Number Walk Program

Source code can be found at: http://carma.newcastle.edu.au/summer/jake/
Outcomes

• Number Walk GUI
  – Developed with python “Tk interface” module
  – Base program provided by Fran Aragon
  – Built from console program to standard user interface
  – Incorporated many new functions
• Learnt programming languages:
  – Python
  – Java
  – C++
• Pseudo L-series Analysis
Objective Of Number Walk

• Assess the normality and randomness of numbers “at a glance”:
  – Visualise numbers
  – Quantitative measurements
  – User friendly interface
  – Compatible data output
Drawing The Visualisations

- Number read as a string
- Bases 2 through 37 are supported
- For base $b > 2$:
  - $\phi = \frac{360}{b}$
  - Walk direction for digit $i$ given by $i \times \phi$ degrees from x axis counterclockwise
  - Each step is one pixel width
- Base 2 is a special case
Normality

• Given a natural number $b$, a real number is said to be base $b$ normal if, in the base $b$ expansion of the number, the limiting frequency of a string of digits of length $m$ is

$$\frac{1}{b^m}$$

• Random sequences are normal, but other numbers are normal too.
Measuring Randomness: Distance From Origin

For a random walk:

\[ D(n) \to \frac{\sqrt{\pi n}}{2} \quad \text{as} \quad n \to \infty \]

Where \( D(n) \) is the distance from the origin after \( n \) steps.

Thus a plot of \( \frac{2}{\sqrt{\pi n}} D(n) \) vs. \( n \) will allow quick indication of randomness.
Measuring Randomness: Fractal Box Dimension

Fractal box dimension of a figure is defined as:

\[ d = \lim_{x \to 0} \frac{\log(N(x))}{\log\left(\frac{1}{x}\right)} \]

Where \( N(x) \) is the number of squares of width \( x \) it takes to cover a figure.

A plot of \( \log(N(x)) \) vs. \( \log\left(\frac{1}{x}\right) \) gives the approximate box dimension as the gradient.
Selected Walks
This is a spiral in base 4 designed to test the program.
\[ C_{10} = 0.1234567891011213141516 \ldots \]
$C_4 = 0.123101112132021222330 \ldots$
\[ \alpha_{2,3} = \sum_{n=1}^{\infty} \frac{1}{3^n 2^{3n}} \]
pi_b6 1000.0k steps
e_b4_1000.0k steps

e = \exp(1) = \text{sum}(1/n!, n=0...\infty)

2.7... in base 1
Euler's Constant (gamma) in base 10
= \lim_{n \to \infty} \left[ \left( \sum_{k=1}^{n} \frac{1}{k} \right) - \ln (n) \right]

source: http://www.numberworld.org/constants.html

\gamma = \lim_{n \to \infty} \left[ \left( \sum_{k=1}^{n} \frac{1}{k} \right) - \ln (n) \right]
This is the rational number = 
3624360069/7000000001
EB(2) = \sum_{n=1}^{\infty} \frac{1}{2^n - 1}
Genome Walker

- The program also features genome walking capabilities.
- Reads sequence of nucleotides and plots with the convention shown to the right.
Aphid lethal paralysis virus bg_8.655k steps

Aphid lethal paralysis virus base 4:
- dim = 1.2105381837

Aphid lethal paralysis virus in base 4
- Distance Normed Avg = 11.5022714829

Data
Least Squares Line
d=1.2 lines
RandWalk Result

Graph showing distance over number of steps.
E_coli_536_base_4:
dim ~ 1.39237075799

A gene from E.Coli bacteria
HIV-1_bg 11.632k steps
\[ EB(2) = \sum_{n=1}^{\infty} \frac{1}{2^n - 1} \]
Demonstration
Conclusion

• The normality of naturally occurring irrational numbers is difficult to analyse.

• Number Walk was built to help direct proofs of normality or non-normality:
  – Provides a tool for distinctly visualising the base-n expansion of a number.
  – Flexible and user friendly.
  – Can be used in imaging genome sequences.

• The images raise interesting questions:
  – Why do pi and e produce such vastly different images to more contrived numbers and genomes?
  – What does this mean?