Practical session 4

Exercise 1 For a natural number n let $\nu(n)$ be the number of primes which divide n (we do not count multiplicity though it would make little difference). Let $\omega(n)$ be a function which tends to infinity arbitrarily slowly as n tends to infinity. Then the number of integers $x \in \{1, 2, \ldots, n\}$ for which

$$|\nu(x) - \ln\ln n| > \omega(n)\sqrt{\ln\ln n} \tag{1}$$

is $o(n)$.

In the solution of the exercise we will use (without proof) the following fact from Number Theory.

Fact 0.1. *For any $n \in \mathbb{N}$ it holds that*

$$\sum_{p \le n} \frac{1}{p} = \ln\ln n + O(1),$$

where the sum extends over all prime numbers $p \le n$.

Solution

Choose x from $\{1, \ldots, n\}$ uniformly at random. The main idea is to show that the distribution of $\nu(x)$ is such that both its expectation and variance are roughly $\ln\ln n$. We could then apply Chebyshev's inequality to estimate the probability that x does not satisfy Equation (1). For every prime $p \le n$ let X_p be the indicator random variable for the event "$p|x$", namely

$$X_p = \begin{cases} 1 & p|x \\ 0 & p\nmid x. \end{cases}$$

Let $N = n^{1/10}$ and let $X = \sum_{p \le N} X_p$ denote the number of primes $p \le N$ that divide x. This specific choice of N is purely technical and will be clarified during the proof. We first prove the following simple claim

Claim 0.2. *For any integer $1 \le x \le n$, the number of primes $p > N$ which divide x is at most 10.*

Proof. Write

$$x = p_1^{k_1} \cdot \ldots \cdot p_m^{k_m} \le n$$

for primes $p_1, \ldots, p_m$, and let t be the number of primes p_i that are larger than N. Assume without loss of generality that those primes are $p_1, \ldots, p_t$. Then

$$n^{t/10} \le N^t \cdot p_{t+1}^{k_{t+1}} \cdot \ldots \cdot p_m^{k_m} \le x \le n,$$

implying that $t \le 10$. □

It follows from Claim 0.2 that $\nu(x) - 10 \le X \le \nu(x)$. Therefore, it suffices to prove that

$$\Pr\left(|X - \ln\ln n| > \omega(n)\sqrt{\ln\ln n}\right) = o(1)$$

holds for any function $\omega(n)$ which tends to infinity as n tends to infinity.

In the remainder of the proof we will show that for every $\lambda > 0$ it holds that

$$\Pr\left(|X - \ln\ln n| > \lambda\sqrt{\ln\ln n} + O(1)\right) < \lambda^{-2} + o(1).$$

We start with estimating $\mathbb{E}(X)$. For any prime p we have that

$$\mathbb{E}(X_p) = \Pr(p \text{ divides } x) = \frac{\lfloor n/p \rfloor}{n}.$$

Since $y - 1 \le \lfloor y \rfloor \le y$ for all $y \in \mathbb{R}$, it follows that

$$\frac{1}{p} - \frac{1}{n} \le \mathbb{E}(X_p) \le \frac{1}{p},$$

which we can write as

$$\mathbb{E}(X_p) = \frac{1}{p} - O(1/n).$$

Therefore

$$\mathbb{E}(X) = \sum_{p \le N} \frac{1}{p} - O(1/n) = \ln\ln n + O(1) - O(N/n) = \ln\ln n + O(1),$$

where the first equality holds by the linearity of expectation and the second equality holds by Fact 0.1.

Next, we estimate $\mathrm{Var}(X)$. Recall that

$$\mathrm{Var}(X) = \sum_{p \le N} \mathrm{Var}(X_p) + \sum_{p \ne q \le N} \mathrm{Cov}(X_p, X_q).$$

For any prime p it holds that

$$\mathrm{Var}(X_p) = \frac{\lfloor n/p \rfloor}{n}\left(1 - \frac{\lfloor n/p \rfloor}{n}\right) \le \frac{\lfloor n/p \rfloor}{n} \le \frac{1}{p}.$$

Therefore

$$\sum_{p \le N} \mathrm{Var}(X_p) \le \sum_{p \le N} \frac{1}{p} = \ln\ln n + O(1), \tag{2}$$

where the equality holds by Fact 0.1. We now show that the sum of covariances is very small, namely, we show that

$$\sum_{p \ne q \le N} \mathrm{Cov}(X_p, X_q) = o(1).$$

For distinct primes p and q we have that $X_p X_q = 1$ if and only if $p|x$ and $q|x$. Since p and q are primes, this happens if and only if $pq|x$. Hence

$$\begin{aligned}
\operatorname{Cov}(X_p, X_q) &= \mathbb{E}(X_p X_q) - \mathbb{E}(X_p)\,\mathbb{E}(X_q) \\
&= \frac{\lfloor n/pq \rfloor}{n} - \frac{\lfloor n/p \rfloor}{n} \cdot \frac{\lfloor n/q \rfloor}{n} \\
&\leq \frac{1}{pq} - \left(\frac{1}{p} - \frac{1}{n}\right) \cdot \left(\frac{1}{q} - \frac{1}{n}\right) \\
&= \frac{1}{np} + \frac{1}{nq} - \frac{1}{n^2} \\
&\leq \frac{1}{n}\left(\frac{1}{p} + \frac{1}{q}\right).
\end{aligned}$$

Thus

$$\begin{aligned}
\sum_{p \neq q \leq N} \operatorname{Cov}(X_p, X_q) &\leq \frac{1}{n} \sum_{p \neq q \leq N} \left(\frac{1}{p} + \frac{1}{q}\right) \\
&\leq \frac{2N}{n} \sum_{p \leq N} \frac{1}{p} \\
&\leq \frac{2N}{n} (\ln \ln n + O(1)) \\
&= \frac{2n^{1/10}}{n} (\ln \ln n + O(1)) \\
&= O(n^{-9/10} \cdot \ln \ln n) \\
&= o(1), \qquad (3)
\end{aligned}$$

where the third inequality is by Fact 0.1. Note that a similar calculation can be used to prove that $\sum_{p \neq q} \operatorname{Cov}(X_p, X_q) \geq -o(1)$ (do it!), but this is not necessary for our proof. Combining (2) and (3) shows that

$$\operatorname{Var}(X) = \sum_{p \leq N} \operatorname{Var}(X_p) + \sum_{p \neq q \leq N} \operatorname{Cov}(X_p, X_q) \leq \ln \ln n + O(1).$$

Given any $\lambda > 0$, it follows by Chebyshev's inequality that

$$\begin{aligned}
\Pr\left(|X - \ln \ln n| > \lambda \sqrt{\ln \ln n} + O(1)\right) &= \Pr\left(|X - \mathbb{E}(X)| > \lambda \sqrt{\ln \ln n} + O(1)\right) \\
&\leq \frac{\operatorname{Var}(X)}{\left(\lambda \sqrt{\ln \ln n} + O(1)\right)^2} \\
&\leq \frac{\ln \ln n + O(1)}{\left(\lambda \sqrt{\ln \ln n} + O(1)\right)^2} \\
&= \lambda^{-2} + o(1).
\end{aligned}$$

Since this was done for every $\lambda > 0$, we conclude that the number of integers $x \in \{1, 2, \ldots, n\}$ for which

$$|\nu(x) - \ln \ln n| > \omega(n) \sqrt{\ln \ln n}$$

is $o(n)$, for any function $\omega(n)$ which tends to infinity as n tends to infinity.