The second moment method is an effective tool in number theory. Let $\nu(n)$ denote the number of primes p dividing n. (We do not count multiplicity though it would make little difference.) The folllowing result says, roughly, that "almost all" n have "very close to" $\ln \ln n$ prime factors. This was first shown by Hardy and Ramanujan in 1920 by a quite complicated argument. We give a remarkably simple proof of Paul Turan [1934], a proof that played a key role in the development of probabilistic methods in number theory.

Theorem 2.1 Let $\omega(n) \to \infty$ arbitrarily slowly. Then number of x in $\{1, \ldots, n\}$ such that

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

is $o(n)$.

Proof. Let x be randomly chosen from $\{1, \ldots, n\}$. For p prime set

$$X_p = \begin{cases} 1 & \text{if } p|x \\ 0 & \text{otherwise} \end{cases}$$

Set $M = n^{1/10}$ and set $X = \sum X_p$, the summation over all primes $p \leq M$. As no $x \leq n$ can have more than ten prime factors larger than M we have $\nu(x) - 10 \leq X(x) \leq \nu(x)$ so that large deviation bounds on X will translate into asymptotically similar bounds for ν. (Here 10 could be any large constant.) Now

$$E[X_p] = \frac{\lfloor n/p \rfloor}{n}$$

As $y - 1 < \lfloor y \rfloor \leq y$

$$E[X_p] = 1/p + O(1/n)$$

By linearity of expectation

$$E[X] = \sum_{p \leq M} \frac{1}{p} + O(\frac{1}{n}) = \ln \ln n + O(1)$$

Now we find an asymptotic expression for $Var[X] = \sum_{p \leq M} Var[X_p] + \sum_{p \neq q} Cov[X_p, X_q]$. As $Var[X_p] = \frac{1}{p}(1 - \frac{1}{p}) + O(\frac{1}{n})$,

$$\sum_{p \leq M} Var[X_p] = \sum_{p \leq M} \frac{1}{p} + O(1) = \ln \ln n + O(1)$$

With p, q distinct primes, $X_p X_q = 1$ if and only if $p|x$ and $q|x$ which occurs if and only if $pq|x$. Hence

$$\begin{array}{l} Cov[X_p, X_q] = E[X_p]E[X_q] - E[X_p X_q] \\ = \frac{\lfloor n/pq \rfloor}{n} - \frac{\lfloor n/p \rfloor}{n}\frac{\lfloor n/q \rfloor}{n} \\ \leq \frac{1}{pq} - (\frac{1}{p} - \frac{1}{n})(\frac{1}{q} - \frac{1}{n}) \\ \leq \frac{1}{n}(\frac{1}{p} + \frac{1}{q}) \end{array}$$

Thus

$$\sum_{p \neq q} Cov[X_p, X_q] \leq \frac{1}{n} \sum_{p \neq q} \frac{1}{p} + \frac{1}{q} \leq \frac{2M}{n} \sum \frac{1}{p}$$

Thus

$$\sum_{p \neq q} Cov[X_p, X_q] = O(n^{-9/10} \ln \ln n) = o(1)$$

That is, the covariances do not affect the variance, $Var[X] = \ln \ln n + O(1)$ and Chebyschev's Inequality actually gives

$$\Pr[|X - \ln \ln n| > \lambda \sqrt{\ln \ln n}] < \lambda^{-2} + o(1)$$

for any constant λ. As $|X - \nu| \leq 10$ the same holds for ν. □

In a classic paper Paul Erdős and Marc Kac [1940] showed, essentially, that ν does behave like a normal distribution with mean and variance $\ln \ln n$. Here is their precise result.

Theorem 2.2. Let λ be fixed, positive, negative or zero. Then

$$\lim_{n \to \infty} \frac{1}{n} |\{x : 1 \leq x \leq n, \nu(x) \geq \ln \ln n + \lambda \sqrt{\ln \ln n}\}| = \int_{\lambda}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

We outline the argument, emphasizing the similarities to Turan's proof. Fix a function $s(n)$ with $s(n) \to \infty$ and $s(n) = o((\ln \ln n)^{1/2})$ - e. g. $s(n) = \ln \ln \ln n$. Set $M = n^{1/s(n)}$. Set $X = \sum X_p$, the summation over all primes $p \leq M$. As no $x \leq n$ can have more than $s(n)$ prime factors greater than M we have $\nu(x) - s(n) \leq X(x) \leq \nu(x)$ so that it suffices to show Theorem 2.2 with ν replaced by X. Let Y_p be independent random variables with $\Pr[Y_p = 1] = p^{-1}$, $\Pr[Y_p = 0] = 1 - p^{-1}$ and set $Y = \sum Y_p$, the summation over all primes $p \leq M$. This Y represents an idealized version of X. Set

$$\mu = E[Y] = \sum_{p \leq M} p^{-1} = \ln \ln n + o((\ln \ln n)^{1/2})$$

and

$$\sigma^2 = Var[Y] = \sum_{p \le M} p^{-1}(1-p^{-1}) \sim \ln\ln n$$

and define the normalized $\tilde{Y} = (Y-\mu)/\sigma$. From the Central Limit Theorem (well, an appropriately powerful form of it!) $\tilde{Y}$ approaches the standard normal N and $E[\tilde{Y}^k] \to E[N^k]$ for every positive integer k. Set $\tilde{X} = (X-\mu)/\sigma$. We compare $\tilde{X}, \tilde{Y}$.

For any distinct primes $p_1, \ldots, p_s \le M$

$$E[X_{p_1} \cdots X_{p_s}] - E[Y_{p_1} \cdots Y_{p_s}] = \frac{\lfloor \frac{n}{p_1 \cdots p_s} \rfloor}{n} - \frac{1}{p_1 \cdots p_s} = O(n^{-1})$$

We let k be an arbitrary fixed positive integer and compare $E[\tilde{X}^k]$ and $E[\tilde{Y}^k]$. Expanding, $\tilde{X}^k$ is a polynomial in X with coefficients $n^{o(1)}$. Further expanding each $X^j = (\sum X_p)^j$ - always reducing X_p^a to X_p when $a \ge 2$ - gives the sum of $O(M^k) = n^{o(1)}$ terms of the form $X_{p_1} \cdots X_{p_s}$. The same expansion applies to $\tilde{Y}$. As the corresponding terms have expectations within $O(n^{-1})$ the total difference

$$E[\tilde{X}^k] - E[\tilde{Y}^k] = n^{-1+o(1)} = o(1)$$

Hence each moment of $\tilde{X}$ approach that of the standard normal N. A standard, though nontrivial, theorem in probability theorem gives that $\tilde{X}$ must therefore approach N in distribution. □

We recall the famous quotation of G. H. Hardy:

> 317 is a prime, not because we think so, or because our minds are shaped in one way rather than another, but *because it is so*, because mathematical reality is built that way.

How ironic - though not contradictory - that the methods of probability theory can lead to a greater understanding of the prime factorization of integers.