**Math 105 notes**

C. Pomerance

# 1 The harmonic sum

The harmonic sum is the sum of reciprocals of the positive integers. We know from calculus that it diverges, this is usually done by the integral test. There's a more elementary proof that goes as follows:

$$\sum_{n=1}^{\infty}\frac{1}{n}=\sum_{j=0}^{\infty}\sum_{n=2^j}^{2^{j+1}-1}\frac{1}{n}>\sum_{j=0}^{\infty}\sum_{n=2^j}^{2^{j+1}-1}\frac{1}{2^{j+1}}=\sum_{j=0}^{\infty}\frac{1}{2},$$

which diverges to infinity. However, the integral test is the better way because it can prove fairly good upper and lower bounds for the sum of the first $N$ terms of the harmonic series. We have

$$\frac{1}{n}<\int_{n-1}^{n}\frac{dt}{t},\quad n\geq 2$$

and

$$\frac{1}{n}>\int_{n}^{n+1}\frac{dt}{t},\quad n\geq 1.$$

Thus, adding these inequalities for $n$ up to $N$ gets us

$$\int_{1}^{N+1}\frac{dt}{t}<\sum_{n=1}^{N}\frac{1}{n}\leq 1+\int_{1}^{N}\frac{dt}{t}.$$

(The second inequality treats the term $n=1$ of the sum as just the number 1, and there's no estimate here, in fact, if $N=1$ the partial sum of the harmonic series is 1.) Evaluating the integrals gets us

$$\log(N+1)<\sum_{n=1}^{N}\frac{1}{n}\leq 1+\log N. \tag{1}$$

Losing a little information here, this result could be written as

$$\sum_{n=1}^{N}\frac{1}{n}=\log N+O(1).$$

This means that there is some positive constant $c$ such that

$$\left|\sum_{n=1}^{N}\frac{1}{n}-\log N\right|\leq c$$

for all positive integers $N$. The $O$-notation hides the constant $c$ from view; it takes getting used to!

This result can also be written as

$$\sum_{n \le x} \frac{1}{n} = \log x + O(1), \quad x \ge 1. \tag{2}$$

Here $x$ is a real variable (we usually use $n$ and $N$ for integer variables). In the inequality under the summation sign, it is implicitly assumed that $n \ge 1$. A more precise way to write it is

$$\sum_{n=1}^{\lfloor x \rfloor} \frac{1}{n}.$$

The reason why the estimate holds is that the difference between $\log\lfloor x \rfloor$ and $\log x$ is tiny, in fact it goes to 0 as $x \to \infty$. So the additional error made in writing $\log x$ instead of $\log\lfloor x \rfloor$ can be absorbed into the $O(1)$ term in (2).

This is all wonderful, but can we do better? That is, can we be more precise about the error — what can be said about

$$\sum_{n \le N} \frac{1}{n} - \log N?$$

Let

$$a_n = \frac{1}{n} - \int_n^{n+1} \frac{dt}{t}.$$

Integrating and using properties of log, we get

$$a_n = \frac{1}{n} - \log\left(1 + \frac{1}{n}\right),$$

and expanding via the Taylor series for $\log(1+x)$ which converges for $-1 < x \le 1$, we get

$$a_n = \frac{1}{n} - \left(\frac{1}{n} - \frac{1}{2n^2} + \frac{1}{3n^3} - + \dots\right) = \frac{1}{2n^2} - \frac{1}{3n^3} + - \dots.$$

This is an alternating series where the terms decrease in absolute value, so we see that

$$0 < a_n < \frac{1}{2n^2}.$$

A consequence of this inequality is that the sum

$$\sum_{n=1}^{\infty} a_n$$

converges to a positive constant. Call this constant $\gamma$. It is known as *Euler's constant* or sometimes, the *Euler–Mascheroni constant*. A very recent article about it can be found in the October 2013 issue of the Bulletin of the American Mathematical Society, written by Jeffrey Lagarias. To 10 decimal places, $\gamma = 0.5772156649$. It is not known if $\gamma$ is rational or irrational.

**Theorem 1.** *For $x \geq 1$ we have*

$$\sum_{n \leq x} \frac{1}{n} = \log x + \gamma + O\left(\frac{1}{x}\right).$$

*Proof.* Let $N = \lfloor x \rfloor$. We have

$$\begin{aligned}
\sum_{n \leq x} \frac{1}{n} = \sum_{n \leq N} \frac{1}{n} &= \sum_{n \leq N} \left(a_n + \int_n^{n+1} \frac{dt}{t}\right) \\
&= \sum_{n \leq N} a_n + \int_1^{N+1} \frac{dt}{t} \\
&= \sum_{n=1}^{\infty} a_n - \sum_{n \geq N+1} a_n + \log(N+1) \\
&= \log(N+1) + \gamma - \sum_{n \geq N+1} a_n.
\end{aligned}$$

Let us estimate this remaining sum using the inequality $0 < a_n < \frac{1}{2n^2}$ that we established above. This gives us

$$0 < \sum_{n \geq N+1} a_n < \sum_{n \geq N+1} \frac{1}{2n^2}.$$

This sum can be majorized in two ways, one more formulaic, the other more clever. The forumlaic way is to use the inequality

$$\frac{1}{2n^2} < \int_{n-1}^{n} \frac{dt}{2t^2},$$

so that

$$\sum_{n \geq N+1} \frac{1}{2n^2} < \int_N^{\infty} \frac{dt}{2t^2} = \frac{1}{2N}.$$

The clever way to estimate the sum is to use

$$\frac{1}{2n^2} < \frac{1}{2n(n-1)} = \frac{1}{2}\left(\frac{1}{n-1} - \frac{1}{n}\right)$$

and then use a telescoping series to get exactly the same estimate. In any event, this shows that

$$\sum_{n \geq N+1} a_n = O\left(\frac{1}{N}\right) = O\left(\frac{1}{x}\right).$$

To complete the proof of the theorem we need to show that

$$\log(N+1) = \log x + O\left(\frac{1}{x}\right).$$

But this is easy, using the inequality $\log(1+\theta) \le \theta$, which holds for all $\theta > -1$ and is easily proved using calculus. (We did so in class.) Thus, if $\theta = (N+1)/x - 1$, we have

$$0 < \log(N+1) - \log x = \log\left(\frac{N+1}{x}\right) = \log(1+\theta) \le \theta < \frac{1}{x}.$$

□

Above we introduced $O$-notation, see the book for a fuller description. We also introduced Euler's constant, and we set the tone for manipulating sums and approximating with the Taylor series. These tools will be commonplace in the course.

# 2 The prime harmonic sum

In 1737, Euler proved that the sum of reciprocals of the primes (called the prime harmonic sum) diverges. In fact he showed that while the ordinary harmonic sum diverges like log, the prime harmonic sum diverges like loglog. See the recent article of Paul Pollack for more on what Euler knew or could have known ("Euler and the partial sums of the prime harmonic series" at http://www.math.uga.edu/∼pollack/work.html). In particular, the sum of the reciprocals of the primes diverges.

The book has a proof of the divergence of the prime harmonic series in Chapter 1 that we covered in class. Here's another proof, also covered in class. Let $P(n)$ denote the largest prime factor of $n$ when $n \ge 2$, and let $P(1) = 1$. So, for example, saying that $P(n) \le 2$ is just saying that $n$ is of the form $2^j$, and saying that $P(n) \le 3$ means that $n = 2^j 3^k$ for some non-negative integers $j, k$. Note that

$$\sum_{P(n)\le 2} \frac{1}{n} = \sum_{j=0}^{\infty} \frac{1}{2^j} = 2.$$

(The information under the first summation sign here encodes that $n$ is the dummy variable and that it is running over all positive integers which satisfy the condition.) It's more exciting to consider

$$\sum_{P(n)\le 3} \frac{1}{n} = \sum_{j\ge 0,\, k\ge 0} \frac{1}{2^j 3^k}.$$

This double sum over $j$ and $k$ factors (using the fundamental theorem of arithmetic) as

$$\sum_{j\ge 0} \frac{1}{2^j} \sum_{k\ge 0} \frac{1}{3^k} = 2 \cdot \frac{3}{2} = 3.$$

More generally, we have for any integer $N \ge 2$,

$$\sum_{P(n)\le N} \frac{1}{n} = \prod_{p\le N} \sum_{j\ge 0} \frac{1}{p^j} = \prod_{p\le N} \frac{p}{p-1}. \tag{3}$$

Changing the sum into a product was a trick devised by Euler and it depends intimately on unique factorization into primes.

The sum in (3) on $n$ involves all numbers $n \le N$ and also lots of larger ones too, in fact infinitely many. In any event, we have from (3) that

$$\sum_{n \le N} \frac{1}{n} < \prod_{p \le N} \frac{p}{p-1}.$$

Using that the sum is greater than $\log N$ (see (2)), and taking logs of this last inequality, gets us to

$$\log\log N < \sum_{p \le N} \log\left(\frac{p}{p-1}\right). \tag{4}$$

Again using the Taylor series, we have

$$\frac{1}{p} - \log\left(\frac{p}{p-1}\right) = \frac{1}{p} + \log\left(1 - \frac{1}{p}\right) = -\frac{1}{2p^2} - \frac{1}{3p^3} - \cdots = -A_p,$$

say. Let $\alpha$ denote the sum

$$\alpha = \sum_p A_p = \sum_p \left(\frac{1}{2p^2} + \frac{1}{3p^3} + \ldots\right).$$

(The notation indicates that we have an infinite sum over all primes $p$.) We note that the sum is convergent, since

$$0 < A_p < \frac{1}{2}\left(\frac{1}{p^2} + \frac{1}{p^3} + \ldots\right) = \frac{1}{2p(p-1)}. \tag{5}$$

(Here, we replaced all of the numerical fractions in the terms in $A_p$ with $\frac{1}{2}$, making the sum larger, and then saw a geometric series which can be summed exactly.) So the series that defines $\alpha$ converges with comparison to the series

$$\sum_p \frac{1}{2p(p-1)}.$$

We conclude from this that

$$\begin{aligned}
\sum_{p \le N} \frac{1}{p} &= \sum_{p \le N} \left(\frac{1}{p} - \log\left(\frac{p}{p-1}\right)\right) + \sum_{p \le N} \log\left(\frac{p}{p-1}\right) \\
&= \sum_p \left(\frac{1}{p} - \log\left(\frac{p}{p-1}\right)\right) - \sum_{p > N} \left(\frac{1}{p} - \log\left(\frac{p}{p-1}\right)\right) + \sum_{p \le N} \log\left(\frac{p}{p-1}\right) \\
&= -\alpha - \sum_{p > N} A_p + \sum_{p \le N} \log\left(\frac{p}{p-1}\right).
\end{aligned}$$

Using the inequality (4) for the final sum here, we have then that

$$\sum_{p\le N}\frac{1}{p} > \log\log N - \alpha - \sum_{p>N} A_p.$$

The infinite sum here is, from (5), less than 0, but greater than

$$-\sum_{p>N}\frac{1}{2p(p-1)} > -\frac{1}{2N}.$$

Since the sum appears with a negative sign, we can ignore it in the inequality, getting

$$\sum_{p\le N}\frac{1}{p} > \log\log N - \alpha. \tag{6}$$

So, the big question now is how accurate is this lower bound? Do we have a companion upper bound?

# 3 An upper bound for the partial sums of the prime harmonic series

We begin with the almost trivial estimate

$$N\log N \ge \sum_{n=1}^{N}\log n \ge \sum_{n=1}^{N}\sum_{p|n}\log p.$$

The inner sum here is over the distinct prime factors of $n$. The next step is to interchange the order of summation, which will be a common trick for us. This gets us to

$$N\log N \ge \sum_{p\le N}\log p\sum_{\substack{n\le N\\ p|n}} 1 = \sum_{p\le N}\log p\cdot\left\lfloor\frac{N}{p}\right\rfloor > \sum_{p\le N}\log p\left(\frac{N}{p}-1\right) = \sum_{p\le N}\left(\frac{N\log p}{p}-\log p\right). \tag{7}$$

**Lemma 1** (Erdős, Chebyshev)**.** *For each positive integer $N$, we have*

$$\prod_{p\le N} p \le 4^N.$$

*Proof.* It's true for $N = 1, 2$. If $N$ is the least number for which it fails, then $N \ge 3$ and $N$ is odd (since if $N$ were even it would be composite, and since it is true for the previous odd number, the product remains the same as with the previous odd number). Write $N = 2k+1$. We have the binomial coefficient $\binom{2k+1}{k}$ divisible by every prime in $[k+2, 2k+1]$, and so

$$\prod_{k+2 \le p \le 2k+1} p \le \binom{2k+1}{k} = \frac{1}{2}\left(\binom{2k+1}{k} + \binom{2k+1}{k+1}\right),$$

since the two binomial coefficients here are equal. Their sum is smaller than

$$\sum_{j=0}^{2k+1} \binom{2k+1}{j} = (1+1)^{2k+1} = 2^{2k+1}.$$

Using this in the above, we have

$$\prod_{k+2 \le p \le 2k+1} p < \frac{1}{2} 2^{2k+1} = 2^{2k} = 4^k.$$

Since $k+1 < 2k+1$, the lemma is true for $k+1$, so we have

$$\prod_{p \le 2k+1} p = \prod_{p \le k+1} p \prod_{k+2 \le p \le 2k+1} p < 4^{k+1} 4^k = 4^{2k+1}.$$

Thus, $N = 2k+1$ is not a counterexample afterall, so the Lemma always holds. $\square$

The logarithm of the inequality in Lemma 1 is

$$\sum_{p \le N} \log p \le N \log 4.$$

Using this, moving some terms around in (7), and dividing by $N$ gets us

$$\sum_{p \le N} \frac{\log p}{p} \le \log N + \log 4. \tag{8}$$

The issue now is how to traverse from this estimate to an upper bound for the partial sum of the prime harmonic series. For this we introduce the concept of partial summation.

**Proposition 1** (Partial summation). *Suppose that $f(x)$ is a continuously differentiable real valued function and $a_n$ for $n = 1, 2, \ldots$ is a sequence of real numbers. Then*

$$\sum_{n \le N} a_n f(n) = f(N) \sum_{n \le N} a_n - \int_1^N \sum_{n \le t} a_n f'(t)\, dt.$$

*Proof.* The contribution of a particular term on the left side is $a_n f(n)$. We compute what "$n$" contributes on the right side. After a little thought we see that this is

$$f(N)a_n - \int_n^N a_n f'(t)\,dt = f(N)a_n - (a_n f(N) - a_n f(n)) = a_n f(n).$$

The contributions are the same, so we have proved the identity. □

Partial summation can be used too when we have a prime dummy variable, or more generally for any subsequence of the natural numbers. For example, if one wants to sum

$$\sum_{p \le N} a_p f(p)$$

and use Proposition 1, we can let $a_n = 0$ when $n$ is not prime, so the sum instantly gets transformed to

$$\sum_{n \le N} a_n f(n),$$

since all of the new terms are 0. The sums on the right side of the identity are also the same as if they were restricted to primes, so we have

$$\sum_{p \le N} a_p f(p) = f(N) \sum_{p \le N} a_p - \int_2^N \sum_{p \le t} a_p f'(t)\,dt.$$

Note the lower limit of integration is now 2, since we can ignore the $n = 1$ term, being 0.

We now apply partial summation to the sum in (8):

$$\sum_{p \le N} \frac{1}{p} = \sum_{p \le N} \frac{\log p}{p} \cdot \frac{1}{\log p} = \frac{1}{\log N} \sum_{p \le N} \frac{\log p}{p} + \int_2^N \frac{1}{t \log^2 t} \sum_{p \le t} \frac{\log p}{p}\,dt.$$

Thus, using (8), we have

$$\begin{aligned} \sum_{p \le N} \frac{1}{p} &\le \frac{1}{\log N}(\log N + \log 4) + \int_2^N \frac{\log t + \log 4}{t \log^2 t}\,dt \\ &= 1 + \frac{\log 4}{\log N} + \log\log N - \log\log 2 - \frac{\log 4}{\log N} + \frac{\log 4}{\log 2} = \log\log N + 3 - \log\log 2. \qquad (9) \end{aligned}$$

We have thus proved together with (2) that for $x \ge 2$,

$$\sum_{p \le x} \frac{1}{p} = \log\log x + O(1).$$

I mentioned in class the theorem of Mertens from 1874 that for $x \geq 2$,

$$\sum_{p \leq x} \frac{1}{p} = \log\log x + \gamma - \alpha + O\left(\frac{1}{\log x}\right). \tag{10}$$

In class we worked out some exercises using partial summation. In particular we used the logarithm of the inequality in Lemma 1 to show that

$$\pi(x) = \sum_{p \leq x} 1 = O\left(\frac{x}{\log x}\right).$$

We also showed that Mertens's theorem (10) implies the same thing. Improving the error term slightly in (10) to $o(1/\log x)$ implies, via partial summation, the prime number theorem, namely

$$\pi(x) \sim \frac{x}{\log x} \text{ as } x \to \infty.$$

Saying that the error in (10) is $o(1/\log x)$ means that

$$\frac{\sum_{p \leq x} \frac{1}{p} - (\log\log x + \gamma - \alpha)}{1/\log x} \to 0 \text{ as } x \to \infty.$$

And saying that $\pi(x) \sim x/\log x$ as $x \to \infty$ means that

$$\frac{\pi(x)}{x/\log x} \to 1 \text{ as } x \to \infty.$$