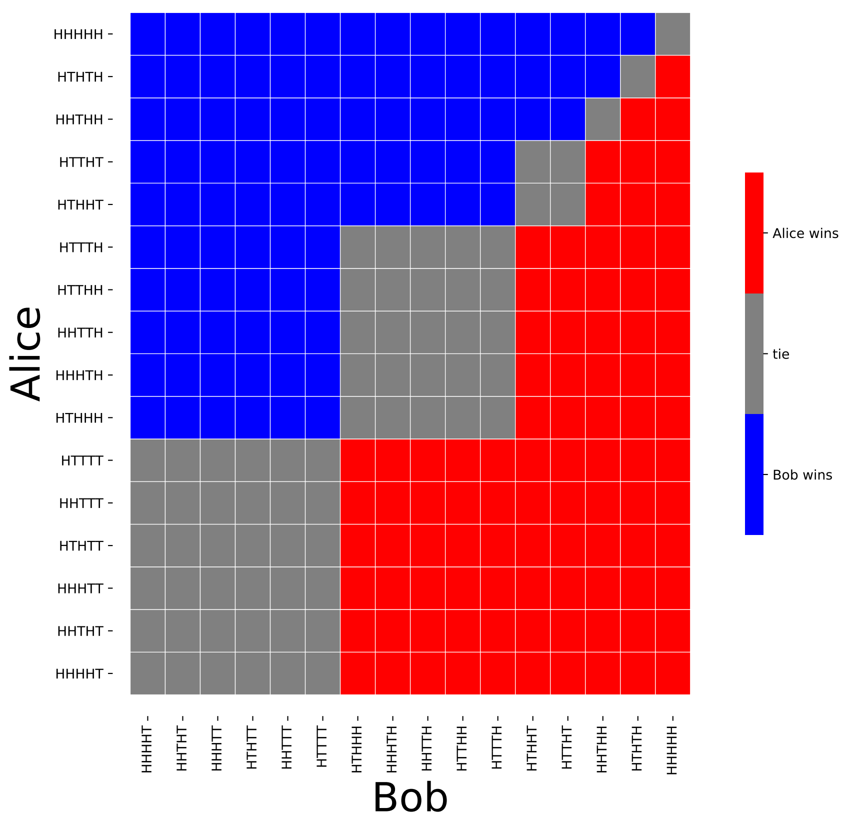

Figure 6: Analogous to Figure 2.