UCSB, Physics 129L, Computational Physics
Lecture notes, Week 5

Zihang Wang (UCSB), zihangwang@ucsb.edu

February 4, 2025

Contents

1 Probability

1.1 Random Variable

A **random variable** is an abstraction of a quantity that depends on random events, and it can be discrete or continuous. In other words, it is a function that maps outcomes in the state space to real numbers. For example, the toss of a coin can result in one of two possible outcomes: "Heads" ($X = 1$) or "Tails" ($X = 0$), and $X = \{0, 1\}$ is an abstraction of elements in the **state space** $\{0, 1\}$ that contains all possible outcomes. Conversely, the stopping direction of a spinning top represents a continuous random variable. If we assume the observations and the associated random variables are independent and identically distributed (i.i.d.), it simplifies many statistical analyses and gives rise to important concepts and applications such as the Central Limit Theorem (CLT), hypothesis testing, Markov chains, Bayesian inference.

As an example, in quantum physics, random variables are often treated as observables, such as energy (E), position (x), momentum (k), and spin (s_z), each of which is typically associated with underlying eigenstates, denoted as $|E\rangle$, $|x\rangle$, $|k\rangle$, and $|s_z\rangle$, respectively. Loosely speaking, the density matrix, constructed from an eigenstate serves as a "random variable", e.g. $x \to |x\rangle\langle x|$.

To demonstrate this idea, let's consider a general quantum state,

$$|\Psi\rangle = \int \mathrm{dx}\psi(x)|x\rangle, \quad \int \mathrm{dx}|\psi(x)|^2 = 1, \quad \rho = |\Psi\rangle\langle\Psi|, \tag{1}$$

where $\psi(x)$ is the position wavefunction, and the second equality ensures the overall normalization. For each random variable, $|x\rangle\langle x|$, the corresponding probability density at particular position can be calculated from the trace, e.g. $\mathrm{Tr}(\rho|x\rangle\langle x|) = |\psi(x)|^2$ where $\mathrm{Tr}(\dots) = \int \mathrm{dx}\langle x|(\dots)|x\rangle$. The probability of observing a particle within in the position window $x \pm \mathrm{dx}$ is given by $P(x) = |\psi(x)|^2\mathrm{dx}$.

It is particularly useful when dealing with ensembles of quantum systems,

$$\rho = \sum_j p_j|\Psi_j\rangle\langle\Psi_j|, \tag{2}$$

where now ρ represents a mixed state composed of multiple pure states $|\Psi_j\rangle\langle\Psi_j|$ with associated probability p_j.

1.2 Probability Distributions

Probability is a measure of the likelihood that an event will occur.

- **Probability Mass Function (PMF)** for discrete X:

$$p(x) = P(X = x). \tag{3}$$

- **Probability Density Function (PDF)** for continuous X:

$$P(a \le X \le b) = \int_a^b f(x)\, dx. \tag{4}$$

- **Cumulative Distribution Function (CDF)**:

$$F(x) = P(X \le x). \tag{5}$$

1.3 Conditional Probability

Given two events A and B with $P(B) > 0$, the probability of A given B is defined as:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{6}$$

This represents the probability of A occurring under the condition that B has already occurred.

1.4 Marginal probability

Marginal probability refers to the probability of an event occurring with respect to a single random variable, ignoring the influence of other variables. It is derived from the joint probability distribution of multiple variables by summing or integrating over the other variables.

In the context of two random variables X and Y, the marginal probability of X is calculated as:

$$P(X) = \sum_{y} P(X, Y = y) \quad \text{(discrete)}$$

or

$$P(X) = \int P(X, Y)\, dY \quad \text{(continuous)}$$

This gives the total probability distribution of X, independent of Y. Similarly, the marginal probability of Y is obtained by summing or integrating over X.

1.5 Independence

Two events A and B are independent if:

$$P(A \cap B) = P(A)P(B). \tag{7}$$

1.6 Expectation and Variance

- **Expectation** (mean):

$$E[X] = \sum_{x} xp(x) \quad \text{(discrete)}, \quad E[X] = \int_{-\infty}^{\infty} xf(x)\, dx \quad \text{(continuous)}. \tag{8}$$

- **Variance**:

$$\mathrm{Var}(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2. \tag{9}$$

1.7 Multiplicity and ensembles

Let's consider a box contains n beads with $m \leq n$ colors, and there are w_i beads for each color type such that $\sum_i w_i = n$. We are interested in possible outcomes when drawing u beads from the box. The random variable X is defined as "u beads with a specific color combination", and the corresponding state space S_u contains all possible color combinations of u beads, $X \in S_u$.

For example, we want to draw three beads $u = 3$ from a box contains three white beads ($w_1 = 3$) and two black beads ($w_2 = 2$), with total of five beads $n = 5$. The state space is given by the following,

$$S_5 = \{X_1, X_2, X_3, X_4, X_5, \dots\} = \{(\text{white}, \text{white}, \text{white}),\\ (white, black, white), (white, white, black), (black, white, white), (black, black, white), \dots\}, \tag{10}$$

and the elements X_i are called "microstates" in the state space. Let's consider the concept of an ensemble. For example, we want to paint a car using the color mixture based on the color of three beads drawn from the box. Each microstate is associated with a color blend, such that,

$$\begin{aligned} &\text{white+white+white=white,}\\ &\text{white+black+white=light gray,}\\ &\text{black+black+white=dark gray.} \end{aligned} \tag{11}$$

There are microstates that have the same color blend: It is call the multiplicity Ω. It captures all degenerate configurations with respect to a color blend. We should note that the color blend is an intensive property that is independent of the system size: as long as you maintain the same ratio, the color bend remains the same (analogous to the energy density). On the other hand, the multiplicity is extensive as it explicitly depends on the number of beads (analogous to the system entropy).

An ensemble is a collection of all possible microstates of a system, consistent with specified macroscopic constraints (such the combined color), along with a probability distribution that assigns weights to each microstate. In particular, if the macroscopic constraints are thermodynamic variables, such as energy, temperature, volume, and particle number, the ensemble is a **thermal ensemble**.

The multiplicity $\Omega_{\text{white}} \cdot \Omega_{\text{black}}$ associated with a color blend is given by the binomial factor,

$$\Omega = \binom{n}{m} = \frac{n!}{m!(n-m)!}, \tag{12}$$

as "n choose m". The number of all possible microstates is 10 with multiplicity,

$$\Omega_{\text{white}} = \binom{3}{3} = 1, \quad \Omega_{\text{lightgray}} = \binom{3}{2}\binom{2}{1} = 6, \quad \Omega_{\text{darkgray}} = \binom{3}{1}\binom{2}{2} = 3. \tag{13}$$

To obtain the probability masses associated with each blend color, we repeat the above process N times and get,

$$\begin{aligned}
\frac{\#\text{white}}{N} &= 1/10, \\
\frac{\#\text{lightgray}}{N} &= 6/10, \\
\frac{\#\text{darkgray}}{N} &= 3/10.
\end{aligned} \tag{14}$$

This simple example illustrates that when dealing with dimension reductions, in most cases, you cannot get away from "data compression" (multiplicity). As a side note, there is a famous computational technique called the "hashing", as the back-bone of the crypto market and blockchain applications. It seems that it gets away from the multiplicity introduced by the dimension reduction, people always afraid that someone else somehow happens to have an identical private key as theirs. So they can take the bitcoin away.

There is a very important assumption we made without an explicit explanation: why the occurrence of microstates are considered equal-probable? This is a very deep and subtle question, rooted in the notion of **ergodicity**: will the system eventually visit all parts of the state space in a uniform manner where we can relate the time averages to ensemble averages. Let's go back to the color blend: for each color blend, will all microstates within become equiprobable when you repetitively sample from the box? In statistical mechanics, **this is a hypothesis that defines the thermal equilibrium for an isolated system: at a given energy surface, all microstates are equiprobable**. This hypothesis is often used as a starting point or fundamental postulate in statistical mechanics.

Let's consider a more complex case where you are asked to make a specific color blend named "very light gray" out of the N drawings you made previously. Let's consider that,

$$\text{very light gray} = \begin{cases} 0.7\text{white} + 0.3 \text{ dark gray} \\ 0.5\text{white} + 0.5 \text{ light gray} \\ 0.6\text{white} + 0.3 \text{ light gray} + 0.1 \text{ dark gray} \\ \ldots \end{cases} \tag{15}$$

where the coefficients are (weight) probability mass functions associated with a given color blend. The next question is, how can we systematically assign weights according to the color blend? This measure can be arbitrary and only depends on the microstate X_i, i.e. it does not need to be the color blend.

Let's consider various measure that ends you with the paint you want:

1) For environmental reasons, the city you lives require additional tax on the black ink you use for painting the car. Therefore, each time you draw three beads, the more black beads inside, it hurts your feeling, and you put in back and draw it again. Therefore, your "very light gray" is 0.5white+0.5 light gray.

2) Your boss want to give you a bonus to cover on the car, and you want to spend less and keep the rest. Since you know black ink is more expensive, you

first draw a few black beads to show the price to your boss, then you draw all white beads. Therefore, your "very light gray" is $0.7\text{white} + 0.3$ dark gray.

We should note that regardless of your drawing preference, the total amount of ink is fixed. It is hard for people to guess your thoughts process when you buy certain things.

You may ask are there a natural way to generalize the above process? Let's say your decide to keep your drawing based on how frequent you get them and how you like. You set your algorithm such that each time you draw a microstate X_i, you slightly tune down (or up) the probability of acceptance based on your preference $E(X_i) \in \mathbb{R}$, scaled by a weight $\beta \in \mathbb{R}$,

$$P_{\text{prefer}}(X_i) \sim \left(1 - \beta\frac{E(X_i)}{N}\right)^N \approx \exp\left(-\beta E(X_i)\right). \tag{16}$$

We name $E(X_i)$ as the energy (willingness) associated with the microstate X_i. There are certain microstates that have equal energy (willingness).

If we set $\beta = 1/k_B T$, it is called a **canonical ensemble** with the following **Boltzmann factor**,

$$P_{\text{prefer}}(X_i) = \frac{1}{Z}\exp\left(-\frac{[E(X_i)]}{k_B T}\right), \tag{17}$$

where Z is a normalization factor,

$$Z = \sum_i \exp\left(-\frac{[E(X_i)]}{k_B T}\right) = \sum_E \Omega_E \exp\left(-\frac{E}{k_B T}\right), \tag{18}$$

where in the last line, we convert the sum in microstates into a sum of energy surfaces.

You should note the difference between the above probability and the probability for drawing a specific microstate X_i with a given energy $E(X_i)$: the latter is given by the multiplicity, following a uniform distribution,

$$P_{\text{draw}}(E(X_i)) \sim \Omega_i, \tag{19}$$

where the associated entropy is,

$$S_{\text{draw}}(E(X_i)) = k_B \ln(P_{\text{draw}}) \sim k_B \ln(\Omega_E), \tag{20}$$

and k_B is the Boltzmann constant.

1.8 Partition Functions and Free energy

The normalization factor,

$$Z = \sum_i \exp\left(-\frac{E(X_i)}{k_B T}\right), \tag{21}$$

is an important quantity called **partition function**. We can express energy and entropy using the probability $P_{\text{prefer}}(X_i)$,

$$-k_B \ln[P_{\text{prefer}}(X_i)] = \frac{E(X_i)}{T} + k_B \ln(Z), \tag{22}$$

with expectation value $\langle \cdots \rangle = \sum_i P_{\text{prefer}}(X_i) \cdots$, such that,

$$\begin{aligned} -k_B \sum_i P_{\text{prefer}}(X_i) \ln[P_{\text{prefer}}(X_i)] &= \frac{\langle E \rangle}{T} + k_B \ln(Z) \\ S &= k_B \ln(Z) + \frac{\langle E \rangle}{T}. \end{aligned} \tag{23}$$

where we introduce the **thermodynamic entropy** (extra minus sign to ensure positivity),

$$\mathcal{S} = -k_B \sum_i P_{\text{prefer}}(X_i) \ln[P_{\text{prefer}}(X_i)]. \tag{24}$$

You draw three beads with probability $P_{\text{draw}}(X_i)$, then the entropy measures how uncertain you are in accepting it based on your preferences.

Let's go back to physical systems. $P_{\text{prefer}}(X_i)$ is the probability of observing a particular state (X_i) with energy $E(X_i)$.

Using Eq. 23, the thermodynamic entropy defines the concept of temperature via the following functional derivative,

$$\frac{\partial \mathcal{S}}{\partial \langle E \rangle} = \frac{1}{T} = \beta. \tag{25}$$

It measures that under one unit of energy boost, how much more chaotic the system becomes. From the same equation Eq. 23, we have,

$$-\frac{\partial \ln(X)}{\partial \beta} = \langle E \rangle. \tag{26}$$

We also have the **Helmholtz free energy** $\langle \mathcal{F} \rangle$,

$$\langle \mathcal{F} \rangle = \langle E \rangle - T\mathcal{S} = -k_B T \ln(Z). \tag{27}$$

It reflects the total "energy" or "information" available within the system. You should note that free energy is defined as a thermodynamical quantity.

Let's look at the expectation value,

$$\langle A \rangle = \sum_i A(X_i) P_{\text{prefer}}(X_i) = \frac{1}{Z} \sum_i A(X_i) \exp\left(-\frac{E(X_i)}{k_B T}\right) = \frac{1}{Z} \sum_E A(E) \exp\left(-\frac{E}{k_B T}\right). \tag{28}$$

Alternatively, it can be directly obtained via the following functional derivative,

$$\langle A \rangle = \left.\frac{\partial \langle F \rangle}{\partial t}\right|_{t=0} = -k_B T \left.\frac{\partial}{\partial t}\right|_{t=0} \ln(Z) = \frac{-k_B T}{Z} \left.\frac{\partial Z}{\partial t}\right|_{t=0}, \tag{29}$$

where the partition function includes a source term,

$$Z = \sum_i \exp\left(-\frac{[E(X_i) + A(X_i)t]}{k_B T}\right), \quad \left.\frac{\partial Z}{\partial t}\right|_{t=0} = -\sum_i \frac{A(X_i)}{k_B T} \exp\left(-\frac{E(X_i)}{k_B T}\right). \tag{30}$$

This gives,

$$\langle A \rangle = \left.\frac{\partial \langle F \rangle}{\partial t}\right|_{t=0} = \frac{1}{Z}\sum_i A(X_i) \exp\left(-\frac{E(X_i)}{k_B T}\right). \tag{31}$$

In statistics, this has an important name: cumulant generating function, which will be soon talking about it.

1.9 Legendre transforms

Now, we have the free energy for canonical ensemble, where the temperature T is fixed (independent) as a preference measure. What if we want to look at different preference measure? Let's consider the concept of **Legendre Transform** that converts a dependent variable to an independent one. Let's start with a general function $f(x)$. Let's consider its differential,

$$\partial f(x) = \frac{\partial f(x)}{\partial x}\partial x = y\partial x, \tag{32}$$

where we set $y = \frac{\partial f(x)}{\partial x}$ as a new variable, and integrate both sides with respect to x,

$$\int \partial f(x) dx = \int y \partial x + g(y), \rightarrow f(x) = xy - g(y), \tag{33}$$

where the minus sign is because of the following symmetry $df/dx = y, dg/dx = x$ (the transformation is cyclic). **y is a Legendre conjugate pair, which is different from the canonical conjugate pair defined in the Hamiltonian formalism**.

1.9.1 Legendre transforms between Lagrangian and Hamiltonian

Let's briefly look at the transformation between the Lagrangian and Hamiltonian formalism via Legendre transform. A **Lagrangian** is defined as $L(q, \dot{q}, t)$. In simple mechanical systems, we have,

$$L(q, \dot{q}, t) = T(\dot{q}) - V(\dot{q}, q), \tag{34}$$

where T is the kinetic energy and V is the potential energy. The general form of the Euler-Lagrange equation is,

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) = \frac{\partial L}{\partial q}. \tag{35}$$

However, as one can see, most systems contain second derivative in time, which is hard to solve in time domain.

The **Hamiltonian** $H(q, p, t)$ is obtained by performing a Legendre transform of L,

$$H(q, p, t) = p\dot{q} - L(q, \dot{q}, t), \tag{36}$$

where $\dot{q}$ is replaced by p using the relation:

$$p = \frac{\partial L}{\partial \dot{q}}. \tag{37}$$

Unlike the Legendre conjugate pair $(p, \dot{q})$, the canonical conjugate pair is given by the momentum and position (p, q). The Hamiltonian $H(q, p, t)$ often represents the **total energy** of the system, e.g. in a simple mechanical system,

$$H = T(\dot{q}) - V(\dot{q}, q), . \tag{38}$$

The Legendre transform allows us to switch from a description using **velocities** (Lagrangian) to one using **momenta** (Hamiltonian). We can derive Hamilton's equations from the Euler-Lagrange equations. We start by differentiating the Hamiltonian with respect to time. Using the chain rule:

$$\frac{dH}{dt} = \sum_i \left(\frac{\partial H}{\partial q_i}\dot{q}_i + \frac{\partial H}{\partial p_i}\dot{p}_i \right). \tag{39}$$

Using the fact that $\dot{q}_i = \frac{\partial H}{\partial p_i}$ (which is derived from the Legendre transformation), we substitute into the expression for $\frac{dH}{dt}$,

$$\frac{dH}{dt} = \sum_i \left(\frac{\partial H}{\partial q_i}\frac{\partial H}{\partial p_i} + \frac{\partial H}{\partial p_i}\dot{p}_i \right). \tag{40}$$

This leads to the **Hamilton's equations**,

$$\dot{q}_i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q_i}, \tag{41}$$

and the above equations are a symmetric first-order formulation of mechanics.

Let's introduce a more general formulation that describes the time evolution of given observables and generalizes the concept of the rate of change of observables, namely the **Poisson Bracket Formulation**. The time evolution of a function $f(q, p)$ in phase space is given by the Poisson bracket with the Hamiltonian $H(q, p)$,

$$\frac{df}{dt} = \{f, H\} = \sum_k \left(\frac{\partial f}{\partial q_k}\frac{\partial H}{\partial p_k} - \frac{\partial f}{\partial p_k}\frac{\partial H}{\partial q_k} \right). \tag{42}$$

The canonical variables are the generalized coordinates q_i and their corresponding generalized momenta p_i. These variables are chosen such that they form a canonical pair and form the **phase space**. The pair (q_i, p_i) is chosen because they satisfy the **canonical conjugation relations**,

$$\{q_i, p_j\} = \delta_{ij}, \quad \{q_i, q_j\} = 0, \quad \{p_i, p_j\} = 0, \tag{43}$$

where $\{,\}$ denotes the Poisson bracket, and δ_{ij} is the **Kronecker delta**, which equals 1 when $i = j$ and 0 otherwise.

For the generalized coordinates q_i and momenta p_i, Hamilton's equations are:

$$\dot{q}_i = \frac{dq_i}{dt} = \{q_i, H\}, \quad \dot{p}_i = \frac{dp_i}{dt} = \{p_i, H\}. \tag{44}$$

In particular, let look at the **Liouville Equation** that describes the flow of the probability distribution in phase space,

$$\frac{\partial \rho}{\partial t} + \{\rho, H\} = 0. \tag{45}$$

The Liouville equation expresses the conservation of probability in phase space under Hamiltonian dynamics. Similarly, the time evolution of quantum density matrix $\hat{\rho} = \sum_i P_i |\psi_i\rangle\langle\psi_i|$ follows the Von Neumann equation, which can be derived directly using the Schrodinger equation,

Taking its time derivative of the density matrix,

$$\frac{d}{dt}\hat{\rho} = \sum_i P_i \left(\frac{d}{dt}|\psi_i\rangle\langle\psi_i| + |\psi_i\rangle \frac{d}{dt}\langle\psi_i| \right), \tag{46}$$

and the Schrodinger equation gives,

$$\frac{d}{dt}|\psi_i\rangle = -\frac{i}{\hbar}\hat{H}|\psi_i\rangle, \quad \frac{d}{dt}\langle\psi_i| = \frac{i}{\hbar}\langle\psi_i|\hat{H}. \tag{47}$$

Substituting the above into the time derivative of $\hat{\rho}$,

$$\frac{d}{dt}\hat{\rho} = -\frac{i}{\hbar}[\hat{H}, \hat{\rho}]. \tag{48}$$

This is the quantum version of the Liouville Equation.

1.9.2 Legendre transforms in thermodynamics

Let's look at the Helmholtz free energy again, $\mathcal{F} = E - T\mathcal{S}$. Following Legendre Transform, we notice that,

$$T = -\frac{\partial \langle F \rangle}{\partial \mathcal{S}}, \tag{49}$$

such that the temperature is a dual variable (or conjugate pair) of the thermodynamic entropy in the canonical ensemble.

However, there are additional contributions in the system so far we neglect. Those are volume V and number of particles N. Let's consider the first law of thermal dynamics,

$$dE = Td\mathcal{S} - PdV + \mu dN, \tag{50}$$

where P is the system pressure and μ is the chemical potential. You should note that all are intensive-extensive pairs, and in the first law, only extensive variables have differentials.

From here, we are able to define various ensembles. A general **Helmholtz free energy** is given by,

$$\mathcal{F} = E - T\mathcal{S} \quad \rightarrow \quad d\mathcal{F} = -PdV + \mu dN - \mathcal{S}dT, \tag{51}$$

and in thermodynamic equilibrium, we must have $d\mathcal{F} = 0$ at constant temperature, volume, and number of particles, namely the **Canonical ensemble**.

Grand Potential is given by the following combination,

$$\Omega_G = \mathcal{F} - \mu N \quad \rightarrow \quad d\Omega_G = d\mathcal{F} - d\mu N - \mu dN = -PdV - SdT - Nd\mu. \tag{52}$$

The grand potential is minimized in equilibrium at constant temperature, volume, and chemical potential, namely the **grand canonical ensemble**.

Gibbs free energy is given by the following combination,

$$G = \Omega_G + PV \quad \rightarrow \quad dG = d\Omega_G + PdV + VdP = VdP - SdT + Nd\mu. \tag{53}$$

Gibbs free energy is minimized in equilibrium at constant temperature, pressure, and number of particles, namely the **Gibbs ensemble**.

We should note that the fixed particle number does not mean that there are a fixed number of available energy states.

1.10 Ensembles with indistinguishable particles: Bosons

While the above discussion is general, particles are distinguishable. Consider an ideal gas composed of N non-interacting (fermionic, bosonic, or classical) particles. In the classical case, the occupation number $n(E)$ at a given energy is directly weighted by the Boltzmann factor,

$$P_{\text{prefer}}(X_i) = \frac{1}{Z}\exp\left(-\frac{E(X_i)}{k_B T}\right), \tag{54}$$

such that $n(E) = NP_{\text{prefer}}(X_i)$.

However, things are different if particles are indistinguishable. Let's consider N non-interacting particles within a fixed volume, each carrying an energy ϵ when moving and has zero energy when stopped (This can be extended further to additional energy levels). Under canonical ensemble, the partition function of the system is given by summing all possible arrangements with state multiplicity $\binom{N}{j}$,

$$Z = \frac{1}{N!}\sum_{j=0}^{N}\binom{N}{j}\exp(-\beta j\epsilon) = \frac{1}{N!}(1+e^{-\beta\epsilon})^N, \qquad P(E=n\epsilon) = \frac{1}{ZN!}\binom{N}{n}\exp(-\beta n\epsilon), \tag{55}$$

where we divide $N!$ since there are N arrangements within. For example, suppose we have a system with $N = 3$ particles, all of which are indistinguishable.

For example, if we want to find the number of ways to place two of these particles in a state with energy ϵ and the third in a different state, using

$$\binom{N}{n} = \binom{3}{2} = 3$$

gives us three configurations:

- First two particles has energy ϵ, third with zero energy.
- First and third has energy ϵ, second with zero energy.
- Second and third has energy ϵ, first with zero energy.

The average number of particles that disappear (e.g. zero energy) is given by,

$$N^C(\text{ground})\sum_{n=0}^{N}(N-n)P(E=n\epsilon) = \frac{1}{ZN!}\sum_{n=0}^{N}(N-n)\binom{N}{n}\exp(-\beta n\epsilon) = \frac{1}{ZN!}N(1+e^{-\beta\epsilon})^{N-1} = \frac{N}{1+e^{-\beta\epsilon}}. \tag{56}$$

Note that at high temperature, $\beta \to 0$, we have an average of $N/2$ from the system. Equal-likely to be in the ground state and the excited state.

In Bose-Einstein statistics, the state with N particles must be considered collectively as a fraction of the states with energy $\epsilon, 2\epsilon, 3\epsilon, \cdots$. The partition function is given by,

$$Z = \sum_{j=0}^{N}\exp(-\beta j\epsilon) = \frac{1-e^{-\beta\epsilon(N+1)}}{1-e^{-\beta\epsilon}}, \tag{57}$$

where the number of particles in the ground state is given by,

$$N(\text{ground}) = \sum_{n=0}^{N}(N-n)P(E=n\epsilon) = \frac{1}{Z}\sum_{n=0}^{N}(N-n)\exp(-\beta n\epsilon) = N - \frac{1}{Z}\frac{\partial Z}{\partial(\beta\epsilon)} = N - \frac{\partial[\ln(Z)]}{\partial(\beta\epsilon)}$$
$$N + \frac{1}{\exp(\beta\epsilon)-1} - \frac{(N+1)}{\exp(\beta\epsilon(N+1))-1} \tag{58}$$

So far, this equation is not defined when energy is zero since we do not consider any change in energy resulting from changes in particle number. Let's add the "chemical potential", an additional energy term that promote ($\mu < 0$) or demotes ($\mu > 0$) particle grouping. It is global such that it does not depend on specific states. Let's see how much it can collect particles purely based on amount of energy if we give them enough particles,

$$\Omega_G = \sum_{j=0}^{N\to\infty} \exp(-\beta[\epsilon j - \mu j]) = \frac{1-e^{-\beta(\epsilon-\mu)(N+1)}}{1-e^{-\beta(\epsilon-\mu)}} = \begin{cases} \frac{1}{1-e^{-\beta(\epsilon-\mu)}} & \beta(\mu-\epsilon) < 0 \\ \frac{1}{e^{-\beta(\epsilon-\mu)}-1}e^{-\beta(\epsilon-\mu)N} \to \infty & \beta(\mu-\epsilon) > 0 \end{cases}. \tag{59}$$

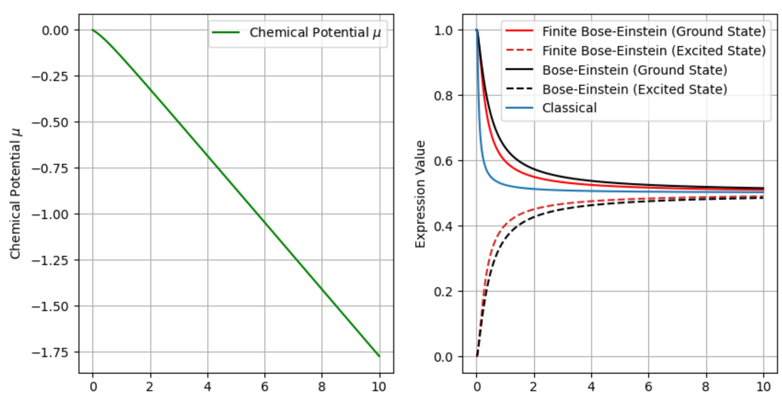
We can see that if $\mu - \epsilon > 0$, the partition function goes to infinity, which makes the system non-normalizable (unless population inversion is occurred, $\beta < 0$). The average number of particles is given by,

$$N(\text{excited state}) = k_B T \frac{\partial}{\partial \mu}\ln(\Omega_G) = \frac{1}{e^{\beta(\epsilon-\mu)}-1}, \qquad N(\text{ground state}) = k_B T \frac{\partial}{\partial \mu}\ln(\Omega_G) = \frac{1}{e^{\beta(0-\mu)}-1}, \tag{60}$$

such that,

$$N(\text{excited state}) + N(\text{ground state}) = N. \tag{61}$$

For a fixed N, a larger gap ϵ increases the stability of the ground state populations. You may notice the difference due to finite system size: it can be corrected using the exact expression above. Form the plot, we can see that for finite system, there is a shift in the occupation fraction.



Let's think about the rate of change in the ground state population. This gives by the derivative with respect to the temperature.

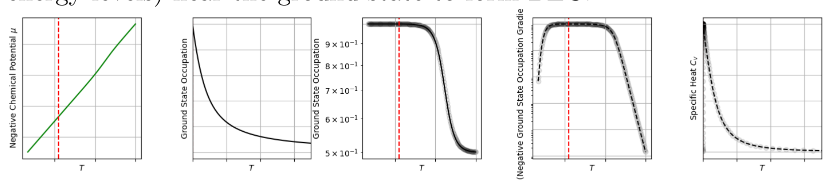
In addition, we consider two features, the first is the variation in energy,

$$\chi(T) = \frac{\partial N(\text{ground})}{\partial T} \leq 0, \tag{62}$$

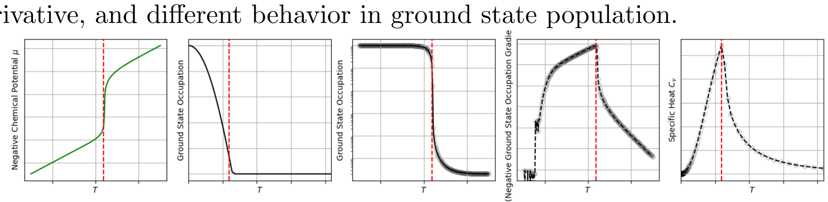
and the specific heat capacity,

$$C_V(T) = \frac{\partial E}{\partial T} \geq 0, \tag{63}$$

where we like to find singularity in its derivative, as a traditional signature of BEC. As you can see from the picture below, no BEC occur in a traditional sense. The reason is clear: we do not give the system enough density of states (or energy levels) near the ground state to form BEC.



On the other hand, let's increase the density of states. As you can see from the picture below, BEC will occur, as identified by the singularity in the derivative, and different behavior in ground state population.



1.11 Ensembles with indistinguishable particles: Fermion

Fermions are particles with half-integer spin (e.g., electrons, protons, neutrons), and they obey the Pauli exclusion principle, which states that no two identical fermions can occupy the same quantum state simultaneously. Let's use the same 2-level system discussed for bosons. In Fermi-Dirac statistics, The state with N particles must be considered collectively as a fraction of the states with $M \geq N$ energy levels $\epsilon, 2\epsilon, 3\epsilon, \cdots M\epsilon$, and the energy is filled one by one until there are N atoms. Therefore, you need to have at least N energy levels to put N atoms in. In fact, the ground state is defined as full occupation of the lowest N energy levels. The partition function is given by (less not considering the spin degenerated case) under canonical ensemble,

$$Z=\left(\sum_{j=0}^{1}\exp(-\beta j\epsilon)\right)\left(\sum_{j=0}^{1}\exp(-2\beta j\epsilon)\right)\cdots\left(\sum_{j=0}^{1}\exp(-\beta Mj\epsilon)\right)=\prod_{m=1}^{M}z_m=\prod_{m=1}^{M}\left(1+\exp(-\beta m\epsilon)\right). \tag{64}$$

Then, the population at a given energy is given by,

$$P(E=m\epsilon)=\frac{\exp(-\beta m\epsilon)}{z_m}=\frac{1}{1+e^{\beta m\epsilon}}, \tag{65}$$

and average occupation is simply given by,

$$N(E=m)=P(E=m\epsilon). \tag{66}$$

As seem in the exponential, it is trivial for the lowest occupations as it is almost impossible to excite any thing out of the lowest energy state. This is one reason that those energy levels are hide under the fermi sea. For this reason, most of time, we do not care them, at least for low energy physics. Let's neglect those below, and introduce a reference point that shifts our focus on somewhere else. The grand canonical ensemble is formulated such that,

$$P(E=m\epsilon)=\frac{\exp(-\beta m\epsilon)}{z_m}=\frac{1}{1+e^{\beta(m\epsilon-\mu)}}, \tag{67}$$

when $m\epsilon=\mu$, $P(E=\mu)=1/2$, which is called fermi-level (or surface).

The probability of finding a particle within the above energy window is given by the ratio,

$$\left(\frac{n_0}{N}\right)_{\text{fermi}}=\frac{g(E)}{e^{(E-\mu)/kT}+1},\ \left(\frac{n_0}{N}\right)_{\text{bose}}=\frac{g(E)}{e^{(E-\mu)/kT}-1}, \tag{68}$$

1.12 Moments, Cumulants, and Ensemble Averages

An analogous concept can be found in statistical physics, which deals with the thermodynamic properties of large ensembles of microstates. In this context, macroscopic physical properties such as temperature, particle number, and energy can be viewed as random variables, depending on ensembles. Consider an ideal gas composed of N non-interacting (fermionic, bosonic, or classical) particles, and we assume there are n_0 numbers of particles that fall within a specific energy range, $E\pm\epsilon$. The probability of finding a particle within the above energy window is given by the ratio,

$$\left(\frac{n_0}{N}\right)_{\text{fermi}}=\frac{g(E)}{e^{(E-\mu)/kT}+1},\ \left(\frac{n_0}{N}\right)_{\text{bose}}=\frac{g(E)}{e^{(E-\mu)/kT}-1}, \tag{69}$$

and

$$\left(\frac{n_0}{N}\right)_{\text{classical}}=g(E)\frac{e^{-(E_0-\mu)/kT}}{\sum_j e^{-(E_j-\mu)/kT}}, \tag{70}$$

where μ is the chemical potential and $g(E)$ is the density of states (multiplicity around a given energy). The v-th moment as,

$$\langle E^v \rangle = \int_{-\infty}^{\infty} E^v \left(\frac{n_0}{N}\right)_X \mathrm{d}E, \tag{71}$$

where X labels the statistics (fermionic, bosonic, or classical). The expectation value of an observable can be written as the sum moments,

$$\langle A \rangle = \int_{-\infty}^{\infty} A(E) \left(\frac{n_0}{N}\right)_X \mathrm{d}E = \sum_{v=0}^{\infty} A_v \langle E^v \rangle, \tag{72}$$

where A has an power expansion, $A(E) = \sum_{v=0}^{\infty} A_v E^v$.

Let's define the moment generating function that is proportional to the system partition function,

$$M(Et) = \sum_E \exp\left([t - \beta]E\right), \quad M(Et)|_{t=0} \sim Z. \tag{73}$$

where t is a small parameter around 0. This treatment is identical to the generation of propagators via the source term J under the Legendre transform. The v-th moment is generated by taking v derivatives to the moment generating function,

$$\langle E^v \rangle = \frac{d^v}{dt^v} M(Et)|_{t=0}. \tag{74}$$

This treatment is the statistical basis of many-body perturbation theory and Feynman diagram.

The v-th moment,

$$\langle E^v \rangle \sim (\langle E^{v/2} \rangle)^2 + \langle E^{v/3} \rangle \langle E^{2v/3} \rangle + (\langle E^{v/3} \rangle)^3 + \ldots, \tag{75}$$

contains polynomial powers of lower-order moments, but does not contribute to non-trivial correlations at v-th order. Cumulants keep the v-th order variations while subtract all lower-order polynomial contributions. Similar to the moment generating function, cumulants are generated via,

$$\langle K^V \rangle = \frac{d^v}{dt^v} \ln[M(Et)]|_{t=0}, \tag{76}$$

and they correspond to the connected Feynman diagrams that are important in higher-order scattering processes.

1.13 Normal Distribution (Gaussian)

Normal distribution (Gaussian) is one of the most important distributions in physics. The normal distribution is given by,

$$p(\mu, \sigma) \sim \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right), \tag{77}$$

where μ, σ are the mean and standard deviation. The integral,

$$\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right) dx = \sqrt{\frac{2\pi}{\sigma^2}}. \tag{78}$$

Let's consider the N-dimensional integral,

$$\int_{-\infty}^{\infty} \prod_{i=1}^{N} dv_i \exp\left(-\frac{1}{2}\mathbf{v}^\dagger \mathbf{A}\mathbf{v} + \mathbf{v}^\dagger \mathbf{w}\right) = (2\pi)^{\frac{N}{2}} \sqrt{\det\left(\mathbf{A}^{-1}\right)} \exp\left(\mathbf{w}^\dagger \mathbf{A}^{-1}\mathbf{w}\right). \tag{79}$$

1.14 Central Limit Theorem (CLT)

The CLT states that when a sufficiently large number of independent and identically distributed (i.i.d.) random variables, $X_1, X_2, \ldots X_N$ from a given (unknown) population, as the sample szie $N \to \infty$, the distribution of the sample mean (also the sum) $\bar{X} = \sum_j X_j/N$ will approximate a normal (Gaussian) distribution, regardless of the original population's distribution,

$$f(\bar{X}) \sim \frac{1}{\sqrt{2\pi\sigma^2/N}} \mathrm{Exp}\left(-\frac{(\bar{X}-\mu)^2}{2\sigma^2/N}\right), \tag{80}$$

where $f(\bar{X})$ is the probability density function (pdf) of the sample mean, μ is the mean of each X_j, and σ^2 is the variance of each X_j. Remember, the variance adds as quadrature. Therefore, an extra N in the denominator. We should also note that it only applies to those distributions who have well-defined mean and variance, (e.g. Cauchy distribution).

CLT has important applications in physics. For example, let's say we want to track the how far and how fast on average a particle move over time T. We record the random variable X_{t_i} (relative position with respect to $X_{t_{i-1}}$) at each time interval t_i,

$$\bar{X}_T = \sum_j X_{t_i}, \quad \bar{V}_T = \frac{\bar{X}_T}{T}. \tag{81}$$

The CLT states that as $T \to \infty$, the distribution of both $\bar{X}_T$ and $\bar{V}_T$ (average position and velocity of the particle) will approach a normal distribution. This phenomenon is known as Brownian motion, a random process governed by random walk, and the corresponding evolutions in the phase space density will follow the deterministic diffusion equation.

It is worth noting the distinction between the solution of the diffusion equation and CLT. The diffusion equation is derived from the principle of conservation of particle number and provides information about the local density of particles at a specific time. In contrast, the random variable used in the CLT characterizes the (average) positions of a single particle. These two concepts, while related, serve distinct purposes in modeling random processes. Although we will not discuss in details, Ergodicity is a concept that relates the behavior of a single-particle time average to many-particle ensemble average. In an ergodic system, the behavior of a single trajectory (e.g., the motion of a single particle) over time is representative of the behavior of the entire ensemble of particles.