e is Transcendental, and mod-n-town

Ivan Aidun

This note was motivated by my attempt to understand this note by Robert Hines describing Hermite's proof of the transcendence of e, and my understanding of the specifics of Hermite's proof come entirely from that note. My exposition about mod-n-town is my own, but I first learned about the problem in a combinatorics class at Budapest Semesters in Mathematics taught by András Gyárfás.

Hermite's proof that e is transcendental

Following Liouville, the strategy to show that a number is irrational or transcendental is to show that it is approximated "too well" by rational numbers. Liouville produced the example number

$$C = \sum_{n=1}^{\infty} 10^{-n!},$$

which is clearly *very* well-approximated by its partial sums. In fact, Liouville's constant is in a sense "maximally transcendental", in the sense that it has irrationality measure ∞. (I'm not going to define irrationality measure, but suffice it to say that rational numbers have irrationality measure 1, algebraic numbers have irrationality measure 2 by Roth's theorem, and transcendental numbers have irrationality measure ≥ 2.)

It is known that the irrationality measure of e is 2 (Borwein and Borwein 1987), so in a sense e is no better approximated by rational numbers than an algebraic number would be. Altho Hermite did not know this, we can take this as evidence that we cannot prove e is transcendental as straightforwardly as we could for Liouville's constant. The idea instead will be to show that, for any given n, we can uniformly approximate $e, e^2, \ldots, e^n$ very well. (We will choose n to be the degree of a putative polynomial having e as a root.) This, plus a little bit of trickery, will give us our desired result.

We will begin by outlining the proof, and then we will go back and fill in the details of the auxiliary constructions (as one must imagine Hermite would have done the first time around).

Outline of Proof

We must get good approximations for the powers of e from somewhere. The most natural place to look would be to use the Taylor series expansion for e^x, but we will actually turn elsewhere. As we will see later, it will be somewhat important to our proof that we have good control over the

prime factors appearing in our approximations, so the factorials appearing in the Taylor series are not very good in this regard.

Instead, observe that for a polynomial $f(x)$, by repeatedly integrating by parts we obtain the identity

$$\int_0^x f(t)e^{-t}\,dt = \sum_{N=0}^{\infty}(f^{(N)}(0) - e^{-x}f^{(N)}(x)).$$

Inspired by this identity, we will define

$$\widetilde{f}(x) = \sum_{N=0}^{\infty} f^{(N)}(x)$$

so that the above identity becomes (after multiplying through by e^x)

$$e^x \int_0^x f(t)e^{-t}\,dt = e^x\widetilde{f}(0) - \widetilde{f}(x).$$

Thus, if we can cleverly choose a polynomial $f(x) \in \mathbb{Q}[x]$ so that the integral is close to 0, then we will get that $e^n \approx \dfrac{\widetilde{f}(n)}{\widetilde{f}(0)}$. (Note: is there another, more enlightening way to view what information $\widetilde{f}$ captures about f? Can I view it as some kind of integral transform with certain nice properties?)

Once we choose a good enough polynomial f, the proof will go like this: suppose that e is the root of a polynomial of degree n, $g(x) = \sum_{k=0}^{n} c_k x^k$, where the $c_k \in \mathbb{Z}$. Then we will have

$$\begin{aligned}
\left|\sum_{k=0}^{n} c_k\widetilde{f}(k)\right| &= \left|\sum_{k=0}^{n} c_k\widetilde{f}(k) - \left(\sum_{k=0}^{n} c_k e^k\right)\widetilde{f}(0)\right| && \text{(since } e \text{ is a root of } g\text{)} \\
&= \left|\sum_{k=1}^{n} c_k(\widetilde{f}(k) - e^k\widetilde{f}(0))\right| \\
&\leq C\sum_{k=1}^{n}\Big|(\widetilde{f}(k) - e^k\widetilde{f}(0))\Big|,
\end{aligned}$$

where $C = \max|c_k|$. If we are able to choose our f so that (1) $\left|\widetilde{f}(k) - e^k\widetilde{f}(0))\right|$ is uniformly small for all $1 \leq k \leq n$, the right hand side can be made smaller than 1. If we can also make it so that (2) $\sum_{k=0}^{n} c_k\widetilde{f}(k)$ is a nonzero integer, we will arrive at a contradiction, and so conclude that e is transcendental.

Choosing a good polynomial

The more difficult part of the requirements we have placed on f is requirement (2), that $\sum_{k=0}^{n} c_k \widetilde{f}(k)$ is a nonzero integer. Making it an integer is not so hard, we can just make sure that $\widetilde{f}(k)$ is an integer, the difficult part is finding a way to guarantee that it is not 0. The method I would think of first would be to show that $\left|\sum_{k=0}^{n} c_k \widetilde{f}(k)\right|$ is bounded away from 0, but this is a non-starter because our argument above crucially depends on this getting close to 0. The only other way I know to show that an integer X is not 0 is to show that there exists a prime number p so that $p \nmid X$. (You can interpret this as showing that the p-adic absolute value $\left|\sum_{k=0}^{n} c_k \widetilde{f}(k)\right|_p$ is bounded away from 0.)

How exactly can we control $\widetilde{f}(k)$? Well, whatever polynomial $f(x)$ we choose will factor into linear terms over $\mathbb{C}$, so we should think about how a factored polynomial interacts with the twiddle transform. If we set $f(x) = c(x-k)^{m_k} h(x)$, where c is a constant and $h(x)$ does not vanish at k, then by the product rule

$$f^{(N)}(x) = \sum_{i=0}^{N} \binom{N}{i} \, c \cdot (m_k)^{\underline{i}} \, (x-k)^{m_k - i} \, h^{(N-i)}(x),$$

where $a^{\underline{b}} = a(a-1)\dots(a-b+1)$ is the "falling factorial" (or Pochhammer symbol). In particular, the above expression shows that

$$f^{(N)}(k) = \begin{cases} 0 & N < m_k \\ \binom{N}{m_k} \, c \cdot (m_k!) \, h^{(N-m_k)}(k) & N \geq m_k. \end{cases}$$

Controlling the prime factors of $h^{(N-m_k)}(k)$ would be difficult, but we have lots of ability to control the prime factors of $\binom{N}{m_k}$ and of $m_k!$ through judicious choice of m_k and of c.

Not wanting to belabor the point overmuch, we will choose a large prime number p, and choose the specific polynomial

$$f(x) = \frac{x^{p-1}(x-1)^p(x-2)^p \dots (x-n)^p}{(p-1)!}.$$

Hopefully the above discussion has instilled enough intuition so that, after enough thought and experimentation, you might have been able to arrive at a polynomial that looked something like this. The crucial features of this polynomial are that the powers and the denominator have been chosen so that $\widetilde{f}(0)$ is not divisible by p, but $\widetilde{f}(k)$ is divisible by p for $1 \leq k \leq n$. Moreover, tho not entirely obvious, when n is fixed and $0 \leq x \leq n$, the denominator $(p-1)!$ grows more quickly with p than the numerator does, which makes this a suitable choice for getting good approximations to e^k. (And since we have infinitely many primes p to choose from, we can make a choice large

enough to obtain whatever bounds we need to make our argument work.)

(Humorous aside: while reading Hines' note, when he points out that the proof relies on the existence of infinitely many primes, my very first thought was "ah yes, the weak form of Dirichlet's theorem".)

Proposition. Let n be fixed. For the polynomial f above, $\widetilde{f}(x)$ has the following properties.

(i) For $0 \leq x \leq n$, as $p \to \infty$, $e^x \widetilde{f}(0) - \widetilde{f}(x) \to 0$ uniformly in x.

(ii) The value $\widetilde{f}(0)$ is an integer, and if $p > n$ it is not divisible by p.

(iii) The value $\widetilde{f}(k)$ is an integer divisible by p for $1 \leq k \leq n$.

Proof. (i) For $0 \leq x \leq n$, we have

$$\begin{aligned}
\left| e^x \widetilde{f}(0) - \widetilde{f}(x) \right| &= e^x \left| \int_0^x f(t) e^{-t} \, dt \right| \\
&\leq e^x \left(x \sup_{[0,x]} \{ f(t) e^{-t} \} \right) \\
&\leq x e^x \sup_{[0,x]} \{ f(t) \} \\
&\leq n e^n \sup_{[0,n]} \{ f(t) \} \\
&\leq n e^n \frac{n^{p-1} (n^p)^n}{(p-1)!} \\
&= e^n \frac{(n^{n+1})^p}{(p-1)!}.
\end{aligned}$$

Taking logs, we have that the log of the numerator asymptotically grows like constant $\cdot\, p$, while the log of the denominator asymptotically grows like $p \log p$ (which I will give a short proof of in an appendix if I have time). Thus, the whole right hand side goes to 0 as p gets large, in a manner independent of x. (Sidenote: we have pretty grotesquely overestimated this quantity, probably by a factor of around $e^n (n/(2e))^{pn}$. However, being more careful about the estimate would only allows us to pick a slightly smaller p, it does not affect the asyptotic behavior of the expression.)

(ii) Using our expression for $f^{(N)}(0)$ from above, with $c = 1/(p-1)!$ and $h(x)$ being the factors

in our chosen $f(x)$ other than x^{p-1}, we have that

$$\begin{aligned}
\widetilde{f}(0) &= \sum_{N=0}^{\infty} f^{(N)}(0) \\
&= \sum_{N=p-1}^{\infty} \binom{N}{p-1} h^{(N-(p-1))}(0) \\
&= \sum_{d=0}^{\infty} \binom{d+p-1}{p-1} h^{(d)}(0) \qquad \text{(setting } d = N-(p-1)\text{)} \\
&= \underbrace{(-1)^{np}(n!)^p}_{d=0 \text{ term}} + \text{terms divisible by } p.
\end{aligned}$$

This is clearly an integer, and if $p > n$ it is not divisible by p. (If I have time, in an appendix I will discuss divisibility properties of binomial coefficients.)

(iii) Using again our expression for $f^{(N)}(k)$, this time with $h(x)$ being the terms other than $(x-k)^p$, we have that

$$\begin{aligned}
\widetilde{f}(k) &= \sum_{N=0}^{\infty} f^{(N)}(k) \\
&= \sum_{N=p}^{\infty} \binom{N}{p} \frac{p!}{(p-1)!} \cdot h^{(N-(p-1))}(0) \\
&= \sum_{N=p}^{\infty} \binom{N}{p} p \cdot h^{(N-(p-1))}(0),
\end{aligned}$$

which is an integer divisible by p.

The only remaining detail to complete our proof that e is transcendental is to note that we can assume $c_0 \neq 0$ (or else we could divide out extra factors of x from g), and that if we choose $p > |c_0|$ then $c_0\widetilde{f}(0)$ is not divisible by p, while $c_k\widetilde{f}(k)$ is divisible by p for $1 \leq k \leq n$, and thus $\sum_{k=0}^{n} c_k\widetilde{f}(k)$ is a nonzero integer.

Oddtown and its cousins

The crucial point above was that we could simultaneously approximate e very well, while maintaining that $\sum c_k\widetilde{f}(k) \neq 0$. This reminds me of a family of combinatorial problems which I know by the name of "the Oddtown problem" (or the generalization "the mod-n-town problem").

The story goes like this: Oddtown is an odd little town, and its residents are enthusiastic about

forming clubs. The clubs in Oddtown have the interesting properties that:

1. every club has an odd number of members, but
2. for every pair of clubs, the number of people who are members of both clubs is always even.

At most how many clubs can there be in Oddtown? (Perhaps you'd like to think about it for a moment before you read on. This parenthetical block of text is unimportant to the exposition, it is merely here to provide your eyes a convenient stopping point if you wish to think about the answer yourself before you continue to read. Are you ready to read the answer? Really really? Okay, in the next paragraph I'm going to reveal the answer.)

A bit of thought reveals that there is at least one configuration of clubs possible: every citizen of Oddtown could be the sole member of their own private club. This way, there would be as many clubs as denizens of Oddtown, each club would have 1 member (an odd number), and every pair of clubs would have 0 members in common (an even number). Could we hope to do better by cleverly assigning residents to clubs? No: there are at most as many clubs in Oddtown as there are residents.

In the generalization, the residents of mod-n-town form clubs so that

1. the number of members of each club is not congruent to 0 mod n, but
2. for every pair of clubs, the number of people who are members of both clubs *is* congruent to 0 mod n.

The same example shows that mod-n-town could have at least as many clubs as residents, and the question is could it possibly have more? The answer is known for n a prime power (I will give the proof below), but unknown for any composite numbers other than prime powers. We do not even know the answer for mod-6-town.

Proposition (mod-p^k-town)**.** In mod-p^k-town, there are at most as many clubs as residents.

Proof. Let there be N residents in mod-p^k-town. Form the matrix A, where $A_{ij} = 1$ if person i is in club j, and 0 otherwise. Then A has N rows, so $\text{rank}(A) \leq N$. We will show that the columns of A are linearly independent over $\mathbb{Z}$, and so the number of columns is at most $\text{rank}(A)$. Note that the hypotheses on the sizes of the clubs translate to

$$c_i \cdot c_j \equiv \begin{cases} b_i \neq 0 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \pmod{p^k},$$

Where c_i denotes the ith column of A. Suppose we had a linear dependence $\sum_j a_j c_j = \mathbf{0}$ with $a_i \in \mathbb{Z}$ not all 0. Dividing out by common factors, we may assume the a_i are setwise coprime, and

in particular, that there is some i so that $p \nmid a_i$. For that i, we have

$$\begin{aligned} 0 = c_i \cdot \mathbf{0} &= c_i \cdot \sum_j a_j c_j \\ &= \sum_j a_j (c_i \cdot c_j) \\ &\equiv a_i b_i \pmod{p^k}. \end{aligned}$$

However, $b_i \not\equiv 0 \pmod{p^k}$, and a_i is a unit mod p^k, so $a_i b_i \not\equiv 0 \pmod{p^k}$.

There are several interesting features of the proof. I find it interesting and surprising that this combinatorics problem has such an elegant solution by linear algebra! But also, relating back to the proof that e is transcendental, the contradiction reached in the proof is that the sum of pairwise products of integers on the one hand is equal to 0, but on the other hand cannot be divisible by p (must be p-adically far from 0). The issue with extending the proof to mod-6-town, or any other mod-composite-town, is that we cannot reach the same kind of contradiction. Attempting the same proof for mod-6-town, we would reach the conclusion that each a_i is divisible by either 2 or by 3. But there is no contradiction there: some of the $a_i b_i$ can be divisible by 2, others by 3, and the whole sum could still be 0. A similar thing happens in Hermite's proof if p is not chosen to be prime, but is just replaced with some large integer.

This similarity makes me wonder whether a solution to the mod-6-town problem could potentially provide a new tool for proving that numbers are irrational or transcendental. I don't have a specific target, nor a specific argument, in mind, but the idea would be that we might have a constant where we can find "good" approximations, but those approximations leave us with little control over the prime factors showing up.