

Survey of Scientific Computing (SciComp 301)

Dave Biersach
Brookhaven National
Laboratory
dbiersach@bnl.gov

Session 12
Continued Fractions,
Chi Squared

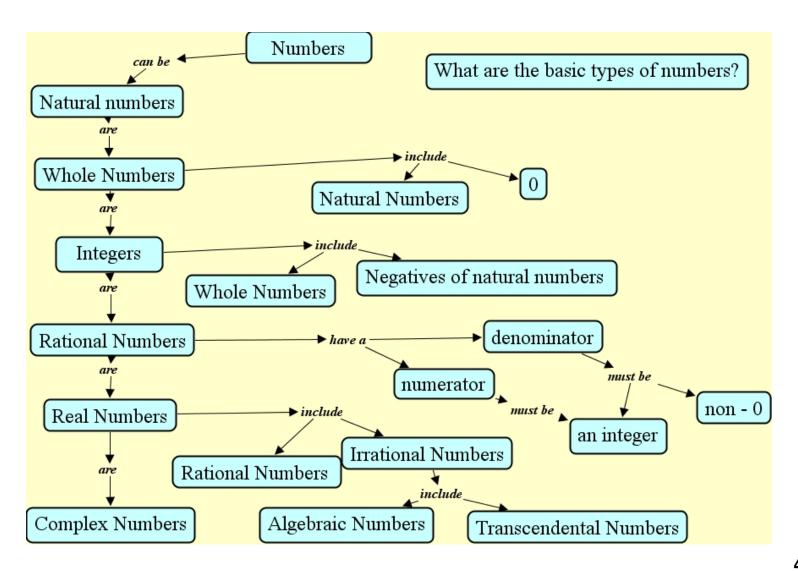
Session Goals

- Gain an appreciation for Continued Fractions in nature
- Understand the three types of CFs: 1) finite, 2) infinite with repeating <u>sequence</u>, 3) infinite with repeating <u>pattern</u>
- Write code to generate a generalized CF for a real number, and how to expand that CF to produce convergents of the original number
- Appreciate the hidden underlying simplicity of the generalized continued fraction for π
- Perform a computational mathematical experiment to determine the solutions to Pell's Equation

Session Goals

- Gain an appreciation for the Normal Distribution
- Investigate if a Normal Distribution can be made from a Uniform Distribution using a Pachinko game
- Use **chi-squared statistic** to determine if a random sample conforms to a reasonable Normal Distribution

Expanding Your Definition of a "Number"

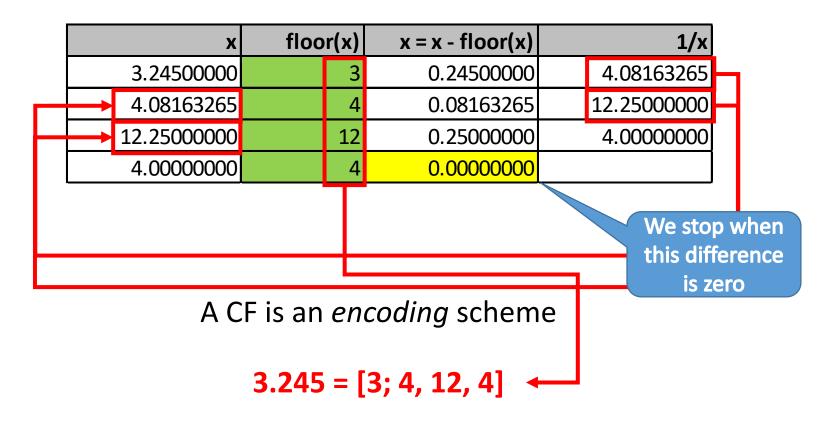


$$x = a_0 + \frac{b_0}{a_1 + \frac{b_1}{a_2 + \frac{b_2}{a_3 + \ddots}}}$$

In a simple continued fraction, all $b_n = 1$

$$3.245 = 3 + \frac{1}{4 + \frac{1}{12 + \frac{1}{4}}}$$

What is the simple CF encoding for **3.245**?



All rational numbers have a CF of finite length!

What is the simple CF encoding for 0.825 (= 33/40)?

x	floo	r(x)	x = x - floor(x)	1/x
0.82500000		0	0.82500000	1.21212121
1.21212121		1	0.21212121	4.71428571
4.71428571		4	0.71428571	1.40000000
1.40000000		1	0.40000000	2.50000000
2.50000000		2	0.50000000	2.00000000
2.00000000		2	0.00000000	

A CF is an *encoding* scheme

$$0.825 = [0; 1, 4, 1, 2, 2]$$

All rational numbers have a CF of finite length!

$$h_n = a_n h_{(n-1)} + b_{(n-1)} h_{(n-2)}$$

$$k_n = a_n k_{(n-1)} + b_{(n-1)} k_{(n-2)}$$

How do we expand (decode) a simple CF?

$$\Delta = \left(\frac{h_n}{k_n} - x\right)$$

n		a	b	h	k	h/k	delta
-2				0	1		
-1	_		1	1	0		
0		0	1	0	1	0.000000000	0.825000000
1		1	1	1	1	1.000000000	-0.175000000
2		4	1	4	5	0.800000000	0.025000000
3		1	1	5	6	0.833333333	-0.008333333
4		2	1	14	17	0.823529412	0.001470588
5		2	1	33	40	0.825000000	0.000000000

Each row gives a better and better approximation (h / k) to the original number x

$$[0; 1, 4, 1, 2, 2] = 0.825 (= 33/40)$$

$\sqrt{2}$ to 3,600 digits

What is the simple CF encoding for $\sqrt{2}$?

х	floor(x)	x = x - floor(x)	1/x
1.41421356	1	0.41421356	2.41421356
2.41421356	2	0.41421356	2.41421356
2.41421356	2	0.41421356	2.41421356
2.41421356	2	0.41421356	2.41421356
2.41421356	2	0.41421356	2.41421356
2.41421356	2	0.41421356	2.41421356

$$\sqrt{2} = [1; \{2\}]$$

Numbers within {} are repeated

All irrational numbers yield an infinite CF with a repeated *sequence* of *finite* length!

There is simple order behind the chaos!

$$h_n = a_n h_{(n-1)} + b_{(n-1)} h_{(n-2)}$$

$$k_n = a_n k_{(n-1)} + b_{(n-1)} k_{(n-2)}$$

$$\Delta = \left(\frac{h_n}{k_n} - x\right)$$

What fraction best approximates $\sqrt{2}$?

n	a	b	h	k	h/k	delta
-2			0	1		
-1		1	1	0		
0	1	1	1	1	1.000000000	0.414213562
1	2	1	3	2	1.500000000	-0.085786438
2	2	1	7	5	1.400000000	0.014213562
3	2	1	17	12	1.416666667	-0.002453104
4	2	1	41	29	1.413793103	0.000420459
5	2	1	99	70	1.414285714	-0.000072152
6	2	1	239	169	1.414201183	0.000012379
7	2	1	577	408	1.414215686	-0.000002124
8	2	1	1393	985	1.414213198	0.00000364
9	2	1	3363	2378	1.414213625	-0.00000063
10	2	1	8119	5741	1.414213552	0.00000011

$$\sqrt{2} \approx 8,119 / 5,741$$

What is the simple CF encoding for $\sqrt{113}$?

x	floor(x)	x = x - floor(x)	1/x
10.63014581	10	0.63014581	1.58693429
1.58693429	1	0.58693429	1.70376823
1.70376823	1	0.70376823	1.42092235
1.42092235	1	0.42092235	2.37573512
2.37573512	2	0.37573512	2.66144940
2.66144940	2	0.66144940	1.51183144
1.51183144	1	0.51183144	1.95376823
1.95376823	1	0.95376823	1.04847275
1.04847275	1	0.04847275	20.63014581
20.63014581	20	0.63014581	1.58693430
1.58693430	1	0.58693430	1.70376822
1.70376822	1	0.70376822	1.42092237
1.42092237	1	0.42092237	2.37573499
2.37573499	2	0.37573499	2.66145027
2.66145027	2	0.66145027	1.51182945
1.51182945	1	0.51182945	1.95377581
1.95377581	1	0.95377581	1.04846442
1.04846442	1	0.04846442	20.63369395
20.63369395	20	0.63369395	1.57804883

Period = 9

$$\sqrt{113} = [10; \{1,1,1,2,2,1,1,1,20\}]$$

$$h_n = a_n h_{(n-1)} + b_{(n-1)} h_{(n-2)}$$

$$k_n = a_n k_{(n-1)} + b_{(n-1)} k_{(n-2)}$$

 $\Delta = \left(\frac{h_n}{k_n} - x\right)$

What fraction best approximates $\sqrt{113}$?

n	а	b	h	k	h/k	delta
-2			0	1		
-1		1	1	0		
0	10	1	10	1	10.000000000	0.630145813
1	1	1	11	1	11.000000000	-0.369854187
2	1	1	21	2	10.500000000	0.130145813
3	1	1	32	3	10.666666667	-0.036520854
4	2	1	85	8	10.625000000	0.005145813
5	2	1	202	19	10.631578947	-0.001433135
6	1	1	287	27	10.629629630	0.000516183
7	1	1	489	46	10.630434783	-0.000288970
8	1	1	776	73	10.630136986	0.000008826
9	20	1	16009	1506	10.630146082	-0.000000270
10	1	1	16785	1579	10.630145662	0.00000151
11	1	1	32794	3085	10.630145867	-0.000000054
12	1	1	49579	4664	10.630145798	0.00000015
13	2	1	131952	12413	10.630145815	-0.000000002

 $\sqrt{113} \approx$ **131,952 / 12,413**

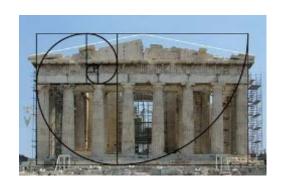
e to 3,600 digits

2.7182818284590452353602874713526624977572470936999595749669676277240766303535475945713821785251664274274663

Continued Fraction for e

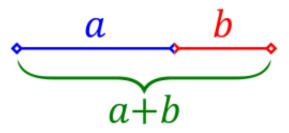
	х	floor(x	()	x = x - floor(x)	1/x
	2.71828183		2	0.71828183	1.39221119
	1.39221119		1	0.39221119	2.54964678
	2.54964678		2	0.54964678	1.81935024
	1.81935024		1	0.81935024	1.22047929
	1.22047929		1	0.22047929	4.53557348
	4.53557348		4	0.53557348	1.86715744
	1.86715744		1	0.86715744	1.15319313
	1.15319313	1		0.15319313	6.52770793
	6.52770793		6	0.52770793	1.89498763
All transcend	dental number	rs 🗀	1	0.89498763	1.11733388
	inite CF with a		1	0.11733388	8.52268767
•			8	0.52268767	1.91318841
repeated <i>patte</i>	ern of finite ier	ngth	1	0.91318841	1.09506427
	1.09506427		1	0.09506427	10.51919947
	10.51919947	10	0	0.51919947	1.92604201
	1.92634201 1.07986461		1	0.92604201	1.07986461
			1	0.07986461	12.52119027
	12.52119027	1	2	0.52119027	1.91868508
	1.91868508		1	0.91868508	1.08851229

 $e = [2; \{1,2n,1\}]$ for n > 0 $e^2 = [7;2,\{1,1,3n,12n+6,3n+2\}]$ for n > 0



The Golden Ratio





a+b is to a as a is to b

$$\frac{a+b}{a} = \frac{a}{b} = \varphi$$

$$1 + \frac{b}{a} = \frac{a}{b} = \varphi$$

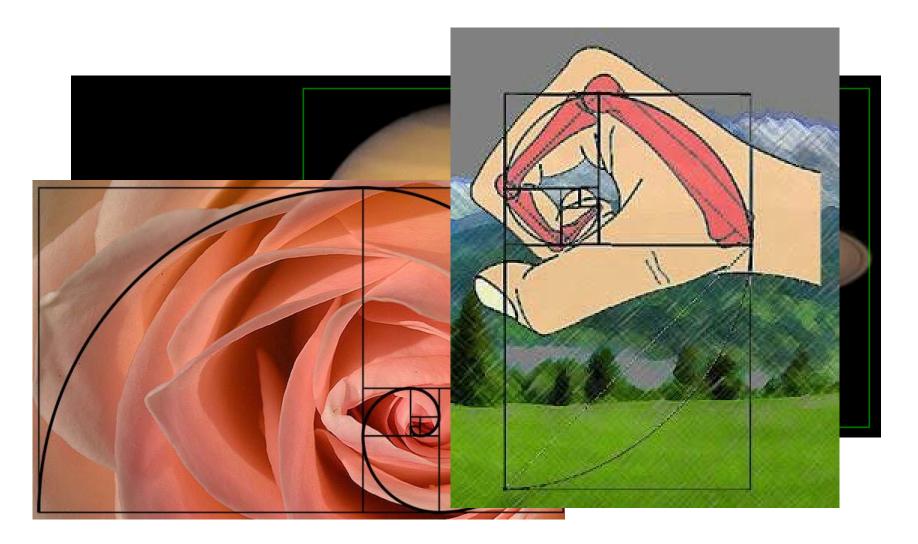
$$1 + \frac{1}{\varphi} = \varphi$$
$$\varphi + 1 = \varphi^2$$

$$\varphi + 1 = \varphi^2$$

$$\varphi^2 - \varphi - 1 = 0$$

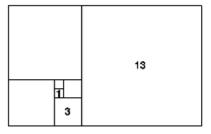
$$\varphi = \frac{1 + \sqrt{5}}{2} = 1.6180339887 \dots$$

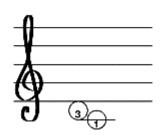
$$\varphi = \frac{1 + \sqrt{5}}{2} = 1.6180339887 \dots$$

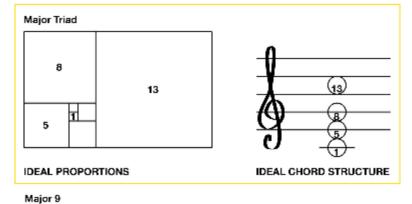


$$\varphi = \frac{1 + \sqrt{5}}{2} = 1.6180339887 \dots$$

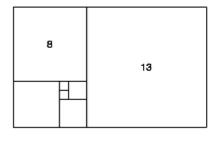
Whole Step

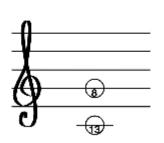


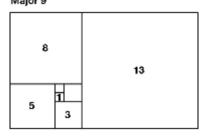


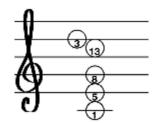


Perfect Fifth





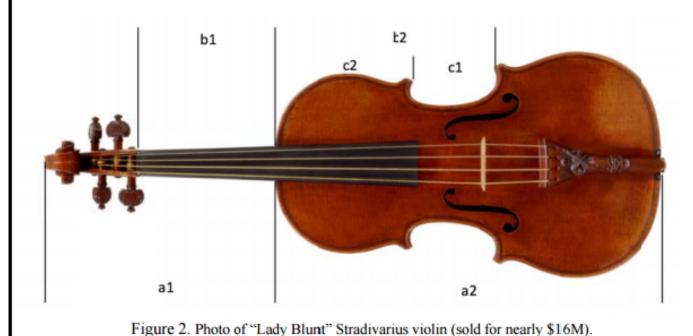




$$\varphi = \frac{1 + \sqrt{5}}{2} = 1.6180339887 \dots$$

The greatest of luthiers, Stradivarius, designed his violins around the golden ratio (ϕ). His violins are the most valuable and precious instruments in the string-playing world because of their exquisite tonal and harmonic qualities, [2]. The Stradivarius violin in Fig. 2 reveals how precisely his instruments are determined by the golden ratio, [3]:

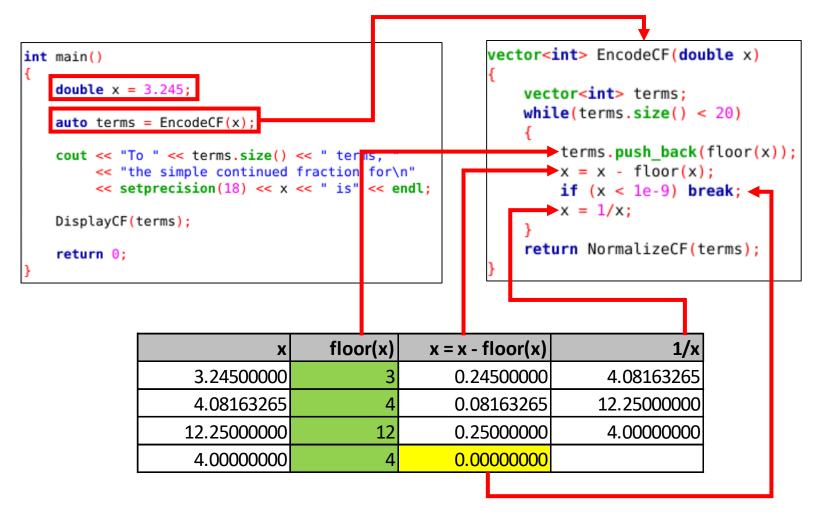
$$\frac{a1+a2}{a2} = \frac{a2}{a1} = \frac{b2}{b1} = \frac{b2}{c2} = \frac{c2}{c1} = \phi$$



φ to 3,600 digits

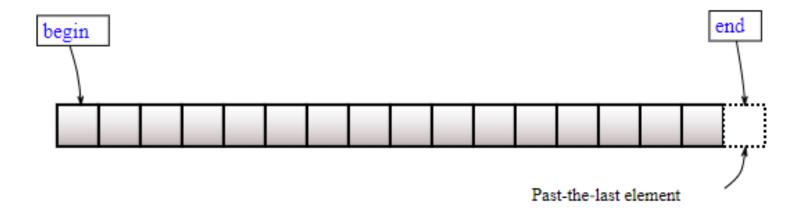
1.6180339887498948482045868343656381177203091798057628621354486227052604628189024497072072041893911374847540

Open Lab 1 – Simple CF Encoding



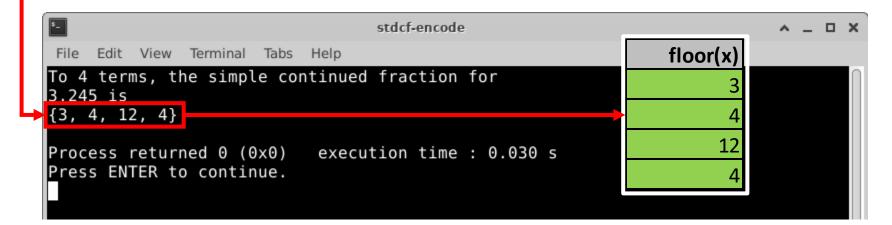
View Lab 1 – Simple CF Encoding

```
void DisplayCF(const vector<int>& terms)
{
    cout << "{":
    auto itr = terms.begin();
    while (true)
    {
        cout << *itr;
        if (++itr == terms.end()) break;
        cout << ", ";
    }
    cout << "}\n";
    return;
}</pre>
```



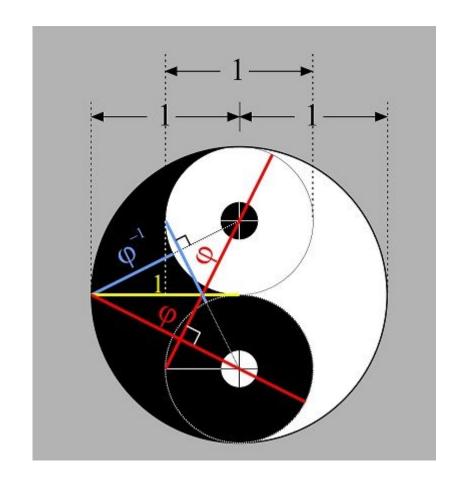
Run Lab 1 – Simple CF Encoding

```
void DisplayCF(const vector<int>& terms)
{
    cout << "{";
    auto itr = terms.begin();
    while (true)
    {
        cout << *itr;
        if (++itr == terms.end()) break;
        cout << ", ";
    }
    cout << "}\n";
    return;
}</pre>
```



Edit Lab 1 –Simple CF Encoding

$$\varphi = \frac{1 + \sqrt{5}}{2} = 1.6180339887$$

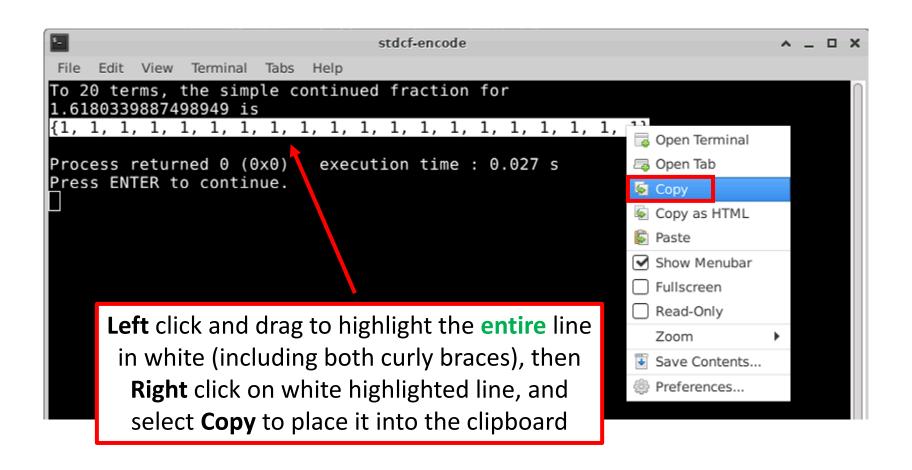


Run Lab 1 – Simple CF Encoding

• Generate the Simple CF for the golden ratio $\frac{1+\sqrt{5}}{2}$

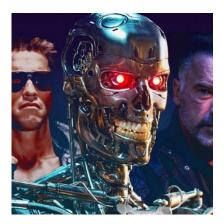
This is Mother
Nature's true *Unit*It is the most simple infinite CF possible!

Check Lab 1 – Simple CF Encoding



C++ Vector Initialization

- Vectors can be defined using the list initializer syntax
 - Elements are comma separated between curly braces
 - First item in list goes into index position 0 in the array
 - The vector is dynamically sized to match the number of elements in the initializer list
- Lab 1 emits source code for Lab 2
- Programs <u>can</u> create programs



#include "stdafx.h" using namespace std; int main() int maxTerms = 20; vector<double> h(maxTerms + 2); vector<double> k(maxTerms + 2); **if** (cf.size() == 0) cout << "Error - Missing cf data!";</pre> return -1; h.at(0) = 0; k.at(0) = 1;h.at(1) = 1: k.at(1) = 0;cout << "Using " << maxTerms << " terms, ";</pre> cout << "the continued fraction expansion is:" << endl;</pre> cout << setw(5) << "a"; cout << right << setw(15) << "h";</pre> cout << right << setw(15) << "k";</pre> cout << setw(20) << "convergent" << endl;</pre> for (int n{ 2 }; n < maxTerms + 2; ++n)</pre> double a = cf.at(n - 2): h.at(n) = a * h.at(n - 1) + h.at(n - 2);k.at(n) = a * k.at(n - 1) + k.at(n - 2);double convergent = h.at(n) / k.at(n); cout << setprecision(0) << right</pre> << setw(5) << a << setw(15) << h[n] << setw(15) << k[n] << setprecision(14) << fixed << setw(20) << convergent << endl;</pre> return 0;

Edit Lab 2 CF Decode

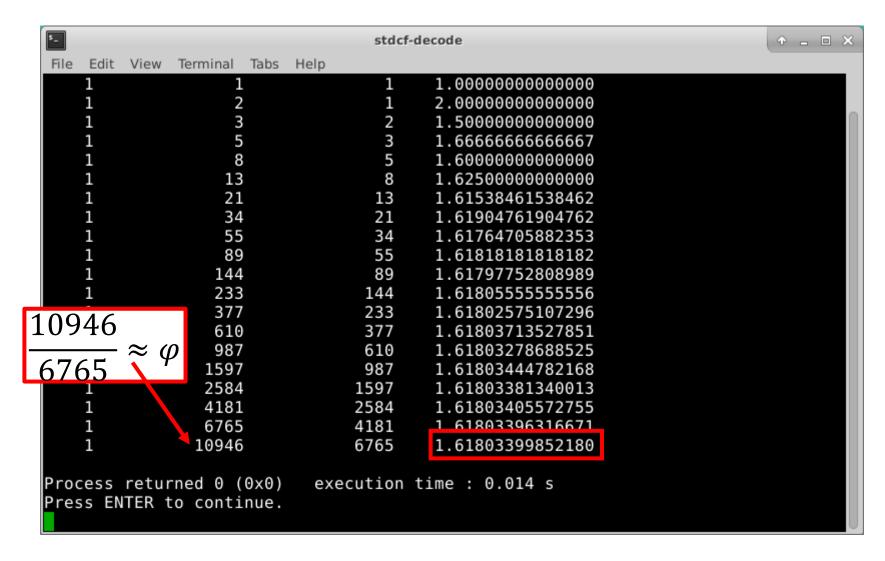
Don't forget the semicolon!

Right click and **paste** in the white line output from Lab 1

$$h_n = a_n h_{(n-1)} + b_{(n-1)} h_{(n-2)}$$

$$k_n = a_n k_{(n-1)} + b_{(n-1)} k_{(n-2)}$$

Run Lab 2 – Simple CF Decoding



π to 3,600 digits

3.1415926535897932384626433832795028841971693993751058209749445923078164062862089986280348253421170679821480

What is the simple CF encoding for π ?

х	floor(x)	x = x - floor(x)	1/x
3.14159265	3	0.14159265	7.06251331
7.06251331	7	0.06251331	15.99659441
15.99659441	15	0.99659441	1.00341723
1.00341723	1	0.00341723	292.63459088
292.63459088	292	0.63459088	1.57581844
1.57581844	1	0.57581844	1.73665853
1.73665853	1	0.73665853	1.35748105
1.35748105	1	0.35748105	2.79735107
2.79735107	2	0.79735107	1.25415271
1.25415271	1	0.25415271	3.93464232
3.93464232	3	0.93464232	1.06992802
1.06992802	1	0.06992802	14.30041960
14.30041960	14	0.30041960	3.32867763
3.32867763	3	0.32867763	3.04249485
3.04249485	3	0.04249485	23.53226532
23.53226532	23	0.53226532	1.87876228
1.87876228	1	0.87876228	1.13796418
1.13796418	1	0.13796418	7.24825805
7.24825805	7	0.24825805	4.02806683

 π = [3;7,15,1,292,1,1,1,2,1,3,1,14,3,3,23,1,1,7....] (no repeated *pattern* of finite length \otimes !)

n	а	b	h	k	h/k	delta
-2			0	1		
-1		1	1	0		
0	3	1	3	1	3.000000000	0.141592654
1	7	1	22	7	3.142857143	-0.001264489
2	15	1	333	106	3.141509434	0.000083220
3	1	1	355	113	3.141592920	-0.000000267
4	292	1	103993	33102	3.141592653	0.00000001
5	1	1	104348	33215	3.141592654	0.000000000
6	1	1	208341	66317	3.141592653	0.000000000
7	1	1	312689	99532	3.141592654	0.000000000
8	2	1	833719	265381	3.141592654	0.000000000
9	1	1	1146408	364913	3.141592654	0.000000000
10	3	1	4272943	1360120	3.141592654	0.000000000

If measuring the circumference of Earth:

22 / 7 = accurate to between this classroom and Washington, DC 355 / 113 = accurate to between this classroom and the main parking lot

If measuring the distance between Earth & Sun: 355 / 113 = accurate to 4 football fields 104348 / 33215 = accurate to the length of my shoe

$$x = a_0 + \frac{b_0}{a_1 + \frac{b_1}{a_2 + \frac{b_2}{a_3 + \ddots}}}$$

In a generalized continued fraction, a_n and b_n can now be any expression

$$x=a_0+\dfrac{1}{a_1+\dfrac{1}{a_2+\dfrac{1}{a_3}}}$$
 In a **simple** CF all the numerators $(b_n)=1$

What is a **generalized** CF encoding for π ?

$$\pi = 3 + \frac{1^{2}}{6 + \frac{3^{2}}{6 + \frac{5^{2}}{6 + \frac{7^{2}}{6 + \cdots}}}}$$

$$\pi = [3; 1, \{(2n+1)^{2} | 6\}]$$

Euler

All the mysterious and unpredictable digits of PI come from this simple generalized CF!!

What is a **generalized** CF expansion for π ?

n	а	b	h	k	h/k	delta	
-2			0	1			
-1		1	1	0			
0	3	1	3	1	3.000000000	0.141592654	
1	6	9	19	6	3.166666667	-0.025074013	
2	6	25	141	45	3.133333333	0.008259320	
3	6	49	1321	420	3.145238095	-0.003645442	
4	6	81	14835	4725	3.139682540	0 00101011/	
5	6	121	196011	62370	3.142712843	Gen CFs	
6	6	169	2971101	945945	3.140881341	converge slow	ly
7	6	225	50952465	16216200	3.142071817	-0.0004791	
8	6	289	9.74E+08	3.1E+08	3.141254824	0.000337	
9	6	361	2.06E+10	6.55E+09	3.141839619	-0.000246965	

$$\pi = [3; 1, \{(2n+1)^2 | 6\}]$$

All the mysterious and unpredictable digits of PI come from this simple generalized CF!!

What is another **generalized** CF encoding for π ?

Biersach



$$\pi = 3 + \frac{1}{7 + \frac{1}{15 + \frac{25}{23 + \frac{65}{31 + \cdots}}}}$$

$$a_0 b_0 b_n a_n$$
 $\downarrow \downarrow \downarrow \downarrow$
 $\pi = [3; 1, \{(8n^2-7) | (8n-1)\}]$

There are infinitely many Generalized CFs but not all converge

What is a **generalized** CF expansion for π ?

n	а	b	h	k	h/k	delta
-2			0	1		
-1		1	1	0		
0	3	1	3	1	3.000000000	0.141592654
1	7	1	22	7	3.142857143	-0.001264489
2	15	25	333	106	3.141509434	0.000083220
3	23	65	8209	2613	3.141599694	-0.000007040
4	31	121	276124	87893	3.141592618	0.00000035
5	39	193	11762125	3744000	3.141593216	-0.00000562
6	47	281	6.06E+08	1.93E+08	3.141593163	-0.00000510

$$\pi = [3; 1, \{(8n^2-7)|(8n-1)\}]$$

This GCF does **not** converge!

What is another **generalized** CF encoding for π ?

Biersach



$$\pi = 2 + \frac{8}{6 + \frac{12}{10 + \frac{32}{14 + \frac{60}{18 + \cdot \cdot}}}}$$

$$\pi = [2; 8, {(4n^2+8n)|(4n+2)}]$$

What is a **generalized** CF expansion for π ?

n	а	b	h	k	h/k	delta
-2			0	1		
-1		1	1	0		
0	2	8	2	1	2.000000000	1.141592654
1	6	12	20	6	3.333333333	-0.191740680
2	10	32	224	72	3.111111111	0.030481542
3	14	60	3776	1200	3.146666667	-0.005074013
4	18	96	81408	25920	3.140740741	0.000851913
5	22	140	2153472	685440	3.141736695	-0.000144041
6	26	192	67387392	21450240	3.141568206	0.000024447
7	30	252	2.44E+09	7.75E+08	3.141596814	-0.000004160
8	34	320	9.98E+10	3.18E+10	3.141591945	0.00000709
9	38	396	4.57E+12	1.45E+12	3.141592775	-0.00000121

$$\pi = [2; 8, {(4n^2+8n)|(4n+2)}]$$

My GCF for π converges faster than Euler's! ©

$$x = a_0 + \frac{b_0}{a_1 + \frac{b_1}{a_2 + \frac{b_2}{a_3 + \ddots}}}$$

In a generalized continued fraction, a_n and b_n can now be any expression

$$\tan(x) = \frac{x}{1 - \frac{x^{2}}{3 - \frac{x^{2}}{3 - \frac{x^{2}}{7 - \ddots}}}}$$

$$\tan(x) = \frac{1}{1 - \frac{x^{2}}{3 - \frac{x^{2}}{3 - \frac{x^{2}}{7 - \frac{x^{2}}{3 - \frac{x^{2}}{7 - \frac{x^{2}}{3 - \frac{$$

Continued Fractions

- CFs may have their own rich arithmetic, algebra, and potentially even their own calculus
 - How can one <u>directly</u> <u>divide</u> two CFs?
 - How can one <u>directly</u> take the sin() of a CF?
 - What does the factorial of a CF look like?
- In many ways a CF is a more accurate representation of an irrational or transcendental number
 - The sum of an infinite series must stop somewhere, and after that point, all of the remaining digits of precision are lost
 - A CF encodes the entire number with no loss of precision
 - What can you discover about CFs?

Pell's Equation

• Your scientist has asked you to write a C++ program to find x & y for a given n (assume $x, y, n \in \mathbb{Z}^+$) such that:

$$x^2 - ny^2 = 1$$

- For every $2 \le n \le 70$, check all $1 \le x \le 70,000$
- Why is there is no need to check for $y > \left| \sqrt{\frac{x^2}{n}} \right|$?
- Do you see any relationship between the specific x & y values that solve the equation for each n value?

int main() DisplayHeader(); const uintmax_t xMax = 70000; for (uintmax t n = 2; $n \le 70$; n++) cout << setw(4) << n;</pre> bool foundSolution = false; uintmax t x = 1; while ((x <= xMax) && !foundSolution)</pre> uintmax_t xSqr = x * x; uintmax t y = 1;uintmax_t yMax = sqrt(xSqr / n); while ((y <= yMax) && !foundSolution)</pre> uintmax_t ySqr = y * y; if (xSqr - n * ySqr == 1)cout << setw(8) << x << setw(8) << y; foundSolution = true; y++; X++; if (!foundSolution) cout << setw(8) << "-" << setw(8) << "-": cout << endl; return 0;

Open Lab 3 Pell's Equation

$$x^2 - ny^2 = 1$$

As soon as a valid solution is found for the current value of **n**, then stop trying any more **x** & **y** values

Hyphens indicate no solution was found in the allowed search space

Run Lab 3 Pell's Equation

\$_			
File	Edit	View	Terminal Tabs
n		X	у
===		====	===
2		3 2	2 1
2 3 4 5 6 7		2	1
4		-	_
5		9 5 8	4
6		5	2
		8	4 2 3 1
8		3	1
9		-	-
10		19	6 3 2
11 12		10	3
12		7	2
13		649	180
14		15	4
15		4	1
16		-	-
17		33	8
18		17	4
19		170	39
20		9	2
21		55	12
22		197	42
23		24	5

\$				
File	Edit	View	Terminal	Tabs
24		5	1	
25		-	-	
26		51	10	
27		26	5	
28		127	24	
29		9801	1820	
30		11	2	
31		1520	273	
32		17	3	
33		23	4	
34		35	6	
35		6	1	
36		-	-	
37		73	12	
38		37	6	
39		25	4	
40		19	3	
41		2049	320	
42		13	2	
43		3482	531	
44		199	30	
45		161	24	
46		24335	3588	
47		48	7	

\$ _			
File	Edit View	Terminal	Tabs
48	7	' 1	
49			
50	99	14	
51	56) 7	
52	649	90	
53	66249	9100	
54	485		
55	89		
56	15		
57	151		
58	19603		
59	536		
60	31	4	
61		-	
62	63		
63	8	3 1	
64		-	
65	129		
66	65		
67	48842		
68	33		
69	7775		
70	251	L 30	

Check Lab 3 – Observations

Which values of **n** have no solution?

$$n = 1, 4, 9, 16, 25, 36, 49, 61, 64, ...$$

Some of the values for x & y are much bigger than for other

close values of **n**:

40	19	3
41	2049	320
42	13	2
43	3482	531
44	199	30
45	161	24
46	24335	3588
47	48	7
48	7	1

 The magnitude of n does not seem to be a good predictor about the magnitude of the x & y values that solve the equation for that specific n

Pell's Equation: Period of Simple CF

Small values for x & y

n	x	у
35	6	1
47	48	7
60	31	4
68	33	4

Period = 2
$$\sqrt{68}$$
 = {8, 4, 16, 4, 16,

Large values for x & y

n	х	у
13	649	180
29	9801	1820
41	2049	320
43	3482	531
46	24335	3588
53	66249	9100
61	1766319049	226153980
67	48842	5967

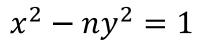
Period = 5
$$\sqrt{29} = \{5, 2, 1, 1, 2, 10, 2, 1, 1, 2, 10,$$

$$\frac{\text{Period} = 5}{\sqrt{53}} = \{7, 3, 1, 1, 3, 14, 3, 1, 1, 3, 14,$$

$$\frac{\text{Period} = 11}{\sqrt{53}} = \frac{11}{\sqrt{53}} = \frac{11}{$$

1, 3, 4, 1, 14, 1,

Pell's Equation: Period of Simple CF



Large values for x & y									
n		х	У						
13	6	49	180						
29	98	01	1820						
41	20	49	320						
43	34	82	531						
46	243	35	3588						
53	662	49	9100						
61	17663190	49	226153980						
67	488	42	5967						

Simple CF

 $\sqrt{13}$ Period = 4

If the period of the simple CF of the \sqrt{n} is large...

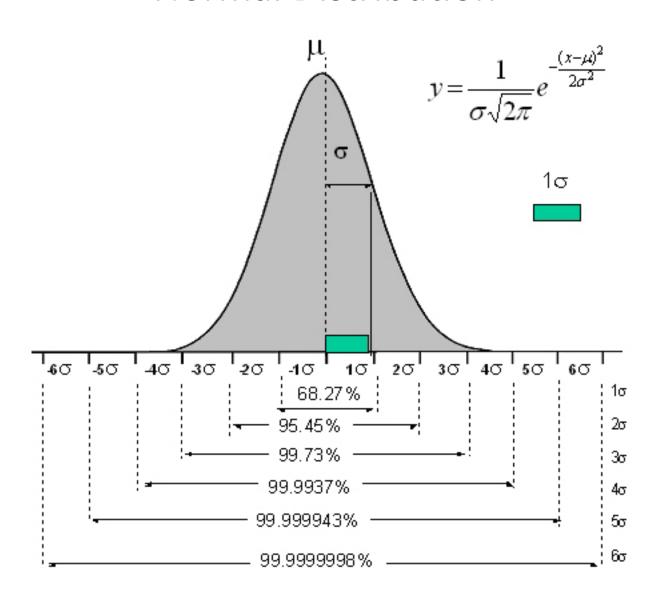
 $\sqrt{61}$ Period = 11

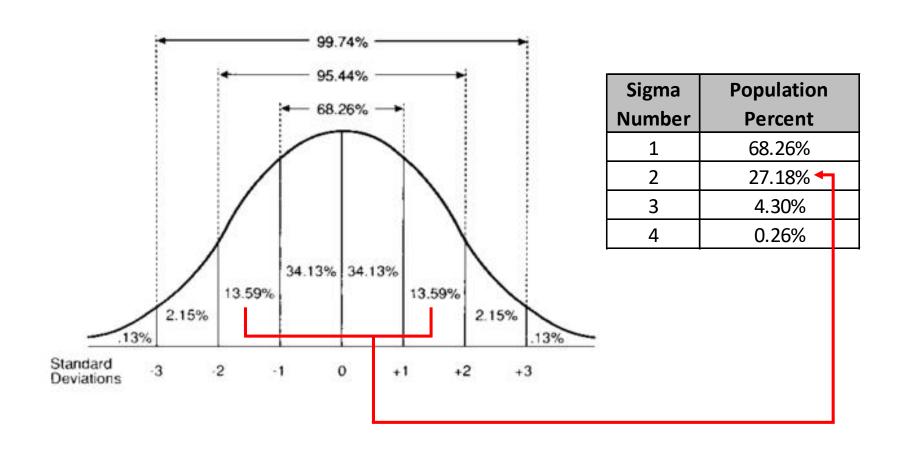
...then the x & y will be large for the solution to Pell's equation

- Until recently, most computer languages only provided a uniform pseudo-random number generator
- Growing up I had heard of a bell curve and I understood the rationale for curving test scores
- However I could not create a normal distribution using my 1978 vintage TRS-80 computer using Bill Gate's first BASIC language interpreter

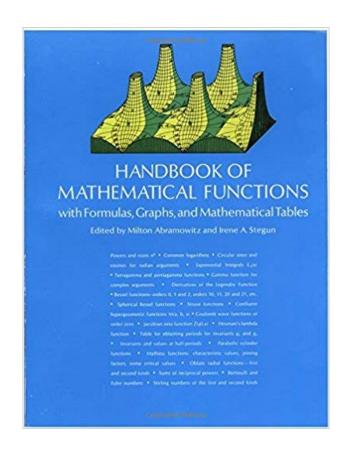


... or could 1?



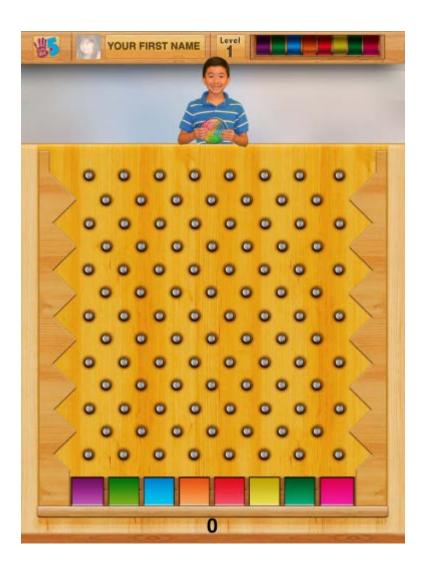


- There are indeed several ways to turn a uniform distribution into a normal distribution
- Developing an accurate functional approximation to a normal curve requires advanced mathematics
- Consider this code from Abramowitz & Stegun's classic Handbook

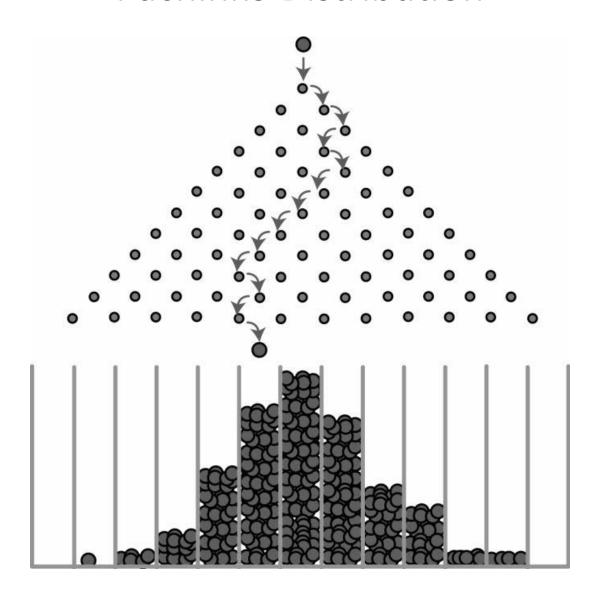


- This was neat but I did not understand it at all!
- Where did all those magic numbers come from?
- I wanted to base my approach after something tangible

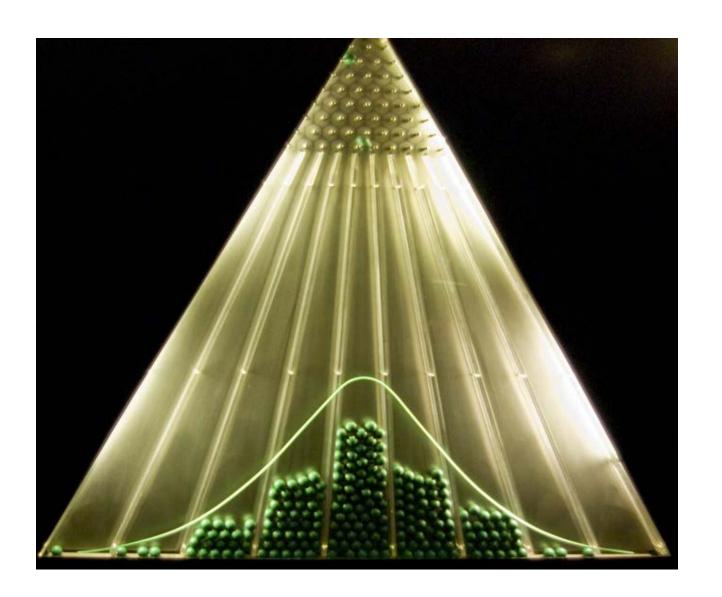
Pachinko Distribution



Pachinko Distribution

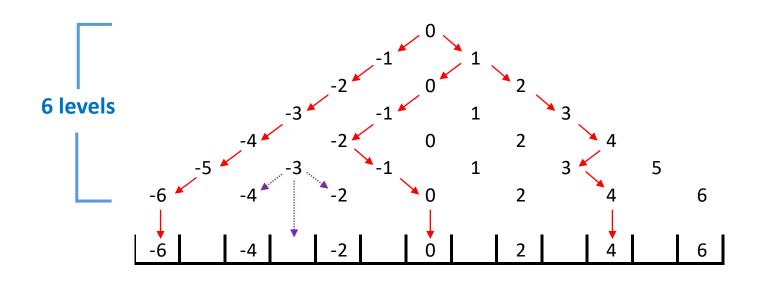


Pachinko Distribution



- We can simulate dropping balls down a Pachinko board where at each level a ball can move one step to left or right
- If we drop enough balls through enough levels, and we accumulate the count of balls at each slot at the bottom, then we should be able to simulate a normal distribution
- We will run a chi-squared test to see if our code simulates a distribution that has a reasonable deviation from the perfect (pure) normal distribution
- If the discrepancies are statistically significant, then we cannot trust that our algorithm is producing a "good enough" normal distribution to use in scientific simulations

Open Lab 4 – Pachinko Normal



int DropBall()

```
The range is ½
the number of
levels

int slot{};
for (int level{}; level < levels; level++) {
    int step = distribution(generator);
    if (step == 0)
        slot--;
    else
        slot++;
}
slot = slot / 2;
return slot;

int slot{};
for (int level{}; level < levels; level++) {
    int slot{};
    int slot{
```

View Lab 4 – Pachinko Normal

```
const int balls{ 1000 };
const int levels{ 10 };
seed_seq seed{ 2016 };
default_random_engine generator(seed);
uniform_int_distribution<int> distribution(0, 1);
double mean{};
double stddev{};

const int sigmas{ 4 };
vector<int> sigCountPachinko(sigmas);
vector<int> sigCountNormal(sigmas);
```

```
int main()
{
    CalcMeanPachinko();

    ResetPachinkoDistribution();

    CalcStdDevPachinko();

    ResetPachinkoDistribution();

    CountBallsPerSigma();

    DisplayBallsPerSigma();

    return 0;
}
```

```
void ResetPachinkoDistribution()
{
    generator.seed(seed);
    distribution.reset();
}

void CalcMeanPachinko()
{
    for (int ball{}; ball < balls; ball++)
        mean += DropBall();
    mean /= balls;
}

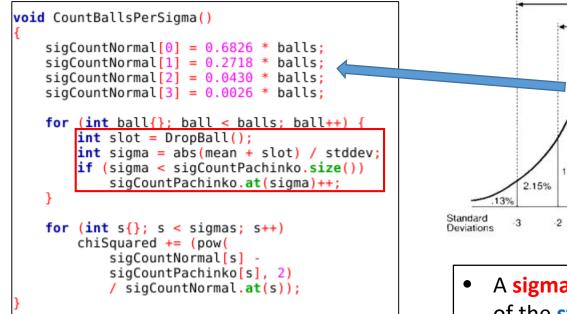
void CalcStdDevPachinko()
{
    double variance{};
    for (int ball{}; ball < balls; ball++)
        variance += pow(DropBall() - mean, 2);
    stddev = sqrt(variance / balls);
}</pre>
```

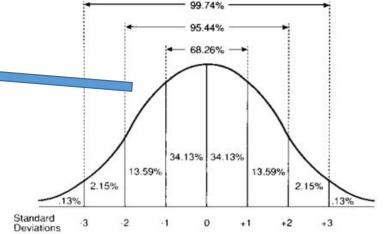
The mean slot should be = 0

View Lab 4 – Pachinko Normal

```
const int balls{ 1000 };
const int levels{ 10 };
```

```
const int sigmas{ 4 };
vector<int> sigCountPachinko(sigmas);
vector<int> sigCountNormal(sigmas);
```





- A sigma is a integral multiple of the standard deviation
- Each slot is belongs to a sigma
- We count the number of balls that fall into each sigma

Chi-Squared Test

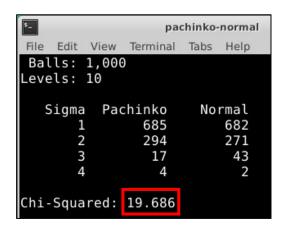
 Does the Pachinko distribution perform reasonably well compared to a perfect normal distribution?

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

```
for (int s{}; s < sigmas; s++)
  chiSquared += (pow(
        sigCountNormal[s] -
        sigCountPachinko[s], 2)
        / sigCountNormal.at(s));</pre>
```



Run Lab 4 – Pachinko Normal



For 4 degrees of freedom, the

19.686 deviation is statistically

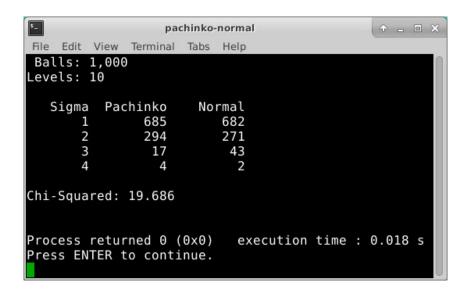
significant (> 9.49) so the proposed algorithm is not generating reasonable normally distributed probabilities!

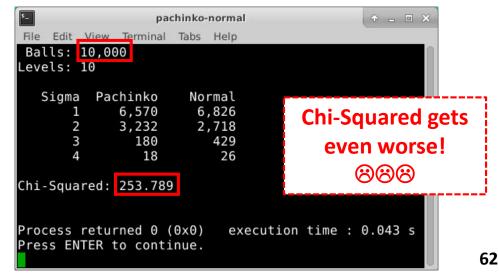
Degrees of Freedom	Probability										
	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.01	0.001
1	0.004	0.02	0.06	0.15	0.46	1.07	1.64	2.71	3.84	6.64	10.83
2	0.10	0.21	0.45	0.71	1.39	2.41	3.22	4.60	5.99	9.21	13.82
3	0.35	0.58	1.01	1.42	2.37	3.66	4.64	6.25	7.82	11.34	16.27
4 5	0.71	1.06	1.65	2.20	3.36	4.88	5.99	7.78	9.49	13.28	18.47
5	1.14	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	15.09	20,52
6	1.63	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	16.81	22.46
7	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	18.48	24.32
8	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	15.51	20.09	26.12
9	3.32	4.17	5.38	6.39	8.34	10.66	12.24	14.68	16.92	21.67	27.88
10	3.94	4.86	6.18	7.27	9.34	11.78	13.44	15.99	18.31	23.21	29.59
	Nonsignificant								S	ignifica	nt

Approaching a Normal Distribution

Maybe we didn't let enough balls drop through to get a good estimate?

What if we used 10x more balls?





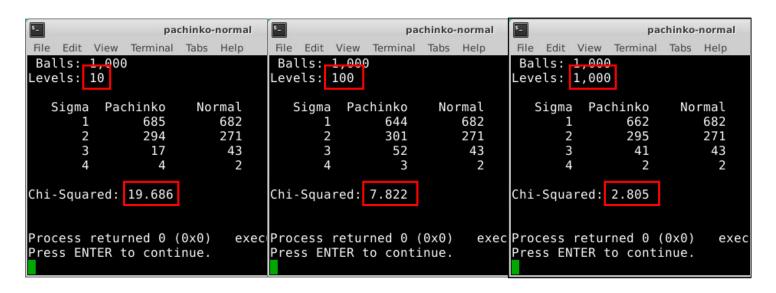
Approaching a Normal Distribution

- So can the Pachinko model accurately mimic a normal distribution? Yes! ...but only under the right circumstances
- It turns out that it is not just the number of balls that are used in the experiment that matters, but also the number of levels in the simulated Pachinko board
- The levels affects how wide (displacement from the center slot) a ball can fall left or right from the topmost (first) pin
- We have to ensure we have enough levels (therefore enough width at the bottom level) to ensure the balls can spread out during their fall to occupy the side slots that represent the higher sigma values

Approaching a Normal Distribution

Increasing the number of levels improves the agreement of the sigma ball count between the Pachinko and perfect normal distribution, thereby decreasing the chi-squared value, until the difference is no longer statistically significant

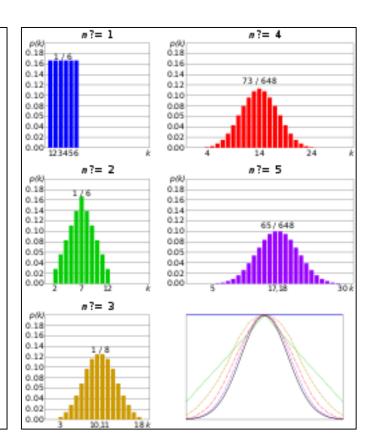
Degrees of Freedom	Probability										
	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.01	0.001
1	0.004	0.02	0.06	0.15	0.46	1.07	1.64	2.71	3.84	6.64	10.83
2	0.10	0.21	0.45	0.71	1.39	2.41	3.22	4.60	5.99	9.21	13.82
3	0.35	0.58	1.01	1.42	2.37	3.66	4.64	6.25	7.82	11.34	16.27
4 5	0.71	1.06	1.65	2.20	3.36	4.88	5.99	7.78	9.49	13.28	18.47
5	1.14	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	15.09	20,52
6 7	1.63	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	16.81	22.46
7	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	18.48	24.32
8	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	15.51	20.09	26.12
9	3.32	4.17	5.38	6.39	8.34	10.66	12.24	14.68	16.92	21.67	27.88
10	3.94	4.86	6.18	7.27	9.34	11.78	13.44	15.99	18.31	23.21	29.59
Nonsignificant							S	ignifica	nt		



Central Limit Theorem

Central limit theorem

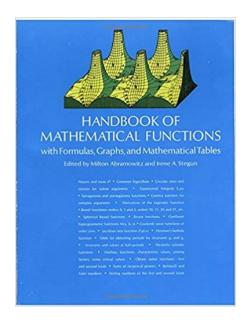
In probability theory, the **central limit theorem** (**CLT**) establishes that, for the most commonly studied scenarios, when independent random variables are added, their sum tends toward a normal distribution (commonly known as a *bell curve*) even if the original variables themselves are not normally distributed. In more precise terms, given certain conditions, the arithmetic mean of a sufficiently large number of iterates of independent random variables, each with a well-defined (finite) expected value and finite variance, will be approximately normally distributed, regardless of the underlying distribution. [1][2] The theorem is a key concept in probability theory because it implies that probabilistic and statistical methods that work for normal distributions can be applicable to many problems involving other types of distributions.



Approximating a Normal Distribution

It is best to use the code from Abramowitz & Stegun, as only **one** call to this function is required to produce a normalized random variable versus to my method:

(1000 balls x 1000 levels = 1M calls!)



The C++ has a built-in normal_distribution()

```
int main()
    std::random device rd{};
    std::mt19937 gen{rd()};
   // values near the mean are the most likely
   // standard deviation affects the dispersion of generated values from the mean
   std::normal distribution<> d{5,2};
    std::map<int, int> hist{};
    for(int n=0; n<10000; ++n) {
        ++hist[std::round(d(gen))];
    for(auto p : hist) {
        std::cout << std::setw(2)
                  << p.first << ' ' << std::string(p.second/200, '*') << '\n';
```

Now you know...

- Rational, Irrational, and Transcendental numbers each have their own style of continued fractions
 - We can take any real number and generate a CF
 - Given a CF, we can *expand* it to regain the original number
- The convergents of a CF are excellent approximations to the original number
- The magnitude of the **x** & **y** values in solutions to Pell's Equation $\{x^2-ny^2=1\}$ is related to the period of the simple continued fraction of \sqrt{n}
- Memorizing thousands of digits of π is okay but I'd rather appreciate its beautifully simple GCF: [3; 1, {6|(2n+1)^2}]

Now you know...

- A perfect Normal distribution ensures that 68.26% of all values fall within one (1) standard deviation from the mean
 - 99.73% of all values in a perfect normal distribution are within **three** (3) standard deviations from the mean
 - The normal distribution is known as the "bell curve"
- There is a way to convert a PRNG created uniform distribution into a normal distribution – but don't use the Pachinko method
- The chi-squared test suggests if the discrepancies between the observed and the expected values are <u>statistically</u> <u>significant</u>