Lesson 1- Probability

What is the number of 1 gram of hydrogen atoms? or
What is the number of molecules of water in a tablespoon?

What is the number of 1 gram of hydrogen atoms? or
What is the number of molecules of water in a tablespoon?

Avogadro's number $N_A = 6.02 \times 10^{23}$

602.000.000.000.000.000.000.000

six hundred two thousand trillions

In statistical mechanics, we work with a large number N of particles and calculate things as a Taylor expansion in $\frac{1}{N}$, often keeping only the leading term ($N \to \infty$).

The key point is not to ask what each particle is doing, impossible and impractical, but rather to ask **what the probability is that a particle is doing something**.

We will be interested in probabilities of states of a system which we write as P_a or $P(a)$. The parameter a represents the **microstate** e.g. the positions $\{\overrightarrow{q_i}\}$ and momenta $\{\overrightarrow{p_i}\}$ of all the particles in a gas, or the square of the wavefunction $|\Psi(\overrightarrow{q})|^2$ in quantum mechanics.

We will sometimes think of a as a discrete index (e.g. if we flip a coin, it can land heads up with $P_H = 1/2$ or tails up with $P_T = 1/2$ and sometimes continuous. In the continuous case, we call $P(x)$ the probability density, so that $\int_{x_1}^{x_2} P(x)dx$ is the probability of finding x values between x1 and x2. **Probability densities only become probabilities when integrated**.

Probabilities distributions are always normalized so that they integrate/sum to 1:

$$\int P(x)dx = 1, \text{ or in discrete } \sum_a P_a = 1$$

Given a probability distribution, we can calculate the expected value of any observable by integrating/summing against the probability. For example, the expected value of x (the mean) is:

$$\bar{x} \equiv \langle x \rangle = \int xP(x)dx$$

or the mean-square:

$$\langle x^2 \rangle = \int x^2 P(x)dx$$

The **variance** of a distribution is the difference between the mean of the square and the square of the mean:

$$Var \equiv \langle x^2 \rangle - \langle x \rangle^2$$

The square root of the variance is called the **standard deviation**

$$\sigma \equiv \sqrt{\langle x^2 \rangle - \langle x \rangle^2}$$

developing intuition for variance is a key to mastering statistics. The key point is that the **expected value** is worthless if you don't know **how likely that value is**.

Let us fix our attention in the Gaussian probability density

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma_0} exp\left(-\frac{(x-x_0)^2}{2\sigma_0^2}\right)$$

A little bit of calculus, let us **prove** for the Gaussian that:

$$\langle x \rangle = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma_0}} exp\left(-\frac{(x-x_0)^2}{2\sigma_0^2}\right) dx = x_0$$

and that

$$\sigma \equiv \sqrt{\langle x^2 \rangle - \langle x \rangle^2} = \sigma_0$$

The standard deviation has an interpretation as the width of a distribution i.e. **how far you can go from the mean** before the **probability has decreased substantially**. For example, in a Gaussian, the probability of finding x between $x_0 - \sigma$ and $x_0 + \sigma$ is 0.68

Proof:

$$\langle x \rangle = \int_{-\infty}^{+\infty} x \cdot \frac{1}{\sqrt{2\pi\sigma}} \cdot \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \mathrm{d}x$$

$$= \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{+\infty} x \cdot \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \mathrm{d}x \ .$$

now substituting $z = x - \mu$

Proof:

$$\langle x \rangle = \int_{-\infty}^{+\infty} x \cdot \frac{1}{\sqrt{2\pi\sigma}} \cdot \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \mathrm{d}x$$

$$= \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{+\infty} x \cdot \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \mathrm{d}x \, .$$

now substituting $z = x - \mu$

$$\langle x \rangle = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{+\infty} (z+\mu) \cdot \exp\left[-\frac{1}{2}\left(\frac{z}{\sigma}\right)^2\right] \mathrm{d}z$$

$$= \frac{1}{\sqrt{2\pi\sigma}} \left(\int_{-\infty}^{+\infty} z \cdot \exp\left[-\frac{1}{2}\left(\frac{z}{\sigma}\right)^2\right] \mathrm{d}z + \mu \int_{-\infty}^{+\infty} \exp\left[-\frac{1}{2}\left(\frac{z}{\sigma}\right)^2\right] \mathrm{d}z \right)$$

$$= \frac{1}{\sqrt{2\pi\sigma}} \left(\int_{-\infty}^{+\infty} z \cdot \exp\left[-\frac{1}{2\sigma^2} \cdot z^2\right] \mathrm{d}z + \mu \int_{-\infty}^{+\infty} \exp\left[-\frac{1}{2\sigma^2} \cdot z^2\right] \mathrm{d}z \right) \, .$$

Proof:

$$
\begin{aligned}
< x > \ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} (z+\mu) \cdot \exp\left[-\frac{1}{2}\left(\frac{z}{\sigma}\right)^2\right] \mathrm{d}z \\
&= \frac{1}{\sqrt{2\pi}\sigma} \left(\int_{-\infty}^{+\infty} z \cdot \exp\left[-\frac{1}{2}\left(\frac{z}{\sigma}\right)^2\right] \mathrm{d}z + \mu \int_{-\infty}^{+\infty} \exp\left[-\frac{1}{2}\left(\frac{z}{\sigma}\right)^2\right] \mathrm{d}z \right) \\
&= \frac{1}{\sqrt{2\pi}\sigma} \left(\int_{-\infty}^{+\infty} z \cdot \exp\left[-\frac{1}{2\sigma^2} \cdot z^2\right] \mathrm{d}z + \mu \int_{-\infty}^{+\infty} \exp\left[-\frac{1}{2\sigma^2} \cdot z^2\right] \mathrm{d}z \right) .
\end{aligned}
$$

by symmetry the first integral is 0, and the second integral was proven at AM2 that:

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$$

so doing a bit of algebra:

$$< x > = \mu$$

For the variance, the proof is similar, but you will end up with the integral

$$\frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} t^2 e^{-t^2} dt$$

that can be solved integrating by parts, and yields for the Gaussian distribution:

$$Var \equiv \langle x^2 \rangle - \langle x \rangle^2 = \sigma^2$$

We will often be interested in situations where the mean is zero. Then the standard deviation is equivalent to the **root-mean-square**

$$\sqrt{\langle x^2 \rangle} = x_{RMS}$$

Example, in a **gas** the velocities point in random directions, so $\langle \overrightarrow{v} \rangle = 0$. Thus the characteristic speed of a gas is characterized not by the mean but by the RMS velocity $v_{RMS} = \sqrt{\langle \overrightarrow{v}^2 \rangle}$.

Suppose now that $P_A(x)$ and $P_B(y)$ are the probabilities of winning x dollars when betting on horse A and y dollars when betting on horse B. What is the probability of getting z total dollars?

Suppose now that $P_A(x)$ and $P_B(y)$ are the probabilities of winning x dollars when betting on horse A and y dollars when betting on horse B. What is the probability of getting z total dollars?

Note that we are looking for the probability of obtaining z total dollars, which is the result of two independent random variables, x and y. then our focus is to know $P_{AB}(z) = P_{AB}(x + y)$.

Let us work a simple case in which x and y are discrete values x={0,1,2} and y={0,2,4}. Let us compute:

$$
\begin{aligned}
P_{AB}(z = 4) &= P(x = 0, y = 4) + P(x = 1, y = 3) + P(x = 2, y = 2) \\
&= 1/3 \cdot 1/3 + 1/3 \cdot 0 + 1/3 \cdot 1/3 = 2/9
\end{aligned}
$$

In general, for two discrete independent random variables:

$$
\begin{aligned}
P_{AB}(z) = P_{AB}(Z = z) &= \sum_{x \in \Omega_X} P_{AB}(X = x, Y = z - x) \\
&= \sum_{x \in \Omega_X} P_A(X = x) P_B(Y = z - x) \\
&= \sum_{x \in \Omega_X} P_A(x) P_B(z - x)
\end{aligned}
$$

This is the definition of the mathematical operation of **convolution** between two functions. We say P_{AB} is the convolution of P_A and P_B and write it as

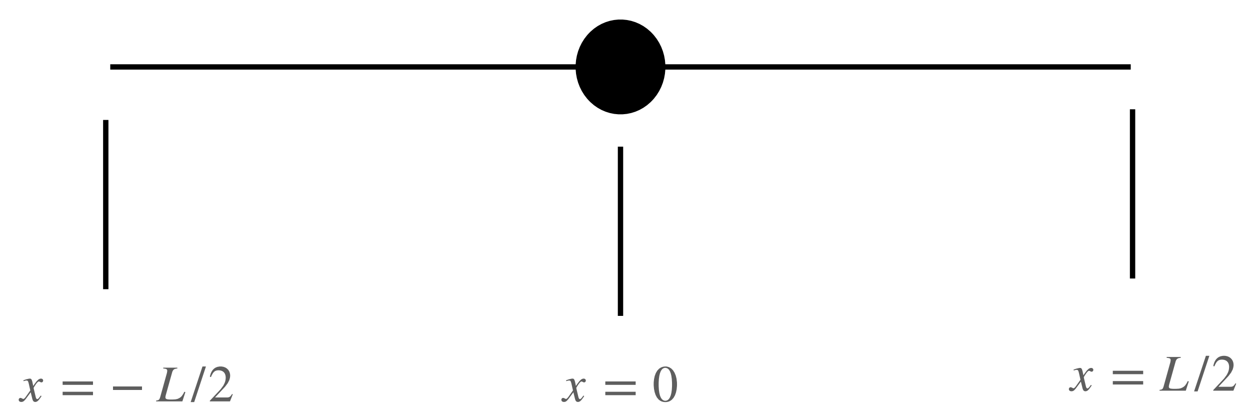
$$P_{AB} = P_A * P_B$$

For random independent continuous variables, we write the convolution as a Riemann integral of the two involved functions:

$$P_{AB}(z) = \int_{-\infty}^{\infty} P_A(x)P_B(x-z)dx = \int_{-\infty}^{\infty} P_A(z-x)P_B(x)dx$$

Note that by definition convolution is commutative.

Exemple: Consider the system of a gas molecule bouncing around in a 1D box of size L centered on $x = 0$. If there are no external forces and no position-dependent interactions, the molecule is equally likely to be anywhere in the box.



Exercise: Consider the system of a gas molecule bouncing around in a 1D box of size L centered on x=0. If there are no external forces and no position-dependent interactions, the molecule is equally likely to be anywhere in the box.

Solution:

$$P(x) = \frac{1}{L}$$

The mean value of the position of the molecule is

$$\langle x \rangle = \int_{-L/2}^{L/2} x \frac{1}{L} dx = 0$$

Exercise: Consider the system of a gas molecule bouncing around in a 1D box of size L centered on x=0. If there are no external forces and no position-dependent interactions, the molecule is equally likely to be anywhere in the box.

Similarly, the mean value of x^2 is

$$\langle x^2 \rangle = \int_{-L/2}^{L/2} x^2 \frac{1}{L} dx = L^2/12$$

so that the standard deviation is:

$$\sigma = \sqrt{\langle x^2 \rangle - \langle x \rangle^2} = L/\sqrt{12} \approx 0.29L$$

Exercise: Consider the system of a gas molecule bouncing around in a 1D box of size L centered on x=0. If there are no external forces and no position-dependent interactions, the molecule is equally likely to be anywhere in the box.

Note that the probability of finding x within $\langle x \rangle \pm \sigma$ is $2\sigma/L = 58\,\%$. It is not $68\,\%$ because the probability distribution is not Gaussian.

Exercise: Suppose now that there is some electric field so that the particles in the box are more likely to be on one side than the other. We might find some funny function for these probabilities, for example:

$$P(x) = \frac{0.74}{L} ln(1 + e^{\frac{2x}{L}})$$

Exercise: Suppose now that there is some electric field so that the particles in the box are more likely to be on one side than the other. We might find some funny function for these probabilities, for example:

$$P(x) = \frac{0.74}{L} ln(1 + e^{\frac{2x}{L}})$$

Solution:

$$\langle x \rangle = 0.59L, \langle x^2 \rangle = 0.42L^2, \sigma = 0.28L$$

and

$$\int_{\langle x \rangle + \sigma}^{\langle x \rangle - \sigma} P(x)dx = 0.6 \quad \text{so } 60\,\% \text{ within } \langle x \rangle \pm \sigma.$$

Law of large numbers: An extremely important result from probability

The average of the results from a set of independent trials varies less and less the more trials are performed

Mathematically, we can state it this way: If $P(x)$ has standard deviation σ, then the probability $P_N(x)$ of finding that the average over N draws from $P(x)$ is x will have standard deviation $\frac{\sigma}{\sqrt{N}}$.

Thus as $N \to \infty$, the standard deviation $\frac{\sigma}{\sqrt{N}} \to 0$.

Law of large numbers: The average of the results from a set of independent trials varies less and less the more trials are performed

Proof:
Lets consider the probability distribution for the center of mass of molecules in a box. Say there are N molecules in the box and the probability function of finding each is $P(x)$. Assume that the probabilities for each molecule are independent having one at x does not tell us anything about where the others might be. For N=2, the center of mass is $x = \dfrac{x_1 + x_2}{2}$, so the mean value of the center of mass, and its deviation is:

$$\begin{aligned}
\langle x \rangle_2 &= \int_{-L/2}^{L/2} dx_1 \int_{-L/2}^{L/2} dx_2 P(x_1) P(x_2) \frac{x_1 + x_2}{2} \\
&= \int_{-L/2}^{L/2} dx_1 \frac{x_1}{2} P(x_1) + \int_{-L/2}^{L/2} dx_2 \frac{x_2}{2} P(x_2) \\
&= \langle x \rangle \\
\langle x^2 \rangle_2 &= \int_{-L/2}^{L/2} dx_1 \int_{-L/2}^{L/2} dx_2 P(x_1) P(x_2) \frac{x_1 + x_2}{2} \\
&= \int_{-L/2}^{L/2} dx_1 \frac{x_1}{2} P(x_1) + \int_{-L/2}^{L/2} dx_2 \frac{x_2}{2} P(x_2) \\
&= \frac{1}{2} \langle x^2 \rangle - \frac{1}{2} \langle x \rangle^2
\end{aligned}$$

$$\sigma_2 = \sqrt{\langle x^2 \rangle_2 - (\langle x \rangle_2)^2} = \sqrt{\frac{1}{2}\langle x^2 \rangle + \frac{1}{2}\langle x \rangle^2 - \langle x \rangle^2}$$

$$= \frac{1}{\sqrt{2}}\sqrt{\langle x^2 \rangle - \langle x \rangle^2} = \frac{\sigma}{\sqrt{2}}$$

if we follow the same sequence of calculations for N particles, you can prove that:

$$\sigma_N = \sqrt{\langle x^2 \rangle_N - (\langle x \rangle_N)^2} = \frac{1}{\sqrt{N}}\sqrt{\langle x^2 \rangle - \langle x \rangle^2} = \frac{\sigma}{\sqrt{N}}$$

For the gas in the box with a flat $P(x) = 1/L$, the expected value of the center of mass is $\langle x \rangle_N = 0$, just like for any individual gas molecules, and the standard deviation is $\sigma_N = \sigma/\sqrt{N} \approx 10^{-11} L/12$. Thus, even though we don't know very well where any of the molecules are, we know the center of mass to extraordinary precision.

The law of large numbers is **the reason that statistical mechanics is possible**: we can compute macroscopic properties of systems (like the center of mass, or pressure, or all kinds of other things) with great confidence even if we don't know exactly what is going on at the microscopic level.

Central limit theorem:

When any probability distribution P(x) is sampled N times the average of the samples approaches a Gaussian distribution as $N \to \infty$ with width scaling like $\sigma \sim \frac{1}{\sqrt{N}}$.

Proof: One way to prove the central limit theorem is by computing *moments*. If you specify the complete set of moments of a function, you know its shape completely.

Moments of a distribution

The n-th moment μ_n about zero of a probability density function f(x) is the expected value of

$$\mu_n = \mathbb{E}[X^n] = \int_{-\infty}^{\infty} x^n P(x) dx$$

and is called a raw moment or crude moment. The moments about its mean $\bar{x}$ are called *central moments*; these describe the shape of the function, independently of translation. The standardized moments are defined as:

$$\frac{\mu_n}{\sigma_n} = \frac{\mathbb{E}[(X - \mu)^n]}{\sigma^n}$$

Moments of a distribution

The first four moments have specific names: mean, variance, skewness, and kurtosis, for higher-order moments there are no specific names.

Skewness measures how **asymmetric** a distribution is around its mean. Kurtosis measures the 4th derivative, which is a measure of **curvature**. More intuitively, higher kurtosis means a probability distribution has a longer tail, i.e. more outliers from the mean. The higher moments do not have simple interpretations.

Moments of a distribution

For the Gaussian probability distribution

$$\frac{1}{\sqrt{2\pi}\sigma_0} e^{\left(-\frac{(x-x_0)^2}{2\sigma_0^2}\right)}$$

the central standardized moments are “easy” to calculate:

$$\bar{x} = 0, \quad \sigma = 1, \quad S = 0, \quad K = 3, \quad M_5 = 0, \quad M_6 = 15, \quad M_7 = 0, \quad M_8 = 105, \ldots$$

It can be proved that, for a Gaussian that, in general:

$$M_n = \begin{cases} 0 & \text{if } n \text{ odd} \\ 2^{\frac{-n}{2}} \frac{n!}{(\frac{n}{2})!} & \text{if } n \text{ even} \end{cases}$$

Moments of a distribution

Now let's compute the moments of the center of mass of our N molecules-in-a-box with probability P(x). We'll do this for a general P(x), but shift the domain so that $\langle x \rangle = 0$ to simplify the formulas. It is trivial to check that the two first moments do coincide with that of the Gaussian. Check the 3rd moment:

$$\langle x^3 \rangle_N = \int_{-L/2}^{L/2} dx_1 \cdots dx_N P(x_1) \cdots P(x_N) \left(\frac{x_1 + \cdots + x_N}{N} \right)^3 = \frac{\langle x^3 \rangle}{N^2}$$

In particular, the skewness goes to zero as $N \to \infty$. That is, the distribution becomes more and more symmetric around the mean as $N \to \infty$.

Moments of a distribution

The 4th moment, the kurtosis, is far more interesting:

$$\langle x^4 \rangle_N = \int_{-L/2}^{L/2} dx_1 \cdots dx_N P(x_1) \cdots P(x_N) \left(\frac{x_1 + \cdots + x_N}{N} \right)^4$$

$$= \frac{\langle x^4 \rangle}{N^3} + \frac{3(N-1)}{N^3} \langle x^2 \rangle \langle x^2 \rangle$$

Remember that:

$$K_N = \frac{\langle (x - \bar{x})^4 \rangle_N}{\sigma_N^4} = \frac{1}{\sigma^4 / N^2} \left(\frac{\langle x^4 \rangle}{N^3} + \frac{3(N-1)}{N^3} \langle x^2 \rangle \langle x^2 \rangle \right)$$

Moments of a distribution and Central Limit Theorem

So that,

$$K_N = \frac{\langle (x-\bar{x})^4 \rangle_N}{\sigma_N^4} = \left(\frac{\langle x^4 \rangle}{\sigma^4 N} + \frac{3\left(1-\frac{1}{N}\right)}{\sigma^4} \langle x^2 \rangle \langle x^2 \rangle \right)$$

and taking into account that $\sigma^2 = \langle x^2 \rangle$ and taking the large *N* limit, we get that: $\lim_{N\to\infty} K_N \to 3$ (Gaussian value for *K*).

In fact for large N, *each one of the moments is equal to that of a Gaussian distribution, which proves the Central Limit Theorem*.

Moments of a distribution and CLT

So that,

$$K_N = \frac{\langle (x-\bar{x})^4 \rangle_N}{\sigma_N^4} = \left(\frac{\langle x^4 \rangle}{\sigma^4 N} + \frac{3\left(1-\frac{1}{N}\right)}{\sigma^4} \langle x^2 \rangle \langle x^2 \rangle \right)$$

and taking into account that $\sigma^2 = \langle x^2 \rangle$ and taking the large *N* limit, we get that: $\lim\limits_{N\to\infty} K_N \to 3$ (Gaussian value for *K*).

In fact for large N, *each one of the moments is equal to that of a Gaussian distribution, which proves the Central Limit Theorem*.

Central limit theorem:

The distribution of the mean of N draws from a probability distribution approaches a Gaussian of width $\dfrac{\sigma}{\sqrt{N}}$ as $N \to \infty$ independent of the original probability distribution, i.e.

$$P_N(x) \to \sqrt{\frac{N}{2\pi\sigma^2}} \exp\left(-N\frac{(x-\bar{x})^2}{2\sigma^2}\right)$$

Physical interpretation

Imagine we have some probability distribution $P(x)$ for molecules in a box, with $-L/2 < x < L/2$. We want to pick N molecules and compute their mean position (center of mass position)

$$x = \frac{1}{N} \sum_j x_j$$

What is the probability distribution $P_N(x)$ that the mean value is x?

Physical interpretation

Let's take the flat distribution $P(x) = 1/L$. For $N = 1$, we pick only molecule with position x_1. Then $x = x_1$ and so $P(x) = 1/L$: any value for the center-of-mass position is equally likely.

Now say $N=2$, so we pick two molecules with positions x_1 and x_2. What is the probability that they will have mean x? For a given x, we need

$$\frac{x_1 + x_2}{2} = x$$

Physical interpretation

For example if $x=0$, then for any x_1 there is an x_2 that works, namely $x_2=-x_1$. However, if the mean is all the way on the edge, $x= L/2$, then not all x_1 work; in fact, we need both x_1 and x_2 to be exactly $L/2$. Thus there are fewer possibilities when x is close to the boundaries of the box than if x is central.

Mathematically, we can write the probability for getting a mean value $x= (x_1+x_2)/2$ as

$$P_2(x) = \int_{-\frac{L}{2}}^{\frac{L}{2}} dx_1 P(x_1) \int_{-\frac{L}{2}}^{\frac{L}{2}} dx_2 P(x_2) \delta(\frac{x_1 + x_2}{2} - x)$$

Physical interpretation

$$P_2(x) = \int_{-\frac{L}{2}}^{\frac{L}{2}} dx_1 P(x_1) \int_{-\frac{L}{2}}^{\frac{L}{2}} dx_2 P(x_2) \delta(\frac{x_1 + x_2}{2} - x)$$

This is another way of writing a convolution $P_2 = P * P$

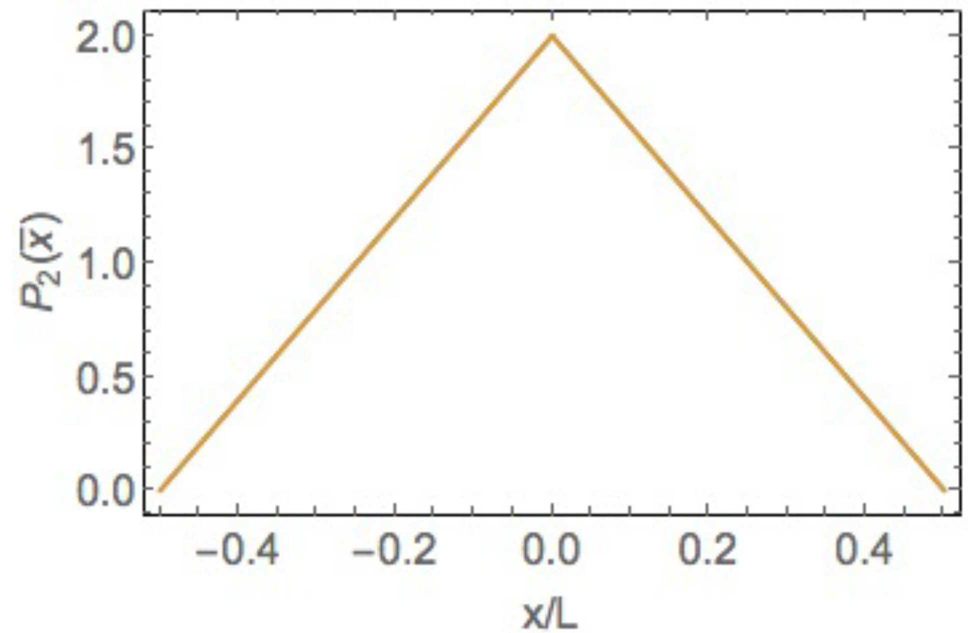
$$P_2(x) = 2 \int_{-\frac{L}{2}}^{\frac{L}{2}} dx_1 P(x_1) \int_{-\frac{L}{2}}^{\frac{L}{2}} dx_2 P(x_2) \delta(x_1 + x_2 - 2x)$$

We can solve for $x_1 = 2x - x_2$. We compute the integral for (x<0)

$$P_2(x) = 2 \int_{-\frac{L}{2}}^{\frac{L}{2}+2x} dx_1 P(x_1) 2P(2x - x_1)$$

Physical interpretation

Taking the flat distribution $P(x) = 1/L$ then
$P_2(x < 0) = 2L + 4x$, and $P_2(x > 0) = 2L - 4x$

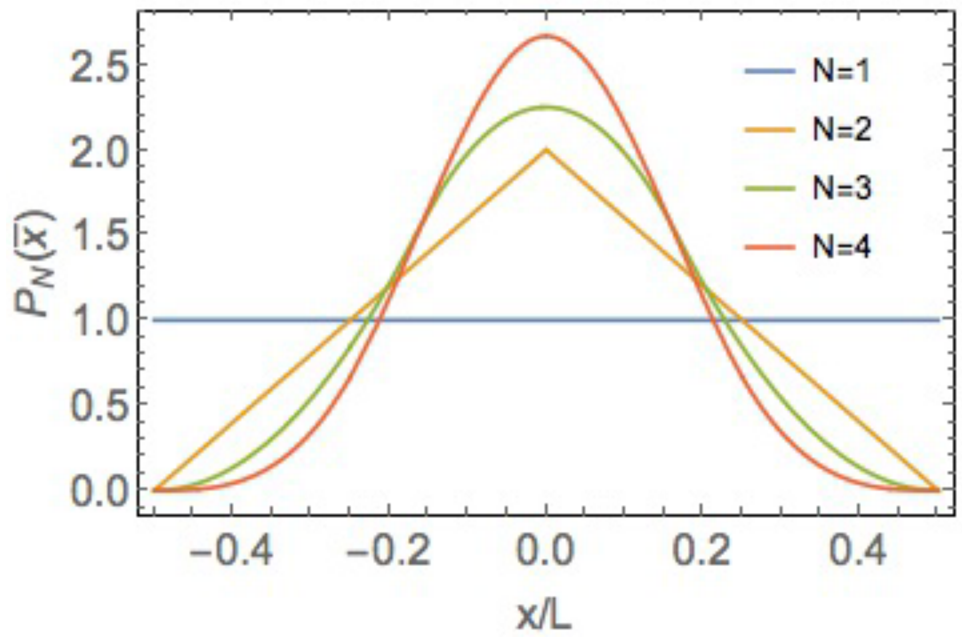


For N =3 we compute

$$P_3(x) = \int_{-\frac{L}{2}}^{\frac{L}{2}} dx_1 P(x_1) \int_{-\frac{L}{2}}^{\frac{L}{2}} dx_2 P(x_2) \int_{-\frac{L}{2}}^{\frac{L}{2}} dx_3 P(x_3) \delta(\frac{x_1 + x_2 + x_3}{3} - x)$$

Physical interpretation

The successive approximations look like



Note: Sometimes we sum the values of the draws from a distribution instead of averaging them. In this case, the mean grows as $\bar{x} \rightarrow N\bar{x}$ and the standard deviation grows like $\sigma \rightarrow \sqrt{N}\sigma$.

Physical interpretation

Generally we have systems composed of enormously large numbers of particles. Because of the central limit theorem, the distribution of any **macroscopic quantity** will be close to a Gaussian around its mean. It tells us that we don't need to worry about the precise details of the microscopic description.

Mathematical trick

It is extremely useful to take logarithms of our probabilities.

Mathematical trick

The Taylor expansion of a Gaussian has an infinite number of terms:

$$e^{-\frac{x^2}{2\sigma^2}} = \sum_{m=0}^{\infty} \frac{1}{m!}\left(-\frac{x^2}{2\sigma^2}\right) = 1 - \frac{x^2}{2\sigma^2} + \frac{1}{2}\left(-\frac{x^2}{2\sigma^2}\right)^2 - \frac{1}{6}\left(-\frac{x^2}{2\sigma^2}\right)^3 + \cdots$$

however if we take the logarithm, only one term remains:

$$\ln e^{-\frac{x^2}{2\sigma^2}} = -\frac{x^2}{2\sigma^2}$$

Then

$$\ln P_N(x) \to -N\frac{(x-\bar{x})}{2\sigma^2} + \ln\sqrt{\frac{N}{2\pi\sigma^2}}$$

More mathematical tricks

Stirling's approximation: We start by

$$\ln N! = \ln N + \ln(N-1) + \ln(N-2) + \cdots + \ln 1 = \sum_{j=1}^{N} \ln j$$

For large N we can approximate the sum by an integral:

$$\ln N! = \sum_{j=1}^{N} \ln j \approx \int_1^N \ln x dx = N \ln N - N - 1 \approx N \ln N - N$$

So as a first approximation:

$$\ln N! \approx N \ln N - N$$

More mathematical tricks

Stirling's approximation, **Euler proof**. We start with this identity:

$$n! = \int_0^\infty x^n e^{-x} dx$$

Proof by induction (case 0 is trivial)

$$\int_0^\infty x^{n+1} e^{-x} dx = \int_0^\infty u dv \text{ with}$$

$$u = x^{n+1} \qquad du = (n+1)x^n dx$$
$$dv = e^{-x} dx \qquad v = e^{-x}$$

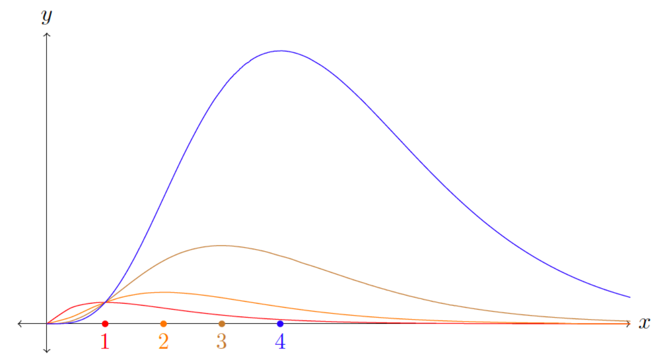
Euler proof

$$\begin{aligned}\int_0^\infty x^{n+1}e^{-x}dx &= uv\,\Big|_0^\infty - \int_0^\infty vdu \\ &= -\frac{x^{n+1}}{e^x}\Bigg|_0^\infty + \int_0^\infty (n+1)e^{-x}x^n dx \\ &= \lim_{b\to\infty} -\frac{b^{n+1}}{e^b} + 0 + (n+1)\int_0^\infty x^n e^{-x}dx \\ &= (n+1)\int_0^\infty x^n e^{-x}dx,\end{aligned}$$

So $(n+1)n! = (n+1)!$

Euler proof

If we look at the plot of $x^n e^{-x}$ for different values of n we observe that the larger n, the more similar to a Gaussian.



Let us then propose a change of variables in Euler's formula $t = (x - n)/\sqrt{n}$ then

$$
\begin{aligned}
n! &= \int_0^\infty x^n e^{-x} dx \\
&= \int_{-\sqrt{n}}^\infty (n + \sqrt{n}t)^n e^{-(n+\sqrt{n}t)} \sqrt{n} dt \\
&= \frac{n^n \sqrt{n}}{e^n} \int_{-\sqrt{n}}^\infty \left(1 + \frac{t}{\sqrt{n}}\right)^n e^{-\sqrt{n}t} dt\,.
\end{aligned}
$$

Euler proof

Let us analyze the previous integrand

$$f_n(t) = \left(1 + \frac{t}{\sqrt{n}}\right)^n e^{-\sqrt{n}t} dt$$

so

$$\log\left(f_n(t)\right) = n \log\left(1 + \frac{t}{\sqrt{n}}\right) - \sqrt{n}t$$

now use Taylor for

$\log(1+x) = x - x^2/2 + O\left(|x|^3\right)$ $\log(1+x) = x - x^2/2 + O\left(|x|^3\right)$

then:

$$\log\left(f_n(t)\right) = n\left(\frac{t}{\sqrt{n}} - \frac{(t/\sqrt{n})^2}{2} + O\left((t/\sqrt{n})^3\right)\right) - \sqrt{n}t =$$

$$-\frac{t^2}{2} + O\left(t^3/\sqrt{n}\right).$$

Euler proof

Summarizing, as $n \to \infty$

$$\left(1+\frac{t}{\sqrt{n}}\right)^{n} e^{-\sqrt{n}t}dt \to e^{-t^2/2}$$

so

$$n! = \frac{n^n\sqrt{n}}{e^n}\int_{-\sqrt{n}}^{\infty}\left(1+\frac{t}{\sqrt{n}}\right)^{n} e^{-\sqrt{n}t}dt \underset{n\to\infty}{\to} \frac{n^n\sqrt{n}}{e^n}\int_{-\infty}^{\infty} e^{-t^2/2}dt$$

and now remember what the value of the last integral is, to obtain the Stirling's estimate

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

Euler proof

Taking log we have a better Stirling's estimate

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

Using logarithms properties:

$$\ln n! \approx n \ln n - \ln e^n + \ln \sqrt{2\pi n} = n \ln n - n + \frac{1}{2} \ln 2\pi n$$