Where does it come from? If $x_1, \dots, x_n$ are IID $g(m, \sigma)$ define the sample mean

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \quad , \quad \text{with } \langle \overline{X} \rangle = m .$$

We know that $\overline{X}$ is gaussian, for large n (anticipating on CLT), but we ignore m and σ, and would like to infer them from the data.

$\frac{\overline{X}-m}{\sigma} \rightsquigarrow g(0,1)$ is true but useless in practice.

We estimate the variance by

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(x_i - \overline{X}\right)^2 \quad ,$$

sole function of known q^{ties}: $x_1, \dots x_n$

$n-1$ here s.t. $\langle S^2 \rangle = \sigma^2$

Then $\frac{\overline{X}-m}{\sqrt{S^2}}$ follows Student's distrib, $\mu = n-1$.

From this, we can get a confidence interval for m, starting from sample.

If $\mu < 2$, the variance does not exist

$\mu < 1$ " mean " " "

This is the case for the time to return to starting position for an unbiased r.w

We will see that $p(\tau) \underset{\tau \to \infty}{\propto} \tau^{-3/2}$ i.e. $\mu = 1/2$

Other examples: turbulence, fracture, earthquake stat, financial crises, distribution of firm sizes, magnetism (Barkhausen noise) ...

Power-law distrib are SCALE INVARIANT $\Leftrightarrow$ SELF-SIMILAR

Means: $\frac{\Pr[X > 10x_0]}{\Pr[X > x_0]}$ does not depend on x_0

$$= \frac{\int_{10x_0}^{\infty} x^{-\mu-1}\, dx}{\int_{x_0}^{\infty} x^{-\mu-1}\, dx} = \frac{(10x_0)^{-\mu}}{x_0^{-\mu}} = 10^{-\mu}$$

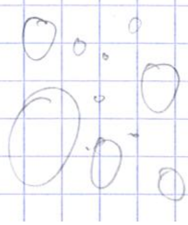
This is at variance with "thin-tailed" distributions, eg $p(x) = \lambda e^{-\lambda x}$, $x > 0$:

$$\frac{e^{-\lambda 10 x_0}}{e^{-\lambda x_0}} = e^{-9\lambda x_0} ,$$

strongly suppressed when $x_0 \nearrow$.

For instance, $\mu \simeq 3/2$ for Pareto wealth distrib, hence

$$\frac{\#\text{ people wealth} > 10 \text{ millions €}}{\#\text{ wealth} > 1 \text{ million €}} = 10^{-1,5} \simeq 30 = \frac{\#\text{ wealth } 10 \text{ billion €}}{\#\ \text{"}\ 1 \text{ billion €}}$$

Thus,  observe a power law distribution collection of droplets on a plate

The picture is zoom independent (in some range)

There is no characteristic scale

Conversely, if the size distribution is gaussian,

zoom in $\Rightarrow$ we see inside a single droplet, or nothing

zoom out $\Rightarrow$ we see a collection of point droplets

NB there is necessarily a lower cutoff to power-law if $x \to \infty$
but there can be an upper cutoff " " " as well $\to$ truncated power law
Ex: distrib of citations to scientific paper; $\mu \simeq 2$: Matthew effect,
rich get richer, but when a paper is "too" famous like Einstein's 1905,
is stops getting cited and becomes textbook material, if not core knowledge.

2°) Dependent and independent variables

Multivariate case: collection of observables, joint pdf $p_n(x_1, \dots x_n)$, $\int p = 1$

If independent: $p_n(x_1 \dots x_n) = \tilde{p}_1(x_1)\tilde{p}_2(x_2)\tilde{p}_3(x_3)\dots$ and

$$\langle x_1^{\alpha_1} \dots x_n^{\alpha_n} \rangle = \langle x_1^{\alpha_1} \rangle \dots \langle x_n^{\alpha_n} \rangle$$

$$\Rightarrow C_{ij} \equiv \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle = 0$$

Thus independence $\Rightarrow$ decorrelation

For a gaussian law $p(x_1, \dots x_n) \propto \exp\left(-\frac{1}{2}\sum_{i,j} x_i C^{-1}_{ij} x_j\right)$ matrix element i,j of C^{-1}

and here, no correlation $\Rightarrow C$ diagonal $\Rightarrow C^{-1}$ diagonal $\Rightarrow p$ factorizes $\Rightarrow$ independ

But in g^{al} no 2 pt correlation $\not\Rightarrow$ independence

Besides 2-point correlations are blind to important correlations. Ex, take multivariate Student

$$p(x_1, x_2) = \frac{1}{2\pi}\left[\frac{1}{1 + \frac{x_1^2 + x_2^2}{\mu}}\right]^{\frac{1+\mu}{2}}$$

$$\langle x_1 x_2 \rangle = 0 = \underbrace{\langle x_1 \rangle}_{0}\underbrace{\langle x_2 \rangle}_{0} \Rightarrow C_{12} = 0 \text{ but } \langle x_1^2 x_2^2 \rangle - \langle x_1^2 \rangle \langle x_2^2 \rangle \neq 0$$

x_1 and x_2 are not independent

3°) Sums of random variables and the central limit theorem

One often meets situations where the statistics of a sum of r.v matters: elections, q^{tis} like the pressure, a stock index... The statistics of the max of a number of r.v also is important: record in sport, climate extremes, network devise for peak consumption... The 2 questions are related (see 4°), when CLT breaks) and remarkably, rather universal can be obtained

3°) a) The Central Limit Theorem (CLT)

Take a sample $(x_1, \dots x_n)$ IID from $p(x)$, with mean m and variance σ^2

$$S_n \equiv \sum_{i=1}^{n} x_i \quad ; \quad \langle S_n \rangle = n m \quad ; \quad V(S_n) = n\sigma^2$$

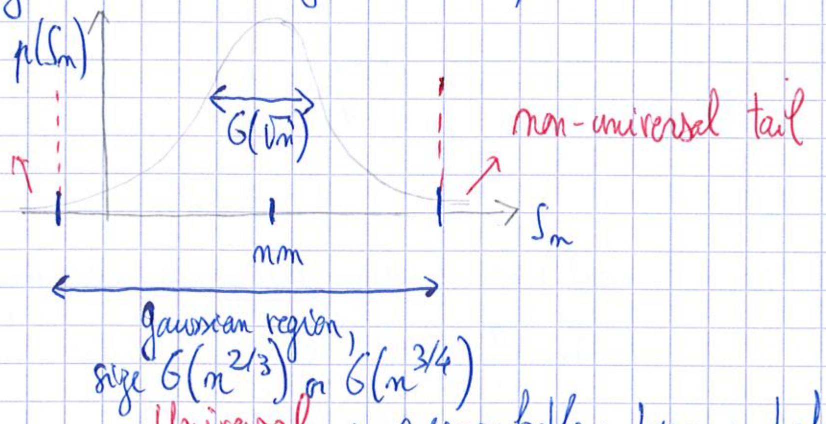
Then:
$$\lim_{n\to\infty} \Pr\left[a \leqslant \frac{S_n - nm}{\sigma\sqrt{n}} \leqslant b \right] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\xi^2/2} \, d\xi$$

In other words: $\dfrac{S_n - \text{its mean}}{\text{its std deviation}}$, which is by def of 0 mean and unit variance becomes gaussian $g(0,1)$ for $n \to \infty$.

$$S_n \simeq nm + \sigma\sqrt{n}\, \xi_n \quad \text{and} \quad \xi \to g(0,1)$$

$$\Rightarrow \boxed{\frac{S_n}{n} \simeq m + \frac{\sigma}{\sqrt{n}} \xi_n} \xrightarrow[n\to\infty]{} m \qquad \text{law of large numbers}$$

Tells nothing about the "tails" (large / small S_n limit). One often reads that the CLT "central" region is where ξ_n is $G(1)$ → wrong. It can go up to $G(n^{1/6})$ in general and $G(n^{1/4})$ for symmetric pdf [s.t. $p(x) = p(2m - x)$]
$p(x)$ symm around m
Hence the tails of the gaussian regime do matter, before we leave the central CLT region, and enter "the non CLT tails".
→ mostly visible with large deviations, seen later



Universal → remarkable. Large n behavior is oblivious of the details of $p(x)$. Large scale randomness creates strict regularity

Generalizations ⊖ CLT holds beyond IID case, if all x_i have different $p_i(x_i)$
Under mild hypothesis (at least all moments finite)

$$\frac{1}{\sqrt{\sum_i \sigma_i^2}} \sum_{i=1}^{n} (x_i - m_i) \xrightarrow[n\to\infty]{} g(0,1)$$

⊖ CLT also survives (moderate) correlations, see below

3°) b) Where does the CLT come from?

From the fact that the gaussian is a fixed point of the convolution operation, in the functional sense. This fixed point is attractive, for all pdf having finite moments.

Take $X_1 \rightsquigarrow g(m_1, \sigma_1)$; $X_2 \rightsquigarrow g(m_2, \sigma_2)$: $X_1+X_2 \rightsquigarrow g(m_1+m_2, \sqrt{\sigma_1^2+\sigma_2^2})$

as can be seen from the characteristic function $\langle e^{ikX_1} \rangle = \exp(m_1 k - \frac{k^2}{2}\sigma_1^2)$

In g^{al}, pdf p_{S_n} of $S_n = X_1 + \dots + X_n$, X_i not gaussian, but IID, mean m, $V(X) = \sigma^2$

$$\hat{p}_{S_n}(k) = \langle e^{ik\sum X_i} \rangle = \prod_{i=1}^{n} \langle e^{ikX_i} \rangle = [\hat{p}(k)]^n$$

$$\Rightarrow p_{S_n}(s) = \int \frac{dk}{2\pi} e^{-iks} [\hat{p}(k)]^n = \int \frac{dk}{2\pi} e^{-iks + n \log \hat{p}(k)}$$

$$= \int_{-\infty}^{+\infty} \frac{dk}{2\pi} \exp\left\{-iks + n\left[ikm + -\frac{k^2}{2} \overset{\sigma^2}{c_2} + \sum_{\ell \geq 3} \frac{(ik)^\ell}{\ell!} c_\ell\right]\right\}$$

$z = \frac{s - nm}{\sigma\sqrt{n}}$

$$= \int_{-\infty}^{+\infty} \frac{dk}{2\pi} \exp\left\{-ikz\sigma\sqrt{n} - \frac{1}{2}nk^2\sigma^2 + n\sum_{\ell\geq 3} \frac{(ik)^\ell}{\ell!} c_\ell\right\};$$

$q = k\sqrt{n}\sigma$

$$= \frac{1}{2\pi} \frac{1}{\sigma\sqrt{n}} \int_{-\infty}^{+\infty} dq \exp\left\{-iqz - \frac{1}{2}q^2 + \sum_{\ell\geq 3} \frac{(iq)^\ell}{\ell!} \frac{n}{\sigma^\ell n^{\ell/2}}\right\}$$

$\propto n^{1-\ell/2}$

We switch from pdf of S_n to that of z (that may depend on n): $p_{S_n}(s)\, ds = p_Z(z)\, dz$, $ds/dz = \sigma\sqrt{n}$

$$p_Z(z) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} dq \exp\left\{-iqz - \frac{1}{2}q^2 + \sum_{\ell\geq 3} \text{terms in } n^{1-\ell/2}\right\}$$

$$\xrightarrow[n\to\infty]{} \frac{1}{2\pi} \int_{-\infty}^{+\infty} dq \exp\left(-iqz - \frac{1}{2}q^2\right)$$

$$= \frac{1}{\sqrt{2\pi}} \exp(-z^2/2)$$

4°) When the CLT goes wrong

Will happen when looking at
- large deviations: tail region, see ch 2
- broad distributions (fat tails)
- correlated variables

a) The fat tails problem

The gaussian distrib has a charact function in $e^{ikm - k^2\sigma^2/2}$, stable under product.

But one can imagine other charact functions that have the same property, like $e^{-\gamma|k|}$. Well: is it really a charact function, meaning the FT of a pdf?

We have to check this :

$$\int \frac{dk}{2\pi} e^{-\gamma|k|} e^{-ikx} = \int_0^\infty \frac{dk}{2\pi} e^{-k(\gamma+ix)} + \int_{-\infty}^0 \frac{dk}{2\pi} e^{k(\gamma-ix)} \ldots$$

$$= \frac{1}{2\pi}\left(\frac{1}{\gamma+ix} + \frac{1}{\gamma-ix}\right) = \frac{1}{\pi}\frac{\gamma}{\gamma^2+x^2}$$

yes! It is a pdf → Cauchy distr ($\mu = 1$)

If we sum such variables : $e^{-\gamma|k|} \times e^{-\gamma|k|} \times \cdots \times e^{-\gamma|k|}$,
we will never get some $e^{+ikm - k^2\sigma^2/2}$

thus the sum of Cauchy variables is never gaussian : remains of Cauchy type.

Can be generalized to functions of type $e^{-\gamma|k|^\mu}$ (and more)

that can be shown to be charact. functions for pdf with broad distrib, and Lévy index μ

$$p(x) \propto \frac{1}{x^{\mu+1}}, \; x \to \infty \; ; \quad \text{large } x \Leftrightarrow \text{small } k$$

$$\langle e^{ikx} \rangle = 1 - \gamma|k|^\mu + \ldots \quad k \to 0$$

Note that $e^{-\gamma|k|^\mu}$ is not analytic while we know that if all moments do exist for the underlying pdf (FTransform), it should be analytic. Thus all moments do not exist.

Take $\mu < 2$: variance not defined. The sum of such "mysterious" variables will never converge to a gaussian, but some other fixed point of the convolution operation, called Lévy stable law. These laws are defined from their charact. function, but lack often an explicit representation. We have seen the exception of Cauchy distrib, explicit. Also explicit is Frechet distribution

$$p(x) = \frac{1}{\sqrt{\pi}} \frac{e^{-1/x}}{x^{3/2}}, \quad \text{having } \mu = 1/2, \; \langle e^{ikx} \rangle = e^{-\sqrt{|k|}}$$

The Lévy-stable laws decay as $|x|^{-\mu-1}$ and are characterized by another param β, in addition to μ ; quantifying the left-right asymmetry

$$L_{\beta,\mu}(x) \underset{|x|\to\infty}{\propto} \frac{1 \pm \beta A}{|x|^{\mu+1}} + \mathcal{O}\left(\frac{1}{|x|^{1+2\mu}}\right) \; (*), \quad 0 < \mu < 2, \; -1 \leq \beta \leq 1$$

Two other param need to be introduced, to account for a shift and a rescaling of x : they are not essential, do not affect the shape (core features)

For $\mu \to 2$, $L_{\beta,\mu} \to$ gaussian. Not obvious from (*)