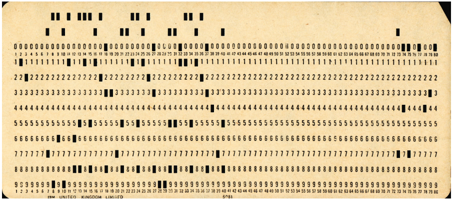
Inrtoduwtion to Erorr Dwtetcion and Erorr Czrrectmon

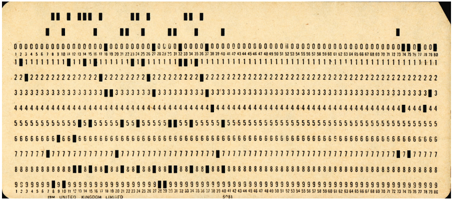
Setevn .J Mzlwer
sjm1@williams.edu
http://www.williams.edu/Mathematics/sjmiller



Gzoanga Univrseity, Setpmebre 52, 2025

Introduction to Error Detection and Error Correction

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



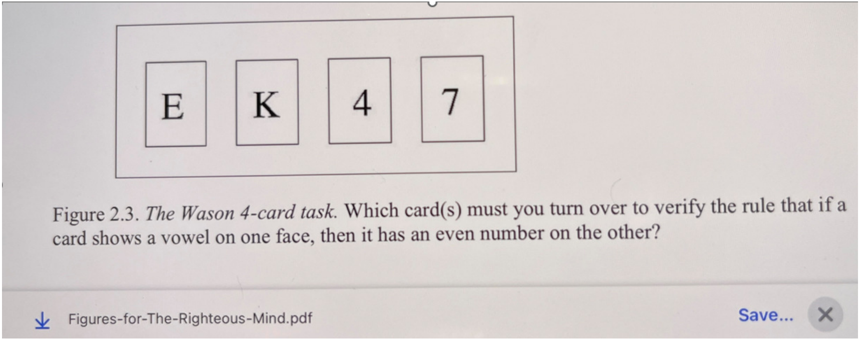
Gonzaga University, September 25, 2025

Introduction

Goals

- Use math from classes (number theory, group theory).
- Discuss challenges in real world applications.
- Creating research questions.

Wason 4 Card Test



Cryptography Basics

Enough to send 0's and 1's:

- ◇ A = 00000, B = 00001, C = 00010, . . .
  Z = 11010, 0 = 11011, 1 = 11100, ...

Two major issues:

- ◇ Transmit message so only desired recipient can read.
- ◇ Ensure correct message received.

Cryptography Basics

Enough to send 0's and 1's:

◇ A = 00000, B = 00001, C = 00010, . . .
Z = 11010, 0 = 11011, 1 = 11100, ...

Two major issues:

◇ Transmit message so only desired recipient can read.

◇ Ensure correct message received.

Bit Error Dangers: RSA

If receive wrong bit in RSA, message completely different.

Secret: $p = 15217$, $q = 17569$, $d = 80998505$.

Bit Error Dangers: RSA

If receive wrong bit in RSA, message completely different.

Secret: $p = 15217$, $q = 17569$, $d = 80998505$.
Public: $N = pq = 267347473$, $e = 3141593$.

Bit Error Dangers: RSA

If receive wrong bit in RSA, message completely different.

Secret: $p = 15217$, $q = 17569$, $d = 80998505$.
Public: $N = pq = 267347473$, $e = 3141593$.
Note: $ed = 1 \bmod (p-1)(q-1)$.

Bit Error Dangers: RSA

If receive wrong bit in RSA, message completely different.

Secret: $p = 15217$, $q = 17569$, $d = 80998505$.
Public: $N = pq = 267347473$, $e = 3141593$.
Note: $ed = 1 \bmod (p-1)(q-1)$.
Message: $M = 195632041$, send $M^e \bmod N$ or
$X = 121209473$.
Decrypt: $X^d \bmod N$ or 195632041.

Bit Error Dangers: RSA

If receive wrong bit in RSA, message completely different.

Secret: $p = 15217$, $q = 17569$, $d = 80998505$.
Public: $N = pq = 267347473$, $e = 3141593$.
Note: $ed = 1 \bmod (p-1)(q-1)$.
Message: $M = 195632041$, send $M^e \bmod N$ or
$X = 121209473$.
Decrypt: $X^d \bmod N$ or 195632041.

Imagine receive $\widetilde{X} = 121209483$.
Message 195632041
Decrypts 121141028, only two digits are the same!

Outline

Will concentrate on Error Detection and Correction.

- How do you detect an error?
- How do you fix an error?

Check Digit

Check Digit

If easy to read again, just need to detect error.

Check Digit

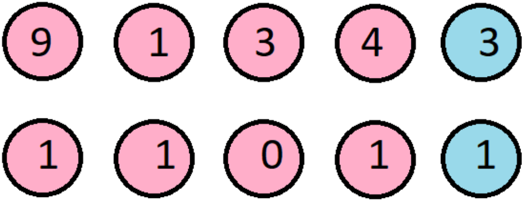
If easy to read again, just need to detect error.

Think scanner at a supermarket....

Check Digit

If easy to read again, just need to detect error.

Think scanner at a supermarket....

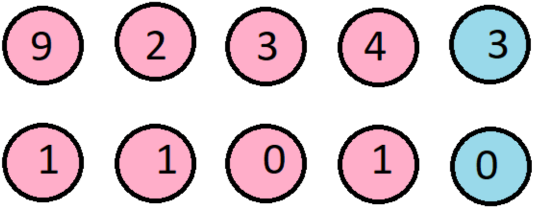


Last digit makes sum 0 mod 10 (or 0 mod 2).

Check Digit

If easy to read again, just need to detect error.

Think scanner at a supermarket....



Last digit makes sum 0 mod 10 (or 0 mod 2).

Next Steps

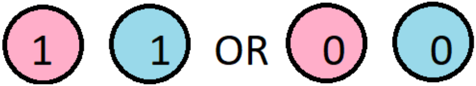
More involved methods detecting more: The Verhoeff algorithm catches single digit errors and flipping adjacent digits: https://en.wikipedia.org/wiki/Verhoeff_algorithm.

Want to detect where the error is:

Next Steps

More involved methods detecting more: The Verhoeff algorithm catches single digit errors and flipping adjacent digits: https://en.wikipedia.org/wiki/Verhoeff_algorithm.

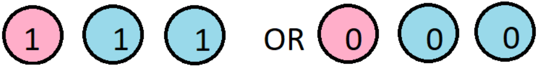
Want to detect where the error is: Tell me twice!



Majority Rules

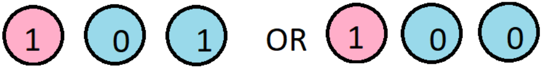
Tell Me Three Times

Tell Me Three Times detects and *probably* corrects (need probability of an error small).



Tell Me Three Times

Tell Me Three Times detects and *probably* corrects (need probability of an error small).



Tell Me Three Times

Crucially uses binary outcome: https://www.youtube.com/watch?v=RerJWv5vwxc and https://www.youtube.com/watch?v=vWCGs27_xPI.

What is the problem with this method?

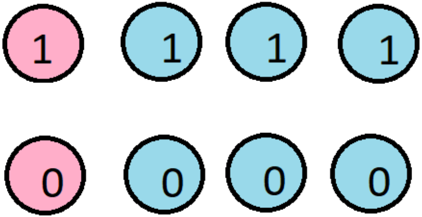
Tell Me Three Times

Crucially uses binary outcome: https://www.youtube.com/watch?v=RerJWv5vwxc and https://www.youtube.com/watch?v=vWCGs27_xPI.

What is the problem with this method?
Only one-third is information, if two errors is wrong!

What else can we do? Is it better? With respect to what metric?

Tell Me n Times

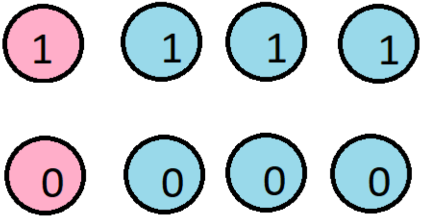


Tell Me Four Times: only 25% of message is data (general case just $1/n$).

Want to correct errors but still send a lot of information.

What's a success?

Tell Me *n* Times



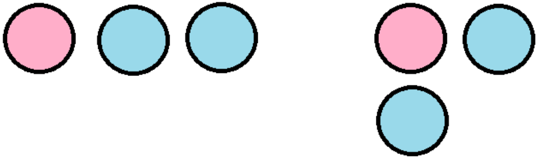
Tell Me Four Times: only 25% of message is data (general case just $1/n$).

Want to correct errors but still send a lot of information.

What's a success? Greater than 50% is data.

Tell Me Three Times (revisited)

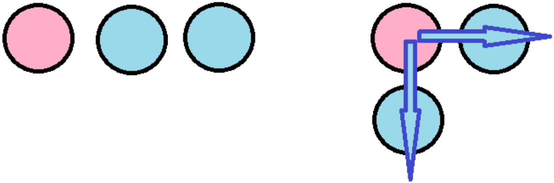
Let's revisit Tell Me Three Times:



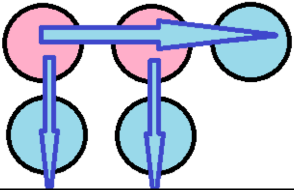
How should we do two data points?
How many check digits do you expect?

Tell Me Three Times (revisited)

Let's revisit Tell Me Three Times:

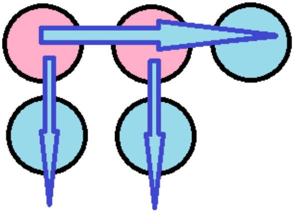


How should we do two data points?
How many check digits do you expect?



Two of Five

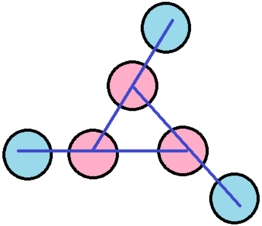
This is better: 2 of 5 or 40% of message is data!

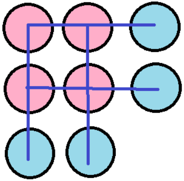


Unfortunately still below 50%.

How many data points should we try next: 3, 4, 5, ...?
Suggestions?

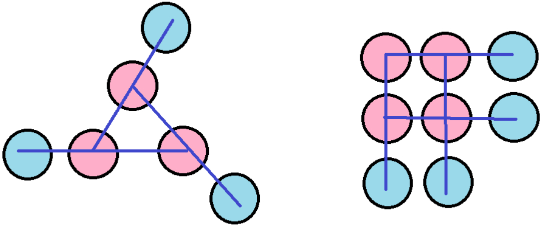
Three and Four Bits of Data





Which is better?

Three and Four Bits of Data

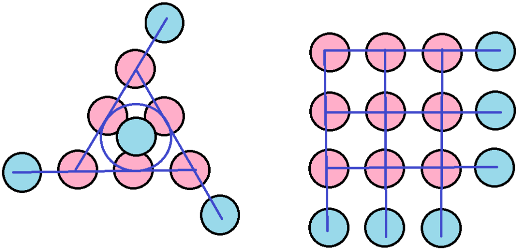


Which is better? Both 50% but fewer needed with triangle.

What should we do next: 5, 6, 7, 8, 9, ...?

Triangle and Square Numbers

$T_n = n(n+1)/2$ and $S_n = n^2$.

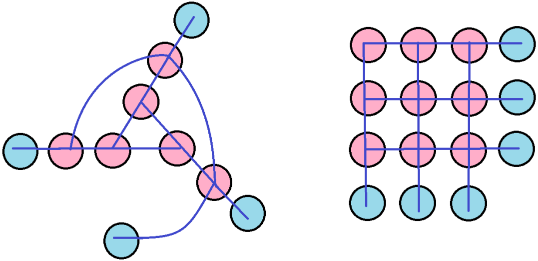


Both give 60% of the message is data. Can we continue?

Data on exactly two lines, check bits on one.

Triangle and Square Numbers

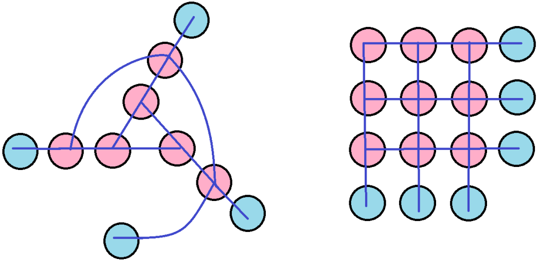
$T_n = n(n+1)/2$ and $S_n = n^2$.



Both give 60% of the message is data. Can we continue?

Data on exactly two lines, check bits on one.

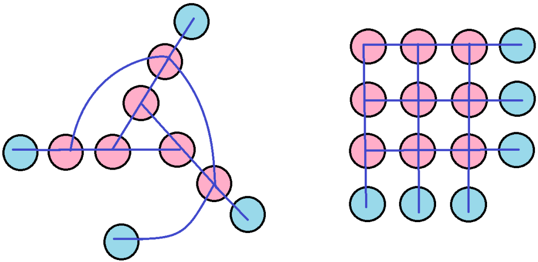
Triangle and Square Systems



Triangle: $T_n = n(n+1)/2$ data, $n+1$ check, so $(n+2)(n+1)/2$ bits total and $n/(n+2)$ information.

Square: $S_n = n^2$ data, $2n$ check, so $n^2 + 2n$ bits total and $n/(n+2)$ information.

Triangle and Square Systems

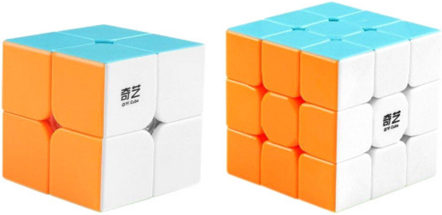


Can get as high a percentage information as desire, at a cost of longer string (and thus more likely to have two errors).

Generalizations

What is a better geometry to use?

Generalizations

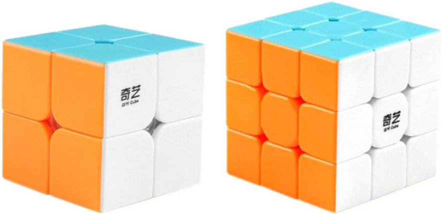


$2 \times 2 \times 2$: 8 data points, 6 check bits (for planes): info is $8/14 \approx 57\%$.

$3 \times 3 \times 3$: 27 data points, 9 check bits (for planes): info is $27/36 = 75\%$.

For 6×6 data square info is $36/48 = 75\%$, for T_7 is $28/36 \approx 77.78\%$.

Generalizations

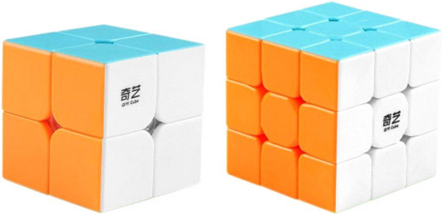


$4 \times 4 \times 4$: 64 data points, 12 check bits: info is $64/76 \approx 84.21\%$.

For 9×9 data square info is $81/99 \approx 81.82\%$.

For T_{11} triangle: 66 data points, info is $66/79 \approx 83.54\%$.

Generalizations



$n \times n \times n$: n^3 data points, $3n$ check bits: info is $n^2/(n^2+3)$.

Better percentage is information for large n; how should we generalize?

Generalizations

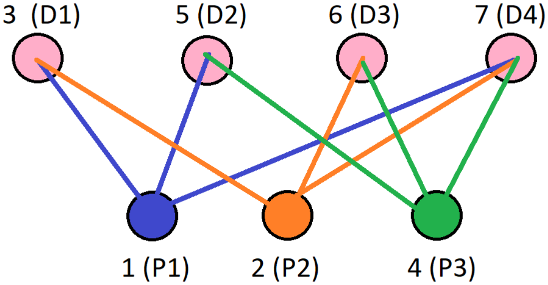
What is a better geometry to use?

Other Approaches

Hamming Codes: Can send a message with 7 bits, 4 are data, and can correct one error: https://en.wikipedia.org/wiki/Hamming_code.

Extended binary Golay code: Can send a message with 24 bits, 12 are data, can correct any 3-bit errors and can detect some other errors: https://en.wikipedia.org/wiki/Binary_Golay_code.

Manhamming



- If no errors, all correct.
- If only one color error, is P1, P2 or P3.
- If just blue and orange is D1.
- If just blue and green is D2.
- If just orange and green is D3
- If all wrong is D4.

Comparison

Say want to transmit around $2^{12} = 4096$ bits of data.

Can do a square and cube; the Hamming code will do $2^{12} - 1 - 12$.

- Square: 4096 out of 4224 data: 96.9697%.
- Cube: 4096 out of 4144 data: 98.8417%.
- Hamming: 4083 out of 4095 data: 99.707%.

All converge to 100%, difference narrows as size increases.

Challenge Problems

Even if these are known, value in trying to solve yourself.

- For a given n, what is the fewest number of check digits one needs to successfully transmit n data digits and be able to correct up to one error?
- Now assume there can be up to two errors....
- Now assume there can be up to k errors....

Happy to chat: sjm1@williams.edu.

Interleaving

Say transmit

0111101001010101010101010101010101011110...

but a localized burst of noise, receive

0111011101010101010101010101010101011110...

Interleaving

Transmit every fourth:

- 01000000001 $\mapsto$ 00000000001
- 10111111111 $\mapsto$ 11111111111
- 11000000001 $\mapsto$ 11000000001
- 10111111110 $\mapsto$ 11111111110

Steganography

Can you see the cat in the tree?

Transmitting Images

How to transmit an image?

- Have an $L \times W$ grid with LW pixels.
- Each pixel a triple, maybe (Red, Green, Blue).
- Often each value in $\{0, 1, 2, 3, \ldots, 2^n - 1\}$.
- $n = 8$ gives 256 choices for each, or 16,777,216 possibilities.

Steganography

Steganography: Concealing a message in another message: https://en.wikipedia.org/wiki/Steganography.

Steganography

Steganography: Concealing a message in another message: https://en.wikipedia.org/wiki/Steganography.

Take one of the colors, say red, a number from 0 to 255.

Write in binary: $r_7 2^7 + r_6 2^6 + \cdots + r_1 2 + r_0$.

If change just the last or last two digits, very minor change to image.

Can hide an image in another.

If just do last, can hide a black and white image easily....

Can you see the cat in the tree?

Can you see the cat in the tree?

Bonus: 0's, 1's and **Cookie Monster**

Pre-requisites: Combinatorics Review

- $n!$: number of ways to order n people, order matters.
- $\frac{n!}{k!(n-k)!} = n\mathrm{C}k = \binom{n}{k}$: number of ways to choose k from n, order doesn't matter.
- Stirling's Formula: $n! \approx n^n e^{-n} \sqrt{2\pi n}$.

Previous Results

Fibonacci Numbers: $F_{n+1} = F_n + F_{n-1}$;
First few: $1, 2, 3, 5, 8, 13, 21, 34, 55, 89, \dots$.

Previous Results

Fibonacci Numbers: $F_{n+1} = F_n + F_{n-1}$;
First few: $1, 2, 3, 5, 8, 13, 21, 34, 55, 89, \dots$.

Zeckendorf's Theorem

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

Previous Results

Fibonacci Numbers: $F_{n+1} = F_n + F_{n-1}$;
First few: $1, 2, 3, 5, 8, 13, 21, 34, 55, 89, \dots$.

Zeckendorf's Theorem

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

Example: $51 = ?$

Previous Results

Fibonacci Numbers: $F_{n+1} = F_n + F_{n-1}$;
First few: $1, 2, 3, 5, 8, 13, 21, 34, 55, 89, \dots$.

Zeckendorf's Theorem

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

Example: $51 = 34 + 17 = F_8 + 17$.

Previous Results

Fibonacci Numbers: $F_{n+1} = F_n + F_{n-1}$;
First few: $1, 2, 3, 5, 8, 13, 21, 34, 55, 89, \dots$.

Zeckendorf's Theorem

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

Example: $51 = 34 + 13 + 4 = F_8 + F_6 + 4$.

Previous Results

Fibonacci Numbers: $F_{n+1} = F_n + F_{n-1}$;
First few: $1, 2, 3, 5, 8, 13, 21, 34, 55, 89, \dots$.

Zeckendorf's Theorem

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

Example: $51 = 34 + 13 + 3 + 1 = F_8 + F_6 + F_3 + 1$.

Previous Results

Fibonacci Numbers: $F_{n+1} = F_n + F_{n-1}$;
First few: 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, ...

Zeckendorf's Theorem

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

Example: $51 = 34 + 13 + 3 + 1 = F_8 + F_6 + F_3 + F_1$.

Previous Results

Fibonacci Numbers: $F_{n+1} = F_n + F_{n-1}$;
First few: $1, 2, 3, 5, 8, 13, 21, 34, 55, 89, \dots$.

Zeckendorf's Theorem

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

Example: $51 = 34 + 13 + 3 + 1 = F_8 + F_6 + F_3 + F_1$.
Example: $83 = 55 + 21 + 5 + 2 = F_9 + F_7 + F_4 + F_2$.
Observe: 51 miles $\approx$ 82.1 kilometers.
Observe: Write 51 as 101001010_{Fib}.

Old Results

Central Limit Type Theorem

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

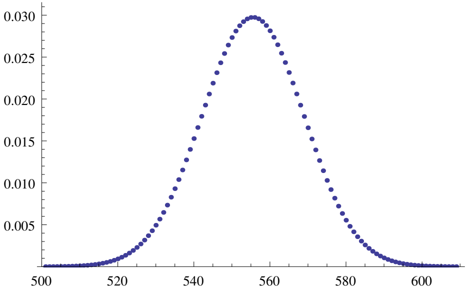


Figure: Number of summands in $[F_{2010}, F_{2011})$; $F_{2010} \approx 10^{420}$.

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

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

Theorem (Zeckendorf Gap Distribution)

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

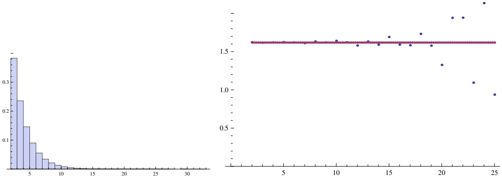


Figure: Distribution of gaps in $[F_{1000}, F_{1001})$; $F_{2010} \approx 10^{208}$.

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

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

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

Preliminaries: The Cookie Problem

The Cookie Problem

The number of ways of dividing C identical cookies among P distinct people is $\binom{C+P-1}{P-1}$.

Preliminaries: The Cookie Problem

The Cookie Problem

The number of ways of dividing C identical cookies among P distinct people is $\binom{C+P-1}{P-1}$.

Proof: Consider $C + P - 1$ cookies in a line.
Cookie Monster eats $P - 1$ cookies: $\binom{C+P-1}{P-1}$ ways to do.
Divides the cookies into P sets.

Preliminaries: The Cookie Problem

The Cookie Problem

The number of ways of dividing C identical cookies among P distinct people is $\binom{C+P-1}{P-1}$.

Proof: Consider $C + P - 1$ cookies in a line.
Cookie Monster eats $P - 1$ cookies: $\binom{C+P-1}{P-1}$ ways to do.
Divides the cookies into P sets.
Example: 8 cookies and 5 people ($C = 8$, $P = 5$):

Preliminaries: The Cookie Problem

The Cookie Problem

The number of ways of dividing C identical cookies among P distinct people is $\binom{C+P-1}{P-1}$.

Proof: Consider $C + P - 1$ cookies in a line.
Cookie Monster eats $P - 1$ cookies: $\binom{C+P-1}{P-1}$ ways to do.
Divides the cookies into P sets.
Example: 8 cookies and 5 people ($C = 8$, $P = 5$):

Preliminaries: The Cookie Problem

The Cookie Problem

The number of ways of dividing C identical cookies among P distinct people is $\binom{C+P-1}{P-1}$.

Proof: Consider $C+P-1$ cookies in a line.
Cookie Monster eats $P-1$ cookies: $\binom{C+P-1}{P-1}$ ways to do.
Divides the cookies into P sets.
Example: 8 cookies and 5 people ($C=8$, $P=5$):

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 $\binom{C+P-1}{P-1}$.

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

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

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

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

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

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

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

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

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

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

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

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

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

$$N = F_{i_1} + F_{i_2} + \cdots + F_{i_{k-1}} + F_n,$$
$$1 \leq i_1 < i_2 < \cdots < i_{k-1} < i_k = n,\ i_j - i_{j-1} \geq 2.$$

$$d_1 := i_1 - 1,\ d_j := i_j - i_{j-1} - 2\ (j > 1).$$
$$d_1 + d_2 + \cdots + d_k = n - 2k + 1,\ d_j \geq 0.$$

Cookie counting $\Rightarrow p_{n,k} = \binom{n-2k+1\ +\ k-1}{k-1} = \binom{n-k}{k-1}$.

Thank You! sjm1@williams.edu

Thank you!

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 approximates 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

$$\begin{aligned} f_{n+1}(k) &= \binom{n-k}{k} \frac{1}{F_n} \\ &= \frac{(n-k)!}{(n-2k)!k!} \frac{1}{F_n} = \frac{1}{\sqrt{2\pi}} \frac{(n-k)^{n-k+\frac{1}{2}}}{k^{(k+\frac{1}{2})}(n-2k)^{n-2k+\frac{1}{2}}} \frac{1}{F_n} \end{aligned}$$

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 ϕ 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.

(Sketch of the) Proof of Gaussianity

Model the distribution as centered around the mean by the change of variable $k = \mu + x\sigma$ where μ and σ 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

$$\begin{aligned} N_n &= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{n-k}{k(n-2k)}}\frac{\phi}{\sqrt{5}} \\ &= \frac{1}{\sqrt{2\pi n}}\sqrt{\frac{1-k/n}{(k/n)(1-2k/n)}}\frac{\sqrt{5}}{\phi} \\ &= \frac{1}{\sqrt{2\pi n}}\sqrt{\frac{1-(\mu+\sigma x)/n}{((\mu+\sigma x)/n)(1-2(\mu+\sigma x)/n)}}\frac{\sqrt{5}}{\phi} \\ &= \frac{1}{\sqrt{2\pi n}}\sqrt{\frac{1-C-y}{(C+y)(1-2C-2y)}}\frac{\sqrt{5}}{\phi} \end{aligned}$$

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

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

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

(Sketch of the) Proof of Gaussianity

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

$$\begin{aligned}
\log(S_n) &= \log\left(\phi^{-n}\frac{(n-k)^{(n-k)}}{k^k(n-2k)^{(n-2k)}}\right) \\
&= -n\log(\phi) + (n-k)\log(n-k) - (k)\log(k) \\
&\quad - (n-2k)\log(n-2k) \\
&= -n\log(\phi) + (n-(\mu+x\sigma))\log(n-(\mu+x\sigma)) \\
&\quad - (\mu+x\sigma)\log(\mu+x\sigma) \\
&\quad - (n-2(\mu+x\sigma))\log(n-2(\mu+x\sigma)) \\
&= -n\log(\phi) \\
&\quad + (n-(\mu+x\sigma))\left(\log(n-\mu) + \log\left(1-\frac{x\sigma}{n-\mu}\right)\right) \\
&\quad - (\mu+x\sigma)\left(\log(\mu) + \log\left(1+\frac{x\sigma}{\mu}\right)\right) \\
&\quad - (n-2(\mu+x\sigma))\left(\log(n-2\mu) + \log\left(1-\frac{x\sigma}{n-2\mu}\right)\right) \\
&= -n\log(\phi) \\
&\quad + (n-(\mu+x\sigma))\left(\log\left(\frac{n}{\mu}-1\right) + \log\left(1-\frac{x\sigma}{n-\mu}\right)\right) \\
&\quad - (\mu+x\sigma)\log\left(1+\frac{x\sigma}{\mu}\right) \\
&\quad - (n-2(\mu+x\sigma))\left(\log\left(\frac{n}{\mu}-2\right) + \log\left(1-\frac{x\sigma}{n-2\mu}\right)\right).
\end{aligned}$$

(Sketch of the) Proof of Gaussianity

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

$$
\begin{aligned}
= \quad & -n\log(\phi) + (n-k)\log\left(\frac{n}{\mu} - 1\right) - (n-2k)\log\left(\frac{n}{\mu} - 2\right) + (n-(\mu+x\sigma))\log\left(1 - \frac{x\sigma}{n-\mu}\right) \\
& - (\mu + x\sigma)\log\left(1 + \frac{x\sigma}{\mu}\right) - (n - 2(\mu + x\sigma))\log\left(1 - \frac{x\sigma}{n-2\mu}\right) \\
= \quad & -n\log(\phi) + (n-k)\log(\phi+1) - (n-2k)\log(\phi) + (n-(\mu+x\sigma))\log\left(1 - \frac{x\sigma}{n-\mu}\right) \\
& - (\mu + x\sigma)\log\left(1 + \frac{x\sigma}{\mu}\right) - (n - 2(\mu + x\sigma))\log\left(1 - \frac{x\sigma}{n-2\mu}\right) \\
= \quad & n(-\log(\phi) + \log\left(\phi^2\right) - \log(\phi)) + k(\log(\phi^2) + 2\log(\phi)) + (n-(\mu+x\sigma))\log\left(1 - \frac{x\sigma}{n-\mu}\right) \\
& - (\mu + x\sigma)\log\left(1 + \frac{x\sigma}{\mu}\right) - (n - 2(\mu + x\sigma))\log\left(1 - 2\frac{x\sigma}{n-2\mu}\right) \\
= \quad & (n-(\mu+x\sigma))\log\left(1 - \frac{x\sigma}{n-\mu}\right) - (\mu + x\sigma)\log\left(1 + \frac{x\sigma}{\mu}\right) \\
& - (n - 2(\mu + x\sigma))\log\left(1 - 2\frac{x\sigma}{n-2\mu}\right).
\end{aligned}
$$

(Sketch of the) Proof of Gaussianity

Finally, we expand the logarithms and collect powers of $x\sigma/n$.

$$
\begin{aligned}
\log(S_n) &= (n-(\mu+x\sigma))\left(-\frac{x\sigma}{n-\mu}-\frac{1}{2}\left(\frac{x\sigma}{n-\mu}\right)^2+\dots\right) \\
&\quad -(\mu+x\sigma)\left(\frac{x\sigma}{\mu}-\frac{1}{2}\left(\frac{x\sigma}{\mu}\right)^2+\dots\right) \\
&\quad -(n-2(\mu+x\sigma))\left(-2\frac{x\sigma}{n-2\mu}-\frac{1}{2}\left(2\frac{x\sigma}{n-2\mu}\right)^2+\dots\right) \\
&= (n-(\mu+x\sigma))\left(-\frac{x\sigma}{n\frac{(\phi+1)}{(\phi+2)}}-\frac{1}{2}\left(\frac{x\sigma}{n\frac{(\phi+1)}{(\phi+2)}}\right)^2+\dots\right) \\
&\quad -(\mu+x\sigma)\left(\frac{x\sigma}{\frac{n}{\phi+2}}-\frac{1}{2}\left(\frac{x\sigma}{\frac{n}{\phi+2}}\right)^2+\dots\right) \\
&\quad -(n-2(\mu+x\sigma))\left(-\frac{2x\sigma}{n\frac{\phi}{\phi+2}}-\frac{1}{2}\left(\frac{2x\sigma}{n\frac{\phi}{\phi+2}}\right)^2+\dots\right) \\
&= \frac{x\sigma}{n}n\left(-\left(1-\frac{1}{\phi+2}\right)\frac{(\phi+2)}{(\phi+1)}-1+2\left(1-\frac{2}{\phi+2}\right)\frac{\phi+2}{\phi}\right) \\
&\quad -\frac{1}{2}\left(\frac{x\sigma}{n}\right)^2 n\left(-2\frac{\phi+2}{\phi+1}+\frac{\phi+2}{\phi+1}+2(\phi+2)-(\phi+2)+4\frac{\phi+2}{\phi}\right) \\
&\quad +O\left(n\left(x\sigma/n\right)^3\right)
\end{aligned}
$$

(Sketch of the) Proof of Gaussianity

$$\begin{aligned}
\log(S_n) &= \frac{x\sigma}{n} n \left(-\frac{\phi+1}{\phi+2}\frac{\phi+2}{\phi+1} - 1 + 2\frac{\phi}{\phi+2}\frac{\phi+2}{\phi} \right) \\
&\quad -\frac{1}{2}\left(\frac{x\sigma}{n}\right)^2 n(\phi+2)\left(-\frac{1}{\phi+1} + 1 + \frac{4}{\phi}\right) \\
&\quad +O\left(n\left(\frac{x\sigma}{n}\right)^3\right) \\
&= -\frac{1}{2}\frac{(x\sigma)^2}{n}(\phi+2)\left(\frac{3\phi+4}{\phi(\phi+1)} + 1\right) + O\left(n\left(\frac{x\sigma}{n}\right)^3\right) \\
&= -\frac{1}{2}\frac{(x\sigma)^2}{n}(\phi+2)\left(\frac{3\phi+4+2\phi+1}{\phi(\phi+1)}\right) + O\left(n\left(\frac{x\sigma}{n}\right)^3\right) \\
&= -\frac{1}{2}x^2\sigma^2\left(\frac{5(\phi+2)}{\phi n}\right) + O\left(n\left(x\sigma/n\right)^3\right).
\end{aligned}$$

(Sketch of the) Proof of Gaussianity

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

$$\begin{aligned} \log S_n &= -\frac{1}{2}x^2 \\ S_n &= e^{-\frac{1}{2}x^2}. \end{aligned}$$

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

$$\begin{aligned} f_n(k)dk &= N_n S_n dk \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}x^2} \sigma dx \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx. \end{aligned}$$

□