

LECTURE 1: Noise in Physical Systems

- Fluctuations occur in almost all natural phenomena, especially if the phenomena in question are time-varying.
 - And if fluctuations arise in a human-made device, we call it noise.
 - Here we will use the term fluctuations to include noise.
 - Understanding noise is central to understanding the design and performance of almost any device.
 - Noise sets essential limits on how small a bit can reliably be stored and on how fast it can be sent; effective designs must recognize these limits in order to approach them.
-
- Our 1st step will be an introduction to random variables and some of their important probability distributions, then we will turn to noise generation mechanisms, and close with some more general thermodynamic insights into noise.
 - Although noise can be surprisingly interesting in its own right, we will focus on concepts that lay the foundation for topics we will explore.
 - But along the way, I will also share examples of noise that have yet to be systematically studied - An open research question, if you may.
 - Eventually, we will move towards a principled exploration of fluctuations in generic physical systems by using these concepts to study the linear response regime of systems nominally removed from thermal equilibrium, and hence explore advances in strongly driven systems which exhibit large magnitude, strongly correlated fluctuations with nonlinear responses.

- Throughout, we will keep an eye out for what we can learn about a physical system or its underlying process from the fluctuations it exhibits.
- Finally, I'm an experimental physicist by training. My approach will not be towards establishing theoretical rigor. Instead, I will exclusively focus on physical intuition gained from the theoretical results and how experimental data are analyzed.
- So, you'll be seeing plenty of experimental examples once we are past the conceptual stage.

* RANDOM VARIABLES:

** Expectation Values:

- Consider a fluctuating quantity $x(t)$, such as the output from a noisy amplifier.
- If x is a random variable, it is drawn from a probability distribution $\phi(x)$.
- This means that it is not possible to predict the value of x at any instant, but knowledge of the distribution does let precise statements be made about the average value of quantities that depend on x .
- The expected value of a function $f(x)$ can be defined by an integral either over time or over a distribution:

$$\langle f(x) \rangle = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} f(x(t)) dt$$

$$= \int f(x) \phi(x) dx \longrightarrow \textcircled{1}$$

(or a sum if the distribution is discrete).

- Taking $f(x) = 1$ shows that a probability distribution must be normalized: $\int_{-\infty}^{+\infty} 1 \cdot f(x) dx = 1 \rightarrow ②$

If $f(x)$ exists and is independent of time then the distribution is said to be stationary. More will be said a little later on this.

- The moments of a distribution are the expectation values of the powers of the observable $\langle x^n \rangle$.

- The 1st moment is the average: $\langle x \rangle = \int x f(x) dx \rightarrow ③$

and the mean square deviation from this is the variance:

$$\begin{aligned}\sigma^2 &= \langle (x - \langle x \rangle)^2 \rangle \\ &= \langle x^2 - 2x\langle x \rangle + \langle x \rangle^2 \rangle \\ &= \langle x^2 \rangle - \langle x \rangle^2 \rightarrow ④\end{aligned}$$

- The square root of the variance is the standard deviation σ .

- The probability distribution contains no information about the temporal properties of the observed quantity; a useful probe of this is the autocovariance function:

$$\langle x(t)x(t-\tau) \rangle = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\tau/2}^{\tau/2} x(t)x(t-\tau) d\tau \rightarrow ⑤$$

- If the autocovariance function is normalized by the variance then it is called the autocorrelation function, ranging from +1 for perfect correlation to 0 for no correlation to -1 for perfect anticorrelation.

- The rate at which it decays as a function of τ provides one way to determine how quickly a function is varying.
- Another way to look at it is to say the autocorrelation function provides a quantitative estimate of the memory in a given signal on average.
- Very soon we will introduce the concept of mutual information, a much more general way to measure the relationships among variables.

SPECTRAL THEOREMS:

The Fourier transform of a fluctuating quantity is:

$$X(f) = \lim_{T \rightarrow \infty} \int_{-T/2}^{T/2} e^{i2\pi ft} x(t) dt \quad \rightarrow (6)$$

and the inverse transform is

$$x(t) = \lim_{T \rightarrow \infty} \int_{-T/2}^{T/2} e^{-i2\pi ft} X(f) df \quad \rightarrow (7)$$

- The Fourier transform is also a random variable.

- The Power Spectral Density (PSD) is defined in terms of the Fourier transform by taking the average value of the square magnitude of the transform:

$$\begin{aligned} S(f) &= \langle |X(f)|^2 \rangle = \langle X(f) X^*(f) \rangle \\ &= \lim_{T \rightarrow \infty} \int_{-T/2}^{T/2} e^{i2\pi ft} x(t) dt \int_{-T/2}^{T/2} e^{-i2\pi f t'} x(t') dt' \quad \rightarrow (8) \end{aligned}$$

- X^* is the complex conjugate of X , replacing i with $-i$, and we shall assume that x is real.

- The power spectrum might not have a well-defined limit for a non-stationary process; wavelets and Wigner functions are examples of time-frequency transforms that retain both temporal and spectral information for non-stationary signals.

- The Fourier transform is defined for negative as well as positive frequencies.

- If the sign of the frequency is changed, the imaginary or sine component of the complex exponential changes sign while the real or cosine part does not.

- For a real-valued signal this means that the transform for negative frequencies is equal to the complex conjugate of the transform for positive frequencies.

- Since the power spectrum is used to measure energy as a function of frequency, it is usually reported as the single-sided power spectral density found by adding the square magnitudes of the negative- and positive-frequency components.

- For a real signal these are identical, and so the single-sided density differs from the two-sided density by an (occasionally omitted) factor of 2.

- The Fourier transform can also be defined with the 2π in front.

$$X(w) = \lim_{T \rightarrow \infty} \int_{-T/2}^{T/2} e^{-iwt} x(t) dt$$

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iwt} X(w) dw$$

- ν measures the frequency in cycles per second; ω measures the frequency in radians per second (2π radians = 1 cycle).
- Defining the transform in terms of ν eliminates errors that arise from forgetting to include the 2π in the inverse transform or in converting from radians to cycles per second, but it is less conventional in literature for historical reasons.
- We will use whichever is convenient for a problem.

- The power spectrum is simply related to the autocorrelation function by the Wiener-Khinchin Theorem, found by taking the inverse transform of the power spectrum:

$$\begin{aligned}
 & \int_{-\infty}^{+\infty} S(f) e^{-i2\pi f t} df = \int_{-\infty}^{+\infty} \langle X(t) X^*(t) \rangle e^{-i2\pi f t} df \\
 &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\infty}^{+\infty} \int_{-T/2}^{T/2} e^{i2\pi f t} x(t) dt \int_{-T/2}^{T/2} e^{-i2\pi f t'} x(t') e^{-i2\pi f t} dt' \\
 &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\infty}^{+\infty} \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} e^{i2\pi f (t-t'-\tau)} x(t) x(t') dt dt' d\tau \\
 &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} \delta(t-t'-\tau) x(t) x(t') dt dt' \quad \xrightarrow{\text{⑩}} \\
 &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t) x(t-\tau) dt = \langle x(t) x(t-\tau) \rangle
 \end{aligned}$$

using the Fourier transform of a delta function

$$\int_{-\infty}^{+\infty} e^{i2\pi xy} dx = \delta(y)$$

$$\int_{-\infty}^{+\infty} f(x) \delta(x-x_0) dx = f(x_0) \quad \xrightarrow{\text{⑪}}$$

- One way to derive these relations is by taking the delta function to be the limit of a Gaussian with unit norm as its variance goes to zero.

- The Wiener-Khinchin theorem shows that the Fourier transform of the autocorrelation function gives the power spectrum; knowledge of one is equivalent to the other.

An important example of this is white noise: a memoryless process with a delta function autocorrelation will have a flat spectrum, regardless of the probability function distribution for the signal.

As the autocorrelation function decays more slowly, the power spectrum decays more quickly (see fig 1 for some examples).

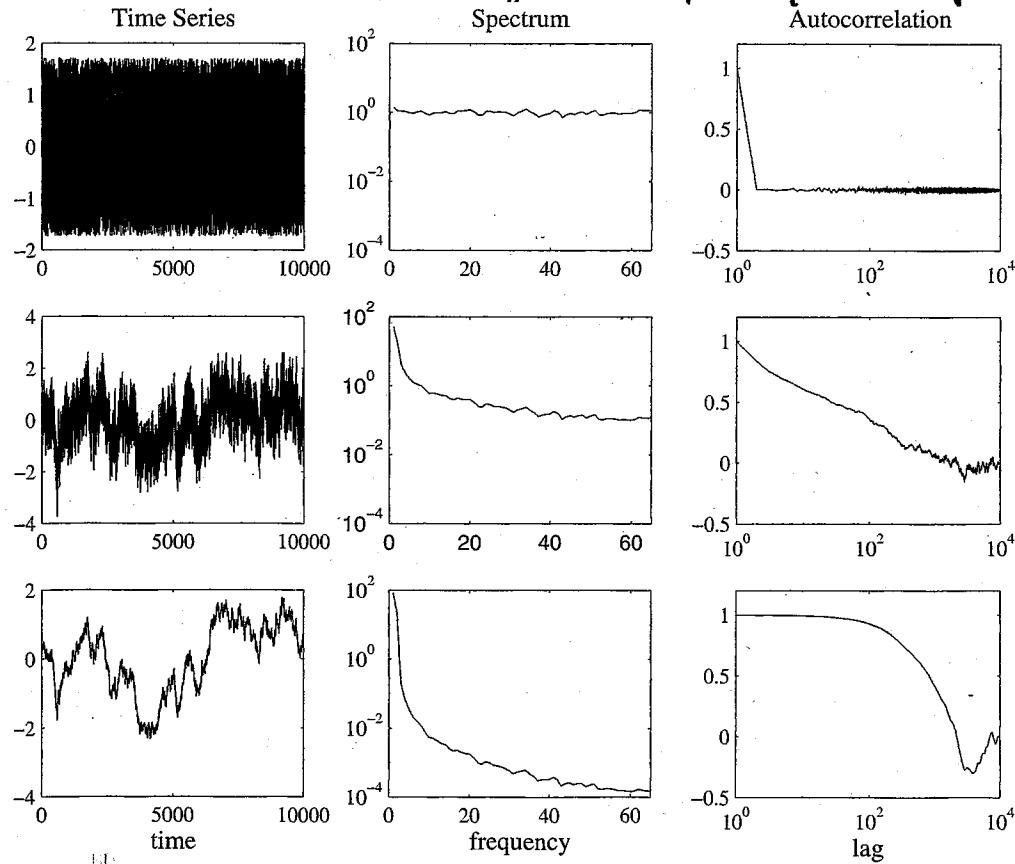


Fig 1: Illustration of the Wiener-Khinchin theorem: as the power spectrum decays more quickly, the autocorrelation function decays more slowly.

- Taking $T=0$ in the Wiener-Khinchin Theorem yields Pasewals Theorem:

$$\langle x(t)x(t-\tau) \rangle = \int_{-\infty}^{+\infty} S(f) e^{-i2\pi f t} df = \int_{-\infty}^{+\infty} \langle |X(f)|^2 \rangle e^{-i2\pi f t} df$$

$$\rightarrow \langle |x(t)|^2 \rangle = \int_{-\infty}^{+\infty} \langle |X(f)|^2 \rangle df \rightarrow 12$$

- The average value of the square of the signal (which is equal to the variance if the signal has zero mean) is equal to the integral of the power spectral density.

- This means that true white noise has an infinite variance in the time domain, although the finite bandwidth of any real system will roll off the frequency response, and hence determine the variance of the measured signal.

- If the division by T is left off in the limiting process defining the averages on both sides of the Pasewals theorem, then it reads that the total energy in the signal equals the total energy in the spectrum (the integral of the square of the magnitude).

* PROBABILITY DISTRIBUTIONS:

- So far we have taken the probability distribution $p(x)$ to be arbitrary, we did not delve into the functional form of $p(x)$.

- In practice, three probability distributions occur so frequently that they receive most attention: binomial, Poisson, and Gaussian.

- Their popularity is due in equal parts to the common conditions that give rise to them and to the convenience of working with them.

- The latter reason sometimes outweighs the former, leading these distributions to be employed far from where they apply.
- For example, many physical systems, particularly those driven strongly far away from thermal equilibrium where very nonlinear (and not linear) responses abide, have long-tailed distributions that fall off much more slowly than these ones do.
- We will look at a class of these long-tailed distributions, viz. power-law tailed distributions later on.

** BINOMIAL:

- Consider many trials of an event that can have one outcome with probability p (such as flipping a coin and seeing a head), and an alternative with probability $1-p$ (such as seeing a tail).
- In n trials, the probability $p_n(x)$ to see x heads and ~~($n-x$)~~ ($n-x$) tails, independent of the particular order in which they were seen, is found by adding up the probability for each outcome-times the number of equivalent arrangements:

$$p_n(x) = \binom{n}{x} p^x (1-p)^{n-x} \rightarrow ⑬$$

Where $\binom{n}{x} = \frac{n!}{(n-x)! x!} \rightarrow ⑭$

(read "n choose x").

- This is the binomial distribution.

- The 2nd line follows by dividing the total number of distinct arrangements of n objects ($n!$) by the number of equivalent distinct arrangements of heads $x!$ and tails $(n-x)!$.

- The easiest way to convince yourself that this is correct is to exhaustively count the possibilities for a small case.
- Small case because you'll see a combinatorial blowup with increasing number.

** Poisson:

- Now consider events such as radioactive decays that occur randomly in time.
- Divide time into n very small intervals so that either there are no decays or one decay in any one interval, and let p be the probability of seeing a decay in an interval.
- If the total number of events that occur in a given time is recorded, and this is repeated many times to form an ensemble of measurements, then the distribution of the total number of events recorded will be given by the binomial distribution.

If the number of events intervals n is large, and the probability p is small, the binomial distribution can be approximated by using $\ln(1+x) \approx x$ for small x and Stirling's approximation for large n :

$$n! \approx \sqrt{2\pi n} \frac{n^n}{e^{-n}}$$

$$\ln n! \approx n \ln n - n \quad \rightarrow 15$$

to find the Poisson distribution:

$$p(x) = \frac{e^{-N} N^x}{x!} \quad \rightarrow 16$$

where $N = np$ is the average number of events.

- This distribution is very common for measurements that require counting independent measurements of an event.

- Naturally, it is normalized:

$$\sum_{x=0}^{\infty} \frac{e^{-N} N^x}{x!} = e^{-N} \underbrace{\sum_{x=0}^{\infty} \frac{N^x}{x!}}_{e^N} = 1 \longrightarrow 17$$

- If x is drawn from a Poisson distribution then its factorial moments, defined by the following equation, have a simple form:

$$\langle x(x-1)\dots(x-m+1) \rangle = N^m \longrightarrow 18$$

- This relationship is one of the benefits of using the Poisson distribution approximation.

- With it, it is easy to show that $\langle x \rangle = N$ and $\sigma = \sqrt{N}$, which in turn implies that the relative standard deviation in a Poisson random variable is:

$$\frac{\sigma}{\langle x \rangle} = \frac{1}{\sqrt{N}} \longrightarrow 19$$

- The fractional error in an estimate of the average value will decrease as the square root of the number of samples.

- This important result provides a good way to make a quick estimate of the expected error in a counting experiment.

** Gaussian: The Gaussian or Normal distribution

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \longrightarrow 20$$

has a mean μ , a standard deviation σ , and the integral form from $-\infty$ to $+\infty$ is 1.

- The partial integral of a Gaussian is an error function:

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_0^y e^{-x^2/2\sigma^2} dx = \frac{1}{2} \operatorname{erf}\left(\frac{y}{\sqrt{2\sigma^2}}\right) \rightarrow (21)$$

- Since the Gaussian is normalized, $\operatorname{erf}(\infty) = 1$.

The Gaussian distribution is common for many reasons.

One way to derive it is from an expansion around the peak of the binomial distribution for large n .

$$p(x) = \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x}$$

$$\ln p(x) = \ln n! - \ln(n-x)! - \ln x! + x \ln p + (n-x) \ln(1-p) \rightarrow (22)$$

Finding the peak by treating these large integers as continuous variables and setting the first derivative to zero shows that this has a maximum at $x \approx np$, and then expanding in a power series around the maximum gives the coefficient of the quadratic term to be $-1/(2np(1-p))$.

Because the lowest non-zero term will dominate the higher orders for large n , this is therefore approximately a Gaussian with mean np and variance $np(1-p)$.

In the next section we will also see that the Gaussian distribution emerges via the Central Limit Theorem as the limiting form for an ensemble of variables with almost any distribution.

- for these reasons, it is often safe (and certainly common) to assume that an unknown distribution is Gaussian.

- The Fourier transform of a Gaussian has a particularly simple form, namely a Gaussian with the inverse of the variance.

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{-x^2/2\sigma^2} e^{ikx} dx = e^{-k^2\sigma^2/2} \quad (23)$$

Remember this: you should never need to look up the transform of a Gaussian, just invert the variance.

Because of this relationship, the product of the variance of a Gaussian and the variance of its Fourier transform will be a constant; this is the origin of many classical and quantum uncertainty relationships.

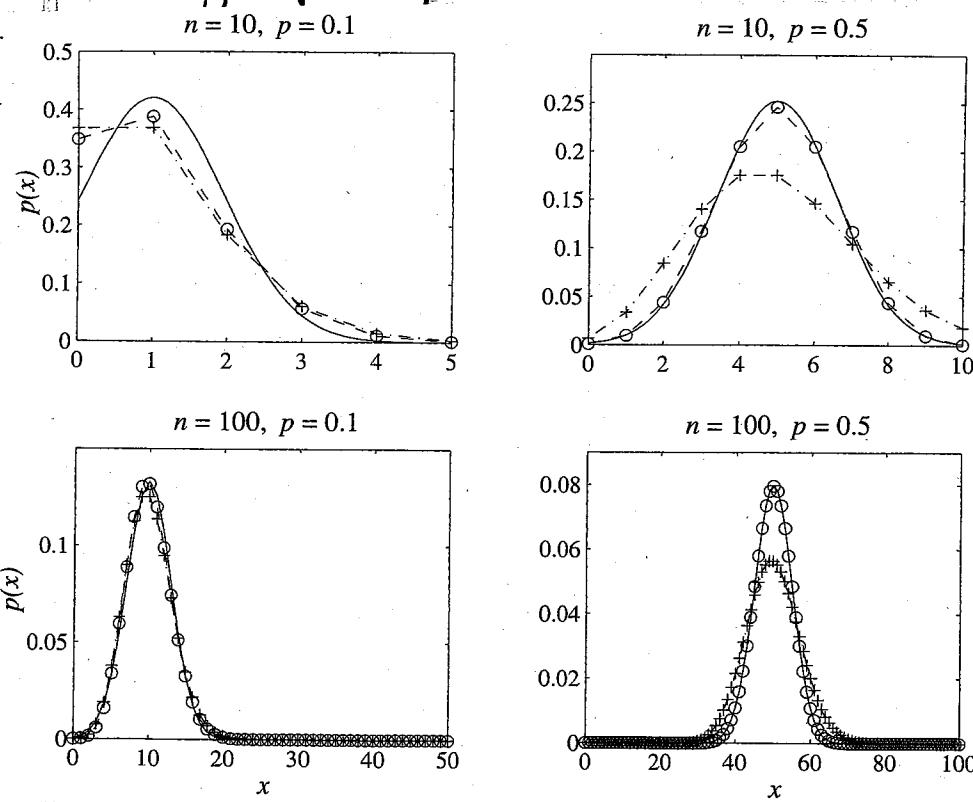


Fig 2

Fig 2: Comparison of the binomial (○), Poisson (+) and Gaussian (-) distributions: n is the number of trials and p the probability of seeing an event. By definition, the binomial distribution is correct. For a small probability of seeing an event, the Poisson

distribution is a better approximation (although the difference is small for a large number of events), while for a large probability of seeing an event the Gaussian distribution is closer.

Pg 13

- figure 2 compares the binomial, Poisson, and Gaussian distributions for $n=10$ and 100 , and for $p=0.1$ and 0.5 , showing where they are and are not good approximations.

****CENTRAL LIMIT THEOREM:** What is the probability distribution in a room full of people talking?

- This may sound like a nonsensical question, because the answer will depend on how many people there are, and on what is being said in what language.
- The remarkable result from the Central Limit Theorem is that if there is a large number of people in the room, then the distribution for their sum will approximately be Gaussian, independent of the details of what they say.
- If two random variables x_1 and x_2 are added (perhaps the sound from two random people), the probability distribution for their sum $y = x_1 + x_2$ is found by counting all the outcomes that give the same final result, weighted by the joint probability for that event:

$$p(y) = \int_{-\infty}^{+\infty} p_1(x) p_2(y-x) dx$$

$$\equiv p_1(x) * p_2(x)$$

- The distribution for the sum is the convolution of the individual distributions, and considering the average of N variables

$$y = \frac{x_1 + x_2 + \dots + x_N}{N} \rightarrow \text{25}$$

that are independent and identically distributed (abbreviated as iid).

- The distribution of y is equal to the distribution of x convolved with itself N times, and since taking a Fourier transform turns convolution into multiplication, the Fourier transform of the distribution of y is equal to the product of the transforms of the distribution of x .

- It is convenient to take the transform of a probability distribution by using the characteristic function, which is the expectation value of a complex exponential

$$\langle e^{iky} \rangle = \int_{-\infty}^{+\infty} e^{iky} p(y) dy \quad \rightarrow (26)$$

- The characteristic function is equal to the Fourier transform of the probability distribution, and when evaluated with time-dependent quantities it plays an interesting role in studying the dynamics of a system.

- Now let's look at the characteristic function for the deviation of y from the average value $\langle x \rangle$:

$$\begin{aligned} \langle e^{ik(y-\langle x \rangle)} \rangle &= \left\langle e^{ik(x_1 + x_2 + \dots + x_N - N\langle x \rangle)/N} \right\rangle \\ &= \left\langle e^{ik[(x_1 - \langle x \rangle) + \dots + (x_N - \langle x \rangle)]/N} \right\rangle \\ &= \left\langle e^{ik(x - \langle x \rangle)/N} \right\rangle \\ &= \left\langle 1 + ik(x - \langle x \rangle) - \frac{k^2}{2N^2}(x - \langle x \rangle)^2 + O\left(\frac{k^3}{N^3}\right) \right\rangle^N \\ &= \left[1 + 0 - \frac{k^2 \sigma^2}{2N^2} O\left(\frac{k^3}{N^3}\right) \right]^N \\ &\approx e^{-k^2 \sigma^2 / N^2} \quad \rightarrow (27) \end{aligned}$$

- This derivation assumes that the variance $\sigma^2 = \langle (x - \langle x \rangle)^2 \rangle$ exists, and drops terms of 3rd order & higher in the Taylor series expansion of the exponential because they will become vanishingly small compared to the lower-order terms in the limit $N \rightarrow \infty$.

- The last line follows because an exponential can be written as

$$\lim_{N \rightarrow \infty} \left(1 + \frac{x}{N}\right)^N = e^x \quad \rightarrow 28$$

which can be verified by comparing the Taylor series of both sides.

- To find the probability distribution for y we now take the inverse transform

$$\begin{aligned} p(y - \langle x \rangle) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-k^2/2N} e^{-ik(y - \langle x \rangle)} dk \\ &= \sqrt{\frac{N}{2\pi}} e^{-N(y - \langle x \rangle)^2/2\sigma^2} \end{aligned} \quad \rightarrow 29$$

(remember that the Fourier transform of a Gaussian is also a Gaussian).

- This proves the Central Limit Theorem.

- The average of N iid variables has a Gaussian distribution with a standard deviation σ/\sqrt{N} reduced by the square root of the number of variables just as with the Poisson process statistics and can be a surprisingly good approximation even with just tens of samples.

- The Central Limit Theorem also contains the Law of Large Numbers: in the limit $N \rightarrow \infty$, the average of N random variables approaches the mean of their distribution.

- Although this might appear to be a trivial insight, lurking behind it is the compressibility of data that is so crucial to digital coding.