The Conjecture
of
Birch
and
Swinnerton-Dyer

(Overheads for a talk at Ohio State, 11/10/2005; they weren’t used due to technical difficulties)

Warren Sinnott
Diophantine problems

\[ x^2 + y^2 = z^2 \] (the “Pythagorean” equation)

\[ x^2 - N y^2 = 1 \] (“Pell’s” Equation)
\( N \) is a given integer, not a square.

\[ x^3 = y^2 + N \] (One of Fermat’s challenges to English mathematicians was to show that when \( N = 2 \) the only positive integer solution is \( x = 3, y = 5 \).)

\[ x^N + y^N = z^N \] (Fermat’s Last Theorem)
\( N \) is an integer > 2.

(from Diophantus:) If a rational number is the difference of two positive rational cubes then it is the sum of two positive rational cubes.

(from a 10th century Arabic mss.) Given a natural number \( N \), does there exist a right triangle with rational sides and area \( N \)? (“congruence number problem”)

Hilbert’s 10th problem

Find an algorithm to decide whether a polynomial equation \( f(x, y, z, \ldots) = 0 \) (with integer coefficients) has any integer solutions.

Matijasevič (following work of J. Robinson, M. Davis, and others) 1970: There is no such algorithm.

Still open: the “rational” form of Hilbert’s 10th problem: Find an algorithm to decide whether \( f(x, y, z, \ldots) = 0 \) has any rational solutions.

Even the following is still open:

Find an algorithm to decide, given integers \( a, b \), whether the equation \( y^2 = x^3 + ax + b \) has a solution in the rational numbers.
Consider the problem of finding the *rational* zeroes to (absolutely irreducible) polynomial equations in two variables (with integer or rational coefficients): \( f(x, y) = 0 \). Roughly, the problem gets harder as the degree of the \( f \) increases.

But the correct measure of the “difficulty” of solving \( f(x, y) = 0 \) is the *genus* of the equation.

Consider the set \( X(\mathbb{C}) = \{(x, y) \in \mathbb{C}^2 : f(x, y) = 0\} \) of complex solutions to \( f(x, y) = 0 \). If we complete \( X \) (add several points at \( \infty \)) and desingularize it, we get a compact Riemann surface \( \hat{X} \); topologically, \( \hat{X} \) is a compact oriented surface, and we let \( g \) be its genus (the number of holes...).
\begin{align*}
g = 0 & \quad g = 1 & \quad g = 2 \\
\end{align*}
A brief and incomplete outline of what is known about rational solutions to $f(x, y) = 0$:

$$g = 0$$

This is the case for example if $\text{deg } f = 1$ or $2$. The set of $X(\mathbb{Q})$ of rational solutions to $f(x, y) = 0$ is either empty or infinite. If $\text{deg } f = 1$, then there are infinitely many points (which form a “1-parameter family.”) When $\text{deg } f = 2$, it is a problem to decide whether $X(\mathbb{Q})$ is empty, and the problem is solved by the “Hasse principle:” $X(\mathbb{Q})$ is non-empty if and only if there are real solutions and $p$-adic solutions for each prime $p$, i.e.

$$X(\mathbb{Q}) \neq \emptyset \iff X(\mathbb{R}) \neq \emptyset \text{ and } X(\mathbb{Q}_p) \neq \emptyset \text{ for all primes } p$$

Moreover, as soon as we have one solution $(A, B) \in X(\mathbb{Q})$ we get infinitely many, and we can parametrize them by the rational points on a line:
\[ g = 1 \]

This case occurs, for example, if \( f(x, y) = y^2 - x^3 - ax - b \) and \( x^3 + ax + b \) has distinct roots. The set \( X(\mathbb{Q}) \) of rational solutions to \( f(x, y) = 0 \) can be finite (including possibly empty) or infinite. There are no algorithms at present to decide which. But if we allow “points at infinity” the set \( X(\mathbb{Q}) \), when nonempty, can be made into an abelian group. (E.g. for \( f(x, y) = y^2 - x^3 - ax - b \), there is one point at infinity, which serves as the identity for the group.)

\[ g > 1 \]

Remarkably little is known in general beyond one spectacular result, due to Faltings: \( X(\mathbb{Q}) \) is finite (including possibly empty). But we have no effective procedure for deciding whether \( X(\mathbb{Q}) \) is empty or for enumerating its elements if it is non-empty.
We consider the case \( g = 1 \) in more detail.

A curve \( E : y^2 = x^3 + ax + b \) is called an *elliptic curve* when \(-4a^3 - 27b^2 \neq 0\) (which guarantees that the roots of \( x^3 + ax + b \) are distinct). One can make the points of \( E \) with values in any field into a group: given points \( P \) and \( Q \) on \( E \), we can construct the line through \( P \) and \( Q \); this line will intersect \( E \) in a third point \( R \), and a group law on \( E \) is then determined by the condition \( P + Q + R = O \) (where \( O \) is the point at infinity and the identity of the group).

(When \( P = Q \), we use the tangent line to the curve at \( P \). There are other special cases to consider as well.)

A key feature of the situation: if \( P \) and \( Q \) have coordinates in a field \( F \), so does \( P + Q \). So \( E(\mathbb{Q}) \), \( E(\mathbb{R}) \), and \( E(\mathbb{C}) \) all become groups under this construction.
the point at infinity (O) is the identity
the lines through O are the vertical lines
reflection across the x-axis is negation
the group law is defined by "collinear points sum to O"
This picture yields the following formulas:

If \( P = (x_1, y_1) \), \( Q = (x_2, y_2) \) are distinct points and \( x_1 \neq x_2 \), then the coordinates \((x_3, y_3)\) of \( P + Q \) are

\[
\left( \left( \frac{y_2 - y_1}{x_2 - x_1} \right)^2 - x_1 - x_2, \left( \frac{y_2 - y_1}{x_2 - x_1} \right) x_3 - \frac{y_1 x_2 - y_2 x_1}{x_2 - x_1} \right)
\]

If \( P \) and \( Q \) are distinct points with \( x_1 = x_2 \), then \( P + Q = O \), the point at infinity.

If \( P = Q \), but \( y_1 = y_2 \neq 0 \), then the coordinates of \( P + Q = 2P \) are

\[
\left( \left( \frac{3x_1^2 + a}{2y_1} \right)^2 - 2x_1, \left( \frac{3x_1^2 + a}{2y_1} \right) x_3 + y_1 - \left( \frac{3x_1^2 + a}{2y_1} \right) x_1 \right)
\]

Finally, if \( P = Q \) and \( y_1 = y_2 = 0 \), then \( P + Q = 2P = O \).
The picture shows $E(\mathbb{R})$. $E(\mathbb{C})$ is a torus: we have
$$\mathbb{C}/L \cong E(\mathbb{C})$$
for some lattice $L \subseteq \mathbb{C}$, the isomorphism being given by
the Weierstrass $\wp$-function for $L$.

We want to understand $E(\mathbb{Q})$.

Theorem (Mordell, 1922) $E(\mathbb{Q})$ is a finitely generated
abelian group.

So $E(\mathbb{Q}) \cong T \oplus \mathbb{Z}^r$, where $T$ is finite abelian. $T$ is easy
to compute:

Theorem (Lutz, Nagell c. 1935) If $(x, y)$ is a torsion
point, then $x$ and $y$ are integers and either $y = 0$ or
$y^2 \mid 4a^3 + 27b^2$.

What about $r$? ($r$ called the rank of $E$) We can compute
an upper bound for $r$ but there’s no known bound for the
heights of the generators of $E(\mathbb{Q})$. (So unless the upper
bound is the rank, we don’t know when to stop looking.)

If $P = (x, y) \in E(\mathbb{Q})$, the height of $P$, denoted $H(P)$,
is the maximum size of the numerator and denominator
of $x$ and $y$. 
How can we determine $r$?

Digression (?):

Consider $E(\mathbb{F}_p) = \{(x, y) \in \mathbb{F}_p^2 : y^2 \equiv x^3 + ax + b \mod p\} \cup \{O\}$. If $p \neq 2$ and $p \nmid 4a^3 + 27b^2$, then the formulas above make $E(\mathbb{F}_p)$ into a group. What is $N_p = \#E(\mathbb{F}_p)$?

For each $x = 0, 1, \ldots, p - 1$ we get

- no points if $x^3 + ax + b$ is not a square mod $p$
- one point if $x^3 + ax + b \equiv 0 \mod p$
- two points if $x^3 + ax + b$ is a nonzero square mod $p$

plus one for the point at infinity.

Since a randomly chosen nonzero element of $\mathbb{F}_p$ is equally as likely to be a square as a non-square, the first and third possibilities might tend to be equally likely, which suggests that $N_p = \#E(\mathbb{F}_p)$ should be about $p + 1$. In fact,

Theorem (Hasse, 1934) $|p + 1 - N_p| \leq 2\sqrt{p}$. (For $p > 2, p \nmid 4a^3 + 27b^2$.)
In the late 1950s, Birch and Swinnerton-Dyer had the happy thought (suggested by work of Siegel on quadratic forms in the 1930s) that if $r = \text{rank} E(\mathbb{Q})$ is large (> 0) then we should get more points in $E(\mathbb{F}_p)$ than expected. (There is a “reduction map” $E(\mathbb{Q}) \rightarrow E(\mathbb{F}_p)$).

Or maybe, if there are more points in lots of $E(\mathbb{F}_p)$’s than there should be, we have a better chance of being able to “piece them together” into a rational point on $E$.

In any case, they tried calculating

$$\pi_E(x) = \prod_{p \leq x} \frac{N_p}{p}.$$ 

for various elliptic curves $E$, on the idea (hope?) that this would grow more rapidly when $r = r_E$ is positive. Here are the results for some curves of the form $E_d : y^2 = x^3 - d^2 x$: 
\[ \pi_E(x) = \prod_{p \leq x} \frac{N_p}{p} \]

Figure 2. Birch and Swinnerton-Dyer data for \( y^2 = x^3 - d^2x \)

(Source: Rubin, Silverberg: Ranks of elliptic curves, BAMS 39 4 2002)
This leads to the conjecture that $\log \pi_E(x)$ grows like $r_E \log \log x$:

Birch Swinnerton-Dyer Conjecture (First form):

For any elliptic curve defined over $\mathbb{Q}$, 

$$\pi_E(x) \sim C_E (\log x)^{r_E},$$

for some constant $C_E$, with $r_E$ the rank of $E(\mathbb{Q})$. 
Another digression (?): zeta and $L$-functions:

The Riemann zeta function $\zeta(s)$ has a number of striking properties — an expression as a product (“Euler product”) over the primes, analytic continuation to $\mathbb{C}$ (except for a simple pole at $s = 1$), a functional equation relating $\zeta(s)$ and $\zeta(1 - s)$.

The Euler product has the form

$$\zeta(s) = \prod_p \left(1 - \frac{1}{p^s}\right)^{-1}$$

$$= \prod_{\text{maximal ideals } P \text{ of } \mathbb{Z}} \left(1 - \frac{1}{[\mathbb{Z} : P]^s}\right)^{-1}$$

If $k$ is an number field, then we can define analogously

$$\zeta_k(s) = \prod_{\text{maximal ideals } \mathfrak{p} \text{ of } \mathcal{O}} \left(1 - \frac{1}{[\mathcal{O} : \mathfrak{p}]^s}\right)^{-1}$$

—which has the same striking properties.
Quite generally, if $A$ is any ring of finite type over $\mathbb{Z}$ (i.e. $A = \mathbb{Z}[a_1, \ldots, a_n]$ for some $a_i \in A$ and with $\mathbb{Z}$ the image of $\mathbb{Z}$ in $A$), then $A/P$ is a finite field for any maximal ideal of $A$, so that we could define

$$
\zeta(A, s) = \prod_{\text{maximal ideals } P \text{ of } A} \left(1 - \frac{1}{[A : P]^s}\right)^{-1}
$$

and ask about its properties.

E.g. take $A = \mathbb{F}_p[x]$: 

$$
\zeta(\mathbb{F}_p[x], s) = \prod_{\text{monic irreducibles } \pi(x)} \left(1 - \frac{1}{p^{(\deg \pi)s}}\right)^{-1}
= \sum_{\text{monic polynomials } m(x)} \frac{1}{p^{(\deg m)s}}
= \sum_{n=0}^{\infty} p^n \frac{1}{p^{ns}}
= \frac{1}{1 - \frac{1}{p^{s-1}}}.
$$

and therefore (now taking $A = \mathbb{Z}[x]$)

$$
\zeta(\mathbb{Z}[x], s) = (!) \prod_p \zeta(\mathbb{F}_p[x], s) = \zeta(s - 1)
$$
If we take $A = \mathbb{F}_p[x, y]/(y^2 - x^3 - ax - b)$, we get a “zeta function” attached to the elliptic curve $E \mod p$. In the 1930s, Hasse showed that

$$
\zeta(E/\mathbb{F}_p, s) = \frac{1 - a_p x + px^2}{(1 - x)(1 - px)}
$$

where $x = p^{-s}$, and $a_p = p + 1 - N_p$. (This is not exactly $\zeta_A(s)$, but takes into account the point at $\infty$.)

(Note that the zeroes of $\zeta(E/\mathbb{F}_p, s)$ occur where $p^{-s}$ is a root of $1 - a_p x + px^2$. If you use Hasse’s estimate $|a_p| \leq 2\sqrt{p}$, you find that the zeroes of $\zeta(E/\mathbb{F}_p, s)$ occur on the line $\Re(s) = 1/2$.)

Hasse suggested multiplying these $\zeta(E/\mathbb{F}_p, s)$ together to get

$$
\zeta(E/\mathbb{Q}, s) = \prod_p^* \zeta(E/\mathbb{F}_p, s)
$$

$$
= \zeta(s)\zeta(s-1) \prod_p 1 - a_p p^{-s} + p^{1-2s}
$$

(the “*” means that things need to be adjusted at the finite number of primes $p$ where $p = 2$ or $p \mid 4a^3 + 27b^2$)
(Note that this function is essentially \(\zeta(A, s)\), where now \(A = \mathbb{Z}[x, y]/(y^2 - x^3 - ax - b)\).

The function

\[
L(E/\mathbb{Q}, s) = \prod_p^*(1 - a_p p^{-s} + p^{1-2s})^{-1}
\]

is called the Hasse-Weil \(L\)-function of \(E\). It only converges for \(\Re(s) > 3/2\), but if we formally set \(s = 1\) we find

\[
L(E/\mathbb{Q}, 1) = \prod_p^* \frac{p}{N_p},
\]

since \(N_p = p + 1 - a_p\). This suggests that \(L(E/\mathbb{Q}, 1)\) should vanish if \(r_E > 0\) and perhaps should vanish to order \(r_E\). This is the second form of the Birch Swinnerton-Dyer Conjecture:
Birch Swinnerton-Dyer Conjecture (Second Form):

For any elliptic curve defined over $\mathbb{Q}$,

$$\text{ord}_{s=1} L(E/\mathbb{Q}, s) = r_E$$

with $r_E$ the rank of $E(\mathbb{Q})$.

Note that this presumes that $L(E/\mathbb{Q}, s)$ can be analytically continued at least to $s = 1$; it is now known that $L(E/\mathbb{Q}, s)$ can be analytically continued to the entire complex plane, for all elliptic curves defined over $\mathbb{Q}$, by work of Wiles, Taylor, Breuil, Conrad, and Diamond.
Here’s a heuristic argument that relates the two forms, and “explains” the growth rate \((\log x)^r\): the usual zeta function has a simple pole at \(s = 1\); and standard arguments allow one to deduce from this that
\[
\prod_{p \leq x} \left(1 - \frac{1}{p}\right)^{-1} \approx \log x
\]
and therefore
\[
\prod_{p \leq x} \left(1 - \frac{1}{p}\right)^r \approx \frac{1}{(\log x)^r},
\]
which arises from \(1/\zeta(s)^r\), which has a zero of order \(r\) at \(s = 1\). By analogy one might expect
\[
\prod_{p \leq x}^* \frac{p}{N_p} \approx \frac{1}{(\log x)^r},
\]
if \(L(E, s)\) has a zero of order \(r\) at \(s = 1\).

(the “\(\approx\)” above means the ratio tends to a nonzero constant.)
What’s known?

- If \( \text{ord}_{s=1} L(E/\mathbb{Q}, s) = 0 \) or 1, then the second form of the conjecture is valid. (Gross, Zagier, and Kolyvagin)
  
  So, for example, if \( L(E/\mathbb{Q}, 1) \neq 0 \), then the only rational solutions to the equation \( y^2 = x^3 + ax + b \) correspond to torsion points and can therefore be determined by the Lutz/Nagell theorem.

  And if \( L(E/\mathbb{Q}, 1) = 0 \) but \( L'(E/\mathbb{Q}, 1) \neq 0 \), then there is a rational solution \( P = (x_0, y_0) \) to \( y^2 = x^3 + ax + b \) such that every solution is a multiple of \( P \) plus a torsion point ("multiple" and "plus" in the sense of the group law on \( E \)).

- The first form implies the second. (Dorian Golfd)