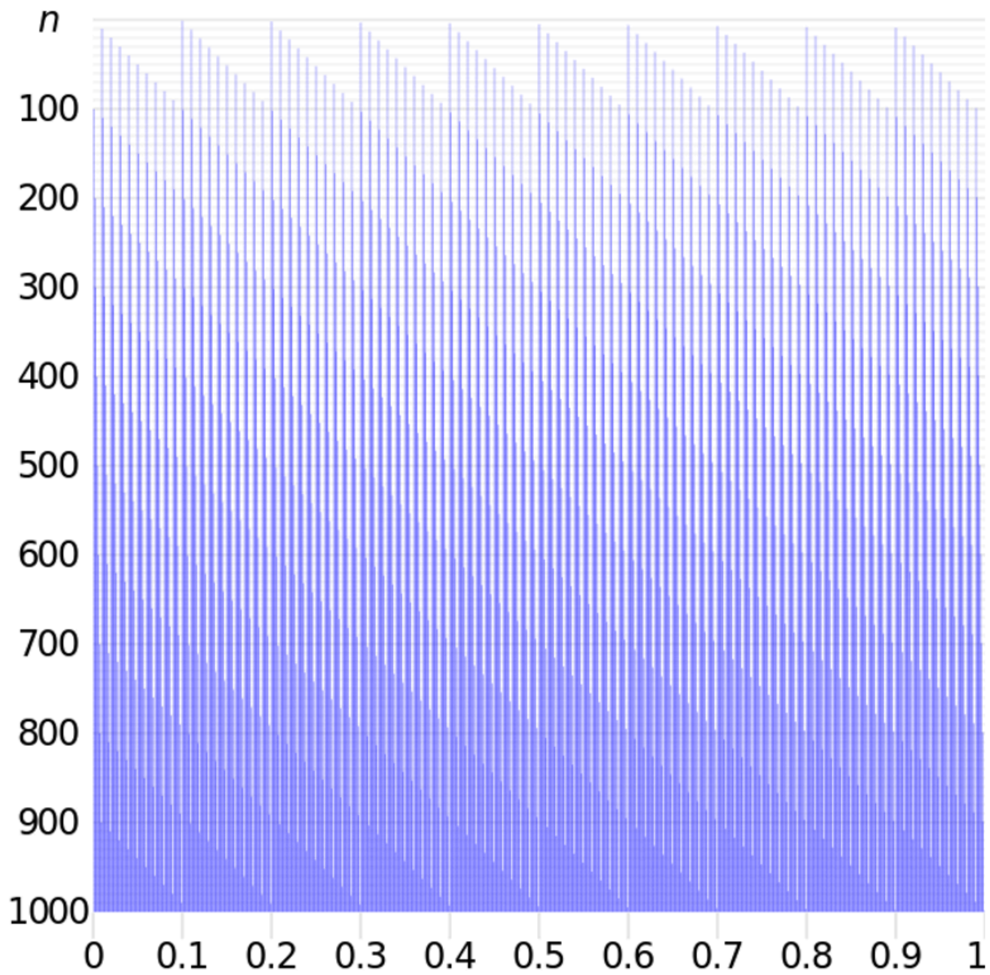
Equidistributed Sequences

Optimal Stick Breaking

AM3000: Bachelorproject

Stephan Wellner



TU Delft

Equidistributed Sequences

Optimal Stick Breaking

by

Stephan Wellner

To obtain the degree of Bachelor of Science
at the Delft University of Technology,
to be defended publicly on Thursday August 1, 2024 at 10:00 AM.

Student number:	5218152	
Project duration:	April 22, 2024 – August 1, 2024	
Thesis committee:	Dr. R.J. (Robbert) Fokkink,	TU Delft, supervisor
	Dr. A. (Anurag) Bishnoi,	TU Delft
	Dr. ir. J.T. (Theresia) van Essen,	Coordinator

This thesis is confidential and cannot be made public until August 1, 2024.

Cover:	Illustration of the filling of the unit interval (horizontal axis) using the first n terms of the decimal Van der Corput sequence, for n from 0 to 999 (vertical axis)
Style:	TU Delft Report Style

TU Delft

Preface

In mathematics, our quest to understand often takes us on complex journeys where simple ideas reveal complexities. One fascinating topic is the study of equidistributed sequences. These sequences help us understand how numbers are spread out evenly, and they have important applications in various fields.

Equidistributed sequences are sequences where the values are evenly spread out over an interval. What makes these sequences fascinating is not just their definition but the many ways they appear in different mathematical situations. They help us understand how numbers can be distributed in a way that seems random but is actually uniform over large scales.

In this report, I look at both well-known discoveries and try to uncover new insights. My approach combines theoretical study and practical (numerical) analysis. I also focus on the work of Nicolaas Govert de Bruijn and Paul Erdös in their article "Sequences of Points on a Circle."

As I delve deeper into this subject, I must recognize the incredible guidance I've received. I am truly grateful to my supervisor, Dr. R.J. (Robbert) Fokkink, whose knowledge and inspiration have been crucial in shaping this work. His thoughtful feedback and constant support have been invaluable throughout this journey.

I want to give a big thank you to Dr.ir. J.T. (Theresia) van Essen, the Bachelor End Project Coordinator. Her dedication and availability have been incredibly helpful while I was writing this report. Her support made finishing this project much easier, and I truly appreciate her guidance.

With this project, I am finishing my bachelor's degree at TU Delft, marking both an end and a new start in my academic journey. I hope this report shows the completion of my undergraduate studies and the start of a lifelong journey in understanding mathematics.

Thank you.

Stephan Wellner
Delft, July 2024

Contents

1 Introduction

In 1935, J.G. van der Corput asked a question about how to evenly distribute sequences of points [13]. By 1945, Van Aardenne-Ehrenfest found an answer, which encouraged more mathematical research. The problem they explored was how to place points evenly on a circle and keep this even distribution when adding new points. In other words, van der Corput wondered if points could be placed on a line segment in such a way that any two subintervals of the same length would have almost the same number of points. Van Aardenne showed that perfect equal distribution is impossible, leading to more studies on the irregularities in point distribution in different areas of mathematics.

A lot of mathematical work has been done on studying sequences of points in the interval $[0, 1)$ *that are evenly distributed. This report looks at the idea that a sequence is evenly distributed if its first* n *points divide a circle into intervals that are roughly equal in length. The sequence* $X_k = log_2(2k-1)$ *mod* 1 *(explained in section 2.2) was introduced by De Bruijn and Erdös in this context. We will show that the way the gaps between points in this sequence are structured is uniquely optimal in a certain way.*

This thesis aims to uncover new insights through a mix of theoretical and numerical studies. The first sections focus on the work of Nicolaas Govert de Bruijn and Paul Erdös, especially their 1949 paper "Sequences of Points on a Circle" [5]. Later sections will present numerical findings and compare different methods for achieving the best way to split a circle evenly.

In short, this thesis aims to reproduce mathematical theories about evenly distributed sequences and the best ways to divide intervals. By merging these theories with new numerical techniques, it hopes to provide a new viewpoint on a classic topic, inspiring more investigation and study in this fascinating area of mathematics.

2

Sequences of Points on a Circle

In this chapter, we explore De Bruijn and Erdös's 1949 study on the lengths of subintervals created when a circle is divided [5]. We reproduce and expand their findings. This involves focusing on the sequence they introduced for optimal distribution, and analysing the distribution of consecutive intervals resulting from the division of the circle. After that, we will answer a question posed by by A. Ostrowski [10], [9] without knowing that De Bruijn and Erdös already had answered that question, and we give G. H. Toulmin's proof (found independently) to answer the question [14].

2.1. Introduction

Consider a circle with radius $\frac{1}{2\pi}$. The circumference of the circle is $2\pi \cdot \frac{1}{2\pi} = 1$. We call this the unit circle. The unit circle is the set of all complex numbers with absolute value equal to 1. It can be parametrized by angles θ (in radians), where each angle θ corresponds to the complex number cos(θ) + i sin(θ) on the unit circle.

The set of numbers modulo 1 consists of real numbers in the range $[0, 1)$, where any integer part is removed. This set represents the fractional parts of real numbers.

The unit circle and the set of numbers modulo 1 are closely related concepts in mathematics.

There is an isomorphism between the complex numbers with absolute value 1 and the real numbers modulo 1. This isomorphism is given by the exponential map $\theta \to e^{i\theta}$, which maps real numbers θ to points $e^{i\theta}$ on the unit circle (recall Euler's formula: $e^{i\theta} = cos(\theta) + isin(\theta)$).

Consider sequences $\{a\}$ of points $a_1, a_2, a_3, ...$ on the unit circle. The points $a_1, ..., a_n$ break the circle into n sticks (they are intervals but we will call them sticks to help the reader visualise the process) with total length equal to 1.

Let $M_n^1(a)$ and $m_n^1(a)$ be the length of the largest and smallest stick, respectively. The subscript n indicates that we have broken the unit circle into n sticks, while the superscript 1 denotes that we look at the maximum and minimum lengths of 1 such stick.

Proposition 1.

$$nM_n^1(a) \geq 1 \geq nm_n^1(a). \tag{2.1}$$

Proof. The average length of one stick is $\frac{1}{n}$. Therefore the length of the largest stick is greater than or equal to $\frac{1}{n}$ (equal if the length of all the sticks are equal to $\frac{1}{n}$). While the length of the smallest stick is lower than or equal to $\frac{1}{n}$ (again, equal if the length of all the sticks are equal to $\frac{1}{n}$). So,

$$M_n^1(a) \geq \frac{1}{n} \geq m_n^1(a)$$

and the claim follows by multiplying by n. □

In the same way, $M_n^r(a)$ and $m_n^r(a)$ denote the maximum and minimum sum of lengths of r consecutive sticks after dividing the unit circle into n sticks.

Proposition 2.

$$nM_n^r(a) \geq r \geq nm_n^r(a). \tag{2.2}$$

Proof. The average length of r consecutive sticks is $\frac{r}{n}$. Hence we get,

$$M_n^r(a) \geq \frac{r}{n} \geq m_n^r(a)$$

and the claim follows by multiplying by n. □

Corollary 1. *The bounds in* (2.2) *after dividing by r are tighter than the ones in* (2.1)*.*

Proof. Picking the longest segment of r consecutive sticks and dividing the total length by r will always give us a number that's smaller than the length of the longest individual stick. However, if the longest segment of r consecutive sticks is made up of r sticks that are all the largest length of an individual stick. Then the result of dividing their total length by r will be exactly the same as the length of each of those sticks. That is,

$$\frac{M_n^r(a)}{r} \leq M_n^1(a).$$

It works analogously for the smallest stick length. □

We set

$$\limsup_{n\to\infty} nM_n^r(a) = \Lambda_r(a)$$

$$\liminf_{n\to\infty} nm_n^r(a) = \lambda_r(a)$$

$$\limsup_{n\to\infty} \frac{M_n^r(a)}{m_n^r(a)} = \mu_r(a). \tag{2.3}$$

We aim to find the greatest lower bound (g.l.b.) for $\Lambda_1(a)$, meaning $\inf \Lambda_1(a)$ over all sequences (a), and the largest upper bound (l.u.b.) for $\lambda_1(a)$, meaning $\sup \lambda_1(a)$ over all sequences (a). From this we will be able to derive the greatest lower bound for $\mu_1(a)$ easily.

We start with the greatest lower bound for $\Lambda_1(a)$. Let $\Lambda_1 =$ g.l.b. $\Lambda_1(a)$. Let $I_n^1 \geq I_n^2 \geq ... \geq I_n^n$ be the stick lengths in decreasing order after breaking n times.

Proposition 3. *For any $1 < k \leq n$ we have $I_n^k \leq I_{n+1}^{k-1}$.*

Proof. Let's say we break the j-th stick of the sequence. Two cases have to be considered here:

1. $j > k$.

   By breaking the j-th stick of the sequence, both resulting parts will be shorter than I_n^j (the length of the original stick). Since the two parts of the broken j-th longest stick are shorter than I_n^j, they will be placed after the $(j-1)$-th position in the new sequence of stick lengths. The new sequence of stick lengths is the sequence we obtain after breaking $n+1$ times. This leaves the first k sticks unperturbed. Hence, they are the same in the original, and new sequences of stick lengths. This means that $I_{n+1}^k = I_n^k$. We also know, by how the sequence is defined, that $I_{n+1}^{k-1} \geq I_{n+1}^k$. Putting these two together we get that $I_{n+1}^{k-1} \geq I_n^k$.
2. For $j \leq k$, we consider 3 subcases:

   (a) Both resulting parts are smaller than I_n^k.

   Then they will be placed behind in the new sequence. So I_n^k becomes I_{n+1}^{k-1}. Hence $I_n^k = I_{n+1}^{k-1}$ and the claim is true.

(b) 1 resulting part is larger than I_n^k and the other part is smaller.

Then the k-th longest stick stays the k-th longest stick in the new sequence. Hence, $I_n^k = I_{n+1}^k$ and, by the definition of the sequence, $I_{n+1}^{k-1} \geq I_{n+1}^k$. Putting these together we obtain $I_{n+1}^{k-1} \geq I_n^k$.

(c) Both resulting parts are larger than I_n^k.

Then both parts will be in front of the k-th largest stick in the new sequence. Hence I_n^k becomes I_{n+1}^{k+1} as there is one more stick in front of the original k-th longest stick now. So $I_n^k = I_{n+1}^{k+1}$ and, by the definition of the sequence, $I_{n+1}^{k-1} \geq I_{n+1}^{k+1}$. Putting these together we obtain $I_{n+1}^{k-1} \geq I_n^k$.

□

By induction, for any $1 \leq k \leq n$ we have

$$I_n^k \leq I_{n+k-1}^1 = M_{n+k-1}^1.$$

Theorem 1. *Let* $\Lambda_1 = \inf\{\Lambda_1(a) : $ *sequences* $a\}$*, then* $\Lambda_1 \geq \frac{1}{\ln 2}$.

Proof. Let $\Lambda_{1,n} =$ Max $\{kM_k^1 : n \leq k < 2n\}$; then

$$1 = \sum_{k=1}^{n} I_n^k \leq \sum_{k=1}^{n} M_{n+k-1}^1 \leq \Lambda_{1,n} \sum_{k=1}^{n} \frac{1}{n+k-1}.$$

To find $\sum_{k=1}^{n} \frac{1}{n+k-1}$ we use a known result [3]:

$$\sum_{k=1}^{n} \frac{1}{k} \approx \ln n + \gamma \text{ where } \gamma \approx 0,577 \text{ is the Euler–Mascheroni constant.}$$

We rewrite $\sum_{k=1}^{n} \frac{1}{n+k-1}$ as

$$\sum_{k=1}^{2n-1} \frac{1}{k} - \sum_{k=1}^{n-1} \frac{1}{k}$$

which tends to

$$\begin{aligned} &\ln(2n-1) + \gamma - (\ln(n-1) + \gamma) \\ &= \ln(2n-1) - \ln(n-1) \\ &= \ln \frac{2n-1}{n-1} \to \ln 2 \text{ as } n \to \infty. \end{aligned}$$

So as $n \to \infty$ we have

$$\sum_{k=1}^{n} \frac{1}{n+k-1} \to \ln 2$$

giving

$$\limsup_{n\to\infty} \Lambda_{1,n} \geq \frac{1}{\ln 2}$$

which implies that

$$\limsup_{n\to\infty} nM_n^1(a) \geq \frac{1}{\ln 2}$$

as proved by [8]. Let $\Lambda_1 =$ g.l.b. $\Lambda_1(a)$, then we have found

$$\Lambda_1 \geq \frac{1}{\ln 2}.$$

□

For the largest upper bound for $\lambda_r(a)$, A. Ostrowski [10], [9] has asked: what is the least number $\lambda_1(a)$ for which we can say that

$$\liminf_{n\to\infty} nm_n^1(a) \leq \lambda_1(a)$$

for every sequence? He asked this without knowing that De Bruijn and Erdös already had solved this [5]. He proved in [9], Theorem 1 that

$$\lambda_1(a) \leq \frac{1}{\ln 4}.$$

We provide the independent proof by G. H. Toulmin [14].

Lemma 1. *Let $\lambda_1 = \sup\{\lambda_1(a) : \text{sequences } a\}$, then given any $\epsilon > 0$, there is an n_0 such that*

$$nm_n^1(a) > \lambda_1(a) - \epsilon \quad \forall\, n \geq n_0. \tag{2.4}$$

Proof. The proof follows directly from the definition of the limit inferior. □

Definition 1. *The **n-th subdivision** is the division of $[0,1]$ into n sticks by the points $a_2, a_3, ..., a_n$ ($a_1 = 0$) arranged in order of magnitude; $m_n^1(a)$ denotes the length of the shortest stick of the n-th subdivision. We have labeled $I_n^1, I_n^2, ..., I_n^n$ as the stick lengths in decreasing order after breaking n times in Proposition 3.*

Let N be any integer $\geq n_0$ (to be chosen subsequently).

Lemma 2. *For any integer r, $0 \leq r \leq N$, we can choose $2r$ distinct sticks of the $(N + r)$-th subdivison which have, respectively, lengths greater than the numbers*

$$\frac{\lambda_1(a)-\epsilon}{N+1}, \frac{\lambda_1(a)-\epsilon}{N+1}, \frac{\lambda_1(a)-\epsilon}{N+2}, \frac{\lambda_1(a)-\epsilon}{N+2}, \frac{\lambda_1(a)-\epsilon}{N+3}, ..., \frac{\lambda_1(a)-\epsilon}{N+r}, \frac{\lambda_1(a)-\epsilon}{N+r}.$$

Proof. Apply induction. We prove, in fact, that we can choose the $2r$ sticks in such a way that any stick of the $(N+r)$-th subdivision which is not chosen must have been already a stick of the N-th subdivision.

For $r = 0$ there are no sticks to choose, so the statement is trivially true.

Assume the lemma is true for $r-1$. That is, for the $(N+r1)$-th subdivision, we can choose $2(r1)$ distinct sticks, satisfying the length conditions. We need to prove that the Lemma holds for r.

If a_{N+r} does not lie in any of the sticks chosen at the $(r-1)$-th stage, simply add the two new sticks of which a_{N+r} is an endpoint. Each of these new sticks has length greater than $\frac{\lambda_1(a)-\epsilon}{N+r}$, satisfying the length condition.

If a_{N+r} does lie in a stick previously chosen, still add the two new sticks of which it is an endpoint, and replace the stick destroyed (by a_{N+r}) by any stick not yet included in the set. We need $2r$ sticks, and we have already chosen $2(r1)$ sticks. Therefore, the number of new sticks we need is $2r2(r1) = 2$. Since the total number of new sticks at stage $N + r$ is $N + r$ and we've chosen $2(r1)$ previously, there are:

$$(N + r) - 2(r - 1) = N - r + 2 \geq 2$$

sticks available to replace the destroyed one. The replacement stick must have been a stick in the N-th subdivision. Hence, it has a length greater than $\frac{\lambda_1(a)-\epsilon}{N}$, satisfying the length condition.

To ensure the last r sticks are in the collection, consider the structure of the subdivision: each subdivision point a_i generates two new sticks. By including the two new sticks at each step and ensuring any replacement stick is chosen from those not yet included, the last r sticks will naturally be part of the final collection because we are always including the latest sticks generated at each step. Therefore, by induction, we ensure that for any $0 \leq r \leq N$, we can choose $2r$ distinct sticks from the $(N + r)$-th subdivision, satisfying the length conditions. □

Theorem 2. *Let $\lambda_1 = \sup\{\lambda_1(a) : \text{sequences } a\}$, then $\lambda_1 \leq \frac{1}{\ln 4}$.*

Proof. Since $\frac{1}{N+r} > \frac{1}{N+x}$ for $r < x \leq r+1$,

$$\sum_{r=1}^{N} \frac{1}{N+r} > \int_{1}^{N+1} \frac{dx}{N+x} = \ln(N+x)|_1^{N+1} = \ln \frac{2N+1}{N+1} = \ln(2 - \frac{1}{N+1}), \tag{2.5}$$

The expression $\ln(2 - \frac{1}{N+1})$ gets closer to $\ln 2$ as N increases because $\frac{1}{N+1}$ becomes smaller. Specifically,

$$\lim_{N\to\infty} \ln(2 - \frac{1}{N+1}) = \ln 2 > \frac{\ln 2}{1 + 2\epsilon \ln 2}. \tag{2.6}$$

The last inequality is true since the denominator $1 + 2\epsilon \ln 2$ is greater than 1. From (2.5) and (2.6) there must exist some sufficiently large $N \geq n_0$ such that

$$\sum_{r=1}^{N} \frac{1}{N+r} > \frac{\ln 2}{1 + 2\epsilon \ln 2}. \tag{2.7}$$

This holds as long as $\epsilon > \frac{1}{2 \ln 2(N+1) - 1}$, the explanation for ϵ can be found in Appendix C.

Now apply Lemma 2 with $r = N$, since the $2N$ sticks chosen must in this case be all the sticks of the division of the $2N$-th subdivision, their total length is 1; hence Lemma 2 and (2.7) give

$$\begin{aligned} 1 &> \sum_{r=1}^{N} \frac{2(\lambda_1(a) - \epsilon)}{N+r} \\ \Leftrightarrow \frac{1}{2(\lambda_1(a) - \epsilon)} &> \sum_{r=1}^{N} \frac{1}{N+r} \\ &> \frac{\ln 2}{1 + 2\epsilon \ln 2} \\ \Leftrightarrow \frac{2(\lambda_1(a) - \epsilon) \ln 2}{1 + 2\epsilon \ln 2} &< 1 \\ \Leftrightarrow 2(\lambda_1(a) - \epsilon) \ln 2 &< 1 + 2\epsilon \ln 2 \\ \Leftrightarrow \lambda_1(a) - \epsilon &< \frac{1}{2 \ln 2} + \epsilon, \end{aligned}$$

i.e.

$$\lambda_1(a) - \epsilon < \frac{1}{\ln 4} + \epsilon;$$

but ϵ can be arbitrarily small, hence

$$\lambda_1(a) \leq \frac{1}{\ln 4}$$

Let $\lambda_1 = l.u.b.\ \lambda_1(a)$, then we have found

$$\lambda_1 \leq \frac{1}{\ln 4}.$$

□

We can easily determine $\mu_1 = g.l.b.\ \mu_1(a)$,

$$\mu_1 = \frac{\frac{1}{\ln 2}}{\frac{1}{\ln 4}} = \frac{\ln 4}{\ln 2} = \frac{2 \ln 2}{\ln 2} = 2.$$

2.2. A Sequence with Optimal Values

Let's have a look at the sequence $a_n = \log_2(2n-1)$ reduced mod 1. The sequence $a_1, ..., a_n$ has n terms:

$$\log_2(1) \text{ mod } 1, \log_2(3) \text{ mod } 1, \log_2(5) \text{ mod } 1, ..., \log_2(2n-1) \text{ mod } 1. \tag{2.8}$$

Lemma 3. *No two of the a_n's are congruent modulo 1.*

Proof. We start with the expression for the logarithms

$$a_n = \log_2(2n-1)$$

$\log_2(2n-1)$ can be expressed as

$$\log_2(2n-1) = k_n + p_n$$

where k_n is an integer and $p_n \in [0,1)$ is the fractional part. Hence, $a_n = p_n$.

We need to show that no two a_n's have the same fractional part. Assume for contradiction that there exist two different terms a_i and a_j such that

$$\log_2(2i-1) \mod 1 = \log_2(2j-1) \mod 1.$$

This implies

$$\log_2(2i-1) = k_i + p_i \quad \text{and} \quad \log_2(2j-1) = k_j + p_j$$

where k_i, k_j are integers and $p_i, p_j \in [0,1)$ are fractional parts such that

$$p_i = p_j.$$

We can rewrite the above logarithmic equations in exponential form

$$2^{k_i+p_i} = 2i-1 \quad \text{and} \quad 2^{k_j+p_j} = 2j-1.$$

Given that $p_i = p_j$, we have

$$2^{k_i+p_i} = 2^{k_j+p_j}.$$

This can be rewritten as

$$2^{k_i} \cdot 2^{p_i} = 2^{k_j} \cdot 2^{p_j}$$

Since $p_i = p_j$:

$$2^{k_i} \cdot 2^{p_i} = 2^{k_j} \cdot 2^{p_i}$$

This simplifies to

$$2^{k_i} = 2^{k_j}.$$

The above equality $2^{k_i} = 2^{k_j}$ implies $k_i = k_j$, meaning k_i and k_j are the same integer. This in turn implies that $2i-1 = 2j-1$ (as the fractional parts p_i and p_j are also equal), or $i = j$, which contradicts our assumption that $i \neq j$.

We conclude that no two terms a_n in the sequence $\log_2(2n-1) \mod 1$ are congruent modulo 1. □

Lemma 4. *No two of the terms in*

$$\log_2 n, \log_2(n+1), ..., \log_2(2n-1) \tag{2.9}$$

are congruent modulo 1.

Proof. For all $x \in \{n, n+1, ..., 2n-1\}$ we can divide x by 2 until we end up in the interval $(1,2)$. Say you have to divide k times. Then

$$\frac{x}{2^k} \in (1,2).$$

For all $x \in \{n, n+1, ..., 2n-1\}$ this k is the same, as $2n-1 < 2n$. Only from $2n$ onwards you can divide one more time by 2. Now we look at

$$\frac{n}{2^k}, \frac{n+1}{2^k}, ..., \frac{2n-1}{2^k}.$$

Which are pairwise not congruent modulo 1 as they are all in the interval $(1,2)$ and different:

$$n \neq n+1 \neq ... \neq 2n-1$$

hence

$$\frac{n}{2^k} \neq \frac{n+1}{2^k} \neq ... \neq \frac{2n-1}{2^k}.$$

We can conclude that no pair of terms in (2.9) are congruent modulo 1. □

Theorem 3. *The a_n's in* (2.8) *occur in the order of the terms in* (2.9)*. That is, there exists a bijection between*

$$\log_2(1) \textit{ mod } 1, \log_2(3) \textit{ mod } 1, \log_2(5) \textit{ mod } 1, ..., \log_2(2n-1) \textit{ mod } 1$$

and

$$\log_2 n, \log_2(n+1), ..., \log_2(2n-1).$$

Proof. Sequence (2.8) contains $\frac{2n-1-1}{2}+1 = n$ terms, and (2.9) contains $2n-1-n+1 = n$ terms as well. Therefore, if we find an injection between the two sets of the same cardinality, it will be a bijection [2]. Moreover, in both sequences, no two numbers are congruent modulo 1. We find a bijection between these two sequences: start at the end with the term $\log_2(2n-1)$, this term is (simply) mapped to $\log_2(2n-1) \mod 1$. $2n-2$ is now even, hence divisable by 2, which gives $n-1$. We map $\log_2(2n-2)$ to $\log_2(n-1) \mod 1$. If we divide by 2 and the resulting term is still even, we continue dividing by 2 until we obtain an odd number. A pattern emerges as we observe this mapping process.

We establish a clear mapping between the elements of the two sequences. For $x \in \{n, n+1, ..., 2n-1\}$ the map is given by

$$f(\log_2(x)) = \log_2(\frac{x}{2^m}) \mod 1, \text{ with } m \text{ being the largest power of } 2 \text{ that divides } x.$$

To prove that $f(x)$ is injective, we need to show that if $x \neq y$, then $f(x) \neq f(y)$.

Let $k \neq k' \in$ the terms of (2.8). Assume for the sake of contradiction that $k \neq k'$ and $f(k) = f(k')$. Then

$$\begin{aligned} \log_2(\frac{k}{2^m}) \mod 1 &= \log_2(\frac{k'}{2^{m'}}) \mod 1 && \text{Assume } f(k) = f(k') \\ \Leftrightarrow \log_2(\frac{k}{2^m}) &= \log_2(\frac{k'}{2^{m'}}) && \text{Lemma 3} \\ \Leftrightarrow \frac{k}{2^m} &= \frac{k'}{2^{m'}} && \text{WLOG } k \not> 2k' \\ \Leftrightarrow 2^{m'} \cdot k &= 2^m \cdot k' \\ \Leftrightarrow k &= 2^{(m-m')} \cdot k' \end{aligned}$$

Here is our contradiction as if $m > m'$ then $k = 2^{(m-m')} \cdot k'$ implies that k' is at least $2k$. On the other hand, if $m = m'$ then $k = k'$ which contradicts our assumption. We conclude that $f(x)$ is injective, hence bijective. □

The lengths of the sticks defined by the a_n's are

$$\log_2(n+1) - \log_2 n, \log_2(n+2) - \log_2(n+1), ..., \log_2(2n-1) - \log_2(2n-2), \log_2(2n) - \log_2(2n-1)$$

$$= \log_2 \frac{n+1}{n}, \log_2 \frac{n+2}{n+1}, ..., \log_2 \frac{2n-1}{2n-2}, \log_2 \frac{2n}{2n-1}$$

Theorem 4. *For the sequence $a_n = \log_2(2n-1)$ mod 1, $\Lambda_1(a) = \frac{1}{\ln 2}$.*

Proof. The longest stick is $\log_2(\frac{n+1}{n})$. Hence,

$$\begin{aligned} nM_n^1(a) &= n\log_2(\frac{n+1}{n}) \\ &= n\log_2(1+\frac{1}{n}) \\ &= n\frac{\ln(1+\frac{1}{n})}{\ln 2}. \end{aligned}$$

For $n \to \infty$, $nM_n^1(a)$ goes to the limit $\frac{1}{\ln 2}$. □

Theorem 5. *For the sequence $a_n = \log_2(2n-1)$ mod 1, $\lambda_1(a) = \frac{1}{\ln 4}$.*

Proof. The shortest stick is $\log_2(\frac{2n}{2n-1})$. Hence,

$$\begin{aligned} nm_n^1(a) &= n\log_2(\frac{2n}{2n-1}) \\ &= n\log_2(1+\frac{1}{2n-1}) \\ &= n\frac{\ln(1+\frac{1}{2n-1})}{\ln 2}. \end{aligned}$$

For $n \to \infty, nm_n^1(a)$ goes to the limit $\frac{1}{2\ln 2} = \frac{1}{\ln 4}$. □

Theorem 6. *For the sequence $a_n = \log_2(2n-1)$ mod 1, $\mu_1(a) = 2$.*

Proof. From Theorem 4 and Theorem 5 it follows that

$$\frac{M_n^1(a)}{m_n^1(a)}$$

goes to the limit

$$\frac{\frac{1}{\ln 2}}{\frac{1}{\ln 4}} = \frac{\ln 4}{\ln 2} = \frac{2\ln 2}{\ln 2} = 2.$$

□

We see that for this sequence we attain the bounds we have found in the previous section.

2.3. Lower Bound for $\Lambda_r(a)$

Now that we have analysed a sequence that attains the bounds if we look at the longest and shortest stick; we can try and find bounds if we look at consecutive sticks. In this section, we recreate De Bruijn and Erdös's method [5] for finding the lower bound of $\Lambda_r(a)$. We follow their general approach but sometimes use different proofs. Some of the lemmas are presented for just one stick to make them easier to understand. This is done to help prepare the reader for the more complex case of r consecutive sticks.

Lemma 5. *Let $a_1, a_2, ..., a_n$ be a sequence, $n \in \mathbb{N}$, and ρ is such that*

$$kM_k^1(a) < \rho \qquad (n \leq k < 2n). \tag{2.10}$$

Then

$$1 < \rho(\frac{1}{n} + \frac{1}{n+1} + ... + \frac{1}{2n-1}).$$

Proof. Let the sticks broken by $a_1, a_2, ..., a_n$ be $I_1, I_2, ..., I_n$ arranged in descending order and their corresponding lengths $\alpha_1, \alpha_2, ..., \alpha_n$, so that

$$\alpha_1 \geq \alpha_2 \geq ... \geq \alpha_n \tag{2.11}$$

and

$$\alpha_1 + \alpha_2 + ... + \alpha_n = 1. \tag{2.12}$$

Now we break $n-1$ more times at $a_{n+1}, a_{n+2}, ..., a_{2n-1}$. Any break splits up at most one stick I_i $(i \in \{1, ..., n\})$. Therefore, after breaking once more at a_{n+1}, the length of the largest stick is still α_1 if a_{n+1} is not in I_1. If a_{n+1} is in I_1 then the largest stick is I_2 with length α_2. By repeating this train of thought, after $a_{n+1}, ..., a_{n+p}$ with $1 \leq p \leq n-1$ have been placed, at least one stick with length $\geq \alpha_{p+1}$ is left over. Thus

$$\begin{aligned} M_n^1(a) &\geq \alpha_1 \\ M_{n+1}^1(a) &\geq \alpha_2 \\ &\vdots \\ M_{2n-1}^1(a) &\geq \alpha_n. \end{aligned} \tag{2.13}$$

Now

$$\begin{aligned} 1 &= \alpha_1 + \alpha_2 + ... + \alpha_n && (2.12) \\ &\leq M_n^1(a) + M_{n+1}^1(a) + ... + M_{2n-1}^1(a) && (2.13) \\ &< \frac{\rho}{n} + \frac{\rho}{n+1} + ... + \frac{\rho}{2n-1} && (2.10) \\ &= \rho(\frac{1}{n} + \frac{1}{n+1} + ... + \frac{1}{2n-1}) && \text{factor out } \rho \end{aligned}$$

□

Lemma 6. *For at least one k $(n \leq k < 2n)$ we have*

$$kM_k^1(a) \geq (\frac{1}{n} + \frac{1}{n+1} + ... + \frac{1}{2n-1})^{-1} = \sigma_n.$$

Proof.

$$\begin{aligned} 1 &= \alpha_1 + \alpha_2 + ... + \alpha_n && (2.12) \\ &\leq M_n^1(a) + M_{n+1}^1(a) + ... + M_{2n-1}^1(a) && (2.13) \text{ for every } \alpha_i \\ &< \frac{\rho}{n} + \frac{\rho}{n+1} + ... + \frac{\rho}{2n-1} && (2.10) \\ &= \rho(\frac{1}{n} + \frac{1}{n+1} + ... + \frac{1}{2n-1}) && \text{factor out } \rho \end{aligned} \tag{2.14}$$

so from (2.14),

$$(\frac{1}{n} + \frac{1}{n+1} + ... + \frac{1}{2n-1})^{-1} < \rho$$

It follows that the claim is true for at least one k, otherwise the bound will never be attained. □

We have $\sigma_n < \frac{1}{\ln 2}$ and $\sigma_n \to \frac{1}{\ln 2}$, so $\Lambda_1(a) \geq \frac{1}{\ln 2}$. This holds for any sequence and the lower bound is attained for the sequence in section 2.2.

We now prove the previous result (Lemma 6) for r consecutive sticks.

Lemma 7. *For at least one k $(rn \leq k < (r+1)n)$ we have*

$$kM_k^r(a) \geq (\frac{1}{rn} + \frac{1}{rn+1} + ... + \frac{1}{rn+n-1})^{-1} = \tau_n.$$

Proof. Let

$$kM_k^r(a) < \rho. \tag{2.15}$$

Let $I_1, I_2, ..., I_{rn}$ be all the r-sticks, (i.e., every I_i consists of r sticks) arranged in descending order and their corresponding lengths $\alpha_1, \alpha_2, ..., \alpha_{rn}$, so that

$$\alpha_1 \geq \alpha_2 \geq ... \geq \alpha_{rn} \tag{2.16}$$

and

$$\alpha_1 + \alpha_2 + ... + \alpha_{rn} = r. \tag{2.17}$$

Now by the same train of thought as in Lemma 5, after breaking p more times with $1 \leq p \leq n-1$, at least one r-stick with length $\geq \alpha_{rp+1}$ is left over. Thus,

$$\begin{aligned} M_{rn}^r(a) &\geq \alpha_1 \\ M_{rn+1}^r(a) &\geq \alpha_{r+1} \\ &\vdots \\ M_{rn+n-1}^r(a) &\geq \alpha_{r(n-1)+1}. \end{aligned} \tag{2.18}$$

Now

$$\begin{aligned} r &= \alpha_1 + \alpha_2 + ... + \alpha_{rn} && (2.17) \\ &\leq rM_{rn}^r(a) + rM_{rn+1}^r(a) + ... + rM_{(r+1)n-1}^r(a) && (2.18) \\ &< r\frac{\rho}{rn} + r\frac{\rho}{rn+1} + ... + r\frac{\rho}{(r+1)n-1} && (2.15) \\ &= r\rho(\frac{1}{rn} + \frac{1}{rn+1} + ... + \frac{1}{(r+1)n-1}) && \text{factor out } r\rho \end{aligned}$$

Hence,

$$(\frac{1}{rn} + \frac{1}{rn+1} + ... + \frac{1}{(r+1)n-1})^{-1} < \rho$$

It follows that the claim is true for at least one k, otherwise the bound will never be attained. □

Lemma 8. $\frac{1}{\ln(1+\frac{1}{r})} > r$.

Proof. We need to show that $\ln(1+\frac{1}{r}) < \frac{1}{r}$. To show this, we can use the Taylor series expansion for $\ln(1+x)$, which is valid for $|x| < 1$:

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - ...$$

For $x = \frac{1}{r}$, the series becomes:

$$\ln(1+\frac{1}{r}) = \frac{1}{r} - \frac{1}{2r^2} + \frac{1}{3r^3} - ...$$

Since $r > 0$, the higher-order terms $(\frac{1}{2r^2}, \frac{1}{3r^3}, \text{etc.})$ are all positive and become increasingly smaller. Therefore

$$\ln(1+\frac{1}{r}) < \frac{1}{r}.$$

Thus, the original inequality

$$\frac{1}{\ln(1+\frac{1}{r})} > r$$

is proven to be true. □

Theorem 7. $\Lambda_r(a) \geq \frac{1}{\ln(1+\frac{1}{r})} > r$.

Proof. We have $\tau_n < \frac{1}{\ln(1+1/r)}$ and $\tau_n \to \frac{1}{\ln(1+1/r)}$, so $\Lambda_r(a) \geq \frac{1}{\ln(1+1/r)}$. Together with Lemma 8 the theorem is proven. This holds for any sequence. □

2.4. Upper Bound for $\lambda_r(a)$

This section explains the upper bound for $\lambda_r(a)$ according to De Bruijn and Erdös [5]. We follow their general method, but occasionally use different proofs. Additionally, some of the lemmas are stated for just one stick, making them easier to grasp. This helps the reader for the more complicated scenario involving r consecutive sticks.

Definition 2. *For any set A, a function $\phi : A \to A$ is called a* ***Permutation*** *of A if ϕ is bijective.*

Definition 3. *A* ***Cyclic order*** *is a way of arranging the elements of a set on a circle (with a chosen direction, say clockwise or counterclockwise).*

Lemma 9. *Let n be a natural number, $\{a_1, ..., a_{2n}\}$ be a sequence of points, $\{a_{k_1}, a_{k_2}, ..., a_{k_{2n}}\}$ the cyclic order of these points on the circle (i.e., $k_1, ..., k_{2n}$ is a permutation of $1, ..., 2n$), and suppose that ρ is such that*

$$km_k^1(a) > \rho \qquad (n < k \leq 2n). \tag{2.19}$$

Set $k_{2n+1} = k_1$. If $k_i^* =$ Max $\{k_i, k_{i+1}, n+1\}$, then

$$1 > \rho \sum_{i=1}^{2n} \frac{1}{k_i^*}. \tag{2.20}$$

Proof. The stick $a_{k_i}, a_{k_{i+1}}$ is one of the sticks determined by $a_1, ..., a_{k_i^*}$. It follows that its length is greater than $\frac{\rho}{k_i^*}$,

$$\begin{aligned} 1 &= \sum_{i=1}^{2n} [a_{k_i}, a_{k_{i+1}}] \\ &\geq \sum_{i=1}^{2n} m_{k_i^*}^1 \\ &> \sum_{i=1}^{2n} \frac{\rho}{k_i^*}. \end{aligned}$$

Where the last inequality follows from (2.19). This completes our proof. □

Lemma 10. *For at least one k $(n < k \leq 2n)$ we have*

$$km_k^1(a) \leq (\frac{2}{n+1} + \frac{2}{n+2} + ... + \frac{2}{2n})^{-1} = \kappa_n.$$

Proof. We have $n < k_i^* \leq 2n \; \forall \; i$, and any k $(n+1 < k \leq 2n)$ occurs at most twice as it can be at the left end or at the right end of a stick. We want to minimise

$$\sum_{i=1}^{2n} \frac{1}{k_i^*}.$$

k_i^* comes from from $\{n+1, ..., 2n\}$ where the denominators $> n+1$ can occur at most twice. We always choose the largest k as it is in the denominator and we want it to be as small as possible. Even if $\{n+2, ..., 2n\}$ are all chosen twice, we still have to add 2 more terms, we take $\frac{1}{n+1}$ twice. Hence,

$$\begin{aligned} \sum_{i=1}^{2n} \frac{1}{k_i^*} &\geq \frac{1}{2n} + \frac{1}{2n} + \frac{1}{2n-1} + \frac{1}{2n-1} + ... + \frac{1}{n+1} + \frac{1}{n+1} \\ &= \frac{2}{2n} + \frac{2}{2n-1} + ... + \frac{2}{n+1} \\ &= \sum_{n+1}^{2n} \frac{2}{k}. \end{aligned} \tag{2.21}$$

From Lemma 9 and (2.21) we get

$$\rho < (\frac{2}{n+1} + \frac{2}{n+2} + ... + \frac{2}{2n})^{-1}$$

It follows that the claim is true for at least one k, otherwise the bound will never be attained. $\square$

We have $\kappa_n > \frac{1}{\ln 4}$ and $\kappa_n \to \frac{1}{\ln 4}$, so $\lambda_1(a) \leq \frac{1}{\ln 4}$. This holds for any sequence and the upper bound is attained for the sequence in section 2.2.

We will now repeat this approach (Lemma 9 and Lemma 10) for r consecutive sticks.

Lemma 11. *Suppose that ρ is such that*

$$km_k^r(a) > \rho \qquad (rn < k \leq (r+1)n). \tag{2.22}$$

If $k_i^* =$ Max $\{k_i, k_{i+1}, rn+1\}$, then

$$r > \rho \sum_{i=1}^{(r+1)n} \frac{1}{k_i^*}. \tag{2.23}$$

Proof. The r-stick $a_{k_i}, a_{k_{i+1}}$ is one of the r-sticks determined by $a_1, ..., a_{k_i^*}$. It follows that its length is greater than $\frac{\rho}{k_i^*}$,

$$\begin{aligned} r &= \sum_{i=1}^{(r+1)n} [a_{k_i}, a_{k_{i+1}}] \\ &\geq \sum_{i=1}^{(r+1)n} m_{k_i^*}^r \\ &> \sum_{i=1}^{(r+1)n} \frac{\rho}{k_i^*}. \end{aligned}$$

Where the last inequality follows from (2.22). This completes our proof. $\square$

Lemma 12. *For at least one k $(rn < k \leq (r+1)n)$ we have*

$$km_k^r(a) \leq r(\frac{r+1}{nr+1} + ... + \frac{r+1}{(r+1)n-1})^{-1} = \omega_n.$$

Proof. We have $rn < k_i^* \leq (r+1)n \; \forall \; i$, and any k $(rn+1 < k \leq (r+1)n)$ occurs at most $r+1$ times as it can be positioned at the end of each stick. When moving from left to right through the sticks, each stick has a left end. k can be placed at the left end of each stick. Additionally, for the last stick, k can also be at its right end. We want to minimise

$$\sum_{i=1}^{(r+1)n} \frac{1}{k_i^*}.$$

k_i^* comes from from $\{rn+1, ..., (r+1)n\}$ where the denominators $> rn+1$ can occur at most $r+1$ times. We always choose the largest k as it is in the denominator and we want it to be as small as possible. Even if $\{rn+2, ..., (r+1)n\}$ are all chosen $r+1$ times, we still have to add $r+1$ more terms, we take $\frac{1}{rn+1}$, $r+1$ times. Hence,

$$\begin{aligned}\sum_{i=1}^{(r+1)n}\frac{1}{k_i^*} &\geq \frac{1}{(r+1)n}+\frac{1}{(r+1)n}+\frac{1}{(r+1)n}+...+\frac{1}{rn+1}+\frac{1}{rn+1}\\ &=\frac{r+1}{(r+1)n}+\frac{r+1}{(r+1)n-1}+...+\frac{r+1}{rn+1}\\ &=\sum_{rn+1}^{(r+1)n}\frac{r+1}{k}. \end{aligned}\tag{2.24}$$

From Lemma 11 and (2.24) we get

$$\rho < r(\frac{r+1}{rn+1}+\frac{r+1}{rn+2}+...+\frac{r+1}{(r+1)n})^{-1}$$

It follows that the claim is true for at least one k, otherwise the bound will never be attained. □

Lemma 13. $\frac{r}{r+1}/\ln(1+\frac{1}{r}) < r$.

Proof.

$$\frac{r}{r+1}/\ln(1+\frac{1}{r}) = \frac{r}{(r+1)\cdot\ln(1+\frac{1}{r})}$$

and

$$\ln(1+x) = x-\frac{x^2}{2}+\frac{x^3}{3}-...$$

which is the Taylor series expansion for $\ln(1+x)$. This is valid for $|x|<1$.

For $x=\frac{1}{r}$, the series becomes:

$$\ln(1+\frac{1}{r}) = \frac{1}{r}-\frac{1}{2r^2}+\frac{1}{3r^3}-...$$

Since $r>0$, the higher-order terms $(\frac{1}{2r^2}, \frac{1}{3r^3}, \text{etc.})$ are all positive and become increasingly smaller. Therefore,

$$\ln(1+\frac{1}{r}) < \frac{1}{r}.$$

Thus, we get

$$\frac{r}{(r+1)\cdot\ln(1+\frac{1}{r})} < \frac{r}{(r+1)\cdot\frac{1}{r}} = \frac{r^2}{1+r} = r\cdot\frac{r}{r+1} < r\cdot 1 = r$$

since $\frac{r}{r+1}<1$ for all $r>0$. □

Theorem 8. $\lambda_r(a) \leq \frac{r}{r+1}/\ln(1+\frac{1}{r}) < r$.

Proof. We have $\omega_n > \frac{r}{r+1}/\frac{1}{\ln(1+1/r)}$ and $\omega_n \to \frac{r}{r+1}/\frac{1}{\ln(1+1/r)}$, so $\lambda_r(a) \leq \frac{r}{r+1}/\frac{1}{\ln(1+1/r)}$. Together with Lemma 13 the theorem is proven. This holds for any sequence. □

2.5. Lower Bound for μ_r

In this section we follow the proof of De Bruijn and Erdös [5] for finding a lower bound for μ_r.

Lemma 14. *Let $a_1, a_2, ..., a_n$ be a sequence of (breaking) points on a circle of circumference 1. For $r\geq 1, n\geq 1$ we have*

$$\frac{M_n^r(a)}{m_{n+1}^r(a)} \geq 1+\frac{1}{r}.\tag{2.25}$$

Proof. Suppose that $r > 1$. Let $I_1, I_2, ..., I_n$ be the sticks determined by $a_1, a_2, ..., a_n$. Let I_{k_0} be the stick that a_{n+1} breaks and let

$$I_{k_{-r+1}}, I_{k_{-r+2}}, ..., I_{k_0}, I_{k_1}, ..., I_{k_{r-1}} \tag{2.26}$$

be consecutive on the circle. Note that if $2r - 1 > n$ the k_i are not all different.

Let M_1 be the maximum length of the sum of r consecutive sticks from the set (2.26). Denote the length of stick I_{k_i} by β_i. Let γ_1 and γ_2 be the lengths of the two sticks into which I_{k_0} breaks by a_{n+1}.

By the pigeonhole principle at least one of the lengths $\beta_{-r+1}, \beta_{-r+2}, ..., \beta_{-1}, \beta_1, ..., \beta_{r-1}$, say β_j is $\geq \frac{M_1 - \beta_0}{r-1}$. The principle states that if n items are put into m containers, with $n > m$, then at least one container must contain more than one item [4]. In this context, the pigeonhole principle helps us understand that when distributing the total sum $M_1 - \beta_0$ among $r - 1$ sticks, at least one of the sticks must be large enough to ensure that the total sum is achieved. We may suppose that $j > 0$. Now we have

$$m_{n+1}^r(a) \leq \beta_{j-r+1} + \beta_{j-r+2} + ... + \beta_{-1} + \gamma_1 + \gamma_2 + \beta_1 + ... + \beta_{j-1} \tag{2.27}$$

and from $\beta_j \geq \frac{M_1 - \beta_0}{r-1}$ we derive

$$\begin{aligned}
\beta_j &\geq \frac{M_1}{r-1} - \frac{\beta_0}{r-1} \\
\Leftrightarrow -\beta_j &\leq -\frac{M_1}{r-1} + \frac{\beta_0}{r-1} \\
\Leftrightarrow M_1 - \beta_j &\leq M_1 - \frac{M_1}{r-1} + \frac{\beta_0}{r-1} \\
&\leq \frac{(r-1)M_1 - M_1}{r-1} + \frac{\beta_0}{r-1} \\
&\leq \frac{r-2}{r-1}M_1 + \frac{\beta_0}{r-1}
\end{aligned}$$

For the largest r-stick M_1 containing I_{k_0}; I_{k_j} is positioned on the left or right side of the r-stick. After breaking I_{k_0}, I_{k_j} is no longer part of the r-stick after $n + 1$ breaks. The smallest r-stick, after $n + 1$ breaks, is less than or equal to $M_1 - \beta_j$, where β_j is the length of I_{k_j}.

Hence,

$$m_{n+1}^r(a) \leq M_1 - \beta_j \leq \frac{r-2}{r-1}M_1 + \frac{\beta_0}{r-1}. \tag{2.28}$$

On the other hand it follows from

$$\begin{aligned}
m_{n+1}^r(a) &\leq \gamma_2 + \beta_1 + ... + \beta_{r-1} \leq M_1 - \gamma_1 \\
m_{n+1}^r(a) &\leq \beta_{-r+1} + ... + \beta_{-1} + \gamma_1 \leq M_1 - \gamma_2
\end{aligned}$$

that

$$2m_{n+1}^r(a) \leq 2M_1 - \beta_0$$

so

$$m_{n+1}^r(a) \leq M_1 - \frac{1}{2}\beta_0 \tag{2.29}$$

Trivially we have, $M_1 \leq M_n^r(a)$.

If $\beta_0 \leq \frac{2M_1}{r+1}$ then from (2.28)

$$\begin{aligned} m^r_{n+1}(a) &\leq \frac{r-2}{r-1}M_1 + \frac{2M_1}{(r+1)(r-1)} \\ &= \frac{r(r-1)}{(r-1)(r+1)}M_1 \\ &= \frac{r}{r+1}M_1 \\ &\leq \frac{r}{r+1}M^r_n(a) \end{aligned}$$

If $\beta_0 \geq \frac{2M_1}{r+1}$ then from (2.29)

$$\begin{aligned} m^r_{n+1}(a) &\leq M_1 - \frac{1}{2}\frac{2M_1}{r+1} \\ &\leq M_1 - \frac{M_1}{r+1} \\ &= \frac{r}{r+1}M_1 \\ &\leq \frac{r}{r+1}M^r_n(a). \end{aligned}$$

Which is the same result. This proves (2.25) for $r > 1$ as in both cases the claim is proven to be true. If $r = 1$, (2.25) follows from

$$m^r_{n+1}(a) = m^1_{n+1}(a) \leq \text{Min }\{\gamma_1, \gamma_2\} \leq \frac{1}{2}\beta_0 \leq \frac{1}{2}M^1_n(a) = \frac{1}{2}M^r_n(a).$$

Since in this case $(r = 1)$,

$$\begin{aligned} 1 \leq& \frac{1}{2}\frac{M^r_n(a)}{m^r_{n+1}(a)} \\ \Rightarrow& 1 + \frac{1}{r} = 2 \leq \frac{M^r_n(a)}{m^r_{n+1}(a)}. \end{aligned}$$

□

Lemma 15. *For at least one* k $(nr \leq k \leq nr + n)$ *we have*

$$\frac{M^r_k(a)}{m^r_k(a)} \geq \frac{1+\frac{1}{r}}{(1+\frac{1}{k})^2} = \delta_n.$$

Proof. Assume that n is a natural number and that for $nr \leq k \leq nr + n$ we have

$$\frac{M^r_k(a)}{m^r_k(a)} < \frac{1+\frac{1}{r}}{(1+\frac{1}{k})^2}. \tag{2.30}$$

By (2.25)

$$\frac{m^r_{k+1}(a)}{M^r_k(a)} \leq \frac{r}{r+1}. \tag{2.31}$$

Combining (2.30) and (2.31) we get for $nr \leq k < nr + n$

$$\frac{m^r_{k+1}(a)}{m^r_k(a)} < \frac{1+\frac{1}{r}}{(1+\frac{1}{k})^2} \cdot \frac{r}{r+1} = \frac{1}{(1+\frac{1}{k})^2} = \frac{1}{(\frac{k+1}{k})^2} = \frac{k^2}{(k+1)^2}. \tag{2.32}$$

By considering the product of the ratios of (2.32) over two consecutive steps

$$\frac{m^r_{k+2}(a)}{m^r_{k+1}(a)} \cdot \frac{m^r_{k+1}(a)}{m^r_k(a)} < \frac{(k+1)^2}{(k+2)^2} \cdot \frac{k^2}{(k+1)^2} = \frac{k^2}{(k+2)^2}.$$

This telescoping behavior helps in simplifying the product of ratios. We can generalize the telescoping process over multiple steps. For any positive integer n, we have

$$\frac{m^r_{k+n}(a)}{m^r_k(a)} < \frac{k^2}{(k+n)^2}.$$

Set $k = rn$. Then,

$$\frac{m^r_{rn+n}(a)}{m^r_{rn}(a)} < \frac{(rn)^2}{(rn+n)^2} = \frac{r^2n^2}{n^2(r+1)^2} = \frac{r^2}{(r+1)^2}. \tag{2.33}$$

On the other hand, by (2.30)

$$\begin{aligned} m^r_{rn+n}(a) &> \frac{(1+\frac{1}{k})^2}{1+\frac{1}{r}} \cdot M^r_{rn+n}(a) \\ &= (1+\frac{1}{k})^2 \cdot \frac{r}{r+1} \cdot M^r_{rn+n}(a) \\ &> \frac{r}{r+1} \cdot M^r_{rn+n}(a) \\ &\geq \frac{r}{r+1} \cdot \frac{r}{rn+n} \\ &= \frac{r^2}{(r+1)^2} \cdot \frac{1}{n}. \end{aligned} \tag{2.34}$$

By Proposition 2 we have $m^r_{rn}(a) \leq \frac{1}{n}$, which gives

$$\frac{1}{m^r_{rn}(a)} \geq n. \tag{2.35}$$

Combining (2.35) with (2.34) we get

$$\frac{m^r_{rn+n}(a)}{m^r_{rn}(a)} > \frac{r^2}{(r+1)^2} \cdot \frac{1}{n} \cdot n = \frac{r^2}{(r+1)^2}$$

which contradicts (2.33). Hence our assumption is false and the Lemma is proven to be true. □

Theorem 9. $\mu_r \geq 1 + \frac{1}{r}$.

Proof. We have $\delta_n < 1 + \frac{1}{r}$ and $\delta_n \to 1 + \frac{1}{r}$, so $\mu_r(a) \geq 1 + \frac{1}{r}$. □

With this proof we have reproduced all bounds provided by De Bruijn and Erdös [5]. Namely for μ, Λ, and λ for single sticks and r consecutive sticks.

The problem of μ_r, Λ_r, and λ_r is closely related to a problem concerning "just distributions" solved by Mrs van Aardenne-Ehrenfest [1]. All De Bruijn and Erdös could prove is that $\mu_r \geq 1 + \frac{1}{r}$ and analogous inequalities for Λ_r and λ_r. They conjecture that the expressions

$$r(\mu_r - 1) \quad , \quad r(\Lambda_r - 1) \quad , \quad r(1 - \lambda_r)$$

tend to infinity if $r \to \infty$ (the expressions unbounded).

3

Numerical Findings

In this chapter, I use a powerful tool that Erdös and de Bruijn did not have: a computer. The primary aim here is to uncover certain results through numerical methods. However, it's important to recognize that due to the limited numerical research conducted on this problem (and time constraints), not all questions will have definitive answers within this text. Consequently, I will occasionally pose questions for the reader to contemplate or investigate independently. My intention with this thesis is to encourage readers to explore this subject further and contribute to its growing body of research.

We start with a circle of length 1. For simplicity, in our code, we will consider it as a line of length 1, but keep in mind that it originates from a circle. This means the first and last sticks of our line are next to each other. We will break this line into sticks, with the cutting points determined by various stick breaking strategies.

Let's first clarify what we are willing to find out. We want the largest length of two consecutive sticks to be as close to the smallest length of two consecutive sticks as possible. In other words, minimise

$$M_n^2(a) - m_n^2(a).$$

This ensures to keep the distribution of two consecutive sticks as uniform as possible. We consider four stick breaking strategies: Random Cutting (3.2), Maximum Stick Random Cutting (3.3), Maximum Stick in Half Cutting (3.4), and Cutting at $\log_2(2n-1)$ mod 1 (3.5). Currently, we don't know which strategy will yield an even distribution of sticks.

Focusing on two consecutive sticks is convenient because extending the theory to r consecutive sticks simply requires changing a few variables in the code. To compare the lengths of sticks, we define a measure to evaluate the uniformity. Instead of analyzing μ_2 (see (2.3)), the ratio between the longest and shortest lengths of two consecutive sticks, we introduce a new measure, the R-value (see section 3.1). We analyze the R-values for the different stick breaking strategies. Our ultimate goal is to find a strategy with a low R-value. The following section provides definitions and explanations leading to how the R-value is defined and used in our analysis.

Definition 4. *A* $\mathbf{2}$***-stick*** *are two consecutive sticks, and a* $\mathbf{r}$***-stick*** *are* r *consecutive sticks.*

Definition 5. $M(n)$ *is the longest* 2*-stick after* n *breaks, and* $m(n)$ *is the shortest* 2*-stick after* n *breaks.*

Definition 6. *The* $\mathbf{R}$***-value*** $R(n)$ *(after* n *breaks) of a stick-breaking strategy is defined as the length of the longest* 2*-stick minus the length of the shortest* 2*-stick times the amount of breaks* n*,*

$$R(n) = (M(n) - m(n)) * n.$$

We multiply by n *to normalise as otherwise* R *will be too small when we look at a large amount of breaks.*

3.1. R-Value

Recall that the goal is to find a stick breaking strategy with an even distribution of 2-sticks. In order to measure that we will use the R-value. An evenly distributed strategy would have $M(n)$ and $m(n)$, defined as in Definition 6, very close to each other. This is the same as saying that their difference should be as close as possible to zero, hence minimised. Our mission becomes analysing for which strategy the R-values are as low as possible.

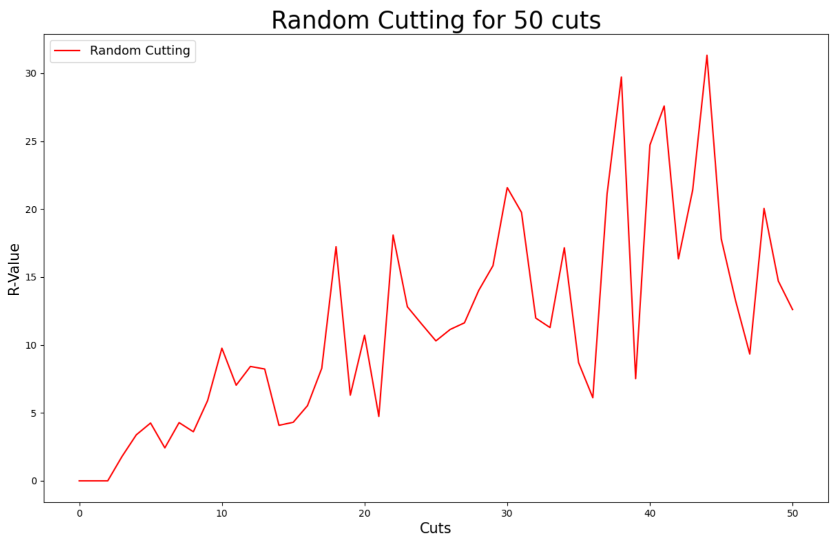
For every strategy, the calculate_r function in Appendix A computes the R-value as follows: the script first sorts the stick-lengths in the correct order. Then it checks what sum of two consecutive sticks is the largest. Note that I have introduced a Boolean variable (is_last_item) for when we arrive at the last (single) stick. Because in that case our 2-stick consists of the last stick and the first stick.

Now that we have found $M(n)$ and $m(n)$ we can compute our R-value by computing $(M(n) - m(n)) * n$.

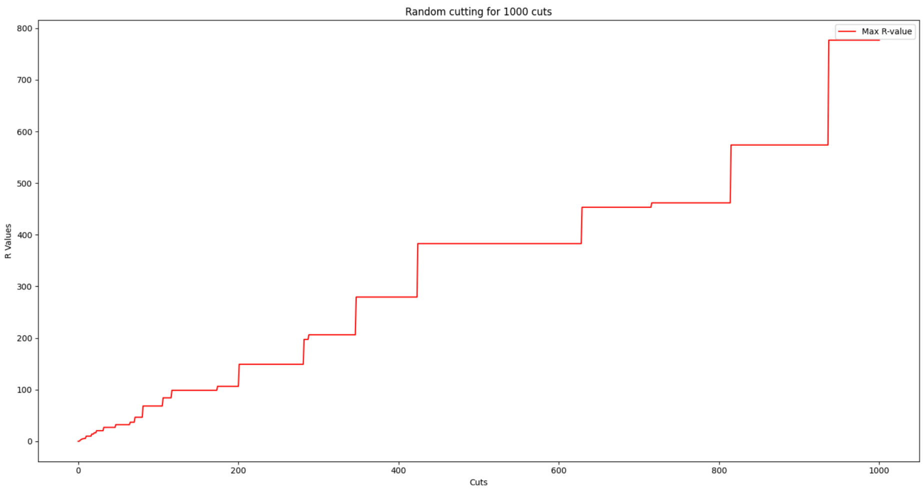
Note that if $n = 1$, i.e., we cut our circle once, we will only have one stick of length 1. For one stick the problem is not defined as we look at the length of two consecutive sticks. Therefore, I have hardcoded that $R(1) = 0$ for one stick.

3.2. Strategy 1: Random Cutting

Put simply, this strategy consists of breaking the stick unformly at random. Then it computes the R-value with the calculate_r function. This strategy is called random_cutting in the code in Appendix A.



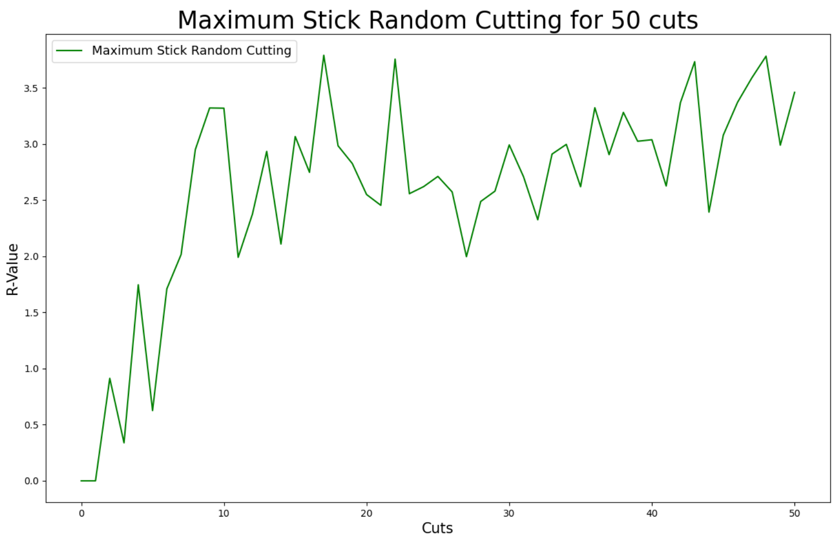
This is a plot of the R-values for the Random Cutting Strategy up to 50 cuts. We see no pattern or recurring behaviour, and we did not expect any as everything is random in this strategy. We now will look at more cuts with this strategy but we do not expect good (low) R-values.



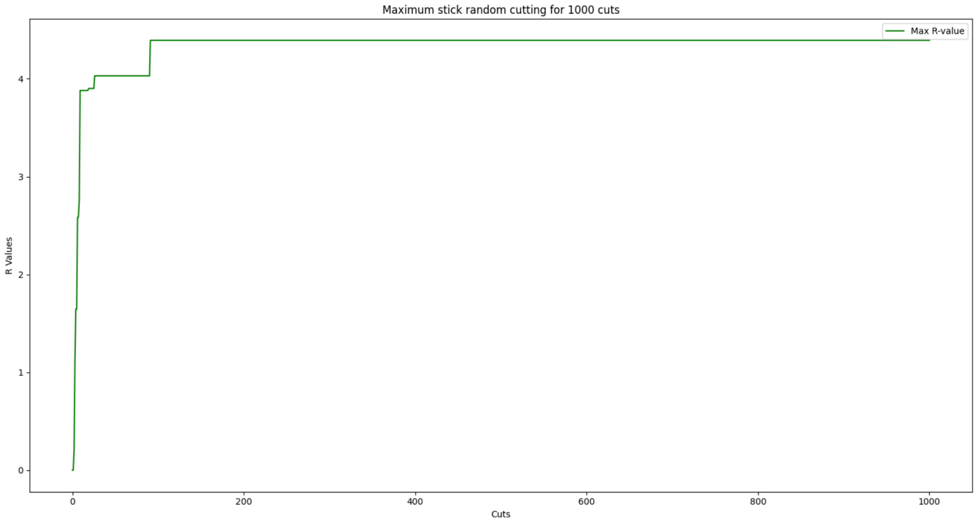
The R-values for the Random Cutting Strategy up to 1000 cuts are very fluctuating and hence not convenient to watch in a graph. You can smooth it out by plotting the "moving" maximum. For each cut number n plot max $R(m)$ for $m < n$, the highest R-value up to that point. This plot shows the "moving" maximum of the R-values till 1000 cuts. As expected this strategy gives us very high R-values with one of them having an R-value over 700. We can imagine that for $n \to \infty$ amount of cuts, the R- value will shoot up to ∞ too. We conclude that this is a very bad strategy.

3.3. Strategy 2: Maximum Stick Random Cutting

This strategy operates in the same way as the previous strategy except that it cuts the first largest single-stick instead of a random stick. More specifically, the code picks the first largest single-stick. If there are more than one largest single-sticks, it picks the first one, meaning the one that is the closest to zero. Then, it breaks this stick into 2 new sticks. For this it takes uniform at random a value between 0 and the length of the largest stick, say new_stick_length. The other stick length is naturally the length of the largest stick minus new_stick_length. Now all we have to do is put our two new sticks back into our list and compute the R-value with the calculate_r function. The strategy is called max_stick_cutting in the code in Appendix A. This procedure is named Kakutani's Splitting Procedure, the reader can find more information about it in [6] and [11].



The plot illustrates the R-values for the Maximum Stick Random Cutting Strategy up to 50 cuts. We can already see that we have found a much better strategy as the R-values stay below $3,7$ compared to a stunning R-value of 30 for the Random Cutting Strategy. This observation holds for up to 50 cuts. Keep in mind that because we do some things randomly, we can obtain different R-values for the same number of cuts if we run the code again. We still do not see any recurring behaviour on the graph.

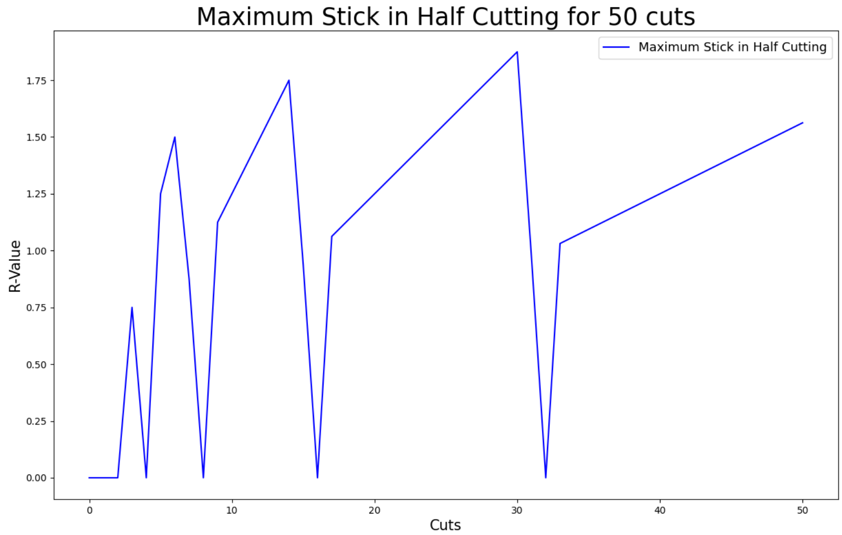


This graph shows the ”moving” maxmimum R-values for the Maximum Stick Random Cutting Strategy up to 1000 cuts. We can observe that all the maximum R-values stay below $4,5$. The R-values are low for a few cuts and then rapidly increase to $4,5$ and then stay under this bound. A question that is raised with this graph is the following: Do the R-values converge to $4,5$? Perhaps to 4? I will leave this question open for my fellow mathematicians.

This is how far I will go with this strategy. However, I suggest a new strategy here: to cut somewhere in the largest 2-stick. I believe this may improve the R-values as the largest single stick is not always part of the largest 2-stick (and in the Maximum Stick Random Cutting Strategy we always cut in the largest 1-stick).

3.4. Strategy 3: Maximum Stick in Half Cutting

This strategy takes the largest (single-)stick and cuts it in half. Specifically, the code picks the first largest stick. Then, it breaks this stick in half. Now we get two sticks of the same length which are half the length of the largest stick. It remains to put our two new sticks back into our list of stick lengths and compute the R-values with the calculate_r function. This strategy is called max_stick_in_half_cutting in the code in Appendix A.



This method is deterministic, and not at random. These are the ten (starting at 1) first R-values of this method

$$0, 0, \frac{3}{4}, 0, \frac{5}{4}, \frac{6}{4}, \frac{7}{8}, 0, 1, \frac{9}{8}, \frac{10}{8}$$

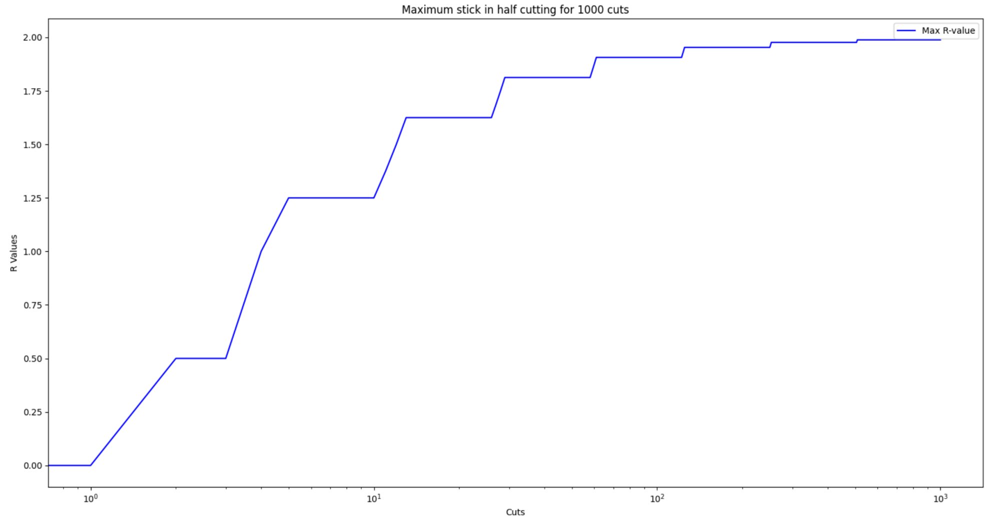
The graph illustrates the R-values for the Maximum Stick in Half Cutting Strategy up to 50 cuts. We see a recurring behaviour. Analysing this, we start at $R(n) = 0$ for $n = \{0, 1, 2\}$ cuts as expected. Then we have a first spike of the R-value at 3 cuts. This is because the cutting process has resulted in sticks of different lengths.

At $n = 4$, the R-value drops to zero again. This occurs because after the fourth cut, the sticks are of equal lengths again, so $M(4) = m(4)$. The pattern of spikes and drops to zero continues. Each time the number of cuts results in unequal lengths of sticks, $R(n)$ spikes. When the number of cuts results in equal lengths, $R(n)$ resets to zero.

For example, at $n = 8$, the R-value is zero again because all the parts are of the same length, $M(8) = m(8)$. Between the resets, the R-value increases linearly. This linear increase occurs because $M(n)m(n)$ remains constant while n increases linearly.

The maximum R-value does not exceed 2. This is because the longest stick M can be at most twice as long as the shortest stick m due to the halving nature of the cuts. The graph shows a repeating behavior where spikes and resets alternate, reflecting the cyclical nature of stick lengths balancing out after certain numbers of cuts. This pattern is expected to continue with more cuts.

Note that at $n = 4$ you have the numbers $0, 1/4, 1/2, 3/4$. You have four choices to continue. Take the first stick, then you get $0, 1/8, 1/4, 1/2, 3/4$. The 2-sticks have lengths $1/4, 3/8, or 1/2$. Now you can continue with $0, 1/8, 1/4, 3/8, 1/2, 3/4$, but also with $0, 1/8, 1/4, 1/2, 5/8, 3/4$. The second way gives a better R-value. By cutting another (not the first) largest single-stick you get a better R-value. We do not further investigate this thought.



For more cuts we can smooth the graph out again via the ”moving” maximum. It seems logarithmic, so we plot it on a logarithmic scale.

The graph shows the ”moving” maximum R-values for the Maximum Stick in Half Cutting Strategy up to 1000 cuts. As anticipated, the behavior remains consistent with the pattern observed in the earlier plot.

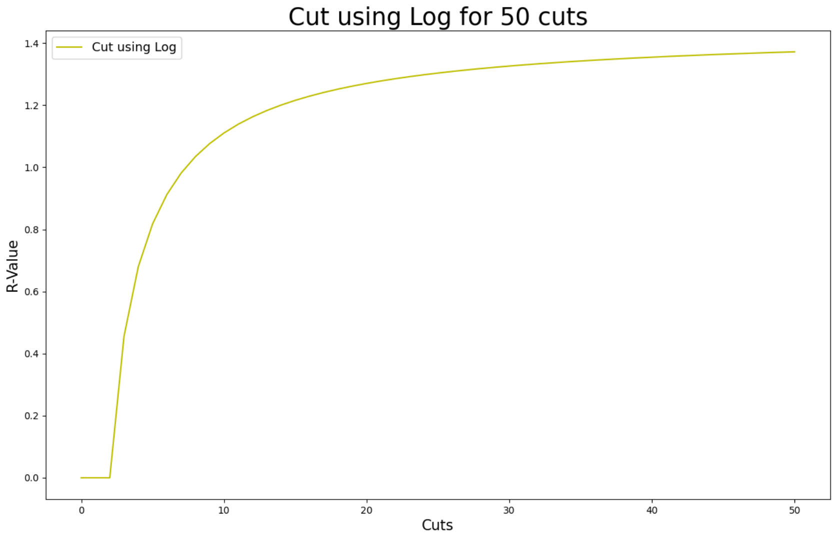
The cyclical pattern of spikes and resets in R-values persists (for the R-value graph, not the ”moving” maximum graph). Moreover, The maximum R-value achieved is 2, which aligns with our expectation. The strategy we use resembles the van der Corput method. For further reading and a deeper understanding of the underlying principles, I recommend reading [7].

3.5. Strategy 4: Cutting at $\log_2(2n-1)$ mod 1

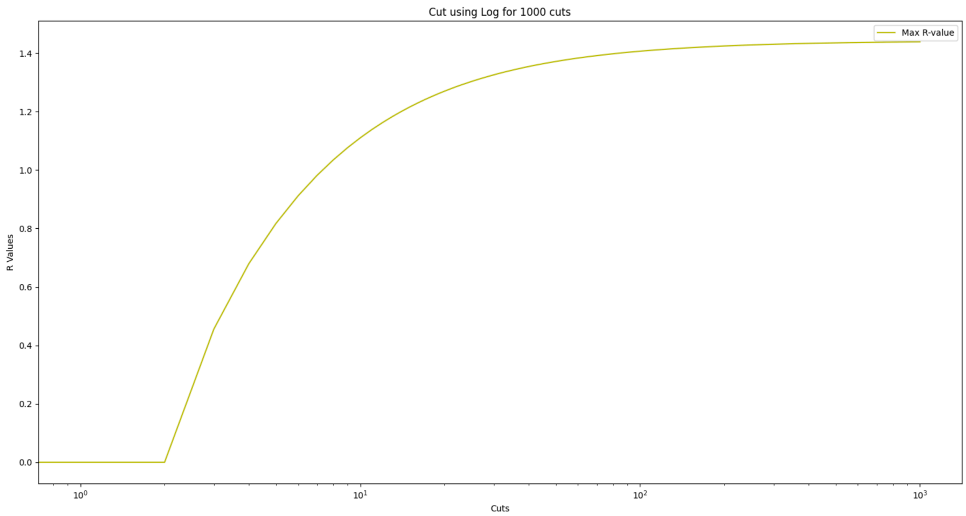
Recall from Section 2.2 that the sequence $log_2(2n-1)$ gave the best possible $\mu_1(a)$. It makes sense to analyse this sequence and determine the R-values for this strategy. To program this, another approach is needed. For every strategy up until now we knew directly what the lengths of the sticks where. However, now we know where we cut our line (of length 1) but we don't know what stick it breaks and hence we don't know its length.

First I have coded a ”dummy” function that simply returns the value of $log_2(2n-1)$ mod 1. This function is called calculate_cut_place in Appendix A. I have made this function so that if the reader wants to compute R-values for another function, he/she will simply have to modify the code in this ”dummy” function.

To compute the R-value for a specific amount of cuts n we first have to compute $log_2(2n-1)$ mod 1 for the values from 1 up until n. Once we got that, we sort them from the largest to the smallest (we also include 1 which is the largest possible value), say this gives $1, l_1, l_2, ..., l_{n-1}, l_n, 0$. Now computing the lengths of the sticks is easy, the lengths are $1-l_1, l_1-l_2, ..., l_n-l_{n-1},$ and l_n-0. We swap the lengths around, yielding $l_n-0, l_n-l_{n-1}, ..., l_1-l_2, 1-l$. These are the lengths of the sticks in correct order. Now we can compute the R-value with the calculate_r function. This process takes place in the function cut_using_logaritm_algoritm in Appendix A.

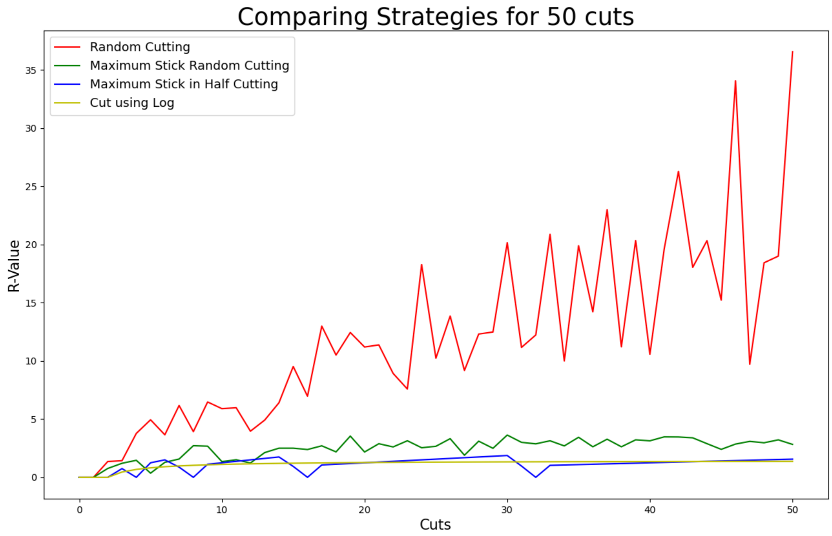


This plot displays the R-values for the Cutting at $\log_2(2n-1)$ mod 1 Strategy up to 50 cuts. The R-values increase in a Logarithmic way and seem to be bounded by $1,4$. This sequence has optimal μ_1 (for 1-sticks), hence it is very evenly distributed. It seems logic to think that it would also have a good distribution of 2-sticks. This is true as there seem to be an asymptote at $1,4$ so we again have found a better strategy. For 1000 cuts we do not expect much to change, the tail of the graph will probably look like a horizontal line.

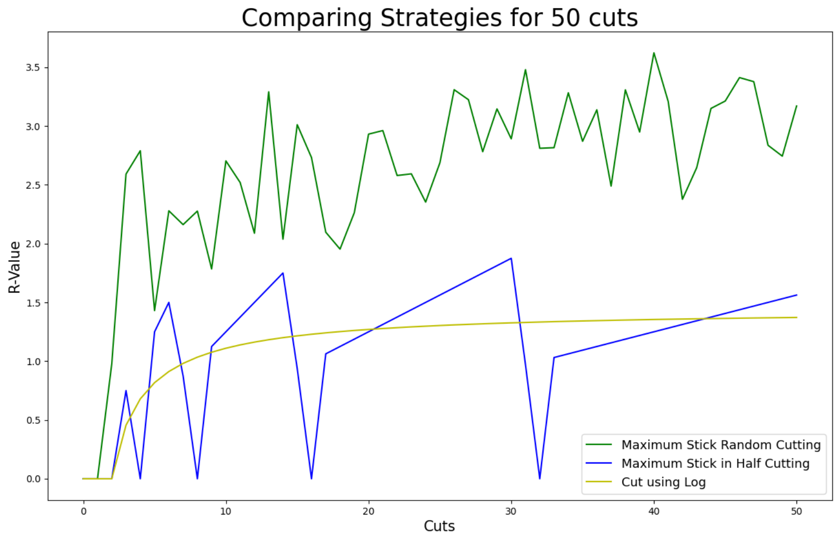


The plot shows the R-values for the Cutting at $\log_2(2n-1)$ mod 1 Strategy up to 1000 cuts (note that the x-axis is lograithmic). The graph surely has passed the previously established asymptote of $1,4$. I leave my fellow mathematicians explain why that happens and find the correct value of the asymptote. I will not dive deeper in the analysis of the logarithmic strategy. However, it seems a good approach to try things with this formula. Like changing the base of the logarithm, or changing the formula inside the brackets. I will leave your imagination up to that.

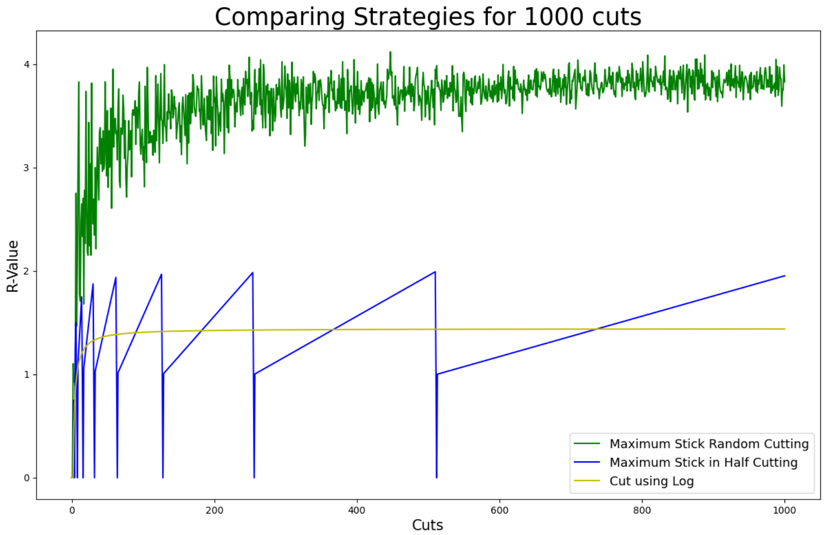
3.6. Conclusion



The plot above compares all of our strategies. We can see what has been discussed in previous sections. The Random Cutting Strategy dominates the graph, making it difficult to observe the other strategies in detail. For 1000 cuts it gets worse. Hence, I provided the graph which compares the 4 strategies for 1000 cuts in Appendix B. To better evaluate the remaining three strategies, we will now remove the Random Cutting Strategy from the plot, allowing for a clearer and more insightful comparison.



In this graph, we observe that among these three strategies, the Maximum Stick Random Cutting Strategy performs the worst, as its R-values are consistently higher than those of the other two strategies. This outcome is expected since this strategy involves randomly cutting the largest stick without much thought or optimization.



This graph extends the previous comparison to 1000 cuts, and the same explanation applies. The behavior remains consistent, confirming that the Maximum Stick Random Cutting Strategy is ineffective. We conclude that this strategy is not optimal.

Between the two remaining strategies, the Logarithmic strategy is the best. Even if the Maximum Stick in Half Cutting strategy sometimes has lower R-values, it is about the limsup of $R(n)$. Which is lowest for the Logarithmic strategy.

4

Endnote

To conclude, the problem for single sticks is mostly solved, as shown by the mathematical proofs in Chapter 2. However, when we look at consecutive sticks, it becomes clear that the stick breaking problem still needs more research. I believe that for anyone interested in exploring this area further, my report will be an essential starting point. Additionally, the code included in this report will be very useful for researchers who want to dig deeper into this topic. With some modifications, the script can be adapted to provide new insights and potentially discover more effective strategies for solving the problem.

Although this research has its limitations, it provides a solid foundation for ongoing exploration in the field. Future studies can build on the concepts and methods presented here, leading to a better understanding of stick breaking strategies and their distributions.

I am very grateful for the opportunity to work on this topic. This project has been a significant learning experience, and I hope it will inspire further investigations and discoveries in this fascinating area of study.

References

[1] T. van Aardenne-Ehrenfest. “Proof of the impossibility of a just distribution of an infinite sequence of points over an interval”. In: *Proc. Kon. Ned. Akad. v. Wetensch.* 48 (1945), pp. 266–271.

[2] Alozano. *an injection between two finite sets of the same cardinality is bijective*. 2022. URL: https://planetmath.org/aninjectionbetweentwofinitesetsofthesamecardinalityisbijective (visited on 06/19/2024).

[3] See website for authors. *Harmonic Series (mathematics)*. 2024. URL: https://en.wikipedia.org/wiki/Harmonic_series_(mathematics) (visited on 05/13/2024).

[4] See website for authors. *Pigeonhole principle*. 2024. URL: https://en.wikipedia.org/wiki/Pigeonhole_principle (visited on 06/14/2024).

[5] N. G. de Bruijn and P. Erdös. “Sequences of Points on a Circle”. In: *Proceedings of the Section of Sciences of the Koninklijke Nederlandse Akademie van Wetenschappen te Amsterdam* 52.1 (1949), pp. 14–17.

[6] I. Carbone and A. Volčič. “Kakutani’s Splitting Procedure in Higher Dimension”. In: *Rend. Istit. Mat. Univ. Trieste* XXXIX.1 (2007), pp. 119–126.

[7] M. R. Iacò. “Low-Discrepancy Sequences: Theory and Applications”. In: *ANALISI MATEMATICA* XXVII.1 (2015), pp. 1–86.

[8] metamorphy. *StackExchange*. 2022. URL: https://math.stackexchange.com/questions/3335549/how-to-prove-that-limsup-n-to-inftyn-m-n-geq-1-ln-2-for-the-following-sequ (visited on 05/01/2024).

[9] A. Ostrowski. “Eine Verschärfung des Schubfächerprinzips in einem linearen Intervall”. In: *Arch. Math.* 8 (1957), pp. 1–10.

[10] A. Ostrowski. “Zum Schubfächerprinzip in einem linearen Intervall”. In: *Mathematische Miazellen* XXVI (1957), p. 60.

[11] R. Pyke and W. R. van Zwet. “Weak Convergence results for the Kakutani interval Splitting procedure”. In: *The Annals of Probability* 32.1A (2004), pp. 380–423.

[12] I. Surname, I. Surname, and I. Surname. *The Title of the Book*. 8th ed. City, State or Country: Publisher, 2000.

[13] R. Tijdeman. “Gelijkverdeelde rijen bestaan niet”. In: *Mathematisch Instituut* 5/19.2 (2018), pp. 90–92.

[14] G. H. Toulmin. “Subdivision of an Interval by a Sequence of Points”. In: *Arch. Math.* VIII (1957), pp. 158–161.

A

Code

```python
import pandas as pd
import math
import random
```

```python
def calculate_r(sticks, amount_cuts = 4):
large_m = 0
small_m = 1

is_last_item = False

for index, stick in enumerate(sticks):
    if(len(sticks) == 1):
        return 0
    if(index == len(sticks) - 1):
        is_last_item = True

    total = stick + sticks[0 if is_last_item else index+1]
    if(total > large_m):
        large_m = total
    if(total <= small_m):
        small_m = total

return (large_m - small_m) * amount_cuts
```

```python
def random_cutting(amount_cuts = 4):
    parts = [1]
    for _ in range(amount_cuts):
        parts_copy = parts.copy()
        i = random.randrange(len(parts_copy))
        stick_length = parts[i]
        del parts[i]
        new_stick_length = random.uniform(0, stick_length)
        other_new_stick_length = stick_length - new_stick_length
        parts = parts_copy[0:i] + [new_stick_length, other_new_stick_length] + parts_copy[i+1:]

    return calculate_r(parts, amount_cuts)
```

```python
def max_stick_cutting(amount_cuts = 4):
    parts = [1]
    for _ in range(amount_cuts):
        parts_copy = parts.copy()

        stick_length = max(parts)
        i = parts.index(stick_length)
        parts.remove(stick_length)

        new_stick_length = random.uniform(0, stick_length)
        other_new_stick_length = stick_length - new_stick_length
```

```python

        parts = parts_copy[0:i] + [new_stick_length, other_new_stick_length] + parts_copy[i
            +1:]

    return calculate_r(parts, amount_cuts)
```

```python
def max_stick_in_half_cutting(amount_cuts = 4):
    parts = [1]
    for _ in range(1, amount_cuts):
        parts_copy = parts.copy()

        stick_length = max(parts)
        i = parts.index(stick_length)
        parts.remove(stick_length)

        new_stick_length = stick_length / 2

        parts = parts_copy[0:i] + [new_stick_length, new_stick_length] + parts_copy[i+1:]
    return calculate_r(parts, amount_cuts)
```

```python
def calculate_cut_place(amount_cuts):
    return math.log2((2*amount_cuts)-1) % 1
```

```python
def cut_using_logaritm_algoritm(amount_cuts = 4):
    if(amount_cuts == 0):
        return calculate_r([1], 0)

    places_to_cut = [1]
    all_sticks = []

    for i in range(1, amount_cuts+1):
        place_to_cut = calculate_cut_place(i)
        places_to_cut.append(place_to_cut)
    places_to_cut = sorted(places_to_cut)[::-1]

    for index, val in enumerate(places_to_cut):
        if(index == len(places_to_cut)-1):
            break
        all_sticks.append(val - places_to_cut[index+1])

    all_sticks = all_sticks[::-1]
    return calculate_r(all_sticks, amount_cuts)
```

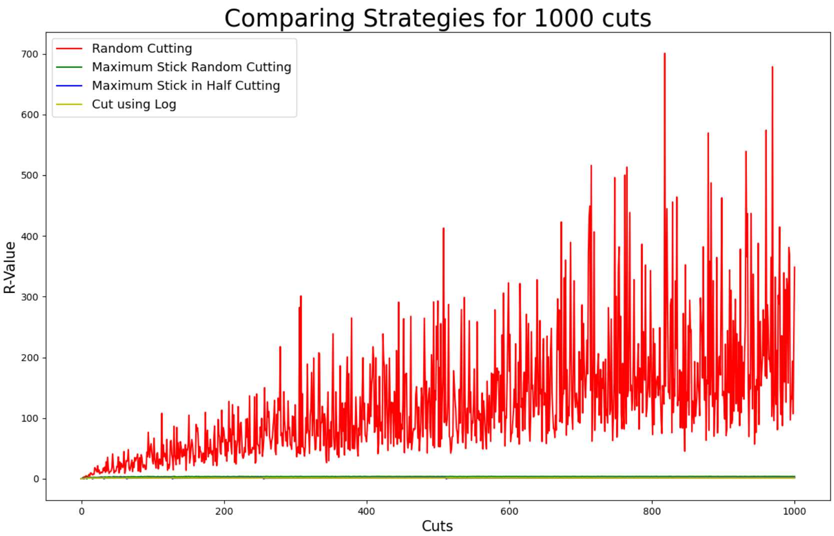
```python
# This code is simply to display the R-values in a simple table as sometimes it can be hard
    to read the exact R-value from the plot. One column shows the exact R-values, and the
    other shows the amount of cuts corresponding to it.
def show_r_values_in_range(amount_cuts):
    df = pd.DataFrame(columns=['r_values', 'amount_of_cuts'])

    for i in range(1, amount_cuts + 1):
        r_value = cut_using_logaritm_algoritm(i)
        df = pd.concat([df, pd.DataFrame({'r_values': [r_value], 'amount_of_cuts': [i]})],
            ignore_index=True)

    return df
```

```python
import matplotlib.pyplot as plt
from main import random_cutting, max_stick_cutting, max_stick_in_half_cutting,
    cut_using_logaritm_algoritm

def generate_data_for_algorithm(func, amount_of_cuts = 4):
    r_values = []
    for i in range(amount_of_cuts + 1):
        r_values.append(func(i)) # cut_using_logaritm_algoritm(2)
    print(r_values)
    return r_values

def create_plot():
    amount_of_cuts = 20
    data_random_cutting = generate_data_for_algorithm(random_cutting, amount_of_cuts)
```

```python
    data_max_stick_cutting = generate_data_for_algorithm(max_stick_cutting, amount_of_cuts)
    data_max_stick_in_half_cutting = generate_data_for_algorithm(max_stick_in_half_cutting, amount_of_cuts)
    data_sequence_algoritm = generate_data_for_algorithm(cut_using_logaritm_algoritm, amount_of_cuts)

    x_values = range(amount_of_cuts + 1)

    plt.plot(x_values, data_random_cutting, label='Random Cutting', color='r')
    plt.plot(x_values, data_max_stick_cutting, label='Max Stick Cutting', color='g')
    plt.plot(x_values, data_max_stick_in_half_cutting, label='Max Stick in Half Cutting', color='b')
    plt.plot(x_values, data_sequence_algoritm, label='Sequence algorithm', color='y')

    plt.xlabel('Cuts')
    plt.ylabel('R-Value')
    plt.title('Comparison of Stick Breaking Strategies')
    plt.legend()
    plt.show()

create_plot()
```

B

Comparing Strategies



C

Finding ϵ

We first want to approximate $\ln(2 - \frac{1}{N+1}) = \ln 2$ when N is large. When N is large, $\frac{1}{N+1}$ is small. Let $x = \frac{1}{N+1}$. Hence, we need to approximate $\ln(2 - x)$ for small x.

The natural logarithm function $\ln(1 + y)$ for y close to 0 can be expanded using a Taylor series:

$$\ln(1 + y) = y - \frac{y^2}{2} + \frac{y^3}{3} - ...$$

For very small y, we often use the first-order approximation:

$$\ln(1 + y) \approx y.$$

We want to approximate $\ln(2 - x)$. Rewrite it in a form that uses the series expansion for $\ln(1 + y)$:

$$\ln(2 - x) = \ln(2(1 - \frac{x}{2})).$$

Using the properties of logarithms, this can be split into:

$$\ln(2 - x) = \ln 2 + \ln(1 - \frac{x}{2}).$$

Since x is small, $\frac{2}{x}$ is also small. We can apply the first-order Taylor expansion to $\ln(1 - \frac{x}{2})$:

$$\ln(1 - \frac{x}{2}) \approx -\frac{x}{2}.$$

Therefore,

$$\ln(2 - x) \approx \ln 2 - \frac{x}{2}.$$

Recall that $x = \frac{1}{N+1}$. Substituting x back, we get:

$$\ln(2 - \frac{1}{N+1}) \approx \ln 2 - \frac{1}{2(N+1)}.$$

This needs to be greater than $\frac{\ln 2}{1+2\epsilon\ln 2}$:

$$\ln 2 - \frac{1}{2(N+1)} > \frac{\ln 2}{1+2\epsilon\ln 2}$$

$$1 - \frac{1}{2(N+1)\ln 2} > \frac{1}{1+2\epsilon\ln 2}$$

$$\frac{2\ln 2(N+1)-1}{2(N+1)\ln 2} > \frac{1}{1+2\epsilon\ln 2}$$

$$1+2\epsilon\ln 2 > \frac{2(N+1)\ln 2}{2\ln 2(N+1)-1}$$

$$2\epsilon\ln 2 > \frac{2(N+1)\ln 2}{2\ln 2(N+1)-1} - 1$$

$$2\epsilon\ln 2 > \frac{2(N+1)\ln 2 - 2\ln 2(N+1)+1}{2\ln 2(N+1)-1}$$

$$\epsilon > \frac{1}{2\ln 2(N+1)-1}$$