Proof. Let $x_1, \dots, x_m$ be a subset of $m = f(n)$ integers in $[n]$ all of whose sums are distinct. Let $I_1, \dots, I_m$ be independent random variables, each taking values 0 and 1 with equal probability 1/2. Consider the random variable $X = I_1 x_1 + \cdots + I_m x_m$. Then

$$\mathrm{E}[X] = \frac{x_1 + \cdots + x_m}{2} \quad \text{and} \quad \mathrm{Var}[X] = \frac{x_1^2 + \cdots + x_m^2}{4} \le \frac{n^2 m}{4}.$$

Setting $Y := X - \mathrm{E}[X]$ and using Chebyshev's inequality with $t := 2\sqrt{\mathrm{Var}[X]} \le n\sqrt{m}$, after reversing the inequality we obtain

$$\Pr[|Y| \le t] \ge 1 - \frac{1}{4} = 0.75.$$

On the other hand, due to the assumption that all sums of $x_1, \dots, x_m$ are distinct, the probability that X takes a particular value is either 0 or 2^{-m}. In particular, $\Pr[Y = s] \le 2^{-m}$ for every integer s in the interval $[-t, t]$. Since there are only $2t + 1$ such integers, the union bound implies that

$$\Pr[|Y| \le t] \le 2^{-m}(2t + 1).$$

Comparing the above inequalities and remembering that $t \le n\sqrt{m}$ leads to $0.75 \cdot 2^m \le 2t + 1 \le 2n\sqrt{m} + 1$, it follows that $2^m/\sqrt{m} \le Cn$ for a constant C, and the desired upper bound on $m = f(n)$ follows. □

21.3 Prime factors

Number theory has its foundation in the Fundamental Theorem of Arithmetic, which states that every integer $x > 1$ can be written uniquely in the form

$$x = p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r},$$

where the p_i's are primes and the k_i's are positive integers. Given x, we are interested in the number r of *prime factors* of x, that is, in the number of distinct primes p_i in such a representation of x. This number of primes dividing x is usually denoted by $\nu(x)$.

An important result in number theory, due to Hardy and Ramanujan (1917) states that almost every integer number between 1 and n has about $\ln\ln n$ prime factors. "Almost all" here means all but $o(n)$ numbers.

Theorem 21.3. *Let $\alpha = \alpha(n)$ be an arbitrarily slowly growing function. Then almost all integers x in $[n]$ satisfy $|\nu(x) - \ln\ln n| \le \alpha\sqrt{\ln\ln n}$.*

Proof (due to Turán 1934). Throughout this proof, let p, q denote prime numbers. We need two well known results from number theory, namely,

$$\sum_{p \le x} \frac{1}{p} \le \ln \ln x + O(1)\,, \tag{21.4}$$

$$\pi(x) = (1 + o(1)) \frac{x}{\ln x}\,, \tag{21.5}$$

where $\pi(x)$ denotes the number of primes smaller than x.

We now choose x randomly from the set $\{1, \ldots, n\}$. For prime p, let X_p be the indicator random variable for the event that p divides x, and let $X = \sum_{p \le x} X_p$; hence, $X = \nu(x)$.

Since x can be chosen in n different ways, and in $\lfloor n/p \rfloor$ cases it will be divisible by p, we have that

$$\mathrm{E}\left[X_p\right] = \frac{\lfloor n/p \rfloor}{n} \le \frac{1}{p}\,,$$

and by (21.4) we also have

$$\mathrm{E}\left[X\right] \le \sum_{p \le x} \frac{1}{p} \le \ln \ln n + O(1)\,.$$

Now we bound the variance

$$\mathrm{Var}\left[X\right] = \sum_{p \le x} \mathrm{Var}\left[X_p\right] + \sum_{p \ne q \le n} \mathrm{Cov}\left(X_p X_q\right) \le \mathrm{E}\left[X\right] + \sum_{p \ne q \le n} \mathrm{Cov}\left(X_p X_q\right)\,,$$

since $\mathrm{Var}\left[X_p\right] \le \mathrm{E}\left[X_p\right]$. Observe that $X_p X_q = 1$ if and only if both p and q divide x, which further implies that pq divides x. In view of this we have

$$\begin{aligned}
\mathrm{Cov}\left(X_p X_q\right) = \mathrm{E}\left[X_p X_q\right] - \mathrm{E}\left[X_p\right] \mathrm{E}\left[X_q\right] &= \frac{\lfloor n/(pq) \rfloor}{n} - \frac{\lfloor n/p \rfloor}{n} \cdot \frac{\lfloor n/q \rfloor}{n} \\
&\le \frac{1}{pq} - \Big(\frac{1}{p} - \frac{1}{n}\Big)\Big(\frac{1}{q} - \frac{1}{n}\Big) \\
&\le \frac{1}{n}\Big(\frac{1}{p} + \frac{1}{q}\Big)\,.
\end{aligned}$$

Then by (21.5)

$$\sum_{p \ne q \le n} \mathrm{Cov}\left(X_p X_q\right) \le \frac{2\pi(n)}{n} \sum_{p \le n} \frac{1}{p} = O\Big(\frac{\ln \ln n}{\ln n}\Big) \to 0\,.$$

Applying Chebyshev's inequality with $t = \alpha \sqrt{\ln \ln n}$ yields the desired result. □