Advanced Statistical Mechanics
Part III Physics, Unversity of Cambridge

Lawrence Wu

February 13, 2023

Contents

Chapter 1

Lecture 1: Revision of equilibrium stochastic physics

1.1 Prerequisite: Canonical coordinates and Phase space

In mathematics and classical mechanics, canonical coordinates are sets of coordinates on phase space which can be used to describe a physical system at any given point in time. In dynamical system theory, a phase space is a space in which all possible states of a system are represented, with each possible state corresponding to one unique point in the phase space.

1.2 Micro-canonical ensemble

In statistical mechanics, the microcanonical ensemble is a statistical ensemble that represents the possible states of a mechanical system whose total energy is exactly specified. The system is assumed to be isolated in the sense that it cannot exchange energy or particles with its environment, so that (by conservation of energy) the energy of the system does not change with time. In the Micro-canonical ensemble, the Hamiltonian for the system is usually defined as a function of the canonical coordinates momentum p and position q:

$$\mathcal{H} = \mathcal{H}(p, q) \tag{1.1}$$

which allows the different states of the system to be represented in phase space $\mathbb{P}$ and $\mathbb{Q}$. The energy of system E would be the weighted mean of $\mathcal{H}$ in phase space. Rather than solving for the behaviour of a particular set of initial conditions, let us hypothesise that the energy is all we need to describe the equilibrium state. This leads us to a statistical mechanical description of the equilibrium state of our system as an ensemble of all possible states with energy E—the microcanonical ensemble. Now consider the scenario where, for a given E, we introduce a small positive change dE in energy and the volume of the allowed states in phase space is changed. We define:

$$\Omega(E)\, dE = \int d\mathbb{P}d\mathbb{Q} \quad \text{(for states with } E < \mathcal{H}(p, q) < E + dE) \tag{1.2}$$

$$= \int d\mathbb{P}d\mathbb{Q}\ \Theta(E + dE - H) - \Theta(E - H) \quad (\Theta \text{ is the Heaviside step function}) \tag{1.3}$$

$$= dE \cdot \int d\mathbb{P}d\mathbb{Q}\ \left[\frac{\partial \Theta(\mathrm{E} - H)}{\partial E} \right] \tag{1.4}$$

$$\tag{1.5}$$

and we reach the conclusion:

> **Definition 1.2.1: Number of states with hamiltonian $\mathcal{H}$ equal to given energy E.**
>
> $$\Rightarrow \Omega(E) = \int d\mathbb{P}d\mathbb{Q}\ \delta(E - \mathcal{H}(\boldsymbol{p}, \boldsymbol{q})$$
>
>)

We then define the "entropy" S to be:

> **Definition 1.2.2: Entropy**
>
> $$S(E) = k_B \ln \Omega(E)$$

Now let's talk about probability $\rho(E)$. But how do we define it and how is it useful? Let's consider a quantity of the system $X(\boldsymbol{p}, \boldsymbol{q})$ and we're trying to deduce its expectation value $\langle X \rangle$. We'll want to integrate over all states with the constraint $\mathcal{H} = E$:

$$\langle X(\boldsymbol{p}, \boldsymbol{q}) \rangle = \frac{1}{\Omega(E)} \int d\mathbb{P}d\mathbb{Q}\ X(\boldsymbol{p}, \boldsymbol{q})\delta(E - \mathcal{H}(\boldsymbol{p}, \boldsymbol{q})) \tag{1.6}$$

where we can read out the normalised probability $\rho(E)$ to be:

> **Definition 1.2.3: Normalised probability**
>
> $$\rho(E) = \frac{1}{\Omega(E)}\delta(E - \mathcal{H}(\boldsymbol{p}, \boldsymbol{q}))$$

There is a useful spooky relation when it comes to relating Ωs for different energies:

> **Question 1**
>
> Check that for a system consisting of two subsystems with total energy E:
>
> $$\Omega(E) = \int \Omega_1(E_1)\,\Omega_2(E - E_1)\,dE_1$$
>
> as should be intuitively clear

1.3   Canonical ensemble

In comparison with the micro-canonical ensemble, a canonical ensemble is a statistical ensemble that represents the possible states of a mechanical system in thermal equilibrium with a heat bath at a fixed temperature. The system can exchange energy with the heat bath so that the states of the system will differ in total energy. Since the energy is not fixed, we assign a probability P to each distinct state that's related to their energy:

$$P \propto e^{-\beta E} \tag{1.7}$$

and for P to be normalised, we define the normalisation factor Z aka the partition function to be:

Where we often call the value in the exponent the "Free energy" of a microstate.

After normalisation, our normalised probability becomes:

From the relation above, we can see that the probability is maximised when the free energy F is minimised. And for a large system, this gives the equilibrium state due to the central limit theorem. (We will go further into this in the later lectures.)

In conclusion, minimising the free energy allow us to find the balance between the two incompatible demands in statistical physics: 1. Minimising energy E 2. Maximising entropy S(E), and ignoring the fluctuation in the system about this preferred microstate. But in this course, we aim to look at two non-equilibrium questions: 1. How long does it take to reach such an equilibrium state from an initial state? 2. Are there barriers and metastabilities states?

1.4   Different types of systems (Hohenberg-Halperin 1977)

1.4.1   Model A: No conservation laws

An order parameter ϕ is defined to be a variable that can be used to describe a system and gives a complete description of the hamiltonian. In other words, the Hamiltonian can be written in the form $\mathcal{H}(\phi)$. And when there are no conservation laws in a system for this parameter ϕ, we call it a Model A system. For example, the magnetisation of a magnetic system is an order parameter of Model A since it does not have to be conserved at all. For models like this we can make a classical mechanical analogy (Think of ϕ as the cartesian coordinates of a particle):

$$\text{"driving force"} = \text{"friction force"} \tag{1.12}$$

$$-\frac{\partial \mathcal{H}(\phi)}{\partial \phi} = \gamma \frac{\partial \phi}{\partial t} \tag{1.13}$$

And this will be the principal which controls the evolution of the system towards equilibrium. This is in contrast with another type of system that we call Model B:

1.4.2 Model B: Conserved fields

For this type of system, we have a conservation law of the form:

$$\int \phi(\boldsymbol{x})d\boldsymbol{x} = constant \tag{1.14}$$

and for this type of system, since we can only change the distribution of this parameter but not the total quantity, the characteristic equation will be the classical continuity equation:

$$\frac{\partial \phi}{\partial t} = -D\nabla^2 \frac{\partial \mathcal{H}(\phi)}{\partial \phi} \tag{1.15}$$

where the flux J is:

$$J = -D\nabla \frac{\partial \mathcal{H}}{\partial \phi} \tag{1.16}$$

1.4.3 Model C-F...

Model C describes a field ϕ that is coupled to a conserved auxiliary field and such analysis goes down all the way to Model F. There's no need to know all of them and the above sections are just raising awareness that these mechanical analysis exists.

Chapter 2

Lecture 2: Basic definitions of probability theory

2.1 Random Variables and Probabilty

As the title suggests, this lecture covers the basic of probability theory so there is a common language that we can use to describe stochastic systems. Let's start with the definition of "Random variables" or "Stochastic variables". A random variable x with a microscopic probability $P(x)$ has the definitions:

Definition 2.1.1: EV and Variance

$$\text{Expectation value: } \langle x \rangle = \int x \cdot P(x)dx \tag{2.1}$$

$$\text{Variance: } \sigma^2 = \langle x^2 \rangle - \langle x \rangle^2 \tag{2.2}$$

We then introduce the concept of characteristic functions $\phi_x(k)$ that most of you probably have not seen before, which is the expectation value of e^{ikx}:

Definition 2.1.2: Characteristic function

$$\phi_x(k) = \langle e^{ikx} \rangle_{P(x)} = \int e^{ikx} P(x)dx \tag{2.3}$$

---

where we can see is just simply the Fourier Transform of $P(x)$! But if we expand the exponential in terms of the Taylor series:

$$\phi_x(k) = \int e^{ikx} P(x)dx \tag{2.4}$$

$$= \int \sum_{n=0}^{\infty} \frac{1}{n!}(ikx)^n P(x)dx \tag{2.5}$$

$$= \sum_{n=0}^{\infty} \frac{(ik)^n}{n!} \langle x^n \rangle \tag{2.6}$$

$$\tag{2.7}$$

where we call $\langle x^n \rangle$ the nth moment of the probability $P(X)$.

> **Definition 2.1.3: Moments of P(x)**
>
> $$n^{th}\ moment : \langle x^n \rangle = \int x^n P(x)dx \tag{2.8}$$

We can also express the characteristic equation in terms of an exponential of a sum:

$$\phi_x(k) = \exp\left[\sum_{n=0}^{\infty} \frac{(ik)^n}{n!} C_n\right] \tag{2.9}$$

where we call the coefficient C_n the cumulants of P(x) and we can calculate them directly from a probability distribution $P(x)$:

> **Definition 2.1.4: Cumulants**
>
> $$C_n = \frac{d^n}{d(ik)^n} \ln \phi_x(k)\ |_{k=0} \tag{2.10}$$

> **Example 2.1.1** (C_1)
>
> $$C_1 = \frac{d}{d(ik)} \ln\left(\sum_{n=0}^{\infty} \frac{(ik)^n}{n!} \langle x^n \rangle\right)_{k=0} \tag{2.11}$$
>
> $$= \left[\frac{1}{i} \frac{i \sum_m \frac{1}{(m-1)!}(ik)^{m-1} \langle x^m \rangle}{\sum_n (ik)^n/m! \langle x^n \rangle}\right]_{k=0} \tag{2.12}$$
>
> $$= \langle x \rangle \tag{2.13}$$

> **Question 2:** C_2
>
> Check that $C_2 = \sigma^2 = \langle x^2 \rangle - \langle x \rangle^2$

due to the Fourier transform nature of the characteristic function, we can see that if $P(x)$ is a Gaussian, then so is $\phi_x(k)$:

$$P = exp\left(-\frac{(x-x_0)^2}{2\sigma^2}\right) \Leftrightarrow \phi_x(k) = e\left(-\frac{\sigma^2 k^2}{2} + ikx_0\right) \tag{2.14}$$

2.2 Random Walk

Let's now consider a linked sequence y_N of random variables x_i with identical probability distributions $P(x)$ (For example the length of a polymer chain with N identical links randomly folded), we call this a "random walk" or a "stochastic process":

Definition 2.2.1: Random walk

$$y_N = \sum_{i=1}^{N} x_i \tag{2.15}$$

$$\langle y_N \rangle = N \langle x \rangle \tag{2.16}$$

$$\sigma_y^2 = \langle y_N^2 \rangle - \langle y_N \rangle^2 = N\sigma^2 \tag{2.17}$$

$$\tag{2.18}$$

where the variance relationship came from:

$$\langle y^2 \rangle - \langle y \rangle^2 = \sum_i \sum_j \langle x_i x_j \rangle - \sum_i \langle x_i \rangle \sum_j \langle x_j \rangle \tag{2.19}$$

$$= N \langle x^2 \rangle + N(N-1)\langle x \rangle^2 - N^2 \langle x \rangle^2 \tag{2.20}$$

$$= N \left(\langle x^2 \rangle - \langle x \rangle^2 \right) = N\sigma^2 \tag{2.21}$$

Now we introduce the normalised random walk $S_N = y_N/N$ where $\langle S_N \rangle = \langle x_N \rangle$ and $\sigma_S^2 = \sigma_x^2/N$, and the central limit theorem suggests that no matter what $P(x)$ is, $P(S_N)$ will always converge to a Gaussian:

Theorem 2.2.1 Central Limit theorem

For a normalised random walk S_N where:

$$S_N = \frac{y_N}{N} \tag{2.22}$$

$$\langle S_N \rangle = \langle x \rangle \tag{2.23}$$

$$\sigma_S^2 = \frac{\sigma_x^2}{N} \tag{2.24}$$

$$\tag{2.25}$$

The probability distribution for S_N: $P(S)$ converges to a Gaussian with the same EV and Variance for large N.

Proof: (This is a very rough proof)
We start with the characteristic function $\phi_S(k)$:

$$\phi_s(k) = \langle e^{iks} \rangle = \left\langle exp\left(i\frac{k}{N} \sum_m x_m \right) \right\rangle \tag{2.26}$$

$$= \left\langle e^{\frac{ik}{N}x} \right\rangle_{P(x)}^N = \left[\phi_x(\frac{k}{N}) \right]^N \tag{2.27}$$

On the other hand, we can also express $\phi_S(k)$ and $\phi_x(k)$ as:

$$\phi_s(k) = \exp\left[\sum_{m=0}^{\infty} \frac{(ik)^m}{m!} C_m(s) \right] \tag{2.28}$$

$$= \left[\phi_x(k/N) \right]^N \tag{2.29}$$

$$= \left[\exp\left(\sum_{m=0}^{N} \frac{(ik)^m}{m!N^m} C_m(x) \right) \right]^N \tag{2.30}$$

$$= \exp\left(\sum_{m=0}^{N} \frac{(ik)^m}{m!N^{m-1}} C_m(x) \right) \tag{2.31}$$

From here, it is not hard to convince yourself that as N increases the terms in the sum with $m > 1$ will slowly disappear. So if we consider only the first two terms in the sum:

$$\phi_S(k) \approx \exp\left(N + ikC_1(x) - \frac{k^2}{2N}C_2(x)\right) \tag{2.32}$$

$$= \exp\left(N + ik\langle x\rangle - \frac{k^2}{2N}\sigma_x^2\right) \tag{2.33}$$

This is just the Fourier Transform of a Gaussian with the mean and variance mentioned above! Therefore, FTing it back to $P(S)$ gives us our desired result. ☺

2.3 Random walk in time

If we replace N in the random walk with time t, we reach a stochastic process $y(t)$ where we're randomly drifting around over continuous time. In this section, we go through a few key properties of this stochastic process.

2.3.1 probabilities

We can then define a set of probabilities:

> **Definition 2.3.1: Probabilities for random walks in time**
>
> $$P_1(y,t) : the\ probability\ of\ being\ at\ value\ y\ at\ time\ t \tag{2.34}$$
> $$P_2(y_1,y_2|t_1,t_2) : probability\ of\ being\ at\ (y_1,t_1)\ and\ then\ (y_2,t_2) \tag{2.35}$$
> $$\vdots\quad and\ so\ on \tag{2.36}$$

2.3.2 Normalisation

These probabilities are normalised by:

$$1 = \int dy_1 \ldots dy_N\ P_N\left(y_1, y_2 \ldots y_N \mid t_1, t_2 \ldots t_N\right) \tag{2.37}$$

$$\tag{2.38}$$

which is saying if we consider all possible y_i at the given times $t = t_i$, we should have a total probability of 1.

> **Example 2.3.1**
>
> $$1 = \int dy P_1(y,t) \tag{2.39}$$

2.3.3 Reduction

> **Theorem 2.3.1** Reduction
>
> The probability P_N can be reduced into P_{N-1} by reduction:
>
> $$\int dy_N P_N\left(y_1 y_2 \ldots y_N \mid \cdots\right) = P_{N-1}\left(y_1, \cdots y_{N-1}\right) \tag{2.40}$$

$$\int P_2 \left(y_1 y_2 \mid t_1 t_2 \right) dy_2 = P_1 \left(y_1 \mid t_1 \right) \tag{2.41}$$

2.3.4 Correlation functions

Correlation functions are defined in the following ways:

Definition 2.3.2: Correlation functions

$$\langle y(t) \rangle = \int y \cdot P(y, t) dy \tag{2.42}$$

$$\langle y_1 \left(t_1 \right) y_2 \left(t_2 \right) \rangle = \int y_1 y_2 p_2 \left(y_1 y_2 \mid t_1 t_2 \right) dy_1 dy_2 \tag{2.43}$$

$$etc \cdots \tag{2.44}$$

2.3.5 Stationary process

A stationary process is defined when $P_1(y, t)$ is invariant under time shift, in other words, not time dependent. For such process, the correlation function $\langle y(t) \rangle$ has no time dependence and the second correlation $\langle y_1 \left(t_1 \right) y_2 \left(t_2 \right) \rangle$ will only depend on $\Delta t = t_1 - t_2$.

2.3.6 Conditional probability

For the case where $t_2 > t_1$ it is very intuitive to think the probability $\langle y_1 \left(t_1 \right) y_2 \left(t_2 \right) \rangle$ can be written in terms of the product of $P_1(y_1 | t_1)$ and some other function G, where G is the conditional probability of $y = y_2$ when $t = t_2$ given that $y = y_1$ when $t = t_1$.

Definition 2.3.3: The propagator G

$$P_2 \left(y_2 y_1 \mid t_2 t_1 \right) = G(2 \mid 1) P_1 \left(y_1 \mid t_2 \right) \tag{2.45}$$

2.3.7 Markov Process

A Markov process is a stochastic process in which the probability of each event depends only on the state attained in the previous event, or we say it's "memoryless". In our case, the probability of all P_i depends only on the single probability P_1 and our propagator G since the process has no "memory" of what happened before. Therefore if we want to calculate for example P_4, all we need is P_1 and then apply the propagator 3 more times.

Theorem 2.3.2 Markov Process

Any higher level probability P_i can be constructed by using just the single probability P_1 and the propagator G.

2.3.8 Evolution relation

If we consider the setting mentioned in the propagator section, we can actually combine the reduction relation Eq.(2.40) with the propagator Eq.(2.45) to reach something called the evolution relation:

$$P_1(y_2 t_2) = \int G(y_2 y_1 \mid t_2 t_1) P_1(y_1 t_1) \, dy_1 \tag{2.46}$$

2.3.9 Kolmogorov-Chapman relation

The last relation we're gonna talk about in this section is the Kolmogorov-Chapman relation. Now imagine we start with state 0 (y_0, t_0), go through a transition state 1 (y_1, t_1) at a given time $t = t_1$ and end up in state 2 (y_2, t_2). The propagator between state 0 and state 2 can be expressed by:

$$G(y_2, y_0 \mid t_2, t_0) = \int G(y_2, y_1 \mid t_2, t_1) G(y_1, y_0 \mid t_1, t_0) \, dy_1 \tag{2.47}$$

Chapter 3

Lecture 3:The Poisson Process

As we mentioned in the last chapter, a stationary Markov process is a process where 1. its higher probabilities P_i can all be derived from the single probability P_1 and the propagator G 2. The probabilities are all invariant with respect to time shifts. Two stationary Markov processes are particularly important - the Poisson Process and the Wiener Process:

1. Poisson Process:

    - Independent steps
    - Only forward steps
    - Either go forward or not move after each step.

2. Wiener Process:

    - Independent steps
    - Can go both forward and backward
    - Must go either forward or backwards after each step, must move!

3.1   Poisson Process

Let's give some examples so we can have a clearer understanding. Imagine shooting a target with a pistol, let's call every time a bullet is fired a "step". So the number of bullets that hit the target is a Poisson Process since after each step the number can only increase by one (go forward 1 step) or not change if we miss (not move).

Formulating the above example in maths, we have a status of the system $N(t)$ and a parameter ν which is the rate of a single forward step (N=n-1 to n) to happen (or else we don't move). The probability P_+ to make a step in time dt is $P_+ = \nu \cdot dt$. And $N(t)$ is our stochastic process. And the most common answers that we're looking for when using this model are: 1. How far forward would we go in time t? 2. How much time would it take for us to reach a certain point n?

Definition 3.1.1: Poisson Process

1. Status of the system (integer): $N(t)$

2. Rate of going forward: ν

3. Probability to go forward a step in time dt: $P_+ = \nu \cdot dt$

If the system is in the state $N(t) = n - 1$ and we look into this single step going from (n-1) to (n), we can ask two questions: 1. What is the total probability $W(t)$ that the system makes this step after time t? 2.

Equivalently, what is the "survival probability" $S(t) = 1 - W(t)$ that the system still remains in the state $N = n - 1$ after time t? To obtain $S(t)$, we look at the relation between $S(t)$ and $S(t + dt)$:

$$S(t + dt) = S(t) \cdot [1 - vdt] \tag{3.1}$$

$$\frac{dS}{dt} = -v \cdot S \tag{3.2}$$

by using $S(t = 0) = 1$ we get:

$$S(t) = e^{-vt} \tag{3.3}$$

Definition 3.1.2: Survival Probability

$$S(t) = e^{-vt}$$

We can then get the average time it takes for the process to make this single step $\langle t_1 \rangle$ by using $W(t)$ and the probability density $w(t)$:

First define the probability density $w(t)$ of making the step:

$$w(t) = \frac{dW(t)}{dt} = -\frac{\partial S}{\partial t} = ve^{-vt} \tag{3.4}$$

Then the average time it takes to go forward this single step would be:

$$\langle t_1 \rangle = \int t \cdot w(t)dt = \frac{1}{v} \tag{3.5}$$

and with a similar method, we can also get:

$$\sigma_t^2 = \langle t_1^2 \rangle - \langle t_1 \rangle^2 = \frac{1}{v^2} \tag{3.6}$$

Question 3: Simulation excercise

We are constantly flipping a biased coin and the stochastic process is the total number of heads obtained. The constant step time Δt is the time it takes for each flip and the biased probability of flipping a head from the coin is p_+, we can express it in terms of the usual Poisson process rates:

$$p_+ = v\Delta t \quad \text{and} \quad p_0 = 1 - v\Delta t \tag{3.7}$$

What's the probability $P(k, N)$ of having exactly k heads after these N steps (or after time $T = N\Delta t$)? (k is just like our position in a Poisson process, and $N = T/\Delta t$ is effectively a type of expression for total time T.)

We can see trivially that this is a binomial process without the need for simulation where:

$$p(k, N) = \frac{N!}{k!(N - b)!} \cdot p_+^k \cdot (1 - p_+)^{N-k} \tag{3.8}$$

The interesting thing is that if you run a simulation with large enough N (small enough time step), p(k,N) reduces to:

$$p(k, \lambda) = \frac{\lambda^k}{k!}e^{-\lambda} \tag{3.9}$$

Where $\lambda = Np_+ = vT$ is the expectation value of heads in the overall session. This is exactly the Poisson distribution!

3.2 The Poisson distribution

But how did we reach this result? Let's try to derive it using a standard Poisson process with external parameters position and time (n,t) and the rate ν. First, let's start with the probability of being at position n at time t+dt:

$$P(n, t + dt) = P(n, t) \cdot [1 - \nu dt] + P(n - 1, t) \cdot \nu dt \tag{3.11}$$

The first term on the RHS represents the process where it was already at position n and carry on staying at the same spot after this small time interval, and the second term corresponds to it being at position n-1 and hopping forward into our desired state n in the small time interval. Using this relation, we can find:

$$\frac{dP(n, t)}{dt} = [P(n, t + dt) - P(n, t)] / dt \tag{3.12}$$

$$= \nu \cdot [P(n - 1, t) - P(n, t)] \tag{3.13}$$

In the square brackets of the second line, the first term and second terms can be considered as the "flux in" and "flux out" of the position n. This is often called the "Master Equation". Recall from the previous chapter the characteristic function $\phi_x(k) = \langle e^{ikx} \rangle$, we can define a "generating function" $g(k, t)$ using a similar logic:

$$g(k, t) = \sum_{n=0}^{\infty} p(n, t) k^n \tag{3.14}$$

If we differentiate it in respect of t:

$$\frac{\partial g(k, t)}{\partial t} = \sum_{n=0}^{\infty} k^n \cdot v[p(n - 1, t) - p(n, t)] \tag{3.15}$$

and reindex n into m-1 and regroup the terms

$$= \sum_{m=0}^{\infty} v \left(k^{m+1} - k^m \right) P(m, t) \tag{3.16}$$

$$= v(k - 1) \sum_{n=0}^{\infty} k^n P(n, t) \tag{3.17}$$

we can see the sum is the same as $g(k, t)$! So we have:

$$\frac{\partial g}{\partial t} = v(k - 1) \cdot g \tag{3.18}$$

By using the following initial conditions, we obtain the exponential form of g:

$$P(n, t = 0) = \delta_{n,0} \Rightarrow g(t = 0) = 1 \tag{3.19}$$

$$g(k, t) = e^{(k-1)vt} \tag{3.20}$$

Now expand the exponential and compare it with the definition of g (Eq.3.14):

$$g(k,t) = e^{-vt} \sum_{n=0}^{\infty} \frac{1}{n!} (vt)^n \cdot k^n \tag{3.21}$$

$$= \sum_{n=0}^{\infty} p(n,t) k^n \tag{3.22}$$

We reach our conclusion:

$$P(n,t) = \frac{1}{n!} (vt)^n e^{-vt} \tag{3.23}$$

Which is the same as Eq. 3.10. We can then ask some simple questions about the expected values or probability of the process:

> **Example 3.2.1**
>
> What is the expectation value of the occurring time of the 10th event?

Solution: Considering each step is independent, the answer is simply:

$$\langle t_{10} \rangle = 10 \langle t_1 \rangle = 10/v \tag{3.24}$$

> **Example 3.2.2**
>
> What is the probability that the 10th event occurs after time t?

Solution:
There are two ways of doing this: 1. Consider the probability $P(t_{10} > t)$: Which we want to integrate the probability that the 10th event happens at exactly $t = t'$ over $t' = t \sim \infty$:

$$P(t_{10} > t) = \int_t^{\infty} P(t_{10} = t') dt' \tag{3.25}$$

And we write this probability as a product of having exactly 9 events and having another one during the time interval dt':

$$= \int_t^{\infty} \frac{(vt')^q}{q!} e^{-vt'} \cdot v dt' \tag{3.26}$$

2. The second way is much simpler, we just consider $P(n(t) < 10)$, which is:

$$P(n(t) < 10) = \sum_{j=0}^{9} \frac{(vt)^j}{j!} e^{-vt} \tag{3.27}$$

3.3 Evolution of the Master Equation

Using the evolution relation (Eq. 2.46), we can write out the evolution relation for any stationary Markov process distribution (not just Poisson):

(The integral has been replaced by a sum since m is discrete)

$$P(n, t + \Delta t) = \sum_m G(n, t + \Delta t \mid m, t) \cdot P(m, t) \tag{3.28}$$

16

And if we subtract $P(n,t)$ from both sides:

$$\frac{\partial P(n,t)}{\partial t} = \frac{1}{\Delta t}\left[\sum_m G(n,t+\Delta t m,t)P(m,t) - P(n,t)\cdot\sum_m G(m,t+\Delta t\mid n,t)\right] \qquad (3.29)$$

where:

$$\sum_m G(m,t+\Delta t\mid n,t) = 1 \qquad (3.30)$$

we get:

$$= \sum_m \frac{G(n,t+\Delta t\mid m,t)}{dt}P(m,t) - \sum_m \frac{G(m,t+\Delta t\mid u,t)}{dt}P(u,t) \qquad (3.31)$$

$$= \sum_m [w_{nm}P(m,t) - w_{mn}P(n,t)] \qquad (3.32)$$

Where $w_f i$ is the transition probabilities between initial state i to final state f.

$$w_{fi} = \text{limit }_{dt\to0}\,\frac{1}{dt}G(j,t+dt\mid i,t) \qquad (3.33)$$

If we are considering the Poisson process, we need to imply the restrictions of $n > m$ and $n < m$ respectively on the two sums, but what we derived here is more general and did not include this restriction.

Chapter 4

Lecture 4: The Wiener process

As the title suggests, we will be looking at the other stationary Markov process that we mentioned at the beginning of the last chapter, the Wiener process.

4.1   The Wiener process

So what is the Wiener process? Let's start with the definition of a random walk (Section 2.2), where you have a random variable x (a single step) with probability density $P(x)$ and a linked sequence y_N, which can be converted into a time-dependent function $y(t)$ with a stepping rate. We call this the Wiener process when the probability distribution of independent step $P(x)$ is the Gaussian probability. And we usually use the expression $W(t)$ to represent a Wiener process (rather than $y(t)$).

> **Definition 4.1.1: Wiener process**
>
> A random walk $W(t)$ where the increment ΔW has the Gaussian probability of variance $\sigma^2 = \Delta t$
>
> $$\Delta W = W(t + \Delta t) - W(t) \tag{4.1}$$
>
> $$P(\Delta w) = \frac{1}{\sqrt{2\pi \Delta t}} \exp\left(-\frac{\Delta w^2}{2\Delta t}\right) \tag{4.2}$$

And then there are a few important properties of the Wiener process:

4.1.1   Characteristic function

It is easy to check (Fourier Transform) that the characteristic function ϕ of ΔW is:

> **Theorem 4.1.1** Wiener Characteristic Function
>
> $$\phi(k, t_0 + \Delta t) = \left\langle e^{ik(w_0 + \Delta w)} \right\rangle_{P(\Delta w)} \tag{4.3}$$
>
> $$= \exp\left(ikw_0 - \frac{1}{2}k^2 (\Delta t - t_0)\right) \tag{4.4}$$
>
> where $w(t_0) = w_0$.

4.1.2 Scaling

If we speed up (or slow down) the rate of the process by a scaling factor c, this process can be linearly scaled into a Wiener-Process:

> **Theorem 4.1.2** Scaling invariance
>
> $$V(t) = \frac{1}{\sqrt{c}}W(ct) \tag{4.5}$$
>
> is a Wiener process. The scaling factor $1/\sqrt{c}$ came from the exponent in Eq. (4.2) since we need to keep the variance of Δv as Δt. (If time flows faster and we're taking more steps per time, the steps must be smaller to keep the overall variance the same.)

4.1.3 Time inversion

This is a bit of a weird one to be interpreted physically.

> **Theorem 4.1.3** Time inversion invariance
>
> $$V(t) = t \cdot W(1/t) \tag{4.6}$$
>
> is also a Wiener process.

4.1.4 Time reversal

If we reverse the time and relabel $W(t = A)$ to be the new origin, this is still a wiener process:

> **Theorem 4.1.4** Time reversal invariance
>
> $$V(t) = W(A) - W(A - t) \tag{4.7}$$
>
> is a Wiener process.

4.1.5 Probability distribution

Using the propagator $G = P(\Delta w)$ from the definition of a Wiener process:

$$P(w + \Delta w, t + \Delta t) = \int_{-\infty}^{\infty} G(\Delta w, \Delta t)P(w, t)dw \tag{4.8}$$

And using the relation $\Delta w = w(t + \Delta t) - w(t)$, we can change the variables and change the integral $dw = -d\Delta w$ (think $w(t + \Delta t)$ as a constant)

$$= -\int_{\infty}^{-\infty} G(\Delta w, \Delta t)P(w(t + \Delta t) - \Delta w, t)d\Delta w \tag{4.9}$$

flipping the interval of integration back

$$= \int_{-\infty}^{\infty} G(\Delta w, \Delta t)P(w(t + \Delta t) - \Delta w, t)d\Delta w \tag{4.10}$$

taking out the $-\Delta w$ term since it's $<< w(t)$

$$= \int_{-\infty}^{\infty} G(\Delta w, \Delta t) P(w(t + \Delta t), t) dww \tag{4.11}$$

$$- \int G(\Delta w, \Delta t) \Delta w \frac{\partial P}{\partial w} d\Delta w \tag{4.12}$$

$$+ \frac{1}{2} \int G(\Delta w, \Delta t) \Delta w^2 \frac{\partial^2 P}{\partial w} d\Delta w \tag{4.13}$$

The integral in the first line over G is 1 (P doesn't depend on Δw), due to normalisation over all possible Δw. The second line is 0 since it's a product of an even function G and odd function Δw:

$$P(w + \Delta w, t + \Delta t) - P(w(t + \Delta t), t) = \frac{1}{2} \int G(\Delta w, \Delta t) \Delta w^2 d\Delta w \frac{\partial^2 p}{\partial w} \tag{4.14}$$

by using $\Delta w = w(t + \Delta t) - w(t)$ and the definition of variance:

$$\frac{\partial P(w, t)}{\partial t} = \frac{1}{2} \frac{\sigma_{\Delta w}^2}{\Delta t} \cdot \frac{\partial^2 P(w, t)}{\partial w^2} \tag{4.15}$$

wrapping the terms up with D, we reach the diffusion equation:

$$\frac{\partial P(w, t)}{\partial t} = D \frac{\partial^2 P(w, t)}{\partial w^2} \tag{4.16}$$

From the derivation above, we conclude that the diffusion equation (4.16) is how $P(w, t)$ evolves over time. There is no universal solution to this equation since it depends on the initial boundary condition $P(w, t = 0)$. Although there is a special case where we start from a single point w_0, setting the boundary condition $P(w, t = 0)$ to be $\delta(w = w_0)$. We get a Gaussian probability distribution:

Theorem 4.1.5 Probabilty distribution of W(t)

$$P(w, t) = w_0 + \frac{1}{\sqrt{2\pi t}} e^{-x^2/(2t)} \tag{4.17}$$

with variance $\sigma_w^2 = t$.

It's easy to check that Eq. (4.17) is a solution of the diffusion differential equation and converges to a delta function $P(w) = \delta(w - w_0)$ at t=0. (Plug it in yourself!) Also, by using the variance $\sigma_{\Delta w}^2$ of Δt and Eq. (2.17), it's not hard to be convinced after time t ($N = t/\Delta t$ steps), the probability distribution of W(t) : $P(w, t)$ will still a Gaussian distribution if we start from a single point.

The equation for the global functions (such as the diffusion equation for the global probability distribution $P(w, t)$) are called "kinetic" or Macroscopic. And on a microscopic scale of $W(t)$, they're governed by "Stochastic differential equations", SDE.

4.2   Stochastic differential equations

Here's a fun little fact. With so many "Physics-inspired" algorithms/models nowadays, the solution of Brownian motion was actually "trading-inspired". In 1900, Louis Bachelier presented the method of stochastic analysis of stock and option markets in his PhD thesis "The theory of speculation". Then in 1905 Albert Einstein and Marian Smoluchowski brought the solution of Brownian motion to the attention of physicists and presented it as a way to indirectly confirm the existence of atoms and molecules. Later in 1911, Langevin proposed a more general approach to formulating and analysing SDEs.

Let's start with something that we're all familiar with, the stochastic differential equation for Brownian Motion:

$$m\frac{du}{dt} = -\gamma u + \xi(t) \tag{4.18}$$

where u is the velocity of the particle and ξ is a stochastic force. More generally, any SDE can be written like this:

$$\frac{dx}{dt} = \underbrace{F(x)}_{\text{drift term}} + \underbrace{\sigma(x)\Delta W(t)}_{\text{diffusion term}} \tag{4.19}$$

where W is the normalised Wiener process, and $\sigma(x)$ is the amplitude which may depend on x. The reason why the second term is called the diffusion term is that if we have only the second term, the differential equation can be written in the diffusion equation form with diffusion equation $D = \frac{1}{2}\sigma^2$.

Mathematically (mathematician uses this notation because they don't trust the physicists with maths), the same SDE above can be written as:

$$dx_t = \mu(x,t)dt + \omega(x,t)dw_t \tag{4.20}$$

since we have a wiener term dw_t, if the steps dt is small enough, we expect dx_t to be Gaussian as well (have gaussian probability distribution).

4.3 Example: Geometric Brownian Motion

The difference between geometric and typical Brownian motion is that for each step Δt, the value of the motion is multiplied by the step x rather than added. Just like a geometric vs an arithmetic series, hence the "geometric" in the name.

> **Definition 4.3.1: Geometric Brownian Motion**
>
> A stochastic process such that:
>
> $$dS_t = \mu \cdot S_t \cdot dt + \sigma \cdot S_t dw_t \tag{4.21}$$

which is also known as the "percentage drift" or the "percentage volatility".

To solve this we rearrange the equation above into:

$$\frac{dS_t}{S_t} = \mu \cdot dt + c \cdot dw_t \tag{4.22}$$

which you will very intuitively write into:

$$d\ln S_t = \cancel{d(\mu t + \sigma w_t)} \quad \text{Wrong!} \tag{4.23}$$

But this second line above is actually WRONG!!! $\int \sigma dw_t$ is not actually σw_t! The problem stems from the cheeky term $\int dw_t$. Although in our notation we wrote the infinitesimal interval dw_t in the integral, the function w_t is not actually a smooth continuous differentiable function (The mathematicians lied to us rip). The solution that we deal with this has a name called "Itoh Calculus".

4.4 Itoh Lemma

In this section, it's all mathematician BS, so we used some big words to blend in. So the Itoh lemma was formed to solve the following question:

If we have a stochastic process $x(t)$ with SDE:

$$dx_t = \mu dt + \sigma dw_t \tag{4.24}$$

and a fucntion $f(x)$ that depends on x. What is the SDE for f?

First we start with the definition of df:

$$df = f(x + dx) - f(x) \tag{4.25}$$

which the terms may all have explicit time dependence. We Taylor expand them:

$$= \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial x}dx + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}dx^2 + \cdots\cdots \tag{4.26}$$

then using $dx_t = \mu(x,t)dt + \omega(x,t)dw_t$ from Eq(4.20) and only keeping terms linear in dt:

(Note that $\langle dw_t^2 \rangle = dt$ for fast steps with small time interval)

$$= \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial x}(\mu\delta t + \sigma dw) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}\left(\mu^2 \cancelto{0}{dt^2} + 2\mu\sigma\cancelto{0}{dtdw_t} + \sigma^2 dw_t^2\right) \tag{4.27}$$

Now:

$$df = \left(\frac{\partial f}{\partial t} + \mu\frac{\partial f}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2 f}{\partial x^2}\right)\cdot dt + \underbrace{\sigma\frac{\partial f}{\partial x}\cdot dw_t}_{\text{diffusion term}} \tag{4.28}$$

Lenma 4.4.1 Itoh Lemma

If we have a stochastic process $x(t)$ with SDE:

$$dx_t = \mu dt + \sigma dw_t \tag{4.29}$$

and a fucntion $f(x)$ that depends on x. The SDE for f is:

$$df = \left(\frac{\partial f}{\partial t} + \mu\frac{\partial f}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2 f}{\partial x^2}\right)\cdot dt + \sigma\frac{\partial f}{\partial x}\cdot dw_t \tag{4.30}$$

Now applying the same analysis to Geometric Brownian Motion Eq.(4.21) and consider the function $f(S) = \ln S_t$:

$$d(\ln S) = \frac{\partial f}{\partial S}dS + \frac{1}{2}\frac{\partial^2 f}{\partial S^2}dS^2 \tag{4.31}$$

$$= \frac{dS}{S} - \frac{1}{2S^2}dS^2 \tag{4.32}$$

$$= \frac{dS}{S} - \frac{1}{2S^2}\left(\mu^2 S^2\cancelto{0}{dt^2} + 2\mu\sigma S^2\cancelto{0}{dtdw} + \sigma^2 S^2 dw^2\right) \tag{4.33}$$

$$= \frac{dS}{S} - \frac{\sigma^2}{2}dt \tag{4.34}$$

which gives:

$$S = S_0 \exp\left(\left(u - \frac{\sigma^2}{2}\right)t + \sigma w_t\right) \tag{4.35}$$

> **Definition 4.4.1: Solution of Geometric Brownian Motion**
>
> $$S(t) = S_0 \exp\left(\left(u - \frac{\sigma^2}{2}\right)t + \sigma w_t\right) \tag{4.36}$$

From this we can see two interesting things:

1. Since dw_t is a small step Wiener process, we have a diffusive growth $e^{\sigma w_t}$

2. If we solved the differential equation without the diffusion term we would've reached the solution:

$$dS = \mu s\, dt \rightarrow S = S_0 e^{\mu t} \tag{4.37}$$

   But instead we have a shift in the deterministic drift term, giving us a lower growth rate:

$$S \propto \exp\left(\mu - \frac{\sigma^2}{2}\right)t \tag{4.38}$$

Try using the code on the GBM Wikipedia page to produce GBM graphs in Python!

Chapter 5

Lecture 5

In this lecture, we will be using the two things we introduced in the last lecture, Etoh's lemma and Geometric Brownian Motion, to study the Black-Scholes equation, which won the Nobel prize for economics in 1997. So what is the Black-Scholes equation and how is it useful? It is a mathematical analysis of dynamics and prediction of option prices in volatile markets that are stochastic. The term "option" might not be familiar to some readers, in simple terms, the B-S equation gives investment managers a tool to allocate their risk and cash. For example, buying more stocks is allocating more risk and less cash.

Interestingly, there are theories that suggest that the 1987 financial crisis was caused by a group of traders that conspired to punish the early adopters of the B-S analysis. Even more interestingly, one of the main causes of the 2008 financial crisis is actually too much trust in the B-S analysis and incorrect use of it (Since they didn't pick the ASM option in Lent like you obviously lol).

5.1 Black-Scholes Equation

So first we use a stochastic variable $S(t)$ to present the value of a given stock, which is governed by the SDE of Geometric Brownian Motion (GBM):

$$dS = \mu S dt + \sigma S dw \tag{5.1}$$

since in stocks we usually describe the growth in percentage growth and not linear growth. An "Option" is a contract when you (the seller) agree to sell me (the buyer) a certain stock $\mathcal{S}$, at a future time t, at the current price if the future price satisfies certain conditions (For example reach a trigger price E). So you see the seller hopes the stock to depreciate and then the opposite for the buyer. You might think, why would the seller agree to a such weird contract? They already have the stock in hand and can sell them at any time that they want! The only reason that a seller will be interested in such a contract is if the buyer (me) pays a small extra fee for this. We denote this fee as another stochastic variable $V(s, t)$, which depends on the value of the stock S and time t. A "Portfolio" Π is a combination of assets (cash, stocks, options etc.) that I currently have:

$$\Pi = -V(s_t) + \alpha S \tag{5.2}$$

where the $-V$ came from the amount that you have already paid for the options contract, and αS from selling a portion α of that same stock at the same time to "hedge" aka decrease risk exposure (Just like the hedge in "Hedge fund"). In continuous trading, we have small increments of time, therefore:

$$\Delta \Pi = -\Delta V(s) + \alpha \Delta S \tag{5.3}$$

Now, we can use the Itoh Lemma 4.30:

$$dV(s) = \left(\frac{\partial V}{\partial t} + \frac{\partial V}{\partial s} ds + \frac{1}{2} \frac{\partial^2 V}{\partial s^2} ds^2 \right) \tag{5.4}$$

24

And subbing in:

$$ds = \mu s\,dt + \sigma s\,dw \tag{5.5}$$

We get:

$$dV(s) = \left(\frac{\partial v}{\partial t} + \mu s \frac{\partial v}{\partial s} + \frac{1}{2} 6^2 s^2 \frac{\partial^2 v}{\partial s^2} \right) dt + \sigma s \frac{\partial V}{\partial s} dW \tag{5.6}$$

Then substitute dV and dS into $d\Pi$:

$$d\Pi = -\frac{\partial v}{\partial t} dt - \mu s \frac{\partial v}{\partial s} dt - \frac{1}{2} \sigma^2 s^2 \frac{\partial^2 v}{\partial s^2} dt + \alpha \mu s\,dt \underbrace{- \sigma s \frac{\partial v}{\partial s} dW + \alpha \sigma s\,dW}_{\text{stochastic part}} \tag{5.7}$$

With our stochastic description of $d\Pi$, the key step of reaching the BS equation is to choose the factor α to be $\partial V/\partial S$ so the volatility terms above cancel out and we have no volatility in $d\Pi$. Then, we make a decision on what we want to achieve for $d\Pi$ (The goal of our investments), a steady geometric growth:

$$d\Pi = r \cdot \Pi\,dt \tag{5.8}$$

Getting us the relation with rate r:

$$r \cdot \overbrace{(\alpha S - V)}^{\Pi} dt = \left(-\frac{\partial V}{\partial t} - \cancel{\mu s \frac{\partial V}{\partial s}} - \frac{1}{2} \sigma^2 s^2 \frac{\partial^2 V}{\partial S^2} + \cancel{\alpha \mu s} \right) dt \tag{5.9}$$

Finally we have:

$$\frac{\partial V(s,t)}{\partial t} + \frac{1}{2} \sigma^2 s^2 \frac{\partial^2 V}{\partial s^2} = r \left(V - s \frac{\partial V}{\partial s} \right) \tag{5.10}$$

This is the B-S equation! It allows us to solve for $V(s,t)$ for any given r and σ. The lack of dependence on μ makes it even more powerful since if we look back to our SDE Eq.5.1 for $S(t)$, it is μ that makes the growth of each stock different and the dW term is just geometric stochastic term that represents the market volatility and does not relate to the growth.

What about r? What are the restrictions on this risk-free geometric growth? The "desired legislation" should have this rate of risk-free growth fixed (limited). We will not go into too much detail about the impact of r in the overall economy since we're physicists not econ students. In reality, we use the rate of US Treasury Bonds r_0 (effectively the interest rate you get by lending money to the US government) since they are such a global superpower and the capitalists believe lending money to them is risk-free lol.

5.2 A simple case

Let's consider the case where $r = 0$ and see how the Black-Scholes equation is useful if we were a hedge fund manager back in the 1980s skrskr:

$$\frac{\partial V}{\partial t} + \frac{\sigma^2}{2} S^2 \frac{\partial^2 V}{\partial S^2} = 0 \tag{5.11}$$

Later in the course, you will see that this is actually the "super diffusion" equation (Another great naming from the physicists ofc) since the diffusion constant $D = \frac{\sigma^2}{2} S^2$ grows as S^2, the effective position (non-linear diffusion).

Using this equation, we can solve for the fair value of the contract V at the current time and buy all options that are underpriced and sell all the overpriced ones.

(Note: There's another interesting question to ask: Do we really know the value of volatility σ? To this day, the measurement or modelling of σ is still a very competitive field between hedge funds and a lot of different analyses are still being proposed.)

With all that waffling above, how do we actually solve for $V(s,t)$ using this equation? Using the current price($t = 0$) of a stock S and the trigger price$E$ of the stock at the end of the contract ($t = T$), we define:

$$x = \ln S/E \tag{5.12}$$

$$S = Ee^x \tag{5.13}$$

$$\tau = T - t \tag{5.14}$$

re-express V into:

$$V(Ee^x, T - \tau) = Z(x, \tau) \tag{5.15}$$

we get:

$$\frac{\partial z}{\partial \tau} - \frac{\sigma^2}{2}\frac{\partial^2 z}{\partial x^2} + \left(\frac{\sigma^2}{2} - r\right)\frac{\partial z}{\partial x} + r = 0 \tag{5.16}$$

Which is a homogeneous differential equation! In order to solve this, there's a beautiful solution by introducing a new variable u:

$$u(x, \tau) = z(x, \tau) \cdot e^{\alpha x + \beta \tau} \tag{5.17}$$

plugging it into the PDE above:

$$\frac{\partial u}{\partial \tau} - \frac{\sigma^2}{2}\frac{\partial^2 u}{\partial x^2} + A(\alpha, \beta) \cdot \frac{\partial u}{\partial x} + \underbrace{B(\alpha, \beta)}\cdot u = 0 \tag{5.18}$$

choose an α and β that yields $A = B = 0$: (try it yourself!)

$$\alpha = r/6^2 - 1/2 \tag{5.19}$$

$$\beta = r/2 + 6^2/8 + r^2/26^2 \tag{5.20}$$

and all we have left is:

$$\frac{\partial u}{\partial \tau} - \frac{\sigma^2}{2}\frac{\partial^2 u}{\partial x^2} = 0 \tag{5.21}$$

Amazing right? The last step is to simply solve this by using Green's function and its initial condition:

$$u(x, \tau) = \int \underbrace{\frac{1}{\sqrt{2\pi\sigma^2\tau}} \exp\left(-\frac{(x-w)^2}{2\sigma^2\tau}\right)}_{\text{Green's function}} \cdot \underbrace{u(w, 0)}_{\text{initial condition}} \, dw \tag{5.22}$$

and hence we can backroll and find Z, and $V(s, t)$.