From Fibonacci Quilts to Benford's Law through Zeckendorf Decompositions

Steven J. Miller (sjm1@williams.edu)

http://www.williams.edu/Mathematics/sjmiller/public_html

AMS Special Session on Difference Equations, San Antonio, January 10, 2015



Introduction

Goals of the Talk

- Generalize Zeckendorf decompositions
- Analyze gaps (in the bulk and longest)
- Patterns and new recurrences
- Some open problems (if time permits)



Collaborators and Thanks

Collaborators:

Gaps (Bulk, Individual, Longest): Olivia Beckwith, Amanda Bower, Louis Gaudet, Rachel Insoft, Shiyu Li, Philip Tosteson.

Kentucky Sequence, Fibonacci Quilt: Joint with Minerva Catral, Pari Ford, Pamela Harris & Dawn Nelson.

Benfordness: Andrew Best, Patrick Dynes, Xixi Edelsbunner, Brian McDonald, Kimsy Tor, Caroline Turnage-Butterbaugh & Madeleine Weinstein.

Supported by:

NSF Grants DMS1265673, DMS0970067, DMS1347804 and DMS0850577, AIM and Williams College.

Previous Results

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

Zeckendorf's Theorem

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

Example:

$$2014 = 1597 + 377 + 34 + 5 + 1 = F_{16} + F_{13} + F_{8} + F_{4} + F_{1}$$
.

Lekkerker'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.

Old Results

Intro

Central Limit Type Theorem

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

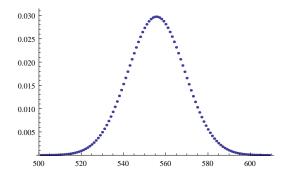


Figure: Number of summands in $[F_{2010}, F_{2011})$; $F_{2010} \approx 10^{420}$.

Intro

Definition of Benford's Law

A dataset is said to follow Benford's Law (base *B*) if the probability of observing a first digit of *d* is

$$\log_B\left(1+\frac{1}{d}\right)$$
.

- More generally probability a significant at most s is $log_B(s)$, where $x = S_B(x)10^k$ with $S_B(x) \in [1, B)$ and $k \in \mathbb{Z}$.
- Find base 10 about 30.1% of the time start with a 1, only 4.5% start with a 9.

Gaps

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

Distribution of Gaps

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

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

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

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

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

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

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

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

Distribution of Gaps

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

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

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

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

Individual: Similar questions about gaps for a fixed $m \in [F_n, F_{n+1})$: distribution of gaps, longest gap.

New Results: Bulk Gaps: $m \in [F_n, F_{n+1})$ and $\phi = \frac{1+\sqrt{5}}{2}$

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

Theorem (Zeckendorf Gap Distribution)

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

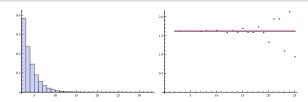


Figure: Distribution of gaps in $[F_{2010}, F_{2011})$; $F_{2010} \approx 10^{420}$.

Kentucky and Quilts

Fair coin: largest gap tightly concentrated around $\log n / \log 2$.

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

Prob
$$(L_n(m) \le f(n)) \approx e^{-e^{\log n - f(n) \cdot \log \phi}}$$

•
$$\mu_n = \frac{\log\left(\frac{\phi^2}{\phi^2+1}\right)n}{\log\phi} + \frac{\gamma}{\log\phi} - \frac{1}{2} + \text{Small Error.}$$

• If f(n) grows **slower** (resp. **faster**) than $\log n/\log \phi$, then $\operatorname{Prob}(L_n(m) \leq f(n))$ goes to **0** (resp. **1**).

Intro

Main Results

Theorem (Distribution of Bulk Gaps (SMALL 2012))

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 $c_i \ge 1$ for all $1 \le i \le L$. Then

$$P(g) \; = \; \begin{cases} 1 - (\frac{a_1}{C_{Lek}})(2\lambda_1^{-1} + a_1^{-1} - 3) & \text{: } g = 0 \\ \lambda_1^{-1}(\frac{1}{C_{Lek}})(\lambda_1(1 - 2a_1) + a_1) & \text{: } g = 1 \\ (\lambda_1 - 1)^2 \left(\frac{a_1}{C_{Lek}}\right)\lambda_1^{-g} & \text{: } g \geq 2. \end{cases}$$

15

Main Results

Theorem (Longest Gap (SMALL 2012))

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

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

Kentucky Sequence and Quilts with Minerva Catral, Pari Ford, Pamela Harris & Dawn Nelson

Rule: (s, b)-Sequence: Bins of length b, and:

- cannot take two elements from the same bin, and
- if have an element from a bin, cannot take anything from the first s bins to the left or the first s to the right.

Rule: (s, b)-Sequence: Bins of length b, and:

- cannot take two elements from the same bin, and
- if have an element from a bin, cannot take anything from the first s bins to the left or the first s to the right.

Fibonaccis: These are (s, b) = (1, 1).

Kentucky: These are (s, b) = (1, 2).

[1, 2], [3, 4], [5, 8], [11, 16], [21, 32], [43, 64], [85, 128].

Rule: (s, b)-Sequence: Bins of length b, and:

- cannot take two elements from the same bin, and
- if have an element from a bin, cannot take anything from the first s bins to the left or the first s to the right.

Fibonaccis: These are (s, b) = (1, 1).

Kentucky: These are (s, b) = (1, 2).

[1, 2], [3, 4], [5, 8], [11, 16], [21, 32], [43, 64], [85, 128].

• $a_{2n} = 2^n$ and $a_{2n+1} = \frac{1}{3}(2^{2+n} - (-1)^n)$: $a_{n+1} = a_{n-1} + 2a_{n-3}, a_1 = 1, a_2 = 2, a_3 = 3, a_4 = 4$

Kentucky and Quilts

Rule: (s, b)-Sequence: Bins of length b, and:

- cannot take two elements from the same bin, and
- if have an element from a bin, cannot take anything from the first s bins to the left or the first s to the right.

Fibonaccis: These are (s, b) = (1, 1).

Kentucky: These are (s, b) = (1, 2).

[1, 2], [3, 4], [5, 8], [11, 16], [21, 32], [43, 64], [85, 128].

- $a_{2n} = 2^n$ and $a_{2n+1} = \frac{1}{3}(2^{2+n} (-1)^n)$: $a_{n+1} = a_{n-1} + 2a_{n-3}, a_1 = 1, a_2 = 2, a_3 = 3, a_4 = 4.$
- \bullet $a_{n+1} = a_{n-1} + 2a_{n-3}$: New as leading term 0.

Gaussian Behavior

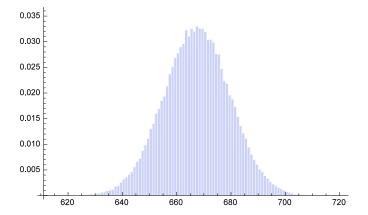


Figure: Plot of the distribution of the number of summands for 100,000 randomly chosen $m \in [1, a_{4000}) = [1, 2^{2000})$ (so m has on the order of 602 digits).

Gaps

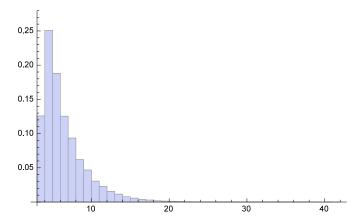


Figure: Plot of the distribution of gaps for 10,000 randomly chosen $m \in [1, a_{400}) = [1, 2^{200})$ (so m has on the order of 60 digits).

Gaps

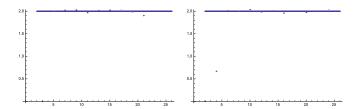
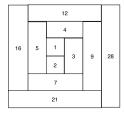


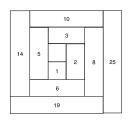
Figure: Plot of the distribution of gaps for 10,000 randomly chosen $m \in [1, a_{400}) = [1, 2^{200})$ (so m has on the order of 60 digits). Left (resp. right): ratio of adjacent even (resp odd) gap probabilities.

Again find geometric decay, but parity issues so break into even and odd gaps.

The Fibonacci (or Log Cabin) Quilt: Work in Progress



1, 2, 3, 4, 5, 7, 9, 12, 16, 21, 28, ...



Generalizations

1, 2, 3, 5, 6, 8, 10, 14, 19, 25, 33, ...

- $a_{n+1} = a_{n-1} + a_{n-2}$, non-uniqueness (average number of decompositions grows exponentially).
- In process of investigating Gaussianity, Gaps,
 K_{min}, K_{ave}, K_{max}, K_{greedy}.

Average Number of Representations

- d_n : the number of FQ-legal decompositions using only elements of $\{a_1, a_2, \ldots, a_n\}$.
- c_n requires a_n to be used, b_n requires a_n and a_{n-2} to be used.

n	d _n	Cn	b_n	a _n
1	2	1	0	1
2	3	1	0	2 3
3	4	1	0	
4	2 3 4 6 8	2	1	4
2 3 4 5 6 7	8	2	1	4 5 7
6	11	3	1	
7	15	4	1	9
8	21	2 2 3 4 6	2	12
9	30	9	3	16

Table: First few terms. Find $d_n = d_{n-1} + d_{n-2} - d_{n-3} + d_{n-5} - d_{n-9}$, implying $d_{\text{FQ;ave}}(n) \approx C \cdot 1.05459^n$.

Greedy Algorithm

 h_n : number of integers from 1 to $a_{n+1} - 1$ where the greedy algorithm successfully terminates in a legal decomposition.

n	a _n	h _n	$ ho_{n}$
1	1	1	100.0000
2	2	2	100.0000
3	3	3	100.0000
4	4	4	100.0000
5	5	5	83.3333
6	7	7	87.5000
10	21	25	92.5926
11	28	33	91.6667
17	151	184	92.4623

Table: First few terms, yields $h_n = h_{n-1} + h_{n-5} + 1$ and percentage converges to about 0.92627.

Benfordness in Interval

Joint with Andrew Best, Patrick Dynes, Xixi Edelsbunner, Brian McDonald, Kimsy Tor, Caroline Turnage-Butterbaugh and Madeleine Weinstein

Benfordness in Interval

Theorem (SMALL 2014): Benfordness in Interval

The distribution of the summands in the Zeckendorf decompositions, averaged over the entire interval $[F_n, F_{n+1})$, follows Benford's Law.

Benfordness in Interval

0000 000

Theorem (SMALL 2014): Benfordness in Interval

The distribution of the summands in the Zeckendorf decompositions, averaged over the entire interval $[F_n, F_{n+1})$, follows Benford's Law.

Example

Looking at the interval $[F_5, F_6) = [8, 13)$

$$8 = 8$$
 $= F_5$
 $9 = 8 + 1 = F_5 + F_1$
 $10 = 8 + 2 = F_5 + F_2$
 $11 = 8 + 3 = F_5 + F_3$
 $12 = 8 + 3 + 1 = F_5 + F_3 + F_1$

Density of S

For a subset Sof the Fibonacci numbers, define the density q(S, n) of S over the interval $[1, F_n]$ by

$$q(S,n) = \frac{\#\{F_j \in S \mid 1 \leq j \leq n\}}{n}.$$

Asymptotic Density

If $\lim_{n\to\infty} q(S,n)$ exists, define the asymptotic density q(S) by

$$q(S) = \lim_{n \to \infty} q(S, n).$$

21

Needed Input

Intro

Let S_d be the subset of the Fibonacci numbers which share a fixed digit d where $1 \le d < B$.

Theorem: Fibonacci Numbers Are Benford

$$q(S_d) = \lim_{n \to \infty} q(S_d, n) = \log_B \left(1 + \frac{1}{d}\right).$$

Proof: Binet's formula, Kronecker's theorem on equidistribution of $n\alpha \mod 1$ for $\alpha \notin \mathbb{Q}$.

Random Variables

Intro

Random Variable from Decompositions

Let $X(I_n)$ be a random variable whose values are the the Fibonacci numbers in $[F_1, F_n)$ and probabilities are how often they occur in decompositions of $m \in I_n$:

$$P\{X(I_n)=F_k\}:= egin{cases} rac{F_{k-1}F_{n-k-2}}{\mu_nF_{n-1}}, & ext{if } 1\leq k\leq n-2 \ rac{1}{\mu_n}, & ext{if } k=n \ 0, & ext{otherwise,} \end{cases}$$

where μ_n is the average number of summands in Zeckendorf decompositions of integers in the interval $[F_n, F_{n+1})$.

Approximations

Gaps

00000 000

Estimate for $P\{X(I_n) = F_k\}$

$$P\{X(I_n) = F_k\} = \frac{1}{\mu_n \phi \sqrt{5}} + O\left(\phi^{-2k} + \phi^{-2n+2k}\right).$$

Constant Fringes Negligible

For any *r* (which may depend on *n*):

$$\sum_{r < k < n-r} P\{X(I_n) = F_k\} = 1 - r \cdot O\left(\frac{1}{n}\right).$$

34

Estimating $P\{X(I_n) \in S\}$

Set
$$r := \left| \frac{\log n}{\log \phi} \right|$$
.

Density of S over Zeckendorf Summands

We have

$$P\{X(I_n) \in S\} = \frac{nq(S)}{\mu_n\phi\sqrt{5}} + o(1) \rightarrow q(s).$$

35

Remark

- Stronger result than Benfordness of Zeckendorf summands.
- Global property of the Fibonacci numbers can be carried over locally into the Zeckendorf summands.
- If we have a subset of the Fibonacci numbers S with asymptotic density q(S), then the density of the set S over the Zeckendorf summands will converge to this asymptotic density.

Benfordness of Random and Zeckendorf Decompositions
Joint with Andrew Best, Patrick Dynes, Xixi Edelsbunner, Brian
McDonald, Kimsy Tor, Caroline Turnage-Butterbaugh and
Madeleine Weinstein

Random Decompositions

Theorem 2 (SMALL 2014): Random Decomposition

If we choose each Fibonacci number with probability q, disallowing the choice of two consecutive Fibonacci numbers, the resulting sequence follows Benford's law.

Example: n = 10

$$F_1 + F_2 + F_3 + F_4 + F_5 + F_6 + F_7 + F_8 + F_9 + F_{10}$$

$$= 2 + 8 + 21 + 89$$

$$= 120$$

Intro

Choosing a Random Decomposition

Select a random subset A of the Fibonaccis as follows:

- Fix $q \in (0, 1)$.
- Let $A_0 := \emptyset$.
- For n > 1, if $F_{n-1} \in A_{n-1}$, let $A_n := A_{n-1}$, else

$$A_n = \begin{cases} A_{n-1} \cup \{F_n\} & \text{with probability } q \\ A_{n-1} & \text{with probability } 1 - q. \end{cases}$$

• Let $A := \bigcup_n A_n$.

Main Result

Theorem

With probability 1, A (chosen as before) is Benford.

Stronger claim: For any subset S of the Fibonaccis with density d in the Fibonaccis, $S \cap A$ has density d in A with probability 1.

Preliminaries

Lemma

The probability that $F_k \in A$ is

$$p_k = \frac{q}{1+q} + O(q^k).$$

Using elementary techniques, we get

Lemma

Define $X_n := \#A_n$. Then

$$E[X_n] = \frac{nq}{1+q} + O(1)$$

$$Var(X_n) = O(n).$$

41

Expected Value of Y_n

Define $Y_{n,S} := \#A_n \cap S$. Using standard techniques, we get

Lemma

$$\mathbb{E}[Y_n] = \frac{nqd}{1+q} + o(n).$$

$$\operatorname{Var}(Y_{n,S}) = o(n^2).$$

Define $Y_{n,S} := \#A_n \cap S$. Using standard techniques, we get

Lemma

Gaps

0000 000

$$\mathbb{E}[Y_n] = \frac{nqd}{1+q} + o(n).$$

$$Var(Y_{n,S}) = o(n^2).$$

Immediately implies with probability 1 + o(1)

$$Y_{n,S} = \frac{nqd}{1+q} + o(n), \lim_{n\to\infty} \frac{Y_{n,S}}{X_n} = d.$$

Hence $A \cap S$ has density d in A, completing the proof.

Zeckendorf Decompositions and Benford's Law

Theorem (SMALL 2014): Benfordness of Decomposition

If we pick a random integer in $[0, F_{n+1})$, then with probability 1 as $n \to \infty$ its Zeckendorf decomposition converges to Benford's Law.

Proof of Theorem

- Choose integers randomly in $[0, F_{n+1})$ by random decomposition model from before.
- Choose $m=F_{a_1}+F_{a_2}+\cdots+F_{a_\ell}\in[0,F_{n+1})$ with probability

$$p_m = \begin{cases} q^{\ell}(1-q)^{n-2\ell} & \text{if } a_{\ell} \leq n \\ q^{\ell}(1-q)^{n-2\ell+1} & \text{if } a_{\ell} = n. \end{cases}$$

• Key idea: Choosing $q = 1/\varphi^2$, the previous formula simplifies to

$$p_m = \begin{cases} \varphi^{-n} & \text{if } m \in [0, F_n) \\ \varphi^{-n-1} & \text{if } m \in [F_n, F_{n+1}), \end{cases}$$

use earlier results.

References

References

References

- Bower, Insoft, Li, Miller and Tosteson, The Distribution of Gaps between Summands in Generalized Zeckendorf Decompositions, preprint. http://arxiv.org/pdf/1402.3912.
- Beckwith, Bower, Gaudet, Insoft, Li, Miller and Tosteson: The Average Gap Distribution for Generalized Zeckendorf Decompositions: The Fibonacci Quarterly 51 (2013), 13–27. http://arxiv.org/abs/1208.5820
- Kologlu, Kopp, Miller and Wang: On the number of summands in Zeckendorf decompositions: Fibonacci Quarterly 49 (2011), no. 2, 116–130. http://arxiv.org/pdf/1008.3204
- Miller and Wang: From Fibonacci numbers to Central Limit Type Theorems: Journal of Combinatorial Theory, Series A 119 (2012), no. 7, 1398–1413. http://arxiv.org/pdf/1008.3202
- Miller and Wang: Survey: Gaussian Behavior in Generalized Zeckendorf Decompositions: To appear in CANT 2011 Proceedings. http://arxiv.org/pdf/1107.2718

Generalizations

Positive Linear Recurrence Sequences

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

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

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.

Intro

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:

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

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:

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

A geometric series!

Generalized Generating Function

Let $f > j_L$. The number of $x \in [H_n, H_{n+1})$, with longest gap < fis 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 i_i are defined as above.

What are the extra obstructions?

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

Kentucky and Quilts

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

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

and

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

where $\epsilon_i(n)$ tends to zero in the limit, and Small Error comes from the Euler-Maclaurin Formula.