The Sato-Tate Conjecture

Dan Fretwell

6th July 2013

Outline of talk

A large part of number theory is concerned with solving Diophantine equations, polynomial equations to be solved over the integers ($\mathbb{Z}$) or the rationals ($\mathbb{Q}$).

For example in two variables we can use modular arithmetic to solve (over $\mathbb{Z}$):

1. Linear Diophantine equations: $ax + by = c$ for $a, b, c \in \mathbb{Z}$,
2. Quadratic Diophantine equations: conics of the form $ax^2 + bxy + cy^2 = d$ for $a, b, c, d \in \mathbb{Z}$.

A large part of number theory is concerned with solving Diophantine equations, polynomial equations to be solved over the integers ($\mathbb{Z}$) or the rationals ($\mathbb{Q}$).

For example in two variables we can use modular arithmetic to solve (over $\mathbb{Z}$):

1. Linear Diophantine equations: $ax + by = c$ for $a, b, c \in \mathbb{Z}$,
2. Quadratic Diophantine equations: conics of the form $ax^2 + bxy + cy^2 = d$ for $a, b, c, d \in \mathbb{Z}$.

A non-trivial Diophantine

Simple sounding problems can produce more difficult Diophantines though!

Congruent Number Problem (CNP)

Which positive integers can be the area of a right angled triangle?

This question is boring if we allow real valued side lengths but becomes very interesting if we only allow rational side lengths.

For example 6 is the area of a $(3, 4, 5)$-triangle and 5 is the area of a $(\frac{20}{3}, \frac{3}{2}, \frac{41}{6})$-triangle. We call 5 and 6 congruent numbers.

Fermat was able to prove that 1 is not a congruent number!

We have the following theorem about congruent numbers:

Theorem

Let d be a positive integer.

Then d is a congruent number if and only if the Diophantine equation $y^2 = x^3 - d^2x$ has infinitely many rational solutions.

Problem: How does one show such an equation has infinitely many rational solutions?

We have the following theorem about congruent numbers:

Theorem

Let d be a positive integer.

Then d is a congruent number if and only if the Diophantine equation $y^2 = x^3 - d^2x$ has infinitely many rational solutions.

Problem: How does one show such an equation has infinitely many rational solutions?

The Cannonball Problem

Another example is the following problem:

The Cannonball Problem

Imagine you have a square-pyramidal stack of cannonballs. How many must you have to be able to rearrange into a square?

Solution

You must have exactly 1 or $4900 = 70^2$ cannonballs.

It turns out that the above solution arises from studying the integer solutions of the Diophantine equation $y^2 = x^3 - 36x$.

The Cannonball Problem

Another example is the following problem:

The Cannonball Problem

Imagine you have a square-pyramidal stack of cannonballs. How many must you have to be able to rearrange into a square?

Solution

You must have exactly 1 or $4900 = 70^2$ cannonballs.

It turns out that the above solution arises from studying the integer solutions of the Diophantine equation $y^2 = x^3 - 36x$.

The Cannonball Problem

Another example is the following problem:

The Cannonball Problem

Imagine you have a square-pyramidal stack of cannonballs. How many must you have to be able to rearrange into a square?

Solution

You must have exactly 1 or $4900 = 70^2$ cannonballs.

It turns out that the above solution arises from studying the integer solutions of the Diophantine equation $y^2 = x^3 - 36x$.

Outline of talk

What is an Elliptic Curve?

For the purposes of this talk, elliptic curves are a special family of cubic curves that are of the form:

$$y^2 = x^3 + Ax + B,$$

where $A, B \in \mathbb{Z}$ are such that $4A^3 + 27B^2 \neq 0$.

The second condition looks baffling but it just guarantees the cubic has distinct roots. It turns out that if you fail this condition then your curve is basically ”the same” as a line and so is much simpler for us to handle.

Applications

Elliptic curves have been found to be very popular in the past 30-40 years. Here are some uses of them:

1. Wiles' proof of Fermat's Last Theorem.
2. Modern cryptography. Chances are you have one assigned to you without knowing it!
3. They have been found to contain many mysterious links with modular forms, integer factorisation, lattices, sphere packings, string theory,...

From a pure perspective they are the next type of Diophantine you would want to be able to solve after linear ones and conics.

How does one go about finding rational solutions to such curves? In practice it is very hard! (Hence why the CNP is tough to solve).

What about integer solutions? Again this is very hard but we have the following remarkable theorem:

Siegel's Theorem

Every elliptic curve has only finitely many integer solutions.

Unfortunately the proof of the above theorem is an existence proof, it doesn't tell us how to find the solutions.

How does one go about finding rational solutions to such curves? In practice it is very hard! (Hence why the CNP is tough to solve).

What about integer solutions? Again this is very hard but we have the following remarkable theorem:

Siegel's Theorem

Every elliptic curve has only finitely many integer solutions.

Unfortunately the proof of the above theorem is an existence proof, it doesn't tell us how to find the solutions.

How does one go about finding rational solutions to such curves? In practice it is very hard! (Hence why the CNP is tough to solve).

What about integer solutions? Again this is very hard but we have the following remarkable theorem:

Siegel's Theorem

Every elliptic curve has only finitely many integer solutions.

Unfortunately the proof of the above theorem is an existence proof, it doesn't tell us how to find the solutions.

Idea: Reduction mod p.

Example

Let E be the elliptic curve with equation $y^2 = x^3 + x + 1$. Then for each prime p (apart from two "bad" primes 2 and 31) we can reduce mod p and get elliptic curves E_p.

The points are as follows (for $p = 3, 5$ and 7):
$E_3 = \{(0, \pm 1), (1, 0)\}$,
$E_5 = \{(0, \pm 1), (2, \pm 1), (3, \pm 1), (4, \pm 2)\}$,
$E_7 = \{(0, \pm 1), (2, \pm 2)\}$.

Note:

- Not every x gives a solution, for example on E_3 there is no point with $x = 2$.
- If an x does give a solution then you almost always get a pair, unless you happen to hit $y = 0$ (very rare).

Idea: Reduction mod p.

Example

Let E be the elliptic curve with equation $y^2 = x^3 + x + 1$. Then for each prime p (apart from two "bad" primes 2 and 31) we can reduce mod p and get elliptic curves E_p.

The points are as follows (for $p = 3, 5$ and 7):
$E_3 = \{(0, \pm 1), (1, 0)\}$,
$E_5 = \{(0, \pm 1), (2, \pm 1), (3, \pm 1), (4, \pm 2)\}$,
$E_7 = \{(0, \pm 1), (2, \pm 2)\}$.

Note:

- Not every x gives a solution, for example on E_3 there is no point with $x = 2$.
- If an x does give a solution then you almost always get a pair, unless you happen to hit $y = 0$ (very rare).

Naive intuition: If the solution counts $N_p := |E_p|$ tend to be "small" then E is unlikely to have many integer points.

The curve in our previous example has:

$$N_3 = 3 \quad N_5 = 8 \quad N_7 = 4 \quad N_{11} = 13 \quad N_{13} = 17 \quad N_{17} = 17.$$

We notice that N_p is increasing steadily with p. Our intuition is incorrect!

In fact it can be shown that there are only 4 integer solutions on E, given by $\{(0, \pm 1), (72, \pm 611)\}$.

Naive intuition: If the solution counts $N_p := |E_p|$ tend to be "small" then E is unlikely to have many integer points.

The curve in our previous example has:

$$N_3 = 3 \quad N_5 = 8 \quad N_7 = 4 \quad N_{11} = 13 \quad N_{13} = 17 \quad N_{17} = 17.$$

We notice that N_p is increasing steadily with p. Our intuition is incorrect!

In fact it can be shown that there are only 4 integer solutions on E, given by $\{(0, \pm 1), (72, \pm 611)\}$.

Naive intuition: If the solution counts $N_p := |E_p|$ tend to be "small" then E is unlikely to have many integer points.

The curve in our previous example has:

$$N_3 = 3 \quad N_5 = 8 \quad N_7 = 4 \quad N_{11} = 13 \quad N_{13} = 17 \quad N_{17} = 17.$$

We notice that N_p is increasing steadily with p. Our intuition is incorrect!

In fact it can be shown that there are only 4 integer solutions on E, given by $\{(0, \pm 1), (72, \pm 611)\}$.

The Hasse bound

So how big do we expect N_p to be for a fixed p?

Well, clearly $N_p \leq p^2$ although as we just saw N_p is usually quite close to p. The Hasse bound tells us about the deviation of N_p from p (ignoring a few "bad" primes).

Hasse bound

Let E be an elliptic curve. Then $|N_p - p| \leq 2\sqrt{p}$ for each "good" prime. In other words N_p almost always lies inbetween $p - 2\sqrt{p}$ and $p + 2\sqrt{p}$.

The Hasse bound

So how big do we expect N_p to be for a fixed p?

Well, clearly $N_p \leq p^2$ although as we just saw N_p is usually quite close to p. The Hasse bound tells us about the deviation of N_p from p (ignoring a few "bad" primes).

Hasse bound

Let E be an elliptic curve. Then $|N_p - p| \leq 2\sqrt{p}$ for each "good" prime. In other words N_p almost always lies inbetween $p - 2\sqrt{p}$ and $p + 2\sqrt{p}$.

The Hasse bound

So how big do we expect N_p to be for a fixed p?

Well, clearly $N_p \leq p^2$ although as we just saw N_p is usually quite close to p. The Hasse bound tells us about the deviation of N_p from p (ignoring a few "bad" primes).

Hasse bound

Let E be an elliptic curve. Then $|N_p - p| \leq 2\sqrt{p}$ for each "good" prime. In other words N_p almost always lies inbetween $p - 2\sqrt{p}$ and $p + 2\sqrt{p}$.

Random walks...

The Hasse bound has a probabalistic interpretation!

For each x from 0 to $p-1$ the value of $x^3 + Ax + B$ either is a square mod p or it isn't. In either case, for each x we get a contribution of $1 + \epsilon_x$ to N_p (where $\epsilon_x \in \{0, \pm 1\}$).

Thus:

$$N_p = \sum_{i=0}^{p-1}(1 + \epsilon_x) = p + \sum_{i=0}^{p-1} \epsilon_x$$

.

The Hasse bound is really telling us that the sum:

$$N_p - p = \sum_{i=0}^{p-1} \epsilon_x$$

behaves like a random walk. The significance of $\sqrt{p}$ in the bound is from the expectation of such a walk!

Random walks...

The Hasse bound has a probabalistic interpretation!

For each x from 0 to $p-1$ the value of $x^3 + Ax + B$ either is a square mod p or it isn't. In either case, for each x we get a contribution of $1 + \epsilon_x$ to N_p (where $\epsilon_x \in \{0, \pm 1\}$).

Thus:

$$N_p = \sum_{i=0}^{p-1}(1 + \epsilon_x) = p + \sum_{i=0}^{p-1} \epsilon_x$$

.

The Hasse bound is really telling us that the sum:

$$N_p - p = \sum_{i=0}^{p-1} \epsilon_x$$

behaves like a random walk. The significance of $\sqrt{p}$ in the bound is from the expectation of such a walk!

Random walks...

The Hasse bound has a probabalistic interpretation!

For each x from 0 to $p-1$ the value of $x^3 + Ax + B$ either is a square mod p or it isn't. In either case, for each x we get a contribution of $1 + \epsilon_x$ to N_p (where $\epsilon_x \in \{0, \pm 1\}$).

Thus:

$$N_p = \sum_{i=0}^{p-1}(1 + \epsilon_x) = p + \sum_{i=0}^{p-1} \epsilon_x$$

.

The Hasse bound is really telling us that the sum:

$$N_p - p = \sum_{i=0}^{p-1} \epsilon_x$$

behaves like a random walk. The significance of $\sqrt{p}$ in the bound is from the expectation of such a walk!

Outline of talk

Ok, so we have a lovely bound for the numbers $a_p = N_p - p$.

Refined intuition: If the shifted solution counts a_p are mainly negative (i.e. $N_p < p$) then E is unlikely to have many integer points.

Question: How are the a_p distributed within the Hasse bound as the prime p varies?

Problem: As p increases so does $2\sqrt{p}$, giving bigger and bigger intervals.

Answer: Normalize by studying the numbers $b_p = \frac{a_p}{2\sqrt{p}}$ instead. All of these values satisfy $|b_p| \leq 1$ by the Hasse bound.

In fact since the b_p's lie between -1 and 1, we can associate an angle θ_p to each via $b_p = \cos(\theta_p)$, where each $\theta_p \in [0, \pi]$.

Ok, so we have a lovely bound for the numbers $a_p = N_p - p$.

Refined intuition: If the shifted solution counts a_p are mainly negative (i.e. $N_p < p$) then E is unlikely to have many integer points.

Question: How are the a_p distributed within the Hasse bound as the prime p varies?

Problem: As p increases so does $2\sqrt{p}$, giving bigger and bigger intervals.

Answer: Normalize by studying the numbers $b_p = \frac{a_p}{2\sqrt{p}}$ instead. All of these values satisfy $|b_p| \leq 1$ by the Hasse bound.

In fact since the b_p's lie between -1 and 1, we can associate an angle θ_p to each via $b_p = \cos(\theta_p)$, where each $\theta_p \in [0, \pi]$.

Ok, so we have a lovely bound for the numbers $a_p = N_p - p$.

Refined intuition: If the shifted solution counts a_p are mainly negative (i.e. $N_p < p$) then E is unlikely to have many integer points.

Question: How are the a_p distributed within the Hasse bound as the prime p varies?

Problem: As p increases so does $2\sqrt{p}$, giving bigger and bigger intervals.

Answer: Normalize by studying the numbers $b_p = \frac{a_p}{2\sqrt{p}}$ instead. All of these values satisfy $|b_p| \leq 1$ by the Hasse bound.

In fact since the b_p's lie between -1 and 1, we can associate an angle θ_p to each via $b_p = \cos(\theta_p)$, where each $\theta_p \in [0, \pi]$.

The Sato-Tate conjecture

The Sato-Tate conjecture tells us how these angles θ_p are distributed in $[0, \pi]$. It shows that even our new intuition is incorrect!

There are a rare class of elliptic curves that have their θ_p's uniformly distributed on $[0, \pi]$. Sato-Tate deals with the rest.

Sato-Tate

Let E be a "typical" elliptic curve and take $\alpha, \beta \in [0, \pi]$ with $\alpha \leq \beta$. Then for a randomly chosen prime p:

$$P(\alpha \leq \theta_p \leq \beta) = \frac{2}{\pi} \int_{\alpha}^{\beta} \sin^2(\theta) d\theta.$$

In other words the θ_p's are uniformly distributed with respect to the circular measure $\frac{2}{\pi}\sin^2(\theta)d\theta$.

The Sato-Tate conjecture

The Sato-Tate conjecture tells us how these angles θ_p are distributed in $[0, \pi]$. It shows that even our new intuition is incorrect!

There are a rare class of elliptic curves that have their θ_p's uniformly distributed on $[0, \pi]$. Sato-Tate deals with the rest.

Sato-Tate

Let E be a "typical" elliptic curve and take $\alpha, \beta \in [0, \pi]$ with $\alpha \leq \beta$. Then for a randomly chosen prime p:

$$P(\alpha \leq \theta_p \leq \beta) = \frac{2}{\pi} \int_{\alpha}^{\beta} \sin^2(\theta) d\theta.$$

In other words the θ_p's are uniformly distributed with respect to the circular measure $\frac{2}{\pi}\sin^2(\theta)d\theta$.

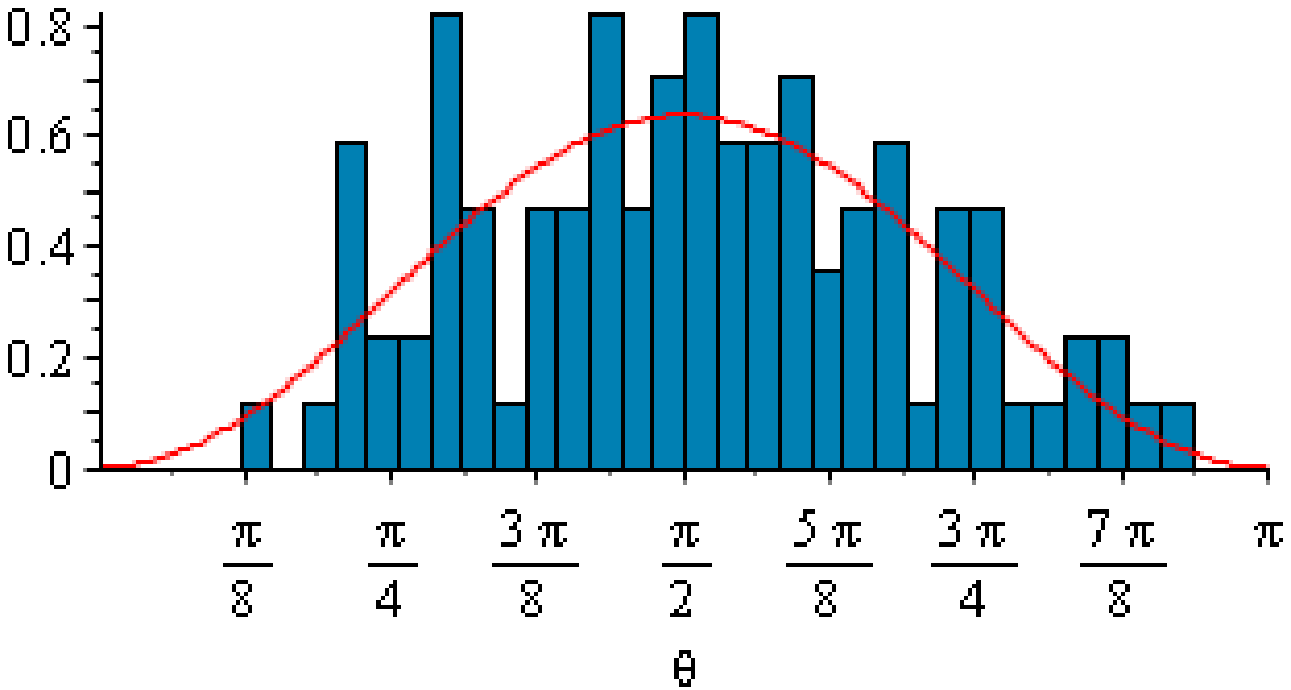
Evidence

Returning to the elliptic curve $E : y^2 = x^3 + x + 1$, it is extremely easy to calculate tables of θ_p values by hand:

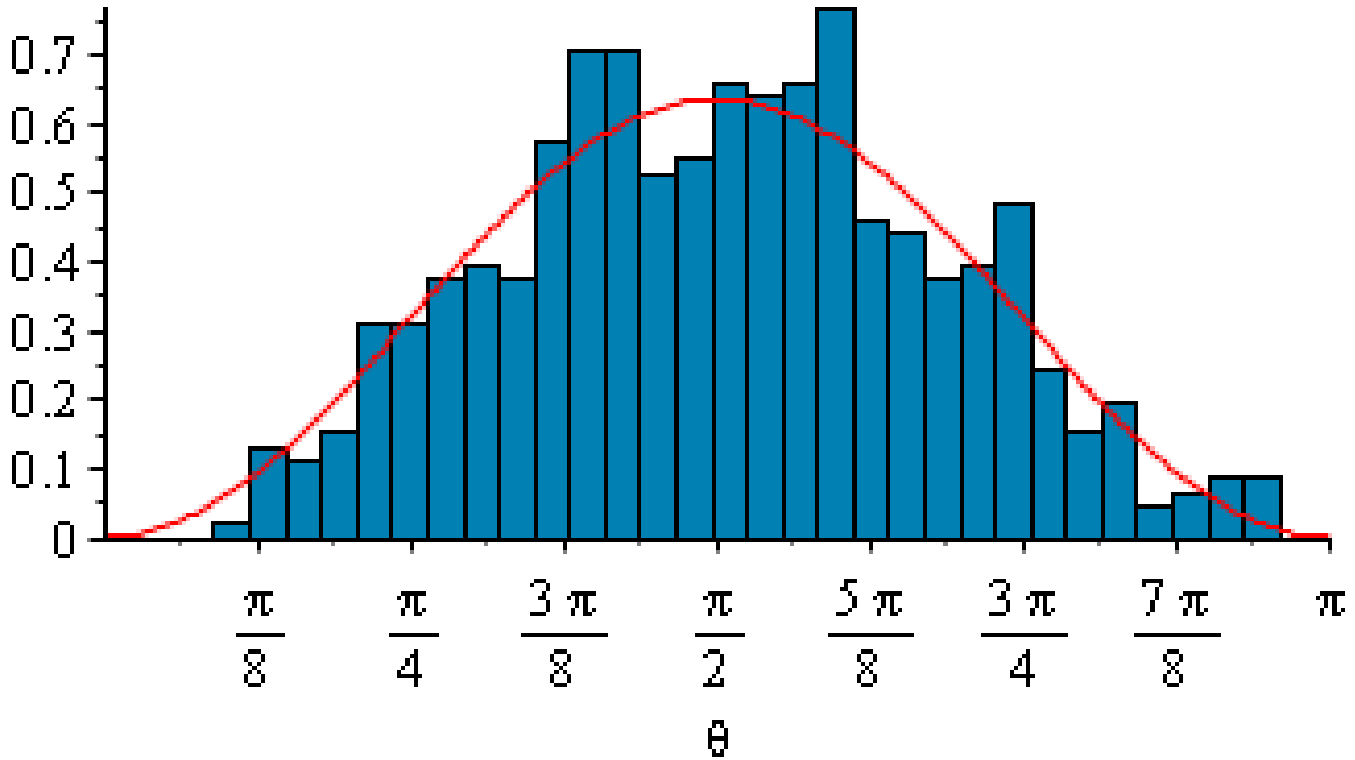
Prime	3	5	7	11	13	17
N_p	3	8	4	13	17	17
a_p	0	-3	3	-2	-4	0
b_p	0	-0.672	0.567	-0.302	-0.556	0
θ_p	1.571	2.308	0.968	1.878	2.160	1.571

If we use a computer to calculate more then we get the following pleasing histograms (the red curve is the graph of $y = \frac{2}{\pi}\sin^2(\theta)$ on $[0, \pi]$):

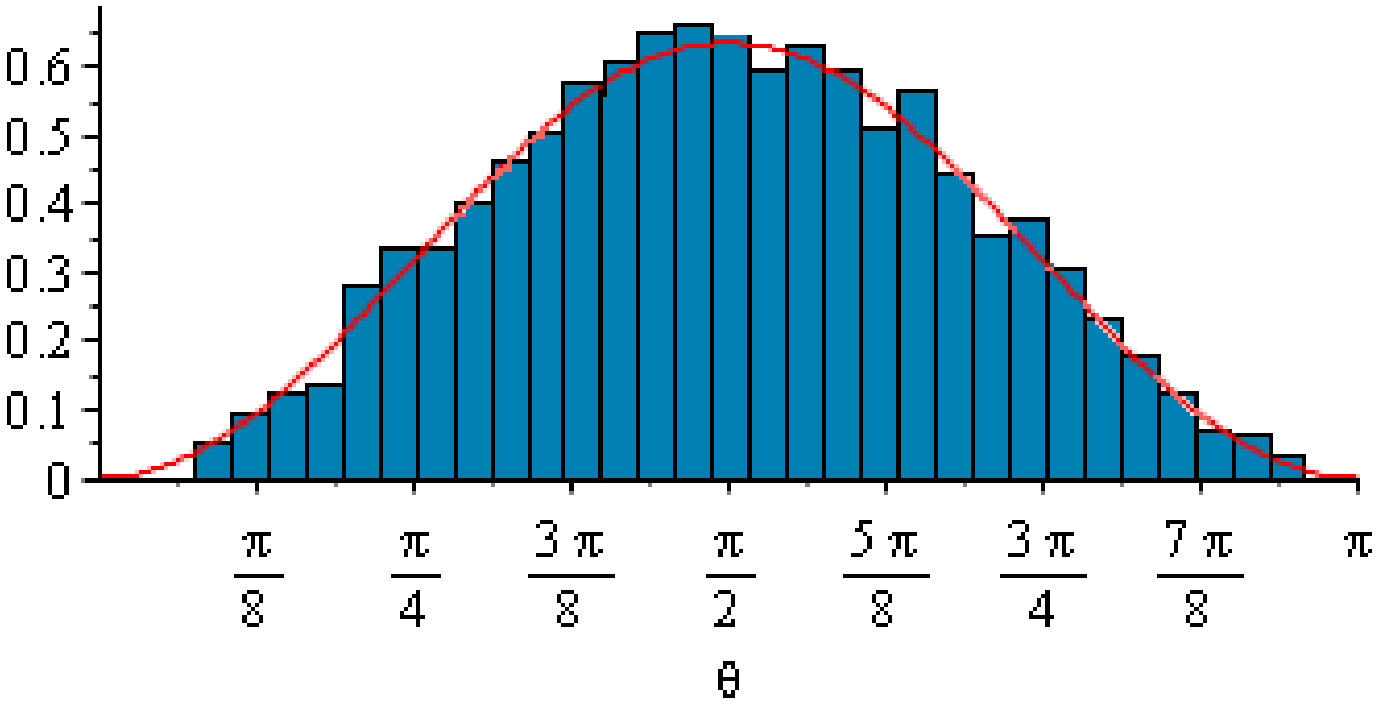
For the first 100 ”good” primes:



For the first 1000 ”good” primes:



For the first 10000 ”good” primes:



Already we can observe a quick convergence to the correct distribution!

Thanks for listening!