Wrong Proof

Feb 14, 2024 18.200 Recitation — Lecturers: Ankur Moitra & Peter Shor

---

Theorem 1. *Suppose that we flip a fair coin n times, for $n \geq 4$. Consider the following two events:*

- *A: the first and last coin flips are the same,*
- *B: there is at least one contiguous subsequence of HTH or THT (here* contiguous subsequence *means these are flips i, $i+1$, and $i+2$ for some i).*

Then the probability that both event A and event B happen is

$$\mathbb{P}(A \wedge B) = \frac{1}{2} - \frac{f_n}{2^n}. \tag{1}$$

Here f_n is the nth Fibonacci number, defined by $f_{i+1} = f_i + f_{i-1}$ for $i \geq 1$ and $f_0 = f_1 = 1$.

For example, for $n = 5$, there are 2^5 possible sequences of 5 coin flips, of which the following 8 are in both A and B:

HHHTH	HHTHH	HTHHH	HTHTH
TTTHT	TTHTT	THTTT	THTHT.

Thus, for $n = 5$, the probability that both event A and event B happen is $\frac{8}{2^5} = \frac{1}{4}$. This result is the same as $\frac{1}{2} - \frac{f_n}{2^n}$, since $f_5 = 8$.

To prove the theorem, we calculate $\mathbb{P}(A)$ and $\mathbb{P}(B)$ and multiply the two together. To justify this multiplication, we first show in Lemma 1 that events A and B are independent.

Lemma 1. *For $n \geq 4$, the events A and B given in the theorem are independent.*

Proof. Consider a sequence of n coin flips in B. The subsequence HTH or THT is of length 3, and so cannot overlap both ends of the n coin flips, since there are at least four coin flips. Thus, either the first coin or the last coin is not contained in this sequence. This coin has equal probability of being heads or tails, and thus has a probability $\frac{1}{2}$ of matching the other end of the string of n coin flips. This shows that

$$\mathbb{P}(A|B) = \frac{1}{2}.$$

Since $\mathbb{P}(A) = \frac{1}{2}$, we have that $\mathbb{P}(A|B) = \mathbb{P}(A)$, proving that the two events are independent. $\square$

Remark. The proof of Lemma 1 could instead have used similar reasoning to show that $\mathbb{P}(B|A) = \mathbb{P}(B)$: because $n \geq 4$, whether the ends of the sequence are the same or different has no effect on subsequences of length 3. Thus A and B are independent.

The next lemma helps us to determine $\mathbb{P}(B)$, the probability that there is at least one contiguous subsequence of HTH or THT. Instead of determining $\mathbb{P}(B)$ directly, we consider B's complement: that is, the set of sequences of length n that contain neither HTH nor THT. For example, for $n = 4$, these sequences are

$$\begin{array}{ccccc} \text{HHHH} & \text{HHHT} & \text{HHTT} & \text{HTTH} & \text{HTTT} \\ \text{TTTT} & \text{TTTH} & \text{TTHH} & \text{THHT} & \text{THHH.} \end{array}$$

For the proof of the lemma, we'll call this set of sequences S_n.

Lemma 2. *The number of sequences of length n that contain neither HTH nor THT is $2f_n$.*

Proof. Let S_n be the set of sequences of length n that contain neither HTH nor THT. We proceed by induction on n.

The base case is straightforward: for $n = 1$, there are two sequences of length 1, neither of which contains a forbidden sequence. Thus,

$$|S_1| = 2 = 2f_1.$$

Now, suppose the lemma holds for all S_i where $1 \leq i < n$. We will prove it holds for S_n. We show this by giving a bijection between S_n and $S_{n-1} \cup S_{n-2}$, as illustrated in Figure 1. Once we have a bijection, we will know that

$$|S_n| = |S_{n-1}| + |S_{n-2}|,$$

so we can apply the inductive hypothesis to obtain

$$|S_n| = 2f_{n-1} + 2f_{n-2}.$$

The definition of the Fibonacci numbers then yields $|S_n| = 2f_n$, as desired.

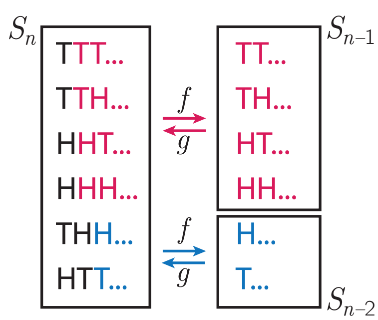

Figure 1: A bijection between S_n and $S_{n-1} \cup S_{n-2}$. By definition, S_n has no sequences beginning with THT or HTH.

To create the bijection, first we specify the map f from S_n to $S_{n-1} \cup S_{n-2}$. If $x \in S_n$, then

$$f(x) = \begin{cases} x \text{ with the first coin flip removed} & \text{if } x \text{ begins with TT or HH,} \\ x \text{ with the first two coin flips removed} & \text{if } x \text{ begins with TH or HT.} \end{cases}$$

We first verify that f maps S_n to $S_{n-1} \cup S_{n-2}$. The map f is valid for all elements of S_n because x must begin with one of HH, TT, TH, or HT. Furthermore, $f(x) \in S_{n-1} \cup S_{n-2}$ because, since x did not contain a contiguous subsequence HTH or THT, neither does $f(x)$.

We need to show that f gives us a bijection. To do so, we construct the inverse $g = f^{-1}$, which maps $S_{n-1} \cup S_{n-2}$ to S_n. If $x \in S_{n-1} \cup S_{n-2}$, then

$$g(x) = \begin{cases} \mathrm{T}x & \text{if } x \in S_{n-1} \text{ and } x \text{ begins with T,} \\ \mathrm{H}x & \text{if } x \in S_{n-1} \text{ and } x \text{ begins with H,} \\ \mathrm{TH}x & \text{if } x \in S_{n-2} \text{ and } x \text{ begins with H,} \\ \mathrm{HT}x & \text{if } x \in S_{n-2} \text{ and } x \text{ begins with T.} \end{cases}$$

The function g maps $S_{n-1} \cup S_{n-2}$ to S_n because it is valid for all elements of S_{n-1} and S_{n-2} and it creates a sequence of length n that does not contain HTH or THT. Furthermore, g is the inverse of f because the coin flips added by g are exactly those removed by f, so $f(g(x)) = x$ and $g(f(x)) = x$.

Because f and g provide a bijection between S_n and $S_{n-1} \cup S_{n-2}$, we can conclude that

$$|S_n| = 2f_{n-1} + 2f_{n-2} = 2f_n,$$

thus completing the inductive proof. □

We now have all the ingredients for a proof of the theorem.

Proof of Theorem 1. It is straightforward to see that

$$\mathbb{P}(A) = \frac{1}{2}.$$

Lemma 2 shows that the probability of $\bar{B}$, the complement of B, is $\mathbb{P}(\bar{B}) = \frac{2f_n}{2^n}$, so

$$\mathbb{P}(B) = 1 - \mathbb{P}(\bar{B}) = 1 - \frac{2f_n}{2^n}.$$

In Lemma 1, we showed that A and B were independent, so

$$\mathbb{P}(A \wedge B) = \mathbb{P}(A)\mathbb{P}(B) = \frac{1}{2}\big(1 - \frac{2f_n}{2^n}\big) = \frac{1}{2} - \frac{f_n}{2^n}.$$

□