The Kubilius probalistic model of integers

3. 3-9-06

3-4-15
#19

The prob. that a random integer, $1 \le n \le x$, is divisible by prime p is $\lfloor \frac{x}{p} \rfloor / \lfloor x \rfloor = \frac{1}{p} + O(\frac{1}{x})$. The events that a random $n \in [1, x]$ is divisible by primes p and q are nearly independent, e.g.

$$\mathbb{P}[\, p|n \text{ and } q|n \,] = \frac{1}{\lfloor x \rfloor} \lfloor \frac{x}{pq} \rfloor \approx \frac{1}{pq} \approx \mathbb{P}[p|n]\, \mathbb{P}[q|n]. \qquad (\mathbb{P} = \text{probability})$$

But, if say $p > \sqrt{x}$ and $q > \sqrt{x}$ then $\mathbb{P}[\, p|n \text{ and } q|n \,] = 0$.

Kubilius model (simple version)

random variables X_p, p-primes, independent

$$\mathbb{P}[X_p = 1] = \frac{1}{p}, \quad \mathbb{P}[X_p = 0] = 1 - \frac{1}{p}$$

X_p approximates the prob. that a random integer is divis. by p.

expectation $\mathbb{E}[X_p] = \frac{1}{p}$

variance $\underbrace{\mathbb{V}[X_p]}_{\mathbb{E}[(X_p - \frac{1}{p})^2]} = \frac{1}{p}\left(1 - \frac{1}{p}\right)^2 + \left(1 - \frac{1}{p}\right)\left(\frac{1}{p}\right)^2 = \frac{1}{p} - \frac{1}{p^2}$.

Sums

introduce r.v. $S_z = \sum_{p \le z} X_p$: models the number of prime divisors that are $\le z$ of a random integer.

expectation

$$\mathbb{E}[S_z] = \sum_{p \le z} \mathbb{E}[X_p] = \sum_{p \le z} \frac{1}{p} = \log\log z + O(1).$$

so, on average we should expect $\omega(n, z) \approx \log\log z$.

variance

$$\mathbb{V}[S_z] = \sum_{p \le z} \mathbb{V}[X_p] = \sum_{p \le z} \left(\frac{1}{p} - \frac{1}{p^2}\right) = \log\log z + O(1)$$

$$= \mathbb{E}\left[\left(S_z - \sum_{p \le z} \frac{1}{p}\right)^2\right]$$

thus, we should expect $(\omega(n,z) - \log\log z)^2 \approx \log\log z$ on average, i.e. $\omega(n, z) = \log\log z + O(\sqrt{\log\log z})$ most of the time.

12. 9-19-11

Refinement of the model - include prime powers

Y_{p^k} : $\mathbb{P}[Y_{p^k}=1] = \frac{1}{p^k} - \frac{1}{p^{k+1}}$

$$\mathbb{P}[Y_{p^k}=0] = 1 - \left(\frac{1}{p^k} - \frac{1}{p^{k+1}}\right)$$

$$\mathbb{E}[Y_{p^k}] = \frac{1}{p^k} - \frac{1}{p^{k+1}}$$

For fixed p, these are mutually exclusive

$$\mathbb{V}[Y_{p^k}] = \left(\frac{1}{p^k} - \frac{1}{p^{k+1}}\right)\left(1 - \frac{1}{p^k} + \frac{1}{p^{k+1}}\right)$$

random variable $T_z = \sum_{\substack{p \le z \\ k \ge 1}} Y_{p^k}$ models behavior of $\Omega(n,z)$.

9-21-2011 (new)

For $k \in \mathbb{N}$, introduce new random events
$$E_k(z) = \{ \forall p^\alpha \| k,\ Y_{p^\alpha} = 1 \text{ and } \forall p \nmid k \text{ with } p \le z,\ X_p = 0 \}.$$
Likewise, set
$$D_k(x,z) = \{ n \le x : k | n,\ P^-(n/k) > z \} = \{ n \le x : p^\alpha \| k \Rightarrow p^\alpha \| n,\ p \nmid k \text{ and } p \le z \Rightarrow p \nmid n \}$$
for $P^+(k) \le z$, and let $N_k(x,z) = |D_k(x,z)|$.

<u>Theorem K</u> (Kubilius model)

Let $\mathcal{K} \subseteq \{ k : P^+(k) \le z \}$, $x \ge z \ge e^3$. Then
$$\mathbb{P}(E_k(z) \text{ for some } k \in \mathcal{K}) = \frac{1}{x} \sum_{k \in \mathcal{K}} N_k(x,z) + O(e^{-u}), \quad u = \frac{\log x}{\log z},$$
uniformly in $x, z, \mathcal{K}$.

<u>Lemma K1</u> Uniformly for $e^3 \le z \le x$, $P^+(k) \le z$ and $k \le \sqrt{x}$,
$$N_k(x,z) = \Phi\left(\frac{x}{k}, z\right) = \frac{x}{k} \prod_{p \le z} \left(1 - \frac{1}{p}\right) \left\{ 1 + O(e^{-u}) \right\}, \quad u = \frac{\log x}{\log z}$$

<u>Proof</u> Use the Fundamental Lemma from sieve methods (e.g. Theorem BH.4 from my Sieve Methods notes). This gives
$$\Phi\left(\frac{x}{k}, z\right) = \frac{x}{k} \prod_{p \le z} \left(1 - \frac{1}{p}\right) \left\{ 1 + O(s^{-\frac{1}{3}s}) \right\} + O(D),$$
where D is arbitrary subject to $D \ge z$, and $s = \frac{\log D}{\log z}$.
We take $D = x^{1/10}$. Then $s^{-\frac{1}{3}s} \ll e^{-10s} = e^{-u}$ and
$$e^{-u} \frac{x}{k} \prod_{p \le z} \left(1 - \frac{1}{p}\right) \gg x^{-1/3} \cdot \sqrt{x} (\log z)^{-1} \gg x^{1/10} = D. \blacksquare$$

3-6-15 #20

<u>Lemma K2</u> Uniformly for $x \ge z \ge e^3$,
$$\sum_{\substack{k > \sqrt{x} \\ P^+(k) \le z}} \frac{1}{k} \ll e^{-u} \log z, \quad u = \frac{\log x}{\log z}.$$

<u>Proof</u> Let $\alpha = \frac{2}{\log z}$, so that $0 < \alpha \le \frac{2}{3}$. The sum ~~in question~~ is

OK
$$\le \sum_{P^+(k) \le z} \left(\frac{k}{\sqrt{x}}\right)^\alpha \frac{1}{k} = x^{-\alpha/2} \prod_{p \le z} \left(1 + \frac{p^\alpha}{p(1 - p^{\alpha - 1})}\right)$$
$$\le e^{-u} \exp\left\{ \sum_{p \le z} \frac{p^{\alpha-1}}{1 - p^{\alpha-1}} \right\}.$$

As $p^\alpha = e^{\alpha \log p} = 1 + O(\alpha \log p)$ since $\alpha \log p \le 2$,
the above sum on p is $\displaystyle\sum_{p \le z} \frac{1}{p}(1 + O(\alpha \log p))(1 + O(p^{-1/3})) = \log_2 z + O(1)$. $\blacksquare$

Proof of Theorem K

First,

$$\mathbb{P}(E_k(z) \text{ for some } k \in \mathcal{K}) = \sum_{\substack{k \in \mathcal{K} \\ k \le \sqrt{x}}} \mathbb{P}(E_k(z)) + O\Big(\sum_{\substack{k > \sqrt{x} \\ p^+(k) \le z}} \mathbb{P}(E_k(z))\Big)$$

Since $\mathbb{P}(E_k(z)) = \prod_{p^\alpha \| k} \left(\frac{1}{p^\alpha} - \frac{1}{p^{\alpha+1}}\right) \prod_{\substack{p \le z \\ p \nmid k}} \left(1 - \frac{1}{p}\right) = \frac{1}{k} \prod_{p \le z} \left(1 - \frac{1}{p}\right)$,

$$\sum_{\substack{k > \sqrt{x} \\ p^+(k) \le z}} \mathbb{P}(E_k(z)) \ll \frac{1}{\log z} \sum_{\substack{k > \sqrt{x} \\ p^+(k) \le z}} \frac{1}{k} \ll e^{-u}$$

by Lemma K2. Therefore,

$$(1) \qquad \mathbb{P}(E_k(z) \text{ for some } k \in \mathcal{K}) = \sum_{\substack{k \in \mathcal{K} \\ k \le \sqrt{x}}} \mathbb{P}(E_k(z)) + O(e^{-u}).$$

Next, by Lemma K1,

$$\sum_{k \in \mathcal{K}} N_k(x,z) = \sum_{\substack{k \in \mathcal{K} \\ k \le \sqrt{x}}} N_k(x,z) + O\Big(\sum_{\substack{k > \sqrt{x} \\ p^+(k) \le z}} N_k(x,z)\Big)$$

$$(2) \qquad = x \prod_{p \le z}\left(1 - \frac{1}{p}\right)\{1 + O(e^{-u})\} \sum_{\substack{k \in \mathcal{K} \\ k \le \sqrt{x}}} \frac{1}{k} + O\Big(\sum_{\substack{k > \sqrt{x} \\ p^+(k) \le z}} N_k(x,z)\Big)$$

$$= x\{1 + O(e^{-u})\} \sum_{\substack{k \in \mathcal{K} \\ k \le \sqrt{x}}} \mathbb{P}(E_k(z)) + O\Big(\sum_{\substack{k > \sqrt{x} \\ p^+(k) \le z}} N_k(x,z)\Big).$$

Since $\sum_{p^+(k) \le z} N_k(x,z) = \lfloor x \rfloor$, by Lemma K1,

$$\lfloor x \rfloor - \sum_{\substack{k > \sqrt{x} \\ p^+(k) \le z}} N_k(x,z) = \sum_{\substack{k \le \sqrt{x} \\ p^+(k) \le z}} N_k(x,z) = x \prod_{p \le z}\left(1 - \frac{1}{p}\right) \sum_{\substack{k \le \sqrt{x} \\ p^+(k) \le z}} \frac{1}{k}\{1 + O(e^{-u})\}$$

$$(3) \qquad = x\{1 + O(e^{-u})\}\Big(1 - \prod_{p \le z}\left(1 - \frac{1}{p}\right) \sum_{\substack{k > \sqrt{x} \\ p^+(k) \le z}} \frac{1}{k}\Big)$$

$$= x\{1 + O(e^{-u})\}\big(1 - O(e^{-u})\big).$$

Therefore, by (1), (2) and (3),

$$\frac{1}{x} \sum_{k \in \mathcal{K}} N_k(x,z) = (1 + O(e^{-u}))\Big(\sum_{\substack{k \in \mathcal{K} \\ k \le \sqrt{x}}} \mathbb{P}(E_k(z)) + O(e^{-u})\Big)$$

$$= \mathbb{P}(E_k(z) \text{ for some } k \in \mathcal{K}) + O(e^{-u}). \blacksquare$$

14.
9-21-11

Theorem LCLT (Lyapunov, 1901)

Suppose $X_1, X_2, \dots$ are independent random variables with $\mathbb{E} X_i = \mu_i$, $\mathbb{V} X_i = \sigma_i^2$ for each i, and define $S_n = X_1 + \dots + X_n$,

$$m_n = \mu_1 + \dots + \mu_n = \mathbb{E} S_n, \quad s_n^2 = \sigma_1^2 + \dots + \sigma_n^2 = \mathbb{V} S_n.$$

If $s_n \to \infty$ as $n \to \infty$ and

$$\text{(Lya)} \quad \lim_{n\to\infty} \frac{1}{s_n^3} \sum_{j=1}^{n} \mathbb{E}\left(|X_i - \mu_i|^3\right) = 0,$$

then for every $\lambda \in \mathbb{R}$,

$$\text{(Norm)} \quad \lim_{n\to\infty} \mathbb{P}\left(S_n \le m_n + \lambda s_n\right) = \Phi(\lambda) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\lambda} e^{-\frac{1}{2}t^2}\, dt.$$

Corollary UBCLT

If $X_1, X_2, \dots$ are uniformly bounded random variables ($\exists\, C$ s.t. $\mathbb{P}(|X_i| \le C) = 1$ for every i) and $s_n \to \infty$ as $n \to \infty$, then (Norm) holds.

Proof if $\mathbb{P}(|X_i| \le C) = 1$ for all i, then $|\mu_i| \le C$ for all i and

$$\mathbb{E}\left(|X_i - \mu_i|^3\right) \le \mathbb{E}\left(2C|X_i - \mu_i|\right) = 2C\sigma_i^2.$$

Therefore,

$$\sum_{j=1}^{n} \mathbb{E}\left(|X_i - \mu_i|^3\right) \le 2C s_n^2$$

and (Lya) follows from $s_n \to \infty$ as $n \to \infty$.

Corollary S_z

Let X_p (p prime) be the random variables from the (basic) Kubilius model, and let $S_z = \sum_{p \le z} X_p$. If $\lim_{z\to\infty} g(z) = \lambda$, then

$$\lim_{z\to\infty} \mathbb{P}\left(S_z \le \log_2 z + g(z)\sqrt{\log_2 z}\right) = \Phi(\lambda).$$

Proof Let $\lambda' < \lambda < \lambda''$. Since $\mathbb{E} S_z = \log_2 z + O(1)$ and $\mathbb{V} S_z = \log_2 z + O(1)$, for large enough z we have

$$\mathbb{P}(S_z \le \mathbb{E}S_z + \lambda'\sqrt{\mathbb{V}S_z}) \le \overbrace{\mathbb{P}(S_z \le \log_2 z + g(z)\sqrt{\log_2 z})}^{f(z)} \le \mathbb{P}(S_z \le \mathbb{E}S_z + \lambda''\sqrt{\mathbb{V}S_z}).$$

By Corollary UBCLT, we find that

$$\limsup_{z\to\infty} f(z) \le \Phi(\lambda''), \quad \liminf_{z\to\infty} f(z) \ge \Phi(\lambda').$$

As λ' and λ'' are arbitrary and Φ is continuous, $\lim_{z\to\infty} f(z) = \Phi(\lambda)$. ∎

Theorem (Erdős-Kac, 1939)

For each fixed λ,

$$\lim_{x\to\infty} \frac{1}{x}\#\{n\le x : \omega(n) \le \log_2 x + \lambda\sqrt{\log_2 x}\} = \Phi(\lambda).$$

Proof Let $f(x) = \frac{1}{x}\#\{n \le x : \omega(n) \le \log_2 x + \lambda\sqrt{\log_2 x}\}$. Put $z = x^{1/\log_3 x}$. Then $\omega(n) - \log_3 x \le \omega(n,z) \le \omega(n)$. By Theorem K,

$$(*) \qquad \frac{1}{x}\#\{n\le x : \omega(n,z) \le y\} = \mathbb{P}(S_z \le y) + O\left(\frac{1}{\log_2 x}\right),$$

uniformly for $y \ge 0$. Here we take $\mathcal{K} = \{k : P^+(k) \le z,\ \omega(k) \le y\}$, so that

$$\sum_{k\in\mathcal{K}} N_k(x,z) = \#\{n\le x ; \omega(n,z) \le y\}, \quad \bigcup_{k\in\mathcal{K}} E_k(z) = (S_z \le y).$$

3-9-15 #21

We have $\log_2 x = \log_2 z + O(\log_4 z)$. If $\omega(n) \le \log_2 x + \lambda\sqrt{\log_2 x}$, then $\omega(n,z) \le \log_2 x + \lambda\sqrt{\log_2 x} = \log_2 z + g_1(z)\sqrt{\log_2 z}$, where $g_1(z) \to \lambda$ $(z\to\infty)$. By $(*)$, for large z,

$$(f_1) \qquad f(x) \le \mathbb{P}(S_z \le \log_2 z + g_1(z)\sqrt{\log_2 z}) + O\left(\frac{1}{\log_2 x}\right).$$

Likewise, defining $g_2(z) = \lambda - \frac{\log_3 z}{\sqrt{\log_2 z}}$, if $\omega(n,z) \le \log_2 z + g_2(z)\sqrt{\log_2 z}$, then

$$\omega(n) \le \log_2 z - \log_3 z + \lambda\sqrt{\log_2 z} + \log_3 x \le \log_2 x + \lambda\sqrt{\log_2 x}.$$

By $(*)$ this implies that

$$(f_2) \qquad f(x) \ge \mathbb{P}(S_z \le \log_2 x + g_2(z)\sqrt{\log_2 z}) + O\left(\frac{1}{\log_2 x}\right).$$

Finally by (f_1) (f_2) ($\overset{\text{cor.}}{\text{and}}$ S_z) and the Squeeze Theorem, $\lim f(x) = \Phi(\lambda)$. ∎