Edited 2025

Maxwell-Boltzmann Distribution

by Tony Shing

Overview:

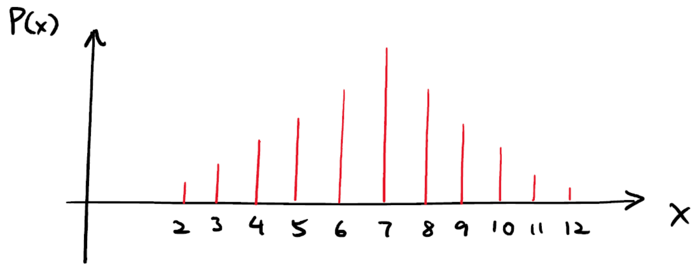
- Prerequisite: Probability, extending to continuous random variables
- Boltzmann factor
- Deriving Maxwell-Boltzmann distribution

1 Probability Distribution: Discrete & Continuous

1.1 Probabilty Mass Function (PMF)

For random variables that can be picken from a finite number cases, we describe them using the **probability mass function** (PMF), $P(x)$, or most of the time we just call them probabilities.

E.g. By throwing 2 dices and sum their values, we obtain a list of different probabilities for getting any integer between 2 to 12.



- **Normalization requriement**: The probability of all possible outcomes add up to 1.

$$\sum_{\text{all possible } x_i} P(x_i) = 1$$

- **Expected value**: i.e. The weighted average of all outcomes according to the probabilities, which is the average value you would get after repeating the process ∞ times.

$$\mathrm{E}[x] \stackrel{\text{def}}{=} \sum_{\text{all possible } x_i} x_i P(x_i)$$

- **Variance**: It is the average of (distance)^2 of all outcomes relative to the expected value. Taking square root gives the **standard derivation** (S.D.) $= \sqrt{\text{Variance}}$.

$$\mathrm{Var}[x] \stackrel{\text{def}}{=} \underbrace{\mathrm{E}}_{\text{Then take average}}[\underbrace{(x-\mathrm{E}[x])^2}_{\substack{(x-\mathrm{E}[x]) \,=\, \text{x's distance from the expected value} \\ \text{Square for taking magnitude only}}}]$$

Side note:

In a special case where there are N different outcomes and every outcome has equal probability to occur,

$$P(x_i) = \frac{1}{N}$$

Then the expected value, variance and S.D. will reduce to the mean and S.D. formula that we can find in high school textbook.

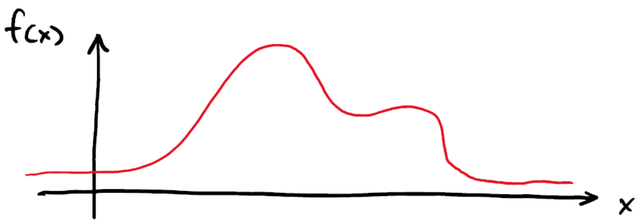
$$\mathrm{E}[x] = \sum_{\text{all possible } x_i} x_i P(x_i) = \frac{\sum_i x_i}{N} \stackrel{\text{def}}{=} \bar{x}$$

$$\mathrm{Var}[x] = \mathrm{E}\left[(x - \mathrm{E}[x])^2\right] = \frac{\sum_i (x_i - \mathrm{E}[x])^2}{N} = \frac{\sum_i (x_i - \bar{x})^2}{N}$$

$$\text{S.D.} = \sqrt{\mathrm{Var}[x]} = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{N}}$$

1.2 Probability Density Function (PDF)

For random variables that can apprear as any value within an interval, we describe them using the **probability density function** (PDF).



Pay attention here: the output value of the PDF, $f(x)$, is NOT the probability of getting x. This is because there are always infinitely many real number along any interval of real numbers. The probability of getting an exact number is limited to 0.

Instead, PDF can only describe the probability of getting a value within an interval $[x_0, x_0 + \mathrm{d}x]$.

$$\underbrace{P(x_0 < x < x_0 + \mathrm{d}x)}_{\text{Prob. inside the interval}} = \underbrace{f(x)}_{\text{height}}\underbrace{\mathrm{d}x}_{\text{width}}$$

As interpreted, the probability is represented by the area under curve $f(x)\,\mathrm{d}x$, not the height $f(x)$. This is also why it is named "density" function - the probability per "length" of x.

The formula for PMF (discrete cases) can be extended to PDF (continuous cases).

- **Normalization requriement**: The probability of all possible cases add up to 1. From the condition for PMF, we can extend the sum to integral for PDF.

$$\sum_{\text{all possible } x_i} P(x_i) = 1 \qquad \Rightarrow \qquad \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$$

- **Expected value**: The weighted average of all outcomes according to the probabilities. Can be understand by the weighted sum interpretation of integration.

$$\mathrm{E}[x] \stackrel{\text{def}}{=} \sum_{\text{all possible } x_i} x_i P(x_i) \qquad \Rightarrow \qquad \mathrm{E}[x] \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x$$

- **Variance**: Same as in the discrete case, it is the average of (distance)^2 of all outcomes relative to the expected value.

$$\mathrm{Var}[x] \stackrel{\text{def}}{=} \mathrm{E}\big[(x - \mathrm{E}[x])^2\big] \qquad \Rightarrow \qquad \mathrm{Var}[x] \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} (x - \mathrm{E}[x])^2 f(x)\,\mathrm{d}x$$

  Taking square root gives the standard derivation (S.D.) $= \sqrt{\mathrm{Var}[x]}$.

---

2 Boltzmann Distribution

2.1 Fluctuation of State Parameters

From previous notes, we have learnt that

- Different states of a thermodynamics system are distinguished by macrostate parameters.
- Change of macrostate is caused by random motion.
- Every *closed* system will eventually evolve into a state with maximum multiplicity, because the probability to observe a macrostate is proportional to its multiplicity.

In general, if we can express the entropy S of system in terms of a macrostate parameter x, We can compute the corresponding $x = x_0$ at the maximum entropy state by solving

$$\begin{cases} \left.\dfrac{\mathrm{d}S}{\mathrm{d}x}\right|_{x=x_0} = 0 & \left(\substack{\text{Entropy is at its maximum} \\ \Rightarrow \text{ Slope } = 0 \text{ at } x = x_0}\right) \\ \left.\dfrac{\mathrm{d}^2S}{\mathrm{d}x^2}\right|_{x=x_0} < 0 & \left(\substack{\text{Slope keeps decreasing} \\ \text{near the maximum}}\right) \end{cases}$$

However, because of random motion, even if the system has reached its maximum entropy state, its state parameters can still fluctuate. Every so often, the parameter x may deviate by some small amount Δx, causing the system entering a lower entropy state. But this change will revert very soon again due to random motion.

For example, in the 2-boxes system with 1000 balls, its macrostates can be distinguished by one state parameter - N_R = the number of balls in the right box. Because the balls are free to move (randomly) between the boxes, N_R must fluctuate, even when it has reached the maximum entropy state 500L-500R.

- One of the ball may temporarily move from the left box to the right box. The system's state change to 499L-501R.
- After some time, another ball moves from the right box to the left box. The system's state returns to 500L-500R.

(add figure here: moving ball in and out)

In fact, we can even determine the probability distribution of observing fluctuation Δx. Recall that the probability of observing a macrostate with parameter x is proportional to the number of microstates (i.e. multiplicity) it corresponds to,

$$P(x = x_0) \;\propto\; \left(\begin{array}{c}\text{\# of microstates that}\\ \text{leads to observing } x = x_0\end{array}\right)$$

$$= \frac{W(x = x_0)}{Z} = \frac{e^{\frac{S(x=x_0)}{k}}}{Z}$$

where Z is just some constant for normalization (because probability should add up to 1).

$$Z = \begin{cases} \displaystyle\sum_{\substack{\text{all possible}\\ x_i}} W(x_i) = \sum_{\substack{\text{all possible}\\ x_i}} e^{\frac{S(x_i)}{k}} & \text{if } x \text{ must take discrete values} \\ \displaystyle\int_{-\infty}^{\infty} W(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} e^{\frac{S(x)}{k}}\,\mathrm{d}x & \text{if } x \text{ can take continuous values} \end{cases}$$

Let $x = x_0$ when the system is at its maximum entropy state, the fluctuation in entropy ΔS can be related to the fluctuation of the state parameter Δx by Taylor series expansion:

$$S(x_0 + \Delta x) = S(x_0) + \Delta S$$

$$= \underbrace{S(x_0)}_{\text{A constant}} + \underbrace{\cancelto{0}{\left.\frac{\mathrm{d}S}{\mathrm{d}x}\right|_{x=x_0}}}_{\substack{S\,=\,\text{max.}\\ \text{so slope }=0}} (\Delta x) + \frac{1}{2}\underbrace{\left.\frac{\mathrm{d}^2 S}{\mathrm{d}x^2}\right|_{x=x_0}}_{\substack{S\,=\,\text{max.}\\ 2^{\text{nd}}\text{ derivative } <0}} (\Delta x)^2 + \cdots$$

The probability distribution of $x = x_0 + \Delta x$ is therefore approximately

$$P(x) = \frac{e^{\frac{S(x)}{k}}}{Z}$$

$$\approx \underbrace{\frac{1}{Z}\cdot e^{\frac{S(x_0)}{k}}}_{\substack{\text{Both are constant}\\ \text{can combine}}} \cdot e^{\frac{1}{2k}\frac{\mathrm{d}^2 S}{\mathrm{d}x^2}\big|_{x_0}(\Delta x)^2}$$

$$= \frac{1}{Z'}\cdot e^{\frac{1}{2k}\frac{\mathrm{d}^2 S}{\mathrm{d}x^2}\big|_{x_0}(x-x_0)^2}$$

This equation is in the form of **normal distribution** (also called Gaussian distribution), whose standard formula is written as

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

A normal distribution is governed by two variables, μ and σ, which are the mean and S.D. of the distribution respectively.

(add figure here: gaussian distribution)

By comparison, we can identify that the distribution of x follows a normal distribution whose

- Mean $= x_0$
  On average, fluctuate of x will be near $x = x_0$ which is the value of x when the system is in the maximum entropy state.
- S.D. $= -\sqrt{k\left(\frac{\mathrm{d}^2 S}{\mathrm{d}x^2}\Big|_{x_0}\right)^{-1}}$
  The magnitude of fluctuation depends on the relation of x with entropy S.

> *Side note:*
>
> Recall in Clausius theorem - entropy of a closed system must either be a constant or increase along any process. But because macrostate parameters are always fluctuating due to random process, entropy is also subject to fluctuation. Clausius theorem should be more accurately written as:
>
> $$\Delta S \geq 0 \xRightarrow{\text{More accurately}} \Delta S_{\text{avg.}} \geq 0$$

2.2 Boltzmann Factor

Consider a closed system, i.e. the total energy conserves within the system.
The fluctuation for internal energy is a bit more interesting - because energy of a closed system must be a constant, there shall not be fluctuation for the total internal energy. However, we can consider the fluctuation of internal energy on individual objects in the system.
A direct consequence of entropy fluctuation is that objects in a closed system are allowed to exchange energy (as long as total energy conserves).
Now we can apply the fluctuation However one thing very different between energy and other state parameters is that - energy must conserve in any scenario.
In the simplest 2-boxes model, without energy consideration, we used $x = n_L =$ number of balls in the left box as the state parameter.
when two such model is put together, total multiplicity is just a
When x=nL, nL can take whatever value,
energy is a conserved quantity if one object has high energy then the other must have low energy
Multiplicity under Energy Constraint
Recall that in the 2-boxes system with energy difference, the multiplicity of the system is a function to internal energy. If the system is closed, i.e. its internal energy is constant because it is not in thermal contact with anything, then neither its multiplicity or entropy can vary.

(add figure here: bar chart S)

But in reality, the only closed system is the whole universe because 100% heat insulation from heat radiation does not exist. Every object in the universe is constantly under thermal contact with other objects - sometimes you may observe its internal energy to be some value U_1, and sometimes you may observe it to be another value U_2.

What is the probability to observe that the object carry internal energy U_i?

Previously, we have only dealt with closed system - any objects that involve in heat exchange are considered as part of the system.

2.3 The Boltzmann Factor

We model the object as thermally contacting with a HUGE heat bath (i.e. the universe), where the only form of energy exchange is by heat transfer.

(add figure here: heat bath contact)

We require the following assumptions:

1. The heat bath is so HUGE that any energy exchange has no effect on it. Its state parameters (e.g. temperature, entropy,...) remains constant all the time.
2. The total energy of the object + heat bath is conserved.
$$U_{\text{total}} = U_{\text{bath}} + U_{\text{obj}} = \text{const.}$$
3. Energy carried by the heat bath is much greater than the energy carried by the object. i.e.
$$U_{\text{bath}} \gg U_{\text{obj}}$$

Let the entropy of the heat bath be S_{bath}, which must be some function to its internal energy U_{bath} (although we do not know what this function is). We can make a Taylor expansion by

$$\begin{aligned} S_{\text{bath}} &= S_{\text{bath}}(U_{\text{bath}}) \\ &= S_{\text{bath}}(U_{\text{total}} - U_{\text{obj}}) \\ &\approx S_{\text{bath}}(U_{\text{total}}) - \left.\frac{\mathrm{d}S_{\text{bath}}}{\mathrm{d}U}\right|_{U=U_{\text{total}}} \cdot U_{\text{obj}} \\ &= S_{\text{bath}}(U_{\text{total}}) - \frac{U_{\text{obj}}}{T_{\text{bath}}} \end{aligned}$$

The total multiplicity of the {heat bath + object} is thus

$$\begin{aligned} S_{\text{total}} &= S_{\text{bath}} + S_{\text{obj}} \\ W_{\text{total}} &= e^{\frac{S_{\text{bath}}+S_{\text{obj}}}{k}} \\ &\approx e^{\frac{S_{\text{bath}}(U_{\text{total}})-\frac{U_{\text{obj}}}{T_{\text{bath}}}}{k}} \cdot e^{\frac{S_{\text{obj}}}{k}} \\ &= e^{\frac{S_{\text{bath}}(U_{\text{total}})}{k}} \cdot e^{-\frac{U_{\text{obj}}}{kT_{\text{bath}}}} \cdot e^{\frac{S_{\text{obj}}}{k}} \\ &= \underline{e^{\frac{S_{\text{bath}}(U_{\text{total}})}{k}}} \cdot e^{-\frac{U_{\text{obj}}}{kT_{\text{bath}}}} \cdot \underline{W_{\text{obj}}} \end{aligned}$$

Now the total multiplicity is expressed purely a function of U_{obj}. This is the total multiplicity (i.e. number of microstates) when the object has a particular value of internal energy U_{obj} with the heat bath maintained at temperature T_{bath}:

$$W_{\text{total}}(U_{\text{obj}}) = (\text{const.}) \cdot e^{-\frac{U_{\text{obj}}}{kT_{\text{bath}}}} \cdot W_{\text{obj}}(U_{\text{obj}})$$

To find the probability which the object has a particular value of internal energy U_{obj}, we simply divide it by the total multiplicity (total number of microstates), which is summed from all possible U_{obj}.

$$P(U_{\text{obj}}) = \frac{W_{\text{total}}(U_{\text{obj}})}{\displaystyle\sum_{\substack{\text{all possible} \\ U_{\text{obj}}}} W_{\text{total}}(U_{\text{obj}})}$$

But the thing we are actually interested in is the probability to find the object having centain value of internal energy. Recall the relation between multiplicity and probability:

$$W(U) =$$

Let the entropy-energy relation of the heat bath and the object be some functions (which we may or may not know its formula),

$$, \qquad S_{\text{obj}} = S_{\text{obj}}(U_{\text{obj}})$$

2.4 Example: System with Discrete Energy States

---

3 Density of State of Ideal Gas

The most important application of the above ideas is to find the energy-probability relation for states in ideal gas, when heat exchange is allowed.

i.e. States with which energy are most probable to be found, given that it is at temperature T?

Since energy of ideal gas can be any real number, we first need to adapt the formula from last section for continuous distribution.

$$\sum_{\substack{\text{all possible} \\ U}} e^{-\frac{U}{kT}} W(U) \qquad \rightarrow \qquad \int_0^\infty e^{-\frac{U}{kT}} W(U)\, \mathrm{d}U$$

W
What is the $W(U)$ for ideal gas?

U_{total} is a constant
so this term is just a constant

Multiplicity of the object
is some function of U_{obj}

3.1 Entropy of Ideal Gas

3.2 Average Speed of Ideal Gas

In textbooks you can find 2 different formulas related to the average speed about ideal gas. How are they derived?

1. **Average Speed** = Expected value to $|v|$, i.e. $\mathrm{E}(|v|)$.

$$\begin{aligned}
\mathrm{E}(|v|) &= \int_0^\infty |v| \cdot P(|v|)\,\mathrm{d}|v| \\
&= \int_0^\infty |v| \cdot \sqrt{\frac{2m^3}{\pi k^2 T^3}} |v|^2 e^{-\frac{m|v|^2}{2kT}}\,\mathrm{d}|v| \\
&= \sqrt{\frac{2m^3}{\pi k^2 T^3}} \left(\frac{2kT}{m}\right)^2 \int_0^\infty \left(\sqrt{\frac{m}{2kT}}v\right)^3 e^{-\left(\sqrt{\frac{m}{2kT}}v\right)^2}\,\mathrm{d}\left(\sqrt{\frac{m}{2kT}}v\right) \\
&= \sqrt{\frac{32kT}{\pi m}} \int_0^\infty x^3 e^{-x^2}\,\mathrm{d}x \\
&= \sqrt{\frac{32kT}{\pi m}} \cdot \frac{1}{2} \\
&= \sqrt{\frac{8kT}{\pi m}} \stackrel{\text{def}}{=} v_{\text{avg}}
\end{aligned}$$

2. **Root Mean Square Speed** = $\sqrt{\text{Expected value to } |v|^2}$, i.e. $\sqrt{\mathrm{E}(|v|^2)}$.

$$\begin{aligned}
\mathrm{E}(|v|^2) &= \int_0^\infty |v|^2 \cdot P(|v|)\,\mathrm{d}|v| \\
&= \int_0^\infty |v|^2 \cdot \sqrt{\frac{2m^3}{\pi k^2 T^3}} |v|^2 e^{-\frac{m|v|^2}{2kT}}\,\mathrm{d}|v| \\
&= \sqrt{\frac{2m^3}{\pi k^2 T^3}} \left(\frac{2kT}{m}\right)^3 \int_0^\infty \left(\sqrt{\frac{m}{2kT}}v\right)^3 e^{-\left(\sqrt{\frac{m}{2kT}}v\right)^2}\,\mathrm{d}\left(\sqrt{\frac{m}{2kT}}v\right) \\
&= \sqrt{\frac{64k^2T^2}{\pi m^2}} \int_0^\infty x^4 e^{-x^2}\,\mathrm{d}x \\
&= \sqrt{\frac{64k^2T^2}{\pi m^2}} \cdot \frac{3\sqrt{\pi}}{8} \\
&= \frac{3kT}{m} \\
\sqrt{\mathrm{E}(|v|^2)} &= \sqrt{\frac{3kT}{m}} \stackrel{\text{def}}{=} v_{\text{rms}}
\end{aligned}$$

Note: Because the average KE (which is something we can measure) is calculated as

$$\begin{aligned}
\text{Avg. KE} &= \mathrm{E}\left(\frac{1}{2}m|v|^2\right) \\
&= \frac{1}{2}m\mathrm{E}(|v|^2) \\
&= \frac{1}{2}mv_{\text{rms}}^2
\end{aligned}$$

The root mean square speed is way more useful than the average speed.

— The End —