SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

Finding the Maximum of a Sum of Two Random Walks with Fixed Endpoints

av

Tove Nordmark

2024 - K20

Finding the Maximum of a Sum of Two Random Walks with Fixed Endpoints

Tove Nordmark

Självständigt arbete i matematik 15 högskolepoäng, grundnivå

Handledare: Daniel Ahlberg

2024

1 Abstract

This text concerns the problem of finding the argument maximum of a sum of two random walks with fixed endpoints at 0, by choosing an appropriate subset of indices based on realizations of one of these walks. A measure of the effectiveness of a strategy for choosing such a set of indices has been proposed and a numerical study has been made. Furthermore, a subset asymptotically containing $1/4$ of the indices which contains the argument maximum with probability 1 has been identified.
Denna text behandlar problemet att hitta index av maxpunkter för en summa av två slumpvandringar med fixa ändpunkter med värdet 0 genom att välja en delmängd av index baserat på ett utfall av en av slumpvandringarna. Ett mått på effektiviteten av en sådan strategi har föreslagits och en numerisk studie har utförts. Vidare har en delmängd identifierats, med asymptotisk storlek $1/4$ av antalet index och som med sannolikhet 1 innehåller åtminstone ett index för en maxpunkt.

2 Acknowledgements

I would like to extend my utmost gratitude to my supervisor, Daniel, for having had the patience to help me turn my arguments from intuition into rigorous mathematics.

Contents

3 Introduction

3.1 Background

The background necessary for this text mostly corresponds to a that of a typical undergraduate. Firstly, the reader ought to be comfortable with the concept of random variables and more generally the set theoretic foundation for probability as defined by Kolmogorov's axioms. A review of the subject may be found in for example [2].
Secondly, the reader should feel comfortable with elementary discrete mathematics such as the multiplicative principle, binomial coefficients and the comparison of sizes of sets by bijection. An overview may be found in [1].
Thirdly, the reader should be familiar with basic aspects of analysis, such as the definition of limits, the values of standard limits and so-called **big-O** notation. These ought to be treated in any analysis textbook for a first course in univariate analysis.
Finally, there are also references in the text to some definitions and results which are not necessarily familiar to students of an undergraduate level. One is the **countable subadditivity** of measures. An explanation may be found in [4]. In addition, a proof in the text rests on the **Borel-Cantelli lemma** which is outlined in [2].
Though this text concerns random walks, it doesn't use any results specifically relating to the theory of them. A further outlook into the topic, and into the relation between random walks and Brownian motion may be found in [3]. The book also contains exercises concerning **Brownian bridges** which are related to the concept of random walk bridges dealt with in this text.

3.2 Problem Outline

This text is essentially concerned with a single problem concerning discrete so-called **random walk bridges**. Let $\{Y_i^{\text{known}}\}_{i=0}^{2k}$ and $\{Y_i^{\text{unknown}}\}_{i=0}^{2k}$ be two discrete random walks of length $2k$ with fixed endpoints at 0, that is,

$$Y_0^{\text{known}} = Y_0^{\text{unknown}} = 0 \quad Y_{2k}^{\text{known}} = Y_{2k}^{\text{unknown}} = 0$$

In the jargon, we say that $\{Y_i^{\text{known}}\}_{i=0}^{2k}$ and $\{Y_i^{\text{unknown}}\}_{i=0}^{2k}$ are uniformly sampled among all random walk paths that start and end at zero. For a rigorous definition, see section 3.3.
Suppose now that are shown an outcome of the first walk $\{Y_i^{\text{known}}\}_{i=0}^{2k}$. Our task is now to systematically construct a set M of indices of the walk which with a certain probability, say $1/2$, contains the argument maximum of the sum

$$\{S_i\}_{i=0}^{2k} := \{Y_i^{\text{known}} + Y_i^{\text{unknown}}\}_{i=0}^{2k}$$

Specifically, we wish to find a strategy that results in choices of M which have a minimal expected/asymptotic cardinality.

3.3 Definitions

In the following, products of sets are formed using the Cartesian product. Define the set of outcomes of series of increments of walks of length $2k$

$$\Omega_k = \left\{ (x_1, x_2, \dots, x_{2k}) \in \{-1, 1\}^{2k} \,\middle|\, \sum_{i=0}^{2k} x_i = 0 \right\} \tag{3.1}$$

In other words, Ω_k is the set of sequences consisting of series of length $2k$ consisting of k copies of -1 and k copies of 1. For each sequence $\{x_i^{\text{known}}\}_{i=0}^{2k}$ in Ω_k, we define the outcome $\{y_i^{\text{known}}\}_{i=0}^{2k}$ of the walk $\{Y_i^{\text{known}}\}_{i=0}^{2k}$ by taking the cumulative sum of the increments,

$$y_0^{\text{known}} = 0, \qquad y_i^{\text{known}} = \sum_{j=1}^{i} x_j \quad (i = 1, 2, \ldots, 2k) \tag{3.2}$$

We define the outcomes of the walks $\{Y_i^{\text{unknown}}\}_{i=0}^{2k}$ in the corresponding way. We now let the probability of each outcome of a walk of a given length k be equal, by sampling the sequences of increments by a uniform distribution on the set Ω_k. For each k, the known and unknown walks are sampled independently. In the terminology of random walks, this means that $\{Y_i^{\text{known}}\}_{i=0}^{2k}$ and $\{Y_i^{\text{known}}\}_{i=0}^{2k}$ are simple and symmetric.

4 The Number of Peaks of a Random Walk with Fixed Endpoints

In the following section we shall find the distribution, expected value and asymptotic distribution of the size of the random set

$$M_k = \{i \in \{1, 2, \ldots, 2k-1\} : Y_i = Y_{i-1} + 1, Y_{i+1} = Y_{i-1}\} \tag{4.1}$$

We shall denote the size of this set by N_k. In this way, N_k may be regarded as the number of “peaks” of a random walk. To begin with, we shall prove a useful result.

Theorem 1. *For each positive integer k, the probability function of N_k is given by*

$$\mathbb{P}(N_k = j) = \frac{\binom{k}{j}^2}{\binom{2k}{k}} \tag{4.2}$$

Proof. The proof rests on finding the sizes of the sets

$$M_{k,j} := \{s \in \Omega_k : N_k = j\} \tag{4.3}$$

We shall show that for each pair of subsets of $\{1, 2, \ldots, k\}$, each with cardinality j, there is one and only one sequence $s \in \Omega_k$ that satisfies $N_k = j$.

Let $s = (x_1, x_2 \ldots x_{2k})$ be a sequence of increments for which the corresponding walk satisfies $N_k = j$. Let m_k denote the corresponding realization of M_k, and let $x_0 = 0$. For each $i = 1, 2, \ldots, 2k$, define

$$n^+(i) = |\{m \le i | x_j = 1\}|, \qquad n^-(i) = |\{m \le i | x_j = -1\}| \tag{4.4}$$

where $|\cdot|$ denotes cardinality. Next, define the following sets:

$$A^+ = \left\{n \in \{0, 1, 2, \ldots, 2k\} | \exists i \in m_k : n^+(i) = n\right\}, \qquad A^- = \left\{n \in \{0, 1, 2, \ldots, 2k\} | \exists i \in m_k : n^-(i+1) = n\right\} \tag{4.5}$$

We will endeavor to show that these form a pair of subsets of $\{1, \ldots, k\}$ containing j distinct elements each. To this end, let $i_1 < i_2 < \ldots i_j$ be an enumeration of m_k. Define

$$\Delta_l^+ := n^+(i_l) - n^+(i_{l-1}) \tag{4.6}$$

$$\Delta_l^- := n^-(i_l) - n^-(i_{l-1}) \tag{4.7}$$

It may now be realized that Δ_l^+ are positive for all $2 \leq l \leq j$, because of the fact that $x_{i_l} = 1$ for all $1 \leq l \leq j$, so that $n^+(i_l)$ is strictly greater than $n^+(i_{l-1})$ for all $2 \leq l \leq j$. Furthermore, Δ_l^- is strictly positive for all such l because $x_{i_l+1} = -1$ for all $1 \leq l \leq j$, so that $n^-(i_l)$ is strictly greater than $n^-(i_{l-1})$. Thus all j elements in A^+ and A^- are distinct. The fact that $x_{i_1} = 1$ and $x_{i_1+1} = -1$ together with the definition (4.5) further imply that all elements of A^+ and A^- are greater than or equal to 1. Furthermore, all elements are less than or equal to k since there are exactly k increments equal to 1, and exactly k increments equal to -1. Thus, (A^+, A^-) form a pair of subsets of $\{1, 2, \ldots, k\}$ with precisely j distinct elements each.
We have thus found a map, defined by (4.47), from $M_{k,j}$ to pairs of subsets of $\{1, 2, \ldots, k\}$ containing j elements each. We shall now show that this map is injective. To see this, we note that since all increments x_i for $1 \leq i \leq 2k$ belong to $\{1, -1\}$, we have the following:

$$\{m \leq i | x_m = 1\} \cup \{m \leq i | x_m = -1\} = \{m \leq i | x_m \in \{1, -1\}\} \tag{4.8}$$
$$= \{1, 2, \ldots, i\} \tag{4.9}$$

Since the sets in the above union are disjoint and by the definition (4.4), we thus obtain

$$n^+(i) + n^-(i) = i \tag{4.10}$$

In particular,

$$i \in m_k \iff n^+(i_l) + n^-(i_l + 1) = i + 1 \tag{4.11}$$

We now note that a sequence of increments $(x_1, \ldots, x_{2k})$ is uniquely determined by the indices m_k of peaks together with the height differences $y_{i_l}^{\text{known}} - y_{i_{l-1}}^{\text{known}}$ between peaks, where we let $i_0 = 0$. Now, suppose that we have two non-identical sequences of increments $(x_1, \ldots, x_{2k})$ and $(x'_1, \ldots, x'_{2k})$, and denote the corresponding walks by $\{y_i^{\text{known}}\}_{i=0}^{2k}$, $\{y'^{\text{known}}_i\}_{i=0}^{2k}$, respectively. Denote the corresponding pairs of sets formed by (4.5) by (A^+, A^-) and (A'^+, A'^-). Now, since the sequences of increments are non-identical, either the sets m_k, m'_k of indices of peaks are different, or the height differences $y_{i_l}^{\text{known}} - y_{i_{l-1}}^{\text{known}}$ and $y'^{\text{known}}_{i_l} - y'^{\text{known}}_{i_{l-1}}$ differ for some $l = 0, 1, \ldots j$. In the former case, equation (4.11) and the definition (4.5) imply that (A^+, A^-) and (A'^+, A'^-) must differ. In the latter case, the pairs of sets (A^+, A^-) and (A'^+, A'^-) must again differ because at least for some peak with index i_l, one of $n^+(i_l), n^-(i_l)$ must be different between the two walks.

Since the number of pairs of subsets of $\{1, 2, \ldots, k\}$ with precisely j distinct elements each is $\binom{k}{j}^2$ by the definition of binomial coefficients and the multiplicative principle, this means that

$$|M_{k,j}| \leq \binom{k}{j}^2 \tag{4.12}$$

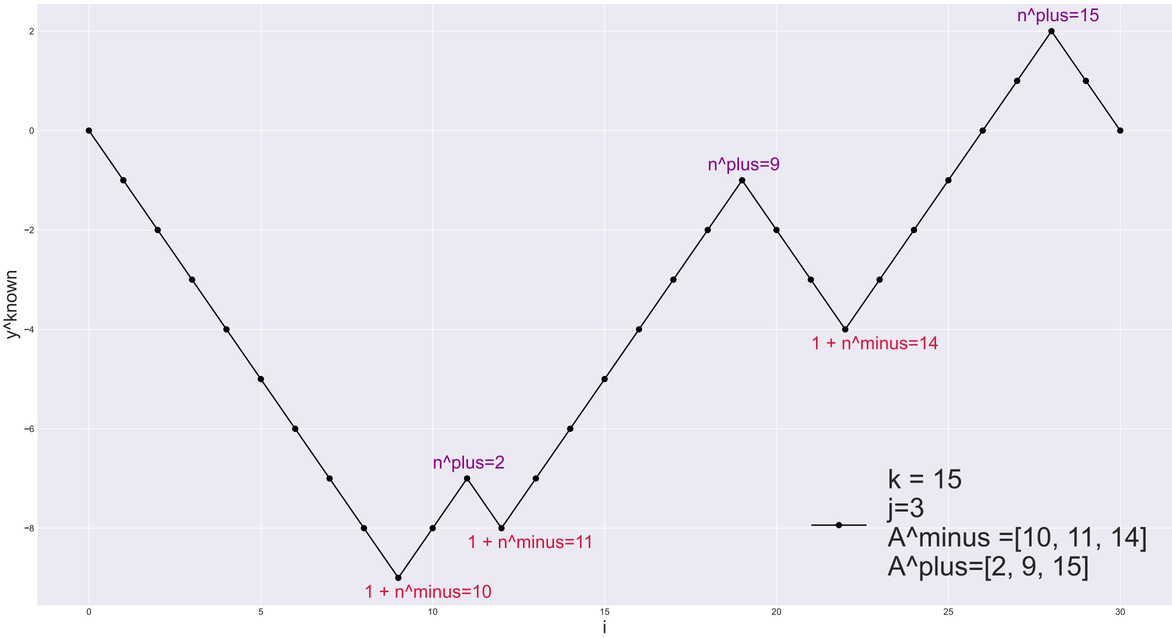

Figure 1: An illustration of the relationship between k, j, the pair of sets (A^+, A^-) and the values of n^+ and n^- at the points at which the walk changes direction. The walk was generated from a choice of (A^+, A^-) using the rules (4.13) and (4.14).

We shall now prove the equality of the two quantities by finding an injective map from pairs of subsets of the set $\{1, 2, \ldots, k\}$ with j elements each, to $M_{k,j}$. To this end, let A^+ and A^- be two subsets of the set $\{1, 2, \ldots, k\}$ with j elements each. Form the sequence $(x_1, x_2, \ldots, x_{2k})$ in the following way: Choose the first element according to the rule

$$x_1 = \begin{cases} 1 & 1 \in A^- \\ -1 & \text{otherwise} \end{cases} \tag{4.13}$$

and the remaining elements according to the rule

$$x_{i+1} = \begin{cases} -1 & x_i = 1 \text{ and } n^+(i) \in A^+ \\ 1 & x_i = -1 \text{ and } n^-(i) + 1 \in A^- \cup \{k\} \\ x_i & \text{otherwise} \end{cases} \qquad (i = 1, 2, \ldots, 2k-1) \tag{4.14}$$

In this way, the pair (A^+, A^-) determines the height differences between peaks, leading to a unique set m_k of indices of peaks, and thus the map defined by (4.5) is injective. For a visual example of how the sets (A^+, A^-) generate a unique walk with j maxima, see figure (1). We find that

$$|M_{k,j}| \geq \binom{k}{j}^2 \tag{4.15}$$

Together with (4.12), we thus have

$$|M_{k,j}| = \binom{k}{j}^2 \tag{4.16}$$

In addition, the total size of the set Ω_k is given by*

$$|\Omega_k| = \binom{2k}{k} \tag{4.17}$$

Using the classical definition of probability, equations (4.16) and (4.17) now yield

$$\mathbb{P}(N_k = j) = \frac{|M_{k,j}|}{|\Omega_k|} \tag{4.18}$$

$$\iff \mathbb{P}(N_k = j) = \frac{\binom{k}{j}^2}{\binom{2k}{k}} \tag{4.19}$$

and the result is established.

□

Since the distribution of N_k is symmetric about $k/2$, its expected value is $k/2$. As a curiosity, this implies that

$$\sum_{j=1}^{k} j \binom{k}{j}^2 = \frac{k}{2}\binom{2k}{k} \tag{4.20}$$

This result is also attainable through an algebraic argument, which is given below for the interested reader.

Lemma 1. *For each positive integer k, the following identity holds:*

$$\sum_{j=1}^{k} j \binom{k}{j}^2 = \frac{k}{2}\binom{2k}{k} \tag{4.21}$$

Proof. The argument is inspired by a derivation of a similar identity in [1]. We consider the identity

$$\frac{d}{dx}(1+x)^{2k} = 2(1+x)^k \frac{d}{dx}\left((1+x)^k\right) \tag{4.22}$$

By the binomial theorem, the left hand side of (4.22) becomes

$$\frac{d}{dx}(1+x)^{2k} = \sum_{i=1}^{2k} \binom{2k}{i} i x^{i-1} \tag{4.23}$$

Similarly, the right hand side of (4.22) is given by

$$2(1+x)^k \frac{d}{dx}\left((1+x)^k\right) = 2\left(\sum_{j=1}^{k}\binom{k}{j} x^j\right)\left(\sum_{j=1}^{k}\binom{k}{j} j x^{j-1}\right) \tag{4.24}$$

Now, two polynomials are equal if and only if their coefficients are equal. In particular, by (4.22), the coefficients before x^{k-1} in equations (4.23), (4.24) must be equal. Thus, by an elementary computation,

$$k\binom{2k}{k} = 2\sum_{j=1}^{k} j \binom{k}{j}\binom{k}{k-j} \tag{4.25}$$

*To see this, notice the fact that the indices of the k copies of 1 in a given element of Ω_k forms a subset of $\{1, \ldots, 2k\}$ of size k

Using the property $\binom{k}{j} = \binom{k}{k-j}$, we obtain

$$\sum_{j=1}^{k} j \binom{k}{j}^2 = \frac{k}{2}\binom{2k}{k} \tag{4.26}$$

as desired. □

We shall use the expression for the distribution of N_k established by theorem 1 to find its expected value and asymptotic distribution, but firstly we shall need the following result, taken from the field of information theory:

Lemma 2. *Let $H(p)$ denote the binary entropy function,*

$$H(p) := -p\log_2 p - (1-p)\log_2(1-p), \quad 0 < p < 1 \tag{4.27}$$

Then for all $n \geq 2$ and for all $1 \leq m < n$, the following holds

$$\log_2 \binom{n}{m} = nH\left(\frac{m}{n}\right) + O(\log_2 n) \tag{4.28}$$

Proof. We shall make use of *Stirling's approximation* in the form

$$\log_2(n!) = n\log_2 n - n\log_2 e + O(\log_2 n) \tag{4.29}$$

Which holds for positive integers n. The proof is now a simple matter of gathering terms and using elementary properties of logarithms. Let n be a positive integer and m be some integer between 0 and n. Define $\alpha = m/n$. By (4.29) and the definition of binomial coefficients as well as by the properties of the O-symbol, we have

$$\begin{aligned}\log_2\binom{n}{m} &= n\log_2 n - n\log_2 e + O(\log_2 n)\\ &\quad - n\alpha\log_2(\alpha n) + n\alpha\log_2 e + O(\log_2 n)\\ &\quad - n(1-\alpha)\log_2((1-\alpha)n) + n(1-\alpha)\log_2 e) + O(\log_2 n)\end{aligned} \tag{4.30}$$

$$\begin{aligned}\iff \log_2\binom{n}{m} &= n\left(\log_2 n - \alpha\log_2(\alpha n) - (1-\alpha)\log_2((1-\alpha)n)\right)\\ &\quad + (\alpha + (1-\alpha) - 1)n\log_2 e + O(\log_2 n)\end{aligned} \tag{4.31}$$

$$\iff \log_2\binom{n}{m} = n\left(\log_2 n - \alpha\log_2(\alpha n) - (1-\alpha)\log_2((1-\alpha)n)\right) + O(\log_2 n) \tag{4.32}$$

Rearranging the terms and using the property $\log a - \log b = \log(a/b)$, we may rewrite the above equation in the following way:

$$(4.31) \iff \log_2\binom{n}{m} = n\left(-(\alpha\log_2(\alpha n) - \alpha\log_2 n) - ((1-\alpha)\log_2((1-\alpha)n) - (1-\alpha)\log_2 n)\right) + O(\log_2 n) \tag{4.33}$$

$$= n\left(-\alpha\log_2(\alpha) - (1-\alpha)\log_2(1-\alpha)\right) + O(\log_2 n) \tag{4.34}$$

Recalling the definition (4.27) of the binary entropy function and that $\alpha := m/n$, we obtain the result

$$(4.34) \iff \log_2\binom{n}{m} = nH\left(\frac{m}{n}\right) + O(\log_2 n) \tag{4.35}$$

Q.E.D □

We may now prove the following:

Theorem 2.

$$\frac{N_k}{k} \overset{a.s.}{\to} \frac{1}{2} \quad as \quad k \to \infty \tag{4.36}$$

where $\overset{a.s.}{\to}$ *denotes convergence almost surely.*

Proof. We shall begin by proving that the series

$$\sum_{k=1}^{\infty} \mathbb{P}\left(N_k \leq k\left(\frac{1}{2} - \varepsilon\right)\right) \tag{4.37}$$

converges to a finite number for each $\varepsilon > 0$, from which the theorem will naturally follow by the *Borel-Cantelli lemma*[2], as well as the symmetry of the probability function of N_k about $\frac{k}{2}$ (see theorem 1). In the case when $\epsilon \geq 1/2$, the convergence is obvious as the terms are all 0. With this in mind, let ε be a number strictly between 0 and $1/2$. Let k be any positive integer. By theorem 1.

$$\mathbb{P}\left(N_k \leq k\left(\frac{1}{2} - \varepsilon\right)\right) = \frac{1}{\binom{2k}{k}} \sum_{j=0}^{k\left(\frac{1}{2}-\varepsilon\right)} \binom{k}{j}^2 \tag{4.38}$$

By the properties of binomial coefficients, the largest term in the above sum is for $j = \lfloor k\left(\frac{1}{2} - \varepsilon\right)\rfloor$. Since there are $\lfloor k\left(\frac{1}{2} - \varepsilon\right)\rfloor + 1$ terms in the sum, we thus obtain the following upper bound:

$$\mathbb{P}\left(N_k \leq k\left(\frac{1}{2} - \varepsilon\right)\right) \leq \frac{\lfloor k\left(\frac{1}{2} - \varepsilon\right)\rfloor + 1}{\binom{2k}{k}} \binom{k}{\lfloor k\left(\frac{1}{2} - \varepsilon\right)\rfloor}^2 \tag{4.39}$$

We now take the logarithm of both sides and obtain the asymptotic expression

$$\log_2 \mathbb{P}\left(N_k \leq k\left(\frac{1}{2} - \varepsilon\right)\right) \leq \log_2(\lfloor k\left(\frac{1}{2} - \varepsilon\right)\rfloor + 1) + 2\log\binom{k}{\lfloor k\left(\frac{1}{2} - \varepsilon\right)\rfloor} - \log_2\binom{2k}{k} \tag{4.40}$$

Using lemma 2, equation (4.40) now yields the following inequality:

$$\log_2 \mathbb{P}\left(N_k \leq k\left(\frac{1}{2} - \varepsilon\right)\right) \leq \log_2(\lfloor k\left(\frac{1}{2} - \varepsilon\right)\rfloor + 1) + \log_2(2k+1) - 2k\left(H\left(\frac{1}{2}\right) - H\left(\frac{\lfloor\left(\frac{1}{2} - \varepsilon\right)k\rfloor}{k}\right)\right) \tag{4.41}$$

We now define

$$\alpha_\varepsilon := 2\left(1 - H\left(\frac{1}{2} - \varepsilon\right)\right) \tag{4.42}$$

Now, since $H(1/2) = 1$, H is strictly increasing on the interval $(0, 1/2)$ and $\lfloor x\rfloor \leq x$ for all positive real numbers x, we find that

$$0 < \alpha_\varepsilon \leq 2\left(H\left(\frac{1}{2}\right) - H\left(\frac{\lfloor\left(\frac{1}{2} - \varepsilon\right)k\rfloor}{k}\right)\right) \tag{4.43}$$

Thus, equation (4.41) implies that

$$\mathbb{P}\left(N_k \leq k\left(\frac{1}{2} - \varepsilon\right)\right) \leq \left(k\left(\frac{1}{2} - \varepsilon\right) + 1\right)(2k+1) \cdot 2^{-\alpha_\varepsilon k} \tag{4.44}$$

where we have taken the (strictly increasing) exponential on both sides of the inequality and used the previously mentioned fact that $\lfloor k\left(\frac{1}{2} - \varepsilon\right)\rfloor \leq k\left(\frac{1}{2} - \varepsilon\right)$.

By a standard result, this establishes that there exist positive real numbers C_ε and β_ϵ, and a positive integer n such that for all $k \geq n$,

$$\mathbb{P}\left(N_k \leq k\left(\frac{1}{2} - \varepsilon\right)\right) \leq C_\varepsilon \cdot e^{-\beta_\varepsilon k} \tag{4.45}$$

Thus, the sum

$$\sum_{k=1}^{\infty} \mathbb{P}\left(N_k \leq k\left(\frac{1}{2} - \varepsilon\right)\right) \tag{4.46}$$

converges to a real number. Therefore, by the *Borel-Cantelli lemma*, the probability that the events $(N_k \leq k\left(\frac{1}{2} - \varepsilon\right))$ occur for an infinite number of k is 0. Now, since by theorem 1, the probability function of N_k is symmetric about $k/2$, we may similarly argue that the probability of $\left(N_k \geq k\left(\frac{1}{2} + \varepsilon\right)\right)$ occurring for an infinite number of k is also 0. Thus, if we define $A_k(\varepsilon)$ as the event

$$A_k(\varepsilon) := \left\{\frac{N_k}{k} \notin \left(\frac{1}{2} - \varepsilon, \frac{1}{2} + \varepsilon\right)\right\} \tag{4.47}$$

it must be the case that for each $0 < \varepsilon < 1/2$,

$$\mathbb{P}\left\{A_k(\varepsilon) \quad \text{i.o.}\right\} = 0 \tag{4.48}$$

where i.o. denotes that the events occurs for an infinite number of indices $k \geq 1$. This is obviously also true when $\varepsilon \geq 1/2$. In particular, it is true for all positive rational ε. By the countable subadditive property of the probability measure, we thus have

$$0 \leq \mathbb{P}\left(\bigcup_{\varepsilon\in\mathbb{Q},\varepsilon>0} \{A_k(\varepsilon) \quad \text{i.o.}\}\right) \leq \sum_{\varepsilon\in\mathbb{Q},\varepsilon>0} \mathbb{P}\left\{A_k(\varepsilon) \quad \text{i.o}\right\} = 0 \tag{4.49}$$

so that

$$\mathbb{P}\left(\bigcup_{\varepsilon\in\mathbb{Q},\varepsilon>0} \{A_k(\varepsilon) \quad \text{i.o}\}\right) = 0 \tag{4.50}$$

Furthermore, suppose $\varepsilon > 0$ and let $\eta > \varepsilon$. Then if the outcomes N_k/k fall outside of the interval $(1/2 - \eta, 1/2 + \eta)$ for an infinite number of indices k, then they must also fall outside of the interval $(1/2 - \varepsilon, 1/2 + \varepsilon)$ for an infinite number of indices k. Thus it is the case that

$$\{A_k(\eta) \quad \text{i.o.}\} \subseteq \{A_k(\varepsilon) \quad \text{i.o.}\} \tag{4.51}$$

We may therefore write

$$\{A_k(\varepsilon) \quad \text{i.o}\} = \bigcup_{\eta\geq\epsilon} \{A_k(\eta) \quad \text{i.o.}\} \tag{4.52}$$

This in turn implies that

$$\begin{aligned}\bigcup_{\varepsilon\in\mathbb{Q},\varepsilon>0} \{A_k(\varepsilon) \quad \text{i.o}\} &= \bigcup_{\varepsilon\in\mathbb{Q},\varepsilon>0} \bigcup_{\eta\geq\epsilon} \{A_k(\eta) \quad \text{i.o}\} &(4.53)\\ &= \bigcup_{\varepsilon>0} \{A_k(\varepsilon) \quad \text{i.o}\} &(4.54)\end{aligned}$$

We may thus now “fill in” the irrational numbers in equation (4.50), to obtain

$$\mathbb{P}\left(\bigcup_{\varepsilon>0} \{A_k(\varepsilon) \quad \text{i.o.}\}\right) = 0 \tag{4.55}$$

In order to further relate the above statement to the definition of a limit, we use the fact that by our definition (4.47),

$$\Omega\backslash\{A_k(\varepsilon) \quad \text{i.o.}\} = \Omega\backslash\left\{\frac{N_k}{k} \notin \left(\frac{1}{2}-\varepsilon, \frac{1}{2}+\varepsilon\right) \quad \text{i.o.}\right\} = \left\{\exists n \geq 1 : \forall k \geq n : \frac{1}{2}-\varepsilon < \frac{N_k}{k} < \frac{1}{2}+\varepsilon\right\} \tag{4.56}$$

This is the case because if there is a finite number of k such that N_k/k does not fall within a given interval, there must be a largest such k. Thus, for all larger k, N_k/k must fall within the interval. Now, by set arithmetic, Kolmogorov's axioms for probability and equation (4.55), we find that

$$\mathbb{P}\left(\Omega \setminus \bigcup_{\varepsilon>0} \{A_k(\varepsilon) \quad \text{i.o.}\}\right) = 1 \tag{4.57}$$

$$\iff \mathbb{P}\left(\bigcap_{\varepsilon>0} \Omega \setminus \{A_k(\varepsilon) \quad \text{i.o.}\}\right) = 1 \tag{4.58}$$

$$\iff \mathbb{P}\left(\bigcap_{\varepsilon>0}\left\{\exists n \geq 1 : \forall k \geq n : \frac{1}{2}-\varepsilon < \frac{N_k}{k} < \frac{1}{2}+\varepsilon\right\}\right) = 1 \tag{4.59}$$

The above statement is equivalent to

$$\mathbb{P}\left\{\forall\varepsilon > 0 : \exists n \geq 1 : \forall k \geq n : \frac{1}{2}-\varepsilon < \frac{N_k}{k} < \frac{1}{2}+\varepsilon\right\} = 1 \tag{4.60}$$

Thus, by the definition of a limit, we have established the result

$$\mathbb{P}\left\{\frac{N_k}{k} \to \frac{1}{2} \quad \text{as} \quad k \to \infty\right\} = 1 \tag{4.61}$$

and thus, by definition,

$$\frac{N_k}{k} \overset{a.s.}{\to} \frac{1}{2} \qquad \text{as} \qquad k \to \infty \tag{4.62}$$

Q.E.D. □

5 Maxima of a Sum of Random Walks

We shall now turn to the problem of finding the maximum of the sum of two random walks. In the following section, we shall be concerned with the sum

$$\{S_i\}_{i=0}^{2k} := \{Y_i^{\text{known}} + Y_i^{\text{unknown}}\}_{i=0}^{2k} \tag{5.1}$$

of the random walks $\{Y_i^{\text{known}}\}_{i=1}^{2k}$ and $\{Y_i^{\text{unknown}}\}_{i=1}^{2k}$ satisfying

$$Y_0^{\text{known}} = Y_0^{\text{unknown}} = 0, \qquad Y_{2k}^{\text{known}} = Y_{2k}^{\text{unknown}} = 0$$

In particular, we wish to find a subset I_k of indices of $\{Y^{\text{known}}\}_{i=0}^{2k}$ such that I_k contains the argument maximum of the sum $\{S_i\}_{i=0}^{2k}$ with a given probability. The set of indices of "peaks" M_k described in the previous section provides a good starting point, as illustrated by the following theorem:

Theorem 3. *If we view the index of the walk as being modulo* $2k$*, that is, that we view the indices* 0 *and* $2k$ *as identical, we have the following: The union of set of indices*

$$M_k = \left\{i \in \{1, 2, \ldots, 2k-1\} : Y_i^{known} = Y_{i-1}^{known} + 1, Y_{i+1}^{known} = Y_{i-1}^{known}\right\} \tag{5.2}$$

and the set $\{2k\}$ *contains the argument maximum of the sum* $\{S_i\}_{i=0}^{2k}$ *with probability* 1.

Proof. Suppose $\{y_i^{\text{known}}\}_{i=0}^{2k}$ is a realization of the walk $\{Y_i^{\text{known}}\}_{i=0}$, $\{x_i\}_{i=0}^{2k}$ be its sequence of increments, and m_k be the corresponding realization of M_k. Suppose further that t_1 and t_2 are two consecutive elements of m_k, in the sense that $t_1 < t_2$, and that there is no $t \in m_k$ such that $t_1 < t < t_2$. Then there is **no** occurrence of the pattern

$$x_{t_1+i} = 1, x_{t_1+i+1} = -1$$

for $i = 1, 2 \ldots t_2 - t_1$, for if that were the case, then there would exist $t \in m_k$ such that $t_1 < t < t_2$. Therefore, there exists a positive integer j such that

$$x_{t_1+i} = \begin{cases} -1 & 0 < i \leq j \\ 1 & j < i \leq t_2 - t_1 \end{cases} \tag{5.3}$$

Therefore it is also the case that

$$y_{t_1+i}^{\text{known}} = \begin{cases} y_{t_1}^{\text{known}} - i & 0 < i \leq j \\ y_{t_2}^{\text{known}} - ((t_2 - t_1) - i) & j < i \leq t_2 - t_1 \end{cases} \tag{5.4}$$

An illustration of the structure of the known walk between the indices t_1 and t_2 is found in figure 2.

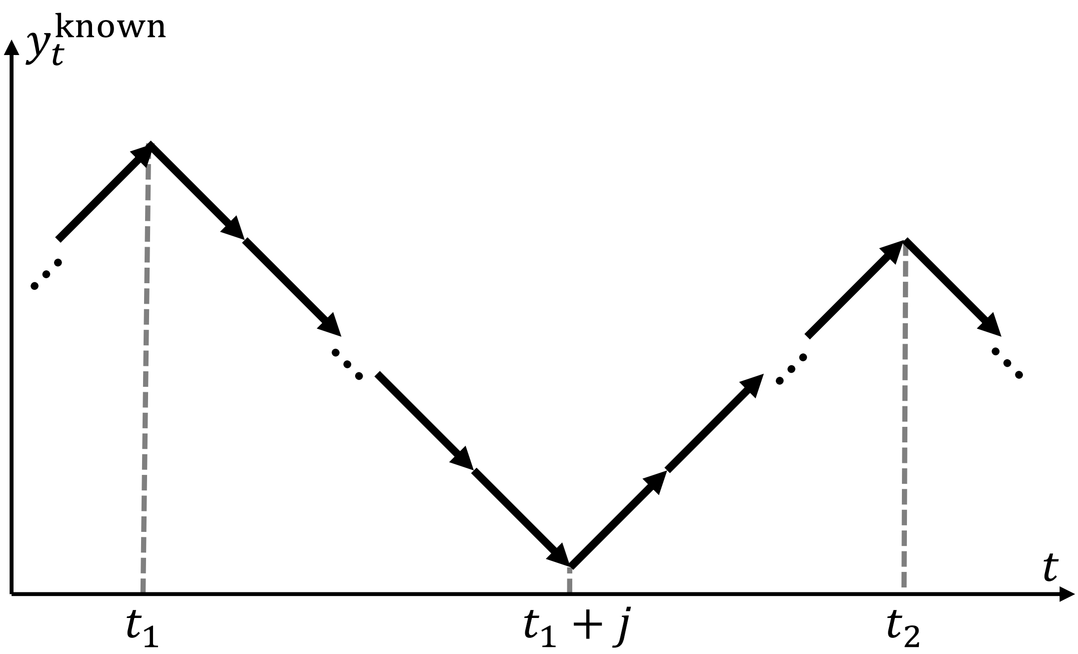

Figure 2: The structure of the known walk between the indices t_1 and t_2.

Now, we shall consider the possible values of the second walk $\{Y_i^{\text{unknown}}\}_{i=0}^{2k}$.
We shall start with the case $i = 1, 2, \ldots, j$ where j is as in (5.4). Since each increment of the second walk either takes the value -1 or 1, the following inequality holds:

$$Y_{t_1+i}^{\text{unknown}} \leq Y_{t_1}^{\text{unknown}} + i \tag{5.5}$$

Using (5.4), we have in the case $0 < i \leq j$:

$$\begin{aligned} y_{t_1+i}^{\text{known}} + Y_{t_1+i}^{\text{unknown}} &\leq y_{t_1+i}^{\text{known}} + Y_{t_1}^{\text{unknown}} + i & (5.6) \\ &= (y_{t_1}^{\text{known}} - i) + Y_{t_1}^{\text{unknown}} + i & (5.7) \\ &= y_{t_1}^{\text{known}} + Y_{t_1}^{\text{unknown}} \quad 0 < i \leq j & (5.8) \end{aligned}$$

In the case $i = j+1, j+2, \ldots, t_2 - t_1 - 1$, we can instead use the inequality

$$Y_{t_1+i}^{\text{unknown}} \leq Y_{t_2}^{\text{unknown}} + ((t_2 - t_1) - i) \tag{5.9}$$

Again, this holds because the value of each increment is either 1 or -1. Using (5.4), the inequality can be rewritten in the following way for the case $i = j+1, j+2, \ldots, t_2 - t_1$:

$$\begin{aligned} y_{t_1+i}^{\text{known}} + Y_{t_1+i}^{\text{unknown}} &\leq y_{t_1+i}^{\text{known}} + Y_{t_2}^{\text{unknown}} + ((t_2 - t_1) - i) && (5.10)\\ &= (y_{t_2}^{\text{known}} - ((t_2 - t_1) - i)) + Y_{t_2}^{\text{unknown}} + ((t_2 - t_1) - i) && (5.11)\\ &= y_{t_2}^{\text{known}} + Y_{t_2}^{\text{unknown}} \quad j < i \leq t_2 - t_1 && (5.12) \end{aligned}$$

Now, we may combine the inequalities (5.8) and (5.12) to obtain:

$$y_t^{\text{known}} + Y_t^{\text{unknown}} \leq \max\{y_{t_1}^{\text{known}} + Y_{t_1}^{\text{unknown}}, y_{t_2}^{\text{known}} + Y_{t_2}^{\text{unknown}}\} \quad t_1 \leq t \leq t_2 \tag{5.13}$$

It is also of interest to find a corresponding inequality for the values of the random walk about 0. To do this, let t_{min} and t_{max} be the smallest and largest elements of m_k. Firstly, consider the case where

$$y_{2k-1}^{\text{known}} = -1, y_1^{\text{known}} = -1 \tag{5.14}$$

In this case, if we view the walk as being modular, identifying 0 with 2k, we may construct additional inequalities as follows: We note that we have the same situation between t_{max} and $2k$, and between 0 and t_{min} that we had between two consecutive elements of m_k. We may thus form the inequalities

$$y_t^{\text{known}} + Y_t^{\text{unknown}} \leq \max\{y_{t_{\text{max}}}^{\text{known}} + Y_{t_{\text{max}}}^{\text{unknown}}, 0\} \quad t_{\text{max}} \leq t \leq 2k \tag{5.15}$$

and

$$y_t^{\text{known}} + Y_t^{\text{unknown}} \leq \max\{0, y_{t_{\text{min}}}^{\text{known}} + Y_{t_{\text{min}}}^{\text{unknown}}\} \quad 0 \leq t \leq t_{\text{min}} \tag{5.16}$$

In the case when (5.14) does not hold, we instead have the same situation between t_{max} and t_{min} as between two consecutive elements of m_k, leading to the inequality

$$y_t^{\text{known}} + Y_t^{\text{unknown}} \leq \max\{y_{t_{\text{max}}}^{\text{known}} + Y_{t_{\text{max}}}^{\text{unknown}}, y_{t_{\text{min}}}^{\text{known}} + Y_{t_{\text{min}}}^{\text{unknown}}\} \quad t \in \{0, 1, \ldots, t_{\text{min}}\} \cup \{t_{\text{max}}, t_{\text{max}}+1, \ldots, 2k\} \tag{5.17}$$

Now, we may finish the proof. Let t_m be a point at which the sum $\{S_i\}_{i=0}^{2k}$ attains its maximum. Now, if m_k contains t_m, or $t_m = 2k$, we are done. If that is not the case, then there are two possibilities: Either t_m falls between two consecutive elements t_1 and t_2 of m_k, or t_m falls between t_{max} and t_{min}[†]. In the former case, by the inequality (5.13), the sum $\{S_i\}_{i=0}^{2k}$ must also attain its maximum value at either t_1 or t_2.

In the latter case, either both inequalities (5.15) and (5.16) hold, in which the sum of the walks attains its maximum at $2k$, t_{min} or t_{max}, or inequality (5.17) holds in which case the sum of the walks attains its maximum at t_{min} or t_{max}. In any case, there is some element of $m_k \cup \{2k\}$ at which the sum of the walks attains its maximum.

Since this holds for each realization m_k of M_k, the random set $M_k \cup \{2k\}$ must contain the argument maximum of $\{S_i\}_{i=0}^{2k}$ with probability 1, Q.E.D.

□

[†]" between" in the"modular arithmetical" sense that t_m is larger than t_{max} or smaller than t_{min}

6 Measures of the Effectiveness of a Strategy

As previously stated, the goal of our analysis is to find a strategy which, based on $\{Y_i^{\text{known}}\}_{i=0}^{2k}$ produces a “small” subset of indices containing the argument maximum of the sum $\{S_i\}_{i=0}^{2k}$ with a fixed probability, say 1/2.

Now, if a given probability of success is desired one has to generate a subset of ones set of potential argument maxima, the size of which corresponds to the desired probability of success. If we assume that the elements of this subset are chosen at random, we can find a connection between the subset size corresponding to a given probability of success and the number of argument maxima in the set of potential peaks. Say the set of potential argument maxima is M, and that we know that the number of argument maxima in M is n_{max}. If we choose n elements at random from M, the number of argument maxima in a randomly sampled subset of M follows a hypergeometric distribution, $\text{Hypergeometric}(|M|, n_{\text{max}}, n)$. In particular, the probability that a randomly sampled subset of M with n elements contains at least one argument maximum, given n_{max}, is

$$p_n = 1 - \frac{\binom{m - n_{\text{max}}}{n}}{\binom{m}{n}} \tag{6.1}$$

Since this is a probability conditioned on the number n_{max} of argument maxima in M, calculating the probability of success for a given subset size n requires information about the distribution of n_{max}. This problem shall not be further explored in the text. However, the formula (6.1) provides a relatively numerically efficient way to estimate the probability of success from a sample of simulated walks, for each value of n.

6.1 Numerical Analysis of the Effectiveness of a Strategy

With section 6 in mind, one can now easily numerically evaluate a strategy. Firstly, one may easily simulate samples of the random walks $\{Y_i^{\text{known}}\}_{i=0}^{2k}$ and $\{Y_i^{\text{unknown}}\}_{i=0}^{2k}$ by repeatedly shuffling a list containing k copies of 1 and k copies of -1. One may then for a given strategy of selecting a subset estimate the probability of containing the argument maximum of the sum, for a given number of elements chosen from the subset, by using formula 6.1. That is, one estimates the expected value of p_n.

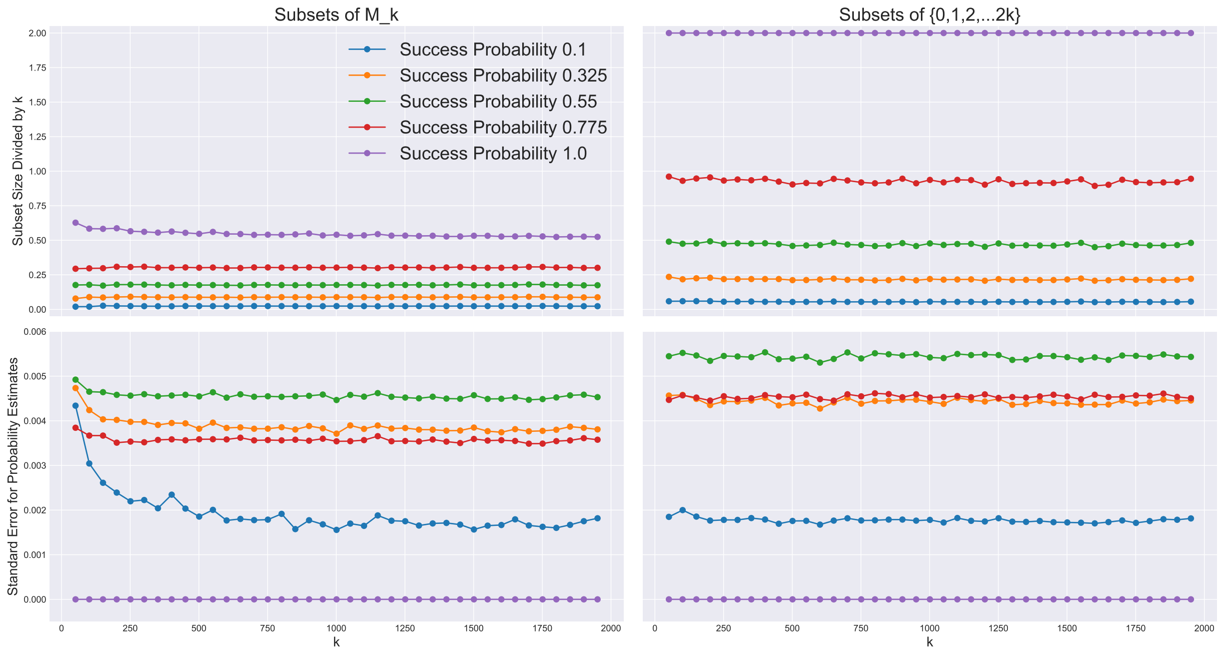

Figure 3: The number of elements of a subset needed to attain a given probability of success. Each estimate is generated from a sample of 2000 random walks.

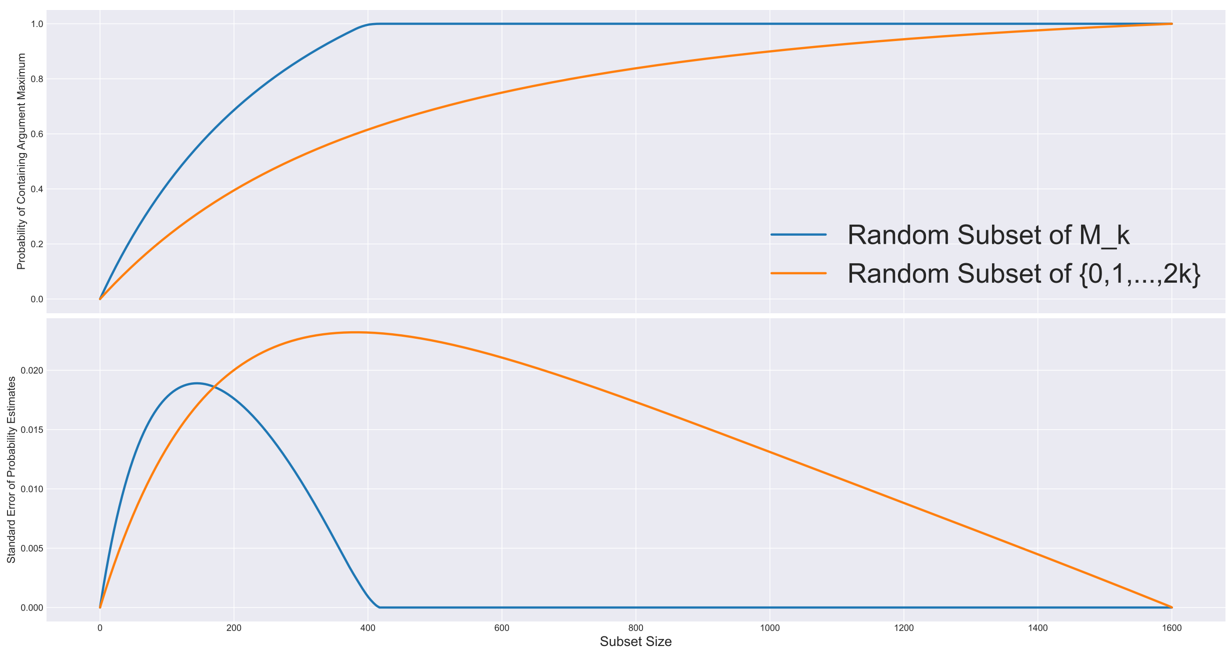

Figure 4: The probability of finding the argument maximum when picking elements from M_k or from the total set of indices in the case $k = 800$. A sample of 100 random walks was used to generate the estimates of success probabilities.

7 Discussion

The theoretical achievements of this paper lie in providing an explicit probability function for the size of the subset of indices of peaks as well as showing the (almost surely) limit of this size. It has also been shown by a simple argument that the set of indices of peaks of the known walk together with the endpoints $\{0, 2k\}$ contains the argument maximum of the sum of the known and unknown walks surely and almost surely. Furthermore, a measure of the effectiveness of a strategy in terms of how the subsets scale in k has been discussed and a numerical method devised to empirically estimate this measure.

For a given strategy, there must, for each success probability, say $1/2$, be some law giving the size n of an appropriate subset for that success probability. There is a question of what form this law might take. Figure (3) indicates that for the strategies of picking random subsets of size n from either the set of peaks or the total set of indices, the value n corresponding to each success probability scales linearly in k. Furthermore, an analytical argument shows that for a success probability of 1/2, for an optimal strategy, n can not scale slower than $\sqrt{2k}$. A reasonable guess, therefore, is that a large class of strategies for picking a subset of size $n_{1/2}$ corresponding to a success probability of $1/2$, exhibit relationships of the following form:

$$n_{1/2} = a_0 \cdot k^{a_1} \tag{7.1}$$

where a_0 and a_1 are some constants. As per the above discussion, a value of a_1 close to 1 would characterize an inefficient strategy and a value close to $1/2$ would signify a near maximally efficient strategy. Determining the optimal value of a_1 is an open problem.

A possible extension of the strategy of choosing indices of peaks would be to consider a larger neighborhood around a given index, and see for example if the increments of the walk are consistently equal to 1 for smaller indices and -1 for larger indices in the neighborhood. This may either result in a smaller subset also scaling linearly in k or potentially a more powerful strategy.

8 Appendix

8.1 Code

Python code for generating a random walk bridge with a specified length and number of peaks, from a random choice of A^+ and A^-

```python
def generate_rw_from_set_pair(k, j):
    """Generates a random walk bridge of length 2k with a specified number j of peaks.
    Params:
    -------
    k: int
        Half of the length of the walk
    j: int
        The desired number of peaks
    Returns:
    --------
    (np.array(size=j), np.array(size=j), np.array(size=2k)):
        A tuple consisting of (A+, A-, walk)"""


    # Random choice of (A+, A-)
    A_plus = np.random.choice(np.array(list(range(1, k + 1))), size=j, replace=False)
    A_minus = np.random.choice(np.array(list(range(1, k + 1))), size=j, replace=False)
```

```python
n_plus = np.zeros(2 * k, int)
n_minus = np.zeros(2 * k, int)
x = np.zeros(2 * k, int)
inverse_peak_indices = []
peak_indices = []

# Starting rule
if 1 in A_minus:
    x[0] = 1
    n_plus[0] = 1
else:
    x[0] = -1
    n_minus[0] = 1

# Iteration rule
for i in range(1, 2 * k):
    if x[i - 1] == 1 and n_plus[i - 1] in A_plus:
        x[i] = -1
        n_minus[i] = n_minus[i - 1] + 1
        n_plus[i] = n_plus[i - 1]
        peak_indices.append(i)
    elif x[i - 1] == -1 and (n_minus[i - 1] + 1 in A_minus or n_minus[i - 1] == k):
        x[i] = 1
        n_plus[i] = n_plus[i - 1] + 1
        n_minus[i] = n_minus[i - 1]
        inverse_peak_indices.append(i)
    else:
        if x[i - 1] == -1:
            x[i] = -1
            n_minus[i] = n_minus[i - 1] + 1
            n_plus[i] = n_plus[i - 1]
        else:
            x[i] = 1
            n_plus[i] = n_plus[i - 1] + 1
            n_minus[i] = n_minus[i - 1]

walk = np.zeros(2 * k + 1, int)
# The walk is the cumulative sum of increments
for i in range(2 * k):
    walk[i + 1] = walk[i] + x[i]

return A_plus, A_minus, walk
```

References

[1] Norman L. Biggs. *Discrete Mathematics*. Oxford University Press, 2002.

[2] Allan Gut. *An Intermediate Course in Probability*. Springer, 2009.

[3] Peter Mörters and Yuval Peres. *Brownian Motion*. Cambridge University Press, 2012.

[4] Satish Shirali. *A Concise Introduction to Measure Theory*. Springer, 2010.