**Math 259: Introduction to Analytic Number Theory**

Čebyšev (and von Mangoldt and Stirling)

It is well known[1] that for any prime $p$ and positive integer $x$ the exponent of $p$ in $x!$ is

$$c_p(x) := \left\lfloor \frac{x}{p} \right\rfloor + \left\lfloor \frac{x}{p^2} \right\rfloor + \left\lfloor \frac{x}{p^3} \right\rfloor + \cdots = \sum_{k=1}^{\infty} \left\lfloor \frac{x}{p^k} \right\rfloor,$$

the sum being finite because eventually $p^k > x$. It was Čebyšev's insight that one could extract information about $\pi(\cdot)$ from the resulting formula

$$x! = \prod_p p^{c_p(x)},$$

or equivalently

$$\log x! = \sum_p c_p(x)\log(p) = \sum_{n=1}^{\infty} \left\lfloor \frac{x}{n} \right\rfloor \Lambda(n), \tag{Č}$$

where $\Lambda(n)$ is the *von Mangoldt function*

$$\Lambda(n) := \begin{cases} \log p, & \text{if } n = p^k \text{ for some positive integer } k \text{ and prime } p; \\ 0, & \text{otherwise.} \end{cases}$$

For instance, Čebyšev was able to come close enough to the Prime Number Theorem to prove "Bertrand's Postulate": every interval $(x, 2x)$ with $x > 1$ contains a prime. (See [H&W 1996, p.343–4] for Erdös's simplification of Čebyšev's proof; this simplified proof is also on the Web: http://forum.swarthmore.edu/dr.math/problems/kuropatwa.4.3.97.html.)

To make use of (Č) we need to estimate

$$\log x! = \sum_{n=1}^{x} \log n$$

for large $x$. We do this by in effect applying the first few steps of symmetrized Euler-Maclaurin summation. For any $\mathcal{C}^2$ function $f$ we have (by integrating by parts twice)

$$\int_{-1/2}^{1/2} f(y)\,dy = f(0) + \frac{1}{2}\left[\int_{-1/2}^{0} f''(y)(y+1/2)^2\,dy + \int_{0}^{1/2} f''(y)(y-1/2)^2\,dy\right]$$

[1] If only thanks to the perennial problems along the lines of "how many zeros end 1998! ?".

$$= f(0) + \frac{1}{2}\int_{-1/2}^{1/2} f''(y)\langle y + \frac{1}{2}\rangle^2\, dy,$$

where $\langle z\rangle$ is the distance from $z$ to the nearest integer. Thus

$$\sum_{k=1}^{N} f(k) = \int_{1/2}^{N+\frac{1}{2}} f(y)\, dy + \frac{1}{2}\int_{\frac{1}{2}}^{N+\frac{1}{2}} f''(y)\langle y + 1/2\rangle^2\, dy.$$

Taking $f(y) = \log(y)$ and $N = x$ we thus have

$$\log x! = (x + \frac{1}{2})\log(x + \frac{1}{2}) + \frac{1}{2}\log 2 - x - \frac{1}{2}\int_{\frac{1}{2}}^{x+\frac{1}{2}} \langle y + 1/2\rangle^2 \frac{dy}{y^2}.$$

The integral is

$$-\frac{1}{2}\int_{\frac{1}{2}}^{\infty} \langle y + 1/2\rangle^2 \frac{dy}{y^2} \ + \ O(1/x),$$

and the other terms are

$$(x + \frac{1}{2})\log x - x + \frac{1}{2}(\log 2 - 1) + O(1/x).$$

Thus we have

$$\log x! = (x + \frac{1}{2})\log x - x + C + O(1/x) \tag{S}$$

for some absolute constant $C$. Except for the determination of $C$ (which turns out to be $\frac{1}{2}\log(2\pi)$, as we shall see about a week hence), this formula is of course *Stirling's approximation* to $x!$. Stirling's approximation extends to an asymptotic series for $x!/[(x/e)^x\sqrt{2\pi x}\,]$ in inverse powers of $x$, but for our purposes $\log x! = (x + \frac{1}{2})\log x - x + O(1)$ is more than enough. In fact, since for the time being we're really dealing with $\log\lfloor x\rfloor!$ and not $\log\Gamma(x+1)$, the best honest error term we can use is $O(\log x)$.

Now let

$$\psi(x) := \sum_{1\le n\le x} \Lambda(n).$$

Then from (Č) and (S) we have

$$\sum_{k=1}^{\infty} \psi(x/k) = (x + \frac{1}{2})\log x - x + O(1).$$

This certainly suggests that $\psi(x) \sim x$, and lets us prove upper and lower bounds on $\psi(x)$ proportional to $x$, for instance

$$\psi(x) \le \log x! - \sum_{m=1}^{\infty} 2^{-m}\log\lfloor x/2^m\rfloor! = \left[\sum_{m=1}^{\infty}\frac{m}{2^m}\log 2\right] x + O(\log^2 x) =$$

$$= (2\log 2)x + O(\log^2 x)$$

(since $x \geq 1 + \sum_{m=1}^{\infty} \lfloor x/2^m \rfloor$ for $x \geq 1$) and[2]

$$\psi(x) \geq \sum_{k=1}^{\infty} (-1)^{k-1} \psi(x/k) = \log \frac{(2x)!}{x!^2} = (\log 2)x + O(\log x).$$

It is true that we're ultimately interested in $\pi(x)$, not $\psi(x)$. But it is easy to get from one to the other. For one thing the contribution to $\psi(x)$ of prime powers $p^k$ with $k > 1$ is negligible — certainly less than $\sum_{k=2}^{\infty} \lfloor x^{1/k} \rfloor \log x \ll x^{1/2} \log x$. The remaining sum, $\sum_{p<x} \log p$, can be expressed in terms of $\pi(x)$ and vice versa using partial summation, and we find:

$$\psi(x) = \log(x)\pi(x) - \int_1^x \pi(y) \frac{dy}{y} + O(x^{1/2} \log x),$$

$$= f(0) + \frac{1}{2} \int_{-1/2}^{1/2} \pi(x) = \frac{\psi(x)}{\log x} + \int_3^x \psi(y) \frac{dy}{y \log^2 y} + O(x^{1/2} \log x).$$

It follows that the Prime Number Theorem $\pi(x) \sim x/\log x$ holds if and only if $\psi(x) \sim x$, and good error terms on one side imply good error terms on the other. It turns out that we can more readily get at $\psi(x)$ than at $\pi(x)$; for instance $\psi(x)$ is quite well approximated by $x$, while the "right" estimate for $\pi(x)$ is not $x/\log x$ but that plus $\int^x dy/\log^2 y$, i.e. the "logarithmic integral" $\int^x dy/\log y$. It is in the form $\psi(x) \sim x$ that we'll actually prove the Prime Number Theorem.

Exercises:

Since our upper and lower asymptotic bounds $\log 2, \log 4$ on $\psi(x)/x$ are within a factor of 2 of each other, they do not quite suffice to prove Bertrand's Postulate. But any improvement *would* prove that $\pi(2x) > \pi(x)$ for sufficiently large $x$, from which the proof for all $x$ follows by just exhibiting a few suitably spaced primes. It turns out that better bounds are available starting from (Č). For instance, show that $\psi(x) < (\frac{1}{2}\log 12)x + O(\log^2 x)$. Can you obtain Čebyšev's bounds of 0.9 and 1.1? In fact it is known that the upper and lower bounds can be brought arbitrarily close to 1, but alas the only known proof of that fact depends on the Prime Number Theorem!

Here is another elementary approach: let $P(u)$ be any nonzero polynomial with integer coefficients of degree $d$; then

$$\int_0^1 f(u)^{2n}\, du \geq 1/\mathrm{lcm}(1, 2, \ldots, 2dn+1) = \exp(-\psi(2dn+1)) > \frac{1}{\prod_{p<2dn} p}.$$

[2] This is essentially the same tactic of factoring $\binom{2n}{n}$ that Čebyšev used to prove $\pi(2x) > \pi(x)$.

Thus

$$\sum_{p<2dn} \log p > 2n \log \min_{0<u<1} 1/|P(u)|.$$

For instance, taking $f(u) = u - u^2$ we find (at least for $4|x$) that $\sum_{p<x} \log x < x \log 4$. This is essentially the same (why?) as Čebyšev's trick of factoring $\binom{2n}{n}$, but suggests different sources of improvement; try $f(u) = (u - u^2)(1 - 2u)$ for example. [Unfortunately here the upper bound cannot be brought down to $1+\epsilon$; see [Montgomery 1994, Chapter 10] — thanks to Madhav Nori for bringing this to my attention.]

Show that

$$\sum_{p\leq x} \log p = \psi(x) - \psi(x^{1/2}) - \psi(x^{1/3}) - \psi(x^{1/5}) + \psi(x^{1/6}) \cdots = \sum_{k=1}^{\infty} \mu(k)\psi(x^{1/k}),$$

where $\mu$ is the *Möbius function* taking the product of $r \geq 0$ *distinct* primes to $(-1)^r$ and any non-square-free integer to 0.

**References**

[H&W 1996] Hardy, G.H., Wright, E.M.: *An Introduction to the Theory of Numbers*, 5th ed. Oxford: Clarendon Press, 1988 [AB 9.88.10 / QA241.H37].

[Montgomery 1994] Montgomery, H.L.: *Ten lectures on the interface between analytic number theory and harmonic analysis.* Providence: AMS, 1994 [AB 9.94.9].