Lecture Notes 1: Platonic Convergence and the Central Limit Theorem

1) An erroneous notion of limit:

Take the standard formulation of the Central Limit Theorem (Feller 1971, Vol. II;Grimmet & Stirzaker, 1982):

Let X_1, X_2,... be a sequence of independent identically distributed random variables with mean m & variance σ^2 satisfying $m< \infty$ and $0 <\sigma^2 <\infty$, then

$$\frac{\sum_{i=1}^{N} X_i - N\,m}{\sigma \sqrt{N}} \xrightarrow{D} \text{Gaussian }(0,\ 1) \text{ as } N \to \infty$$

Where $\xrightarrow{D}$ is converges "in distribution".

Taking convergence for granted provides a plain illustration of the severe disease of Platonicity --or working backwards from theory to practice. Effectively we are dealing with a double problem.

1) The first, as uncovered by Jaynes, comes from to the abuses of formalism & measure theory:

- Jaynes 2003 (p.44):"The danger is that the present measure theory notation presupposes the infinite limit already accomplished, but contains no symbol indicating which limiting process was used (...) Any attempt to go directly to the limit can result in nonsense".

  Granted Jaynes is still too Platonic in general and idealizes his convergence process (he also falls headlong for the Gaussian by mixing thermodynamics and information). But we accord with him on this point --along with the definition of probability as information incompleteness, about which in later sessions.

2) The second problem is that we do not have a "clean" limiting process --the process cannot be idealized. It is very rare to find permanent idealized conditions that allow for temporal aggregation.

Now how should we look at the Central Limit Theorem? Let us see how we arrive to it assuming "independence".

2) The Problem of Convergence

The CLT works in a specific way:It does not fill-in uniformily,but in a near-Gaussian way--indeed,disturbingly so. Simply,whatever your distribution (assuming one mode),your sample is going to be skewed to deliver more central observations,and fewer tail events.The consequence is that,under aggregation,the sum of these variables will converge "much" faster in the body of the distribution than in the tails. As N,the number of observations increases,the Gaussian zone should cover more grounds... but not in the "tails".

You can see it very easily with two very broad uniform distributions,say with a lower bound a and an upper bound b, b-a very large. As you convolute, you will see the peakedness in the center, which means that more observations will fall there (see Appendix).

This quick note shows the intuition of the convergence and presents the difference between distributions. (See Appendix)

Take the sum of of random independent variables X_i with **finite variance** under distribution φ(X). Assume 0 mean for simplicity (and symmetry, absence of skewness to simplify).

A better formulation of the Central Limit Theorem (Kolmogorov et al,x)

$$P\left[-u \leq Z = \frac{\sum_{i=0}^{n} X_i}{\sqrt{n}\,\sigma} \leq u\right] = \frac{1}{\sqrt{2\pi}} \int_{-u}^{u} e^{-\frac{Z^2}{2}} \, dZ$$

So the distribution is going to be:

$$\left(1 - \int_{-u}^{u} e^{-\frac{Z^2}{2}} \, dZ\right) \text{ for } -u \leq z \leq u$$

inside the "tunnel" [-u,u] --the odds of falling inside the tunnel itself

and

$$\int_{-\infty}^{u} \varphi'[n]\,(Z)\, dz + \int_{u}^{\infty} \varphi'[n]\,(Z)\, dz$$

outside the tunnel [-u,u]

Where φ'[n] is the n-summed distribution of φ.

How φ'[n] behaves is a bit interesting here --it is distribution dependent. And it depends on the initial distribution!

Bouchaud-Potters Treatment of Width of the Tunnel [-u,u]

(in class derivation)

■ 3) Using Log Cumulants & Observing Gaussian Convergence

The normalized cumulant of order *n*, C(n) is the derivative of the log of the characteristic function ϕ which we convolute N times divided by the second cumulant (i,e., second moment).

$$C(n, N) = \frac{(-i)^n\, \partial^n \log(\phi^N)}{\left(-\partial^2 \log(\phi)^N\right)^{n-1}} \;/.\; z \to 0$$

Since C(N+M)=C(N)+C(M), the additivity of the Log Characteristic function under convolution makes it easy to see the speed of the convergence to the Gaussian.

Fat tails implies that higher moments implode --not just the 4th .

Table of Normalized Cumulants -Speed of Convergence (Dividing by σ^n where n is the order of the cumulant).

Distribution	Normal$[\mu, \sigma]$	Poisson(λ)	Exponential(λ)	$\Gamma(a, b)$
PDF	$\frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\ \sigma}$	$\frac{e^{-\lambda}\ \lambda^x}{x!}$	$e^{-x\lambda}\ \lambda$	$\frac{b^{-a}\ e^{-\frac{x}{b}}\ x^{a-1}}{\Gamma(a)}$
N−convoluted Log Char acteristic	$N \log\left(e^{i z \mu - \frac{z^2 \sigma^2}{2}}\right)$	$N \log(e^{(-1+e^{i z})\lambda})$	$N \log(\frac{\lambda}{\lambda - i z})$	N $\log((1 - i\, b\, z)^{-a})$
2nd Cum	1	1	1	1
3 rd	0	$\frac{1}{N \lambda}$	$\frac{2\lambda}{N}$	$\frac{2}{a\, b\, N}$
4 th	0	$\frac{1}{N^2\ \lambda^2}$	$\frac{3!\, \lambda^2}{N^2}$	$\frac{3!}{a^2\ b^2\ N^2}$
5 th	0	$\frac{1}{N^3\ \lambda^3}$	$\frac{4!\ \lambda^3}{N^3}$	$\frac{4!}{a^3\ b^3\ N^3}$
6 th	0	$\frac{1}{N^4\ \lambda^4}$	$\frac{5!\, \lambda^4}{N^4}$	$\frac{5!}{a^4\ b^4\ N^4}$
7 th	0	$\frac{1}{N^5\ \lambda^5}$	$\frac{6!\ \lambda^5}{N^5}$	$\frac{6!}{a^5\ b^5\ N^5}$
8 th	0	$\frac{1}{N^6\ \lambda^6}$	$\frac{7!\, \lambda^6}{N^6}$	$\frac{7!}{a^6\ b^6\ N^6}$
9 th	0	$\frac{1}{N^7\ \lambda^7}$	$\frac{8!\, \lambda^7}{N^7}$	$\frac{8!}{a^7\ b^7\ N^7}$
10 th	0	$\frac{1}{N^8\ \lambda^8}$	$\frac{9!\, \lambda^8}{N^8}$	$\frac{9!}{a^8\ b^8\ N^8}$

Distribution	Mixed Gaussians (Stoch Vol)	StudentT(3)	StudentT(4)	□
PDF	$p\ \frac{e^{-\frac{x^2}{2\sigma_1{}^2}}}{\sqrt{2\pi}\ \sigma_1} + (1-p)\ \frac{e^{-\frac{x^2}{2\sigma_2{}^2}}}{\sqrt{2\pi}\ \sigma_2}$	$\frac{6\sqrt{3}}{\pi\, (x^2+3)^2}$	$12\, (\frac{1}{x^2+4})^{5/2}$	□
N − convoluted Log Characteristic	$N \log\left(p\, e^{-\frac{z^2\ \sigma_1{}^2}{2}} + (1-p)\, e^{-\frac{z^2\ \sigma_2{}^2}{2}}\right)$	N $\left(\log\left(\sqrt{3}\ \lvert z\rvert + 1\right) - \sqrt{3}\ \lvert z\rvert\right)$	$N \log($ $2\, \lvert z\rvert^2\ K_2(2\, \lvert z\rvert))$	□
2nd Cum	1	1	1	□
3 rd	0	Ind	□	□
4 th	$-\left(3\,(-1+p)\ p\,(\sigma_1^2 - \sigma_2^2)^2\right) / \left(N^2\ (p\,\sigma_1^2 - (-1+p)\,\sigma_2^2)^3\right)$	Ind	Ind	□
5 th	0	Ind	Ind	□
6 th	$\left(15\,(-1+p)\ p\,(-1+2\,p)\ (\sigma_1^2 - \sigma_2^2)^3\right) / \left(N^4\ (p\,\sigma_1^2 - (-1+p)\,\sigma_2^2)^5\right)$	Ind	Ind	□

Note: On "Infinite Kurtosis"- Discussion

Note on Chebyshev's Inequality and upper bound on deviations under finite variance

A lot of idiots talk about finite variance not considering that it still does not mean much. Consider Chebyshev's inequality:

$$P[X > \alpha] \le \frac{\sigma^2}{\alpha^2}$$

$$P[X > n\,\sigma] \le \frac{1}{n^2}$$

Which effectively accommodate power laws but puts a bound on the probability distribution of large deviations --but still significant.

The Effect of Finiteness of Variance

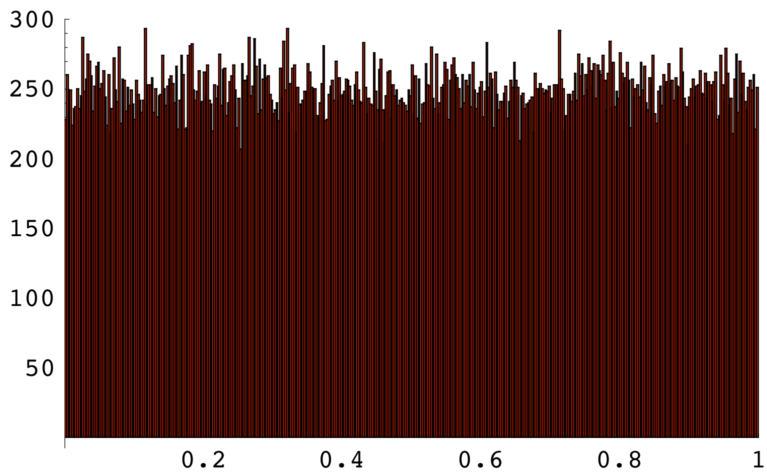
This table shows the probability of exceeding a certain σ for the Gaussian and the lower on probability limit for any distribution with finite variance.

Deviation	Gaussian	Chebyshev Upper Bound
3	$7. \times 10^{2}$	9
4	$3. \times 10^{4}$	16
5	$3. \times 10^{6}$	25
6	$1. \times 10^{9}$	36
7	$8. \times 10^{11}$	49
8	$2. \times 10^{15}$	64
9	$9. \times 10^{18}$	81
10	$1. \times 10^{23}$	100

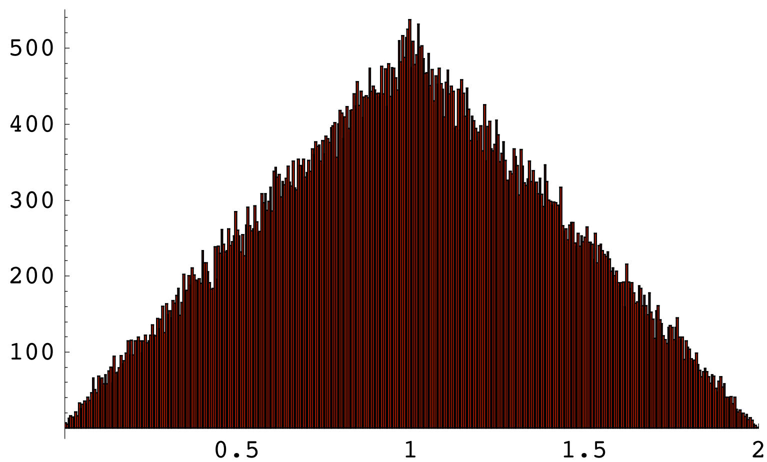
■ Calculations

■ Calculations

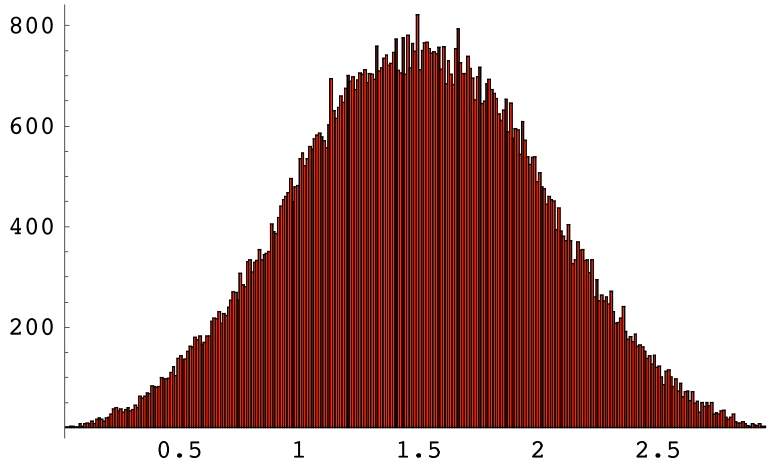
■ APPENDIX to 2- How We Converge Mostly in the Center - A Tutorial



N=2



N=3



Out[417]= - Graphics -

Example: Uniform Distribution

f (x) = { 1   0 ≤ x ≤ 1

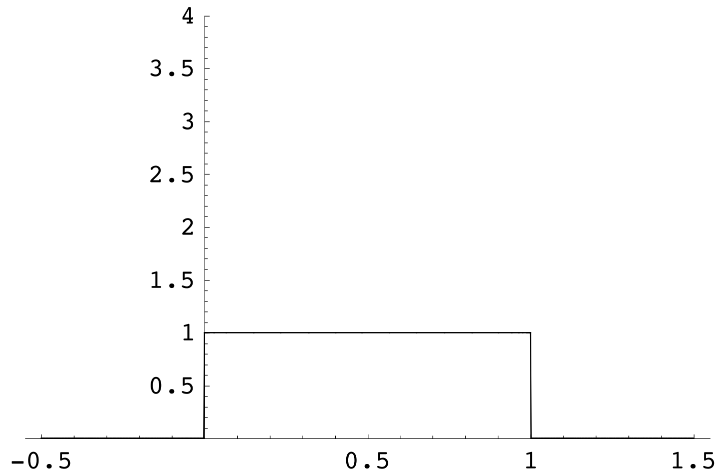
By Convoluting 2, 3, 4 times iteratively:

$$f2\,(z2) = \int_{-\infty}^{\infty} (f\,(z-x))\,(f\,x)\,dx = \begin{cases} 2 - z2 & 1 < z2 < 2 \\ z2 & 0 < z2 \leq 1 \end{cases}$$

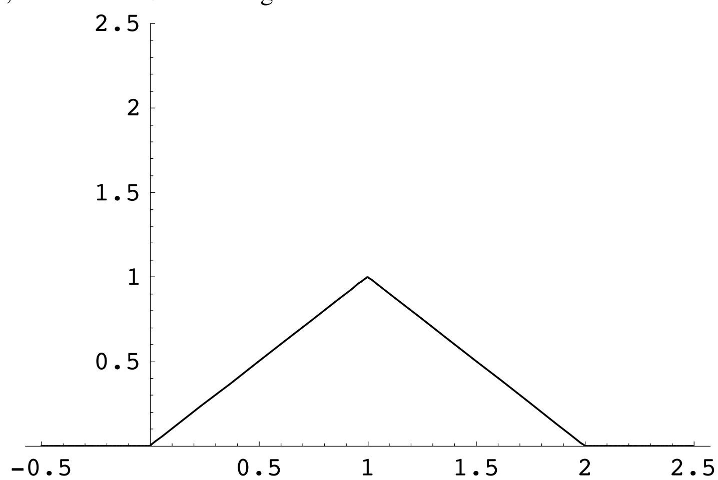
$$f3\;z3 = \int_0^3 (f2\,(z3 - x2))\,(f\,x2)\,dx2 = \begin{cases} \frac{z3^2}{2} & 0 < z3 \leq 1 \\ -(z3-3)\,z3 - \frac{3}{2} & 1 < z3 < 2 \\ -\frac{1}{2}\,(z3-3)\,(z3-1) & z3 = 2 \\ \frac{1}{2}\,(z3-3)^2 & 2 < z3 < 3 \end{cases}$$

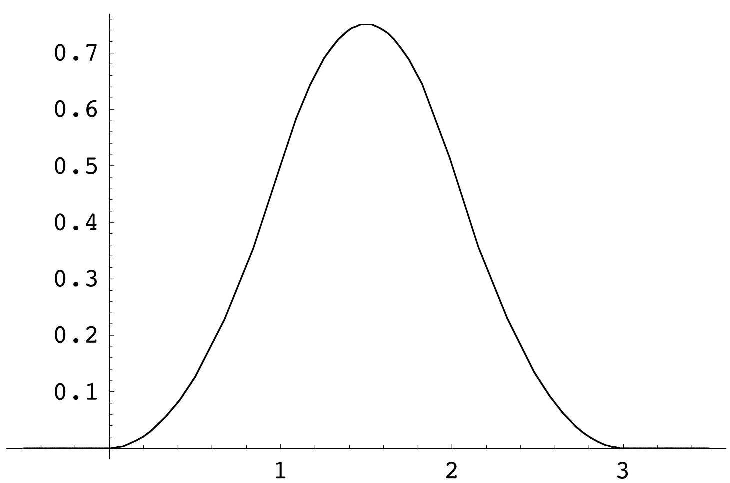
$$f4\;x = \int_0^4 (f3\,(z4 - x))\,(f\,x3)\,dx3 = \begin{cases} \frac{1}{4} & z4 = 3 \\ \frac{1}{2} & z4 = 2 \\ \frac{z4^2}{4} & 0 < z4 \leq 1 \\ \frac{1}{4}\,(-z4^2 + 4\,z4 - 2) & 1 < z4 < 2 \lor 2 < z4 < 3 \\ \frac{1}{4}\,(z4-4)^2 & 3 < z4 < 4 \end{cases}$$

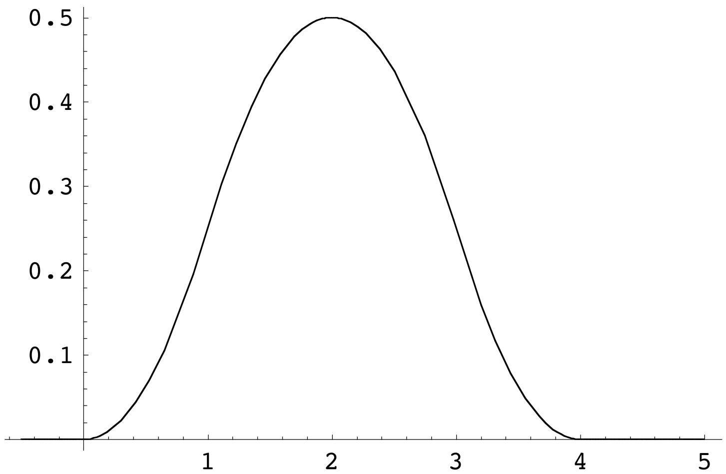
A simple Uniform Distribution



We can see how quickly, after one single addition, the net probabilistic "weight" is going to be skewed to the center of the distribution, and the vector will weight future densities..







Discussion

Gaussian Case

Paretan Case

Dealing With the Distribution of the Summed distribution φ