University of Cincinnati

Date: 5/11/2011

I, Benjamin E Merkel, hereby submit this original work as part of the requirements for the degree of Master of Science in Mathematical Sciences.

It is entitled:
Probabilities of Consecutive Events in Coin Flipping

Student's name: Benjamin E Merkel

This work and its defense approved by:

Committee chair: Stephan Pelikan, PhD
Committee member: Donald French, PhD
Committee member: Joanna Mitro, PhD
Probabilities of Consecutive Events in Coin Flipping

A thesis submitted to the
Graduate School
of the University of Cincinnati
in partial fulfillment of the
requirements for the degree of
MASTERS OF SCIENCE (MS)
in the Department of Mathematical Sciences
of the College of Arts and Sciences
2011
by
Benjamin Earnest Merkel
B.A., University of Cincinnati, 2009

Committee Chair: Dr. Stephan Pelikan

Thesis Committee:
Dr. Donald A. French
Dr. Joanna Mitro
Abstract

The motivation of my thesis came from a problem I heard on Radiolab, a podcast distributed through National Public Radio. In the podcast, the two hosts asked the question, “What is the probability of flipping seven consecutive tails when flipping a coin a hundred times?” They approximated the probability to being 1/6. From a mathematical point of view, this seems like too simple of an answer because there are $2^{100}$ cases one must consider.

In my thesis, I first go about finding an exact probability to this initial question. Afterwards, I show how one can answer a generalized version of this question, where the number of flips and the number of consecutive events are variable. Additionally, I show how to find the probability of consecutive heads or tails occurring.

By answering these questions, I learned calculation techniques using matrix multiplication. These methods are shared in the paper. Lastly, I go into some of the underlying mathematics in this matrix multiplication and how it relates to their related recursive sequences.
# Table of Contents

The Question .................................................................1

Some Initial Strategies ......................................................2

Underestimation ............................................................2

Overestimation ..............................................................3

Brute Force .................................................................3

Simplification of Brute Force .............................................4

Specific Solution ..........................................................8

Extended Answer ..........................................................11

Calculation Techniques ..................................................14

Generalized Answer .......................................................15

"Nacci Matrices" ..........................................................16

References .................................................................20
1 The Question

In June of 2010, a group of friends and I were eating dinner together in Columbus. One friend in particular, we’ll call her Steph, knew about my interest in mathematics and presented us with a math problem she had heard on the radio. The problem came from Radiolab, a podcast on NPR, in which the two hosts, Jad Abumrad and Robert Krulwich, were studying stochasticity (or randomness). In one of their studies, they were trying to gain a better understanding of what randomness looks like. To do this, they visited Deborah Nolan, a statistics professor from Berkley. She separated students into two groups. Professor Nolan instructed the first group to flip a coin one hundred times and record their results. Then the second group was instructed to guess results of a hundred coin flips. In other words, one group was recording results of coin flips while the other was fabricating them.

Afterwards, the results were posted anonymously on the board and within seconds, Professor Nolan was able to tell which results were real and which were made up. The reason behind this seemingly miraculous feat was that one of the results, the actual flips, had a streak of seven tails in a row. The group that was making up results would never make a streak that long because intuitively, it seems way too unlikely to happen. This brings us to the question that was presented to my friends and me:

What is the probability of flipping seven consecutive tails when flipping a fair coin one hundred times?

Like solving most mathematical problems, this is just a specific case of a larger question, namely:

What is the probability of flipping $m$ consecutive tails when flipping a fair coin $n$ times?

where $n$ is greater than or equal to $m$. The purpose of this paper is to explain how to answer these two questions and explore some of the related problems and applications that arose in seeking a solution. But before we go about tackling this generalized question, let us look at some attempts of how others tried to solve this problem.
2 Some Initial Strategies

2.1 Underestimation

First let us look at how Radiolab, the podcast which presented the problem, attempted to answer the specific question (the probability of seven consecutive tails in 100 flips). They consulted Jay Koehler, a professor in finance, to help them find a solution. Koehler’s approach was to break down the hundred coin flips into fourteen sets of seven flips, since seven times fourteen is approximately one hundred. For each of these sets of seven flips, the probability of getting seven tails in a row is exactly 1 in 128. Therefore, the probability of not getting seven tails in a row is \( \frac{127}{128} \). So the probability of NOT flipping seven tails in a row for each of the fourteen sets is \( \frac{127}{128} \) or 89.6%. That means the probability of getting seven consecutive tails, according to this argument, is at least 10.4%.

Curiously, Radiolab, or rather Jay Koehler, quotes the probability being about one-sixth, or 16.6%. Since they don’t explicitly say how they arrived at this percentage, we’ll just have to take this number on good faith. Either way, both percentages are underestimates of the actual percentages, and furthermore, are not very accurate, as we will later see.

The source of the inaccuracy comes from lumping a hundred separate incidents into fourteen isolated groups. For instance, let us consider just the first fourteen flips. According to the first method, the only cases that have seven consecutive tails are when flips one through seven are tails, flips eight through fourteen are tails, or both. The case where flips two through eight are tails isn’t necessarily counted as a case of seven consecutive tails. Nor is the case of flips three through nine, four through ten, etc. That’s why this method provides a low estimation. It certainly counts some of the cases we’re looking for, but by no means all of them.

So in general, using this method, the probability of flipping seven tails in a row in \( n \) flips is at least \( 1 - \left( \frac{127}{128} \right)^n / 7 \). And even more generally, the probability of flipping \( m \) consecutive tails in \( n \) flips is at least \( 1 - \left( \frac{2^m-1}{2^n} \right)^n / m \).

Though a bit conservative, this is actually a decent model. It matches two of our expectations. First, as the number of flips, \( n \), increases, it becomes more probable that a string of \( m \) consecutive tails will occur. Secondly, if \( m \) is equal to \( n \), we get that the probability of flipping \( m \) tails in a row is \( 1 - \frac{2^m-1}{2^n} = \frac{1}{2^n} \), which is the correct answer. Additionally, as \( m \) and \( n \) both increase to infinity, we find that the probability of flipping \( m \) consecutive tails in \( n \) flips approaches zero. This too should match what we expect. If we flip a coin three billion
times, we should think the event of flipping two billion consecutive tails as being practically impossible.

Quality of the model aside, what we can take away from this is that the probability of flipping seven tails in a row in one hundred flips is at least greater than 10%, or one in ten.

2.2 Overestimation

Now let us look at an alternate approach. Let us think of the seven consecutive tails as a set event with 93 other flips that can be either heads or tails. So one estimate of the number of possibilities can be found by multiplying 94 (the number of different places the seven consecutive tails could be located: e.g. 1-7, 2-8, up to 94-100) by \(2^{93}\) (the number of possibilities the other 93 flips could be).

The resulting probability using this method of estimation then is \(\frac{94 \times 2^{93}}{2^{100}}\), or \(\frac{94}{2^7} \approx 73.44\%\). This, unlike the first method, is an overestimation. In fact, it turns out that this method is a pretty drastic overestimate. For instance, the single case where all one hundred flips are tails is counted 94 times (once for each place where you count the consecutive seven tails).

Let us again consider the general probability formula using this method. The probability of flipping seven consecutive tails in n flips is \(\frac{(n-6)2^{n-7}}{2^n}\), and the probability of flipping m consecutive tails in n flips using this method is \(\frac{(n-m+1)2^{n-m}}{2^m}\), or \(\frac{n-m+1}{2^m}\).

This model is not as nice as our previous one. If m and n are equal, we still get that the probability of flipping m consecutive tails is \(\frac{1}{2^m}\). However, consider what happens as the number of flips increases. If n is sufficiently large, we get that the probability of flipping m consecutive tails is greater than 100%. This is definitely a bad sign, and hints to the shoddiness of this model. So let us consider yet again another approach.

2.3 Brute Force

So now we know that the likelihood of flipping seven consecutive tails in one hundred flips is somewhere between 10.4% and 73.44%. Unfortunately, this is still quite a large interval, so we still have work to do. The first two approaches were pretty easy to calculate, but (as we will see later) not very accurate. What about instead trying a counting method that is accurate, but not necessarily efficient? Let us use the brute force method.

Starting with the situation of 100 coin flips, there are \(2^{100}\), or approximately \(1.27 \times 10^{30}\) cases to consider. Going through case by case is impractical, so let us simplify the problem. Consider the probability of flipping seven coins in a row...
in seven flips, and then finding the probability out of eight flips, nine flips, etc.? With seven flips, we have 128 possibilities, with only one of these possibilities being a successful one (T-T-T-T-T-T-T). Thus, the probability of flipping seven tails in a row in seven flips is 1 in 128.

Now, let us consider the probability of flipping seven tails in a row when flipping a coin eight times. Here, we now have 256 possibilities, with three successful possibilities:

\[
\begin{align*}
H-T-T-T-T-T-T & \quad T-T-T-T-T-T-T-H & \quad T-T-T-T-T-T-T-T \\
T-T-T-T-T-T-T & \quad T-T-T-T-T-T-T-H & \quad T-T-T-T-T-T-T-T \\
\end{align*}
\]

Therefore, the probability of flipping seven tails in a row out of eight flips is \(\frac{3}{256}\). Notice that this is slightly more likely than our first case which was \(\frac{1}{128}\), or \(\frac{2}{256}\). As we jump up to 9 coin flips, we can seeing that counting the possibilities becomes increasingly difficult, so we can partition our counting by the number of heads in the set. The results are:

\[
\begin{align*}
T-T-T-T-T-T-T & \quad H-T-T-T-T-T-T-T & \quad T-H-T-T-T-T-T-T & \quad T-T-T-T-T-T-T-T \\
T-T-T-T-T-T-T & \quad T-T-T-T-T-T-T-H & \quad T-T-T-T-T-T-T-H & \quad H-H-T-T-T-T-T-T & \quad H-T-T-T-T-T-T-H & \quad T-T-T-T-T-T-T-T \\
H-H-T-T-T-T-T-T & \quad H-T-T-T-T-T-T-H & \quad T-T-T-T-T-T-T-H & \quad T-T-T-T-T-T-T-H & \quad H-T-T-T-T-T-T & \quad T-T-T-T-T-T-T-T \\
\end{align*}
\]

So altogether, we have exactly eight cases where there are seven consecutive tail flips. Thus, the probability of flipping seven tails in a row out of nine flips is \(\frac{8}{512}\), again, slightly more probable than the previous case of eight flips. If we were to continue this method of counting, we would see the following:

\[
\begin{align*}
n = 7 : & \quad 1/128 \quad 0.78\% \\
n = 8 : & \quad 3/256 \quad 1.17\% \\
n = 9 : & \quad 8/512 \quad 1.56\% \\
n = 10 : & \quad 20/1024 \quad 1.95\% \\
n = 11 : & \quad 48/2048 \quad 2.34\% \\
\end{align*}
\]

This method soon becomes quite tedious and even by the time you get to the case of 15 flips (let alone 100 flips), this method becomes impractical.

### 2.4 Simplification of Brute Force

There is a pattern embedded in our previous results, though it may not be obvious. Let us try to simplify the problem again by trying to find the probability of getting two tails in a row instead of seven. Performing the same method of counting as before gives us the following results:

\[
\begin{align*}
n = 2 : & \quad 1/4 \\
n = 3 : & \quad 3/8 \\
n = 4 : & \quad 8/16 \\
n = 5 : & \quad 19/32 \\
n = 6 : & \quad 43/64 \\
\end{align*}
\]
These calculations (1, 3, 8, 19, 43, ...) can be found in the On-Line Encyclopedia of Integer Sequences under sequence A008466 as the number of tosses having a run of two or more [tails], (where a search for the previous sequence of numbers for seven consecutive tails [1, 3, 8, 20, 48] yields no results).

If we take a cursory glance at these results, it appears as if there is no discernible pattern to how the terms are generated. But if one looks at the difference between the numerator and denominator of each term, there’s a surprising result. As a new sequence, we would have:

\[(4 - 1), (8 - 3), (16 - 8), (32 - 19), (64 - 43), ...\] or 3, 5, 8, 13, 21, ...

The result is we get entries of the Fibonacci sequence! In fact, if we wanted to generate the probability of flipping two tails in a row in n flips, we would simply need to calculate \(1 - \frac{F_n}{2^n}\), where \(F_n\) is the n\(^{th}\) entry in the Fibonacci sequence (which is defined as \(F_n = F_{n-1} + F_{n-2}\), with \(F_0 = 0\) and \(F_1 = 1\)). But why does the Fibonacci sequence show up here? Why should that recursive sequence have any bearing on the probability of coin flipping? This seems particularly odd since each coin flip is an act independent from the others, while the Fibonacci sequence is something that builds upon itself.

The trick is that we were looking at the problem the wrong way around. We can gain more understanding by instead asking the question, What’s the probability that we will NOT flip two tails in a row? Take, for example, the situation of flipping a coin twice. What is the probability of NOT flipping two tails in a row? Here, we only have four possibilities (H-H, H-T, T-H, and T-T); three of which are cases where consecutive tail flips do not occur. Now, if we flipped a coin three times, and wanted to find all the cases where there were no two consecutive tails flipped, we could find all of these cases by extending our results from the previous case. In other words, if we look at any set of three flips that does not have two consecutive tails flipped, we know the first two of these flips would also not have two consecutive tails, and would therefore, fall into one of the three cases we found before.

This is where the recursive process comes in (and thus, where the Fibonacci sequence comes in). Consider the three cases before broken into two sets:

\[H_2 = \{H-H, T-H\} \text{ and } T_2 = \{H-T\}\]

Here, any element could have a heads flipped next and be an example of a string of three flips without a run of two tails (e.g. H-H-H, T-H-H, and H-T-H), but only the elements in \(H_2\) could have a tails flipped next and still not have two consecutive tails. So now, we would have:


Now we have some understanding that lets us go about proving this relationship.
Theorem 1 The probability of flipping 2 consecutive tails when flipping a fair coin \( n \) times is \( 1 - \frac{F_{n+2}}{2^n} \)

Proof

Here, \( F_{n+2} \) is the \( (n+2) \)th entry of the Fibonacci sequence, which is the recursive sequence where each term is the sum of the previous two and the first two terms are 0 and 1. The first ten terms of the Fibonacci sequence are: 0, 1, 2, 3, 5, 8, 13, 21, and 34. Now let us make the following definitions:

\[
A_n := \{\text{sequences of } n \text{ flips with no two consecutive tail flips}\} \\
H_n := \{\text{elements of } A_n \text{ whose last flip is a heads}\} \\
T_n := \{\text{elements of } A_n \text{ whose last flip is a tails}\}
\]

Clearly, we can see \( H_n \) and \( T_n \) partition \( A_n \) since a string of flips must either end in a heads or tails, so \( A_n = H_n \cup T_n \) with \( H_n \cap T_n = \emptyset \). Another way we could express this relationship is \( |A_n| = |H_n| + |T_n| \), where \( |A_n| \) means the number of elements in the set \( A_n \), \( |H_n| \) means the number of elements in \( H_n \), etc.

Now notice that any of the elements in \( A_n \) can have a heads flipped next, and then be an element of \( A_{n+1} \), or more specifically, \( H_{n+1} \). We can express this symbolically by saying: \( |H_{n+1}| = |A_n| \). Additionally, any element of \( H_n \) can have a tails flipped next and be an element of \( T_{n+1} \), so \( |T_{n+1}| = |H_n| \).

Using these identities and looking at \( |A_{n+2}| \), we get:

\[
|A_{n+2}| = |H_{n+2}| + |T_{n+2}| = |A_{n+1}| + |H_{n+1}| = |A_{n+1}| + |A_n|
\]

or simply

\[
|A_{n+2}| = |A_{n+1}| + |A_n|
\]

Additionally, we have the initial conditions that \( |A_1| = 2 \), \( |A_2| = 3 \) and \( |A_3| = 5 \); almost the same definition of the Fibonacci sequence, with only a difference in the indexing. Looking at the Fibonacci sequence we see that \( F_1 = 2 \), \( F_2 = 3 \) and \( F_3 = 5 \). So the number of strings of \( n \) flips with no two consecutive tail flips is \( |A_n| \), or the \( (n+2) \)th entry in the Fibonacci sequence. Therefore, the probability of NOT flipping two consecutive tails is \( \frac{F_{n+2}}{2^n} \). Thus, the probability of flipping two consecutive tails when flipping a fair coin \( n \) times is \( 1 - \frac{F_{n+2}}{2^n} \).

If we switch gears and look at the probability of getting three consecutive tails, or rather the probability of NOT getting three consecutive tails, we find again that there is an embedded recursive relationship as we go from the case of 3 flips to 4 flips to 5 flips and so on. This time, however, it mirrors the Tribonacci numbers, which is a sequence of numbers where each term is generated by summing the previous three numbers in the sequence (as opposed to the previous two with the Fibonacci sequence), with the first three terms being \( T_0 = 0 \), \( T_1 = 0 \), \( T_2 = 1 \).
The only difference in proving this is relationship from our previous proof is in the way we partition our sets. We would now partition all possible sets of n flips into three categories:

\[
A_n := \{\text{sequences of } n \text{ flips with no three consecutive tail flips}\}
\]
\[
H_n := \{\text{elements of } A_n \text{ whose last flip is a heads}\}
\]
\[
T_n := \{\text{elements of } A_n \text{ that end with exactly one tail flipped}\}
\]
\[
2T_n := \{\text{elements of } A_n \text{ that end with exactly two tails flipped}\}
\]

Similar to the earlier proof, substituting a set of equalities would yield the relationship \(|A_{n+3}| = |A_{n+2}| + |A_{n+1}| + |A_n|\), with \(|A_1| = 2, |A_2| = 4\) and \(|A_3| = 7\). We'll omit the details of this proof, and instead provide a rigorous look at the cases of flipping seven tails in a row.
3 Specific Solution

So now let try to answer our initial, specific question: What is the probability of flipping seven tails in a row in 100 flips?

**Theorem 2** The probability of flipping 7 consecutive tails when flipping a fair coin 100 times is \(1 - \frac{A_{100}}{2^{100}}\)

Here, \(A_n\) is the \(n^{th}\) term in the recursive sequence defined by the following characteristics:

\[|A_{n+7}| = |A_{n+6}| + |A_{n+5}| + |A_{n+4}| + |A_{n+3}| + |A_{n+2}| + |A_{n+1}| + |A_n|\]

and

\[|A_1| = 2^1, |A_2| = 2^2, |A_3| = 2^3, ..., |A_6| = 2^6, |A_7| = 127\]

**Proof**

First let us begin by partitioning all the possibilities of \(n\) flips:

- \(A_n := \{\text{sequences of } n \text{ flips with no seven consecutive tail flips }\}\)
- \(H_n := \{\text{elements of } A_n \text{ whose last flip is a heads}\}\)
- \(T_n := \{\text{elements of } A_n \text{ that end with exactly one tail flipped}\}\)
- \(2T_n := \{\text{elements of } A_n \text{ that end with exactly two tails flipped}\}\)
- \(3T_n := \{\text{elements of } A_n \text{ that end with exactly three tails flipped}\}\)
- \(4T_n := \{\text{elements of } A_n \text{ that end with exactly four tails flipped}\}\)
- \(5T_n := \{\text{elements of } A_n \text{ that end with exactly five tails flipped}\}\)
- \(6T_n := \{\text{elements of } A_n \text{ that end with exactly six tails flipped}\}\)

Now let us try to define some relationships as we did previously. For instance, for any given sequence of flips in \(A_n\), a heads could be flipped next and it would continue to not have seven consecutive tails. So by adding an H to the end of each string in \(A_n\), we would have constructed an element of \(H_{n+1}\). So a way to express this fact is \(|A_n| = |H_{n+1}|\).

Now let us try a similar approach, but with adding a tails flip at the end of each sequence instead of a heads flip. It should become immediately apparent that we cannot add a tails to any of the strings in \(6T_n\), since it would create a sequence of flips with a run of seven tails (and thus, not be an element in \(A_{n+1}\)).

So let us consider what happens when a tails is added to any of elements of the other sets. Any element of \(H_n\) would become an element of \(T_{n+1}\) when adding a tails flip. Any element of \(T_n\) would become an element of \(2T_{n+1}\), and so on. From these two observations, we can produce the following equalities:
consecutive tails, so every possibility would be an element of seven consecutive terms to be able to generate the sequence. If the number possibility of seven tails in a row out of 128, so

\[ |H_{n+1}| = |A_n| \]
\[ |T_{n+1}| = |H_n| \]
\[ |2T_{n+1}| = |T_n| \]
\[ |3T_{n+1}| = |2T_n| \]
\[ |4T_{n+1}| = |3T_n| \]
\[ |5T_{n+1}| = |4T_n| \]
\[ |6T_{n+1}| = |5T_n| \]

This will give us enough information for us to define a recursive relationship for this process. Let us start by looking at \( A_{n+7} \):

\[
|A_{n+7}| = |H_{n+7}| + |T_{n+7}| + |2T_{n+7}| + |3T_{n+7}| + |4T_{n+7}| + |5T_{n+7}| + |6T_{n+7}|
\]
\[
= |A_{n+6}| + |H_{n+6}| + |T_{n+6}| + |2T_{n+6}| + |3T_{n+6}| + |4T_{n+6}| + |5T_{n+6}|
\]
\[
= |A_{n+6}| + |A_{n+5}| + |H_{n+5}| + |T_{n+5}| + |2T_{n+5}| + |3T_{n+5}| + |4T_{n+5}|
\]
\[
= |A_{n+6}| + |A_{n+5}| + |A_{n+4}| + |H_{n+4}| + |T_{n+4}| + |2T_{n+4}| + |3T_{n+4}|
\]
\[
= |A_{n+6}| + |A_{n+5}| + |A_{n+4}| + |A_{n+3}| + |H_{n+3}| + |T_{n+3}| + |2T_{n+3}|
\]
\[
= |A_{n+6}| + |A_{n+5}| + |A_{n+4}| + |A_{n+3}| + |A_{n+2}| + |H_{n+2}| + |T_{n+2}|
\]
\[
= |A_{n+6}| + |A_{n+5}| + |A_{n+4}| + |A_{n+3}| + |A_{n+2}| + |A_{n+1}| + |H_{n+1}|
\]
\[
|A_{n+7}| = |A_{n+6}| + |A_{n+5}| + |A_{n+4}| + |A_{n+3}| + |A_{n+2}| + |A_{n+1}| + |A_n|
\]

This result shouldn’t seem too surprising. It depends on essentially the same argument as the Fibonacci result we saw previously. Here’s another reason why this recursive relationship should make sense: If we were concerned with not making a run of seven tails, which flips should we be most interested in? Assuming that the sequence already does not have a run of seven tails, then we would want to know its previous seven flips. This would tell us how close or far away we are from having a run of seven tails. We wouldn’t be concerned with whether it was heads or tails eighteen flips ago. Likewise, the recursive formula derived only needs information from the last seven flips in order to generate the next terms in the sequence.

Now that we have determined a recursive equation for \( |A_n| \), we need to know seven consecutive terms to be able to generate the sequence. If the number of flips is less than seven, then every string of flips would fail to have seven consecutive tails, so every possibility would be an element of \( A_n \), so \( |A_1| = 2^1 \), \( |A_2| = 2^2 \), ..., \( |A_6| = 2^6 \). When there are seven flips exactly, there is only one possibility of seven tails in a row out of 128, so \( |A_7| = 127 \). This gives us enough initial data to generate the following sequence of values for \( |A_n| \):

2, 4, 8, 16, 32, 64, 127, 253, 504, 1004, 2000, 3984, 7936, 15808, ...
This is actually a known sequence. These numbers are called Heptanacci numbers. It is sequence A122189 on the On-Line Encyclopedia of Integer Sequences. So, again, this tells us the number of cases where there are not seven consecutive tails flipped. So for instance, if we wanted to know the probability of flipping seven consecutive tails when flipping a fair coin ten times, we would first look for the tenth term in this sequence: 1004. Then, to find the probability of not flipping seven tails in a row, we would divide this number by the total number of possibilities, or 1024, to get 1004/1024, or $\frac{251}{256} \approx 98.05\%$. Lastly, to find the probability of flipping seven consecutive tails, we simply need to look at the complement of our previous term, or $\frac{1024-1004}{1024} = \frac{20}{1024} = \frac{5}{256} \approx 1.95\%$.

Since this is a known sequence, the first 200 terms of the Heptanacci numbers can be found with little difficulty. That way, we can perform this same process for the case of 100 flips to find the answer to our initial question. Again, as before, the only difference between our sequence and the sequence of the Heptanacci numbers is the indexing. The Heptanacci numbers has 2 as its second term, as opposed to the first. So when looking for the 100th term in our sequence, were looking for the 101st term in the Heptanacci numbers, or 865,145,690,433,457,063,670,671,045,568. Thats over 865 octillion cases of a hundred flips of coins that do not have seven tails in a row, which may seem like an extremely high number. Fortunately for us, its out of $2^{100}$ possibilities, which yields a nice comparison. So the probability of flipping seven tails in a row when flipping a fair coin one hundred times is:

$$1 - \frac{H_{101}}{2^{100}} = 1 - \frac{865145690433457063670671045568}{2^{100}} \approx 31.75\%$$

For the sake of being precise, and answering our initial question completely, the exact probability is:

31.7520387496605807487237045121323554887839703005
7537699099245998013429925777018070220947265625%

But for all intents and purposes the probability of flipping seven tails in a row is about a 1 in 3 chance. Lets compare this to our predictions at the beginning of the paper. This percentage is twice as likely as what the people on Radiolab approximated (1 in 6 probability). It is three times as likely as our under estimate (10.4%) and under half as much as our overestimate (73.44%).
4 Extended Answer

Now it may seem like we should be finished with this question, but there is still a point I'd like to bring up. Thinking back to the context of the question and how it originated, the hosts of Radiolab wanted to find out the probability because they were surprised when they flipped seven tails in a row. The source of their surprise was not from the fact that it was tails, but rather that there was a run of seven of them. The two hosts, Jad and Robert, would have been just as surprised if there were seven heads in a row instead. This brings up an interesting extension to original question:

What is the probability of flipping seven consecutive heads OR seven consecutive tails when flipping a fair coin one hundred times?

We can again come up with a range for this particular probability, but this time using the results we just found. Since the coin we’re flipping is assumed to be fair, the probability of flipping seven tails in a row out of one hundred flips is exactly the same as the probability of flipping seven heads in a row; both being about 31.75%. So we have already know the answer to new question lies somewhere between 31.75% (the case where runs of heads and runs of tails only occur together) and 63.5% (the case where runs of heads and runs of tails never occur together). To find a more precise solution, we will again turn to recursive relationships. This time, we will be keeping track of the number of heads at the end of flips as well as the number of tails. So let us introduce the following notation for $A_n$ and a corresponding partition:

$$A_n := \{ \text{sequences of n flips with no seven consecutive tail flips} \}$$

$$H_n := \{ \text{elements of } A_n \text{ that end with exactly one head flipped} \}$$

$$2H_n := \{ \text{elements of } A_n \text{ that end with exactly two heads flipped} \}$$

$$3H_n := \{ \text{elements of } A_n \text{ that end with exactly three heads flipped} \}$$

$$4H_n := \{ \text{elements of } A_n \text{ that end with exactly four heads flipped} \}$$

$$5H_n := \{ \text{elements of } A_n \text{ that end with exactly five heads flipped} \}$$

$$6H_n := \{ \text{elements of } A_n \text{ that end with exactly six heads flipped} \}$$

$$T_n := \{ \text{elements of } A_n \text{ that end with exactly one tail flipped} \}$$

$$2T_n := \{ \text{elements of } A_n \text{ that end with exactly two tails flipped} \}$$

$$3T_n := \{ \text{elements of } A_n \text{ that end with exactly three tails flipped} \}$$

$$4T_n := \{ \text{elements of } A_n \text{ that end with exactly four tails flipped} \}$$

$$5T_n := \{ \text{elements of } A_n \text{ that end with exactly five tails flipped} \}$$

$$6T_n := \{ \text{elements of } A_n \text{ that end with exactly six tails flipped} \}$$

Now again, we need to identify some relationships among these categories,
so we can form some sort of recursive relationship. Let us start by looking at
the set $6H_{n+1}$. The only elements in this set must have been elements in $5H_n$
if you take off the last flip. We can quantify this by saying $|6H_{n+1}| = |5H_n|$. Similarly, we have $|6T_{n+1}| = |5T_n|$. Using the same rationale, we can determine the following:

$$
|6H_{n+1}| = |5H_n|, \quad |6T_{n+1}| = |5T_n|
$$

$$
|5H_{n+1}| = |4H_n|, \quad |5T_{n+1}| = |4T_n|
$$

$$
|4H_{n+1}| = |3H_n|, \quad |4T_{n+1}| = |3T_n|
$$

$$
|3H_{n+1}| = |2H_n|, \quad |3T_{n+1}| = |2T_n|
$$

$$
|2H_{n+1}| = |H_n|, \quad |2T_{n+1}| = |T_n|
$$

$$
|H_{n+1}| = |6T_n| + |5T_n| + |4T_n| + |3T_n| + |2T_n| + |T_n|
$$

$$
|T_{n+1}| = |6H_n| + |5H_n| + |4H_n| + |3H_n| + |2H_n| + |H_n|
$$

Now notice that $|A_n| = |H_{n+1}| + |T_{n+1}|$ since $A_n$ is partitioned by $H_n, 2H_n, \ldots, 5T_n, 6T_n$. Using this equality and the ones we found previously, we can now derive the following relationship:

$$
|A_{n+6}| = (|H_{n+6}| + |T_{n+6}|) + (|2H_{n+6}| + |2T_{n+6}|) + \ldots + (|6H_{n+6}| + |6T_{n+6}|)
$$

$$
= |A_{n+5}| + (|H_{n+5}| + |T_{n+5}|) + \ldots + (|5H_{n+5}| + |5T_{n+5}|)
$$

$$
= |A_{n+5}| + |A_{n+4}| + \ldots + (|4H_{n+4}| + |4T_{n+4}|)
$$

.$$=

$$
|A_{n+6}| = |A_{n+5}| + |A_{n+4}| + |A_{n+3}| + |A_{n+2}| + |A_{n+1}| + |A_n|
$$

Curiously, we’ve obtained the same relationship for Hexanacci numbers: the sequence we would have found and used if we were answering the question, What is the probability of flipping six tails in a row when flipping a fair coin $n$ times? This time, however, we have slightly different initial conditions. When the number of flips is less than seven, every possibility is an element of $A_n$, so $A_1 = 2^1$, $A_2 = 2^2$, $\ldots$, $A_6 = 2^6$. When there are exactly seven flips, there are two possibilities of sequences with a run of seven heads or tails, namely H-H-H-H-H-H, and T-T-T-T-T-T-T. Since there are 128 total possibilities, that leaves 126 different sequences of flips that have no run of seven heads or tails in a row, so $A_7 = 126$.

This gives us enough information to start calculating, and so on. The sequence we calculate is:

$$
2, 4, 8, 16, 32, 64, 126, 250, 496, 984, \ldots
$$

Interestingly enough, this is just the Hexanacci numbers multiplied by two (and with a slightly different indexing). This is because the initial conditions
we have are precisely twice the value of the initial conditions of the Hexanacci numbers. So again, we’re in luck. Since the Hexanacci numbers are a well known sequence, it’s easy to find the first couple hundred terms of the sequence. Using this information, we can find that

\[ A_{100} = 580156959829221175647260088196 \]

So to find the probability of flipping either seven heads or seven tails in a row, we’d need to calculate \( 1 - \frac{A_{100}}{2^{100}} \). Computing this yields about 54.23%. So in a surprising turn of events, flipping seven heads or tails in a row, an event that might seem rare or magical, is actually more likely to happen than not when flipping a coin one hundred times.
5 Calculation Techniques

In solving the initial problem, there are some additional techniques learned that weren’t explicitly stated in the solution. For instance, there were several points in the paper where we had to find very large terms in a sequence. Initially, the way I found these terms was to simply find a list of the sequences online. This was relatively easy since I was working with well-known sequences. But what would happen if I was trying to find terms of more obscure sequences? There are plenty of ways to compute this using computer programs. One way I found which was particularly interesting was using matrices. So for example, say we wanted to compute high terms of the Fibonacci sequence. We should consider the matrix

\[
A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}
\]

Watch what happens as we raise \( A \) to higher and higher powers:

\[
A^2 = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}, \quad A^3 = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}, \quad A^4 = \begin{bmatrix} 2 & 3 \\ 3 & 5 \end{bmatrix}, \quad A^5 = \begin{bmatrix} 3 & 5 \\ 5 & 8 \end{bmatrix}
\]

\[
A^6 = \begin{bmatrix} 5 & 8 \\ 8 & 13 \end{bmatrix}, \quad A^7 = \begin{bmatrix} 8 & 13 \\ 13 & 21 \end{bmatrix}, \quad A^8 = \begin{bmatrix} 13 & 21 \\ 21 & 34 \end{bmatrix}, \quad A^9 = \begin{bmatrix} 21 & 34 \\ 34 & 55 \end{bmatrix}
\]

As we can already see, by looking at entries in each successive \( A^n \), we find the Fibonacci sequence. What’s even better is the indexing matches up matches exactly so that in \( A^n \), the \( a_{2,1} \) entry is the \( n^{th} \) term in the Fibonacci sequence. Using computer software, we can instantly calculate \( A^n \) and display the \( a_{2,1} \) entry. Thus we can find any specific term of this sequence with ease. What if we wanted to calculate terms of the Tribonacci sequence? Here, the matrix is

\[
A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}
\]

We could then calculate the \( n^{th} \) term of the Tribonacci sequence by looking at the \( a_{3,1} \) entry of \( A_n \). But we needn’t stop here. Say we wanted to calculate a term in the recursive relationship:

\[
A_m = 5A_{m-1} - 3A_{m-2}, \text{ with the initial conditions } A_1 = 1, A_2 = 2
\]

We could use a similar technique as before to find the term. This time, we’d have

\[
A = \begin{bmatrix} 0 & 1 \\ -3 & 1 \\ 5 \end{bmatrix}
\]

It may be difficult to notice the pattern on a 2 by 2 matrix, but we keep 1s on the superdiagonal and change the last row in accordance with our
recursion: the coefficient of \( A_{m-1} \) should be the \( a_{m,2} \) entry and the coefficient of \( A_{m-2} \) should be the \( a_{m,1} \) entry. To take into account the initial conditions (something we didn’t have to do with the Fibonacci/Tribonacci numbers), we multiply \( A^n \) on the right by the vector \( \begin{bmatrix} 1 \\ 2 \end{bmatrix} \), and in this case, we would get the \((n + 2)^{th}\) term of our recursion in the second entry of the vector.

6 Generalized Answer

Now we’ve amassed enough information to provide a general solution to our initial question:

What is the probability of flipping \( m \) consecutive tails when flipping a fair coin \( n \) times?

The \( m \) in this case will correspond to recursive sequence where each term is generated by summing the previous \( m \) terms with the initial conditions \( A_1 = 2^1 \), \( A_2 = 2^2 \), \( A_3 = 2^3 \), ..., \( A_{m-1} = 2^{m-1} \) and \( A_m = 2^m - 1 \).

Then the number of strings that do not have \( m \) consecutive tails will be the \( n^{th} \) term of this recursive sequence. So the probability of flipping \( m \) consecutive tails is \( 1 - \frac{a_m}{2^n} \). Finally, we can calculate \( A_n \) efficiently by looking at the \( a_{m,1} \) entry in \( n^{th} \) power of the \( m \) by \( m \) matrix with 1s on the superdiagonal, 1s on the last row, and 0s everywhere else. Given a specific \( m \) and \( n \), this relationship can be proven following a similar proof as Theorem 1 or Theorem 2.
Let us now focus our attention to the \( m \times m \) matrices (which I'll refer to as Nacci matrices) that produce the Nacci sequences we've seen thus far. Each one of these matrices corresponds to a recursive sequence, and each one of these recursive sequences corresponds to the number of cases where consecutive streaks do not appear. So it would benefit us to know some of the properties of these Nacci matrices, since it would relate information to us about the recursive sequences we used in answering the general question.

Let us start by calculating the determinant of a Nacci matrix. To start with a simple example, let's look at the Nacci matrix

\[
A_2 := \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}.
\]

Here, \( \det(A_2 - \lambda I) \) is \( \lambda^2 - \lambda - 1 \). Next, looking at

\[
A_3 := \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix},
\]

and calculating the determinant of \( \det(A_3 - \lambda I) \), we find it's \((-\lambda)\det(A_2 - \lambda I) - (1)(-1) = -\lambda^3 + \lambda^2 + \lambda + 1\). As you may be able to tell, there's again a pattern here. We can prove this inductively:

**Lemma 3** For a Nacci matrix, \( A_m \), \( \det(A_m - \lambda I) = (-1)^m(\lambda^m - \sum_{j=0}^{m-1} \lambda^j) \)

**Proof** We've already shown this is true for \( m = 2 \). So let us assume it holds true for \( m = k \). Now let us consider the case of \( m = k+1 \).

\[
\det(A_{k+1} - \lambda I) = (1)(-\lambda)(\det(A_k - \lambda I) + (1)(1)\det(B_k)
\]

where \( B_k \) is the \( m \times m \) matrix obtained when deleting the first row and second column of \( A_{k+1} \). An important fact about this new matrix \( B_k \) is that \( \det(B_k) = (-1)\det(B_{k-1}) \), where \( B_{k-1} \) is the matrix obtained by now deleting the 1st row and 2nd column of \( B_k \). Repeating this process over and over eventually wittles \( B_k \) down to

\[
\begin{bmatrix} 0 & 1 \\ 1 & 1 - \lambda \end{bmatrix},
\]

whose determinant is -1. So

\[
\begin{align*}
\det(A_{k+1} - \lambda I) &= (\lambda)((-1)^k \lambda^k + (-1)^{k+1} \sum_{j=0}^{k-1} \lambda^j) + (-1)\det(B_k) \\
&= (-1)^{k+1}(\lambda^{k+1}) + (-1)^{k+2} \sum_{j=1}^{k} \lambda^{j+1} + (-1)(-1)^{k-2}\det\left(\begin{bmatrix} 0 & 1 \\ 1 & 1 - \lambda \end{bmatrix}\right) \\
&= (-1)^{k+1}(\lambda^{k+1}) + (-1)^{k+2} \sum_{j=0}^{k} \lambda^{j}
\end{align*}
\] (1) (2) (3)
Therefore, by induction, this claim is true for all Nacci matrices where \( m \in \mathbb{N} \).

Now that we’ve proven the claim, we should then turn our attention to solving the equation:

\[
f(x) = (-1)^m x^m + (-1)^{m+1} \sum_{j=0}^{m-1} x^j = (-1)^{m} (x^m - x^{m-1} - x^{m-2} - \ldots - x^2 - x - 1) = 0
\]

Solving this equation will give us the eigenvalues of the Nacci matrix of dimension \( m \times m \). First note that \( f(2) = (-1)^m (2^m - 2^{m-1} - \ldots - 4 - 2 - 1) = (-1)^m (1) = (-1)^m \), no matter what our choice of \( m \) is. It’s reasonable to expect then a root would be near \( x = 2 \), and leads us to the following claim.

**Lemma 4** Let \( \lambda \) be an eigenvalue (real or complex) of a Nacci matrix. Then \( ||\lambda|| \leq 2 \).

**Proof** Assume to the contrary that \( \lambda \) is a real eigenvalue such that \( |\lambda| \geq 2 \). Since \( \lambda \) is an eigenvalue, it is a root for the equation

\[
f(\lambda) = (-1)^m (\lambda^m - \lambda^{m-1} - \lambda^{m-2} - \ldots - \lambda^2 - \lambda - 1) = 0 \rightarrow (\lambda^m - \lambda^{m-1} - \lambda^{m-2} - \ldots - \lambda^2 - \lambda - 1) = 0
\]

But if \( \lambda \geq 2 \), then

\[
f(\lambda) = \lambda^m - \lambda^{m-1} - \lambda^{m-2} - \ldots - \lambda^2 - \lambda - 1 \geq \lambda^{m-1} - \lambda^{m-2} - \ldots - \lambda^2 - \lambda - 1 \geq \lambda^{m-2} - \lambda^{m-3} - \ldots - \lambda^2 - \lambda - 1 \geq \lambda^{m-3} - \lambda^{m-4} - \ldots - \lambda^2 - \lambda - 1 \leq \lambda - 1 \geq 1
\]

Therefore, \( f(\lambda) \geq 1 \), so \( \lambda \) is not an eigenvalue for the \( m \) by \( m \) Nacci matrix.

Similarly, if \( \lambda \leq 2 \), then \( \lambda = \mu \) for some \( \mu \geq 2 \). If \( m \) is even, then we get

\[
f(\lambda) = f(-\mu) = \mu^m + \mu^{m-1} - \mu^{m-2} + \mu^{m-3} - \ldots - (-1)^m (\mu^3 - \mu^2 + \mu - 1)
\]

By similar reasoning, if \( m \) is odd, we get that \( f(\lambda) < -f(\mu) = -1 \). Either way, \( f(\lambda) \neq 0 \), which is a contradiction. So if \( \lambda \) is real, we have that \( |\lambda| < 2 \).

Let us consider that last possibility: \( \lambda \) is a complex number. Assume to the contrary that \( \lambda \) is an eigenvalue such that \( |\lambda| \geq 2 \). Then we could write \( \lambda = ||\lambda|| (\cos \theta + i \sin \theta) \) for some \( \theta \) and consider \( f(\lambda) \). Following a similar argument as our first case, we could show that \( ||f(\lambda)|| \leq 1 \), which implies that \( f(\lambda) \neq 0 \), and therefore, \( \lambda \) is not an eigenvalue of the \( m \) by \( m \) Nacci matrix.
Thus, we have that the norm of any eigenvalue of a Nacci matrix is bounded by 2.

It turns out that we can do slightly better than this even. Remember from before how \( |f(2)| = 1 \) no matter what \( m \) we consider. This hints to the fact that there is an eigenvalue close to two. Let us remember that an eigenvalue of the Nacci matrix \( A_m \) is a root of the polynomial \( \det(A_m - \lambda I) \), or

\[
f_m(x) = (-1)^m(x^m - (x^{m-1} + x^{m-2} + \cdots + x^2 + x + 1)) = (-1)^m(x^m - \frac{x^m - 1}{x - 1})
\]

We can use Newtons method of approximating roots to then get a better approximation of one of the eigenvalues of \( A_m \). To do this, we need to calculate the first derivative:

\[
f'(x) = (-1)^m(mx^{m-1} - ((mx^{m-1})(x - 1) - \frac{x^m - 1}{(x - 1)^2}) (9)
\]

\[
= (-1)^m\frac{(mx^{m-1})(x - 1)((x - 1) - 1) + (x^m - 1)}{(x - 1)^2} (10)
\]

\[
= (-1)^m\frac{(mx^{m-1})(x - 1)(x - 2) + x^m - 1}{(x - 1)^2} (11)
\]

This may look like a very complicated term, but when evaluated at 2, we get:

\[
f'(2) = (-1)^m(2^m - 1)
\]

Then computing the first approximation of the root using Newtons method, we get:

\[
x_m = 2 - \frac{f_m(2)}{f'_m(2)} = 2 - \frac{(-1)^m}{(-1)^m(2^m - 1)} = 2 - \frac{1}{2^m - 1}
\]

Let us gain a better understanding of what this means. Consider the case of \( m = 2 \). Here, the actual eigenvalue is the golden ratio (\( \approx 1.602 \)), while the upper bound we obtain is \( 2 - \frac{1}{3} = 2 - \frac{1}{\frac{1}{2}} = 2 - \frac{1}{\frac{1}{3}} = 2 - \frac{3}{1} = 2 - 1 = 1.66666... \)

Here is a table comparing the actual roots with our estimated roots for \( f_2(x) \) through \( f_{10}(x) \):

<table>
<thead>
<tr>
<th>( m )</th>
<th>Actual Root</th>
<th>Approximated Root</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1.61803...</td>
<td>1.66666...</td>
</tr>
<tr>
<td>3</td>
<td>1.83929...</td>
<td>1.85714...</td>
</tr>
<tr>
<td>4</td>
<td>1.92756...</td>
<td>1.93333...</td>
</tr>
<tr>
<td>5</td>
<td>1.96595...</td>
<td>1.96774...</td>
</tr>
<tr>
<td>6</td>
<td>1.98358...</td>
<td>1.98412...</td>
</tr>
<tr>
<td>7</td>
<td>1.99196...</td>
<td>1.99212...</td>
</tr>
<tr>
<td>8</td>
<td>1.99603...</td>
<td>1.99607...</td>
</tr>
<tr>
<td>9</td>
<td>1.99803...</td>
<td>1.99804...</td>
</tr>
<tr>
<td>10</td>
<td>1.99901...</td>
<td>1.99902...</td>
</tr>
</tbody>
</table>
One value of knowing these eigenvalues is that it can help us estimate how big the terms in our recursive sequences are getting. So for example, consider the terms of the Fibonacci sequence: $F_{10} = 55 \approx 55.013 \ldots = (1.618 \ldots)(34) = \lambda F_9$. In general, the limit of the ratio of consecutive terms in the Nacci sequences approach the largest eigenvalue of the corresponding Nacci matrix. The proof of this is quite difficult since it rests on being able to calculate the closed form of our recursive sequences, which involves calculating complex roots. We’ll end this paper by providing a claim and proof of this relationship for a simpler case: the Fibonacci sequence.

Lemma 5 $\lim_{n \to \infty} \frac{F_{n+1}}{F_n} = \frac{1+\sqrt{5}}{2} = \varphi$

Proof

As previously mentioned, the proof of this rests on finding the closed form of a recursive sequence. For our case, the closed form of the Fibonacci numbers is:

$$F_n = \frac{1}{\sqrt{5}} \left( \left( \frac{1 + \sqrt{5}}{2} \right)^n - \left( \frac{1 - \sqrt{5}}{2} \right)^n \right)$$

So then

$$\frac{F_{n+1}}{F_n} = \frac{1}{\sqrt{5}} \frac{\left( \frac{1 + \sqrt{5}}{2} \right)^{n+1} - \left( \frac{1 - \sqrt{5}}{2} \right)^{n+1}}{\left( \frac{1 + \sqrt{5}}{2} \right)^n - \left( \frac{1 - \sqrt{5}}{2} \right)^n} = \frac{\left( \frac{1 + \sqrt{5}}{2} \right)^{n+1}}{\left( \frac{1 + \sqrt{5}}{2} \right)^n - \left( \frac{1 - \sqrt{5}}{2} \right)^n} - \frac{\left( \frac{1 - \sqrt{5}}{2} \right)^{n+1}}{\left( \frac{1 + \sqrt{5}}{2} \right)^n - \left( \frac{1 - \sqrt{5}}{2} \right)^n}$$

Notice that $\left| \frac{1 - \sqrt{5}}{2} \right| < 1$, so this term will go towards 0 as $n \to \infty$. That means:

$$\lim_{n \to \infty} \frac{F_{n+1}}{F_n} = \lim_{n \to \infty} \frac{\left( \frac{1 + \sqrt{5}}{2} \right)^{n+1}}{\left( \frac{1 + \sqrt{5}}{2} \right)^n} = \frac{1 + \sqrt{5}}{2} = \varphi$$
References
