THE BOLTZMANN DISTRIBUTION

ZHENGQU WAN

Abstract. This paper introduces some of the basic concepts in statistical mechanics. It focuses how energy is distributed to different states of a physical system, i.e. under certain hypothesis, it obeys the Boltzmann distribution. I will demonstrate three ways that the Boltzmann distribution will arise.

Contents

1. The Motivating Problem

Consider the following game. Suppose there are N people and everyone has some cash. For simplicity, each person's cash is a natural number. Now imagine the following process. At each second we randomly uniformly pick a person A. If A has positive amount of cash, then we randomly uniformly pick a person B(it could be the same person) and transfer a dollar from A to B. Keep this game running for a sufficiently long period of time, will the distribution of wealth among this group of people reach an equilibrium. If so, what will the equilibrium distribution be?

The answer to the above question is affirmative. Let X_i denote the wealth of the i-th person. Let u denote the average wealth so that Nu is a positive integer. The game can be modeled by a Markov chain where the state space is

$$S = \left\{ (X_1, ...X_N) : \ X_i \in \mathbb{Z}_{\geqslant 0}, \ \sum_{i=1}^{n} X_i = nu \right\}$$

Theorem 1.1. *The Markov chain is transitive and periodic.*

Proof. Starting at any initial state $X = (X_1, ..., X_N)$, it is possible to transfer all the cash to the first person and thus reach state $(Nu, 0, ..., 0)$ and vice versa. Thus

Date: 2017, July, 13.

the Markov chain is transitive. Since it has positive probability for the state X to remain unchanged, the Markov chain is periodic. $\square$

Theorem 1.2. *The transition probability from any state to any of its neighboring states is $\frac{1}{N^2}$. Thus the stationary distribution of this Markov chain is the uniform distribution π on S.*

Proof. For each state $X = (X_1, ..., X_N)$, to transition to its neighboring state $(X_1, ..., X_i - 1, ..., X_j + 1, ..., X_N)$ (assumming $X_i > 0$), we have to pick i-th person first and then pick j-th person. Since there are N people in total, the probability is $\frac{1}{N^2}$. For any pair neighboring states X and Y, since we have equal transition probability from X to Y and from Y to X, the uniform distribution on S is stationary. $\square$

Since the Markov chain is irreducible and aperiodic, any initial probability distribution of the Markov chain will converge to the uniform distribution π. Therefore, after the game runs for a sufficiently long period of time, we are equally likely to observe any state in S.

To describe the the distribution of wealth, we just have to know how many (in term of expected value) people have cash w for each w. Notice

$$E[\#\text{People with wealth } w] = E[\sum_{i=1}^{N} \mathbf{1}\{X_i = w\}] = \sum_{i=1}^{N} Pr[X_i = w]$$

So we just have to know the distribution of each X_i. We can explicitly compute

$$\begin{aligned} & Pr[X_i = w] \\ = & \frac{\#\text{Ways to distribute } (Nu - w) \text{ wealth to } (N-1)\text{people}}{\#S} \\ = & \binom{Nu - w + N - 2}{N-2} \Big/ \binom{Nu + N - 1}{N-1} \\ = & \frac{(Nu - w + N - 2)!}{(N-2)!(Nu - w)!} * \frac{(N-1)!(Nu)!}{(Nu + N - 1)!} \end{aligned}$$

We can use Stirling's approximation to obtain the asympototic behavior of the above expression, but then we would trap ourselves in massive computations. Instead, we will study the continuous analogue of this problem in the next section, which is easier to deal with.

If the amount of wealth being transferred each time is some small number 2^{-m} instead of 1 dollar, we will obtain the uniform distribution on the finer grid on the simplex.

$$S = \left\{ (X_1, ...X_N) : \; X_i \in 2^{-m}\mathbb{Z}_{\geqslant 0}, \; \sum_{i=1}^{n} X_i = Nu \right\}$$

When m gets larger, the equilibrium distribution converges (weak* convergence) to the uniform probability distribution on the $(N-1)$-dimensional simplex

$$S = \left\{ (X_1, ..., X_n) : \; X_i \in \mathbb{R}_{\geqslant 0}, \; \sum_{i=1}^{N} X_i = Nu \right\}$$

To obtain the probability distribution of individual X_i, we observe that, for $t \in [0, Nu]$, the region $\{X_i \geqslant t\}$ is still a simplex but with size $(1 - \frac{t}{Nu})$ times as large as S. Therefore

$$Pr[X_i \geqslant t] = \left(1 - \frac{t}{Nu}\right)^{N-1}$$

Thus the density function is thus

$$\rho_N(t) = \frac{N-1}{Nu}\left(1 - \frac{t}{Nu}\right)^{N-2}, \ t \in [0, Nu]$$

Send $n \to \infty$, we obtain

$$\rho_\infty = \frac{1}{u}\exp(-\frac{t}{u}), \ t \in [0, +\infty)$$

This is the famous **Boltzmann Distribution** in statistical mechanics, which tells us that it is less likely to find the system in higher energy states, **with the probability being inverse proportional to the exponential of the energy**. In this example, the wealth is the analogue of energy in physics and we see that the number of people with certain wealth decreases exponentially with the wealth.

Another important feature is that these X_i are **Asymptotically Independent**, i.e.the joint distribution of the first k variables $(X_1, ..., X_k)$ tends to be independent when $n \to \infty$. Such feature is analogous to the fact in real world that, position and momentum of different gas molecules can be regard as independent random variables, in spite of their collisions and interactions.

To prove the asymptotical independence, we first observe that, for $t_j \geqslant 0$ and $\sum_j t_j \leqslant Nu$, the region $\{X_1 \geqslant t_1, ..., X_k \geqslant t_k\}$ is still a simplex but with size $(1 - \frac{t_1 + ... t_k}{Nu})$ times as large as S. Therefore

$$Pr[X_1 \geqslant t_1, ..., X_k \geqslant t_k] = \left(1 - \frac{t_1 + ... + t_k}{Nu}\right)^{N-k-1}$$

So we obtain the joint density

$$\rho_N(t_1, ..., t_k) = (\frac{1}{Nu})^k \frac{(N-k-1)!}{(N-2k-1)!}\left(1 - \frac{t_1 + ... + t_k}{Nu}\right)^{N-2k-1}$$

Send $N \to \infty$, we obtain

$$\rho_\infty(t_1, ..., t_k) = \frac{1}{u^k}\exp(-\frac{t_1 + ... + t_k}{u})$$

Note

$$\frac{1}{u^k}\exp(-\frac{t_1 + ... + t_k}{u}) = \prod_{i=1}^{k}\frac{1}{u}\exp(-\frac{t_i}{u})$$

Thus we conclude as $N \to \infty$, the individual wealth $X_1, ..., X_k$ tends to be independent, while each of them obeys the Boltzmann distribution.

2. Boltzmann Distribution Arises from the Principle of Indifference

In this section, we will see how Boltzmann distribution arises from the principle of indifference. We will use the ideal gas model for demonstration.

In the ideal gas model, N particles move and bounce inside a d-dimensional box Ω. The particles themselves do not collide. The state of the ideal gas is completely specified by the positions $X_1, ..., X_N$ and momentums $P_1, ..., P_N$ of all the particles. Once the initial state of the ideal gas is known, the system evolves according to Newtonian mechanics.

Just like in the previous game where we are interested in the distribution of wealth among the people, here we are interested in the distribution of kinetic energy among the gas particles. Although the evolution of ideal gas is deterministic, we still treat X_i and P_i as random variables, because the number of gas particles is huge in reality ($\sim 10^{23}$) and it is impossible to measure the state of the system or to solve the evolution.

Since the position $X_i \in \Omega$ and the momentum $P_i \in \mathbb{R}^d$, the state of the ideal gas belongs to the phase space

$$(\Omega \times \mathbb{R}^d)^N$$

At temperature T, each particle particle has average kinetic energy $\frac{dk_BT}{2}$ (assume the particle is mono-atomic, k_B is the Boltzmann constant) Therefore the total kinetic energy satisfies

$$\sum_{i=1}^{N} \frac{P_i^2}{2m} = \frac{Ndk_BT}{2}$$

Where m is the mass of each gas particle. Therefore at temperature T, the state of the ideal gas belongs to the constant energy hypersurface

$$S = \left\{ (X, P) : \sum_{i=1}^{N} \frac{P_i^2}{2m} = \frac{Ndk_BT}{2} \right\} = \Omega^N \times L$$

Where

$$L = \left\{ P : \sum_{i=1}^{N} \frac{P_i^2}{2m} = \frac{Ndk_BT}{2} \right\}$$

It turns out, by the **principle of indifference** in statistical mechanics, each state on the energy hypersurface is equally likely to be observed as any other state. Therefore (X, P) is the uniform random variable on S. We have the following meaningful consequences:

(1) Each X_i is a uniform random variable in Ω.
(2) The positions of different particles are independent, i.e. $\{X_1, ..., X_N\}$ is independent.
(3) The positions $(X_1, ..., X_n)$ are independent to the momentums $(P_1, ..., P_n)$.
(4) Each P_i obeys the **Boltzmann-Maxwell Distribution** as $N \to \infty$.
(5) The momentums of different particles are **Asymptotically Independent**, i.e. the joint distribution of each $\{P_1, ..., P_k\}$ tends to be independent as $N \to \infty$.

(1), (2) and (3) are not too difficult to obtain. I will prove (4) and (5) as they are analogous to the properties emphasized in the firse section: wealth of individual person tends to obey the Boltzmann distribution and tends to be independent.

Theorem 2.1 (The 4th statement)**.** *The momentum of each particle obeys the* ***Boltzmann-Maxwell distribution*** *as $N \to \infty$, that is, the probability measure on $\mathbb{R}^d$ is given by*

$$dPr = \frac{1}{(2\pi m k_B T)^{\frac{d}{2}}} \exp\left(-\frac{P^2/2m}{k_B T}\right) dP$$

Where dP is the euclidean measure on the momentum space $\mathbb{R}^d$.

Remark 2.2. We observe $P^2/2m$ inside the exponential is the kinetic energy of the particle. The theorem tells us that the particle is less likely the to be observed at higher kinetic energy states. The probability decreases exponentially with the kinetic energy of the particle.

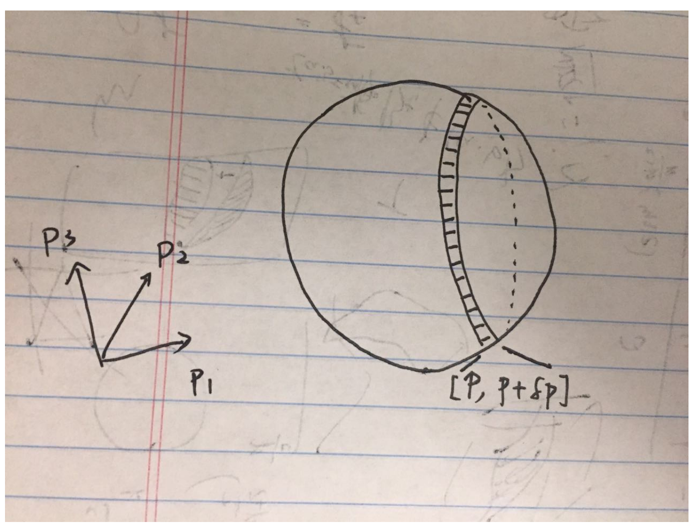
Proof. We first assume $d = 1$. The momentum of different particles is a uniform random variable on the $(N-1)$-dimensional sphere

$$L = \left\{ P : \sum_{i=1}^{N} \frac{P_i^2}{2m} = \frac{N d k_B T}{2} \right\}$$

To obtain the probability distribution of the momentum of the first particle, it suffices to compute

$$Pr[p \leqslant P_1 \leqslant p + \delta p]$$

Notice the region $\{p \leqslant P_1 \leqslant p + \delta p\}$ is nothing but a thin belt on the sphere L. It is topologically a $(N-2)$-dimensional sphere of radius $(N m k_B T - p^2)^{\frac{1}{2}}$ cross a width approximately δp (see the following illustration)



Thus the volume of the region is the $(N-2)$-dimensional circumference of the belt times the width. By geometry of sphere, these two equantities are

$$\text{circumference} = vol(S^{N-2})(N m k_B T - p^2)^{\frac{N-2}{2}}$$

$$\text{width} = \sqrt{\frac{Nmk_BT}{Nmk_BT - p^2}}\delta p \sim \delta p$$

Therefore

$$\begin{aligned}
& Pr[p \leqslant P_i \leqslant p + \delta p] \\
= & \frac{\text{circumference} * \text{width}}{\text{volume of } L} \\
\sim & \frac{vol(S^{N-2})(Nmk_BT - p^2)^{\frac{N-2}{2}}}{vol(S^{N-1})(Nmk_BT)^{\frac{N-1}{2}}}\delta p \\
\sim & \frac{2\pi^{\frac{N-1}{2}}/\Gamma(\frac{N-1}{2})}{2\pi^{\frac{N}{2}}/\Gamma(\frac{N}{2})} \frac{1}{(Nmk_BT)^{\frac{1}{2}}}\left(1 - \frac{p^2}{Nmk_BT}\right)^{\frac{N}{2}}\delta p \\
\sim & \frac{1}{(2\pi mk_BT)^{\frac{1}{2}}}\exp\left(-\frac{p^2/2m}{k_BT}\right)\delta p
\end{aligned}$$

Thus we have obtained the Boltzmann-Maxwell distribution. □

Theorem 2.3 (The 5th statement)**.** *The momentums of different particles are* ***Asymptotically Independent****, i.e. the joint distribution of each* $\{P_1, ..., P_k\}$ *tends to be independent as* $N \to \infty$*.*

Proof. Again we first assume $d = 1$. Now we consider the momentum of first k particles and want to know the probablity

$$Pr[p_i \leqslant P_i \leqslant p_i + \delta p_i, i = 1, 2, ..., k]$$

The region $\{p_i \leqslant P_i \leqslant p_i + \delta p_i, i \in \{1, ..., k\}\}$ is a belt on sphere L. It is topologically a $(N-k-1)$-dimensional sphere of radius $(Nmk_BT - \sum_{i=1}^k p_i^2)^{\frac{1}{2}}$ cross a cube of volume approximately $\delta p_1 ... \delta p_k$. Thus the volume of the region is the $(N-k-1)$-dimensional circumference of the belt times the volume of the cube. By geometry of sphere

$$\text{circumference} = vol(S^{N-k-1})(Nmk_BT - \sum_{i=1}^k p_i^2)^{\frac{N-k-1}{2}}$$

$$\text{volume of the cube} \sim \delta p_1 ... \delta p_k$$

Therefore

$$
\begin{aligned}
& Pr[p_i \leqslant P_i \leqslant p_i + \delta p_i, i = 1, 2, ..., k] \\
= & \frac{\text{circumference} * \text{volume of the cube}}{\text{volume of } L} \\
\sim & \frac{vol(S^{N-k-1})(Nmk_BT - \sum_{i=1}^{k} p_i^2)^{\frac{N-k-1}{2}}}{vol(S^{N-1})(Nmk_BT)^{\frac{N-1}{2}}} \delta p_1 ... \delta p_k \\
& \text{(after routine computation)} \\
\sim & \frac{1}{(2\pi m k_B T)^{\frac{k}{2}}} \exp\left(- \frac{\sum_{i=1}^{k} p_i^2/2m}{k_B T} \right) \delta p_1 ... \delta p_k \\
= & \prod_{i=1}^{k} Pr[p_i \leqslant P_i \leqslant p_i + \delta p_i]
\end{aligned}
$$

Thus we have proved the asymptotical independence. □

Remark 2.4. In the previous two proofs, we assumed $d = 1$. In the general case where d is greater than 1, the momentum space of N copies of d-dimensional gas particles is the same as the momentum space of Nd copies of 1-dimensional gas particles. Thus we still have the Boltzmann distribution of each particle

$$
dPr = \frac{1}{(2\pi m k_B T)^{\frac{d}{2}}} \exp\left(- \frac{\sum_{i=1}^{d} p_i^2/2m}{k_B T} \right) dp_1 ... dp_d
$$

The momentums of different particles still tend to be independent as $N \to \infty$.

Remark 2.5. It is possible that different gas particles inside the box have different masses. In this case, the principle of indifference still guarantees us a uniform measure on the energy hypersurface. But it is no longer a sphere but an ellipsoid, where the ellipsoid is equipped with the canonical symplectic measure of the phase space. We can carry out the similar calculation and find out that the previous statements still hold.

3. Validity of the Principle of Indifference

For the distribution of wealth problem in the first section, we justified that the probability distribution converges to the uniform one on the state space. For statistical mechanics, the uniform measure comes arises from the **Principle of Indifference**. The justification of the principle of indifference is the **Ergodic Hypothesis**. In many physics problems, we are interested in a Hamiltonian system where the system evolves on a compact Hamiltonian level surface S. (For example, the Hamiltonian for the ideal gas is the total kinetic energy. When the box is adiabatic, the state of the system is confined to the surface of constant total kinetic energy.) S is always equipped with a canonical symplectic measure μ that is invariant under the evolution ϕ_t.

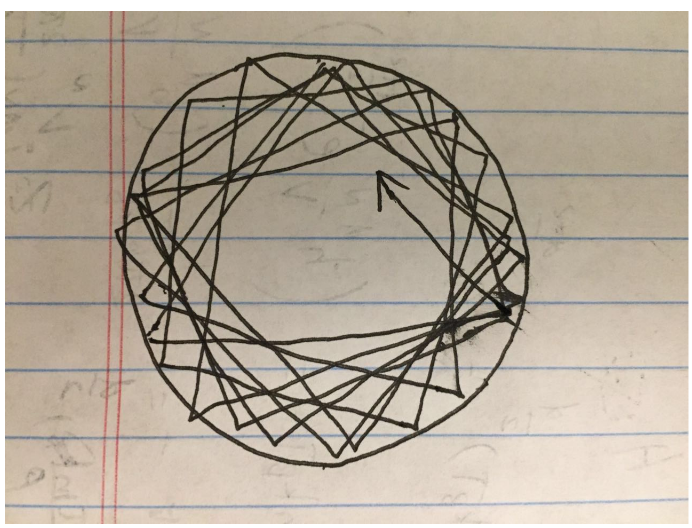
The Ergodic Hypothesis says the system is ergodic. We say the system is ergodic when the only set invariant under the evolution has either zero or full measure. But ergodicity has a more intuitive meaning due to Birkhoff ergodic theorem: starting from almost every point in S, the trajectory visits the whole surface S evenly. In the more precise mathematical term: let $A \subseteq S$ be any measurable subset and $f = \mathbf{1}_A$ be the indicator function supported on A. Then for almost all initial condition $x_0 \in S$, we have

$$\lim_{T \to +\infty} \frac{1}{T} \int_0^T f(\phi_t(x_0))dt = \frac{\mu(A)}{\mu(S)}$$

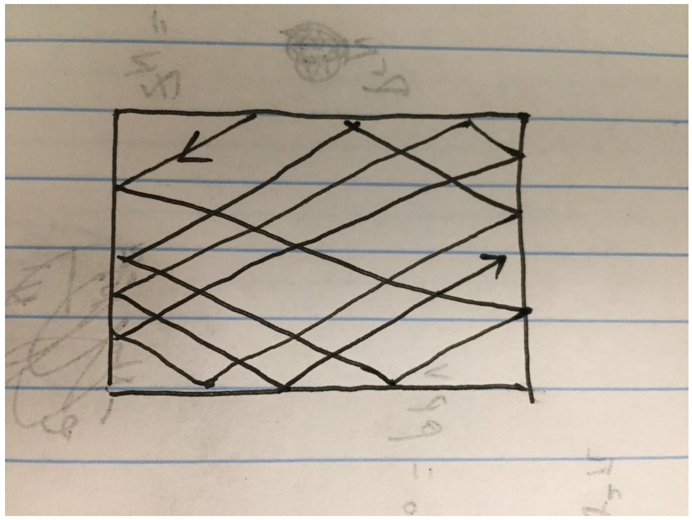
Since the trajectory visit the whole space S evenly, without prior knowledge of the system, we will observe the system in any of the possible states evenly likely. Thus the Ergodic hypothesis justifies the principle of indifference.

Sadly the Ergodic Hypothesis is not always valid. For instance S would better be compact so that μ is a finite measure. In addition, we have two famous examples where the Ergodic Hypothesis fails.

(Example 1) Let B^2 be the unit disk. Consider a particle of mass 1 moving at unit speed inside Ω and the particle collides elastically with ∂B^2. In this case, the hypersurface S inside the phase space is the unit cotangent bundle of B^2, that is $B^2 \times S^1$. The system is not ergodic, because if the particle starts near the boundary and has initial momentum approximately tangential to the boundary, then it will never reach the central region of the disk. (See following diagram for illustration)



(Example 2) Let I^2 be the unit square replacing B^2 in the previous example. Then $S = I^2 \times S^1$. We observe the momentum at x-direction and the momentum at y-direction of the particle does not transfer to each other. Therefore the system cannot be ergodic. (See following diagram for illustration)



Finally we examine the ideal gas model. If we assume the gas particles have zero volume do not collide. Then there will be no transfer of kinetic energy from one particle to another and thus the system cannot be ergodic. Then the momentum of different particles no longer distributes uniformly on the constant energy sphere L. If we assume the gas particles have positive volume. Then the position of different particles no longer distributes uniformly on Ω^N, because two distinct particles cannot be too close to each other. Unfortunately in neither cases, the state of the ideal gas will distribute uniformly on $\Omega^N \times L$.

4. Boltzmann Distribution Arises as the Maximal Entropy Distribution

So far we have seen the Boltzmann distribution arising from uniform distribution on the state space. In this section, we provide an alternative reason why the Boltzmann distribution should arise. We will first demonstrate for the discrete system and then for the continuous system.

Suppose we have N identical particles, each of which is at some state labeled by $M = \{0, 1, 2, ..., m\}$. The total energy of these particles is a fixed constant E. Let $\epsilon = \frac{E}{N}$ be the average energy. The j-th state is associated with an energy level $\epsilon_j \geqslant 0$. We want to learn how is energy distributed among these particles, i.e. how large is n_j the number of particles observed at the j-th state.

The configuration of the system is described by an N-tuple $(x_1, ..., x_N) \in M^N$. The principle of indifference tells us we are equally likely to observe any configuration of the system with total energy E, i.e. each element in

$$S = \left\{ (x_1, ..., x_N) : \sum_{j=1}^{N} \epsilon_{x_j} = E \right\}$$

is equally likely to be observed. Once we know the configuration of the system, we know how many particles are at each j-th state, i.e. n_j. The distribution of $(n_0, ..., n_m)$ is peaked somewhere it corresponds to the maximum number of configurations in S. Let $\#$ be the number of configuratins corresponding to $(n_0, ..., n_m)$. Then $\#$ equals the number of ways to put N objects into containers of size $n_0, ..., n_m$ respectively. Thus

$$\# = \frac{N!}{\prod_i n_i!}$$

We want to find $(n_0, ..., n_m)$ which maximizes $\#$, subject to the conservation of particles and conservation of total energy

$$\begin{cases} \sum_j n_j = N \\ \sum_j n_j \epsilon_i = E \end{cases}$$

Maximizing $\#$ is the same as maximizing $\log(\#)$. Using Stirling's approximation

$$\begin{aligned} &\log(\#) \\ &\sim N\log(N) - N - \sum_j (n_j \log(n_j) - n_j) \\ &= N\log(N) - \sum_j n_j \log(n_j) \\ &= -\sum_j N\frac{n_j}{N}\log(\frac{n_j}{N}) \\ &= -N\sum_j p_j \log(p_j) \end{aligned}$$

Remark 4.1. Boltzmann gave the above quantity $\log(\#)$ a famous name, entropy, while in modern texts, entropy is more often referred to $-\sum_j p_j \log(p_j)$.

We compute the total differential

$$d\log(\#) = -\sum_j \log(n_j) dn_j \quad \text{(We view \# as function of } n_j)$$

The achieve maximum, there are some real numbers α, β such that

$$d\log(\#) = \alpha dN + \beta dE = \sum_i (\alpha + \beta\epsilon_j) dn_j$$

Therefore for all j:

$$-\log(n_j)dn_j = (\alpha + \beta\epsilon_j)dn_j \Rightarrow n_j = \exp(-\alpha - \beta\epsilon_j)$$

Since $\sum_j n_j = N$, we obtain $N\exp(\alpha) = \sum_i \exp(-\beta\epsilon_i)$. Therefore the proportion, a.k.a. probability, of a particle to be at j-state is

$$p_j = \frac{n_j}{N} = \frac{\exp(-\beta\epsilon_j)}{\sum_j \exp(-\beta\epsilon_j)} = \frac{\exp(-\beta\epsilon_j)}{Z}$$

Notice the probability to be at j-th state p_j is inverse proportional to the exponential of energy $\exp(\beta\epsilon_j)$, the Boltzmann distribution appears again!

Remark 4.2. β is called the thermodynamic beta. It is usually positive. It defines the temperature by the relation $\beta = \frac{1}{k_B T}$. Note higher the temperature is, lower β will be and the particle tends to be more probable to be at higher energy state. This is consistant with our usual perception that temperature is a measure of average kinetic energy.

Remark 4.3. Although we only checked the first derivative of $\log(\#)$, routine computation will show $\log(\#)$ is indeed maximized. Besides, $p_j = \frac{\exp(-\beta\epsilon_j)}{Z}$ is also the unique distribution maximizing the entropy $-\sum_j p_j \log(p_j)$ subject to contraint $\sum_j p_j = 1$, $\sum_j p_j\epsilon_j = \epsilon$.

Remark 4.4. All the previous statements and computation holds for countable state space $M = \mathbb{N}$.

(The continuous case) Let ρ be a probability density on phase space X equipped with the Hamiltonian function H. Then it should satisfy

$$1 = \int_X \rho dq dp \tag{4.5}$$

Suppose the average energy of the particle is fixed at ϵ, i.e.

$$\epsilon = \int_X \rho H dq dp \tag{4.6}$$

In the discrete case, the entropy is

$$-\sum_j p_j \log(p_j)$$

The continuous analogue of entropy is

$$-\int_X \rho \log(\rho) dq dp$$

Consider entropy as the linear functional

$$L[\rho] = -\int_X \rho \log(\rho) dq dp$$

We want to maximize the entropy $L[\rho]$ subject to the linear constraints 4.5 and 4.6. If we perturb and consider $L[\rho + t\eta]$ where $\rho + t\eta$ satisfies 4.5 and 4.6, L should be maximized at $t = 0$. Therefore the following identities should hold

$$0 = \frac{d}{dt} L[\rho + t\eta] = -\int_X (\log(\rho) + 1)\eta dq dp = -\int_X \log(\rho)\eta dq dp$$

For all η such that

$$\begin{cases} 0 = \int_X \eta dq dp \\ 0 = \int_X \eta H dq dp \end{cases}$$

Therefore $\log(\rho)$ would better be a linear combination of 1 and H. Thus there are real number α, β such that

$$-\log(\rho) = \alpha + \beta H \Rightarrow \rho = \frac{\exp(-\beta H)}{\exp(\alpha)}$$

Since ρ is a probability measure, we obtain

$$\rho = \frac{\exp(-\beta H)}{\int_X \exp(-\beta H) dqdp} = \frac{\exp(-\beta H)}{Z}$$

Thus the distribution of the particle in its phase space obeys the Boltzmann distribution.

5. Application to Particles in the Box

Now we apply the Boltzmann distribution to the particles in the box. Consider a group of non-interacting identical particles inside a d-dimensional box Ω. Each particle experience a potential V. Then the group of particles obey the Boltzmann distribution: the distribution of each particle in its phase space independently and each particle obeys the Boltzmann distribution

$$dPr = \frac{1}{Z} \exp\bigg(- \frac{H}{k_B T} \bigg) dqdp$$

Where $Z = \int_X \exp(-\frac{H}{k_B T}) dqdp$ is the normalization factor and $H(q,p) = \frac{p^2}{2m} + V(q)$ is the total energy. Now we calculate the expected kinetic energy of each particle. We denote the momentum by $p = (p_1, ..., p_d)$.

Recall that $\int_{\mathbb{R}} e^{-x^2} dx = \sqrt{\pi}$ and $\int_{\mathbb{R}} x^2 e^{-x^2} dx = \frac{\sqrt{\pi}}{2}$.We compute

$$\begin{aligned}
& Z \\
&= \int_M \exp\bigg(- \frac{H(q,p)}{k_B T} \bigg) dqdp \\
&= \int_{\mathbb{R}^d} \exp\bigg(- \frac{p^2}{2mk_B T} \bigg) dp \int_\Omega \exp\bigg(- \frac{V(q)}{k_B T} \bigg) dq \\
&= (2\pi m k_B T)^{\frac{d}{2}} \int_\Omega \exp\bigg(- \frac{V(q)}{k_B T} \bigg) dq
\end{aligned}$$

And the expected kinetic energy along j-th coordinates

$$\begin{aligned}
& \int_M \frac{p_j^2}{2m} \frac{1}{Z} \exp\bigg(- \frac{H(q,p)}{k_B T} \bigg) dqdp \\
&= \frac{1}{Z} \int_{\mathbb{R}^d} \frac{p_j^2}{2m} \exp\bigg(- \frac{p^2}{2mk_B T} \bigg) dp \int_\Omega \exp\bigg(- \frac{V(q)}{k_B T} \bigg) dq \\
&= \frac{1}{Z} (2\pi m k_B T)^{\frac{d}{2}} \frac{k_B T}{2} \int_\Omega \exp\bigg(- \frac{V(q)}{k_B T} \bigg) dq \\
&= \frac{k_B T}{2}
\end{aligned}$$

Therefore the average kinetic energy of each particle is $\frac{dk_B T}{2}$. (Note the result agrees with the formula of average kinetic energy of ideal gas particles) This fact tells us temperature is proportional to the average kinetic energy of each particle.

6. Application to Gas Particles under Gravity

We all know atmosphere at higher altitude gets thinner. We want to study this phenomenon quantitatively. In an idealized situation, the atmosphere is in thermo-equilibrium and has a constant temeprature T. Then according to Boltzmann distribution, the density of gas particles with mass m at height h is

$$\rho(h) = \rho_0 \exp(-\frac{mgh}{k_B T})$$

Where ρ_0 is the density at zero height.

The pressure at height h should equal the gravity of all the gas particles above height h divided by the area. Thus

$$P(h) = \int_h^{+\infty} mg\rho(s)ds = k_B T \rho_0 \exp(-\frac{mgh}{k_B T})$$

On the other hand, the ideal gas law tells us

$$\begin{aligned} & PV = Nk_B T \\ \Rightarrow\ & P = \rho k_B T \quad \text{(Divide by the volume at both side)} \end{aligned}$$

i.e. the pressure of ideal gas is proportional to its density. The ideal gas law tells us

$$P(h) = \rho(h) k_B T = k_B T \rho_0 \exp(-\frac{mgh}{k_B T})$$

Thus the ideal gas law gives a consistent prediction of pressure. In fact, the Boltzmann distribution is the only distribution such that the pressure at height h ballances the gravity of gas particles above such height, and thus let the system stay in equilibrium. This fact leads us the see that Boltzmann distribution arises as the equilibrium distribution.

7. Boltzmann Distribution Arises as the Equilibrium Distribution

Motivated by the example in previous section, consider a group of N particles inside a box Ω and each particle experience a potential V. Suppose the system is at thermo-equilibrium with a fixed temperature T. We want to learn the density of particles at difference places inside Ω.

Recall that $-\nabla V$ is the force experienced by each particle. Thus $\int_A -\rho \nabla V dq$ is the rate of change of momentum in the region $A \subseteq \Omega$. On the other hand, the particles experience pressure which also contribute to the rate of change of momentum. The rate of change of momentum in region X due to pressure is $\int_A -\nabla P dq$ (the Archimedean's law). When the system is at equilibrium, they have to cancel each other

$$-\rho\nabla V = \nabla P$$
$$\Rightarrow -\rho\nabla V = k_B T \nabla\rho$$
$$\Rightarrow \frac{\nabla\rho}{\rho} = -\frac{\nabla V}{k_B T}$$

We can verify that $\rho = C\exp(-\frac{V}{k_B T})$ satisfies the above equation. Thus the density of particles is proportional to $\exp(-\frac{V}{k_B T})$. We arrive at the Boltzmann distribution again.

Acknowledgments

I would like to thank my mentor Clark Butler. He recommended me Calegari's note on statistical mechanics. I found his notes interesting and enlightening. I learned how Boltzmann distribution, under reasonal assumption of the physical system, will arise. Clark also taught me about the dynamical system and a bit ergodic theory and how they are connected to statistical mechanics, helping me finalize this paper.

References

[1] Danny Calegari. *Notes on Thermodynamic Formalism.* http://math.uchicago.edu/ dannyc/courses/thermodynamics_ 2014/thermodynamics_ notes.pdf

[2] Mehran Kardar. *Statistical Physics of Particles.* ISBN-13: 978-0521873420