Distribution of Summands in Zeckendorf Decompositions

Steven J. Miller (sjm1@williams.edu)
http://www.williams.edu/Mathematics/sjmiller/public_html

Special Session on Additive and Combinatorial Number Theory
AMS Sectional, Akron, Ohio, October 21, 2012
Introduction
Goals of the Talk

- Overview of previous results.
- Highlight techniques: combinatorics, generating fns.
- Interesting generalizations.
- Open problems.

Joint with Olivia Beckwith, Amanda Bower, Louis Gaudet, Rachel Insoft, Shiyu Li, Philip Tosteson.
Previous Results

Fibonacci Numbers: \( F_{n+1} = F_n + F_{n-1} \);
\( F_1 = 1, F_2 = 2, F_3 = 3, F_4 = 5, \ldots \).
Previous Results

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

Zeckendorf’s Theorem

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

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

Fibonacci Numbers: \[ F_{n+1} = F_n + F_{n-1} \];
\[ F_1 = 1, \ F_2 = 2, \ F_3 = 3, \ F_4 = 5, \ldots \]

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

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

Lekkerkerkerkerkerkerken’s Theorem (1952)
The average number of summands in the Zeckendorf decomposition for integers in \([F_n, F_{n+1}]\) tends to \(\frac{n}{\sqrt[2]{\phi^2 + 1}} \approx 0.276n\), where \(\phi = \frac{1 + \sqrt{5}}{2}\) is the golden mean.
Old Results

Central Limit Type Theorem
As $n \to \infty$, the distribution of the number of summands in the Zeckendorf decomposition for integers in $[F_n, F_{n+1})$ is Gaussian.

Figure: Number of summands in $[F_{2010}, F_{2011})$; $F_{2010} \approx 10^{420}$. 
New Results: Bulk Gaps: $m \in [F_n, F_{n+1})$ and $\phi = \frac{1+\sqrt{5}}{2}$

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

Theorem (Zeckendorf Gap Distribution)

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

Figure: Distribution of gaps in $[F_{1000}, F_{1001})$: $F_{2010} \approx 10^{208}$. 
New Results: Longest Gap

Theorem (Longest Gap)

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

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

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.
Reinterpreting the Cookie Problem

The number of solutions to $x_1 + \cdots + x_P = C$ with $x_i \geq 0$ is

$$\binom{C+P-1}{P-1}.$$
Preliminaries: The Cookie Problem: Reinterpretation

Reinterpreting the Cookie Problem

The number of solutions to \( x_1 + \cdots + x_P = C \) with \( x_i \geq 0 \) is
\[
\binom{C+P-1}{P-1}.
\]

Let \( p_{n,k} = \# \{ N \in [F_n, F_{n+1}): \text{the Zeckendorf decomposition of } N \text{ has exactly } k \text{ summands} \} \).

For \( N \in [F_n, F_{n+1}) \), the largest summand is \( F_n \).
Preliminaries: The Cookie Problem: Reinterpretation

Reinterpreting the Cookie Problem

The number of solutions to $x_1 + \cdots + x_P = C$ with $x_i \geq 0$ is

\[ \binom{C + P - 1}{P - 1}. \]

Let $p_{n,k} = \# \{ N \in [F_n, F_{n+1}) : \text{the Zeckendorf decomposition of } N \text{ has exactly } k \text{ summands} \}$.

For $N \in [F_n, F_{n+1})$, the largest summand is $F_n$.

\[ N = F_{i_1} + F_{i_2} + \cdots + F_{i_{k-1}} + F_n, \]

\[ 1 \leq i_1 < i_2 < \cdots < i_{k-1} < i_k = n, \quad i_j - i_{j-1} \geq 2. \]

\[ d_1 := i_1 - 1, \quad d_j := i_j - i_{j-1} - 2 \quad (j > 1). \]

\[ d_1 + d_2 + \cdots + d_k = n - 2k + 1, \quad d_j \geq 0. \]

Cookie counting $\Rightarrow p_{n,k} = \binom{n - 2k + 1 + k - 1}{k - 1} = \binom{n - k}{k - 1}$. 
Gaps in the Bulk
Distribution of Gaps

For $F_{i_1} + F_{i_2} + \cdots + F_{i_n}$, the gaps are the differences $i_n - i_{n-1}, i_{n-1} - i_{n-2}, \ldots, i_2 - i_1$.

Example: For $F_1 + F_8 + F_{18}$, the gaps are 7 and 10.
Distribution of Gaps

For $F_{i_1} + F_{i_2} + \cdots + F_{i_n}$, the gaps are the differences $i_n - i_{n-1}, i_{n-1} - i_{n-2}, \ldots, i_2 - i_1$.

Example: For $F_1 + F_8 + F_{18}$, the gaps are 7 and 10.

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

For $F_{i_1} + F_{i_2} + \cdots + F_{i_n}$, the gaps are the differences $i_n - i_{n-1}, i_{n-1} - i_{n-2}, \ldots, i_2 - i_1$.

Example: For $F_1 + F_8 + F_{18}$, the gaps are 7 and 10.

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

What is $P(k) = \lim_{n \to \infty} P_n(k)$?
Distribution of Gaps

For $F_{i_1} + F_{i_2} + \cdots + F_{i_n}$, the gaps are the differences $i_n - i_{n-1}, i_{n-1} - i_{n-2}, \ldots, i_2 - i_1$.

Example: For $F_1 + F_8 + F_{18}$, the gaps are 7 and 10.

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

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

Can ask similar questions about binary or other expansions:

$2012 = 2^{10} + 2^9 + 2^8 + 2^7 + 2^6 + 2^4 + 2^3 + 2^2$. 
Main Results

**Theorem (Base $B$ Gap Distribution)**

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)**

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

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 $P(j) =$

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

Lekkerkerker ⇒ total number of gaps \( \sim F_{n-1} \frac{n}{\phi^2 + 1} \).
Proof of Fibonacci Result

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

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

Lekkerkerker ⇒ total number of gaps \( \sim F_{n-1} \frac{n}{\phi^2 + 1} \).

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

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

Number of choices is $F_{n-k-2-i}F_{i-1}$. 
Calculating $X_{i,i+k}$

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

Number of choices is $F_{n-k-2-i}F_{i-1}$.

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 $F_{i+k}$?

Number of choices is $F_{n-k-1-i}F_{i-1}$.

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 $F_{n-k-i-2}$. 
Determining $P(k)$ for $k \geq 2$

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

- $\sum_{i=0}^{n-k-3} F_i F_{n-k-i-3}$ is the $x^{n-k-3}$ coefficient of $(g(x))^2$, where $g(x)$ is the generating function of the Fibonacci.

- Alternatively, use Binet’s formula and get sums of geometric series.
Determining $P(k)$ for $k \geq 2$

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

- $\sum_{i=0}^{n-k-3} F_i F_{n-k-i-3}$ 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.

$P(k) = \frac{C}{\phi^k}$ for a constant $C$, so $P(k) = \frac{1}{\phi^k}$. 

Proof sketch of almost sure convergence

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

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
Fibonacci Case Generating Function

\[ G_{n,k,f} \] be the number of \( m \in \left[ F_n, F_{n+1} \right) \) with \( k \) nonzero summands and all gaps less than \( f(n) \).
Fibonacci Case Generating Function

$G_{n,k,f}$ be the number of $m \in [F_n, F_{n+1})$ with $k$ nonzero summands and all gaps less than $f(n)$.

$G_{n,k,f}$ 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

$G_{n,k,f}$ be the number of $m \in [F_n, F_{n+1})$ with $k$ nonzero summands and all gaps less than $f(n)$.

$G_{n,k,f}$ is the coefficient of $x^n$ for the generating function

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

Let $m = F_n + F_{n-g_1} + F_{n-g_1-g_2} + \cdots + F_{n-g_1-\cdots-g_{n-1}}$, then

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

Gaps uniquely identify $m$ by Zeckendorf’s Theorem.
The Combinatorics

\[ G_{n,k,f} \text{ is the } n^{th} \text{ coefficient of} \]

\[
\frac{1}{1-x} \left[ x^2 + \cdots + x^{f(n)-2} \right]^{k-1} = \frac{x^{2(k-1)}}{1-x} \left( \frac{1-x^{f(n)-3}}{1-x} \right)^{k-1}.
\]
The Combinatorics

\( G_{n,k,f} \) is the \( n^{th} \) coefficient of

\[
\frac{1}{1-x} \left[ x^2 + \cdots + x^{f(n)-2} \right]^{k-1} = \frac{x^{2(k-1)}}{1-x} \left( \frac{1 - x^{f(n)-3}}{1-x} \right)^{k-1}.
\]

For fixed \( k \) hard to analyze, but only care about sum over \( k \).
The Generating Function

Sum over $k$ gives number of $m \in [F_n, F_{n+1})$ with longest gap $< f(n)$, call it $G_{n,f}$.

It’s the $n^{th}$ coefficient (up to potentially small algebra errors!) of

$$F(x) = \frac{1}{1-x} \sum_{k=1}^{\infty} \left( \frac{x^2 - x^{f-2}}{1-x} \right)^{k-1} = \frac{x}{1-x-x^2 + x^{f(n)}}.$$
Sum over $k$ gives number of $m \in [F_n, F_{n+1})$ with longest gap $< f(n)$, call it $G_{n,f}$.

It’s the $n^{th}$ coefficient (up to potentially small algebra errors!) of

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

Use **partial fractions** and **Rouché** to find the CDF.
Partial Fractions

Write the roots of $x^f - x^2 - x - 1$ as $\{\alpha_i\}_{i=1}^f$, generating function is

$$F(x) = \frac{x}{1 - x - x^2 + x^f(n)} = \sum_{i=1}^{f(n)} \frac{-\alpha_i}{f(n)\alpha_i - 2\alpha_i^2 - \alpha_i} \sum_{j=1}^{\infty} \left( \frac{x}{\alpha_i} \right)^j.$$
Partial Fractions

Write the roots of \( x^f - x^2 - x - 1 \) as \( \{\alpha_i\}_{i=1}^f \), generating function is

\[
F(x) = \frac{x}{1 - x - x^2 + x^f} = \sum_{i=1}^{f(n)} \frac{-\alpha_i}{f(n)\alpha_i^f(n) - 2\alpha_i^2 - \alpha_i} \sum_{j=1}^\infty \left( \frac{x}{\alpha_i} \right)^j .
\]

Take the \( n^{\text{th}} \) coefficient to find the number of \( m \) with gaps less than \( f(n) \).
Partial Fractions

Divide the number of $m \in [F_n, F_{n+1})$ with longest gap $< f(n)$ by the number of $m$, which is $F_{n+1} - F_n = F_{n-1} = 5^{-1/2}(\phi^{n-1} - (1/\phi)^{n-1})$.

**Theorem**

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

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

Now, we find out about the roots of $x^f - x^2 - x + 1$. 
Rouché and Roots

When $f(n)$ is large $z^{f(n)}$ is very small, for $|z| < 1$. Thus, by Rouché’s theorem from complex analysis:

**Lemma**

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 $z_f = \frac{1}{\phi} + \left| \frac{z_f}{z_f + \phi} \right|$, so as $f \to \infty$, $z_f$ converges to $\frac{1}{\phi}$.

We only care about the **smallest root**.
Getting the CDF

As \( f \) grows, only one root goes to \( 1/\phi \). The other roots don’t matter. So,
Getting the CDF

As $f$ grows, only one root goes to $1/\phi$. The other roots don’t matter. So,

**Theorem**

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 $P_f = 0$.

Take logarithms, Taylor expand, result follows from algebra.
References
References

  http://arxiv.org/abs/1208.5820

  http://arxiv.org/pdf/1008.3204

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

  http://arxiv.org/pdf/1107.2718
Appendix: Gaussian Behavior
Generalizing Lekkerkerker: Erdos-Kac type result

**Theorem (KKMW 2010)**

As \( n \to \infty \), the distribution of the number of summands in Zeckendorf’s Theorem is a Gaussian.

**Sketch of proof:** Use Stirling’s formula,

\[
n! \approx n^n e^{-n} \sqrt{2\pi n}
\]

to approximate binomial coefficients, after a few pages of algebra find the probabilities are approximately Gaussian.
(Sketch of the) Proof of Gaussianity

The probability density for the number of Fibonacci numbers that add up to an integer in \([F_n, F_{n+1})\) is

\[ f_n(k) = \binom{n-1-k}{k} / F_{n-1}. \]

Consider the density for the \(n+1\) case. Then we have, by Stirling

\[
 f_{n+1}(k) = \binom{n-k}{k} \frac{1}{F_n} \\
 = \frac{(n-k)!}{(n-2k)!k!} \frac{1}{F_n} = \frac{1}{\sqrt{2\pi}} \frac{(n-k)^{n-k+1/2}}{k^{k+1/2}(n-2k)^{n-2k+1/2}} \frac{1}{F_n}
\]

plus a lower order correction term.

Also we can write \(F_n = \frac{1}{\sqrt{5}} \phi^{n+1} = \frac{\phi}{\sqrt{5}} \phi^n\) for large \(n\), where \(\phi\) is the golden ratio (we are using relabeled Fibonacci numbers where \(1 = F_1\) occurs once to help dealing with uniqueness and \(F_2 = 2\)). We can now split the terms that exponentially depend on \(n\).

\[
 f_{n+1}(k) = \left( \frac{1}{\sqrt{2\pi}} \sqrt{\frac{(n-k)}{k(n-2k)}} \frac{\sqrt{5}}{\phi} \right) \left( \phi^{-n} \frac{(n-k)^{n-k}}{k^k(n-2k)^{n-2k}} \right).
\]

Define

\[
 N_n = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{(n-k)}{k(n-2k)}} \frac{\sqrt{5}}{\phi}, \quad S_n = \phi^{-n} \frac{(n-k)^{n-k}}{k^k(n-2k)^{n-2k}}.
\]

Thus, write the density function as

\[ f_{n+1}(k) = N_n S_n \]

where \(N_n\) is the first term that is of order \(n^{-1/2}\) and \(S_n\) is the second term with exponential dependence on \(n\).
Model the distribution as centered around the mean by the change of variable $k = \mu + x\sigma$ where $\mu$ and $\sigma$ are the mean and the standard deviation, and depend on $n$. The discrete weights of $f_n(k)$ will become continuous. This requires us to use the change of variable formula to compensate for the change of scales:

$$f_n(k)dk = f_n(\mu + \sigma x)\sigma dx.$$ 

Using the change of variable, we can write $N_n$ as

$$N_n = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n-k}{k(n-2k)}} \sqrt{\frac{\phi}{\sqrt{5}}}$$ 

$$= \frac{1}{\sqrt{2\pi n}} \sqrt{\frac{1-k/n}{(k/n)(1-2k/n)}} \sqrt{\frac{\phi}{\sqrt{5}}}$$ 

$$= \frac{1}{\sqrt{2\pi n}} \sqrt{\frac{1-(\mu+\sigma x)/n}{((\mu+\sigma x)/n)(1-2(\mu+\sigma x)/n)}} \sqrt{\frac{\phi}{\sqrt{5}}}$$ 

$$= \frac{1}{\sqrt{2\pi n}} \sqrt{\frac{1-C-y}{(C+y)(1-2C-2y)}} \sqrt{\frac{\phi}{\sqrt{5}}}$$

where $C = \mu/n \approx 1/(\phi + 2)$ (note that $\phi^2 = \phi + 1$) and $y = \sigma x/n$. But for large $n$, the $y$ term vanishes since $\sigma \sim \sqrt{n}$ and thus $y \sim n^{-1/2}$. Thus

$$N_n \approx \frac{1}{\sqrt{2\pi n}} \sqrt{\frac{1-C}{C(1-2C)}} \sqrt{\frac{\phi}{\sqrt{5}}} = \frac{1}{\sqrt{2\pi n}} \sqrt{\frac{(\phi+1)(\phi+2)}{\phi}} \sqrt{\frac{5}{\sqrt{5}}} \sqrt{\frac{\phi}{\phi}} = \frac{1}{\sqrt{2\pi n}} \sqrt{\frac{5(\phi+2)}{\phi}} = \frac{1}{\sqrt{2\pi \sigma^2}}$$

since $\sigma^2 = n \frac{\phi}{5(\phi+2)}$. 

(Sketch of the) Proof of Gaussianity
(Sketch of the) Proof of Gaussianity

For the second term $S_n$, take the logarithm and once again change variables by $k = \mu + x\sigma$,

$$
\log(S_n) = \log \left( \frac{(n-k)^{(n-k)}}{k^k(n-2k)^{(n-2k)}} \right)
$$

$$
= -n\log(\phi) + (n-k)\log(n-k) - (k)\log(k)
- (n-2k)\log(n-2k)
$$

$$
= -n\log(\phi) + (n - (\mu + x\sigma))\log(n - (\mu + x\sigma))
- (\mu + x\sigma)\log(\mu + x\sigma)
- (n - 2(\mu + x\sigma))\log(n - 2(\mu + x\sigma))
$$

$$
= -n\log(\phi)
+ (n - (\mu + x\sigma)) \left( \log(n - \mu) + \log \left( 1 - \frac{x\sigma}{n - \mu} \right) \right)
- (\mu + x\sigma) \left( \log(\mu) + \log \left( 1 + \frac{x\sigma}{\mu} \right) \right)
- (n - 2(\mu + x\sigma)) \left( \log(n - 2\mu) + \log \left( 1 - \frac{x\sigma}{n - 2\mu} \right) \right)
$$

$$
= -n\log(\phi)
+ (n - (\mu + x\sigma)) \left( \log \left( \frac{n}{\mu} - 1 \right) + \log \left( 1 - \frac{x\sigma}{n - \mu} \right) \right)
- (\mu + x\sigma) \log \left( 1 + \frac{x\sigma}{\mu} \right)
- (n - 2(\mu + x\sigma)) \left( \log \left( \frac{n}{\mu} - 2 \right) + \log \left( 1 - \frac{x\sigma}{n - 2\mu} \right) \right).$$
(Sketch of the) Proof of Gaussianity

Note that, since \( n/\mu = \phi + 2 \) for large \( n \), the constant terms vanish. We have \( \log(S_n) \)

\[
\begin{align*}
&= -n \log(\phi) + (n - k) \log \left( \frac{n}{\mu} - 1 \right) - (n - 2k) \log \left( \frac{n}{\mu} - 2 \right) + (n - (\mu + x\sigma)) \log \left( 1 - \frac{x\sigma}{n - \mu} \right) \\
&\quad - (\mu + x\sigma) \log \left( 1 + \frac{x\sigma}{\mu} \right) - (n - 2(\mu + x\sigma)) \log \left( 1 - \frac{x\sigma}{n - 2\mu} \right) \\
&= n(- \log(\phi) + \log(\phi^2) - \log(\phi)) + k(\log(\phi^2) + 2 \log(\phi)) + (n - (\mu + x\sigma)) \log \left( 1 - \frac{x\sigma}{n - \mu} \right) \\
&\quad - (\mu + x\sigma) \log \left( 1 + \frac{x\sigma}{\mu} \right) - (n - 2(\mu + x\sigma)) \log \left( 1 - 2 \frac{x\sigma}{n - 2\mu} \right) \\
&= (n - (\mu + x\sigma)) \log \left( 1 - \frac{x\sigma}{n - \mu} \right) - (\mu + x\sigma) \log \left( 1 + \frac{x\sigma}{\mu} \right) \\
&\quad - (n - 2(\mu + x\sigma)) \log \left( 1 - 2 \frac{x\sigma}{n - 2\mu} \right).
\end{align*}
\]
Finally, we expand the logarithms and collect powers of \( x\sigma / n \).

\[
\log(S_n) = (n - (\mu + x\sigma)) \left( -\frac{x\sigma}{n - \mu} - \frac{1}{2} \left( \frac{x\sigma}{n - \mu} \right)^2 + \ldots \right) \\
- (\mu + x\sigma) \left( \frac{x\sigma}{\mu} - \frac{1}{2} \left( \frac{x\sigma}{\mu} \right)^2 + \ldots \right) \\
- (n - 2(\mu + x\sigma)) \left( -\frac{x\sigma}{n - 2\mu} - \frac{1}{2} \left( \frac{x\sigma}{n - 2\mu} \right)^2 + \ldots \right)
\]

\[
= (n - (\mu + x\sigma)) \left( -\frac{x\sigma}{n} \frac{(\phi+1)}{(\phi+2)} - \frac{1}{2} \left( \frac{x\sigma}{n} \frac{(\phi+1)}{(\phi+2)} \right)^2 + \ldots \right) \\
- (\mu + x\sigma) \left( \frac{x\sigma}{n} \frac{1}{\phi+2} - \frac{1}{2} \left( \frac{x\sigma}{n} \frac{1}{\phi+2} \right)^2 + \ldots \right) \\
- (n - 2(\mu + x\sigma)) \left( -\frac{2x\sigma}{n} \frac{\phi}{\phi+2} - \frac{1}{2} \left( \frac{2x\sigma}{n} \frac{\phi}{\phi+2} \right)^2 + \ldots \right)
\]

\[
= \frac{x\sigma}{n} n \left( - \left( 1 - \frac{1}{\phi+2} \right) \frac{(\phi+2)}{(\phi+1)} - 1 + 2 \left( 1 - \frac{2}{\phi+2} \right) \frac{\phi + 2}{\phi} \right) \\
- \frac{1}{2} \left( \frac{x\sigma}{n} \right)^2 n \left( -2\frac{\phi + 2}{\phi + 1} + \frac{\phi + 2}{\phi + 1} + 2(\phi + 2) - (\phi + 2) + 4 \frac{\phi + 2}{\phi} \right) \\
+ O \left( n(x\sigma/n)^3 \right)
\]
(Sketch of the) Proof of Gaussianity

\[
\log(S_n) = \frac{x \sigma}{n} \left( - \frac{\phi + 1}{\phi + 2} \frac{\phi + 2}{\phi + 1} - 1 + 2 \frac{\phi + 2}{\phi + 2} \right) \\
- \frac{1}{2} \left( \frac{x \sigma}{n} \right)^2 n(\phi + 2) \left( - \frac{1}{\phi + 1} + 1 + \frac{4}{\phi} \right) \\
+ O \left( n \left( \frac{x \sigma}{n} \right)^3 \right)
\]

\[
= - \frac{1}{2} \frac{(x \sigma)^2}{n} (\phi + 2) \left( \frac{3 \phi + 4}{\phi(\phi + 1)} + 1 \right) + O \left( n \left( \frac{x \sigma}{n} \right)^3 \right)
\]

\[
= - \frac{1}{2} \frac{(x \sigma)^2}{n} (\phi + 2) \left( \frac{3 \phi + 4 + 2 \phi + 1}{\phi(\phi + 1)} \right) + O \left( n \left( \frac{x \sigma}{n} \right)^3 \right)
\]

\[
= - \frac{1}{2} x^2 \sigma^2 \left( \frac{5(\phi + 2)}{\phi n} \right) + O \left( n (x \sigma / n)^3 \right)
\]
But recall that

$$\sigma^2 = \frac{\phi n}{5(\phi + 2)}.$$ 

Also, since $\sigma \sim n^{-1/2}$, $n \left(\frac{x\sigma}{n}\right)^3 \sim n^{-1/2}$. So for large $n$, the $O\left(n \left(\frac{x\sigma}{n}\right)^3\right)$ term vanishes. Thus we are left with

$$\log S_n = -\frac{1}{2}x^2$$

$$S_n = e^{-\frac{1}{2}x^2}.$$ 

Hence, as $n$ gets large, the density converges to the normal distribution:

$$f_n(k)dk = N_n S_n dk$$

$$= \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{1}{2}x^2} \sigma dx$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx.$$
Generalizations

Generalizing from Fibonacci numbers to linearly recursive sequences with arbitrary nonnegative coefficients.

\[ H_{n+1} = c_1 H_n + c_2 H_{n-1} + \cdots + c_L H_{n-L+1}, \quad n \geq L \]

with \( H_1 = 1, \) \( H_{n+1} = c_1 H_n + c_2 H_{n-1} + \cdots + c_n H_1 + 1, \quad n < L, \)

coefficients \( c_i \geq 0; \) \( c_1, c_L > 0 \) if \( L \geq 2; \) \( c_1 > 1 \) if \( L = 1. \)

- Zeckendorf: Every positive integer can be written uniquely as \( \sum a_i H_i \) with natural constraints on the \( a_i \)'s (e.g. cannot use the recurrence relation to remove any summand).
- Lekkerkerker
- Central Limit Type Theorem
Generalizing Lekkerkerkerker

Generalized Lekkerkerkerker’s Theorem

The average number of summands in the generalized Zeckendorf decomposition for integers in $[H_n, H_{n+1})$ tends to $Cn + d$ as $n \to \infty$, where $C > 0$ and $d$ are computable constants determined by the $c_i$’s.

$$C = -\frac{y'(1)}{y(1)} = \frac{\sum_{m=0}^{L-1} (s_m + s_{m+1} - 1)(s_{m+1} - s_m) y^m(1)}{2 \sum_{m=0}^{L-1} (m + 1)(s_{m+1} - s_m) y^m(1)}.$$  

$$s_0 = 0, s_m = c_1 + c_2 + \cdots + c_m.$$  

$y(x)$ is the root of $1 - \sum_{m=0}^{L-1} \sum_{j=s_m}^{s_{m+1}-1} x^j y^{m+1}.$  

$y(1)$ is the root of $1 - c_1 y - c_2 y^2 - \cdots - c_L y^L$. 
Central Limit Type Theorem

As \( n \to \infty \), the distribution of the number of summands, i.e., \( a_1 + a_2 + \cdots + a_m \) in the generalized Zeckendorf decomposition \( \sum_{i=1}^{m} a_i H_i \) for integers in \([H_n, H_{n+1})\) is Gaussian.
Example: the Special Case of $L = 1$, $c_1 = 10$

$H_{n+1} = 10H_n$, $H_1 = 1$, $H_n = 10^{n-1}$.

- Legal decomposition is decimal expansion: $\sum_{i=1}^{m} a_i H_i$:
  
  $a_i \in \{0, 1, \ldots, 9\}$ ($1 \leq i < m$), $a_m \in \{1, \ldots, 9\}$.

- For $N \in [H_n, H_{n+1})$, $m = n$, i.e., first term is $a_n H_n = a_n 10^{n-1}$.

- $A_i$: the corresponding random variable of $a_i$. The $A_i$'s are independent.

- For large $n$, the contribution of $A_n$ is immaterial. $A_i$ ($1 \leq i < n$) are identically distributed random variables with mean 4.5 and variance 8.25.

- Central Limit Theorem: $A_2 + A_3 + \cdots + A_n \to$ Gaussian with mean $4.5n + O(1)$ and variance $8.25n + O(1)$. 

Far-difference Representation

**Theorem (Alpert, 2009) (Analogue to Zeckendorf)**

Every integer can be written uniquely as a sum of the \( \pm F_n \)'s, such that every two terms of the same (opposite) sign differ in index by at least 4 (3).

**Example:** \( 1900 = F_{17} - F_{14} - F_{10} + F_6 + F_2 \).

\( K \): # of positive terms, \( L \): # of negative terms.

**Generalized Lekkerkerkerker’s Theorem**

As \( n \to \infty \), \( E[K] \) and \( E[L] \) \( \to n/10 \).

\[
E[K] - E[L] = \phi/2 \approx 0.809.
\]

**Central Limit Type Theorem**

As \( n \to \infty \), \( K \) and \( L \) converges to a bivariate Gaussian.

\[
\text{corr}(K, L) = -(21 - 2\phi)/(29 + 2\phi) \approx -0.551,
\]

\[
\phi = \frac{\sqrt{5} + 1}{2}.
\]
**Binet’s Formula**

\[
F_1 = F_2 = 1; \quad F_n = \frac{1}{\sqrt{5}} \left[ \left( \frac{1 + \sqrt{5}}{2} \right)^n - \left( \frac{-1 + \sqrt{5}}{2} \right)^n \right].
\]

- **Recurrence relation:** \( F_{n+1} = F_n + F_{n-1} \) \hspace{1cm} (1)
- **Generating function:** \( g(x) = \sum_{n>0} F_n x^n \).

\( (1) \Rightarrow \sum_{n\geq2} F_{n+1} x^{n+1} = \sum_{n\geq2} F_{n} x^{n+1} + \sum_{n\geq2} F_{n-1} x^{n+1} \)

\[ \Rightarrow \sum_{n\geq3} F_{n} x^{n} = \sum_{n\geq2} F_{n} x^{n+1} + \sum_{n\geq1} F_{n} x^{n+2} \]

\[ \Rightarrow \sum_{n\geq3} F_{n} x^{n} = x \sum_{n\geq2} F_{n} x^{n} + x^2 \sum_{n\geq1} F_{n} x^{n} \]

\[ \Rightarrow g(x) - F_1 x - F_2 x^2 = x(g(x) - F_1 x) + x^2 g(x) \]

\[ \Rightarrow g(x) = \frac{x}{1 - x - x^2}. \]
Partial Fraction Expansion (Example: Binet’s Formula)

• Generating function: \( g(x) = \sum_{n>0} F_n x^n = \frac{x}{1-x-x^2}. \)

• Partial fraction expansion:

\[
\Rightarrow g(x) = \frac{x}{1-x-x^2} = \frac{1}{\sqrt{5}} \left( \frac{\frac{1+\sqrt{5}}{2}x}{1 - \frac{1+\sqrt{5}}{2}x} - \frac{\frac{-1+\sqrt{5}}{2}x}{1 - \frac{-1+\sqrt{5}}{2}x} \right).
\]

Coefficient of \( x^n \) (power series expansion):

\[
F_n = \frac{1}{\sqrt{5}} \left[ \left( \frac{1+\sqrt{5}}{2} \right)^n - \left( \frac{-1+\sqrt{5}}{2} \right)^n \right] - \text{Binet’s Formula!}
\]

(using geometric series: \( \frac{1}{1-r} = 1 + r + r^2 + r^3 + \cdots \)).
Differentiating Identities and Method of Moments

- **Differentiating identities**
  
  Example: Given a random variable $X$ such that
  \[ \Pr(X = 1) = \frac{1}{2}, \Pr(X = 2) = \frac{1}{4}, \Pr(X = 3) = \frac{1}{8}, \ldots \]
  then what’s the mean of $X$ (i.e., $E[X]$)?

  **Solution:** Let $f(x) = \frac{1}{2}x + \frac{1}{4}x^2 + \frac{1}{8}x^3 + \cdots = \frac{1}{1-x/2} - 1$.
  
  \[
  f'(x) = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4}x + 3 \cdot \frac{1}{8}x^2 + \cdots.
  \]
  
  \[
  f'(1) = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{8} + \cdots = E[X].
  \]

- **Method of moments:** Random variables $X_1, X_2, \ldots$
  
  If the $\ell^{th}$ moment $E[X_n^\ell]$ converges to that of the standard normal distribution $(\forall \ell)$, then $X_n$ converges to a Gaussian.

**Standard normal distribution:**

$2m^{th}$ moment: $(2m - 1)!! = (2m - 1)(2m - 3) \cdots 1$, 

$(2m - 1)^{th}$ moment: $0$. 

New Approach: Case of Fibonacci Numbers

\[ \rho_{n,k} = \# \{ N \in [F_n, F_{n+1}) : \text{the Zeckendorf decomposition of } N \text{ has exactly } k \text{ summands} \}. \]

- **Recurrence relation:**
  \[ N \in [F_{n+1}, F_{n+2}): N = F_{n+1} + F_t + \cdots, t \leq n - 1. \]
  \[ \rho_{n+1,k+1} = \rho_{n-1,k} + \rho_{n-2,k} + \cdots \]
  \[ \rho_{n,k+1} = \rho_{n-2,k} + \rho_{n-3,k} + \cdots \]
  \[ \Rightarrow \rho_{n+1,k+1} = \rho_{n,k+1} + \rho_{n-1,k}. \]

- **Generating function:** \[ \sum_{n,k>0} \rho_{n,k} x^k y^n = \frac{y}{1-y-xy^2}. \]
- **Partial fraction expansion:**
  \[ \frac{y}{1-y-xy^2} = -\frac{y}{y_1(x) - y_2(x)} \left( \frac{1}{y - y_1(x)} - \frac{1}{y - y_2(x)} \right) \]
  where \( y_1(x) \) and \( y_2(x) \) are the roots of \( 1 - y - xy^2 = 0. \)

**Coefficient of** \( y^n: g(x) = \sum_{k>0} \rho_{n,k} x^k. \)
New Approach: Case of Fibonacci Numbers (Continued)

\( K_n \): the corresponding random variable associated with \( k \).

\[ g(x) = \sum_{k>0} p_{n,k} x^k. \]

- Differentiating identities:

\[ g(1) = \sum_{k>0} p_{n,k} = F_{n+1} - F_n, \]

\[ g'(x) = \sum_{k>0} kp_{n,k} x^{k-1}, \quad g'(1) = g(1) E[K_n], \]

\[ (xg'(x))' = \sum_{k>0} k^2 p_{n,k} x^{k-1}, \]

\[ (xg'(x))'|_{x=1} = g(1) E[K_n^2], \]

\[ (x(xg'(x)))'|_{x=1} = g(1) E[K_n^3], \ldots \]

Similar results hold for the centralized \( K_n \):

\[ K'_n = K_n - E[K_n]. \]

- Method of moments (for normalized \( K'_n \)):

\[ E[(K'_n)^{2m}]/(SD(K'_n))^{2m} \to (2m - 1)!!, \]

\[ E[(K'_n)^{2m-1}]/(SD(K'_n))^{2m-1} \to 0. \quad \Rightarrow \quad K_n \to \text{Gaussian}. \]
New Approach: General Case

Let \( p_{n,k} = \# \{ N \in [H_n, H_{n+1}) : \text{the generalized Zeckendorf decomposition of } N \text{ has exactly } k \text{ summands} \} \).

- **Recurrence relation:**
  
  Fibonacci: \( p_{n+1,k+1} = p_{n,k+1} + p_{n,k} \).
  
  General: \( p_{n+1,k} = \sum_{m=0}^{L-1} \sum_{j=s_m}^{s_{m+1}-1} p_{n-m,k-j} \).
  
  where \( s_0 = 0, s_m = c_1 + c_2 + \cdots + c_m \).

- **Generating function:**
  
  Fibonacci: \( \frac{y}{1-y-xy^2} \).
  
  General:

  \[
  \sum_{n \leq L} p_{n,k} x^k y^n - \sum_{m=0}^{L-1} \sum_{j=s_m}^{s_{m+1}-1} x^j y^{m+1} \sum_{n<L-m} p_{n,k} x^k y^n \left/ \frac{1 - \sum_{m=0}^{L-1} \sum_{j=s_m}^{s_{m+1}-1} x^j y^{m+1}}{1} \right. 
  \]
New Approach: General Case (Continued)

- Partial fraction expansion:
  
  **Fibonacci:** 
  \[ \frac{y}{y_1(x) - y_2(x)} \left( \frac{1}{y - y_1(x)} - \frac{1}{y - y_2(x)} \right). \]

  **General:** 
  
  \[
  - \frac{1}{\sum_{j=s_{L-1}}^{s_{L-1}} x^j} \sum_{i=1}^{L} \frac{B(x, y)}{(y - y_i(x)) \prod_{j \neq i} (y_j(x) - y_i(x))}. 
  \]

  \[ B(x, y) = \sum_{n \leq L} p_{n,k} x^k y^n - \sum_{m=0}^{L-1} \sum_{j=s_m}^{s_{m+1}-1} x^j y^{m+1} \sum_{n < L-m} p_{n,k} x^k y^n, \]

  \[ y_i(x): \text{root of } 1 - \sum_{m=0}^{L-1} \sum_{j=s_m}^{s_{m+1}-1} x^j y^{m+1} = 0. \]

  **Coefficient of** \( y^n: \) 
  \[ g(x) = \sum_{n,k>0} p_{n,k} x^k. \]

- Differentiating identities

- Method of moments: implies \( K_n \rightarrow \text{Gaussian}. \)