Notes on Statistical Mechanics and Thermodynamics relevant to Machine Learning

compiled by D. Gueorguiev, 1/7/2025

# Starting Note

This document does not aim to discuss in depth the complete set of topics covered in the classic Statistical Mechanics books. Rather the interest here is in statistical models and underlying physical mechanisms relevant to certain problems in the Machine Learning field in order to establish a clear link to and to provide useful insights to Machine Learning algorithms of interest.

# Introductory Material

## Classical Ideal Gas

The abstraction ‘Ideal Gas’ and the mathematical model behind it have many applications outside the Physical Sciences. Thus we start our journey into the world of Statistical Mechanics and Thermodynamics with the definition of Ideal Gas.

An ideal gas is different from real gases by the absence of interactions between particles which constitute the gas.

While idea gas is unrealistic model for real gasses it can be thought as a limit scenario for real gasses with sufficiently low densities. The simple model of ideal gas does not capture phase transitions present in real gasses.

## Phase Space of a Classical Gas

Let us consider particles contained in some specified volume. Each particle has a well-defined position and momentum. The position of every particle is represented as a point in some abstract space (*configuration space*) with an axis for every coordinate of every particle. The generalized coordinates of this abstract space are given as

(1)

The momentum of every particle can be represented as a point in momentum space – an abstract -dimensional space (*momentum space*), with axes for every component of the momentum of every particle.

(2)

The complete microscopic state of the system can be described by a point in *phase space* – an abstract -dimensional space with axes for every coordinate and every momentum component for all particles. Phase space is the union of configuration space and momentum space, .

(3)

The *kinetic energy* of the -th particle is given by the usual expression . The *total kinetic energy* of the system is the sum of the kinetic energies of all particles in the system.

(4)

Since, by definition, there are no interactions between the particles in an ideal gas, the potential energy of the system is zero.

(5)

**Distinguishability**:

Particles in the system are considered *distinguishable* so that the exchange of two particles results in a different microscopic sate. Thus every point in phase space represents a different microscopic state.

**Thermodynamic Entropy:**

Let us denote with the thermodynamic entropy of a system – this quantity is associated with particular microstate defined by thermodynamic parameters such as temperature, volume, and energy. is related to the number of microstates which can result in a given microstate with the following expression:

( Boltzmann entropy law )

**Independence of Positions and Momenta:**

The underlying assumptions of the ideal gas model is that the positions and the momenta of the particles are independent.

**Separation of Entropy into Two Parts**

Since the positions and momenta are independent, we can express their joint probability as a product of two functions - one of positions only and the other of momenta only.

(6)

According to Boltzmann’s entropy definition and (6) the total entropy will be expressed as a sum of the contributions of the positions and the momenta. The probability distribution in configuration space, , depends only on the volume and the number of particles, . Consequently, the configurational entropy, depends only on and ; that is, .

The probability distribution in momentum space, , depends only on the total energy, , and the number of particles, . Thus, we write:

(7)

The thermodynamic quantities , , and are referred to as *extensive* parameters (aka *observables*, *variables*) because they measure the amount or extent of something. They are to be contrasted with *intensive* parameters, such as temperature or pressure, which do not automatically become bigger for bigger systems.

**Distribution of Particles between Two Subsystems**

Let us consider a composite system consisting of two boxes (subsystems) containing a total of distinguishable, non-interacting particles. We will name the boxes and , with volumes and . The total volume is . The number of particles in is , with being the number of particles in .

We can either constrain the number of particles in each box to be fixed or allow the numbers to fluctuate by making a hole in the wall separating the boxes. The total number of particles is constant in either case.

We are interested in the probability distribution for the number of particles in each subsystem after the constraint implemented as the impenetrable wall in the middle is removed by making a hole in the middle.

For simplicity we assume that the positions of the particles are mutually independent of each other besides being independent on momentum. The probability distribution of the configuration space can be written as:

(8)

If we further assume that a given particle is equally likely to be anywhere in the composite system, the probability of it being in subsystem is .

Figure: Box and Box with volumes and and number of particles and accordingly

If there are particles that are free to go back and forth between the two subsystems, the probability distribution of is given by the binomial distribution

(9)

with the constraint that .

The average value of and from the binomial distribution is

, (10)

Thus

(11)

The width of the probability distribution for is given by the standard deviation.

(12)

Stirling’s approximation: (13)

Gosper’s form of the Stirling approximation:

(14)

(15)

We substitute the simplified expression of (14) in (15) :

(16)

Let us compute the maximum of allowing to be continuous variable:

(17)

(18)

which becomes

(19)

Thus we get

(20)

Similarly one can write:

(21)

If we compare (20) and (21) to (11) we will notice

and

Thus we conclude the value of corresponding to the maximum of the probability distribution using the Gosper’s form of the Stirling approximation is equal to the mean value .

# References

[1] [An Introduction to Statistical Mechanics and Thermodynamics, Robert H. Swendsen, 2012](https://github.com/dimitarpg13/information_theory_and_statistical_mechanics/blob/main/literature/books/An-Introduction-to-Statistical-Mechanics-and-Thermodynamics-Swendsen-2012.pdf)

[2] [Statistical Physics by L.D. Landau and E.M. Lifshitz, 2nd Edition, 1970 (orig. 1958)](https://github.com/dimitarpg13/information_theory_and_statistical_mechanics/blob/main/literature/books/LandauLifshitz-StatisticalPhysics_1958.pdf)