$\zeta(3)$ APPARENTLY IRRATIONAL

JARED WEINSTEIN

Abstract. In June 1978, R. Apéry gave a lecture at the "Journées Arithmétiques" conference at Marseille-Luminy in which he claimed that the real number

$$\zeta(3) = 1 + \frac{1}{2^3} + \frac{1}{3^3} + \dots$$

is irrational. He made skeptics out of many of the attendants by making one shockingly unbelievable assertion after another. First, it was put forth that the recurrence

$$n^3 u_n + (n-1)^3 u_{n-2} = (34n^3 - 51n^2 + 27n - 5)u_{n-1}$$

has one solution $b_n = 1, 5, 73, 1445, \dots$ in *integers*, and another $a_n = 0, 6, 351/4, \dots$ in rationals whose denominator grows very slowly. Second, Apéry claimed that $a_n/b_n \to \zeta(3)$ rapidly enough to prove that $\zeta(3) \notin \mathbf{Q}$. Later, Zagier and Cohen were able to clarify the proof, which was indeed very clever, but not very illuminating. We will present a different proof by F. Beukers, involving modular forms for $\Gamma_1(6)$, which is a bit more illuminating. We will also give an explanation of the first assertion in terms of Picard-Fuchs differential equations associated to families of elliptic curves. These are notes for a talk for the Many Cheerful Facts series at Berkeley given Feb. 25, 2004.

Consider the recursion

$$n^3 u_n + (n-1)^3 u_{n-2} = (34n^3 - 51n^2 + 27n - 5)u_{n-1}.$$

Theorem 1 (Apèry). *Let a_n and b_n be the sequences of rational numbers satisfying the above recursion with initial conditions $a_0 = 0$, $a_1 = 6$, $b_0 = 1$, $b_1 = 5$. Then for all n, $b_n \in \mathbf{Z}$ and $2[1,2,\dots,n]a_n \in \mathbf{Z}$. Furthermore, $a_n/b_n \to \zeta(3)$*

Remark 1. We expect the denominators of u_n to grow like $n!^3$, much faster than exponential, whereas $[1,2,\dots,n]^3 = O(e^{3n})$. Explanation:

$$[1,2,\dots,n] = \prod_{p \le n} p^{\frac{\log n}{\log p}} < \prod_{p \le n} n.$$

Now apply the prime number theorem: $\pi(x) \sim x/\log x$.

Apèry proved the theorem by giving completely explicit expressions for the sequences a_n and b_n. The expressions made the assertion $a_n/b_n \to \zeta(3)$ very easy, but the fact that the expressions satisfied the recurrences proved extremely complicated.

Corollary 1. *(See* [Po]*.) $\zeta(3)$ is irrational.*

Proof. It's easy to see from the recursion that

$$n^3(a_n b_{n-1} - a_{n_1} b_n) = (n-1)^3(a_{n-1}b_{n-2} - a_{n-2}b_{n-1}),$$

It follows very quickly from this (using the initial values) that

$$a_n b_{n-1} - a_{n-1} b_n = \frac{6}{n^3}.$$

Then

$$\begin{aligned}
\zeta(3) - \frac{a_n}{b_n} &= \lim_{N\to\infty} \zeta(3) - \frac{a_n}{b_n} - \left(\zeta(3) - \frac{a_N}{b_N}\right) \\
&= \sum_{k=n}^{\infty} \frac{a_{n+1}}{b_{n+1}} - \frac{a_n}{b_n} \\
&= \sum_{k=n}^{\infty} \frac{6}{k^3 b_k b_{k+1}} \\
\left|\zeta(3) - \frac{a_n}{b_n}\right| &= O(b_n^{-2})
\end{aligned}$$

To estimate the growth of b_n, we divide the recursion above by n^3 and ignore all terms with $1/n$. We find that the $b_n = O(\alpha^n)$, where α is the larger root of $T^2 - 34T + 1$, that is, $\alpha = (1+\sqrt{2})^4$. Now suppose $\zeta(3)$ is rational with denominator D. Then

$$D[1,\ldots,n]^3 \left|a_n - b_n\zeta(3)\right| = O(e^{3n}(1+\sqrt{2})^{-4n}).$$

On the one hand, the LHS is an integer, and on the other, the RHS goes to zero because of the happy accident that

$$e^3 < (1+\sqrt{2})^4.$$

Thus the LHS is actually zero for large enough n, which is in contradiction with (say) the calculation regarding the successive differences between the a_n/b_n. □

The remainder of this paper will attempt to answer the question: why should this recurrence produce (nearly) integral sequences? The answer should in a sense come from geometry. (For a better sense of what we mean by this, and a great deal of interesting speculation besides, see [KZ] and the sources cited therein. For an in-depth analysis of a similar situation involving $\zeta(2)$, see [Be2].)

Our first observation is that if we define the generating function $F(q) = \sum_{n\geq 0} u_n q^n$, then the recursion above is equivalent to the degree 3 differential equation $LF = u_1 - 5u_0$, where L is the differential operator

$$L = t^2(t^2 - 34t + 1)\left(\frac{d}{dt}\right)^3 + t(6t^2 - 153t + 3)\left(\frac{d}{dt}\right)^2 + (7t^2 - 112t + 1)\left(\frac{d}{dt}\right) + (t - 5).$$

For our purposes, "coming from geometry" means that solutions to this equation can be whipped up using modular forms. The group $\mathrm{PSL}_2(\mathbf{R})$ acts on the upper half plane $\mathcal{H}$ via

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} z = \frac{az+b}{cz+d}$$

Let Γ be a subgroup of $\mathrm{PSL}_2(\mathbf{Z})$ of finite index.

Definition 1. A *modular form* for Γ of weight k is a holomorphic function f on $\mathcal{H}$ satisfying

$$f(\gamma z) = (cz+d)^k f(z)$$

for all $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$. We also require that f "has moderate growth at the cusps."

Therefore a modular form of weight 0 is simply a holomorphic function t on the quotient space $\Gamma\backslash\mathcal{H}$. This space can be compactified in a canonical way to obtain a compact Riemann surface $X(\Gamma)$; f then becomes a meromorphic function on $X(\Gamma)$ holomorphic away from the cusps.

Modular forms of weight 2, on the other hand, correspond via $f \mapsto f(z)dz$ to 1-forms on $X(\Gamma)$ which are holomorphic away from the cusps. We make an important observation about these forms. Say $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$. Then γ acts on the three-dimensional vector space spanned by the functions $f_2(z) = z^2 f(z), f_1(z) = zf(z), f(z)$ by the matrix

$$\mathrm{Sym}^2(\gamma) = \begin{pmatrix} a^2 & 2ab & b^2 \\ ac & ad+bc & bd \\ c^2 & 2cd & d^2 \end{pmatrix}.$$

It will be immensely useful to consider f (resp. f_1, f_2) locally as a function of t. That is, to calculate $f(t)$, draw a path p from 0 to t in $X(\Gamma)$, lift the path to $\mathcal{H}$, then evaluate f on the endpoint $p(1)$ of the lifted path. The function $f(t)$ is multivalued: it may depend on the choice of path and we can not deform around branch points of t. If $\gamma \in \Gamma$ then we can consider the *monodromy* action of γ on a branch $f(t)$ of the multivalued function; instead of evaluating f on $p(1)$ we evaluate on $\gamma(p(1))$. Again, the monodromy action of γ on the vector of functions (f_2, f_1, f_0) is as $\mathrm{Sym}^2(\gamma)$. Now note that the action of monodromy commutes with differentiation with respect to t:

$$\frac{d}{dt} f(\gamma z) = \left(\frac{d}{dt} f\right)(\gamma z),$$

so that the same is true of the vector $f_2^{(k)}, f_1^{(k)}, f^{(k)}$ consisting of k-iterated t-derivatives of these three functions.

It is a tautology from linear algebra that

$$\begin{vmatrix} f & f' & f'' \\ f_1 & f_1' & f_1'' \\ f_2 & f_2' & f_2'' \end{vmatrix} f''' - \cdots - \begin{vmatrix} f' & f'' & f''' \\ f_1' & f_1'' & f_1''' \\ f_2' & f_2'' & f_2''' \end{vmatrix} = 0.$$

A priori this "differential equation" isn't that interesting because the coefficients are given in terms of f. But let's see what happens when we apply the action of monodromy to these coefficients. By the above discussion, we see that acting by γ has the same effect as mutliplying by the determinant of the matrix $\mathrm{Sym}^2(\gamma)$, which is 1! Therefore these coefficients are Γ-invariant and are therefore algebraic functions of t. This equation is called a Picard-Fuchs differential equation associated to the mapping $\mathcal{H} \to X(\Gamma)$.

It is not hard to see[1] that the coefficient of f''' is $2f(t)^3/(dt/dz)^3$.

It is time to specialize this picture for the task at hand. The following argument is adapted from [Be1]. Let Γ be the group $\Gamma_0(6)$ consisting of those matrices whose southwest entry is divisible by 6. Here, a fundamental domain for the action of Γ on $\mathcal{H}$ is given below:

[1] Indeed, if F is an $(n-1)$ times differentiable function of z, then the $n \times n$ matrix whose ij-th entry is $(d/dz)^i(z^j f(z))$ has determinant $1!2!\dots(n-1)!f^n$.

It is now easily seen that $X_0(6)$ (the Riemann surface obtained by completing $\Gamma\backslash\mathcal{H}$ is a curve of genus 0. For our function t we will take

$$t(z) = \left(\frac{\eta(z)\eta(6z)}{\eta(2z)\eta(3z)}\right)^1 2$$

.

Here $\eta(z) = q^{1/24}\prod_{n\geq 1}(1-q^n)$ is the Dedekind η-function and $q = e^{2\pi i z}$. The precise definition of t is unimportant; the only facts we need about t are as follows:

(1) $t(z)$ is invariant under $z \mapsto -1/6z$,

(2) $t\colon X_0(6) \to \mathbf{P}^1$ is a double cover of $\mathbf{P}^1$ which is branched[2] over $t(i/\sqrt(6) = (1-\sqrt{2})^4$ and $t(2/5 + i/5\sqrt{6}) = (1+\sqrt{2})^4$,

(3) The values of t at the cusps are $t(i\infty) = 0$ and $t(1/2) = \infty$,

(4) As $t(z+1) = t(z)$, t can be expanded as a Fourier series

$$t = q - 12q^2 + 66q^3 + \dots$$

with integer coefficients.

Our weight 2 form f will be

$$f(z) = \frac{(\eta(2z)\eta(3z))^7}{(\eta(z)\eta(6z))^5} = 1 + 5q + 13q^2 + \dots.$$

We need the fact that $z \mapsto -1/6z$ transforms $f(z)$ into $-6f(z)$. By considering $f(z)dz$ we see that f has no zeroes on $\mathcal{H}$.

The Picard-Fuchs equation for this $f(t)$ has $2f^3/(dt/dz)^3$ as its coefficient of f'''; since this is a holomorphic function invariant under $z \mapsto -1/6z$, it is a rational function of t, determined up to a scalar by its zeroes and poles. This function has no zeroes save a triple zero at $t = \infty$ and poles at the branch points of t, which are at $t = 0$ and $t = (1 \pm \sqrt{2})^4$. We find that (up to a constant), the coefficient is (I hope!) $t^{-3}(t^2 - 34t + 1)^{-3}$. Computing the other coefficients and cancelling denominators should give the equation $LF = 0$ as above.

The punchline, of course, is that in the series development

$$f(t) = \sum_{n\geq 0} b_n t^n,$$

the coefficients b_n match the sequence generated by the recursion, and furthermore they are visibly integers. This is half of Apéry's claim.

We will only sketch the other half, which is the half that concerns $\zeta(3)$.We select a weight 4 modular form G for Γ:

$$G(z) = E_4(z) - 36E_4(36z) - 7(E_4(2z) - 9E_4(z))$$

where $E_4(z) = \sum'_{m,n}(mz+n)^{-4}$ (up to a constant). Say $G(z) = \sum_{n\geq 1}\alpha_n q^n$ with the usual notation. We find that the triple antiderivative $g(z) = \sum_{n\geq 1}\alpha_n q^n/n^3$ of $G(z)$ has the essential property that

$$g(-1/6z) - \zeta(3) = (-1/6z)(g(z) - \zeta(3))$$

[2] The reason t takes algebraic values at these special arguments is a consequence of the theory of complex multiplication.

and therefore that the product

$$f(z)(g(z) - \zeta(3)) = \sum_{n \geq 0} (a_n - b_n \zeta(3)) t^n$$

is invariant under $z \mapsto -1/6z$. This means exactly that the monodromy around the branch point $t = (1 - \sqrt{2})^4$ leaves this product function unchanged. Therefore as a function of t the product can be extended to the region $|t| < (1 + sqrt2)^4$, which is the next branch point out. Thus

$$|a_n - b_n \zeta(3)| < O((1 + \sqrt{2})^{-4n}).$$

Inspection reveals that the a_n have denominators dividing $[1, 2, \ldots, n]^3$. This is sufficient to show that $\zeta(3) \notin \mathbf{Q}$.

References

[Be1] Beukers, F. *Irrationality Proofs using modular forms.* Journées aritmetíques de Besan con (Besan con, 1985). Astérisque No. 147-148 (1987), 271-283.

[Be2] Beukers, F. *Irrationality of π^2, periods of an elliptic curve and* $\Gamma_1(5)$. Diophantine approximations and transcendental numbers. (Luminy, 1982), 47–66, Progr. Math., 31, Birkhäuser, Boston, Mass., 1983.

[Po] van der Poorten, A. *A proof that Euler missed… Apéry's proof of the irrationality of* $\zeta(3)$. Math. Intelligencer 1 (1978/79), no. 4, 195-203.

[KZ] Kontsevich, M. and Zagier, D. *Periods.* Mathematics unlimited–2001 and beyond, 771-808, Springer, Berlin, 2001.