8.3.2021

2. The Erdős-Kac Theorem

The "simplest" additive function is $\omega(n) =$ nb. of $p \mid n$, without multiplicity

$$(\omega(12) = 2)$$

i.e.

$$\omega(p^\nu) = 1, \quad \forall p, \forall \nu \geq 1$$

We saw last week that because

$$\sum_p \frac{1}{p} = +\infty \quad \text{(Euler)}$$

ω does not satisfy the assumptions of Erdős-Wintner.

Nevertheless, we try to understand the values of ω for $n \leq N$, $N \to +\infty$.

Experience of E-W Th. suggests to compare ω on $\Omega_N = \{1, \ldots, N\}$ with

$$\sum_p B_p, \qquad \mathbb{P}(B_p = 1) = \frac{1}{p}$$

(B_p) independent

[since $\omega(n) = \sum_p \underline{B}_p(n)$,

$\underline{B}_p(n) = 1 \iff p \mid n$]

But this series diverges a.s because

$\{B_p = 1\}$ occurs a.s. infinitely often.

almost surely

But if we note that

$$\omega(n) = \sum_{p \leq N} \underline{B}_p(n)$$

$(n \in \Omega_N)$

we can try an approximation of ω on Ω_N by $\sum_{p \leq N} B_p$.

<u>First question:</u> if

$$X_N = \sum_{p \leq N} B_p,$$

what is the asymptotic behavior as $N \to +\infty$?

[B.7.2] Central Limit Theorem

Theorem [Special case of CLT]

If $(Y_n)_{n\geq 1}$ are independent Bernoulli r.v with $\mathbb{P}(Y_n=1)=y_n$, and if $\sum_{n\geq 1} y_n(1-y_n) = +\infty$, then

$$\frac{\sum_{n\leq N} Y_n - \sum_{n\leq N} y_n}{\sqrt{\sum_{n\leq N} y_n(1-y_n)}} \xrightarrow{\text{law}} \mathcal{N}(0,1)$$

$$\mathbb{P}(\mathcal{N}(0,1)\in[a,b]) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-t^2/2}\,dt$$

This suggests strongly:

Theorem (Erdős–Kac, ~1939) [Th. 2.3.1]

Let X_N on Ω_N be defined by $X_N(n)=\omega(n)$. Then

$$\frac{X_N - \log\log N}{\sqrt{\log\log N}} \xrightarrow[N\to\infty]{\text{law}} \mathcal{N}(0,1).$$

Note: $\log\log N$ arises as the asymptotic behavior of $\sum_{p \leq N} \frac{1}{p}$.

Sketch of the proof of the CLT

Th. (Lévy Criterion; simple form)

$d \geq 1$, (X_n) r.v. with values in $\mathbb{R}^d$

X r.v. with values in $\mathbb{R}^d$

Then $X_n \xrightarrow[n \to \infty]{\text{law}} X$

$$\Updownarrow$$

for all $t \in \mathbb{R}^d$, $\varphi_{X_n}(t) \xrightarrow[n \to \infty]{} \varphi_X(t)$

where the <u>characteristic functions</u> are

$$\varphi_{X_n}(t) = \mathbb{E}\left(e^{it \cdot X_n}\right)$$

$$\left[t \cdot x = t_1 x_1 + \dots + t_d x_d\right]$$

Note: (1) $\Downarrow$ easy because $\varphi_X(t) = \mathbb{E}(f(X))$ for $f(x) = e^{it \cdot x}$, continuous + bounded

(2) The "difficult" version of the Lévy Criterion does not require to know that the law of X is a probability measure.

<u>Application to CLT</u>: exploit independance to get

$$\mathbb{E}\left(\exp\left(it \frac{\sum_{n\leq N} Y_n - \sum_{n \leq N} y_n}{\sqrt{\sum_{n\leq N} y_n}} \right) \right)$$

$$\underset{\text{indep.}}{=} \prod_{n \leq N} \mathbb{E}\left[\exp\left(it \frac{Y_n - y_n}{\sqrt{\sum_{n\leq N} y_n}} \right) \right]$$

and then one uses Taylor expansion of exponential to get the CLT naturally (cf. B.7.1).

<u>Exercise</u> - Cheek that ("obvious") adaptations of the idea of proof of

the Schoenberg / Erdős-Wintner Ths do not work here.

We will use the <u>method of moments</u> instead.

<u>Theorem</u> [B.5.5 + B.5.6]

Let (X_n), X be real-valued r.v., s.t.

(1) $\forall n \geq 1$, $\forall k \geq 0$ (integer), $\mathbb{E}(|X_n|^k) < +\infty$ ("$X_n \in L^k(\Omega)$") and $\mathbb{E}(|X|^k) < +\infty$

(2) For $k \geq 0$, integer,

$$\mathbb{E}(X_n^k) \xrightarrow[n\to\infty]{} \mathbb{E}(X^k)$$

(3) The series

$$\sum_{k\geq 0} \frac{\mathbb{E}(|X|^k)}{k!} z^k$$

has positive radius of convergence.

Then $X_n \xrightarrow[n\to\infty]{\text{law}} X$; moreover if

(4) $\mathbb{E}(|X_n|^k) \leq c_k$ for $n \geq 1$, $k \geq 0$

(uniform integrability)

The converse is true : from $X_n \xrightarrow{\text{law}} X$

one can deduce

$\forall k \geqslant 0, \quad \mathbb{E}(X_n^k) \longrightarrow \mathbb{E}(X^k).$

Note : (1) This is easy if all X_n are uniformly bounded (say $|X_n| \leq A$), because then $f(x) = x^k$ can be restricted to a bounded continuous function, and moreover they generate a dense subspace of $\mathcal{C}([-A, A])$ [Weierstrass Theorem]

(2) Without Condition (3), there are known counterexamples.

Proof of the Erdős - Kac Theorem

Step 1 - Truncation

Step 2 - Convergence of moments:

let $$\underline{X}_N = \sum_{p \le N} \underline{B}_p$$

$$X_N = \sum_{p \le N} B_p$$

$$\sigma_N = \sum_{p \le N} \frac{1}{p} \quad (\sim \log \log N)$$

$$\mathbb{E}\left(\left(\frac{\underline{X}_N - \sigma_N}{\sqrt{\sigma_N}}\right)^k\right) = \mathbb{E}\left(\left(\frac{X_N - \sigma_N}{\sqrt{\sigma_N}}\right)^k\right) + (\text{error})$$

Step 3 - Converse of method of moments applies to

$$\frac{X_N - \sigma_N}{\sqrt{\sigma_N}} \quad \left[\xrightarrow[\text{CLT}]{\text{law}} \mathcal{N}(0,1)\right]$$

so we get

$$\mathbb{E}\left(\left(\frac{\underline{X}_N - \sigma_N}{\sqrt{\sigma_N}}\right)^k\right) \xrightarrow[N \to \infty]{} \mathbb{E}(\mathcal{N}^k)$$

and hence convergence in law by direct method of moments.

<u>Step 1</u>. We note that "very few integers $n \leq N$ have many large prime factors": let

$$Q = N^{1/(\log\log N)^{1/3}}$$

Then

$$\omega(n) - (\log\log N)^{1/3} \leq \sum_{\substack{p \mid n \\ p \leq Q}} 1 \leq \omega(n)$$

(because the difference is the number of $p \mid n$ s.t. $Q < p \leq N$, and if there are α of them, then

$$N^{\alpha/(\log\log N)^{1/3}} = Q^{\alpha} \leq n \leq N$$

so $\alpha \leq (\log\log N)^{1/3}$].

One can see then that E-K Th. follows from the variant where we consider

$$\frac{\sum_{p \leq Q} B_p - \sigma_N}{\sqrt{\sigma_N}}.$$

Step 2. Notation change:

centered r.v

$$\underline{X}_N = \sum_{p \le Q} \left(\frac{\underline{B}}{p} - \frac{1}{p} \right)$$

$$X_N = \sum_{p \le Q} \left(B_p - \frac{1}{p} \right)$$

$$\sigma_N = \sum_{p \le N} \frac{1}{p} \sim \sum_{p \le Q} \frac{1}{p}$$

We compute [for fixed $k \geq 0$]

$$\mathbb{E}_N \left(\left(\frac{X_N}{\sqrt{\sigma_N}} \right)^k \right)$$

$$= \frac{1}{\sigma_N^{k/2}} \mathbb{E}_N \left[\left(\sum_{p \le Q} \left(\underline{B}_p - \frac{1}{p} \right) \right)^k \right]$$

$$= \frac{1}{\sigma_N^{k/2}} \sum_{\substack{p_1, \dots, p_k \\ p_i \le Q}} \cdots \sum \mathbb{E}_N \left[\underbrace{\prod_{i=1}^{k} \left(\underline{B}_{p_i} - \frac{1}{p_i} \right)}_{f(n)} \right]$$

where f only depends n modulo $q = p_1 \cdots p_k$

So by Th. 1.3.3 + its Corollary we have

$$\mathbb{E}_N \left(f(n \bmod q) \right) = \mathbb{E}_{\mathbb{Z}/q\mathbb{Z}} (f) + O\left(\frac{\|f\|_1}{N} \right)$$

Here

$$f(x) = \prod_{i=1}^{k} \left(\delta_{p_i}(x) - \frac{1}{p_i} \right)$$

where

$$\delta_{p_i}(x) = \begin{cases} 1 & \text{if } x \bmod p_i = 0 \\ 0 & \text{otherwise} \end{cases}$$

so

$$\|f\|_1 = \sum_x |f(x)| \leq q \leq Q^k$$

$$[= o(N)]$$

and moreover

$$\mathbb{E}_{\mathbb{Z}/q\mathbb{Z}} (f) = \mathbb{E}\left(\prod_{i=1}^{k} \left(B_{p_i} - \frac{1}{p_i} \right) \right)$$

by a small computation.

So we conclude:

$$\mathbb{E}_N \left(\left(\frac{X_N}{\sqrt{\sigma_N}} \right)^k \right)$$

$$= \frac{1}{\sigma_N^{k/2}} \sum_{\substack{p_i \leq Q \\ p_1, \ldots, p_k}} \cdots \sum \left(\mathbb{E}\left(\prod \left(B_{p_i} - \frac{1}{p_i} \right) \right) + O\left(\frac{Q^k}{N} \right) \right)$$

$$= \mathbb{E}\left[\left(\frac{X_N}{\sqrt{\sigma_N}}\right)^k\right] + O\left(\frac{Q^{2k}}{N}\right)$$

(by reversing the computation).

Our choice of $Q = N^{1/(\log\log N)^{1/3}}$ ensures that $\dfrac{Q^{2k}}{N} \xrightarrow[N\to\infty]{} 0$.

<u>Step 3.</u> By the C-L-T we know that

$$\frac{X_N}{\sqrt{\sigma_N}} \xrightarrow[N\to\infty]{\text{law}} \mathcal{N}(0,1)$$

and so we will get

$$\mathbb{E}\left(\left(\frac{X_N}{\sqrt{\sigma_N}}\right)^k\right) \longrightarrow \mathbb{E}(\mathcal{N}^k)$$

<u>if</u> the converse of method of moments applies, and we conclude

$$\boxed{\frac{X_N}{\sqrt{\sigma_N}} \xrightarrow{\text{law}} \mathcal{N}(0,1)}$$

[because one knows that the

moments are not too big so that

$$\sum_{k\geq 0} \frac{\mathbb{E}(|\mathcal{N}|^k)}{k!} z^k$$

has positive radius of convergence (in fact $+\infty$)]

We need to prove that there exists $c_k \geq 0$ s.t. $\mathbb{E}\left(\left(\frac{X_N}{\sqrt{\sigma_N}}\right)^k\right) \leq c_k$ (*) for all $N \geq 1$.

<u>Claim</u>: $\exists$ $c \geq 0$ s.t. for all N

$$\max\left(\mathbb{E}\left(\exp\left(\frac{X_N}{\sqrt{\sigma_N}}\right)\right), \mathbb{E}\left(\exp\left(-\frac{X_N}{\sqrt{\sigma_N}}\right)\right)\right) \leq c$$

[In fact $c = \exp(1)$ works...]

Observe that

$\exists$ c_k, $\quad |x|^k \leq c_k(e^x + e^{-x})$

for all $x \in \mathbb{R}$, to conclude proof of (*)

Let $t \in [-1, 1]$. By independence

$$\mathbb{E}\left(\exp\left(t \frac{X_N}{\sqrt{\sigma_N}}\right)\right)$$

$$= \prod_{p \leq Q} \mathbb{E}\left(\exp\left(\frac{t(B_p - 1/p)}{\sqrt{\sigma_N}}\right)\right)$$

For $N \geq N_0$ (so that $\sigma_{N_0} \geq 1$) we get exp(reals in $[-1, 1]$) and use

$$\forall x \in [-1, 1], \quad e^x \leq 1 + x + x^2$$

so

$$\mathbb{E}\left(\exp\left(\frac{t(B_p - 1/p)}{\sqrt{\sigma_N}}\right)\right)$$

$$\leq 1 + \underbrace{\mathbb{E}\left(\cancel{\frac{t(B_p - 1/p)}{\sqrt{\sigma_N}}}\right)}_{= 0}$$

$$+ \frac{t^2}{\sigma_N} \underbrace{\mathbb{E}\left(\left(B_p - \frac{1}{p}\right)^2\right)}_{\mathbb{V}(B_p)}$$

Using $1 + x \leq e^x$ for $x \in \mathbb{R}$

we get

$$\prod_{p \le Q} \mathbb{E}\left(\exp\left(\frac{t\,(B_p - 1/p)}{\sqrt{\sigma_N}} \right) \right)$$

$$\le \exp\left(\frac{t^2}{\sigma_N} \underbrace{\sum_{p \le Q} \mathbb{V}(B_p)}_{= \sum_{p \le Q} \frac{1}{p}\left(1 - \frac{1}{p}\right)} \right)$$

$$\le \exp\left(t^2 \right)$$