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,

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

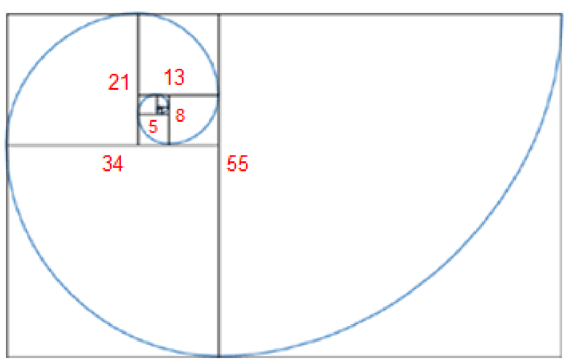
Tangent

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

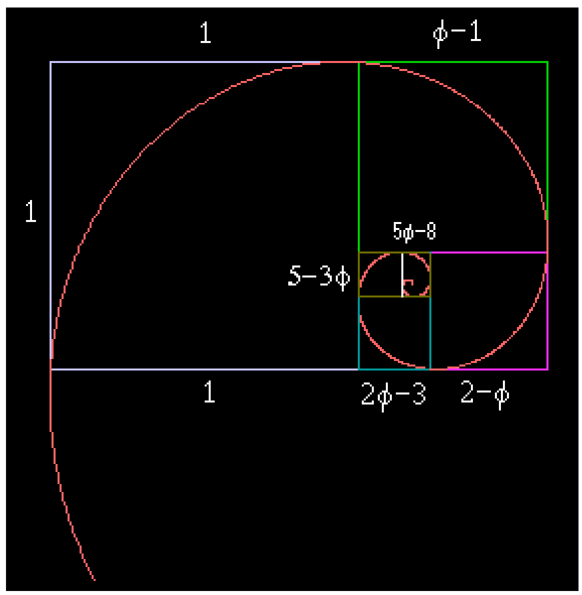
The Fibonacci numbers satisfy a numerous amount of very interesting identities. The reader can find some of them at http://mathworld.wolfram.com/FibonacciNumber.html.

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 representation (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,

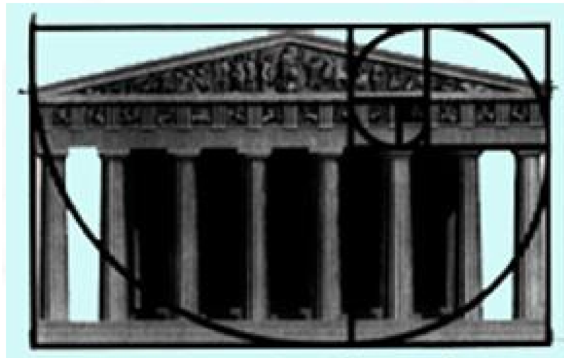
$$\phi = \sqrt{1 + \sqrt{1 + \sqrt{1 + \sqrt{1 + \cdots}}}}$$

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

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

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

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?