Distribution of the Longest Gap in Positive Linear Recurrence Sequences

Shiyu Li (Berkeley) and Phil Tosteson (Williams)

http://web.williams.edu/Mathematics/sjmiller/public_html/jmm2013.html/

MAA General Contributed Paper Session
Research in Number Theory, I
Room 2, Upper Level, San Diego
Wednesday, January 9, 2013, 8:15am

Introduction

Zeckendorf Decompositions

Fibonacci Numbers: $F_{n+1} = F_n + F_{n-1}$;

Zeckendorf Decompositions

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

Zeckendorf Decompositions

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.

Zeckendorf Decompositions

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.

Example:
$2012 = 1597 + 377 + 34 + 3 + 1 = F_{16} + F_{13} + F_8 + F_3 + F_1$.

Zeckendorf Decompositions

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.

Example:
$2012 = 1597 + 377 + 34 + 3 + 1 = F_{16} + F_{13} + F_8 + F_3 + F_1$.

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.

Gaps: Our object of study

Instead of looking at the number of summands, we study the spacings between them, as follows:

Definition: Gaps

If $x \in [F_n, F_{n+1})$ has Zeckendorf decomposition $x = F_n + F_{n-g_1} + F_{n-g_2} + \cdots + F_{n-g_k}$, we define the *gaps* in its decomposition to be $\{g_1, g_1 - g_2, \ldots, g_{k-1} - g_k\}$.

Example:

- $2012 = F_{16} + F_{13} + F_8 + F_3 + F_1$.

Gaps: Our object of study

Instead of looking at the number of summands, we study the spacings between them, as follows:

Definition: Gaps

If $x \in [F_n, F_{n+1})$ has Zeckendorf decomposition $x = F_n + F_{n-g_1} + F_{n-g_2} + \cdots + F_{n-g_k}$, we define the *gaps* in its decomposition to be $\{g_1, g_1 - g_2, \ldots, g_{k-1} - g_k\}$.

Example:

- $2012 = F_{16} + F_{13} + F_8 + F_3 + F_1$.
- Gaps of length 3, 5, and 4.

Previous Results (Miller-Beckwith)

Theorem (Zeckendorf Gap Distribution)

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

Previous Results (Miller-Beckwith)

Theorem (Zeckendorf Gap Distribution)

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

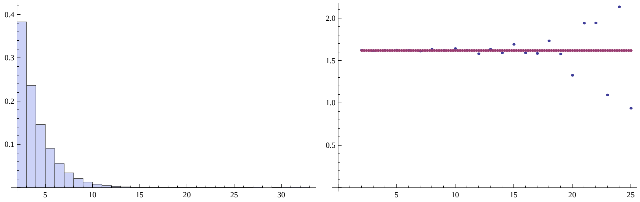

Figure: Distribution of gaps in $[F_{1000}, F_{1001})$; $F_{1000} \approx 10^{208}$.

Our Question

Given a random number m in the interval $[F_n, F_{n+1})$, what is the

probability that m has longest gap equal to r?

Results

Main Result

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

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

Main Result

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

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

Immediate Corollary: If $f(n)$ grows ***slower*** or ***faster*** than $\log n / \log \phi$, then $\text{Prob}(L_n(m) \leq f(n))$ goes to **0** or ***1***, respectively.

Main Result

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

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

Main Result

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

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

And from this sort of analysis we can get the mean:

$$\mu_n = \frac{\log\left(\frac{\phi^2}{\phi^2+1)}n\right)}{\log \phi} + \frac{\gamma}{\log \phi} - \frac{1}{2} + \text{Error}_{MC} + \epsilon_1(n),$$

where $\epsilon(n) \to 0$ for large n.

Fibonacci Case Generating Function

Let $G(n, k, f)$ be the number of m in $[F_n, F_{n+1})$ that have k nonzero summands in their Zeckendorf Decomposition and all gaps less than $f(n)$.

Fibonacci Case Generating Function

Let $G(n, k, f)$ be the number of m in $[F_n, F_{n+1})$ that have k nonzero summands in their Zeckendorf Decomposition and all gaps less than $f(n)$.

Then $G(n, k, f)$ is the coefficient of x^n for the generating function

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

Fibonacci Case Generating Function

Let $G(n,k,f)$ be the number of m in $[F_n, F_{n+1})$ that have k nonzero summands in their Zeckendorf Decomposition and all gaps less than $f(n)$.

Then $G(n,k,f)$ is the coefficient of x^n for the generating function

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

For fixed k, this is surprisingly hard to analyze. We only care about the sum over all k.

The Generating Function

If we **sum** over k we get the total **number** of $m \in [F_n, F_{n+1})$ with longest gap $< f(n)$, call it $G(n, f)$.

The Generating Function

If we **sum** over k we get the total **number** of $m \in [F_n, F_{n+1})$ with longest gap $< f(n)$, call it $G(n, f)$. This is the n^{th} coefficient of

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

The Generating Function

If we **sum** over k we get the total **number** of $m \in [F_n, F_{n+1})$ with longest gap $< f(n)$, call it $G(n, f)$. This is the n^{th} coefficient of

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

We use **partial fractions** and **Rouché** to find the CDF.

Exact CDF

- Expand $F(x)$ using roots of $x^f - x^2 - x - 1$

Exact CDF

- Expand $F(x)$ using roots of $x^f - x^2 - x - 1$
- n^{th} coefficient of this is the number of y with gaps less than $f(n)$; dividing by the total gives us the probability.

Exact CDF

- Expand $F(x)$ using roots of $x^f - x^2 - x - 1$
- n^{th} coefficient of this is the number of y with gaps less than $f(n)$; dividing by the total gives us the probability.

Theorem (Exact CDF)

The proportion of $m \in [F_n, F_{n+1})$ with $L(x) < f(n)$ is exactly

$$\sum_{i=1}^{f(n)} \frac{-\sqrt{5}(\alpha_i)}{f(n)\alpha_i^{f(n)} - 2\alpha_i^2 - \alpha_i} \left(\frac{1}{\alpha_i}\right)^{n+1} \frac{1}{(\phi^n - (-1/\phi)^n)}$$

Exact CDF

- Expand $F(x)$ using roots of $x^f - x^2 - x - 1$
- n^{th} coefficient of this is the number of y with gaps less than $f(n)$; dividing by the total gives us the probability.

Theorem (Exact CDF)

The proportion of $m \in [F_n, F_{n+1})$ with $L(x) < f(n)$ is exactly

$$\sum_{i=1}^{f(n)} \frac{-\sqrt{5}(\alpha_i)}{f(n)\alpha_i^{f(n)} - 2\alpha_i^2 - \alpha_i} \left(\frac{1}{\alpha_i}\right)^{n+1} \frac{1}{(\phi^n - (-1/\phi)^n)}$$

Next, we look at the roots of $x^f - x^2 - x + 1$.

Rouché’s Theorem

A useful consequence of the argument principle is Rouché’s Theorem:

Theorem (Rouché’s Theorem)

Suppose we have two functions f and g on a region K and that $|f(x) - g(x)| < |g(x)|$ for all x on the boundary δK. Then f and g have the same number of roots inside K.

Rouché's Theorem

Analyzing $z^f - z^2 - z + 1$ we obtain:

Lemma (Critical Root Behavior)

For $f \in \mathbb{N}$ and $f \geq 4$, the polynomial $p_f(z) = z^f - z^2 - z + 1$ has exactly one root z_f with $|z_f| < .9$. Further, $z_f \in \mathbb{R}$ and $z_f = \frac{1}{\phi} + \left|\frac{z_f^f}{z_f + \phi}\right|$, so as $f \to \infty$, z_f converges to $\frac{1}{\phi}$.

Rouché's Theorem

Analyzing $z^f - z^2 - z + 1$ we obtain:

Lemma (Critical Root Behavior)

For $f \in \mathbb{N}$ and $f \geq 4$, the polynomial $p_f(z) = z^f - z^2 - z + 1$ has exactly one root z_f with $|z_f| < .9$. Further, $z_f \in \mathbb{R}$ and $z_f = \frac{1}{\phi} + \left|\frac{z_f^f}{z_f + \phi}\right|$, so as $f \to \infty$, z_f converges to $\frac{1}{\phi}$.

As f grows, only one root goes to $1/\phi$. The other roots don't matter. This gives us

Getting the CDF

Theorem (Approximate Cumulative Distribution Function)

If $\lim_{n\to\infty} f(n) = \infty$, the proportion of m with $L(m) < f(n)$ is, as $n \to \infty$

$$\lim_{n\to\infty} (\phi z_f)^{-n} = \lim_{n\to\infty} \left(1 + \left|\frac{\phi z_f^{f(n)}}{\phi + z_f}\right|\right)^{-n}.$$

If $f(n)$ is bounded, then $P_f = 0$.

Getting the CDF

Theorem (Approximate Cumulative Distribution Function)

If $\lim_{n\to\infty} f(n) = \infty$, *the proportion of* m *with* $L(m) < f(n)$ *is, as* $n \to \infty$

$$\lim_{n\to\infty} (\phi z_f)^{-n} = \lim_{n\to\infty} \left(1 + \left|\frac{\phi z_f^{f(n)}}{\phi + z_f}\right|\right)^{-n}.$$

If $f(n)$ *is bounded, then* $P_f = 0$.

We can see the double exponential by taking logarithms, Taylor expanding, and re-exponentiating.

Mean/Variance

Note

$$\mu = \sum_{j=1}^{n} j(\; CDF(j) - CDF(j-1)\;)$$

Using **Partial Summation**, Euler-Maclaurin, and evaluating the resulting integrals, we calculate the mean and variance.

Mean/Variance

Note

$$\mu = \sum_{j=1}^{n} j(\ CDF(j) - CDF(j-1)\ )$$

Using **Partial Summation**, Euler-Maclaurin, and evaluating the resulting integrals, we calculate the mean and variance.

$$\mu_n = \frac{\log\left(\frac{\phi^2}{\phi^2+1)}n\right)}{\log\phi} + \frac{\gamma}{\log\phi} - \frac{1}{2} + \text{Error}_{MC} + \epsilon_1(n),$$

and

$$\sigma_n^2 = \frac{\pi^2}{6\log\phi} - \frac{1}{12} + \text{Error}_{MC}^2 + \epsilon_2(n),$$

where $\epsilon_1(n), \epsilon_2(n)$ go to zero in the limit.

Generalizations

Positive Linear Recurrence Sequences

This method can be greatly generalized to **Positive Linear Recurrence Sequences** ie: linear recurrences with non-negative coefficients. WLOG:

$$H_{n+1} = c_1 H_{n-(j_1=0)} + c_2 H_{n-j_2} + \cdots + c_L H_{n-j_L}.$$

Theorem (Zeckendorf's Theorem for *PLRS* recurrences)

*Any $b \in \mathbb{N}$ has a unique **legal** decomposition into sums of H_n, $b = a_1 H_{i_1} + \cdots + a_{i_k} H_{i_k}$.*

Here **legal** reduces to non-adjacency of summands in the Fibonacci case.

Messier Combinatorics

The **number** of $b \in [H_n, H_{n+1})$, with longest gap $< f$ is the coefficient of x^{n-s} in the generating function:

Messier Combinatorics

The **number** of $b \in [H_n, H_{n+1})$, with longest gap $< f$ is the coefficient of x^{n-s} in the generating function:

$$\frac{1}{1-x}\left(c_1 - 1 + c_2 x^{t_2} + \cdots + c_L x^{t_L}\right) \times$$
$$\times \sum_{k \geq 0} \Bigg[\left((c_1 - 1)x^{t_1} + \cdots + (c_L - 1)x^{t_L}\right)\left(\frac{x^{s+1} - x^f}{1-x}\right) +$$
$$+ x^{t_1}\left(\frac{x^{s+t_2-t_1+1} - x^f}{1-x}\right) + \cdots + x^{t_{L-1}}\left(\frac{x^{s+t_L-t_{L-1}} + 1 - x^f}{1-x}\right)\Bigg]^k .$$

Messier Combinatorics

The **number** of $b \in [H_n, H_{n+1})$, with longest gap $< f$ is the coefficient of x^{n-s} in the generating function:

$$\frac{1}{1-x}\left(c_1 - 1 + c_2 x^{t_2} + \cdots + c_L x^{t_L}\right) \times$$

$$\times \sum_{k \geq 0} \Bigg[\left((c_1 - 1)x^{t_1} + \cdots + (c_L - 1)x^{t_L}\right) \left(\frac{x^{s+1} - x^f}{1-x}\right) +$$

$$+x^{t_1}\left(\frac{x^{s+t_2-t_1+1} - x^f}{1-x}\right) + \cdots + x^{t_{L-1}}\left(\frac{x^{s+t_L-t_{L-1}} + 1 - x^f}{1-x}\right)\Bigg]^k.$$

A geometric series!

Let $f > j_L$. The number of $x \in [H_n, H_{n+1})$, with longest gap $< f$ is given by **the coefficient of s^n** in the generating function

$$F(s) = \frac{1 - s^{j_L}}{\mathcal{M}(s) + s^f \mathcal{R}(s)},$$

where

$$\mathcal{M}(s) = 1 - c_1 s - c_2 s^{j_2+1} - \cdots - c_L s^{j_L+1},$$

and

$$\mathcal{R}(s) = c_{j_1+1} s^{j_1} + c_{j_2+1} s^{j_2} + \cdots + (c_{j_L+1} - 1) s^{j_L}.$$

and c_i and j_i are defined **as above**.

The **coefficients** in the **partial fraction** expansion might blow up from multiple roots.

The **coefficients** in the **partial fraction** expansion might blow up from multiple roots.

Theorem (Mean and Variance for "Most Recurrences")

For x in the interval $[H_n, H_{n+1})$, the mean longest gap μ_n and the variance of the longest gap σ_n^2 are given by

$$\mu_n = \frac{\log\left(\frac{\mathcal{R}(\frac{1}{\lambda_1})}{\mathcal{G}(\frac{1}{\lambda_1})} n\right)}{\log \lambda_1} + \frac{\gamma}{\log \lambda_1} - \frac{1}{2} + Error_{MC}^1 + \epsilon_1(n),$$

and

$$\sigma_n^2 = \frac{\pi^2}{6 \log \lambda_1} - \frac{1}{12} + Error_{MC}^2 + \epsilon_2(n),$$

where $\epsilon_i(n)$ tends to zero in the limit, and $Error_{MC}$ comes from the Euler-Maclaurin Formula.

References

References

- Beckwith, Bower, Gaudet, Insoft, Li, Miller and Tosteson: Bulk gaps for average gap measure.
  http://arxiv.org/abs/1208.5820

- 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

Acknowledgements

Thanks to:

- Williams College
- National Science Foundation
- Professor Steven Miller