Erdös-Kac theorem revisited

Xiaochuan Yang (University of Bath)

March 22, 2021, PIMS, Bath

Joint work with Louis H. Y. Chen (NUS) and Arturo Jaramillo (CIMAT)

- 1940. **Erdös and Kac** initiated a new chapter in number theory with the following CLT

$$Z_n := (\omega(J_n) - \log\log(n))/\sqrt{\log\log(n)} \xrightarrow{\mathcal{L}} N(0,1),$$

where $J_n \sim \mathrm{Unif}[n] = \{1, ..., n\}$ and $\omega(k)$ counts the number of distinct prime divisors of k, e.g. $\omega(24) = \omega(2^3 \times 3) = 2$.

- The original proof is mostly counting, aka sieve method.
- The error (measured in d_{K}) can be assessed optimally by sharp estimates of the characteristic function, where

$$d_{\mathrm{K}}(Z_n, N) = \sup_x |\mathbb{P}[Z_n \leq x] - \mathbb{P}[N \leq x]|.$$

- In this talk, I will present a purely probabilistic perspective for proving the Erdös-Kac theorem for general additive functions, with optimal non-asymptotic error and no use of characteristic functions.

Outline

1. Review of the problem
2. Key observation
3. Results
4. Proof

Error bounds by analytic methods

- **Erdös and Kac (1940)** $d_{\mathrm{K}}(Z_n, N(0,1)) \to 0$.
- **LeVeque (1949)** conjectured the best rate to be $C \log\log(n)^{-1/2}$.
- **Rényi et Turán (1958)** proved it.

  Tools: Perron's formula, Dirichlet series, estimate of ζ around $\{z \in \mathbb{C} \ ; \ \mathrm{Re}(z) = 1\}$, several sophisticated contour integrals and

Esseen's inequality

$$d_{\mathrm{K}}(X, N(0,1)) \leq c_1 \int_{-T}^{T} \left| \mathbb{E}[e^{iuX}] - e^{-u^2/2} \right| \frac{du}{u} + \frac{c_2}{T}.$$

- **Barbour, Kowalski and Nikeghbali (2014)** obtained the optimal error bound by studying the regularity of

$$\mathfrak{g}(\lambda) := \lim_{n\to\infty} \frac{\mathbb{E}[e^{\mathbf{i}\lambda Z_n}]}{\mathbb{E}[e^{\mathbf{i}\lambda U_n}]},$$

where U_n is a well-chosen signed measure (mod-ϕ convergence).

- Using sharp estimate of the characteristic function, **Chhaibi, Delbaen, Méliot and Nikeghbali (2020)** proved mod-Poisson convergence with exact asymptotic

$$\begin{aligned} d_{\mathrm{TV}}(\omega(J_n), M_n) &= \sup |\mathbb{P}[\omega(J_n) \in A] - \mathbb{P}[M_n \in A]| \\ &\sim \frac{c}{\log\log(n))^{1/2}}, \end{aligned}$$

where $M_n \sim \mathrm{Po}(\log\log n)$ and c is known.

Probabilistic attempts

Basic identity

$$\omega(J_n) = \sum_{p \in \mathcal{P}_n} \mathbb{1}(p|J_n) = \sum_{p \leq r_n} \mathbb{1}(p|J_n) + R.$$

- **Billingsley (1969)** method of moments and truncation.
- **Kubilius (1964)** obtained total variation bounds between

$$(\mathbb{1}(p|J_n), p \leq r_n) \text{ and } (\mathrm{Ber}(1/p), p \leq r_n)$$

for the first primes in the series. From this, he deduced

$$d_{\mathrm{K}}(Z_n, N(0,1)) \leq C \frac{\log\log\log(n)}{\sqrt{\log\log(n)}}.$$

The main error comes from R.

- **Harper (2009)** used Chen-Stein method and truncation to obtain the same bound as the one by **Kubilius (1964)**.
- **Arratia (2010)** constructed an elegant coupling

$$\mathbb{P}(J_n = T_n P_n) \geq 1 - \frac{C \log\log(n)}{\log(n)},$$

where T_n est a functional of the **size-biased permutation** of primes and P_n is a random prime. The optimal Wasserstein bound follows, and the argument cannot be extended to obtain Kolmogorov bounds.

- We find a coupling with the harmonic distribution in place of T_n,

$$\mathbb{P}(H_n = k) = \frac{1}{kL_n}, \quad k \in \{1, ..., n\},$$

which turns out to have transparent divisibility structure.

Key message of the talk

The harmonic distribution has very friendly divisibility structure,

and it is at most one prime away from the uniform.

Notation

- Define $\alpha_p(k)$ by the relation

$$k = \prod_{p \in \mathcal{P}_k} p^{\alpha_p(k)},$$

- Introduce a sequence of independent geometric rv indexed by $\mathcal{P}$

$$\mathbb{P}(\xi_p = k) = (1 - p^{-1})p^{-k}, \quad k \in \mathbb{N} \cup \{0\}.$$

- Why are they relevant?

$$\mathbb{P}(\alpha_{p_1}(J_n) \geq k_1, ..., \alpha_{p_i}(J_n) \geq k_i) = \mathbb{P}[p_1^{k_1} \cdots p_i^{k_i} | J_n]$$
$$\to \frac{1}{p_1^{k_1} \cdots p_i^{k_i}} = \mathbb{P}(\xi_{p_1} \geq k_1, ..., \xi_{p_i} \geq k_i).$$

Theorem of conditioning independence (CJY)

Set $\vec{C}(n) := (\alpha_p(H_n); p \in \mathcal{P}_n)$ and $\vec{\xi}(n) := (\xi_p; p \in \mathcal{P}_n)$. Introduce

$$A_n := \Big\{ \prod_{p \in \mathcal{P}_n} p^{\xi_p} \le n \Big\}.$$

Then, we have

$$\mathbb{P}(A_n) = L_n \prod_{p \le n} (1 - p^{-1}) \ge \frac{1}{2}, \quad n \ge 21.$$

$$\mathcal{L}(\vec{C}(n)) = \mathcal{L}(\vec{\xi}(n) \mid A_n).$$

It follows that for ψ additive, i.e. $\psi(jk) = \psi(j) + \psi(k)$ with j, k coprime

$$\psi(H_n) \sim \mathcal{L}\Big(\sum_{p \in \mathcal{P}_n} \psi(p^{\xi_p}) | A_n \Big)$$

Key message: one bit away from uniform

Proposition (CJY)

Let H_n be harmonic and $Q_n \sim \text{Unif}\,\{1\} \cup \mathcal{P}_{n/H_n}$. Then for $n \geq 21$

$$d_{\mathrm{TV}}(H_n Q_n, J_n) \leq \frac{61 \log\log n}{\log n},$$
$$\mathbb{P}[Q_n | H_n] \leq \frac{6.4 \log\log n}{\log n}.$$

It follows that

$$d_{Wass}\Big(\frac{\psi(J_n) - \mu_n}{\sigma_n}, N\Big) \leq d_{Wass}\Big(\frac{\psi(H_n) - \mu_n}{\sigma_n}, N\Big) + c\Big(\frac{\log\log n}{\log n}\Big)^{1/2} + \frac{c}{\sigma_n},$$

where d_{Wass} is the Wasserstein distance.

Main results

Theorem (CJY)

Let ψ be bounded on $\mathcal{P}$ and satisfy

$$\sum_{p\in\mathcal{P}}\sum_{k\geq 2}\frac{\psi(p^k)^2}{p^k}<\infty.$$

Then we have

$$d_{Wass,Kol}(\frac{\psi(J_n)-\mu_n}{\sigma_n},N)\leq\frac{c}{\sigma_n}.$$

If further ψ is integer-valued and $M_n\sim\mathrm{Po}(\mu_n)$, then

$$d_{Kol}(\psi(J_n),M_n)\leq\frac{c}{\sqrt{\mu_n}}+\frac{c}{\mu_n}\sum_{p\in\mathcal{P}_n}\frac{|\psi(p)-1|}{p}.$$

Assume further $\psi(p)=1$, we have $d_{\mathrm{TV}}(\psi(J_n),M_n)\leq c/\sqrt{\mu_n}$.

The condition holds if $k\mapsto\psi(p^k)$ grows uniformly at most polynomially.

Examples:

1. Let $\omega(k)$ be the number of distinct prime factors of $k \in \mathbb{N}$. Then
$$\omega(p^k) = 1.$$
2. Let $\Omega(k)$ be the number of prime divisors of $k \in \mathbb{N}$ counting multiplicites, e.g. $\Omega(2^3) = 3$. Then
$$\Omega(p^k) = k.$$
3. Let $\Theta(k)$ the number of divisors of $k \in \mathbb{N}$ which is multiplicative, i.e. $\Theta(jk) = \Theta(j)\Theta(k)$ for $j, k \in \mathbb{N}$ co-prime. We have $\Theta(p^k) = k + 1$. Define the additive function $\Lambda(k) = \log \Theta(k)$. Then
$$\Lambda(p^k) = \log(k + 1).$$

More generally, any multiplicative function with sub-double-exponential growth in $k \mapsto \text{MultFunc}(p^k)$ will do.

Proof

- We focus on normal approximation in Wasserstein distance.
- By the key message, it suffices to consider the harmonic $\psi(H_n)$ which is coupled with a family of geometric ξ_p.

Baby fact

Let $M_{p,k} \sim \mathrm{Po}(\frac{1}{kp^k})$ then for $p \in \mathcal{P}$

$$\xi_p = \sum_{k \geq 1} k M_{p,k}.$$

- Couple everything in a Poisson process η on $\mathbb{X} = \mathcal{P} \times \mathbb{N} = \{(p,k)\}$ with intensity $\lambda(p,k) = \frac{1}{kp^k}$ by setting

$$\xi_p = \sum_{k \geq 1} k \eta(p,k).$$

- Goal: normal approximation for a functional of Poisson process under conditioning. Use Malliavin calculus and Stein's method.

- linear approximation ("first chaos dominates")

$$\sum_{p\in\mathcal{P}_n}\psi(p^{\xi_p}) = \sum_{p\in\mathcal{P}_n}\psi(p)\xi_p + R_n = \int \psi(p)k1_{p\in\mathcal{P}_n}d\eta(p,k) + R_n$$

  where $\mathbb{E}[|R_n|] \le c$. Let μ_n and σ_n^2 be the mean and variance of the Poisson integral.

The rest of the proof:

1. Stein's method for conditional distribution
2. Integration by parts
3. Error estimate by controlling the add-one-cost (Malliavin derivative)

Step 1 Stein's method

- Consider $f'(x) - xf(x) = h(x) - \mathbb{E}[h(N)]$ with $h \in \mathrm{Lip}_1$. There exists a solution such that $\|f\|, \|f'\|, \|f''\| \leq 2$.
- Let I_n be the indicator of A_n and $W_n = \frac{\int \ldots d\eta - \mu_n}{\sigma_n}$. Then

$$\begin{aligned}
&\mathbb{E}[h(\frac{\psi(H_n) - \mu_n}{\sigma_n})] - \mathbb{E}[h(N)] \\
&= \frac{1}{\mathbb{P}[A_n]}\Big(\mathbb{E}[f'(W + \frac{R_n}{\sigma_n})I_n] - \mathbb{E}[(W + \frac{R_n}{\sigma_n})f(W + \frac{R_n}{\sigma_n})I_n]\Big) \\
&= O(\frac{1}{\sigma_n}) + O(\mathbb{E}[f'(W + \frac{R_n}{\sigma_n})I_n] - \mathbb{E}[Wf(W + \frac{R_n}{\sigma_n})I_n]),
\end{aligned}$$

where we used the key observation, $\mathbb{P}[A_n] \geq 1/2$ and $\mathbb{E}[|R_n|] \leq c$.

Step 2 integration by parts

Duality D - $\tilde{\eta}$

Let η be Poisson with intensity λ and $D_x G = G(\eta + \delta_x) - G(\eta)$. Then

$$\mathbb{E}[\tilde{\eta}(u)G] = \mathbb{E}\int u(x) D_x G \lambda(dx),$$

where $\tilde{\eta} = \eta - \lambda$.

We apply it to $\tilde{\eta}(\rho) = W_n$ and G with

$$\rho(p,k) = \frac{1}{\sigma_n}\psi(p)k1_{p\in\mathcal{P}_n}$$

$$G = f(W_n + \frac{R_n}{\sigma_n})I_n.$$

Then

$$\mathbb{E}[Wf(W + \frac{R_n}{\sigma_n})I_n] = \mathbb{E}\int \rho(x) D_x(f(W_n + \frac{R_n}{\sigma_n})I_n)\lambda(dx).$$

Step 3 controlling the add-one-cost

It is easy to verify $D(FG) = FDG + GDF + DFDG$.

Claim ($D_x I_n$ is small)

$$D_x(f(W_n + \frac{R_n}{\sigma_n})I_n) = I_n D_x(f(W_n + \frac{R_n}{\sigma_n}) + O(\frac{1}{\sigma_n}).$$

Notice $D_x W_n = \rho(x)$ so that

$$\begin{aligned} D_x\Big(f(W_n + \frac{R_n}{\sigma_n})\Big) &= f(W_n + \rho(x) + \frac{R + D_x R_n}{\sigma_n}) - f(W_n + \frac{R_n}{\sigma_n}) \\ &= f'(W_n + \frac{R_n}{\sigma_n})\rho(x) + o(1) \end{aligned}$$

where we used $\|f''\| \leq 2$, leading to

$$\mathbb{E}[Wf(W + \frac{R_n}{\sigma_n})I_n] = \mathbb{E}[f'(W + \frac{R_n}{\sigma_n})I_n] + O(\frac{1}{\sigma_n}).$$

Outlook

Our approach and the neat divisibility property of harmonic distribution may be useful elsewhere.

- **Billinsley ('73)** characterized the possible limits for additive functions. Non-Gaussian, non-Poisson Fluctuation.
- Extreme value statistics e.g. Dickman approximation of the largest prime divisor of a uniform sample.
- Erdös-Kac type theorem for Fourier coefficients of modular forms

Thank you!

Reference: A probabilistic approach to the Erdös- Kac Theorem for additive functions
https://arxiv.org/abs/2102.05094

Normalizing constants for ω

Decomposition

$$\omega(J_n) = \sum_{p \in \mathcal{P}_n} \mathbb{1}(p|J_n),$$

- We compute $\mathbb{E}[\omega(J_n)]$ to see the normalization.

$$\mathbb{E}[\mathbb{1}(p|J_n)] = \mathbb{P}(p|J_n) = \frac{1}{n}\left\lfloor \frac{n}{p} \right\rfloor = \frac{1}{p} + O(n^{-1}),$$

$$\mathbb{E}[\omega(J_n)] = \Big(\sum_{p \in \mathcal{P}_n} \frac{1}{p} \Big) + O(1) = \log\log(n) + O(1),$$

where we have used the Mertens formula.

- To see $\mathbb{V}\mathrm{ar}(\omega(J_n)) \sim \log\log(n)$, naively,

$$\begin{aligned}
\mathbb{E}[\omega(J_n)^2] &= \sum_{p\in\mathcal{P}_n} \mathbb{P}(p|J_n) + \sum_{p\neq q\in\mathcal{P}_n} \mathbb{P}(p|J_n,\ q|J_n) \\
&= \log\log(n) + O(1) + \sum_{p\neq q} \frac{1}{n}\left\lfloor \frac{n}{pq} \right\rfloor \\
&\leq \log\log(n) + O(1) + \sum_{p\neq q} \frac{1}{pq} \\
&= \log\log(n) + O(1) + \Big(\sum_{p\in\mathcal{P}_n} \frac{1}{p}\Big)^2 \\
&= \log\log(n) + O(1) + \mathbb{E}[\omega(J_n)]^2.
\end{aligned}$$

  This gives $\mathbb{V}\mathrm{ar}(\omega(J_n)) \leq \log\log(n) + O(1)$.

- The inequality in red could be too generous. This can be resolved by a truncation argument, leading to $\mathbb{V}\mathrm{ar}(\omega(J_n)) \sim \log\log n$.