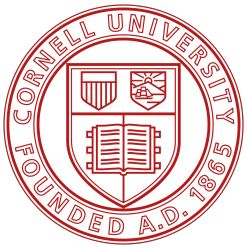


Cornell University

Numb3rs Math Activities

Contents

The TV show Numb3rs begins each week with:

We all use math every day; to predict weather, to tell time, to handle money. Math is more than formulas or equations; it's logic, it's rationality, it's using your mind to solve the biggest mysteries we know.

We have developed materials on the mathematics behind each of the episodes of the series. We welcome comments, suggestions, and contributions to these pages. Send them to the project director, Rick Durrett (rtd1(at)cornell.edu).

We use the original numbering which is on the DVD. This differs somewhat from the numbers used on the NCTM page.

101: Pilot

In this episode, a serial murder-rapist has made numerous attacks in the Los Angeles area. Making some seemingly harmless basic assumptions, Charlie builds a statistical model of the attacker's behavior which helps the FBI stop the murders.

What does practical mean?

A model of the attacker's behavior could be a number of different things. Don and the FBI want to know where the next attack will be. Charlie points out that this might be the incorrect approach to the problem by making an analogy to a sprinkler. He proposes that finding the sprinkler would be easier than deducing the location of the next point where a drop will hit. In many ways, Charlie is probably correct. We can probably assume that the killer has an apartment where he or she spends quite a bit of time. If we can find the most likely neighborhood where the attacker resides, it should be easier on FBI resources than sending a dozen agents to patrol several neighborhoods searching for an attack in progress. A second reason, going back to the analogy of the sprinkler, is that the physics of finding the next sprinkler drop are quite complicated; by this, Charlie means that regardless of how many drops we've seen hit the ground, the area where we should look for the next drop to hit will be quite large. In the best possible model, more data should provide significantly better deductions. Since the sprinkler is stationary; however, the seeming randomness of the action of physics on each droplet will affect Charlie's model less and less as the number of droplets are observed.

We will define a model of the attacker's behavior to be a function $p(x)$ from the addresses in Los Angeles to the unit interval (the interval $[0, 1]$). That is to say that if x is a location in Los Angeles, then $p(x)$ is a number between 0 and 1. Not just any function will do, however. We require the function $p(x)$ to have the following property: when we sum the values $p(x)$ over all addresses, the result is 1. We then call $p(x)$ the probability that x is the killer's address. Where $p(x)$ is higher, the assailant is more likely to reside.

After thinking about the problem briefly, we realize that we have our hands full: there are infinitely many models to choose from. We somehow need to find the right one. However, we don't even have a concept of what right means. In laymen's terms, we need the model to "fit the data." How do we quantify this common notion so that we may make mathematical deductions? Charlie makes a point that when a person tries to make a bunch of points on a plane appear randomly distributed, the result is that the points adjacent to any given point x are all approximately the same distance away from x. Charlie uses this information to produce his model. As one can see from the map Charlie brings to the FBI, the "hot spot" — the most likely area where the perpetrator lives —

which Charlie computes is in what we might imagine is the "center" of the attacks.

Unfortunately, Charlie's model fails The FBI gets DNA samples from every resident of the neighborhood Charlie says they should check, but none of the DNA matches the perpetrator's. So, Charlie has to ask himself if the model he made was good. He sees one data point which appears anomalous. However, any model with the properties we outlined above shouldn't be affected too much by a single data point. Indeed, after fixing the error, Charlie's new model has a smaller hot zone which lies completely within the one the FBI already checked. He is stunned by the realization that his model is bad.

Eventually, Charlie realizes that he has made a classic error. It is sometimes quipped that the only difference between physicists and engineers is that physicists can be sloppy in their approximations. A physicist wanting to produce a set of laws of physics which is as complete as possible will not choose the most complicated set of laws before trying out a simpler set first. In the same way, Charlie chooses the mathematically practical approach by choosing the simplest possible solution to his problem by assuming there would be exactly one hot zone. It is mathematically practical in the sense that solving this complicated problem is a lot easier with one hot zone than two. Generally this is a good way to approach a problem; at worst one learns why the easy approach does not work which hopefully gives some clues as to what the more complicated approach should look like. With two hot zones (one representing home and the other representing work), Charlie's model gives more accurate results: one hot zone is in the same neighborhood as before, the other is in an industrial area, and the hottest parts of the hot zones are quite small. After making the arrest, Don notices that the perpetrator had moved from the original hot zone a few weeks ago, which is why the FBI hadn't found him in their original search.

How did Charlie produce his model?

Of all the possible models, how did Charlie find that specific one? Not many clues are given in the episode as to what method he uses. So, let's consider the following more tractable problem. Suppose we have done the following experiment: we hung a spring from the ceiling and measured the lengths of the spring after attaching a weight. After doing this for several different weights, we have collected a bunch of data. After making a graph of weight versus change in length, we might notice that the points form almost a line (assuming the weights aren't too heavy). This means that if the weight is W and the change in length is L, $W = kL$ for some number k (this relationship is called Hooke's law after the British physicist born in the 1600s). How do we compute k? We could draw in a bunch of lines which seem to approximate our data well and pick the one which is the best. This is called the linear regression problem. But which one is the best? There are many different concepts of that, and the simplest one isn't the one we generally use.

Let M denote the set of all lines through the origin. M is our set of models. Let S denote the set of points which we have computed experimentally. Given any line $W(L)$ in the set M and data point (x, y), compute the quantity $|W(x) - y|$, the vertical distance between the line W and the point (x, y). Add up the quantities $|W(x) - y|$ for each point in S. This gives us a mapping from the set of models to the non-negative real numbers. Generally, a map from a set of functions (in this case, lines) to the real numbers is called a functional. Denote our functional by $A(W)$. If we can find a line for which $A(W) = 0$, then our data points are all colinear. This is generally not going to happen. The next best thing would be to find a line W for which $A(W)$ is as small as possible. In this case, we say that such a line W minimizes $A(W)$.

So now we must find a line which minimizes $A(W)$. When one wishes to minimize a quantity, one generally uses differential calculus. Unfortunately, the absolute value function has no derivative at zero. So, square the distance first! Instead of adding up $|W(x) - y|$ to create $A(W)$, we add up $|W(x) - y|^2$. This is called the method of least squares and is generally the accepted method of solving the linear regression problem. To solve the problem requires a little calculus. Wolfram's website has a relatively good explanation.

Notice that there was nothing special about the line here. Any set of functions M and non-negative functional $A(W)$ would have worked (although there will be technical problems if M and A are not chosen wisely, such as not being able to find a minimizing function inside our set M). So we could have found the quadratic polynomial of best fit or the exponential of best fit in a similar fashion, although solving such a problem will no doubt be vastly more complicated.

This method is a general approach used by mathematicians in a variety of situations. The difficulty is generally in proving the existence of a function which minimizes whichever functional we have decided to work with. In physics, one often hears that objects take the path of least action or least energy. That is to say that rather than solving a very complicated differential equation, we could solve a variational problem instead. This method is so useful that it is basis of theoretical physics.

Random sequences?

Charlie makes a comment that to consciously construct a random sequence is impossible. This is true in many ways. As an example, consider the following property of sequences due to Khinchin. Given any real number x we can find a continued fraction

expansion

$$x = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{\ddots}}}}$$

In fact, as long as x is irrational, the continued fraction expansion is unique. Khinchin's Theorem says that

$$\lim_{n\to\infty} \left(\prod_{i=1}^{n} a_i \right)^{1/n} = K_0$$

for almost every continued fraction. Here almost every refers to a somewhat complicated notion from measure theory. Luckily, in our case it means exactly what it sounds like. What is interesting is that no one has been able to demonstrate a continued fraction which has the property given above. So, truly, we are not very good at coming up with random sequences at all!

102: Sabotage

Don investigates a series of train wrecks that recreate accidents due to railroad negligence. A numerical code is left at the site of each accident. Charlie helps Don to find the terrorist behind the recreations by breaking the codes, which contain statistics of the wrecks occurred in the past.

The art of breaking codes by analyzing patterns is part of a wider mathematical area called Cryptography. In Episode 205 the reader can find a brief explanation of some of the simplest algorithms to *encode* and *decode* information (we also refer the reader to Episode 324). In order to break the code in this episode, Charlie also uses *statistical data analysis*, a technique mentioned in Episode 211 as well.

By the end of the episode, Charlie gives a passionate speech about the way nature communicates to us in terms of mathematical patterns. We quote his exact words here:

> *"Math is the real world, okay, it's everywhere, okay. Can I show you? You see how the petals spiral? The number of petals in each row is the sum of the preceding two rows, the Fibonacci Sequence. It's found in the structure of crystals and the spiral of galaxies and a nautilus shell. What's more, the ratio between each number in the sequence to the one before it is approximately 1.61803, what the Greeks call the Golden Ratio. It shows up in the pyramids of Giza and the Parthenon at Athens, the dimensions of this card. And it's based on a number we can find in a flower. Math is nature's language... its method of communicating directly with us. Everything is numbers."*

Below we will explain the main properties of the **Fibonacci sequence** and the **Golden Ratio**, and we will formally establish the relationship between these two mathematical entities.

The Fibonacci Sequence

The Fibonacci sequence is a sequence of numbers, $(F(n))_{n\geq 0}$, generated recursively in the following way,

$$\begin{aligned} F(0) &= 0 \\ F(1) &= 1 \\ F(n) &= F(n-1) + F(n-2) \text{ for } n \geq 1. \end{aligned}$$

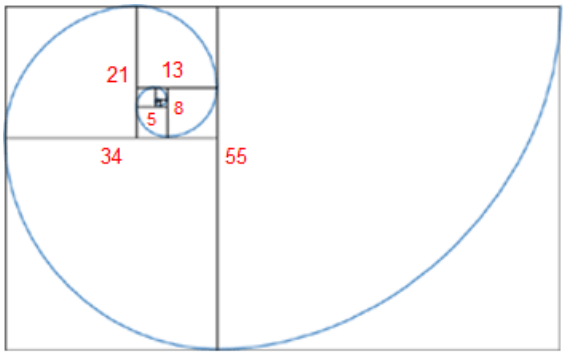
Tangent

The Fibonacci sequence or **Fibonacci numbers** are named after Leonardo of Pisa (about 1175–1250), who was nicknamed Fibonacci (from fillies Bonaccio, i.e. son of Bonaccio). He was investigating (in the year 1202) how fast rabbits breed in ideal circumstances and found out that the number of pairs of rabbits, female and male, in the population increases according to the mentioned sequence.

In words, every number in the sequence is equal to the sum of the two previous ones. The numbers in this sequence get arbitrarily big as n increases. The first 11 numbers in this sequence are shown below:

$$0,\quad 1,\quad 1,\quad 2,\quad 3,\quad 5,\quad 8,\quad 13,\quad 21,\quad 34,\quad 55\ldots.$$

Charlie mentions that this sequence is found in the spiral structure of some flowers and galaxies. Here we explain what he referred to. Suppose that you have two squares of side-length equal to 1, sharing one of the sides. Then you put a square of side-length 2 on top of them. Then you draw a square of side-length 3 on the left of the rectangle drawn before. And you continue this way, adding squares of side-length equal to the sum of the side-lengths of the two previous squares, as shown in the figure on the right. Then, by construction, the lengths of the squares increase according to the Fibonacci sequence, and if you draw a curve joining the opposite vertices of the squares we obtain a spiral as shown in the figure as well.



These spirals appear in nature in numerous examples, such as sea shells, flowers and galaxies. The reader can find more about the appearance of the Fibonacci sequence in nature at http://www.maths.surrey.ac.uk/hosted-sites/R.Knott/Fibonacci/fibnat.html

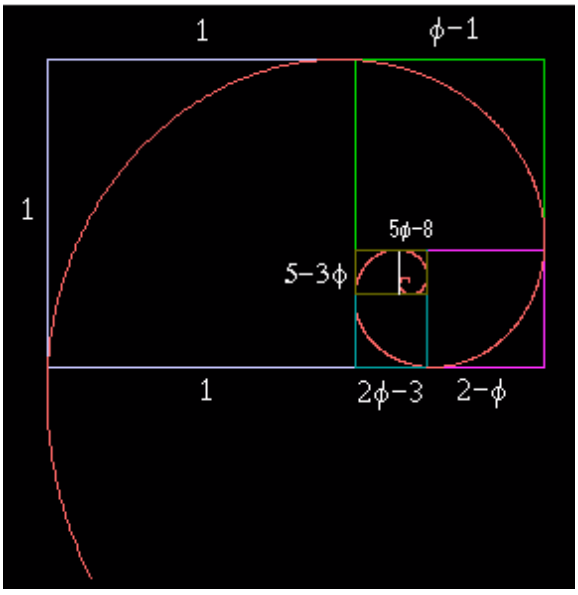
The Fibonacci numbers satisfy a numerous amount of very interesting identities. The reader can find some of them at http://mathworld.wolfram.com/FibonacciNumber.html.

We will concentrate our attention here to its relationship with the golden ratio, which we explain below.

The Golden Ratio

Suppose that you have a rectangle with sides of length 1 and x. Then you partition the rectangle into a square with side length 1 and another rectangle of side lengths 1 and x-1. The golden ratio or golden proportion is the only positive number x such that the two rectangles obtained by this construction are similar, i.e.

$$\frac{x}{1} = \frac{1}{x-1}.$$



The golden ratio is usually denoted by the Greek letter ϕ, and according to the equation above we find that ϕ satisfies the following quadratic equation

$$\phi^2 - \phi - 1 = 0,$$

which implies that $\phi = \dfrac{1+\sqrt{5}}{2} \approx 1.61803398874989\cdots$.

This number possesses many nice properties. For instance its continued fraction rep-

resentation (Episode 101) only contains ones, i.e.

$$\phi = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{\ddots}}}}.$$

Tangent
The golden ratio or golden proportion was known by the ancient Greeks, it occurs in some of the Platonic Solids and it is mentioned in Euclid's Elements. Although this golden proportion is found in some of the Ancient Greek constructions, such as the Parthenon, there is no definite evidence that they were designed by using this mathematical constant (see picture below). The reader can find more about the golden ratio and architecture at Dr. Ron Knott's website.

There is also a nice way of expressing the golden ratio as a limit of radicals in the following way,

$$\phi = \sqrt{1 + \sqrt{1 + \sqrt{1 + \sqrt{1 + \cdots}}}}$$

Regarding the construction at the beginning of the section, if we continue partitioning the successive rectangles into a square and a new rectangle, we deduce that all the rectangles are similar, and if we draw a curve joining the opposite vertices of the squares we obtain a spiral similar to the one generated by the Fibonacci sequence. Below we explain the algebraic relationship between the golden ratio and the Fibonacci sequence, which explains the similarity of the mentioned spirals, The Golden Section in Architecture.

The Relationship Between the Golden Ratio and the Fibonacci Sequence

The quadratic formula satisfied by the golden ration implies that

$$\phi^2 = \phi + 1 \longrightarrow \phi^n = \phi^{n-1} + \phi^{n-2} \text{ for } n \geq 2$$

Since $1-\phi = -\phi^{-1}$ satisfies the same equation, we get that

$$(1-\phi)^n = (1-\phi)^{n-1} + (1-\phi)^{n-2} \text{ for } n \geq 2.$$

Define

$$\widetilde{F}(n) = \frac{\phi^n - (1-\phi)^n}{\sqrt{5}} = \frac{\phi^n - (-\phi)^{-n}}{\sqrt{5}}.$$

The two facts above imply that $(\widetilde{F}(n))_{n\geq 0}$ satisfies the Fibonacci recursion as well and since

$$\widetilde{F}(0) = 0$$
$$\widetilde{F}(1) = \frac{\phi - (1-\phi)}{\sqrt{5}} = \frac{2\phi - 1}{\sqrt{5}} = 1.$$

We conclude that $\widetilde{F}(n) = F(n) = \frac{\phi^n - (1-\phi)^n}{\sqrt{5}} = \frac{\phi^n - (-\phi)^{-n}}{\sqrt{5}}$.

Tangent
Binet's formula is named after the French mathematician Jacques Philippe Marie Binet, who derived it in 1843. However, this formula was already known by Euler, Daniel Bernoulli, and de Moivre more than a century earlier.

This is the famous Binet's formula, which gives us a closed expression for the terms of the Fibonacci sequence in terms of n. This formula allows us to prove properties of the Fibonacci sequence, in particular we can prove that,

$$\lim_{n\to\infty} \frac{F(n+1)}{F(n)} = \lim_{n\to\infty} \frac{\phi^{n+1} - (1-\phi)^{n+1}}{\phi^n - (1-\phi)^n} = \phi.$$

This formula in words tells us that as n increases, the ratio between consecutive terms of the Fibonacci sequence approaches the golden ratio. This is exactly what Charlie is referring to when he says: "…the ratio between each number in the sequence to the one before it is approximately 1.61803, what the Greeks call the Golden Ratio."

103: Vector

In this episode, a deadly virus is spreading across Los Angeles.

What is Graph Theory?

A graph is defined as two sets V and E. V is any non-empty set and is called the vertex set. Elements of E are pairs of elements of V, and E is unsurprisingly called the edge set. Graph theory is the study of properties of these graphs with applications ranging from computers to social networking to epidemiology. Suppose you draw a hundred points on a piece of paper. You pick one special point v and play connect the dots, draw edges from one vertex to the next. Then you start back at v and do it again. You repeat the process a dozen or so times. You then ask your friend to figure out which vertex you started from. This is essentially the problem the FBI and CDC are working on, to find the first infected person(s), only there are hundreds of thousands of people all linked together in the contaminated area, so the problem is vastly more complex.

The problem (and beauty) of graph theory is that many problems can be simply stated and sometimes even relatively simply solved, but from a practical standpoint are all but impossible. In other words, one can prove the existence of something with relative ease, but giving an example of one could be extremely difficult. One such problem is the Ramsey coloring problem. We call a graph complete if every two vertices are joined by an edge. K_n denotes the complete graph on n vertices. A 2-colored graph is a graph G together with a partition of the vertex set into two subsets. Suppose we agree to call the two colors red and blue. Given a 2-colored complete graph G, is there a number $R(r)$ so that if G consists of more than R vertices then G has a complete monochromatic subgraph on r vertices? Here monochromatic means that all the vertices in the subgraph are all in one of the two partitioned subsets. Ramsey proved that yes, there is. Further, he extended the result to the following question: is there a number $R(r, s)$ so that if there are more than R vertices in a 2-colored complete graph G then there is either a complete red subgraph on r vertices or a complete blue subgraph on s vertices. Again the answer is yes. The smallest number $R(r, s)$ is called a Ramsey number. Note that when $r = s$, this new problem corresponds to the original problem. Even for very small numbers, the Ramsey numbers are not known precisely. $R(5, 5)$, for example, is not known precisely, although it is known to be between 43 and 49. The prolific twentieth-century mathematician Erdős expresses the computational difficulty of the problem with the quip,

> "Imagine an alien force, vastly more powerful than us landing on Earth and demanding the value of $R(5,5)$ or they will destroy our planet. In that case, we should marshal all our computers and all our mathematicians and attempt to find the value. But suppose, instead, that they asked for $R(6,6)$, we should attempt to destroy the aliens."

Proofs of very simple theorems are quite complicated. The Four Color Theorem is an example of such a problem. A planar graph is a graph that can be drawn on a plane without having its edges cross. Another way to think of such a graph is by taking the plane, cutting it up into regions, and drawing a graph by putting a vertex inside each region and then drawing edges between vertices whose regions border one another. The Four Color Theorem says that given a planar graph, one can color the vertices with 4 colors so that no adjacent vertices are of the same color. The question comes from the desire to color a map of nations so that no bordering nations are the same color. This is slightly different since nations are not generally contiguous. For example, during the colonial times of the Great Britain, that nation had numerous colonies across the globe. The Four Color Theorem is known for having one of the least elegant proofs of all time. The original proof of the theorem reduced the problem to almost 2000 cases which were then handled using a computer.

Try computing $R(3,3)$.

A bipartite graph is a graph whose vertex set can be partitioned into two sets so that vertices in each partition are only joined by edges to vertices in the other partition. A graph is connected if there is a path between any two vertices (where path has the intuitive meaning). A minimally connected graph is called a tree. Prove that every tree is bipartite. A cycle is exactly what it sounds like. Prove that a cycle with an even number of vertices is bipartite.

A complete bipartite graph is a bipartite graph with the property that each vertex in one partition is joined to every vertex in the other partition, i.e. it is a bipartite graph with the most possible edges. $K_{m,n}$ denotes the complete bipartite graph partitioned into sets of m and n vertices. There is an interesting relationship between a graph being planar and the graphs $K_{3,3}$ and K_5. Learn more about that.

Run of the Yang-Mills

When Charlie visits Larry at the restaurant, Larry makes a very nerdy joke. The punch line is that it's "just a run of the Yang-Mills black hole." Yang-Mills (and the mass-gap problem) is a problem of mathematical physics. Yang-Mills is the underpinning of elementary particle theory that has made many correct and testable predictions, but the

mathematics is not at the level of rigor required of mathematicians. One of the major problems is establishing that there is a Yang-Mills theory with the property that quantum particles have positive mass. This mass-gap problem is one of the $1,000,000 Clay Institute Millennium Problems.

104: Uncertainty Principle

In this episode, a series of bank robberies have occurred, none of them violent. After Charlie helps figure out where the next robbery will occur, however, the robbers open fire and kill an agent. By intervening, the FBI has changed the methods of the robbers and revealed an even more dubious plot.

What is the Heisenberg Uncertainty Principle?

Charlie tells Don that the Heisenberg Uncertainty Principle says that the act of measurement in a system affects the system. This is, as Charlie admits later on, not actually what the principle says. In fact, Charlie's statement is not especially profound: when trying to measure very precisely the location of a piece of dust floating in the air, we will naturally cause the air to move around and thus cause the dust particle to move. This would not have been news to physicists. What the Heisenberg Uncertainty Principle says is that one cannot with perfect accuracy determine the position and velocity of anything — especially not an electron. The products of the errors is on the order of Planck's constant which is about 10^{-34} Joules-seconds. A Joule-second is a unit which is derived from only macroscopic quantities (like a kilogram and a meter). So, in macroscopic terms, the error is completely negligible. The smaller the things we are measuring, the more important the principle becomes.

The Principle, at its most basic roots, has nothing to do with physics at all. In fact, in its general form, it is a statement about the relationship between a function and its Fourier transform. It turns out that the Fourier transform is fundamental to quantum mechanics, so naturally it has physical implications. The American Mathematical Society has an excellent article on the subject which also explains the Fourier transform as well.

What is P vs NP?

The P vs NP problem is a problem from the field of logic (or theoretical computer science). Suppose we have a problem we want a computer to solve. For example, we want a computer to find a solution to the Traveling Salesman problem. Suppose a salesman needs to travel to 50 cities across the globe. We know how much the flights cost between each city and the salesman wants to spend as little as possible flying. This clearly has a solution as there are "only" $50 \cdot 49 \cdot 48 \cdot \cdots \cdot 3 \cdot 2 \cdot 1$ different paths to choose from. No one wants to do this out by hand, so we want to program the computer to do it. But how long will it take? This question is formulated in the following way: suppose that we have n locations for the problem. Is there an algorithm which requires $O(f(n))$ steps? Here the big-O notation means that for all sufficiently large values of n, the number of steps

required by our algorithm will be smaller than our function $f(n)$. In general, we say that a problem is in an f-complexity class if there is an algorithm which solves the problem in $O(f(n))$ steps.

The P-complexity class consists of all YES/NO problems which are solvable in polynomial time. That is P consists of the set of problems for which there exists an algorithm which can check a possible solution to the problem in $O(f(n))$ for some polynomial $f(n)$. We say that P consists of problems for which there is a polynomial time algorithm for **checking solutions**. The NP-complexity class is somewhat more complicated. Basically, though, the NP-class consists of problems for which there is a polynomial time algorithm for **finding** solutions. The P vs NP problem asks if P and NP are actually equal. If the algorithm needed to verify a solution of a problem is in polynomial time, is there an algorithm which solves the problem of **finding a solution** in polynomial time? It seems like a somewhat silly question at first since solving a problem and verifying a solution are almost the same thing to a computer, but it is the one of the biggest open mathematical problems in the world. The Clay Mathematics Institute has offered $1,000,000 prize to anyone who can give a positive or negative answer to whether P=NP.

The astute viewer noticed that Charlie says Minesweeper is NP-complete. It is unclear what Charlie means by this since it has very little bearing on solving the P vs NP problem. The NP-complete problems represent a narrow subclass of NP problems which might not be P. In essence, NP-complete are the hardest NP problems. Richard Kaye proved Minesweeper is NP-complete in the Mathematics Intelligencer. One can visit his webpage on the mathematics of Minesweeper for more information. The Traveling Salesman problem is also NP-complete.

Statistically dead?

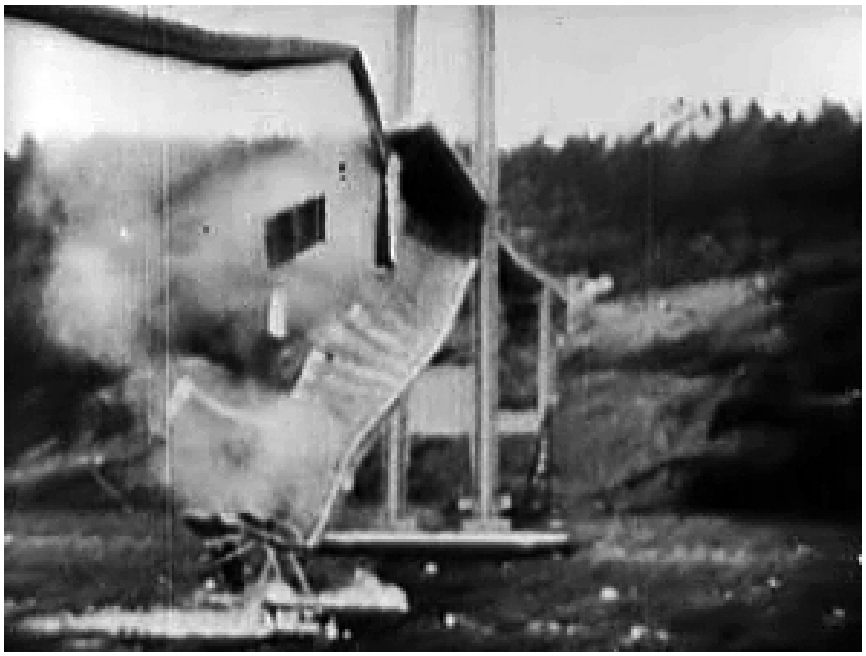
At one point Charlie says to Don that he is statistically dead. Having already been fired upon, Charlie implies that Don is less likely to survive a second attempt. It is probably true that a person who survives being fired upon once are more likely than the average person to be killed in a gun fight during their lifetime. After all, someone who is in a gunfight once is more likely to be in a subsequent gun fight since this set of people includes police officers, military personnel, mercenaries, gang bangers, and so on. Is the real implication Charlie is making correct? Specifically, is his brother is more likely to die from a gunshot wound having already been in a gunfight?

105: Structural Corruption

In this episode, a student who asked Charlie for help has died. Though labeled as a suicide, Charlie insists that Don investigate further.

Turbulence and Flow

The flow of fluids is simultaneously among the most important and most complicated areas of engineering and mathematical physics. The construction of airplanes obviously requires the understanding of air flowing across the wings; it doesn't take much stretch of the imagination to realize that the construction of high speed automobiles and trains requires careful study of aerodynamics not just to reach those high speeds but to prevent the vehicles from being torn to pieces. That buildings and bridges needed to be carefully constructed with air flow in mind came as a shock to many. The Tacoma Narrows Bridge, shown to the right, experienced a phenomena called resonance.



Imagine a young child is sitting on a swing. As is usually the case in homework problems from physics textbooks, you find that attached to the child is a large spring. You stretch and relax the spring at various constant frequencies. As you vary the frequency

of this stretching and relaxing, you notice that if the frequency is very high, the child doesn't swing very high. The same thing happens when you push and pull at very low frequency. As you increase the frequency from low to high, you notice that the child starts swinging higher and higher until at a certain frequency it peaks and then starts to fall down again. The frequency in the middle is the maximum resonance frequency. The oscillatory behavior of the child swinging (a pendulum) and the oscillatory motion of you stretching and relaxing interact with one another. At low and high frequencies, they work against one another too much and the swing won't move very much. At that maximum resonance frequency, though, the oscillation of the rubber band is amplifying the motion of the swing as much as possible.

The same phenomena is found throughout nature. The famed singing woman who can break a wine glass is theoretically possible. When sound is directed toward the glass, the bulb of the wine glass begins to stress and strain, though it looks like it is just shaking. If the sound is at the correct frequency (and of sufficient volume), then the strain from the oscillations will exceed the elasticity of the glass and shatter it. The same volume which shatters the glass will be completely useless at other frequencies. In fact if the wind hits a poorly designed bridge just right, the bridge will begin to twist to the point where it tears itself apart. The problem Charlie is working on in this episode is to determine whether the wind blowing against the side of the building could cause enough strain to buckle the building. Obviously the wind is not strong enough to physically push the building over, but the with the fluid flow across the building, factors like resonance can be huge.

The general equation which governs fluid flow is called the Navier-Stokes equation. This equation is a non-linear partial differential equation. As a result, it exhibits chaotic behavior called turbulence. This means that even though two points are relatively close to one another in position, the forces at each of those two points could be vastly different. For example, in the eye of a hurricane, there is almost no wind at all, and yet just outside this small region, the winds are strong enough to tear the roof off a house. When a plane takes off, the airflow immediately below the plane is powerful enough to throw a car, and yet the wheels of the plane are not destroyed. Chaos is a property which results from non-linearity: solutions to non-linear equations are extremely sensitive on initial conditions. This is one reason why the weather is so hard to predict. Simply approximating wind speed at a point is enough for the resulting weather patterns to be completely different from reality over a very short period of time. This is popularly called the Butterfly Effect - a butterfly flapping its wings in Brazil could be the difference between clear skies and tornados in Texas, to paraphrase the title of a talk given by the physicist Lorenz. The Navier-Stokes equation is so complicated that, despite being generally accepted as correct, no one has found a long-time solution to it. Anyone who can find such a solution (or prove none exists) will receive a $1,000,000 prize from the Clay Institute.

Fluid flow is not just important in construction. Animal flight is even more complicated than airplanes since the wings can move in complicated ways. Why do large birds fly differently from small insects? How does the bumblebee fly? Its flight was not well-understood until the last few years. This glaring inability to explain something so simple was one of the greatest failures of science. Its wing size and number of flaps per minute were not enough to explain how it could keep itself aloft. Its unique style of flight is something noticed long ago. In fact, in 1934 a french entomologist claimed that the flight of the bumblebee was aerodynamically impossible. However, in 2005 high speed digital photography was able to show how the wings of the bee flap. As mentioned in the episode, the way blood flows through the body is an important factor in forensic medicine, let alone medicine proper.

106: Prime Suspect

In this episode there were several mathematical topics that were mentioned. We will discuss the Riemann hypothesis, other Millennium problems, and encryption.

Riemann Hypothesis and Millennium Problems

As was made obvious in the episode, the Riemann Hypothesis is one of the most famous conjectures in mathematics. It was originally stated in an 1859 paper written by Bernhard Riemann, and it involves the Riemann zeta-function defined as the infinite series:

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$$

It can be shown that this series converges if the real part of s (which is a complex number in general) is greater than 1. Then it is possible to *analytically continue* the zeta function so that it is defined for all s with real part not equal to 1. This definition of the zeta function is analytic on its domain of definition, which is the complex analysis version of being differentiable. This basically means that with this formula we can define the function for part of the complex plane, and then we can show that there is a way of extending this to almost all of the complex plane that agrees nicely with the original definition. The Riemann Hypothesis is that all the zeros of the continued zeta function occur either when s is a negative even integer or when the real part of s is $\frac{1}{2}$. It is important in mathematics because it has many deep connections to prime numbers and especially to the distribution of the prime numbers (how often they occur), and there are many interesting results which have been proved to be true assuming that the Riemann hypothesis is true. For more information about various connections between other parts of math, look up the Wikipedia page about it.

As was also mentioned in the episode, the Riemann hypothesis is one of the *Millennium Prize Problems*. These are 7 (actually, the number recently became 6) famous mathematical problems which come with a prize of one million dollars offered by the Clay Mathematics Institute for anyone who solves them. This institute was established by Landon T. Clay and is "dedicated to increasing and disseminating mathematical knowledge." Here is a list of the problems with short descriptions:

- P versus NP: This problem deals with how hard different problems are to solve on a computer based on the length of the input for the problem. (Here the particular problem is fixed and the length of its input is allowed to vary.) A problem S is in the

set P if there is a polynomial $p(x)$ and a computer algorithm which given an input for S of length n can solve the problem in time $p(n)$. An example is the problem of sorting a list of n numbers, which can be done in polynomial time. A problem is an NP problem if given an input and a possible solution, the solution can be checked to see if is an actual solution in polynomial time (again the polynomial is a function of the length of the input). An example of this kind of problem is the zero sum subset problem - given a set of integers is there a subset of them that sum to zero. Clearly if a problem is in P it is also in NP. The major question is whether there is a problem that is in NP but is not in P.

- The Hodge Conjecture: This conjecture is even difficult to state since it involves some very technical definitions. A (very) vague statement is that certain algebraic invariants of algebraic/geometric objects called varieties come from geometric invariants.
- The Riemann Hypothesis: This was discussed above.
- Yang-Mills existence: This is discussed in Episode 103.
- Navier-Stokes existence and smoothness: The Navier-Stokes equations describe in full detail the motion of liquids. Even though they have been known for over a century, it is still not known whether there are smooth (i.e. differentiable infinitely many times) solutions to these equations.
- The Birch and Swinnerton-Dyer conjecture: This conjecture is similar to the Hodge conjecture in that it is difficult to state, but it roughly states that there is a relationship between the number of solutions to a particular type of equation and the order of the zero of the L-function at a particular point (L-functions are related to Riemann's zeta function.)
- The Poincare conjecture: This conjecture has actually been solved in 2006 and received fairly wide media coverage. The two-dimensional version of the the conjecture says that if you have a surface that is simply connected (that means that if you draw a loop on the surface, you can shrink it to a point by sliding the loop continuously along the surface) and closed (it does not go off to infinity, and cannot be stretched so that it does), then it can be deformed into a sphere. The three dimensional version of this took many years to prove.

Encryption

One of the main plot points of the episode is that the bad guys wanted the math professor to give them his solution to the Riemann hypothesis so that they could crack the encryption on major financial data. As was mentioned in the episode, many encryption

algorithms depend on the fact (or apparent fact) that it is very hard to factor large numbers. One of the main encryption algorithms used today is called the RSA algorithm (the name comes from the initials of its inventors. This algorithm is also described Episode 205. As mentioned there, the main difficulty in trying to break the RSA code is that factoring numbers into prime factors can be very difficult. In particular, there is no known algorithm that factors a number n into its prime factors that is in P, i.e. runs in polynomial time as a function of $\log(n)$ (here log is used because the number of bits it takes to store a number n is $\log_2 n$ — the base 2 log of n). It is believed that no such algorithm exists. Interestingly, there is another kind of cryptosystem that is provably difficult in the sense that if an algorithm is devised to decrypt messages in polynomial time, then this algorithm can be adapted to factor integers in polynomial time.

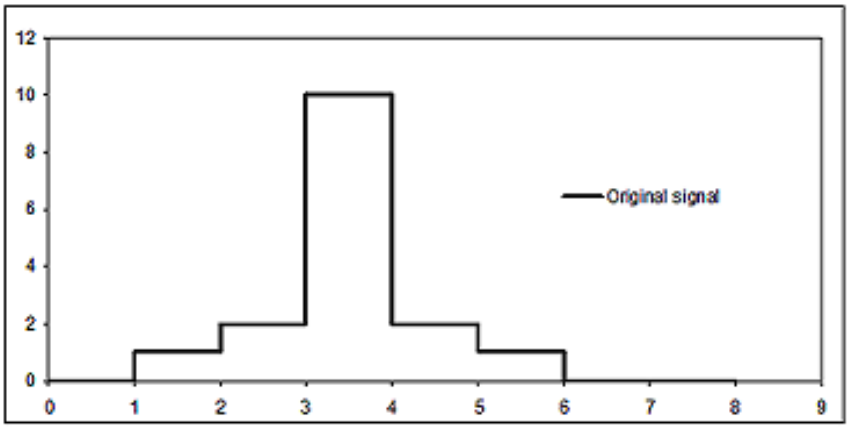
107: Counterfeit Reality

Counterfeiters are producing small denomination bills by using the talent of an artist who they have kidnapped. Don and Charlie are afraid that the artist will be murdered as a liability by the criminals. The only way to find out the identity of the hostage is by running a match analysis between artistic pieces of several missing artists and the work seen in the counterfeit money. In order to do this Charlie uses Wavelet Analysis over the artists' work. We will explain below what Wavelet analysis is and how it is applied to image processing.

What is Wavelet Analysis?

Wavelet analysis refers to the decomposition of finite energy signals into different frequency components by superposition of functions obtained after scaling and translating an initial function known as **Mother Wavelet**. Wavelet Analysis is different from other techniques in that it analyses frequency components with a resolution that matches their scale.

In order to clarify the statement above we present an example of how Wavelet Analysis is used to analyze a discrete signal. Assume that you have a discrete signal given by a vector with 8 components,

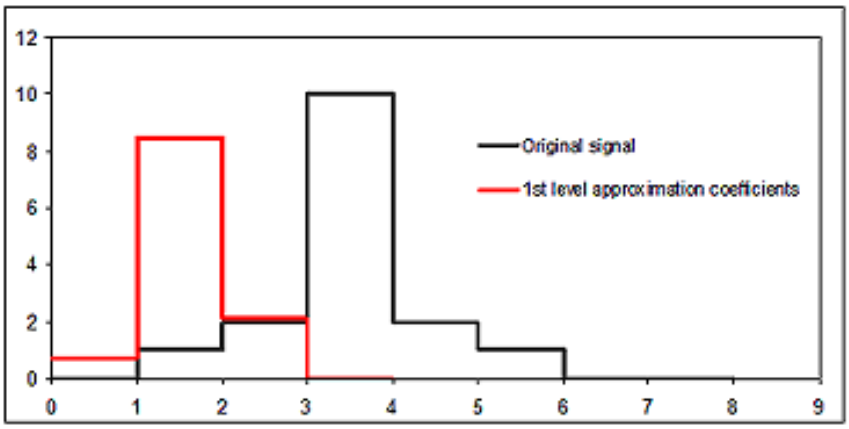


$$X = [0, 1, 2, 10, 2, 1, 0, 0]$$

We show in the figure on the right a graphic representation of these data. Let's call $X(i)$ the ith component of the vector X ($i = 0, 1, 2, \ldots, 7$). In the first step of the Wavelet decomposition of this signal we split our information into two vectors of 4 com-

ponents, say $a(1), d(1)$. The kth component of $a(1)$ is equal to $\frac{X(2k)+X(2k+1)}{\sqrt{2}}$ and the kth component of $d(1)$ is equal to $\frac{X(2k)-X(2k+1)}{\sqrt{2}}$ $(k = 0, 1, 2, 3)$. In our case, $a(1) = \left[\frac{1}{\sqrt{2}}, \frac{12}{\sqrt{2}}, \frac{3}{\sqrt{2}}, 0\right]$ and $d(1) = \left[\frac{-1}{\sqrt{2}}, \frac{-8}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right]$.

The vectors $a(1)$ and $d(1)$ are called the first level **approximation and detail coefficients** of the wavelet decomposition, respectively. On the right we show the graph of the first level approximation coefficients, which reveals that they represent the original signal in a different scale and explains why they are called *approximation* coefficients.



In the next step we follow the exact same procedure over the vector $a(1)$ splitting it into two vectors of 2 components, $a(2)$ and $d(2)$, corresponding to the sums and differences of consecutive terms divided by $\sqrt{2}$, respectively. Then, $a(2) = \left[\frac{13}{2}, \frac{3}{2}\right]$ and $d(2) = \left[\frac{-11}{2}, \frac{3}{2}\right]$.

$a(2)$ and $d(2)$ are called the second level approximation and detail coefficients of the decomposition, respectively. Finally for the third and last step, we follow the same procedure over the vector $a(2)$, and we obtain two one-dimensional vectors $a(3)$ and $d(3)$ (the third level approximation and detail coefficients, respectively) where, $a(3) = \frac{8}{\sqrt{2}}$ and $d(3) = \frac{5}{\sqrt{2}}$.

One of the nice features of the wavelet decomposition is that the reconstruction algorithm is very similar to the decomposition one. Let's assume that we have the approximation coefficient of the third level $a(3)$ and the decomposition coefficients corresponding to all the levels, $d(3), d(2), d(1)$. In order to get $a(2)$ from this information we take sums and

differences of the components of $a(3)$ and $d(3)$ divided by $\sqrt{2}$. We obtain then,

$$a(2) = \left[\frac{\frac{8}{\sqrt{2}} + \frac{5}{\sqrt{2}}}{\sqrt{2}} = \frac{13}{2}, \frac{\frac{8}{\sqrt{2}} - \frac{5}{\sqrt{2}}}{\sqrt{2}} = \frac{3}{2} \right].$$

We proceed analogously with $a(2)$ and $d(2)$ in order to obtain the first level approximation coefficients,

$$a(1) = \left[\frac{\frac{13}{2} - \frac{11}{2}}{\sqrt{2}} = \frac{1}{\sqrt{2}}, \frac{\frac{13}{2} + \frac{11}{2}}{\sqrt{2}} = \frac{12}{\sqrt{2}}, \frac{\frac{3}{2} + \frac{3}{2}}{\sqrt{2}} = \frac{3}{\sqrt{2}}, \frac{\frac{3}{2} - \frac{3}{2}}{\sqrt{2}} = 0 \right].$$

Finally, the same calculations over $a(1)$ and $d(1)$ reconstruct the original information,

$$a(1) = \left[\frac{\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}}}{\sqrt{2}} = 0, \frac{\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}}{\sqrt{2}} = 1, \frac{\frac{12}{\sqrt{2}} - \frac{8}{\sqrt{2}}}{\sqrt{2}} = 2, \frac{\frac{12}{\sqrt{2}} + \frac{8}{\sqrt{2}}}{\sqrt{2}} = 10, \frac{\frac{3}{\sqrt{2}} + \frac{1}{\sqrt{2}}}{\sqrt{2}} = 2, \frac{\frac{3}{\sqrt{2}} - \frac{1}{\sqrt{2}}}{\sqrt{2}} = 1, \frac{0+0}{\sqrt{2}} = 0, \frac{0-0}{\sqrt{2}} = 0 \right].$$

In order to understand the reasoning behind the calculations above it is convenient to see them from a "continuous" point of view. Let's call h the function that takes the value 1 on the interval $[0, 1)$ and 0 otherwise, and g the function that takes the value 1 on $[0, 0.5)$, -1 on $[0.5, 1)$, and 0 otherwise. We can represent the signal X as a combination of translations of the function h as follows,

$$0 \cdot h(t) + 1 \cdot h(t-1) + 2 \cdot h(t-2) + 10 \cdot h(t-3) + 2 \cdot h(t-4) + 1 \cdot h(t-5) + 0 \cdot h(t-6) + 0 \cdot h(t-7).$$

The key formulas behind the wavelet decomposition and reconstruction algorithms are the ones given by the representation of h and g as a combination of translations of versions of h and g in a different scale. More precisely,

$$\begin{aligned}
h(t) &= \frac{1}{\sqrt{2}} \left(\frac{1}{\sqrt{2}} h(2^{-1}t) \right) + \frac{1}{\sqrt{2}} \left(\frac{1}{\sqrt{2}} g(2^{-1}t) \right) \\
h(t-1) &= \frac{1}{\sqrt{2}} \left(\frac{1}{\sqrt{2}} h(2^{-1}t) \right) - \frac{1}{\sqrt{2}} \left(\frac{1}{\sqrt{2}} g(2^{-1}t) \right) \\
\frac{1}{\sqrt{2}} h(2^{-1}t) &= \frac{1}{\sqrt{2}} h(t) + \frac{1}{\sqrt{2}} h(t-1) \\
\frac{1}{\sqrt{2}} g(2^{-1}t) &= \frac{1}{\sqrt{2}} h(t) - \frac{1}{\sqrt{2}} h(t-1).
\end{aligned}$$

We adopt the following notation,

$$h_{i,k} = 2^{i/2} h(2^i t) \text{ and } g_{i,k} = 2^{i/2} g(2^i t) \text{ for } i, k \text{ integers}$$

Under this notation we observe, by using the formulas above, that the first level approximation and detail coefficients of the decomposition correspond to the coefficients in the decomposition of X in terms of the scaled functions $h_{-1,k}$ and $g_{-1,k}$ as shown below,

$$X = \frac{1}{\sqrt{2}}\, h_{-1,0} + \frac{12}{\sqrt{2}}\, h_{-1,1} + \frac{3}{\sqrt{2}}\, h_{-1,2} + 0\, h_{-1,3} - \frac{1}{\sqrt{2}}\, g_{-1,0} - \frac{8}{\sqrt{2}}\, g_{-1,1} + \frac{1}{\sqrt{2}}\, g_{-1,2} + 0\, g_{-1,3}.$$

An analogous argument shows that the second level approximation and detail coefficients correspond to the coefficients in the decomposition of the signal generated by the first level approximation coefficients, in terms of translations of scaled versions of h and g given by

$$h_{-2,k} \text{ and } g_{-2,k} \text{ for } k \text{ an integer.}$$

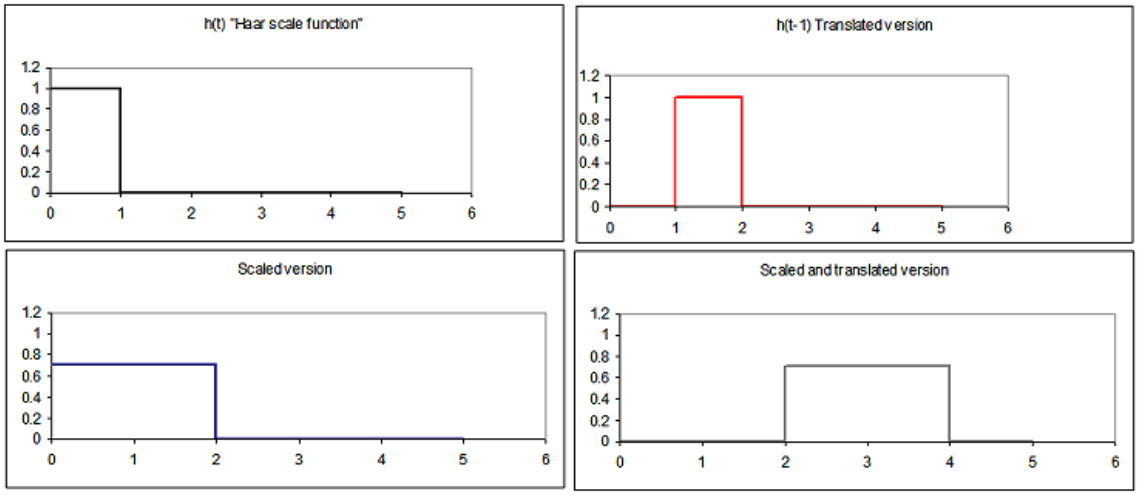
In the **signal processing** terminology, we say that the functions above for i big analyze the small scales that correspond to the high frequencies of the signal (sudden changes in a short period of time) and the scaled functions for i small analyze the big scale features of the signal that correspond to the low frequencies, which give the signal its overall shape. This frequency analysis is local since each approximation and detail coefficient is calculated by using only part of the data and then it is adequate to study **non-stationary** signals (signals that do not repeat into infinity with the same periodicity).

Activity 1

Calculate the approximation and detail coefficients of the wavelet decomposition of the following signal:

$$X = [1, 2, 10, 12, 5, 10, 11, 0]$$

Verify that the original signal can be recovered from your data by using the reconstruction algorithm described above.



We mentioned above that wavelet analysis analyses frequency components with a resolution that matches their scale. In our case, the low frequencies are analyzed with the scaled versions of h and g, and the main feature of these functions is that their support (i.e. the interval where they are not 0) get's larger as i gets smaller (see figures above). This adaptive feature is what differentiates wavelet analysis from other techniques like **Fourier Analysis** where the functions used to analyze different frequencies have fixed support. The scheme presented in this example corresponds to what is called the **Haar Discrete Wavelet Transform**, which is closely related to the **Haar Multiresolution Analysis**; and the functions h and g are called the **Haar Scale and Mother Wavelet functions**, respectively. We refer the reader to "An introduction to wavelets" by Amara Graps and "A wavelet tour of signal processing" by Stephane Mallat to find out more about Fourier Analysis, Multiresolution Analysis and Discrete Wavelet Transforms. We also recommend the introductory exposition of Yves Nievergelt in the book titled "Wavelets Made Easy." Now we present an example that shows how wavelet analysis is used in data compression and explain how wavelet analysis is used to image processing.

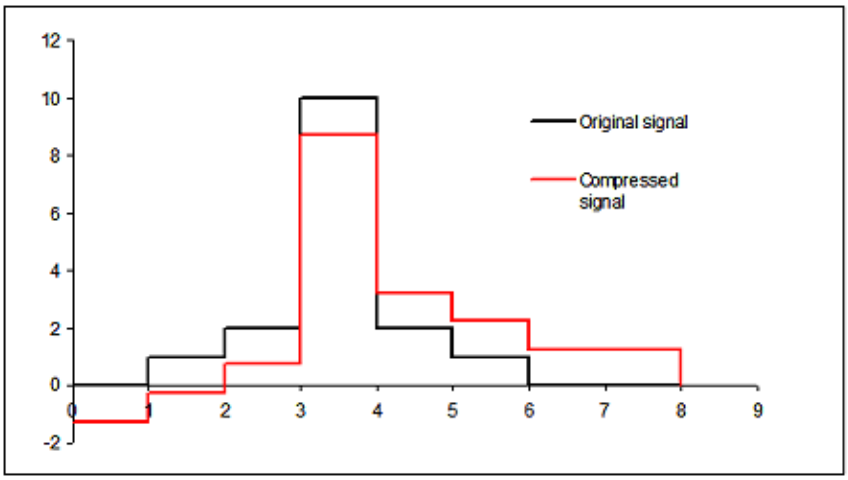
Tangent

The Haar Scale function and Wavelet function are named after Alfred Haar, who proposed them as generators of bases for function spaces in 1902. The main disadvantage of these functions is that they are discontinuous and therefore not appropriate for the analysis of smooth signals. Only after 85 years, in 1987, **Ingrid Daubechies** discovered the first continuous wavelet functions with compact support and with her discovery Wavelet Analysis revolutionized the signal processing world. The Wavelet Transform constitutes a new method to decompose signals. This method represents a more adequate approach to lessen the restriction given by **Heisenberg's uncertainty principle**, which states that it is impossible to know the exact frequency and the exact time of occurrence of this frequency in a signal. This principle is explained in Episode 104 as well.

Data Compression and Image Processing

Assume that after the wavelet decomposition described above we only keep the third level approximation coefficient and the first and second level detail coefficients, or equivalently we drop the value of $d(3)$. This represents a compression of data because we have to store only seven numbers instead of the initial eight. By assuming that $d(3) = 0$ and following the reconstruction algorithm as described above we obtain the new approximation coefficients and compressed signal given by

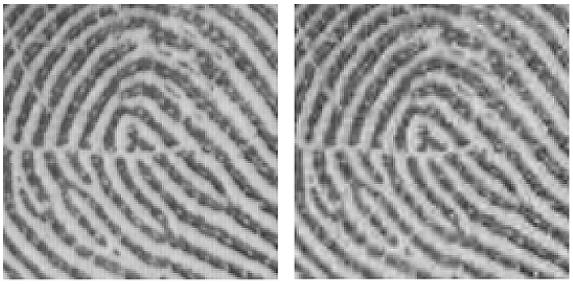
$$\begin{aligned}
\tilde{a}(2) &= [4, 4] \\
\tilde{a}(1) &= \left[\frac{-3}{2\sqrt{2}}, \frac{19}{2\sqrt{2}}, \frac{11}{2\sqrt{2}}, \frac{5}{2\sqrt{2}}\right] \\
\tilde{X} &= \left[\frac{-5}{4}, \frac{-1}{4}, \frac{3}{4}, \frac{35}{4}, \frac{13}{4}, \frac{9}{4}, \frac{5}{4}, \frac{5}{4}\right]
\end{aligned}$$



We show below a graph that reveals some similarities between the original data and the compressed one. In general, data compression using wavelets consists in the storage of a restricted amount of detail coefficients of the wavelet decomposition. These coefficients, as their name suggest, contain the information about the details of the signal but are not important in order to recognize its "big picture" features.

Image Processing

We can think of an image as a rectangular array of numbers, where each number represents the intensity at the corresponding pixel. The easiest way of generalizing the procedure explained before to the two dimensional case is by running the wavelet decomposition first over the rows of the image and then over its columns. Image compression again corresponds to the storage of the approximation coefficients and some of the detail coefficients. This procedure has been found to be very useful and has numerous applications. Perhaps the best-known application of wavelet analysis is used by the FBI, which, since 1993, uses a wavelet transform to compress digitalized fingerprint records (See the figure below).



Original Image Compressed Image (Ratio of comp.=12.9)

Image compression is not the only application of Wavelet Analysis. Since the analysis of the frequencies is done in a local manner there exist many applications of the wavelet transform to other image processing problems such as image restoration and edge detection. In edge detection edges of an image correspond to the occurrence of high frequencies in small portions of the signal, or equivalently to big detail coefficients in small scales. Matlab's wavelet toolbox contains a comprehensive collection of routines to analyze data. All these nice features of the Wavelet transform helped Charlie to recognize similarities, in edges and scales of grey among others, between the artist's work and the one seen in the counterfeit money.

Activity 2

Find the compressed data obtained after dropping the third level detail coefficient of the wavelet decomposition of the signal given in Activity I, and draw a graph of your results.

108: Identity Crisis

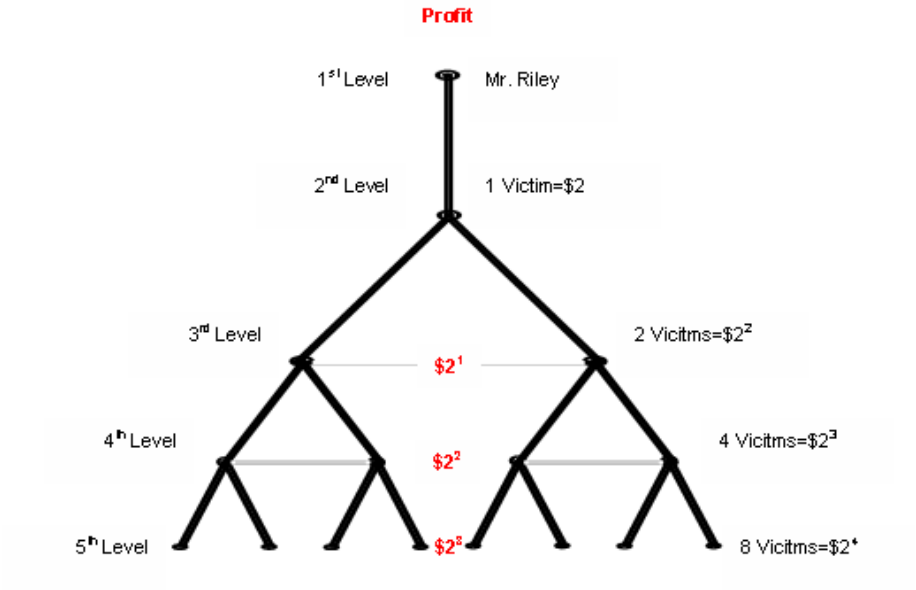
A man accused of stock fraud is found garroted in his apartment and the crime is very similar to one committed a year ago. In this episode Charlie helps his brother to investigate the relation between the two murders in order to find out who is the perpetrator and to determine the innocence of the man arrested one year ago.

Below we will explain the illegal scheme used by the victim to gain money which is related to the concept of Geometric Progression or Geometric sequence.

What is a Pyramid Scheme?

A pyramid scheme is a fraudulent non-sustainable business model that — as its name suggests — takes the form of a pyramid. The scheme could be more complicated than the one presented here; however, the idea behind it is always the same: the people involved in the scheme must recruit others in order to recover their initial investment plus some profit, but the recruitment process is done only with the promise of earning money after recruiting others with no service or product in exchange.

Example 0.1. There is one person on top of the pyramid. This person asks for a fixed amount of money, say \$1, to a second person with the promise that if he convinces two people to join and pay the entrance fee of \$1, he will get \$2 which would double up his/her initial investment. At this point the first two people have made \$1 dollar each without selling any product or giving any service. The other two people have to recruit two people each in order to double up their money. At this point, we are at the fourth level of the pyramid; where already $1+1+2+4=8$ people are involved in the scheme. As the pyramid gets bigger the amount of people increases really fast, at level n there will be $1+1+2+\cdots+2^{n-2}$ persons in the scheme and the bottom 2^{n-2} have to recruit 2^{n-1} new people in order to receive money. This is what makes the scheme fraudulent: since the amount of people involved increases **geometrically**, after a few levels the system will collapse because there are no more people to recruit and then the people at the bottom of the pyramid will loose their initial "investment."



Example 0.2. The scheme presented above is very similar to the one Charlie explained to Don and Terry in the episode. In this case Mr. Riley started by taking \$2 out from one account, then he took \$2 out of other two accounts, he kept \$2 to himself and used the other \$2 to replace the money he had already taken, this corresponds to the third level of the pyramid. Then he took \$2 out of 4 accounts, he used \$4 to replace the money he had taken and kept the other \$4. He kept doing this so that at level n of the pyramid he made 2^{n-2} dollars. The system crashed when he reached the 21st level of the pyramid, and at this point he obtained a profit of \$524.288 ($= \$2^{19}$) as mentioned in the show.

What is a Geometric Progression?

A geometric progression is a sequence of numbers with the property that the quotient between any of the numbers and the previous one is a constant. This quotient is called the **ratio** of the progression. In mathematical notation a geometric progression or geometric sequence is a sequence of the form

$$(a_n)_{n\geq 1} \text{ where, } a_n = ar^n \text{ for all } n.$$

a is an arbitrary number and r corresponds to the ratio of the progression. If the ratio of the progression is a number strictly greater than 1 as time evolves, i.e. as n gets bigger, the terms of the progression become big really fast (this is the case in our examples above where $r = 2$). If the ratio is equal to 1 the progression is constant, i.e. all its terms are actually the same. Finally, if the ratio of the progression is a positive number strictly less than 1 as time evolves the terms get smaller approaching 0.

A concept that is closely related to the concept of Geometric Progression is the concept of **Geometric Series**. A geometric series corresponds to the sum (possibly infinite) of terms of a Geometric Progression. For instance, if you add the first n terms of the geometric progression shown above then you will obtain

$$ar + ar^2 + \cdots + ar^n = a\left(\frac{1 - r^{n+1}}{1 - r} - 1\right).$$

This equation is obtained from the following algebraic equation

$$(1 + r + \cdots + r^n)(1 - r) = 1 - r^{n+1}.$$

And if you add up all the terms of a geometric progression with ratio r, between 0 and 1, then you will obtain,

$$\sum_{i=1}^{\infty} a_n = ar + ar^2 + \cdots + ar^n + \cdots = a\left(\frac{1}{1 - r} - 1\right).$$

This follows as a consequence of the formula above for n terms, and the fact that as n gets bigger then r^{n+1} approaches 0.

Activity 1

(a) By using the formulas presented above calculate how many people are involved in the scheme of Example 1 above after 12 levels.

(b) What is the total profit of Mr. Riley after the 21 levels of the pyramid?

Tangent

A different but similar kind of scheme is what is known as **Ponzi Scheme**. It is named after Charles Ponzi who after emigrating from Italy to the United States in 1903 earned a lot of money by implementing a short-term high return investment scheme, using the currency difference between the United States and foreign countries to buy and sell international mail coupons. In this scheme the actors interact with all the other people involved in the scheme which gives the scheme a non-pyramidal structure. Also the amount of people does not increase geometrically because reinvestment is allowed. This gives the scheme a longer life. However the scheme is fraudulent as well because it relies on the same principle as pyramid schemes: new money from the investors is used to pay off earlier investors until the whole scheme collapses.

Example 0.3 (Doubling Strategy)**.** Geometric Progressions appear in numerous areas of study. Here we provide an example of their appearance in elementary financial theory

through the key concept of **arbitrage**. We say that there exists arbitrage in a certain financial transaction if at least one of the parties involved obtains a risk-less profit after the transaction. For instance, consider the following coin toss game: an *infinitely* rich gambler bets \$1 on tails. If the result after tossing the coin is tails he will double up his money obtaining \$2 and the game ends. If the result is heads he will lose his dollar but he has the chance to bet back again \$2. If in the second round the result is tails, then he will double up the \$2 getting \$4, obtaining a net profit of \$1 and the game will end, otherwise he will lose the \$2 cumulating a loss of $\$1 + \$2 = \$3$. However, since his wealth is infinite, in the later case he can bet \$4 in the third round. If he wins he will get \$8 and his net profit will be \$1, otherwise he will cumulate a total loss of $\$1 + \$2 + \$4 = \$7$. The game keeps going in this direction indefinitely. The probability of the gambler winning \$1 by the end of the game is 1, and hence arbitrage exists. Notice that the amount of money the gambler has to bet in each round follows a geometric progression of ratio 2 and this is why he is assumed to have an unlimited liquidity. In order to avoid arbitrage, finance theory assumes limited liquidity.

Activity 2

Suppose the game described above has not finished after 10 rounds. How much money has the gambler lost up to this point? How much money does he have to bet in the next round?

Fingerprint Analysis

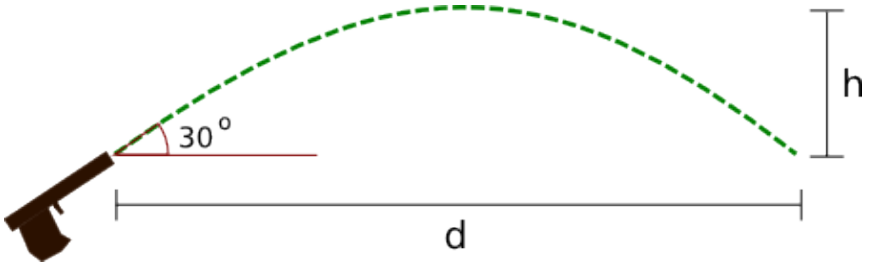
In this episode, Charlie also questions the terminology experts use in fingerprint recognition analysis. He criticizes the assumption often made in this area which states that there are no two people with the same fingerprint structure. However, Charlie explains that this may not be necessarily true because there is no way of knowing this in a deterministic way. He proposes that using probabilistic statements would be more appropriate - as is already the case in DNA match studies. Charlie's critic is founded on the following argument: world's population as of today is approximately 6.5 billion and each person's fingerprint structure contains a very high amount of information. This makes it almost impossible to file all this information in the data bases and demands the use of compressing methodologies in order to store such a high amount of information. In the compression of the fingerprint images only the main characteristics are taken into account and some of the details are not taken into consideration, which makes fingerprint recognition a probabilistic problem rather than a deterministic one. One of the algorithms the FBI has implemented for Fingerprint Compression uses Wavelet Analysis, a technique mentioned in Episode 107 as well.

109: Sniper Zero

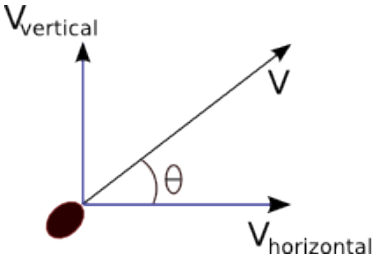
In this episode the FBI investigates a bizarre string of sniper attacks which seem to have little in common. To determine the location of the sniper in each shooting, Charlie uses ballistic trajectory modeling. Exponential growth and regression to the mean are also briefly mentioned, and the first of these we explore in depth below.

Ballistic Trajectory

A bullet, like any other object flying through the air, is subject to the forces of gravity, air resistance, and wind. One way to closely approximate the actual trajectory is to ignore the effects of drag and wind, instead looking only at gravity.



Consider the figure above. A bullet leaves the barrel of a gun inclined at a 30° angle and flies a horizontal distance of d before reaching the starting elevation. The force of gravity acts on the bullet, creating a downward acceleration of $g = 9.8$ m/sec^2 and so influencing the vertical component of the velocity vector (see diagram below) over time. Since we disregard drag, the horizontal component of the velocity does not change.



$|V|$=magnitude of vector V

$V_{vertical} = (0, |V|\sin\theta)$

$V_{horizontal} = (|V|\cos\theta, 0)$

$V = V_{vertical} + V_{horizontal}$

The next activity involves figuring out the equations describing the speed and position of an object in free-fall. These derivations make some use of calculus. Try to follow them and do the exercises, but if you can't, just use the equations mentioned in order to do Activity 2.

Tangent

Technically, the term *velocity* means the vector pointing in the direction of motion with magnitude equal to the *speed* of the object. However, in everyday usage and even in many physics textbooks the term velocity is used to denote both the vector and its magnitude, the speed. It is usually easy to figure out which is being meant from the context: just ask yourself, is the sentence talking about a vector or a scalar?

Activity 1

Let us first consider only the vertical direction of motion. For the sake of brevity, I'll just write $v(t)$ below instead of $v_{\text{vertical}}(t)$.

1. Recall that the acceleration of an object is equal to the instantaneous change in velocity, i.e. $a(t) = v'(t)$. Apply the fundamental theorem of calculus to this equality to deduce that $v(t) - v(0) = gt$, where g is the acceleration due to gravity.

2. We can do even better by applying the same trick to velocity. Namely, we know that $v(t) = y'(t)$, where $y(t)$ is the vertical position of the object at time t. Apply the fundamental theorem of calculus again to show that
$$y(t) = y(0) + v(0)t + \frac{gt^2}{2}$$

3. Now write down an equation for the horizontal position $x(t)$ of the bullet in terms of the initial horizontal velocity $v_{\text{horizontal}}(0)$. (Hint: remember, we disregard drag and wind so the only acting force is gravity.)

Activity 2

Suppose the bullet is fired at an angle of 30° as in the first picture, with a speed of 900m/s from an initial point $x(0) = 0$ and $y(0) = 0$ (i.e. from the origin) at time $t = 0$. Use the equations from activity 1 to answer the following questions.

1. What is the maximum height achieved by the bullet? At what time is this height achieved? [Hint: what is the vertical velocity of the bullet when it's at a peak height?]

2. What is the horizontal distance of the bullet from the origin at the time of peak height? What is the distance to the point at which the bullet is again at the height from which it was fired; that is, $y = 0$?

3. Show that the trajectory of the bullet is a parabola.

Analyzing the general situation, in which both wind and drag affect the path of a bullet, is in fact very complicated. You can get a taste of the difficulties involved by reading the wikipedia article on external ballistics. Furthermore, mathematically recreating the path of a bullet after it has hit a target, thus only knowing its angle of entry, is much harder.

Exponential Growth

Here are a few recent uses of the term exponential growth in the news media:

The company has had a spectacular two years, riding the exponential growth in oil prices that helped to increase profits by a fifth in 2006 to £28.5 million. (*Business Big Shot: Alasdair Locke*, The Times, Dec 20, 2007)

After years of exponential growth, there has recently been a slow down in the Northern Ireland property market. (*Well-known property firms merge*, BBC News, Dec 7, 2007)

Kessler himself came under university scrutiny for alleged financial irregularities. In January 2005, an anonymous source contended he "spent or formally committed all of the reserves of the dean's office and has also incurred substantial long-term debt in the form of lavish salary increases and exponential growth in new, highly compensated faculty and staff directly reporting to him." (*UCSF dean is fired, cites whistle-blowing*, Los Angeles Times, Dec 15, 2007)

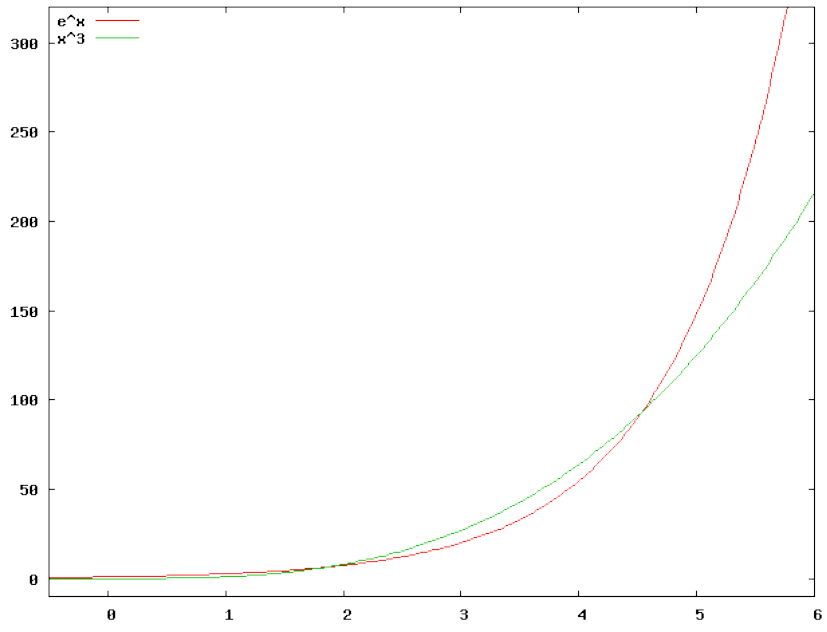
While the above excerpts describe growth in entirely different areas, the one thing they have in common is the use of the term *exponential growth*. In mathematics, we say that quantity x grows exponentially with respect to time t if x satisfies the following differential equation: $\frac{dx}{dt} = kx$, where k is a constant and $\frac{dx}{dt}$ is either a derivative, when t is continuous, or the change in x in a given time interval, when t is discrete. In plain words, this means that x grows exponentially if it increases proportionally to its own value. Most often exponential growth occurs in situations where "x creates more x", typical examples being population growth and compound interest. Exponential growth can occur both when the time intervals are discrete, for example in annual or monthly interest compounding, and when the time variable is continuous, as in continuous compounding of interest or modeling of large populations. The discrete time case is more often encountered in practice and is easier to analyze mathematically, since we don't need to resort to the exponential function.

Activity 3

1. Suppose yesterday you heard that annual inflation was 3% in the last year. If x is the price of a representative basket of goods, and t is measured in years, what is the corresponding proportionality constant k in the exponential growth equation that models the price increase? (Hint: note that in this case $dt = 1$ year.)

2. What if t is measured in days instead?

The reason why such growth is called exponential is that when the time variable t is continuous, we can solve the differential equation $\frac{dx}{dt} = kx$. By separating variables we get $\frac{dx}{x} = k\,dt$, integrating we arrive at $\ln(x) = kt + C$, where C is some constant, and exponentiating both sides, we finally get $x = D\,e^{kt}$, where D is a constant. We can solve for D by plugging in $t = 0$, the starting time, to arrive at the general solution $x(t) = x(0)\,e^{kt}$. Exponential growth is much faster than polynomial, as the example below illustrates in case of e^t versus t^3.



Activity 4

1. Find a constant r so that $2^t = e^{rt}$.
2. Show that 2^t becomes larger than any polynomial in t, for sufficiently large t. (Hint: suppose $p(t) = t^n$ for some positive integer n. For which t is $2t > p(t)$?)
3. Can you think of a function $f(t)$ which grows faster than an exponential function, in the sense of part 2 above?

In practice, when talking about compound interest two quantities are important. One is the annual interest rate, sometimes called the annual percentage rate (APR). The other is the number of compounding periods per year: how many times per year is the interest added to the principal amount. For instance, say you have \$100 credit card debt with an APR of 20%. Usually credit cards compound monthly, so there are 12 compounding periods per year. Thus if you make no payments (and incur no additional penalties or expenses) for a whole year, your debt will *not* simply be $100 + 100 \cdot 0.2 = 120$, which it would if the interest was compounded *only once per year*. Instead, after the first month, you'll owe $100 + 100 \cdot \left(\frac{0.2}{12}\right) = 101.67$ dollars. After the second month, you'll owe $101.67 + 101.67 \cdot \left(\frac{0.2}{12}\right) = 103.36$ dollars, and so on. At the end of the year, with such monthly compounding, you'll owe \$121.94. Might not seem like a huge difference from the once a year compounding sum of \$120, but over longer periods of time, the difference becomes substantial.

Activity 5

1. You open a savings account which earns 2% interest with a deposit of \$1000. Would you rather the interest compound daily or monthly? Write down the formula for the amount of money in the account after a year in both cases. (Hint: write down the expression for the amount of money after one period of compounding, now after two periods (don't simplify!), then three... See the pattern?)
2. Suppose we decide to compound not once a month or a day, but once every split second. In fact, we can let the number of compounding periods go to infinity, thus letting the length of each period approach zero. Use the fact that $e^y = \lim_{n\to\infty}\left(1 + \frac{y}{n}\right)^n$, for any real number y, to show that when the number of compounding intervals goes to infinity, then after t years, your account will have $1000e^{0.02t}$ dollars. This is continuous compounding.

Activity 6

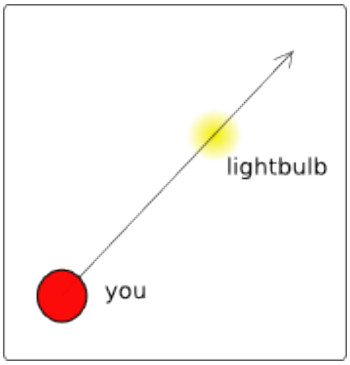
In popular usage, the expression “exponential growth" is often used as a synonym for “very fast growth". There’s no good reason to describe faculty hiring practices, as the third quote in the beginning of this section does, in terms of exponential growth. While an exceptional number of faculty might have been added during Kessler’s tenure as dean, there’s no sense in which “faculty makes more faculty" proportionally to existing numbers. At other times, “exponential growth" can be more accurately described as sigmoidal (remember that strange function used in logistical regression?). While similar in the low range to the exponential function, sigmoidal growth reflects the fact that at some point growth must slow down due to lack of resources.

110: Dirty Bomb

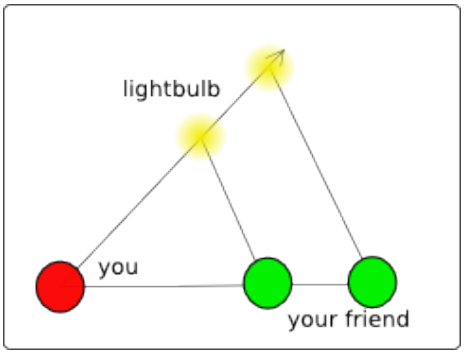
In this episode there was a discussion about triangulation and there was a scene where Charlie explains the prisoner's dilemma to three prisoners in the hope that it will make one of them confess.

Triangulation

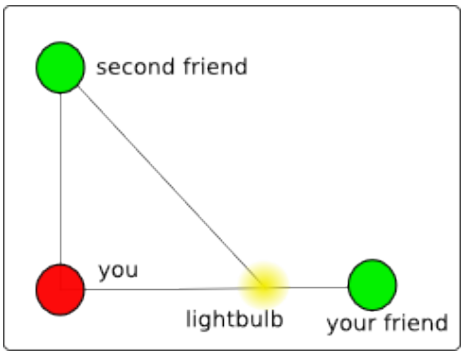
In this episode a truckload of radioactive waste has been hijacked and Charlie uses triangulation of the radiation the waste emits to find where it is. This is mathematically similar to trying to find a lightbulb in a very large field (without moving). If you are standing in a field, then you will be able to see the lightbulb but you won't be able to tell how far away it is. This means you know that it lies somewhere on a particular line that goes through you, which probably wouldn't be particularly useful, since to find the lightbulb without gathering more information you would have to walk along the entire line to get to the lightbulb.



However, let's say you have a friend in the field and both you and your friend have lightbulbs and radios. Then you can report to each other the angle between the friend's lightbulb and the other lightbulb. Is this enough to find the other lightbulb? Not quite, because there is no angle-angle theorem in geometry. In other words, if you adapt coordinates so that you are at the origin, then if you double the distances of the other lightbulb and your friend from the origin, both the angles that you and your friend measure will be the same. The only way to fix this is to measure the distance between you and your friend.

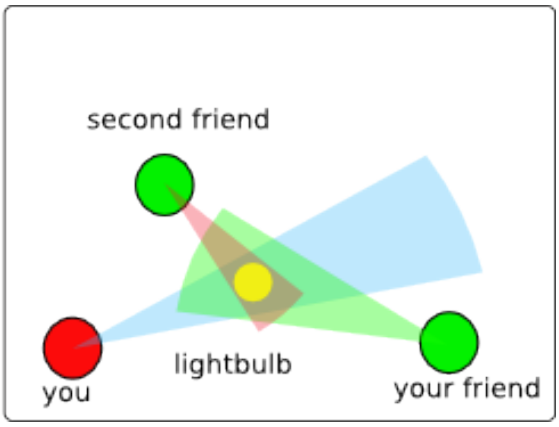


Of course, if you are unlucky there will be a problem with this. There is a chance that the lightbulb will lie on the line between you and your friend. If this is the case, it would be impossible for you and your friend to tell where on the line the lightbulb is. This can be fixed, however, by adding a third friend. If the three friends make sure they aren't all standing on a single line, then they can always find the lightbulb.



Activity 1

Let's say you are standing on the coordinates $(0,0)$ (the origin), your first friend is standing at $(1,0)$, and your second friend is standing at $(0,1)$. Also, assume the lightbulb is in the first quadrant so that both its coordinates are positive.



1. If the angle you see between your first friend and the lightbulb is 45° and the angle he sees between you and the lightbulb is 90°, where is the lightbulb?
2. If both your friends report an angle of 90° between you and the lightbulb, where is the lightbulb?
3. What if both your friends report an angle of 0° between you and the lightbulb?

Of course, in the real world you can't measure exactly what the angle is between your friend and the lightbulb. Any measurement has errors, so all that you will actually be able to say is that the lightbulb is very likely to lie inside a cone whose point is your location. The more accurate your measurement is, the skinnier the cone will be.

Activity 2

In the above picture with errors in measurements, which friend would be better to ask about his estimate of the angle between you and the lightbulb? Why? [Assume each friend has the same error in measuring angles.]

Tangent

The GPS, or Global Positioning System, uses a similar method to determine the exact location of a small handheld unit. Instead of having friends in a field the system has satellites in the sky that have very precise clocks in them. The handheld unit receives signals from the satellites saying what time they think it is and their current location, which allows the handheld unit to calculate the distance from it to each of the satellites. Then the handheld unit is able to mathematically determine its location. Interestingly, the clocks are so accurate that they are able to detect effects from general relativity, and correcting for these effects leads to improvements in positional measurement of tens of meters.

The Prisoner's Dilemma

In this episode, Charlie refers to the Prisoner's Dilemma, which is a particular game that shows that if the players in a game are only looking out for themselves and not colluding with the other players, then the results of the game might be worse for everyone involved. Here's how it works. Suppose that there are two prisoners that were involved in the same crime, and for the police to successfully convict either one of them, they need the testimony of the other one. If neither prisoner testifies, then they will each serve 2 months on minor charges. If one testifies and the other doesn't, then the testifier will go free and the other will serve 12 years. However, if both testify, then each will get 8 years. Now it's obvious that if the prisoners can collude and trust each other, neither will testify. This result would be the best possible outcome in the sense that the total time served would be minimized. However, if each prisoner only acts in his own interest, both of them will testify. This is because if one prisoner testifies, the outcome for him is better than if he didn't testify no matter what the other person does. This situation can be conveniently described in a table where the first number listed is the incarceration time (called a payoff in game theory) of the first prisoner and the second number is the incarceration time of the second prisoner.

	Prisoner 2 Testifies	Prisoner 2 is Silent
Prisoner 1 Testifies	8, 8	0, 12
Prisoner 1 is Silent	12, 0	2 months, 2 months

Activity 3

1. If we have the following table, what are some conditions on the numbers a, b, c, d so that the argument given above still works and the equilibrium solution is the lower right hand square?

	Prisoner 2 Testifies	Prisoner 2 is Silent
Prisoner 1 Testifies	a, a	b, c
Prisoner 1 is Silent	c, b	d, d

2. Use reasoning similar to the argument above to figure out the equilibrium choices for the following game. (The left number is the payoff for player 1, and for this game bigger numbers are better.)

	Player 2, Choice A	Player 2, Choice B	Player 2, Choice C
Player 1, Choice A	4, 9	6, 4	1, 9
Player 1, Choice B	5, 3	9, 5	5, 2
Player 1, Choice C	1, 7	15, 12	10, 8

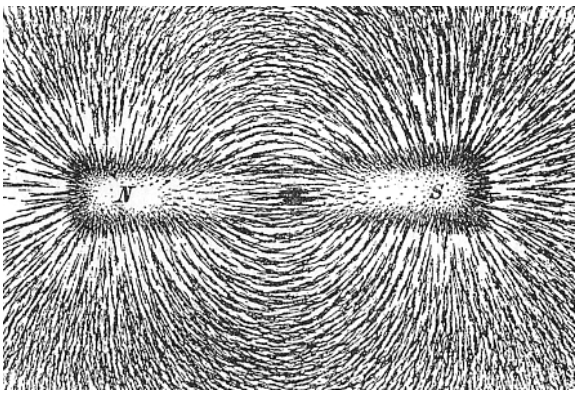
3. Will similar reasoning work for the following game? Why or why not? Is there an equilibrium solution? (An equilibrium solution is a choice for player A and B such that given knowledge of player B's choice, player A wouldn't want to change his choice, and vice versa.)

	Prisoner 2 Testifies	Prisoner 2 is Silent
Prisoner 1 Testifies	$2, -2$	$-2, 2$
Prisoner 1 is Silent	$-2, 2$	$2, -2$

111: Sacrifice

In this episode the data from a murdered researcher's computer is recovered after being "wiped." After a review of the basics of magnetism and binary encoding, we will explore how the two are combined to store data on hard disk drives and how this data might be recovered after being overwritten with other data.

Magnetism

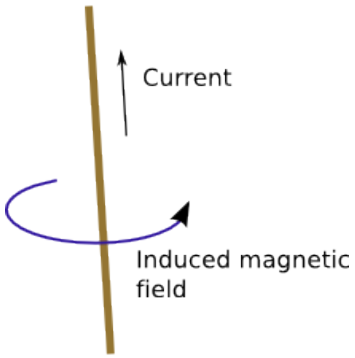


You have probably played with bar magnets, like the one on the right, before, discovering that they have two distinct poles and that like poles of two bar magnets repel while opposite attract. What you might have not realized is that magnetism originates at the subatomic level. Electrons (and other subatomic particles) have an intrinsic property misleadingly called spin; it does not refer to the particles actually spinning in the everyday sense of the word–it's more accurate to think of spin as a property of a particle akin to mass or charge. Nevertheless, as spinning a larger charged object creates a magnetic dipole (i.e. makes the spinning object produce a magnetic field), so does the spin of an electron make it behave like a tiny bar magnet. Additionally, when the electron also orbits an atom, this gives it another distinct magnetic dipole created by this orbiting, usually called the orbital dipole moment. The magnitude and axis of both of these magnetic moments can be expressed as vectors; adding those up we get the total magnetic dipole contributed by one electron. If we sum up these dipoles for each electron in each atom of a small object, and they do not cancel each other out, then that object, like the bar magnet above, produces a magnetic field and is known as a magnet.

Activity 1

1. Use wikipedia or other internet resources to find out about the differences between diamagnetic, paramagnetic, and ferromagnetic elements.
2. What would happen if you take a bar magnet and break it in half near the midway point dividing the north and south poles? Can you use what you've read about ferromagnetism and the discussion above to explain why this happens?

In addition to electrons producing magnetic fields by their spin and orbit around an atomic nucleus, moving electrons also create a magnetic field "around" the direction of motion, a phenomenon called electromagnetic induction. For instance, if a current flows through a straight wire, as on the right, a magnetic field is induced according to the right hand rule. Such a magnetic field can be used to realign the polarity of ferromagnetic materials, allowing the possibility of storing information in magnetic alignment, which is the basis of digital storage media like hard drives and floppy disks.



Binary Encoding

Throughout history humans used symbols, whether pictures, pictograms, or alphabets, to communicate, replicate, and store information. What differentiates older ways of storing information in books or manuscripts from what happens when you select "Save" in your word processor is the form of the encoding. By encoding I mean the representation of information in a way different from its original form. These sentences, for instance, encode information that originally exists as ideas in my mind. Encryption is another example of encoding, as is compression.

Activity 2

1. Julius Caesar used a simple shift encoding to communicate with his generals. In this encoding a letter of the Roman alphabet was replaced by one appearing n after it, for some positive integer n, wrapping at the end of the alphabet to the beginning. Although Caesar originally used $n = 3$, the most famous version of this scheme is rot13, in which $n = 13$, so that A becomes N, B becomes O, etc. The next part of this activity is encoded in rot13. Decode and do it.

2. N fgnaqneq jnl gb rapbqr gur Ratyvfu nycunorg hfrq ol pbzchgref vf NFPVV, Nzrevpna Fgnaqneq Pbqr sbe Vasbezngvba Vagrepunatr. Vg rapbqrf rnpu yrggre nf n guerr-qvtvg ahzore. Ybbx hc na NFPVV gnoyr bayvar, naq qrpbqr gur jbeq va cneg 3 bs guvf npgvivgl.

3. 067 111 110 103 114 097 116 117 108 097 116 105 111 110 115 !

Once we have encoded our alphabet as ASCII numbers, the next step is to encode these numbers using a series of blocks each with only two states, on or off, as in case of a ferromagnetic material which can be aligned along only one axis, with the north pole pointing either up or down. This is accomplished by converting ASCII codes into binary.

Note that $2008 = 8 \cdot 10^0 + 0 \cdot 10^1 + 0 \cdot 10^2 + 2 \cdot 10^3$. By writing "2008" in everyday use we mean for the digits to simply enumerate how many 1s, 10s, 100s, and 1000s are in the number. But if we look at it this way, a natural question arises: why use the powers of 10, instead of say, powers of 3 or 7? Indeed, there's absolutely no objective reason to chose 10 — resulting in what is called the decimal or base 10 numerical system —over other possible bases. The Sumerians, for example, used base 60, hence the 60 seconds per minute, 60 minutes per hour. The following activity should give you a feel for how different bases work and how to convert from one to another.

Activity 3

Let m_n denote the number m in base n. For instance, $5_{10} = 101_2 = 1 \cdot 2^0 + 0 \cdot 2^1 + 1 \cdot 2^2$.

1. Here's an algorithm for converting from base 10 to base 4: take your number, say 103_{10}, divide by 4, recording and discarding the remainder, in this case 3 ($103 = 4 \cdot 25 + 3$). Divide the result by 4 again, recording and discarding remainder, rinse and repeat until you get 0. ($25 = 4 \cdot 6 + 1$, $6 = 4 \cdot 1 + 2$, $1 = 4 \cdot 0 + 1$). Concatenate the remainders in reverse order to get the number base 4, $1213_4 = 103_{10}$. Why does this algorithm work?

2. Design an algorithm to convert from base 4 to base 16.

Activity 4

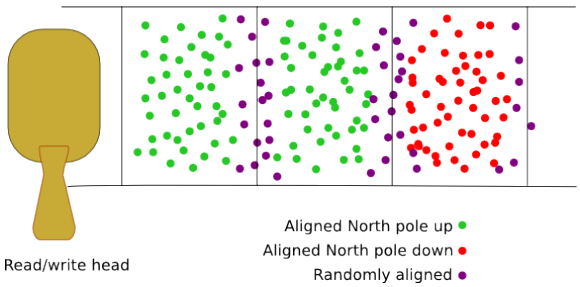
Besides binary (base 2), another common base used by computer scientists and programmers is hexadecimal, or base 16. Why do you think this is? [Hint: See part 2 of the previous activity.]

Hard Drives and Data Recovery

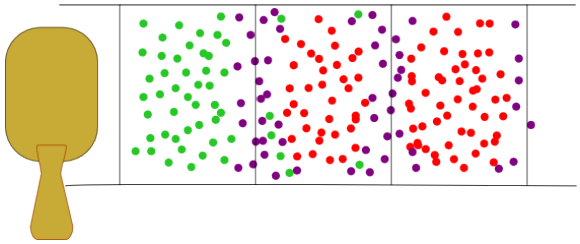
Your computer hard drive consists of layered, circular platters of ferromagnetic material that quickly spin while read/write heads hover above. Those read/write heads are able to read and change the magnetic polarity of tiny areas on a platter's surface. The more such areas are squeezed onto the platters, the larger the storage capacity of the hard drive. The reading and writing operations are accomplished using the principle of electromagnetic induction discussed previously: a current passing through the tip of the head is used to change the polarity of the area below, and the reverse effect–the appearance of a current in response to a changing magnetic field–is used to measure the existing magnetic alignment.

Normally, when you delete something in Windows or on a Mac (and in most modern Linux systems), the files are not deleted but their records are moved from the folder they originally resided into a designated "Trash/Recycle Bin" area. During this operation the actual data on the disk is left untouched, and the files are easily recoverable from trash. If you empty the trash bin, the records of the files get destroyed: the filesystem designates the physical space in which the file resides as empty. However, the actual file, the bits encoded in the magnetic alignment, do not get realigned in any way. Thus the information is actually kept completely intact until it is overwritten, there's just no way to readily access or restore it. However, restoring such non-overwritten files is not particularly difficult; there exist both commercial and free software tools to do exactly that. Things get much more interesting, and contentious, when it comes to restoring information that has been overwritten one or more times.

The most important paper on this subject is Peter Gutmann's Secure Deletion of Data from Magnetic and Solid-State Memory. It describes the following method of recovering overwritten data. Suppose we zoom onto a hard drive's surface and discover that the initial alignment within each area are as illustrated below.



Suppose we now overwrite the middle area: change the alignment to point down.



Due to microscopic imperfections in the read/write head and tiny imprecisions in its movements, there will be a small band of the previous up aligned particles left, separated from the new down aligned particles by a thin buffer band of randomly aligned particles. Overwriting the same area again will create additional microscopic bands of molecules aligned according to the previous magnetic alignment. These bands will trace out the history of previous magnetic alignment of this area of the hard disk, and when examined with a powerful electron microscope can be read. Naturally, the more times an area is overwritten, the "fainter" the outer bands become and the harder it becomes to reconstruct long erased alignments.

This is the theory presented in the paper above, much of which has been confirmed. However, the most important and contentious claim made in the paper, that all this can be done quickly and efficiently enough to actually recover any nontrivial amount of overwritten data, and that such techniques are being used by intelligence agencies, remains unproven. Nonetheless, many software packages have been written which wipe files by overwriting the areas in which they are located repeatedly, thus making it virtually impossible to detect the original alignment.

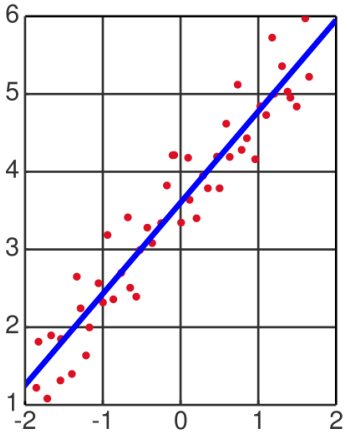
112: Noisy Edge

In this episode an Aerial Anomaly is reported over at least seven locations in Los Angeles. These observations let Charlie plot the aircraft's likely flight path by using a new way of removing noise from noisy signals ("Squish-Squash"). Since this is a sophisticated mathematical method, we first describe a simple analog, the method of least squares.

Least Squares

The path of the Aerial Anomaly that Charlie found was curved, but approximating curves well is difficult. There is a general mathematical principle that to understand something complicated it is best to understand a simple case first, so we will discuss how to fit a line through a group of data points in the plane.

So now our task is the following. Given some collection of data points in the plane $(x_1, y_1), (x + 2, y_2), \ldots, (x_n, y_n)$, find the line $y = ax + b$ that "best" fits our data. Least Squares defines best as the minimum sum of the squared distance between the line $y = ax + b$ and our data points.



Tangent

Carl Friedrich Gauss developed the fundamentals of Least Squares in 1795 at the age of eighteen. It was used to track the movement of the asteroid Ceres in 1801, based on 40 days of observations, after it was lost in the glare of the sun. For more information click here.

Thus, it is our task to find a and b such that

$$S = \sum_{i=1}^{n} (y_i - (ax_i + b))^2$$

is minimized. Now this equation has two parameters that we may vary (a and b), so for us to find the minimum we must find where $\frac{\partial S}{\partial a} = 0$ and $\frac{\partial S}{\partial b} = 0$. (This uses techniques from multivariable calculus. The derivative on the left is a partial derivative with respect to a, which means that we treat a as a variable and all the other variables as constants when we take the derivative.) Note, that normally here we would be finding maxima; however this system of equations has only one minimum (since if there were a max we could make it larger by moving b a little further away).

Taking these derivatives, we get

$$\frac{\partial S}{\partial a} = \sum_{i=1}^{n} -2(y_i - a - bx_i) = 0$$
$$\frac{\partial S}{\partial b} = \sum_{i=1}^{n} -2x_i(y_i - a - bx_i) = 0$$

Rearranging these we get the system of two linear equations with two unknowns.

$$an + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$
$$a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i.$$

After inputting the known values for (x_i, y_i) we may then simply solve for the a and b that give us the Least Square Line of our data set.

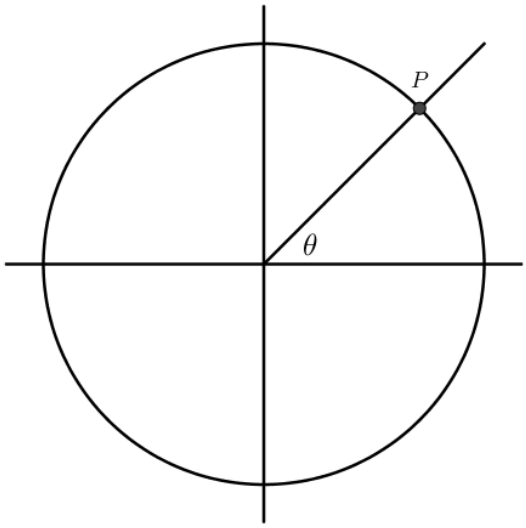
Activity 1
Given the data points $(1,5), (2,7), (3,4)$, and $(4,13)$ find, using the last two linear equations, the Least Squares line.

Activity 2
Given a set of data points that are co-linear in the plane, show that the Least Squares line is indeed the one that passes through these points.

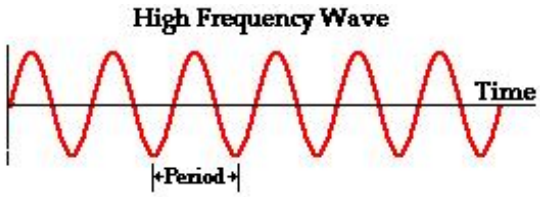
Radars and Signal Processing

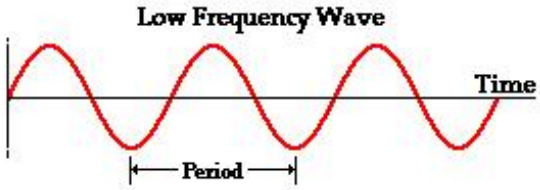
In this episode, the stealth capabilities of the Aerial Anomaly make the object unable to be detected by both civilian and military radars. How is it possible that this could happen? Radars work by emitting and receiving electromagnetic radiation. Typically they use the long-wavelength radio and microwaves, because they reflect off of typical aircraft better and are less likely than shorter-wavelength to scatter or be absorbed before returning. Air traffic control radars rotate on a tower emitting *signals* in every direction at the speed of light. When these signals come into contact with objects they are reflected back at the same speed. Measuring the time that it takes for the signal to return lets us calculate the distance to the object. Also, it is known at what angle this signal was emitted, so we know the approximate location of the object.

Its velocity may also be estimated by shifts in the frequency of the signal. If the echo is of a higher frequency it means that object is traveling towards the radar, and lower frequency implies the direction of travel is away from the radar. The magnitude of the change in frequency indicates the speed of the object.



However, it is not as "easy" as that, the problem is *noise*, which is the mathematical term for errors in measurements. These can be caused by nearby objects, irregularities in the object, or stray radiation. An everyday example of noise is the static on empty TV stations. When trying to find objects such as large jet liners, the reflected *blip* is so strong that the noise is generally insignificant. Airplanes with stealth capabilities reflect so little of the signal that it's hard to tell the difference between their echo and the noise. One nice thing about noise is that it's generally high frequency and the echoes are low frequency. This is also true for the static on tv stations: if you look at any particular pixel in static it will change very frequently, but if you look at a pixel on a tv show, then it will change much less frequently.



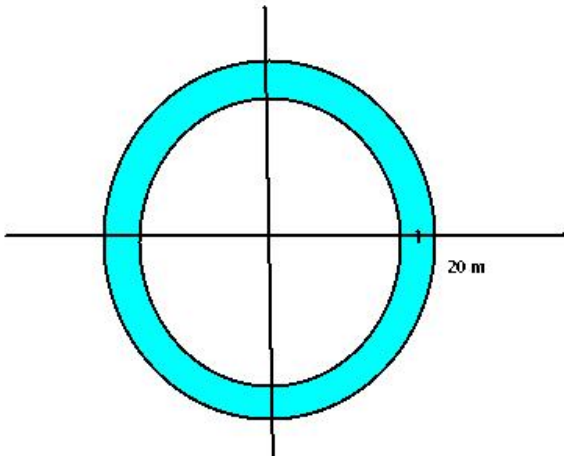


Summing these two waves gives us some idea of what the radar would receive when it found an object. To try to remove the noise, one can use an electronic "low pass filter" (see here for more information). If the signals are being processed digitally, then using real-time Fourier Analysis (see here) can help remove the noise.

In this episode, Charlie compares this situation with an audience clapping. Suppose everyone is clapping at a high frequency, except for one person who is clapping at a lower frequency. The single person's clapping sound would be similar to the bottom wave and the rest of the audience's sound would look like the top wave. Recording equipment would record all of these signals at the same time and we could use either type of filter to find the single clapper's frequency. Also, if the slower clapper is clapping with the same loudness and we had 3 recordings, using the idea of the radar in finding distance, we could find (approximately) their location in the room. However, this is not exact, since it is possible for patterns of high frequency waves to look like low frequency waves. This effect is called aliasing (see here for more information). An example of aliasing called the "wagon wheel effect" is visible in most movies that we see. Video equipment is not continuously recording the movements of a car wheel, but rather taking a sampling of images (at say 30 frames/second). Aliasing makes it seem like the wheel is moving more slowly than it actually is, and can even make it look like it is going backwards.

Another problem with finding signals of planes with stealth capacity is that the signal may be so weak that the filtering throws it out as noise. However, if the plane is moving at a relatively constant speed we can program the detector to look for patterns in the signal, which will be separated by the distance it travels in consecutive sweeps of the radar. It is assumed that noise will not appear in any coherent pattern, so we may test signals for moving patterns. This is where the reference to "Squish-Squash" came up in the episode.

To test the data for correlation we must construct a function g (based on our estimate of the speed of the plane) that gives the probability that two data points p and q are related, so that $g(p-q)$ is the probability that p and q are the same object. In other words, if the plane starts at the origin, $g(p)$ is the probability that it will be at the point p one unit of time later. In Noisy Edge, there were 7 sightings of the plane, so using the location and times of the sightings we could estimate the Aerial Anomaly's speed (say it was 20 m/s). Just for simplicity, assume that, after we found the object's expected flight path, we also had a radar directly under its predicted flight path, and that the radar does a full rotation/sec. Then the function g that we might want to use would look like this.



Here g is 1 in the blue region and 0 outside of the blue region. There are several other more technical things to consider when deciding what the function g should be. After this, if there is a suspected data point we can narrow our search for the plane after one unit of time to the areas for which g has a high value. A recent paper explains how to use this a "Squish-Squash" algorithm to modify the function g to make it more likely that data points will be found. It is called "Continuum Percolation with Unreliable and Spread-Out Connections" by M. Franceschetti et.al., and it appears online here. However, this paper requires a significant amount of background to read.

113: Man Hunt

In this episode, Charlie uses Bayesian inference and Markov chains to help his brother figure out the cause of a bus crash which let two prisoners escape from a prison bus.

Bayesian Inference

Bayesian inference is the process of adjusting your belief in the probability of the truth of some statement based on new events. As a silly example, you may believe that it is extremely unlikely that there are any purple cows in the world. However, if you see a purple cow, then you would obviously change your belief to the belief that there is at least one purple cow in the world.

To give a mathematical basis for this inference process, we need to state Bayes' Theorem. Suppose we have two different events, A and B. We'll write $P(A)$ for the probability that A occurs and $P(B)$ for the probability that B occurs. Also, we'll write $P(A \mid B)$ for the probability that A happens given that B happens. We can rewrite this as $P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}$, that is, $P(A \mid B)$ is the probability that A and B both happen divided by $P(B)$. Then Bayes rule is the following formula:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

This can be proved very simply using the formula for $P(A \mid B)$. Now let's look at an example.

Let's say that Billy Bob has two boxes, one with one fair die in it, labelled with the numbers 1 through 6, and one with two fair dice in it, labelled with the numbers 1 through 6. Let's say he picks one box at random and then rolls the dice in the box. Let's use Bayes' rule to figure out the probability that he picked the box with one die given the fact that the total of the die (or dice) he rolled was a 3. First, we need to translate the words into the symbols we used in the previous paragraph. Let's use A to denote the event that Billy Bob picks the box with one die and B the event that he rolls a 3. Then as we assumed, $P(A) = 0.5$, and we can calculate $P(B) = 0.5 \cdot \left(\frac{1}{6}\right) + 0.5 \cdot \left(\frac{2}{36}\right) = \frac{4}{36} = \frac{1}{9}$. This is because there's a $\frac{1}{2}$ chance he'll pick box 1, and if he does there's a $\frac{1}{6}$ chance he'll roll a 3. Also, there's a $\frac{1}{2}$ chance he'll pick box 2 and a $\frac{2}{36}$ chance he'll roll a 3 (he can get a 1,2 or a 2,1). Now since what we want to figure out the probability that he picked box 1 (event A) given the fact that he rolled a 3 (event B), we want to use Bayes' rule to figure out $P(A \mid B)$. To use the formula we need to figure out what $P(B \mid A)$ is. This is the probability that he rolls a 3 given the fact that he picks box A, so $P(B \mid A) = \frac{1}{6}$. Then $P(A \mid B) = \left(\frac{1}{6}\right) \cdot \left(\frac{1}{2}\right) / \left(\frac{1}{9}\right) = \frac{3}{4}$. This makes sense because if he picks the box with two

dice he is unlikely to get a 3, while if he picks the box with 1 die he is more likely to get a 3.

Tangent

Bayes' rule is attributed to the Reverend Thomas Bayes, a Presbyterian minister who lived from 1701 to 1761 in a work entitled "Essay Towards Solving a Problem in the Doctrine of Chances." This was a response to a work entitled "The Doctrine of Chances" by Abraham de Moivre, one of Bayes' mathematical colleagues. A link to Bayes original paper (which was published posthumously) can be found here.

Activity 1

1. If A is as above and B is the event that Billy Bob rolls a 6, what is $P(A \mid B)$?
2. If A is as above and B is the event that Billy Bob rolls a 5 or a 6, what is $P(A \mid B)$?
3. If A is as above and B is the event that Billy Bob rolls a 1, what is $P(A \mid B)$?

Activity 2

This problem is loosely based on a situation that occurred in the tv show "Let's Make a Deal." Let's say you're on a game show and the host presents you with three doors. One has a shiny new car behind it, and the other two have goats. The object (of course) is to get the car and not to get a goat. Now the game goes like this. First, you pick a door. Then at least one of the two doors you did not pick has a goat behind it. The game show host knows what the shut doors have behind them and he deliberately opens one of the two that you did not pick to show you a goat (if both have a goat, he randomly picks the one to open). Then he lets you either stay with your original choice of door or switch to the other unopened door. Then you get what is behind that door. Now the goal of the problem is to use Bayes' rule to figure out whether it is better to switch or stay put. Let's say you picked door 1 and he opened door 2. Let A be the event that door 3 has a car behind it and let B be the event that he opened door 2.

1. What is $P(A)$? Remember this refers to the probability that A occurs before you know anything that happens.
2. What is $P(B)$? Remember this refers to the probability that B occurs before you know anything that happens.
3. What is $P(B \mid A)$? This is the probability that door 3 has a goat behind it given the fact that door 2 has a goat behind it. (Hint: how many goats and cars are left after door 2 has been opened?)
4. What is $P(A \mid B)$?
5. Now should you switch?

Markov Chains

A Markov chain is a series of random events where the outcome of event n is only dependent on the outcome of the event $n-1$. For example, let's say there is a number line where each integer has a dot on it, and let's say that there is also an evil but useless robot standing on position 0. The robot has a quarter, and he flips it to determine whether he should go left or right. If he gets heads, he'll move to position 1, and if he gets tails he'll move to position -1. This is the first random event. Now he repeats this process over and over to generate a series of random events, where each time if he gets heads he moves right 1 step from his current position and if he gets tails he moves left 1 step from his current position (let's say right is the positive direction). This is a Markov chain because where the robot goes on his next step only depends on his current position and does not depend on where he's been in the past. The output of each event is the current position of the robot.

Now let's study this Markov chain a little bit to see how it behaves. One question we might ask is what is the average position of the robot after n steps. (A more technical way of phrasing this question is to ask what the expected value of the robot is after n steps.) To measure this we could make the robot take n steps, record its position, put it back on zero, repeat these three steps many times, and then average the results. It's actually pretty easy to see that the average should be 0, independent of what n is. To prove this, we can use mathematical induction. Let's say n is 0. Then the robot never moves at all, so its average position is 0. Now let's assume that after n steps the average position of the robot is 0. Then on the $n+1$ the robot has a $\frac{1}{2}$ chance of being at position 1 and a $\frac{1}{2}$ chance of being at position -1. Therefore the average position of the robot after $n+1$ steps is also 0. This proves that no matter what n is, the robots average position after n steps is 0.

Another question we could ask about this Markov chain is what the robot's average distance from the zero position is after n steps. This question is a little trickier, but still doable. Let's let x_n be the robots position after n steps (so x_n is a random variable) and let's define a new random variable $y_n = x_n^2 - n$. Also let's write $E(-)$ for the average value of $-$. Then we have the equations:

$$E(y_n) = E(x_n^2 - n) = E\left(\frac{(x_{n-1}+1)^2 + (x_{n-1}-1)^2}{2} - n\right) = E(x_{n-1}^2 + 1 - n) = E(y_{n-1}).$$

Here the second equality comes from the fact that half the time $E(x_n^2) = (x_n+1)^2$ and the other half of the time $E(x_n^2) = (x_n-1)^2$. Since the average distance of the robot after 0 steps is 0, we see that the average distance of the robot from 0 is $\sqrt{n}$.

201: Judgement Call

In this episode, Charlie discusses Bayes' Theorem and its application to Bayesian filters. Bayesian filters are used for a number of applications, including Spam filtering. This activity discusses conditional probability and Bayes' theorem.

Conditional Probability

In order to understand Bayes' Theorem, we must first understand a little about conditional probability. Given two events, they are either dependent or independent. For example, drawing a card from a deck of 52 and flipping a coin are independent events. Knowing the results of the card draw cannot possibly help you to predict the results of the coin toss, they are essentially unrelated. Drawing two cards without replacement, however, is an example of dependent events. If the first card you draw is red, you know that the probability of drawing a second red card is less than 50% (in particular, 25/51).

Probability Warm-Up

Let's start with a few exercises in probability. Assume you have a standard a 52 card deck.

1. What is the probability of drawing a red queen? Of drawing two red queens without replacement?
2. You are going to draw two cards in succession. Your first card is a red queen. What is the probability that your second card will be a red queen?
3. This time, your first card is a black queen. What is the probability that your second card will be a red queen?

A *conditional probability*, written $P(A \mid B)$, gives the probability that and event A occurs, given B. For example, if B is the event "your first card is a red queen", and A denoted "your second card is a red queen" then $P(A \mid B)$ would be the answer to the second question above. If the event C denotes "your first card is a black queen" then $P(A \mid C)$ would be your final answer.

Bayes' Theorem gives a convenient formula for conditional probabilities. It states: $P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$. This formula may not look useful at first, but oftentimes it is far easier to find one conditional probability than another, or one is given to you. Consider the following example:

Using Bayes' Theorem

You are given two jars containing 5 gumballs each:

- Jar 1 contains 4 red gumballs and 1 blue.
- Jar 2 contains 1 red gumball, 3 blue, and 1 green.

Consider the following problems:

1. You reach into a jar without seeing its number, and you grab a red gumball. Disappointed (that won't turn my tongue a funny color at all!), you put it back. What is the probability that you drew from Jar 1?
2. While you are busily computing probabilities, your friend reaches into Jar 1, takes a gumball, and pops it in his mouth. You then reach into Jar 1 and draw out a red gumball. What is the probability that your friend's gumball is blue?
3. Angry, and realizing the error of your ways, you reach into Jar 2 in the hopes of finding a blue gumball. What are the chances that your wish is satisfied?

Finally satisfied with your choice of candy, you decide to turn your attention to heavier matters. A new generation of evil mutants is evolving right here in the United States. Scientists predict that approximately 0.05% of the population are evil mutants. A doctor has just invented a test for the mutant gene that he claims is "99% accurate." We will assume that this indicates that for 1% of people, the test yields the *incorrect* result, whatever that may be. The doctor submits that the FBI should begin using this test to lock away mutants immediately, but Charlie has some reservations...

Identifying Evil Mutants

Let A represent that a person is NOT a mutant, and let B represent that a person tests positive for the mutant gene.

1. Find $P(B)$. Remember to take the tests of both mutants and normals into account.
2. Find $P(B \mid A)$, taking into account that the test is "99% accurate."
3. Find $P(A \mid B)$, i.e. the probability that a person who tests positive is not a mutant.
4. Do you see a problem (other than constitutional and basic moral issues) with the doctor's plan?

202: Better or Worse

In this episode, Charlie tries to identify a criminal via the car keys in her purse. He explains that the remote used to unlock a car from a distance will be difficult to trace since it uses a *rolling code*. After all, if a remote entry device were to transmit the same signal each time, it would be a relatively simple manner for a thief to intercept and duplicate said signal. Instead, car keys use a mathematical algorithm to send a different n digit number each time the button is pressed, which the locking mechanism can identify as right or wrong.

Rolling Codes

The goal of a rolling code is to use a structured formula to change an n-digit number over time. A number of formulas for this change will be *recursive*. A formula is recursive if the output at a given time depends on previous outputs. For example, if we use a_i to denote the ith output, $a_{i+1} := a_i + 1$ is a recursive formula. If $a_0 = 12$, the resulting sequence of outputs will be 13, 14, 15, 16, etc..

Another common recursive formula is used to generate the Fibonacci sequence. The sequence begins with the numbers 0 and 1. Each successive number is the sum of the previous two. The sequence thus continues 0, 1, 1, 2, 3, 5, 8, 13, 21, etc..

Recursive formulas make for good rolling codes, as a thief would generally need to figure out the recursive formula, the initial input, *and* the number of iterations that have occurred in order to predict the next output. However, to use practically, a sequence cannot be simply increasing as above- the remote key and car would eventually be overloaded by the length of the digits. Thus, the function must somehow control the length of outputs.

> Tangent
>
> The method actually used in remote entry devices is somewhat more complicated than the examples described here. For more information about this method, and the history of attacks on it, see Keeloq.

Modular arithmetic is frequently used to control the length of numbers in computer science. Some of you may be familiar with the term 'modulus' from programming or 'clock arithmetic'. In modular arithmetic, integers that differ by a given positive number are considered equal. For example, $0 = 4 = 8 = 12 \mod 4$. Similarly, $1 = 7 = 13 = 19$

mod 6. You can use modular arithmetic to find shorter expressions of numbers If you are working mod n, simply take the remainder of each number under division by n. The resulting numbers will, in particular, all be smaller than n.

One problem with rolling codes described by *recursive* functions such as these is a finite period. If a code depends only on the previous output, then as soon as the code repeats a single output, it must cycle. Watch out for this problem while working with your rolling codes.

Activity 1

Our new rolling code will be given by taking a number, squaring it, and modding out by n (modding out means taking the remainder under division). This can be rewritten as $a_i + 1 := a_i^2 2 \mod n$.

- Does this generally generate a better code for larger n or smaller n? Why?
- Which inputs make for the shortest periods, and which inputs make for the largest?

Activity 2

Take a number of length n and square it, adding zeroes at the beginning if necessary to obtain a number of length $2n + 1$. Take the middle n digits of this number as the output.

203: Obsession

In the episode Obsession, a photographer is killed while attempting to take tabloid sex photos. In an effort to retrace the photographers steps, the FBI examine his camera. They find a series of photos of a basketball hoop, each showing the hoop's shadow. Since each photograph is timestamped, Charlie was able to use the behavior of the hoop's shadow to pinpoint the longitude and latitude where the photos had been taken.

Spheres and Shadows

In order to make calculations involving the position of the sun in the sky, Charlie used a technique called spherical astronomy. In this subject, all observable planetary bodies and stars are assumed to lie on the celestial sphere, an imaginary rotating sphere of gigantic radius concentric to the Earth. In ancient Greece, this model was believed to be an accurate representation of the universe. Although we now know that these observed rotations are actually caused by the rotation of the earth itself rather than some supersphere, scientists still use the techniques of spherical astronomy, which are often quite accurate and simplify calculations a great deal.

Spherical astronomy works particularly well for calculations involving constellations and other far away objects, as their movement is barely perceptible from Earth (the movement of stationary bodies will be solely determined by the rotation of earth). When doing long term calculations with closer bodies; however, other considerations must be taken into account. For example, the position of the Sun on the celestial sphere (i.e. in relation to all the other stars) is constantly shifting, moving a little under one degree eastward in each 24 hour period. The sun also moves due to the tilting of the earth on its axis. The path traced by the sun on the celestial sphere is called the *ecliptic*.

Tangent

The tilt of the Earth on its axis and the eccentricity of its orbit around the sun results in more complex solar movements than described here. For more information, (and some great pictures) see this article on the analemma, a figure eight path the sun traces in the sky in the course of a year.

Activity 1

1. How long will it take for the Sun to return to its original position in the celestial sphere?
2. The north star is famous for being virtually stationary in the night sky. From the perspective of the celestial sphere, how is Polaris accomplishing this feat?
3. Assume for the sake of simplicity that the sun remains fixed on the celestial sphere during a 24 hour time period, how would you describe the set of points on Earth that experience the Sun *directly* overhead at some time during this period?

The assumption of a celestial sphere makes it relatively easy to describe the position of a celestial body from the perspective of a person standing on the earth. The position of the sun, for example, can be related by two angles: the altitude is the angle between the sun and the observable horizon, whereas the the azimuth is the angle on the horizon measured clockwise from due south. From an observer on the earth's prospective, all points with the same altitude form a circle in the sky at that altitude, whereas all points with the same azimuth form a semicircle in the sky, passing directly overhead and intersecting the horizon at a right angles.

Activity 2

1. When a tall pole casts a shadow, what determines the length of the shadow? What determines it's position?
2. How would a shadow cast at at 3pm compare to a shadow cast at 3:30?
3. Two basketball hoops (each 10 feet tall) are photographed at precisely the same afternoon time. Assume that these basketball hoops are at the same latitude, but one is 200 miles east of the other. How could you differentiate the photographs?
4. Assume a basketball hoop is casting a 5 foot shadow. What is the sun's altitude at the time?

The method of representing points on a sphere that we have used here can be extended to spherical coordinates. You are no doubt accustomed to discussing points in 2-dimensional space, or the *Euclidean plane* as pairs of points (x, y). This idea can be extended by adding a third z-axis. Imagine that the xy-plane that you are used to lies on a chalkboard, the z axis will be coming directly out of the chalkboard, perpendicular to both the x and y axes. Now we can describe all points in three dimensional space as (x, y, z).

Spherical coordinates provide a different, and often useful way of describing the po-

sition of objects in three dimensional space. The basic principle of this coordinate system is that every point lies on *some* sphere centered at the origin, so one need only specify which sphere (using radius), and the precise coordinates of the point on said sphere (the angles of azimuth and altitude). Therefore a point is written in spherical coordinates as (r, θ, ϕ), where θ represents the azimuth, ϕ represents altitude, and r represents the distance between the point and the origin.

Activity 3

1. Describe the set of all points in three space with a radius of 5.
2. Describe the set of all points in three space with an altitude of 45°.
3. Describe the set of points in three space with an azimuth of 90°.
4. Challenge: Come up with a formula for converting the spherical coordinates (r, θ, ϕ) of a point into in Cartesian coordinates: (x, y, z). Note: You will need the distance formula in three dimensions, and more than one right triangle.