11.3.2021

3. Generalizations to other "random integers"

Question: what if we replace, in the Erdős-Kac Th, the integers in

$$\Omega_N = \{1, \dots, N\}$$

by other "random" integers? Do we still have convergence in law? With the same limit?

Example 1 - Fix $f \in \mathbb{Z}[X]$, monic, $\deg f \geq 2$

Take $\Omega_N = \{f(1), \dots, f(N)\} \subset \mathbb{Z}$

and probability measure

$$\mathbb{P}_N(x) = \frac{1}{N} \sum_{\substack{1 \leq n \leq N \\ f(n) = x}} 1$$

Define $X_N(m) = \omega(m)$ for $m \in \Omega_N$.

What about the limiting behavior?

[Keep Ω_N as before, but look at $\omega(f(n))$]

Basic Question is: does Th. 1.3.3 extend, in other words study $f(n) \bmod q$ as n varies for $q \geq 1$.

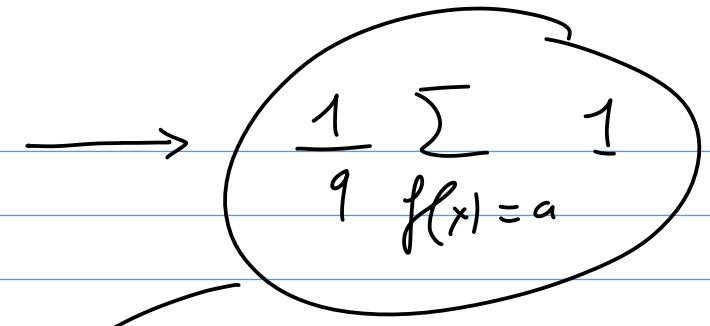
Fix $a \in \mathbb{Z}/q\mathbb{Z}$;

$$\frac{1}{N} \left| \{ n \leq N \mid f(n) \equiv a \bmod q \} \right|$$

$$= \frac{1}{N} \sum_{\substack{n \leq N \\ f(n) \equiv a \bmod q}} 1$$

$$= \sum_{\substack{x \in \mathbb{Z}/q\mathbb{Z} \\ f(x) = a}} \frac{1}{N} \underbrace{\sum_{\substack{n \leq N \\ n \equiv x \,(q)}} 1}_{\frac{1}{q} + O\left(\frac{1}{N}\right)}$$

$$= \frac{1}{q} \sum_{\substack{x \in \mathbb{Z}/q\mathbb{Z} \\ f(x) = a}} 1 + O\left(\frac{\text{nb. of roots of } f(x) = a \,(q)}{N} \right)$$



$\hookrightarrow \mu_f(\{a\})$ where μ_f is the image of the unif. measure mod q by $f: \mathbb{Z}/q\mathbb{Z} \to \mathbb{Z}/q\mathbb{Z}$

Key point : in terms of uniformity w.r.t. N and q, this is not as good as Th. 1.3.3 in applications!

[The support of $\mathbb{P}_N$ has $\sim N$ elements of size up to $N^{\deg(f)}$]

Consequence: (1) No problem for Erdős-Kac because we only used $q \leq Q^k = N^{k/(\log\log N)^{1/3}}$

(2) Consider the question

$$\lim_{N\to\infty} \mathbb{P}_N(n \text{ is squarefree}) = ?$$

$\hookrightarrow \forall p,\ p^2 \nmid n$

One can show (with Möbius function) that for $1 \leq n \leq N$ uniform, we get

$$\text{Proba} \longrightarrow \underbrace{\prod_p \left(1 - \frac{1}{p^2}\right)}_{\text{natural heuristically}} \left(= \frac{6}{\pi^2}\right)$$

because $p^2 \mid n$ in $\{1, \dots, N\}$ with prob. $\simeq \frac{1}{p^2}$.

For values of f, it turns out to be quickly very difficult: There is not a single (irreducible) f of degree ≥ 4 where the answer is known.

Example 2 - "shifted primes"

$$\Omega_N = \{ p - 1 \mid p \leq N \text{ prime} \}$$

$$\mathbb{P}_N(p-1) = \frac{1}{\pi(N)}$$

$\pi(x) = \sum_{p \leq x} 1$ (prime counting function)

$$X_N(p-1) = \omega(p-1)$$

What is the asymptotic behavior?

$p-1$ has some unusual features: $p-1$ is even except when p is 2; $p-1$ is not $\equiv 2 \mod 3$, ...

Q. Are there other more hidden constraints on $p-1$?

Analogue of Th. 1.3.3 asks about

$$\frac{1}{\pi(N)} \sum_{\substack{p \le N \\ p-1 \equiv a \bmod q}} 1 = \frac{1}{\pi(N)} \sum_{\substack{p \le N \\ p \equiv a+1 \bmod q}} 1$$

Not.

$$= \frac{\pi(N; q, a+1)}{\pi(N)}$$

Th. (Dirichlet; de la Vallée-Poussin $\simeq 1896$)

If $(b, q) = 1$, then

$$\frac{1}{\pi(N)} \pi(N; q, b) \xrightarrow[N \to \infty]{} \frac{1}{\varphi(q)} = \frac{1}{|(\mathbb{Z}/q\mathbb{Z})^\times|}$$

This means that the qualitative version of Th. 1.3.3 extends with limit the uniform measure on $(\mathbb{Z}/q\mathbb{Z})^\times$.

On the other hand the uniformity in the quantitative version is known only for $q \leq (\log N)^A$ (for $A > 0$ arbitrary)

[Siegel – Walfisz ; $\simeq$ 1935]

The extension to $q \leq N^\delta$, $\delta > 0$, is essentially equivalent to proving "almost" the Generalized Riemann Hypothesis for Dirichlet L-functions.

→ see Chapter III

Scheduling: next Monday, exercise class
next Thursday, next lecture

Exemple 3. Look at a discrete group, e.g. $G = SL_m(\mathbb{Z})$, $m \geq 2$ and f = pol. function of the entries

(e.g. $f = a_{1,1}$; $f = Tr$; ...)

"Random integers": values of $f(\gamma)$ for $\gamma \in G$ taken uniformly over $\gamma \in G$ with $|a_{ij}| \leq N$.

Q.: What about $\omega(Tr(\gamma))$ for such γ, as $N \to \infty$?

Extending Th. 1.3.3 is not too hard qualitatively; extending it quantitatively with sufficient uniformity (say for Erdős-Kac) is directly related to things like

- Property (T) of Kazhdan
- Homogeneous dynamics
- Expander graphs

4_ Convergence without normalization

Erdős-Kac: $\dfrac{X_N - \log\log N}{\sqrt{\log\log N}} \xrightarrow{\text{law}} N(0,1)$

where $X_N(n) = \omega(n)$, $n \in \Omega_N = \{1, \dots, N\}$.

[The normalization removes any individuality to the r.v.

One might ask if one can replace this by other statements really involving just X_N.

This is indeed possible.

Probabilistically the idea is to compare X_N more directly with

$$X_N = \sum_{p \le N} B_p \quad , \quad (B_p) \text{ indep.}$$

X_N is very close to a Poisson random variable with parameter $\sim \log\log N$

[Recall: Y r.v. is Poisson (λ) if

$$\mathbb{P}(Y = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \geq 0 \ ]$$

A comparison of X_N with Poisson $(\log\log N)$ r.v. was done by Rényi - Turán: they proved

$$\frac{1}{N} \sum_{n \leq N} e^{it w(n)} = (\log N)^{e^{it}-1} \times \Phi(t) \times (1+o(1))$$

$$\mathbb{E}_N\left(e^{it X_N}\right) \qquad \text{(uniformly for } t \in \mathbb{R})$$

and the point is that

$$(\log N)^{e^{it}-1} = \exp\left((e^{it}-1)\log\log N\right)$$

$$= \mathbb{E}\left(e^{it\,\mathrm{Po}(\log\log N)}\right)$$

Easy to see: $\dfrac{\mathrm{Po}(\lambda) - \lambda}{\sqrt{\lambda}} \xrightarrow{\text{law}} N(0,1)$

What is $\Phi(t)$?

$$\Phi(t) = \underbrace{\frac{1}{\Gamma(e^{it})}}_{\Phi_1(t)} \underbrace{\prod_p \left(1-\frac{1}{p}\right)^{e^{it}} \left(1+\frac{e^{it}}{p-1}\right)}_{\Phi_2(t)}$$

K. - Niheghbali (~ 2010) observed :

$\Phi_2(t)$ arises in the formula

$$\mathbb{E}\left(e^{it\sum_{p\le N} B_p}\right) = (\log N)^{e^{it}-1} \Phi_2(t)(1+o(1))$$

$$\prod_{p\le N}\left(\frac{1}{p}e^{it}+1-\frac{1}{p}\right)$$

$\Phi_1(t)$ arises in the formula

$$\mathbb{E}\left(e^{it Y_N}\right) = N^{e^{it}-1} \Phi_1(t)(1+o(1))$$

$$\uparrow$$
$$P_o(\log N)$$

where Y_N = nb. of cycles (counting fixed points) of a uniformly chosen $\sigma \in S_N$

[i.e. $\frac{1}{N!} \sum_{\sigma \in S_N} e^{it\,\ell(\sigma)} = \dots$]

$\ell(\sigma)$ = nb. of cycles

Q. What is the "real" meaning of such "mixtures" of $\sum_{p \le N} B_p$ and Y_N ?

[At least probably not a coincidence...]

Reference: A. Granville, "The anatomy of integers and permutations"

Chapter III (Ch. 5 in the notes)

Chebychev Bias

1853: Chebychev wrote a letter stating "there seem to be more primes $p \leq x$ s.t. $p \equiv 3 \ (4)$ than s.t. $p \equiv 1 \ (4)$."

Strongest interpretation

$$\pi(x; 4, 3) > \pi(x; 4, 1)$$

for $x \geq 3$ is **false** (Leech, 1957, $x = 26881$).

But nevertheless, one can show:

Theorem (Rubinstein – Sarnak; 1994)

If one assumes GRH (+ another very likely conjecture) then

$$\frac{1}{\log X} \int_1^X \mathbb{1}_{(\pi(t;4,3) > \pi(t;4,1))} \frac{dt}{t}$$

$$\longrightarrow_{X \to \infty} 0,99\ldots$$

This confirms quantitatively the observation of Chebychev !