INTRODUCTION TO PROBABILITY THEORY

Thermodynamics: macroscopic description of bodies
How to put it in correspondance with the fact that matter is composed of atoms and molecules? Statistical mechanics!

As the name suggests, stat. mech. is a inherently probabilistic description of physical systems $\rightarrow$ manipulations of probabilities is an important prerequisite.

- A probability space is a set of possible "events" Ω and a function

$$P : \{\text{subset of } \Omega\} \longrightarrow \overbrace{[0,1]}^{\text{unit interval}}$$ such that:

1) $P(\Omega) = 1$ or equivalently $P(\{\}) = 0$
2) the probability of two $\underbrace{\text{disjoint}}_{A \cap B = \emptyset}$ sets of Ω, respectively A and B,
$$\underbrace{P(A \cup B)}_{\text{union of A and B}} = P(A) + P(B)$$

These are the Kolmogorov axioms of probability which state the positivity, normalitation and countable additivity of probability.

- A random variable X is a function from the "space of events" Ω to the "space of outcomes".

ex: Ω all possible dice rolls

E = space of outcomes = $\{1, 2, 3, 4, 5, 6\}$ in this case it is discrete...

$X \quad \omega \in \Omega \longrightarrow X(\omega) = x \in E$

⚠ In practice, often physicists identify X with x. Keep in mind that for mathematicians they are different.

Random variables can be discrete and continuous. If X is discrete (as in the case of the toss of a dice), then the probability is completely specify by $P(x_1), P(x_2), \dots$

Condition of normalization: $\sum_i P(x_i) = 1$

ex. The roll of a fair dice

$E = \{1, 2, 3, 4, 5, 6\}$

$P(x=1) = \frac{1}{6}$, $P(x=1,4) = \frac{2}{6} = \frac{1}{3}$.

I will always consider a real variable if not otherwise specified

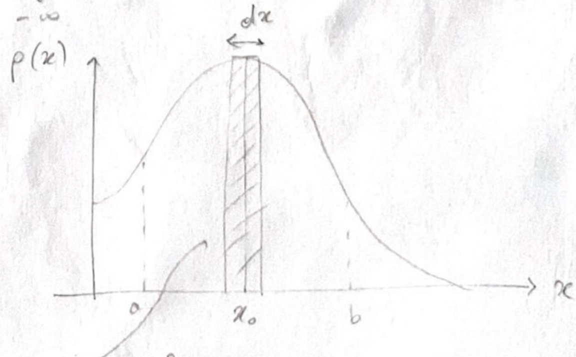
We now focus on continuous random variables. Let x be (the outcome of) a random variable so that $x \in \mathbb{R}$. We define the cumulative probability function $F(y)$ (CPF) as the probability that $x \le y$:

$$\boxed{F(y) = P(x \in [-\infty, y])}$$

When it exists, the probability density function (PDF) $p(x)$ is defined as:

$$p(x) = \frac{dF(x)}{dx} \quad \text{i.e.} \quad F(x) = \int_{-\infty}^{x} dx' \, p(x')$$

Let consider an infinitesimal interval dx:



$$P(x \in [x_0 - dx/2, x_0 + dx/2]) = \underline{p(x_0)\, dx}$$

it is the area of the rectangle

$$P(x \in [a, b]) = \int_a^b p(x)\, dx$$

Condition of normalization : $\int_{-\infty}^{+\infty} \rho(x)\, dx = P(x \in \mathbb{R}) = 1$

Remark : $\rho(x)\, dx$ is a probability $\Rightarrow [\rho(x)] = [x]^{-1}$

so the PDF has a unit !

- Expectation value : Consider an observable $O(x)$, the expectation value is :

value taken by the random variable

$$\langle O(x) \rangle = \int dx\, \rho(x)\, O(x)$$

random variable

average with respect to the relevant probability density

NB : if the variable is discrete

$$\langle O(x) \rangle = \sum_i O(x_i)\, P(x_i)$$

- Mean and variance

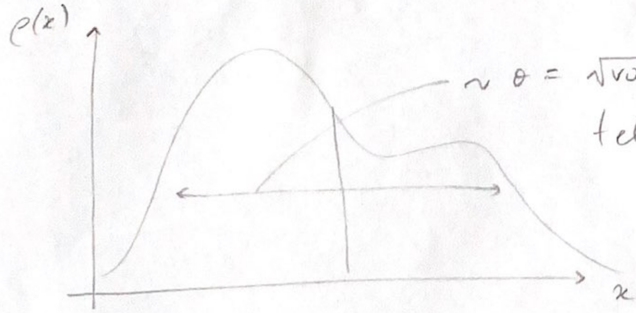
$$\mu = \langle x \rangle = \int dx\, \rho(x)\, x$$

for discrete variable $\langle x \rangle = \sum_i x_i\, p(x_i)$

try by yourself !

$$\text{var}(x) = \sigma^2 = \langle (x-\mu)^2 \rangle = \int dx\, (x-\mu)^2 \rho(x) = \langle x^2 \rangle - \mu^2$$

$\text{var}(\lambda x) = \lambda^2\, \text{var}(x)$

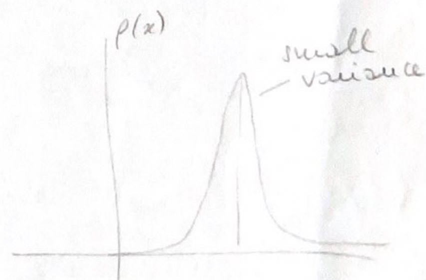


$\sim \sigma = \sqrt{\text{var}(x)}$: standard deviation

tells how wide the distribution is

(notice the dimensions !)

μ — where the center of the distribution is



- Notion of independence

Consider two random variables x, y. Then the joint probability density $p_{xy}(x,y)$ is such that

$$p_{xy}(x,y)\,dx\,dy = P\left(x \in [x, x+dx],\ y \in [y, y+dy]\right)$$

Normalization: $\int dx\,dy\, p_{xy}(x,y) = 1$

We say that the two random variables x, y are <u>independent</u> if

$$\boxed{p_{xy}(x,y) = p_x(x)\, p_y(y)} \quad \forall\, x, y$$

the joint probability density can be factored as the product of the singular probability densities.

ex. Rolling two (fair) dices

$$P_{xy}(x,y) = \frac{1}{36} = P_x(x)\, P_y(y) = \frac{1}{6} \times \frac{1}{6}$$

What does it mean that two variables are independent? It means that in no way the result of x affects the result of y.
If the two variables are not independent, they are said to be <u>correlated</u>.

<u>Theorem</u>:

Given $p_{xy}(x,y)$

(i) $\langle x + y \rangle = \langle x \rangle + \langle y \rangle$ always true

(ii) $\mathrm{var}(x+y) = \mathrm{var}(x) + \mathrm{var}(y)$ this is only true if the two variables are independent!

- <u>Change of variables</u> : Consider a one-to-one function $y = f(x)$

$$\underline{P}(y \in A) = \underline{P}(f(x) \in A) = \int_{f(x)\in A} dx\, p_X(x) = \int_A dy\, p_Y(y)$$

We make a change of variable in the first integral:

this is the definition of $p_Y(y)$
y is itself a random variable!

$y = f(x)$

$dy = f'(x)\,dx \Rightarrow$ Jacobian $\left|\frac{dx}{dy}\right| = \left|\frac{1}{f'(x)}\right|$

$$\Rightarrow \int_{y\in A} dy \left|\frac{dx}{dy}\right| p_X(f^{-1}(y)) = \int_{y \in A} dy\, p_Y(y)$$

$$\Rightarrow \boxed{p_Y(y) = p_X(f^{-1}(y)) \left|\frac{dx}{dy}\right|} \text{ or equivalently, } \boxed{p_Y(y)\,dy = p_X(x)\,dx}$$

Remark: unit measure!

NB If f is not 1-to-1, i.e. $\exists$ multiple x_i s.t. $y = f(x_i)$

$$\Rightarrow p_Y(y) = \sum_i p_X(x_i) \left|\frac{dx}{dy}\right|_{x=x_i}$$

<u>Remark</u>: since integrals are invariant under change of variable, expectation values are invariant.

$$\langle y \rangle_y = \langle f(x) \rangle_x$$

The <u>mode</u> is the most probable value $\neq$ the mean

$$x_{mode} = \arg\max_x p_X(x)$$

This is generally <u>not</u> invariant under a change of parametrization!

$$f(x_{mode}) \neq \arg\max_{y} p_Y(y)$$

[redacted]

[redacted]

- <u>Moments and cumulant</u>

- the moments of a random variable x are defined by :

$$\boxed{m_{x,n} = \langle x^n \rangle}$$

They fully define a probability density, i.e. if two RV have the same moments $\forall n$, then they have the same PDF.

Why ?

- <u>Characteristic function</u>.

the characteristic function is defined as the expectation value of :

$$\boxed{\phi_x(k) = \langle e^{-ikx} \rangle}$$

If x has a PDF, then the characteristic function is the Fourier transform of the PDF :

$$\phi_x(k) = \int dx \; \rho(x)\, e^{-ikx} \quad \text{and} \quad \rho(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} dk \; \phi_x(k)\, e^{+ikx}$$

Let's take the Taylor expansion of $\phi_x(k)$:

$$\phi_x(k) = \sum_{n=0}^{\infty} \frac{k^n}{n!} \left.\frac{\partial^n \phi_x(k)}{\partial k^n}\right|_{k=0} = \int dx \; \rho(x) \sum_{n=0}^{\infty} \frac{(-ik)^n x^n}{n!} =$$

$$= \sum_{n=0}^{\infty} \frac{(-i)^n k^n}{n!} \underbrace{\int dx \; \rho(x)\, x^n}_{\text{these are the moments!}} = \sum_{n=0}^{\infty} \frac{(-i)^n k^n}{n!} m_{x,n}$$

Thus, the moments of the random variable are the coefficient of the Taylor expansion of $\phi_x(k)$:

$$\boxed{m_{x,n} = \frac{1}{(-i)^n} \left.\frac{\partial^n \phi_x(k)}{\partial k^n}\right|_{k=0}}$$

this is why the characteristic function is important!
From it you can construct all the moments of the PDF.

Also, if $x_1, x_2, \dots, x_N$ are N independent random variables

$$\phi_{x_1+x_2+\dots+x_N}(k) = \phi_{x_1}(k) \cdot \phi_{x_2}(k) \cdot \phi_{x_3}(k) \cdots \phi_{x_N}(k) \cdot \;\; \forall k$$

the characteristic function of the sum of independent random variables is the product of their individual characteristic functions.

ex. two independent random variables x, y. $z = x + y$

$$\phi_z(k) = \langle e^{-ikz} \rangle = \langle e^{-ik(x+y)} \rangle = \langle e^{-ikx} \rangle \langle e^{-iky} \rangle = \phi_x(k)\, \phi_y(k)$$

What does it tell about the density? Multiplication in Fourier space is convolution in real space.

$$P_z(z) = \int \frac{dk}{2\pi}\, \phi_z(k)\, e^{+ikz} = \int \frac{dk}{2\pi}\, \phi_x(k)\, \phi_y(k)\, e^{+ikz} = \int \frac{dk}{2\pi} \int dx \int dy\, P_x(x) P_y(y)$$

$$e^{-ikx}\, e^{-iky}\, e^{+ikz} = \int \frac{dk}{2\pi} \int dx \int dy\, P_x(x)\, P_y(y)\, e^{ik(z-x-y)}$$

NB $\delta(x - x_0) = \frac{1}{2\pi} \int dk\, e^{ik(x-x_0)}$

this is a useful trick:

$$\rho(z_0) = \int dz\, \delta(z - z_0)\, \rho(z) = \langle \delta(z - z_0) \rangle_z$$

$$= \int dx\, dy\, P_x(x)\, P_y(y)\, \delta(z - x - y) = P_z(z)$$

$$\Rightarrow \boxed{P_z(z) = \int dx\, P_x(x)\, P_y(z - x)}$$

Remark: Sometimes instead of the Fourier transform it is used the Laplace transform:

$M_x(k) = \langle e^{kx} \rangle$ it is called moment-generating function.

- Cumulant generating function

The cumulant generating function is defined as the log of the characteristic function

$$\boxed{\Psi_x(k) = \log \phi_x(k) = \log \langle e^{-ikx} \rangle}$$

The expansion coefficients are named cumulants

$$\Psi_x(k) = \sum_{n=1}^{\infty} \frac{(-ik)^n}{n!} \langle x^n \rangle_c \quad \text{with} \quad \boxed{\langle x^n \rangle_c = \frac{1}{(-i)^n} \frac{d^n \log \phi_x(k)}{dk^n} \bigg|_{k=0}}$$

Note that the cumulant $\langle x^n \rangle_c$ is a function of the moments $m_{x,n'}$ with n' up to $n'=n$ plus covariances.

$\langle x \rangle_c = \langle x \rangle$ mean

$\langle x^2 \rangle_c = \langle x^2 \rangle - \langle x \rangle^2$ variance

$\langle x^3 \rangle_c = \langle x^3 \rangle - 3\langle x^2 \rangle \langle x \rangle + 2\langle x \rangle^3$ ~ skewness $\gamma = \frac{\mu^3}{\theta^3}$ it measures the asymmetry of the distribution

$\langle x^4 \rangle_c = \langle x^4 \rangle - 4\langle x^3 \rangle\langle x \rangle - 3\langle x^2 \rangle^2 + 12\langle x^2 \rangle\langle x^2 \rangle - 6\langle x \rangle^4$

Kurtosis

There is an elegant diagrammatic way to express the moments as a function of the cumulant (see chap. 2.2 of Kardar's book "Statistical physics of particles"). ✓

Some important distribution

I have introduced earlier the cumulants. They are particularly useful when a comparison between a probability density and the Gaussian distribution is needed. As we will see the Gaussian distribution is characterised by having all zero cumulants of order higher than 2.

• Gaussian or normal distribution $\mathcal{N}(\mu, \theta^2)$

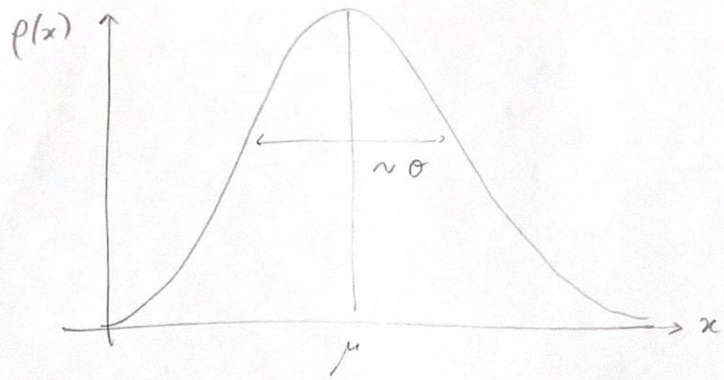
$x \in \mathbb{R}$

$$\boxed{p(x) = \frac{1}{\sqrt{2\pi\theta^2}} e^{-\frac{(x-\mu)^2}{2\theta^2}}}$$

check on the dimension

$[\theta^2] = [x^2]$

$[p(x)] = [x]^{-1}$



Checking normalization: $\displaystyle\int_{-\infty}^{+\infty} dx \, \frac{1}{\sqrt{2\pi\sigma^2}} \, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

change of variable: $y = \dfrac{x-\mu}{\sigma}$

$dy = \dfrac{1}{\sigma} dx$

$$= \underbrace{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} dy \, e^{-y^2/2}}_{\text{Gaussian integral : good to remember}} = \frac{\sqrt{2\pi}}{\sqrt{2\pi}} = 1$$

proof: $I = \displaystyle\int_{-\infty}^{+\infty} dy \, e^{-\alpha y^2} = \sqrt{\frac{\pi}{\alpha}}$

$$I^2 = \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dy \, e^{-\alpha(x^2+y^2)}$$

going to polar coordinates:

$r^2 = x^2 + y^2$

$dx\,dy = r\,dr\,d\theta$

$$= \int_0^{2\pi} d\theta \int_0^{+\infty} dr \, r \, e^{-\alpha r^2} = 2\pi \left[-\frac{e^{-\alpha r^2}}{2\alpha} \right]_0^{+\infty} = 2\pi \left[\frac{1}{2\alpha} \right] = \frac{\pi}{\alpha}$$

$$\Rightarrow \quad I = \sqrt{\frac{\pi}{\alpha}} \quad //$$

Characteristic function:

$$\phi_x(k) = \langle e^{-ikx} \rangle = \int dx \, \frac{1}{\sqrt{2\pi\sigma^2}} \, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, e^{-ikx}$$

① change of variable: $\bar{x} = x - \mu \qquad dx = d\bar{x}$

$$= \frac{e^{-ik\mu}}{\sqrt{2\pi\theta^2}} \int d\bar{x}\, e^{-\frac{\bar{x}^2}{2\theta^2} - ik\bar{x}} = \frac{e^{-ik\mu}}{\sqrt{2\pi\theta^2}} \int d\bar{x}\, e^{-\frac{a}{2}\bar{x}^2 + b\bar{x}}$$

$a = \frac{1}{\theta^2}$ $\quad b = -ik$

Again, a Gaussian integral useful to remember:

$$\int d\bar{x}\, e^{-\frac{a}{2}\bar{x}^2 + b\bar{x}} = \int d\bar{x}\, e^{-\left(\sqrt{\frac{a}{2}}\bar{x} - z\right)^2 + z^2} = \int d\bar{x}\, e^{-\left(\sqrt{\frac{a}{2}}\bar{x} - \frac{b}{\sqrt{2a}}\right)^2 + \frac{b^2}{2a}}$$

$$2\sqrt{\frac{a}{2}}\bar{x}z = b\bar{x}$$

$$\sqrt{2a}\,\bar{x}z = b\bar{x}$$

$$z = \frac{b}{\sqrt{2a}}$$

② $\bar{x}' = \sqrt{\frac{a}{2}}\bar{x} - \frac{b}{\sqrt{2a}}$

$d\bar{x}' = \sqrt{\frac{a}{2}}\, d\bar{x}$

$$= \frac{e^{-ik\mu}}{\sqrt{2\pi\theta^2}} e^{-\frac{k^2\theta^2}{2}} \sqrt{2\theta^2} \int_{-\infty}^{+\infty} d\bar{x}\, e^{-\bar{x}'^2} = \frac{e^{-ik\mu - \frac{k^2\theta^2}{2}}}{\sqrt{\pi}} \sqrt{\pi} =$$

$$= \boxed{e^{-ik\mu - \theta^2\frac{k^2}{2}} = \phi_x(k)}$$

The characteristic function of a Gaussian is a Gaussian!

To get the cumulant we take the log:

$$\psi_x(k) = \log \phi_x(k) = -ik\mu - \frac{\theta^2 k^2}{2}$$

We identify:

$\langle x \rangle_c = \mu$

$\langle x^2 \rangle_c = \theta^2$

$\langle x^3 \rangle_c = 0$

All cumulant higher than 2 are zero.

This makes it easier to compute the moments

$\langle x \rangle = \mu$

$\langle x^2 \rangle = \theta^2 + \mu^2$

$\langle x^3 \rangle = 3(\theta^2+\mu^2)\mu - 2\mu^3 = 3\theta^2\mu + \mu^3$

$$\begin{aligned}
\langle x^4 \rangle &= 4\langle x^3\rangle\langle x\rangle + 3\langle x^2\rangle^2 - 12\langle x^2\rangle\langle x\rangle^2 + 6\langle x\rangle^4 \\
&= 4(3\theta^2\mu+\mu^3)\mu + 3(\theta^2+\mu^2)^2 - 12(\theta^2+\mu^2)\mu^2 + 6\mu^4 \\
&= \cancel{12\theta^2\mu^2} + \underbrace{4\mu^4} + 3\theta^4 + 6\theta^2\mu^2 + \underbrace{3\mu^4} - \cancel{12\theta^2\mu^2} - \underbrace{12\mu^4} + \underbrace{6\mu^4} \\
&= \mu^4 + 6\theta^2\mu^2 + 3\theta^4 \quad //
\end{aligned}$$

All moments can be written in terms of μ and θ^2.

Notice that the sum of two or more gaussian random variables is also a Gaussian random variable.

How to prove? For instance, using the characteristic function:

$z = \sum_i x_i$

$$\phi_z(k) = \prod_i \phi_{x_i}(k) = \prod_i \left(e^{-ik\mu_i - \frac{\theta_i^2 k^2}{2}} \right) = e^{-ik \underbrace{\sum_i \mu_i}_{\mu_z} - \frac{k^2}{2} \underbrace{\sum_i \theta_i^2}_{\theta_z^2}}$$

$\Rightarrow$ z is a Gaussian variable with $\mu_z = \sum_i \mu_i$, $\theta_z^2 = \sum_i \theta_i^2$.

- <u>Binomial distribution</u>

Consider a biased coin that gives head with probability p and tail with probability $1-p$.

n = # of heads after N flips.

$$\boxed{P_N(n) = \underbrace{\binom{N}{n}}\ \underbrace{p^n (1-p)^{N-n}}}$$

tosses

probability of a single sequence (energetic factor)

purely combinatorial factor: how many sequences of tosses of length N contain n heads (energetic factor)

technically this is a "probability mass function"

Binomial formula:
$$(x+y)^N = \sum_{n=0}^{N} \binom{N}{n} x^n y^{N-n}$$

NB all the tosses are independent!

The characteristic function is:

$$\phi_n(k) = \langle e^{-ikn} \rangle = \sum_{n=0}^{\infty} \binom{N}{n} p^n (1-p)^{N-n} e^{-ikn} = \left(pe^{-ik} + 1 - p\right)^N$$

Why? n can be seen as the sum of N independent coin flips (Bernoulli variables)
To each flip corresponds a random variable y_i so that:

$$y_i \begin{cases} 1 & \text{with probability } p \text{ (if the coin lands on head)} \\ 0 & \text{———— } (1-p) \text{ (if the coin lands on tail)} \end{cases}$$

they are independent

$$\langle e^{-ikn} \rangle = \langle e^{-ik\sum_i y_i} \rangle = \langle \prod_i e^{-iky_i} \rangle = \prod_i \langle e^{-iky_i} \rangle = \langle e^{-iky_i} \rangle^N$$

$$= \left[e^{-ik} p + (1-p) \right]^N \quad ✓$$

The cumulant generating function is:

$$\psi_n(k) = N \underbrace{\log \left[pe^{-ik} + 1 - p \right]}_{\psi_y(k)}$$

So we compute the mean and variance:

$$\frac{d\psi}{dk} = \frac{N}{pe^{-ik}+1-p}(-i)e^{-ik}p$$

$$\frac{d^2\psi}{dk^2} = \frac{N(-i)(-i)pe^{-ik}(pe^{-ik}+1-p) - (-i)Ne^{-ik}p(-i)pe^{-ik}}{(pe^{-ik}+1-p)^2}$$

$$\langle n\rangle_c = \frac{1}{(-i)} N(-i)p = Np$$

$$\langle n^2\rangle_c = \frac{1}{\cancel{(-i)^2}} N\cancel{(-i)^2}p + \frac{1}{\cancel{(-i)^2}}\cancel{(-i)^2}p^2N = Np - p^2N = Np(1-p)$$

(Notice that both scale with N: $\frac{\sqrt{\langle n^2\rangle_c}}{\langle n\rangle} \xrightarrow[N\to\infty]{} 0 \Rightarrow$ the distribution of n becomes sharply peaked.)

Actually all cumulants scale with N...

Let's introduce a rescaled variable: $\boxed{\Delta n = \frac{n-\langle n\rangle}{\sqrt{\langle n^2\rangle_c}}}$ such that:

$$\langle \Delta n\rangle = 0$$

$$\langle \Delta n^2\rangle_c = \langle \Delta n^2\rangle - \cancel{\langle \Delta n\rangle^2} = \left\langle\left(\frac{n-\langle n\rangle}{\sqrt{\langle n^2\rangle_c}}\right)^2\right\rangle = \frac{\langle (n-\langle n\rangle)^2\rangle}{\langle n^2\rangle_c} = 1$$

So the rescaled variable has now the first and second cumulant independent of N. All the other cumulant go to zero as $N\to\infty$

ex.

$$\langle \Delta n^3\rangle_c = \langle \Delta n^3\rangle - 3\cancel{\langle \Delta n^2\rangle\langle \Delta n\rangle} + \cancel{2\langle \Delta n\rangle^3}$$

$$= \left\langle\left(\frac{n-\langle n\rangle}{\sqrt{\langle n^2\rangle_c}}\right)^3\right\rangle = \left\langle\frac{(n-\langle n\rangle)^3}{\langle n^2\rangle_c^{3/2}}\right\rangle = \frac{\langle (n-\langle n\rangle)^3\rangle}{N^{3/2}p(1-p)} \sim \frac{N}{N^{3/2}} \to 0$$

Remind of something?

The Gaussian distribution!

$$P(\Delta n = x) = P\left(n = x\sqrt{Np(1-p)} + Np\right)\frac{dn}{d\Delta n}$$

Jacobian! $= \sqrt{Np(1-p)}$

We have seen the formula of the change of variables

$$= \sqrt{Np(1-p)}\binom{N}{X} p^X (1-p)^{N-X}$$

with $X = x\sqrt{Np(1-p)} + Np$

$$= N^{1/2}\frac{N!}{(N-X)!\,X!}\,p^{X+1/2}(1-p)^{N-X+1/2}$$

Taking the logarithm:

$$\frac{1}{2}\log N + (X+1/2)\log p + (N-X+1/2)\log(1-p) + \log N! - \log(N-X)! - \log X!$$

Using Stirling approximation : $\log a! \simeq a\log a - a + \frac{1}{2}\log 2\pi a + O\left(\frac{1}{a}\right)$

Long computation...

$\Gamma(N+1) = N! = \int_0^\infty dx\, x^N e^{-x}$

$$= \frac{1}{2}\log\frac{1}{2\pi} - \frac{x^2}{2}$$

I exponentiate : $\boxed{\sqrt{\frac{1}{2\pi}}\,e^{-x^2/2} = \mathcal{N}(0,1)}$!!

This is an instance of the central limit theorem (very important in stat. mech.!) :

If $x_1, x_2, \dots, x_n$ are independent and identically distributed (iid) random variables, then

$p_i(x_i)$ is the same $\forall i$

$\langle x_i\rangle = \mu$

$\mathrm{var}(x_i) = \sigma^2$

$$\boxed{\frac{x_1 + x_2 + x_3 + \dots + x_N - n\mu}{\sqrt{n}\,\sigma} \xrightarrow[n \to \infty]{} \mathcal{N}(0,1)}$$

this is the mean of the sum $\langle x_1 + x_2 + \dots + x_n \rangle = n\mu$

this is the standard deviation of the sum : $\mathrm{var}(\sum_i x_i) = n\,\sigma^2$

CLT estimates the type of fluctuations expected in the sum of iid random variables in the limit of large number.

It tells that fluctuations around the expected value become Gaussian. Also called normal fluctuations.

An important consequence: for the sample average. $\bar{X} = \frac{1}{N}\sum_i x_i$ the CLT states that it approaches a normal distribution

$$\bar{X} \underset{N\to\infty}{\sim} \mathcal{N}\left(\mu, \frac{\sigma^2}{N}\right)$$

So the width of the distribution vanishes as $N \to \infty$

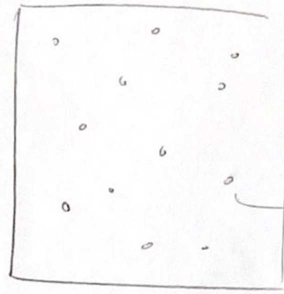
This is the <u>law of large number</u> which tells that the sample average becomes sharply peaked around the true mean value.

Note that there exists a more general CLT which applies to weakly-correlated random variables, i.e. random variables whose correlations decay with N "sufficiently" fast.

Why all of this is important in statistical physics?

ex. N particles in a box:

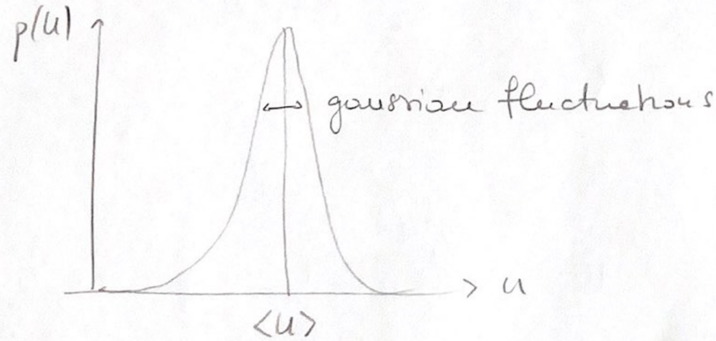
$$U = \sum_{i=1}^{N} \varepsilon_i$$



Problem: we don't know the ε_i's, we cannot measure the energy of every single particle in the system. We simply do not have access to this level of information.

Statistical mechanics: we reason in terms of probability densities (ensemble) $p(\varepsilon_i)$ and treat each ε_i as a random variable

$$U = \sum_{i=1}^{N} \varepsilon_i \xrightarrow[CLT]{} \mathcal{N}\left(N\langle \varepsilon_i \rangle, N\sigma_{\varepsilon_i}^2\right)$$



For very large N, deviations from $\langle U \rangle$ becomes negligible → contact with thermodynamics!

NB In general particles interact... $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N$ are not independent!

example: $\boxed{\langle U \rangle = N c_V T}$ for the ideal gas

- Briefly, the <u>Poisson distribution</u>

We have seen the binomial distribution

$$P_N(n) = \binom{N}{n} p^n (1-p)^{N-n}$$

$$\langle n \rangle = Np = \lambda \quad \text{fixed}$$

it is the average number of events.

$$\lambda = \frac{N}{p} = \text{parameter of the Poisson distribution}$$

$$\boxed{P(k) = \frac{\lambda^k e^{-\lambda}}{k!}}$$

in the limit $N \to \infty$, $p \to 0$ s.t. $\lambda = \text{const}$

describes the distribution of a system with a large number of possible events, each of which is rare.

$$\boxed{\lambda = \langle k \rangle = \text{var}(k)}$$

all cumulants have the same value!

ex. it is known the average rate of an event r. Then $\lambda = rt$

$$\boxed{P(k \text{ events occur in interval } t) = \sum_k \frac{(rt)^k}{k!} e^{-rt}}$$

- Markov chains
- Population dynamics
- Chemical reactions

- $N > 1$ random variables

Relevant physical observables may live in high-dimensional space

e.g. phase space $(\vec{x}_1, \dots, \vec{x}_N, \vec{v}_1, \dots, \vec{v}_N)$ in total $6N$ degrees of freedom

- The joint PDF $\rho(\vec{x})$ of N variables $x_1, \dots, x_N$ is such that

$$\rho(x_1^0, \dots, x_N^0) \underbrace{dx_1 dx_2 \dots dx_N}_{\substack{\text{volume element} \\ \prod_i dx_i = d^N x_i}} = P\left(x_i \in [x_i^0 + dx_i]\right)$$

Normalization: $\int \prod_i dx_i \; \rho(x_1, \dots, x_N) = 1$

- The marginal or unconditional PDF

$\rho(x_1, \dots, x_n) = \int dx_{n+1} \dots dx_N \; \rho(x_1, x_2, \dots, x_n, x_{n+1}, \dots, x_N)$ indicates the probability to obtain $x_1, \dots, x_n$ irrespectively of the other values.

Normalization: $\int dx_1 \dots dx_n \; \rho(x_1, \dots, x_n) = 1$

⚠ I am using the same letter ρ for simplicity...

- The conditional PDF $\rho(x_1, \dots, x_n | x_{n+1}, \dots x_N)$ = probability density to get $x_1, \dots, x_n$ given that the last $N-n$ RV's have value $x_{n+1}, \dots, x_N$.

Connection to the joint distribution:

$$\boxed{\rho(x_1, \dots, x_n | x_{n+1} \dots, x_N) = \frac{\rho(x_1, x_2, \dots, x_N)}{\rho(x_{n+1}, \dots, x_N)}}$$

joint

x_n marginal

As a consequence (simpler form): for event A, B

Bayes theorem: $P(A|B)\,P(B) = P(A,B) = P(B|A)\,P(A)$

- N-dimensional Gaussian distribution

ex. joint distribution of N Gaussian random variables

$(x_1, x_2, \dots x_N)$:

$$p(x_1, x_2, \dots, x_N) = \frac{1}{\sqrt{(2\pi)^N \det C}}\, e^{-\frac{1}{2}\sum_{i,j} C^{-1}_{ij}(x_i-\mu_i)(x_j-\mu_j)}$$

where C is the $(N \times N)$ covariance matrix, real and symmetric:

$$C_{ij} = \langle (x_i-\mu_i)(x_j-\mu_j)\rangle = \underbrace{\langle x_i x_j\rangle_c}_{}$$

it generalizes the notion of variance to multiple dimension.

it has the variances on its diagonal

C is diagonizable and $\det C$ is the product of its eigenvalues

Characteristic function: $\boxed{\varphi_{\vec{x}}(\vec{k}) = e^{-i\sum_j \mu_j k_j - \sum_{j,l} C_{jl} k_j k_l}}$

Cumulants: $\langle x_i\rangle_c = \mu_i$

$\langle x_i x_j\rangle_c = C_{ij}$

All higher cumulants vanish!

Moments: $\langle x_i\rangle = \mu_i$

$\langle x_i x_j\rangle = \langle x_i x_j\rangle_c + \langle x_i\rangle\langle x_j\rangle$

Wick theorem: When $\langle x_i \rangle = 0, \forall i$, then in the expansion of moments in terms of cumulants, only $\langle x_i x_j \rangle_c$ survives. Odd moments vanish while even moments simply correspond to all ways of pairing variables:

$$\langle x_i x_j x_l x_m \rangle = \langle x_i x_j \rangle \langle x_l x_m \rangle + \langle x_i x_l \rangle \langle x_j x_m \rangle + \\ + \langle x_i x_m \rangle \langle x_j x_l \rangle .$$