THE AVERAGE SPACING BETWEEN PRIMES

ROBERT J. LEMKE OLIVER AND FRANK THORNE

13. The Average Spacing Between Primes

We know that there are $\approx \frac{x}{\log x}$ primes $\leq x$. Can we answer the 'inverse' question? If p_n denotes the nth prime, then how large is p_n on average?

Proposition 13.1 (The 'inverse' prime number theorem)**.** *We have*

$$p_n = n \log n + o(n \log n).$$

Proof. It is equivalent to prove that for every $\epsilon > 0$, the inequalities

$$\pi\big((1+\epsilon)n \log n\big) > n,$$
$$\pi\big((1-\epsilon)n \log n\big) < n$$

hold for sufficiently large n. By the prime number theorem (in its weak form, for which we don't need the Riemann Hypothesis) we have

$$\begin{aligned}
\pi\big((1+\epsilon)n \log n\big) &= \frac{(1+\epsilon)n \log n}{\log\big((1+\epsilon)n \log n\big)} + o\left(\frac{(1+\epsilon)n \log n}{\log\big((1+\epsilon)n \log n\big)}\right) \\
&= \frac{(1+\epsilon)n \log n}{\log n + \log\log n + \log(1+\epsilon)} + o(n) \\
&= \frac{(1+\epsilon)n \log n}{\log n} \cdot \left(1 + O\left(\frac{\log\log n}{\log n}\right)\right) + o(n) \\
&= (1+\epsilon)n + o(n),
\end{aligned}$$

exactly as we needed. The reverse inequality follows in the same way. □

We can now prove the folliowing:

Proposition 13.2 (Average spacing between primes)**.** *On average, we have*

$$p_{n+1} - p_n \sim \log n.$$

To borrow a phrase from Ravi Vakil, once we have written out what this means, we will have essentially proven it by accident.

Proof. Let us choose the following precise interpretation of Proposition 13.2. For every large N, we have

$$\frac{1}{N} \sum_{N \leq n < 2N} \big(p_{n+1} - p_n\big) = \log N + o(\log N). \tag{13.1}$$

Note that $\log(2N) = \log N + o(\log N)$, so that on the right side of (13.1) we could replace $\log N$ by $\log(2N)$, or by $\log n$ for any n in the sum.

Date: November 4, 2019.

But this is just a complicated way of writing

$$p_{2N} - p_N = N \log N + o(N \log N).$$

So the result follows straightaway from Proposition 13.1. □

So now we have another interesting statistic to study, the 'normalized spacing between primes'[1] $\alpha(n) := \frac{p_{n+1}-p_n}{\log n}$. This function is 1 on average, and here are some questions we can ask:

- Does $\alpha(n)$ have any lower bound, other than zero?
- Is $\alpha(n)$ bounded from above?
- Is the set of all $\alpha(n)$ dense in $\mathbb{R}$?
- Does the set $\{\alpha(n) \ : n < N\}$ converge to any nice distribution?

These are *not* easy questions. Let's investigate the first. Clearly it follows from the twin prime conjecture. Equally, it follows from the Zhang-Maynard-Tao-Polymath result

$$\liminf_{n\to\infty} (p_{n+1} - p_n) \leq 246.$$

In fact, the proof came earlier – but not *so* much earlier:

Theorem 13.3 (Goldston-Pintz-Yıldırım, 2005)**.** *We have*

$$\liminf_{n\to\infty} \frac{p_{n+1} - p_n}{\log n} = 0.$$

The statement is much weaker than the twin prime conjecture, but it had stubbornly resisted proof for quite some time! This result justly made its authors famous, and the Zhang, Maynard, and Polymath papers are all a further development of the same ideas.

In the opposite direction, Westzynthius proved in 1931 that

$$\limsup_{n\to\infty} \frac{p_{n+1} - p_n}{\log n} = +\infty.$$

This result has been further improved. In 1937, Rankin proved that, for some constant $c > 0$, the inequality

$$p_{n+1} - p_n > \frac{c \log n \log\log n \log\log\log\log n}{(\log\log\log n)^2} \tag{13.2}$$

holds for infinitely many values of n. Paul Erdős offered a $10,000 prize for a proof that the constant c in (13.3) can be taken to be arbitrarily large.

The prize was won by Ford, Green, Konyagin, Tao, and (independently) Maynard, who then improved the inequality to

$$p_{n+1} - p_n > \frac{c \log n \log\log n \log\log\log\log n}{\log\log\log n}. \tag{13.3}$$

Terence Tao offers[2] a $10,000 bounty for a proof that *this* c can be taken arbitrarily large.

Joke 13.4. *What does a drowning analytic number theorist say? – Log, log, log, log, …*

[1][To RLO: Is there any notation you like for this?]

[2]See his blog post, `https://terrytao.wordpress.com/2014/12/16/long-gaps-between-primes/`

13.1. **A probabilistic model.** We now consider, following Cramér, Gallagher, and Soundararajan, the question of what a probabilistic model predicts. Let us first consider the most naive probabilistic model: an integer n is prime with probability $1/\log n$. (Since $\log(n \log n) \approx \log n$, an integer of size close to p_n is also predicted to be prime with the same probability.)

What then, is the probability that

$$\frac{p_{n+1} - p_n}{\log n} \in [\alpha, \beta]$$

for fixed $\beta > \alpha \geq 0$? According to our random model, this should be

$$\sum_{\alpha \log n \leq h \leq \beta \log n} \left(1 - \frac{1}{\log n}\right)^{h-1} \cdot \frac{1}{\log n}.$$

Since $(1 - 1/t)^t \to e^{-1}$, this is asymptotically

$$\sum_{\alpha \log n \leq h \leq \beta \log n} e^{-h/\log n} \cdot \frac{1}{\log n} \approx \int_{\alpha \log n}^{\beta \log n} e^{-t/\log n} \frac{dt}{\log n} = \int_{\alpha}^{\beta} e^{-u} du.$$

This is a beautiful answer! [To do: add probability background] In particular, it answers all of our bullet points above. Unfortunately, it has the disadvantage of being based on faulty assumptions. We have seen that the probabilities of n and $n + k$ are *not* independent; that one has to correct by a 'singular series' factor describing the local behavior modulo every prime. Should we still expect the above?

Let us for the moment push on and address a second question. We have seen that, on average, the interval $[n, n + \log n]$ will contain exactly one prime. But what is the probability of this? Or that it contains more, or fewer?

More generally, what is the probability that the interval $[n, n + \lambda \log n]$ contains exactly k primes? We can apply similar reasoning to this question. Let $M = \lambda \log n + O(1)$ be the number of integers in $[n + \lambda \log n]$. Then our probabilistic model predicts an answer of

$$\begin{aligned} \binom{M}{k} \left(\frac{1}{\log n}\right)^k \left(1 - \frac{1}{\log n}\right)^{M-k} &\approx \frac{(\lambda \log n)^k}{k!} \left(\frac{1}{\log n}\right)^k \left(1 - \frac{1}{\log n}\right)^{\lambda \log n - k} \\ &\approx \frac{\lambda^k}{k!} e^{-\lambda}. \end{aligned} \tag{13.4}$$

Remark. An excellent exercise is to verify all of the above approximations rigorously!

These predictions *should* look intimidatingly difficult to prove. (Remember that the proof of Theorem 13.3 was published in the *Annals of Mathematics* not so long ago.) But a striking result of Gallagher establishes that these predictions *do* hold, if we assume the truth of an earlier prediction.

Namely, we recall the statement of the *Hardy-Littlewood prime tuple conjecture*[3]:

Conjecture 13.5 (Hardy-Littlewood prime tuple conjecture, again)**.** *Let*

$$\mathcal{H} = \{h_1, h_2, \cdots h_k\}$$

be an admissible k-tuple of integers. Then,

$$\#\{n \leq x : \ n + h_1, \ldots, n + h_k \text{ are all prime}\} \sim \mathfrak{S}(\mathcal{H}) \frac{x}{(\log x)^k}. \tag{13.5}$$

[3]To RLO: text copy and paste. To be avoided?

The singular series $\mathfrak{S}(\mathcal{H})$ is defined by

$$\mathfrak{S}(\mathcal{H}) := \prod_p \left(1-\frac{1}{p}\right)^{-k} \cdot \left(1-\frac{\nu_{\mathcal{H}}(p)}{p}\right), \tag{13.6}$$

where, for each prime p, $\nu_{\mathcal{H}}(p)$ is the number of distinct residue classes $(\mathrm{mod}\, p)$ represented by elements of $\mathcal{H}$.

Theorem 13.6 (Gallagher)**.** *Assume that the Hardy-Littlewood conjecture* (13.5) *holds uniformly for all $\mathcal{H}$ with $h_k - h_1 \leq \lambda \log N$.*

Then, we have

$$\#\{n \leq N \ : \ (n, n+h] \text{ contains exactly } k \text{ primes}\} \sim N \frac{\lambda^k}{k!} e^{-\lambda}.$$

So, in other words, not only should we indeed expect (13.4), but it is in fact a consequence of an earlier prediction of the same probabilistic model.

This implies the predicted answer to our first question as well:

Corollary 13.7. *Under the same assumptions as Theorem 13.6, we have*

$$\frac{1}{N}\#\left\{n \leq N \ : \ \frac{p_{n+1}-p_n}{\log n} \in [\alpha, \beta]\right\} \sim \int_\alpha^\beta e^{-t} dt.$$

Our objective in this chapter is to explain all of this. There are several key steps to this:

- First we will prove Gallagher's key technical result: that the *average* value of the singular series $\mathfrak{S}(\mathcal{H})$, as defined in (13.6), is equal to 1. This is to be expected: it asserts that the 'naive' Hardy-Littlewood prime tuple conjecture, although not true for any particular tuple $\mathcal{H}$, should hold on average.
- We then give our reader a crash course in probability. We review the principle that, under suitable conditions, *a probability distribution is determined by its moments*. We will very much see the utility of this fact in the ensuing proof!
- After this, we are set to give Gallagher's proof of Theorem 13.6.
- Finally, we will explain how Theorem 13.6 implies Corollary 13.7.

13.2. **The average value of the singular series.** Suppose that $\mathcal{H} = \{0, h\}$. In that case, we have

$$\mathfrak{S}(\mathcal{H}) := \prod_p \left(1-\frac{1}{p}\right)^{-2} \cdot \left(1-\frac{\nu_{\mathcal{H}}(p)}{p}\right), \tag{13.7}$$

where $\nu_{\mathcal{H}}(p) = 1$ when $p \mid h$ and $\nu_{\mathcal{H}}(p) = 2$ otherwise. Then Gallagher tells us that

$$\lim_{H\to\infty} \frac{1}{H} \sum_{1\leq h\leq H} \mathfrak{S}(\{0,h\}) = 1 + o(1).$$

Even this special case is not trivial! Roughly speaking, one can argue as follows. The average value of $\nu_{\mathcal{H}}(p)$ is $2 - 1/p$, since we have $p \mid h$ for a proportion $1/p$ of h. Replacing $\nu_{\mathcal{H}}(p)$ by its average value, we obtain that

$$\lim_{H\to\infty} \frac{1}{H} \sum_{1\leq h\leq H} \mathfrak{S}(\{0,h\}) \overset{*}{=} \prod_p \left(1-\frac{1}{p}\right)^{-2} \cdot \left(1-\frac{2-\frac{1}{p}}{p}\right) = 1.$$

We can't just 'replace $\nu_{\mathcal{H}}(p)$ by its average value' in a rigorous argument, but at least this gives us an idea of what to expect.

More precisely, Gallagher proved the following.

Theorem 13.8 (Gallagher's singular series average)**.** *For fixed r, we have*

$$\sum_{\substack{1\le h_1,\dots,h_k\le N\\ \text{distinct}}} \mathcal{S}(h_1,\dots,h_k) = N^k\cdot(1+o(1)). \tag{13.8}$$

We now present his proof. Write $\mathcal{H} := \{h_1,\dots,h_k\}$.

Step 1. The singular series as an infinite sum. Gallagher begins by rewriting the singular series

$$\begin{aligned}
\mathfrak{S}(\mathcal{H}) &:= \prod_p \left(1-\frac{1}{p}\right)^{-k}\cdot\left(1-\frac{\nu_{\mathcal{H}}(p)}{p}\right)\\
&= \prod_p \left(\frac{p}{p-1}\right)^{k}\left(1-\frac{\nu_{\mathcal{H}}(p)}{p}\right)\\
&= \prod_p \left(1+\frac{p^k-\nu_{\mathcal{H}}(p)p^{k-1}-(p-1)^k}{(p-1)^k}\right)\\
&=: \prod_p (1+a(\mathcal{H},p)). \qquad (13.9)\\
&=: \sum_{\substack{q\\ \text{squarefree}}} a(\mathcal{H},q). \qquad (13.10)
\end{aligned}$$

We have

$$a(\mathcal{H},p) \ll_k \begin{cases} (p-1)^{-2} & \text{if } \nu_{\mathcal{H}}(p)=k,\\ (p-1)^{-1} & \text{if } \nu_{\mathcal{H}}(p)<k.\end{cases} \tag{13.11}$$

This is not difficult to check, and the would-be expert is advised to develop fluency in this sort of computation. That $a(\mathcal{H},p)\ll (p-1)^{-1}$ is almost obvious – upon multiplying out the polynomials in the numerator, the p^k term cancels out. To see (13.11), think of Taylor expanding the numerator: expand everything out in terms of powers of p, and ignore everything of order smaller than p^{k-1}.

We point out also that we have $\nu_{\mathcal{H}}(p)=k$ unless

$$p \mid D(\mathcal{H}) := \prod_{i<j}(h_i-h_j),$$

which we can think of as the 'discriminant' of the k-tuple $\mathcal{H}$. Therefore, we can easily see from (13.9) that the product defining $\mathfrak{S}(\mathcal{H})$ converges.

Step 2. Truncation and a tail estimate. Gallagher's next step is to rewrite (13.10) as

$$\sum_{\substack{q\\ \text{squarefree}}} a(\mathcal{H},q) = \sum_{\substack{q\le x\\ \text{squarefree}}} a(\mathcal{H},q) + O(E(x,h,k)),$$

for a *tail error estimate* $E(x,N,k)$ satisfying

$$\sum_{\substack{q>x\\ \text{squarefree}}} a(\mathcal{H},q) \leq E(x,N,k),$$

simultaneously for all $\mathcal{H}$ of size k with all $h_i \in [1,N]$.

To do this, write

$$\sum_{\substack{q>x\\ \text{squarefree}}} a(\mathcal{H},q) \leq \sum_{\substack{q>x\\ \text{squarefree}}} \frac{C^{\omega(q)}}{\phi(q)^2}\phi((q,D)),$$

where $\omega(q)$ is the number of prime divisors of q, and where the constant C depends on k only. (It is the implied constant in (13.11).)

We rewrite each $q=de$, where $d \mid D$ and $(e,D)=1$, to write the above as

$$\sum_{d|D} \frac{C^{\omega(d)}}{\phi(d)} \sum_{\substack{e>x/d\\ (e,D)=1}} \frac{C^{\omega(e)}}{\phi(e)^2},$$

where the sum is over squarefree d and e.

We have

$$\sum_{e>Y} \frac{C^{\omega(e)}}{\phi(e)^2} \ll_\epsilon Y^{-1+\epsilon}.$$

(In fact, one can do better here.) This is proved by estimating $\frac{e}{\phi(e)} \ll_\epsilon e^\epsilon$ and $C^{\omega(e)} \ll e^\epsilon$, and then summing the results. (The proof that $C^{\omega(e)} \ll e^\epsilon$ is an excellent exercise!) Therefore, we have

$$\sum_{\substack{q>x\\ \text{squarefree}}} a(\mathcal{H},q) \ll \sum_{\substack{d|D\\ \text{squarefree}}} \frac{C^{\omega(d)}}{\phi(d)} \cdot (x/d)^{-1+\epsilon} \ll D^\epsilon x^{-1+\epsilon} \ll_k N^\epsilon x^{-1+\epsilon}.$$

We therefore have $E(x,N,k) := O_{k,\epsilon}(N^\epsilon x^{-1+\epsilon})$ as our tail estimate, and may conclude that

$$\sum_{\substack{1\leq h_1,\dots,h_k\leq N\\ \text{distinct}}} \mathcal{S}(d_1,\dots,d_k) = \sum_{q\leq x} \sum_{\substack{1\leq h_1,\dots,h_k\leq N\\ \text{distinct}}} a(\mathcal{H},q) + O(N^k(Nx)^\epsilon x^{-1}). \tag{13.12}$$

In particular, looking forward to the end of the proof, we can see that the choice x^δ will suffice for any $\delta > 0$.

Step 3. Conclusion of the proof. Let $V(q)$ denote the set of tuples $(\nu_1,\dots,\nu_m)$, where $m=\omega(q)$ is the number of divisors of the squarefree integer q, and where each ν_i corresponds to one of the prime divisors $p_i \mid q$ and satisfies $1 \leq \nu_i \leq p$.

Then, the inner sum in (13.12) can be written in the form

$$\sum_{v\in V(q)} \prod_{p|q} a(p,v(p)) \left(A(k,N,v) + O(N^{k-1})\right),$$

where:

- $a(p,v(p))$ is defined as in (13.9), namely by

$$a(p,v(p)) := \frac{p^k - v(p)p^{k-1} - (p-1)^k}{(p-1)^k},$$

- $A(k, N, v)$ denotes the number of k-tuples of integers $h_1, \cdots, h_k$, all in $[1, N]$, such that for each $p \mid q$ the h_i occupy exactly $v(p)$ residue classes mod p.

  We don't assume that these integers are necessarily distinct. (The number of tuples with repeats is $\ll N^{k-1}$, handled by the error term above.)

By breaking the sum over $(h_1, \ldots, h_k)$ into arithmetic progressions modulo q, and counting within each individually, we find that

$$A(k, N, v) = \left[\left(\frac{N}{q}\right)^k + O\left(\frac{N}{q}\right)^{k-1}\right] \prod_{p|q} \binom{p}{v(p)} \sigma(k, v(p)),$$

where $\sigma(k, r)$ is the number of maps from $\{1, \ldots, k\}$ onto $\{1, \ldots, r\}$. (This is a 'Stirling number of the second kind'.)

Thus, the inner sum in (13.12) can be further rewritten as

$$\left(\frac{N}{q}\right)^k A(q) + O\left(\left(\frac{N}{q}\right)^{k-1} B(q)\right) + O(N^{k-1} C(q)), \tag{13.13}$$

with

$$\begin{aligned}
A(q) &= \sum_v \prod_{p|q} a(p, v(p)) \binom{p}{v(p)} \sigma(k, v(p)) &&= \prod_{p|q} \left[\sum_{v=1}^{p} a(p, v) \binom{p}{v} \sigma(k, v)\right], \\
B(q) &= \sum_v \prod_{p|q} |a(p, v(p))| \binom{p}{v(p)} \sigma(k, v(p)) &&= \prod_{p|q} \left[\sum_{v=1}^{p} |a(p, v)| \binom{p}{v} \sigma(k, v)\right], \\
C(q) &= \sum_v \prod_{p|q} |a(p, v(p))| &&= \prod_{p|q} \left[\sum_{v=1}^{p} a(p, v)\right].
\end{aligned}$$

We tackle the error terms $B(q)$ and $C(q)$ first. We have the combinatorial identity

$$\sum_{v=1}^{p} \binom{p}{v} \sigma(r, v) = p^r. \tag{13.14}$$

This is 'well known' (to combinatorialists, anyway), and we give a proof in Step 4. So we have

$$B(q) \ll \prod_{p|q} p^k \cdot \max |a(p, v)| \ll \prod_{p|q} \frac{p^r}{p-1} \leq C^{\omega(q)} \frac{q^k}{\phi(q)}.$$

Similarly, but without needing any combinatorial identity, we have

$$C(q) \ll \prod_{p|q} p \cdot \max |a(p, v)| \ll \prod_{p|q} \frac{p}{p-1} \leq C^{\omega(q)} \frac{q}{\phi(q)}.$$

Thus, the total contribution of $B(q)$ and $C(q)$ to (13.13) is

$$\ll C^{\omega(q)} \frac{q^k}{\phi(q)} \cdot \frac{N^{k-1}}{q^{k-1}} \ll N^{k-1} q^{\epsilon}.$$

Summing this over $q \leq x$ yields an error term of $N^{k-1} x^{1+\epsilon}$ in (13.12). Compared with our expected main term of N^k, this is negligible as long as $x < x^{1-\delta}$ for any $\delta > 0$.

Finally, we are left to evaluate $A(q)$. And here we are left with a beautiful surprise.

Proposition 13.9. *For any prime p, we have*

$$A(p) = \sum_{v=1}^{p} a(p,v)\binom{p}{v}\sigma(k,v) = 0.$$

Note that this settles the issue immediately. We have $A(q) = 0$ for $q > 1$, by multiplicativity, and $A(q) = 1$ for $q = 1$, again by multiplicativity. (The empty product equals 1.) Therefore, up to the error terms which we already bounded, all that survives in (13.12) is

$$\sum_{q=1} N^k,$$

where the sum is over only the single element $q = 1$ and, yup, that sum equals N^k. Subject to Proposition 13.9, we are done, and indeed by choosing $x = N^{1/2}$ we see that we obtain

$$\sum_{\substack{1 \le h_1,\dots,h_k \le N \\ \text{distinct}}} \mathcal{S}(h_1,\dots,h_k) = N^k + O(N^{k-1/2+\epsilon}). \tag{13.15}$$

Step 4. Two combinatorial identities. In this section we prove the combinatorial identity (13.14), together with an additional related identity, and then we use them to complete the proof of Proposition 13.9.

Proposition 13.10. *(1) The identity* (13.14) *holds; namely, we have*

$$\sum_{v=1}^{p} \binom{p}{v}\sigma(r,v) = p^r.$$

(2) We have

$$\sum_{v=1}^{p} v\binom{p}{v}\sigma(r,v) = p^{r+1} - (p-1)^r p.$$

Proof. To prove the first identity, classify the set of maps $\{1,\dots,r\} \to \{1,\dots,p\}$ by the size of their image. There are $\binom{p}{v}$ subsets of $\{1,\dots,p\}$ of size p, and for each there are $\sigma(r,v)$ surjections from $\{1,\dots,r\}$ onto it.

To prove the second, write

$$\binom{p}{v} = p\binom{p}{v} - p\binom{p-1}{v}$$

and use the first. □

Now, finally, we have

$$\begin{aligned}
A(p) &= \sum_{v=1}^{p} \frac{p^k - vp^{k-1} - (p-1)^k}{(p-1)^k}\binom{p}{v}\sigma(k,v) \\
&= (p-1)^{-k}\left[(p^k - (p-1)^k)\sum_{\nu=1}^{p}\binom{p}{v}\sigma(k,v) - p^{k-1}\sum_{v=1}^{p} v\binom{p}{v}\sigma(k,v)\right] \\
&= (p-1)^{-k}\left[(p^k - (p-1)^k)p^k - p^{k-1}(p^{k+1} - (p-1)^k p)\right] \\
&= (p-1)^{-k}\cdot 0,
\end{aligned}$$

so that we are done.

13.3. **The Poisson distribution.** Recall the statement of Theorem 13.6, which we now restate in probabilistic terms. Suppose that n is a randomly chosen integer between 1 and N. Then, under the hypotheses of that theorem, the probability that the interval $(n, n+h]$ contains exactly k primes is equal to

$$(1 + o_N(1)) \cdot \frac{\lambda^k}{k!} e^{-\lambda}.$$

In fact, this expression is well known within probability theory.

Definition 13.11. Let X be a discrete random variable. Then X obeys a *Poisson distribution with parameter* λ if we have (for $k = 0, 1, 2, \dots$)

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

This probability distribution models a wide range of situations, and in particular those which resemble the 'random' behavior of the primes. For example, three million people play the lottery. Each has a one-in-a-million chance of holding a winning ticket. How many winners do you expect?

To familiarize ourselves with this definition, let's do a bit of computation. First, we compute that

$$P(X \in \{0, 1, 2, \dots\}) = \sum_{k \geq 0} \frac{\lambda^k}{k!} e^{-\lambda} = e^{\lambda} \cdot e^{-\lambda} = 1.$$

Next, we compute the *expectation* of this random variable, given by

$$\begin{aligned} \mathbb{E}(X) = \sum_{k=0}^{\infty} k \cdot P(X = k) &= \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k}{k!} e^{-\lambda} \\ &= \sum_{k=0}^{\infty} \frac{\lambda^{k+1}}{k!} e^{-\lambda} = \lambda. \end{aligned}$$

Now, we will look at the nth moment, which is the expected value of X^n. It is

$$\begin{aligned} \mathbb{E}(X^n) = \sum_{k=0}^{\infty} k^n \cdot P(X = k) &= \sum_{k=0}^{\infty} k^n \cdot \frac{\lambda^k}{k!} e^{-\lambda} \\ &= \sum_{k=0}^{n} \sigma(n, k) \lambda^k, \end{aligned}$$

where as above $\sigma(n, k)$ is the Stirling number of the second kind – i.e., the number of surjections from $\{1, \dots, n\}$ to $\{1, \dots, k\}$. We will omit the proof of this identity, as proving it is not really our purpose here – although we enthusiastically encourage the dissatisfied reader to pause and attempt to prove this for herself.

The following is a special case of a result well known to probabilists.

Theorem 13.12. *Let X be a discrete random variable, and suppose that the moments $P(X^n)$ coincide with those of the Poisson distribution with parameter λ.*

Then, in fact, X satisfies the Poisson distribution with parameter λ.

Or, subject to appropriate conditions, 'a probability distribution is determined by its moments'. In fact we will use a stronger version of this theorem than stated: if the moments of a probability distribution *converge* to those of Poisson, then the distribution itself converges to Poisson.

For more on this beautiful and important topic, we refer our reader to **????**

13.4. **A sieve result.** We now *briefly* introduce a standard sort of result.

Theorem 13.13. *For each fixed $\lambda > 0$, we have*
(13.16)

$$\{n \leq N \ : n + d_1, n + d_2, \dots, n + d_k \text{ are all prime}\} \leq (1 + o_N(1))2^k k! \mathfrak{S}_{\{d_1,\dots,d_k\}} \cdot \frac{N}{(\log N)^k},$$

uniformly for all $d_1, \dots, d_k \in [0, \lambda \log N]$.

This is much easier to 'understand' than it may initially look. Recall that the Hardy-Littlewood prime tuple conjecture predicts that

$$\{n \leq N \ : n + d_1, n + d_2, \dots, n + d_k \text{ are all prime}\} \sim (1 + o_N(1))\mathfrak{S}_{\{d_1,\dots,d_k\}} \cdot \frac{N}{(\log N)^k},$$

essentially the same as (13.16), except without the additional 'fudge factor' of $2^k k!$. Since we are assuming the Hardy-Littlewood prime tuple conjecture, we *a fortiori* have (13.16). But it is worth knowing that (13.16) can be proved unconditionally. In this form it was proved by Klimov, and it is a very standard sort of result.

The basic idea here, which we will return to (at least briefly), is wrapped up in a machine called the *Selberg sieve*. To *find* primes of a special form – or to prove they exist – is often very difficult. But to bound their number, one can weaken the condition 'is prime'.

For example, consider the set

$$\{n \leq N \ : n + d_1, n + d_2, \dots, n + d_k \text{ all have no prime factors } \geq N^{1/100}\}.$$

This is much more tractable to count than (13.16). (Especially if one only asks for an *upper bound*, as this opens the door to useful tricks.) Up to an error of $N^{1/100}$, this serves as a stand-in for (13.16) – *if* we are content with an upper bound.

13.5. **Conclusion of the proof.** We now finish Gallagher's proof of Theorem 13.6. We have

$$\begin{aligned}\sum_{n \leq N} (\pi(n + h) - \pi(n))^k &= \sum_{n \leq N} \sum_{n < p_1, \dots, p_k \leq n+h} 1 \\ &= \sum_{r=1}^{k} \sigma(k, r) \sum \pi_{d_1,\dots,d_r}(N),\end{aligned}$$

where[4]

$$\pi_{d_1,\dots,d_r}(N) := \{n \leq N \ : n + d_1, n + d_2, \dots, n + d_r \text{ are all prime}\},$$

and where the inner sum is over all r-tuples $d_1 < \cdots < d_r$ with all $d_i \in [1, h]$.

[4]TO DO. Our use of notation isn't consistent with other sections; e.g. k and r are used with different meanings than in other sections. To fix.

We have, by the Hardy-Littlewood conjecture, that

$$\pi_{d_1,\dots,d_r}(N) \sim \mathfrak{S}(\{d_1,\dots,d_r\})\frac{N}{(\log N)^r}.$$

Therefore, applying Gallagher's technical result that the average of this singular series is 1, we have

$$\sum \pi_{d_1,\dots,d_r}(N) \sim \frac{h^r}{r!}\frac{N}{(\log N)^r}.$$

We thus conclude that

$$\frac{1}{N}\sum_{n\leq N}(\pi(n+h)-\pi(n))^k = (1+o_N(1))\frac{h^r}{r!(\log N)^r}\cdot\sum_{r=1}^{k}\sigma(k,r).$$

Plugging in $h=\lambda\log N$, we get that[5]

$$\frac{1}{N}\sum_{n\leq N}(\pi(n+h)-\pi(n))^k = (1+o_N(1))\cdot\sum_{r=1}^{k}\sigma(k,r)\frac{\lambda^r}{r!}.$$

We now return to our earlier question, and prove Corollary 13.7 – namely, that (under our Hardy-Littlewood hypothesis) we have

$$\#\left\{n\leq N \ : \ \frac{p_{n+1}-p_n}{\log n}\in[\alpha,\beta]\right\} = (1+o_N(1))N\int_\alpha^\beta e^{-t}dt.$$

This is now easy:

- The probability that there are no primes in $[p_n, p_n+\alpha\log n]$ is $\sim \frac{\alpha^0}{0!}e^{-\alpha}=e^{-\alpha}$.
- The probability that there are no primes in $[p_n+\alpha\log n, p_n+\beta\log n$ is similarly $e^{\alpha-\beta}$. So the probability that there is at least one prime is $1-e^{\alpha-\beta}$.
- Therefore, the probability that there are no primes in $[p_n, p_n+\alpha\log n]$, and at least one in $[p_n+\alpha\log n, p_n+\beta\log n]$, is asymptotic to

$$e^{-\alpha}\cdot(1-e^{\alpha-\beta}) = e^{-\alpha}-e^{-\beta} = \int_\alpha^\beta e^{-t}dt.$$

[5]There seems to be a different factor of $r!$ here.