Coin Flips, Fibonacci Numbers and Gaps!

Steven Miller and Philip Tosteson

Joint with: Amanda Bower, Rachel Insoft, and Shiyu Li

AMS Baltimore

Introduction

Flipping a Coin

Suppose you have flipped a fair coin n times, and recorded your answer:

e.g. $HTTTHHHHTHTHHHTTHTTTHHT$

- If you pick string of heads at random, how long will it be on average?
- What do you expect the longest run of heads to be?

With added conditions?

Coin flips are analogous to a random string of 0's and 1's. A run of heads is like a run of zeros or a *gap between ones.*

$HTTTHHHHTH = 0111000010$

With added conditions?

Coin flips are analogous to a random string of 0's and 1's. A run of heads is like a run of zeros or a *gap between ones.*

$HTTTHHHHTH = 0111000010$

- Now take all binary strings of length n of 0's and 1, with the restriction: no two 1's are adjacent.

  e.g. 1000101

With added conditions?

Coin flips are analogous to a random string of 0's and 1's. A run of heads is like a run of zeros or a *gap between ones.*

$HTTTHHHHTH = 0111000010$

- Now take all binary strings of length n of 0's and 1, with the restriction: no two 1's are adjacent.

  e.g. 1000101

- Fix one random string. How **long** will a random run of zeroes from that string be?
- For a random string, what do you expect the longest run of 0's to be?

Base 2

There is a **bijection** between **numbers in the interval** $[2^{n+1}, 2^{n+2})$ and **binary strings of length** n:

- Take the binary representation of x,

  e.g write 13 as 1101.

- Remove the first digit (always a 1), so $13 \mapsto 101$.

For Fibonacci Numbers

Fibonacci Numbers: $F_{n+1} = F_n + F_{n-1}$;
$F_1 = 1$, $F_2 = 2$, $F_3 = 3$, $F_4 = 5, \dots$.

Zeckendorf's Theorem

Every positive integer can be written uniquely as a sum of non-consecutive Fibonacci numbers.

Every number has a "base Fibonacci" decomposition:
Example:
$2014 = 1597 + 377 + 34 + 5 + 1 = F_{16} + F_{13} + F_8 + F_4 + F_1$.

We write 2014 as 1001000010001001. Notice, no two ones are adjacent

For more general sequences

This works for arbitrary **linearly recursive sequences with arbitrary nonnegative coefficients.**

$$H_{n+1} = c_1 H_n + c_2 H_{n-1} + \cdots + c_L H_{n-L+1}, \;\; n \geq L$$

with $H_1 = 1$, coefficients $c_i \geq 0$

Theorem (General Zeckendorf Theorem)

For every recurrence sequence H_n there is a notion of a legal decomposition string (of integers). There is a bijection between numbers $x \in [H_n, H_{n+1})$, and legal string of length n.

Legality reduces to non-adjacency in the case of Fibonacci numbers.

Connection to Difference Equations

The **probability questions** from before are actually questions about **Zeckendorf Decompositions!**

- Statistics about coin flips correspond to statistics about binary decompositions
- Random binary strings of **nonadjacent ones** go with Fibonacci numbers
- These probabilistic systems are governed by difference equation!

Results on Zeckendorf Decompositions

Lekkerkerker's Theorem (1952)

The average number of summands in the Zeckendorf decomposition for integers in $[F_n, F_{n+1})$ tends to $\frac{n}{\varphi^2+1} \approx .276n$, where $\varphi = \frac{1+\sqrt{5}}{2}$ is the golden mean.

Central Limit Type Theorem [KKMW]

As $n \to \infty$, the distribution of the number of summands in the Zeckendorf decomposition for integers in $[H_n, H_{n+1})$ is Gaussian (normal), with mean and variance computable constants in the coefficients H_i.

Gaps Between Summands

For $H_{i_1} + H_{i_2} + \cdots + H_{i_n}$, the gaps are the differences:

$$i_n - i_{n-1}, \ i_{n-1} - i_{n-2}, \ \ldots, \ i_2 - i_1.$$

Gaps Between Summands

For $H_{i_1} + H_{i_2} + \cdots + H_{i_n}$, the gaps are the differences:

$$i_n - i_{n-1}, \; i_{n-1} - i_{n-2}, \; \ldots, \; i_2 - i_1.$$

Example: For $H_1 + H_8 + H_{18}$, the gaps are 7 and 10.

Question 1: Gaps in the Bulk/Individual Gap Measures

Definition

Let $P_n(k)$ be the probability that a gap for a decomposition in $[H_n, H_{n+1})$ is of length k.

Big Question: What is $P(k) = \lim_{n\to\infty} P_n(k)$?

Definition

For $m \in (H_n + 1, H_n]$ with $k(m)$ gaps, the individual gap measure associated to m is
$\nu_{m;n}(x) \ := \ \frac{1}{k(m)-1} \sum_{j=2}^{k(m)} \delta(x - (r_j - r_{j-1}))$

More precisely: what is the behavior of the collection of $\nu_{m;n}$ as $n \to \infty$?

Question 2: Longest gap

Definition

For $x \in [H_n, H_{n+1})$ the *longest gap* or $L(x)$ is the max of all the gap lengths of x.

Example: For $x = H_1 + H_6 + H_{18} + H_{22}$, the longest gap is $L(x) = 12$.

Question 2: Longest gap

Definition

For $x \in [H_n, H_{n+1})$ the *longest gap* or $L(x)$ is the max of all the gap lengths of x.

Example: For $x = H_1 + H_6 + H_{18} + H_{22}$, the longest gap is $L(x) = 12$.

Question: How does the distribution $\mathbb{P}(L(x) = k)$ for $x \in [H_n, H_{n+1}))$ behave as $n \to \infty$?

For $H_n = 2^n$, this corresponds to the distribution of the **longest run of heads.**

Results

Previous results

Theorem (Base B Gap Distribution (SMALL 2011))

For base B decompositions, $P(0) = \frac{(B-1)(B-2)}{B^2}$*, and for* $k \geq 1$*,* $P(k) = c_B B^{-k}$*, with* $c_B = \frac{(B-1)(3B-2)}{B^2}$*.*

Theorem (Zeckendorf Gap Distribution (SMALL 2011))

For Zeckendorf decompositions, $P(k) = \frac{1}{\phi^k}$ *for* $k \geq 2$*, with* $\phi = \frac{1+\sqrt{5}}{2}$ *the golden mean.*

Theorem (Zeckendorf Gap Distribution)

Gap measures $\nu_{m;n}$ *converge almost surely to average gap measure where* $P(k) = 1/\phi^k$ *for* $k \geq 2$*.*

Theorem

Let $H_{n+1} = c_1 H_n + c_2 H_{n-1} + \cdots + c_L H_{n+1-L}$ *be a positive linear recurrence of length* L *where* $c_i \geq 1$ *for all* $1 \leq i \leq L$. *Then* $P(j) =$

$$\begin{cases} 1 - (\frac{a_1}{C_{Lek}})(\lambda_1^{-n+2} - \lambda_1^{-n+1} + 2\lambda_1^{-1} + a_1^{-1} - 3) & : j = 0 \\ \lambda_1^{-1}(\frac{1}{C_{Lek}})(\lambda_1(1 - 2a_1) + a_1) & : j = 1 \\ (\lambda_1 - 1)^2 \left(\frac{a_1}{C_{Lek}}\right) \lambda_1^{-j} & : j \geq 2 \end{cases}$$

Theorem (Individual Gap Measure Distribution)

The individual gap measures $\nu_{m;n}$ *converge almost surely to average gap measure.*

Proof of Fibonacci Result

Lekkerkerker $\Rightarrow$ total number of gaps $\sim F_{n-1}\frac{n}{\phi^2+1}$.

Proof of Fibonacci Result

Lekkerkerker $\Rightarrow$ total number of gaps $\sim F_{n-1}\frac{n}{\phi^2+1}$.

Let $X_{i,j} = \#\{m \in [F_n, F_{n+1})$: decomposition of m includes F_i, F_j, but not F_q for $i < q < j\}$.

Proof of Fibonacci Result

Lekkerkerker $\Rightarrow$ total number of gaps $\sim F_{n-1}\frac{n}{\phi^2+1}$.

Let $X_{i,j} = \#\{m \in [F_n, F_{n+1})$: decomposition of m includes F_i, F_j, but not F_q for $i < q < j\}$.

$$P(k) = \lim_{n\to\infty} \frac{\sum_{i=1}^{n-k} X_{i,i+k}}{F_{n-1}\frac{n}{\phi^2+1}}.$$

Calculating $X_{i,i+k}$

How many decompositions contain a gap from F_i to F_{i+k}?

Calculating $X_{i,i+k}$

How many decompositions contain a gap from F_i to F_{i+k}?

$1 \leq i \leq n - k - 2$:

Calculating $X_{i,i+k}$

How many decompositions contain a gap from F_i to F_{i+k}?

$1 \leq i \leq n - k - 2$:

For the indices less than i: F_{i-1} choices. Why? Have F_i, don't have F_{i-1}. Follows by Zeckendorf: like the interval $[F_i, F_{i+1})$ as have F_i, number elements is $F_{i+1} - F_i = F_{i-1}$.

Calculating $X_{i,i+k}$

How many decompositions contain a gap from F_i to F_{i+k}?

$1 \leq i \leq n - k - 2$:

For the indices less than i: F_{i-1} choices. Why? Have F_i, don't have F_{i-1}. Follows by Zeckendorf: like the interval $[F_i, F_{i+1})$ as have F_i, number elements is $F_{i+1} - F_i = F_{i-1}$.

For the indices greater than $i + k$: $F_{n-k-i-2}$ choices. Why? Have F_n, don't have F_{i+k+1}. Like Zeckendorf with potential summands $F_{i+k+2}, \ldots, F_n$. Shifting, like summands $F_1, \ldots, F_{n-k-i-1}$, giving $F_{n-k-i-2}$.

Calculating $X_{i,i+k}$

How many decompositions contain a gap from F_i to F_{i+k}?

$1 \leq i \leq n - k - 2$:

For the indices less than i: F_{i-1} choices. Why? Have F_i, don't have F_{i-1}. Follows by Zeckendorf: like the interval $[F_i, F_{i+1})$ as have F_i, number elements is $F_{i+1} - F_i = F_{i-1}$.

For the indices greater than $i + k$: $F_{n-k-i-2}$ choices. Why? Have F_n, don't have F_{i+k+1}. Like Zeckendorf with potential summands $F_{i+k+2}, \dots, F_n$. Shifting, like summands $F_1, \dots, F_{n-k-i-1}$, giving $F_{n-k-i-2}$.

So total choices number of choices is $F_{n-k-2-i}F_{i-1}$.

Determining $P(k)$

$$\sum_{i=1}^{n-k} X_{i,i+k} = F_{n-k-1} + \sum_{i=1}^{n-k-2} F_{i-1}F_{n-k-i-2}$$

- $\sum_{i=0}^{n-k-3} F_iF_{n-k-i-3}$ is the x^{n-k-3} coefficient of $(g(x))^2$, where $g(x)$ is the generating function of the Fibonaccis.
- Alternatively, use Binet's formula and get sums of geometric series.

Determining $P(k)$

$$\sum_{i=1}^{n-k} X_{i,i+k} = F_{n-k-1} + \sum_{i=1}^{n-k-2} F_{i-1}F_{n-k-i-2}$$

- $\sum_{i=0}^{n-k-3} F_i F_{n-k-i-3}$ is the x^{n-k-3} coefficient of $(g(x))^2$, where $g(x)$ is the generating function of the Fibonaccis.
- Alternatively, use Binet's formula and get sums of geometric series.

$P(k) = C/\phi^k$ for some constant C, so $P(k) = 1/\phi^k$.

Proof sketch of almost sure convergence

- $m = \sum_{j=1}^{k(m)} F_{i_j}$,
  $\nu_{m;n}(x) = \frac{1}{k(m)-1} \sum_{j=2}^{k(m)} \delta\left(x - (i_j - i_{j-1})\right)$.

- $\mu_{m,n}(t) \ = \ \int x^t \mathrm{d}\nu_{m;n}(x)$.

- Show $\mathbb{E}_m[\mu_{m;n}(t)]$ equals average gap moments, $\mu(t)$.

- Show $\mathbb{E}_m[(\mu_{m;n}(t) - \mu(t))^2]$ and $\mathbb{E}_m[(\mu_{m;n}(t) - \mu(t))^4]$ tend to zero.

Key ideas: (1) Replace $k(m)$ with average (Gaussianity); (2) use $X_{i,i+g_1,j,j+g_2}$.

Longest Gap

Longest Gap

For most recurrences, our central result is

Theorem (Mean and Variance of Longest Gap)

Let λ_1 be the largest eigenvalue of the recurrence, γ be Euler's constant, and K a constant that is a polynomial in λ_1. Then the mean and variance of the longeset gap are:

$$\mu_n = \frac{\log(nK)}{\log\lambda_1} + \frac{\gamma}{\log\lambda_1} - \frac{1}{2} + o(1)$$

$$\sigma_n^2 = \frac{\pi^2}{6(\log\lambda_1)^2} + o(1).$$

Strategy

Our argument follows three main steps:

- Find a rational generating function $S_f(x)$ for the number of $m \in (H_n, H_{n+1}]$ with longest gap less than f.
- Obtain an approximate formula for the CDF of the longest gap.
- Estimate the mean and variance using Partial Summation and the Euler Maclaurin Formula.

Fibonacci case

For the fibonacci numbers, our generating function is

$$S_f(x) = \frac{x}{1 - x - x^2 + x^f}.$$

From this we obtain

Theorem (Longest Gap Asymptotic CDF)

As $n \to \infty$, the probability that $m \in [F_n, F_{n+1})$ has longest gap less than or equal to $f(n)$ converges to

$$\text{Prob}\left(L_n(m) \leq f(n)\right) \;\approx\; e^{-e^{\log n - f(n)/\log \phi}}$$

Generating Function 1

For k fixed the number of $m \in [F_n, F_{n+1})$ with k summands and **longest gap less than** f equals the coefficient of

x^n for in the expression

$$\frac{1}{1-x}\left[\sum_{j=2}^{f(n)-2} x^j\right]^{k-1}.$$

Generating Function 2

Why the n^{th} coefficient of $\frac{1}{1-x}\left(\sum_{j=2}^{f(n)-1} x^j\right)^{k-1}$?

Generating Function 2

Why the n^{th} coefficient of $\frac{1}{1-x}\left(\sum_{j=2}^{f(n)-1} x^j\right)^{k-1}$?

Let $m = F_n + F_{n-g_1} + F_{n-g_1-g_2} + \cdots + F_{n-g_1-\cdots-g_{n-1}}$. The gaps uniquely identify m because of Zeckendorf's Theorem! And we have the following:

Generating Function 2

Why the n^{th} coefficient of $\frac{1}{1-x}\left(\sum_{j=2}^{f(n)-1} x^j\right)^{k-1}$?

Let $m = F_n + F_{n-g_1} + F_{n-g_1-g_2} + \cdots + F_{n-g_1-\cdots-g_{n-1}}$. The gaps uniquely identify m because of Zeckendorf's Theorem! And we have the following:

- The sum of the gaps of x is $\leq n$.

Generating Function 2

Why the n^{th} coefficient of $\frac{1}{1-x}\left(\sum_{j=2}^{f(n)-1} x^j\right)^{k-1}$?

Let $m = F_n + F_{n-g_1} + F_{n-g_1-g_2} + \cdots + F_{n-g_1-\cdots-g_{n-1}}$. The gaps uniquely identify m because of Zeckendorf's Theorem! And we have the following:

- The sum of the gaps of x is $\leq n$.
- Each gap is ≥ 2.

Generating Function 2

Why the n^{th} coefficient of $\frac{1}{1-x}\left(\sum_{j=2}^{f(n)-1} x^j\right)^{k-1}$?

Let $m = F_n + F_{n-g_1} + F_{n-g_1-g_2} + \cdots + F_{n-g_1-\cdots-g_{n-1}}$. The gaps uniquely identify m because of Zeckendorf's Theorem! And we have the following:

- The sum of the gaps of x is $\leq n$.
- Each gap is ≥ 2.
- Each gap is $< f$.

Generating Function

If we **sum** over k we get the total **number** of $m \in [F_n, F_{n+1})$ with longest gap $< f$. It's the n^{th} coefficient of

$$F(x) = \frac{1}{1-x} \sum_{k=1}^{\infty} \left(\frac{x^2 - x^{f-2}}{1-x} \right)^{k-1} = \frac{x}{1 - x - x^2 + x^f}.$$

Obtaining the CDF

We analyze asymptotic behavior of the coefficients of

$$S_f(x) = \frac{x}{1 - x - x^2 + x^f}$$

as n, f vary.

- Use a partial fraction decomposition.
- Problem: What happens to the roots of $1 - x - x^2 + x^f$ as f varies?
- Solution: $1 - x - x^2 + x^f$ has a unique smallest root α_f which converges to $1/\phi$ for large f.
- The contribution of α_f dominates, allowing us to obtain an approximate *CDF*.

Numerical Results

Convergence to mean is at best approximately $n^{-\delta}$ for some small $\delta > 0$. **Computing numerics is difficult**:

$F_{n+1} = F_n + F_{n-1}$: Sampling 100 numbers from $[F_n, F_{n+1})$ with $n = 1,000,000$.

- **Mean** predicted : 28.73 **vs.** observed: 28.51
- **Variance** predicted : 2.67 **vs.** observed: 2.44

$a_{n+1} = 2a_n + 4a_{n-1}$: Sampling 100 numbers from $[a_n, a_{n+1})$ with $n = 51,200$.

- **Mean** predicted : 9.95 **vs.** observed: 9.91
- **Variance** predicted : 1.09 **vs.** observed: 1.22

Numerical Results pt 2

$F_{n+1} = F_n + F_{n-1}$: Sampling 20 numbers from $[F_n, F_{n+1})$ with $n = 10,000,000$.

- **Mean** predicted : 33.52 **vs.** observed: 33.60
- **Variance** predicted : 2.67 **vs.** observed: 2.33

$a_{n+1} = 2a_n + 4a_{n-1}$: Sampling 100 numbers from $[a_n, a_{n+1})$ with $n = 102,400$.

- **Mean** predicted : 10.54 **vs.** observed: 10.45
- **Variance** predicted : 1.09 **vs.** observed: 1.10

Future Research

- Generalizing results to all PLRS and signed decompositions.
- Other systems such as f-Decompositions of Demontigny, Do, Miller and Varma.

Acknowledgements

Thanks to...

- NSF Grant DMS0850577
- NSF Grant DMS0970067
- AMS
- Our peers at Williams SMALL REU Summers 2010, 2011, 2012, and 2013

References

References

- Beckwith, Bower, Gaudet, Insoft, Li, Miller and Tosteson: The Average Gap Distribution for Generalized Zeckendorf Decompositions.
  http://arxiv.org/abs/1208.5820

- Bower, Insoft, Li, Miller and Tosteson: The Distribution of Gaps between Summands in Generalized Zeckendorf Decompositions
  http://arxiv.org/abs/1402.3912
- Kologlu, Kopp, Miller and Wang: Gaussianity for Fibonacci case.
  http://arxiv.org/pdf/1008.3204

- Miller - Wang: Gaussianity in general.
  http://arxiv.org/pdf/1008.3202

- Miller - Wang: Survey paper.
  http://arxiv.org/pdf/1107.2718