Notes for Lecture 3

Probability

In statistical mechanics, the entropy emerged as the central quantity. The variety of minimization principles of thermodynamic potentials include the mechanical energy minimization that we know and love from mechanics is also the result of the maximum entropy principle.

Now, one might ask – where does this maximum entropy principle come from? The answer is very simple: from the probability. In this and the next lectures, we will learn some technical matters about the probability. In particular, we must keep our senses keen on large numbers.

3.1 Probability

Let us assume that a variable x has a **probability density function** (PDF) $p(x)$. For $p(x)$ to be a valid PDF, we require that $p(x) \geq 0$ (**non-negativity**) and $\int_{x_{min}}^{x_{max}} p(x) = 1$ (**probability sum rule**) where x_{min} and x_{max} are the physical limits of the variable x (e.g., they can be $-\infty$ and ∞, respectively). The variable x is called a **random variable**, in the sense that its outcome is decided by chance[1]. To simplify the notation, we use $\int dx \equiv \int_{x_{min}}^{x_{max}} dx$, below.

[1] Of course, an exception would be when $p(x) = \delta(x - x_0)$ for a certain value of x_0.

3.1.1 Moments

The n**-th moment** m_n of the PDF is defined as

$$m_n \equiv \langle x^n \rangle = \int dx\, p(x) x^n \qquad n = 0, 1, 2, ... \tag{3.1}$$

The zero-th moment is always 1, by definition. The first moment is the mean value of x. Or, the average value. Or, the expectation value. We will discuss the meaning of higher moments shortly.

For a given $p(x)$, there is no guarantee that all moments are well-defined. For example, for a Lorentzian distribution function, $p(x) = \frac{1}{\pi}\frac{\Gamma}{x^2+\Gamma^2}$, all moments for $n \geq 1$ are not defined.

3.1.2 Function of a random variable

For a random variable x, a function $F(x)$ is also a random variable. Its expectation value is given by the following.

$$\langle F(x) \rangle = \int dx\, p(x) F(x) \tag{3.2}$$

3.1.3 First characteristic function

The **(first) characteristic function** $\tilde{p}(k)$ of $p(x)$ is defined as its Fourier transformation.

$$\tilde{p}(k) \equiv \langle p(x) \rangle = \int dx\, p(x) e^{-ikx} \tag{3.3}$$

Or, the characteristic function is the mean value of the function e^{-ikx}. The inverse transformation is thus given by

$$p(x) = \frac{1}{2\pi} \int dk\, \tilde{p}(k) e^{ikx} \tag{3.4}$$

By expanding the exponential in Eq. 3.3, we get

$$\tilde{p}(k) = \sum_{n=0}^{\infty} \frac{1}{n!} \int dx\, p(x)(-ikx)^n \tag{3.5}$$

$$= \sum_{n=0}^{\infty} \frac{(-ik)^n m_n}{n!} \tag{3.6}$$

In this sense, the characteristic function is called the **moment generating function**.

Note that just because the characteristic function may be well-defined does not mean that it can generate moments. For it to generate moments, the characteristic function must be analytic (cf. Section 3.2.4).

3.1.4 Second characteristic function, and cluster expansion

Take the logarithm[2] of the first characteristic function $\log \tilde{p}(k)$: this defines the **second characteristic function**. If we assume that $\tilde{p}(k)$ is analytic, then $\tilde{p}(k) = 1 + \sum_{n=1}^{\infty} \frac{(-ik)^n m_n}{n!}$, using the fact that $m_0 = 1$ by the probability sum rule. So, then it follows that $\log \tilde{p}(k)$ can also be expanded in powers of k^n, using the expansion $\log(1+x) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} x^n$. So, it is clear then that the second characteristic function is also analytic, and its Taylor expansion starts from $n = 1$. We write

$$\log \tilde{p}(k) = \sum_{n=1}^{\infty} \frac{(-ik)^n \langle x^n \rangle_c}{n!} \tag{3.7}$$

which can be taken as *defining* the **n-th cumulant** $\langle x^n \rangle_c$. So, the second characteristic function is the **cumulant generating function**.

To find the relationship between cumulants and moments, it is the most convenient to take the exponential of Eq. 3.7.

$$\begin{aligned}
\tilde{p}(k) &= \exp\left(\log \tilde{p}(k)\right) \\
&= \exp\left(\sum_{n=1}^{\infty} \frac{(-ik)^n \langle x^n \rangle_c}{n!}\right) \\
&= \prod_{n=1}^{\infty} \exp\left(\frac{(-ik)^n \langle x^n \rangle_c}{n!}\right) \\
&= \prod_{n=1}^{\infty} \sum_{p_n=0}^{\infty} \frac{1}{p_n!} \left(\frac{(-ik)^n \langle x^n \rangle_c}{n!}\right)^{p_n}
\end{aligned}$$

By definition (Eq. 3.6), this must be equal to $\sum_{m=0}^{\infty} (-ik)^m \langle x^m \rangle / m!$. By comparing these two expressions, we deduce that

$$\langle x^m \rangle = m! \sum_{\{p_n | \sum_n n p_n = m\}} \prod_n \frac{\langle x^n \rangle_c^{p_n}}{p_n! (n!)^{p_n}}. \tag{3.8}$$

[2] In this note, the convention is to use log for the natural logarithm. The notation ln is not used. If we ever need to use the log with the base of 10, then it can be written as $\log_{10}$.

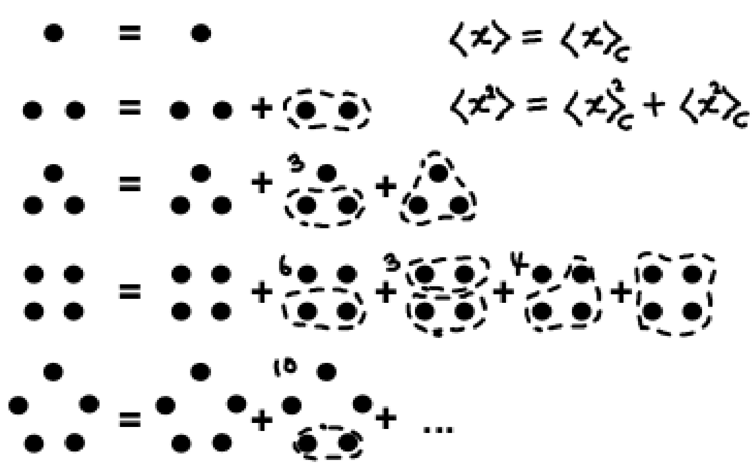
This formula has a simple graphical interpretation. For any m-th moment ($m \geq 1$), we draw m identical dots. Now, consider grouping these points into p_n clusters of size n ($n \geq 1$, $p_n \geq 1$; $p_n = 0$ term can be ignored, since such term means multiplication by 1), which means that $\sum_n np_n = m$. For each such possible grouping, we assign the factor $\prod_n \langle x^n \rangle_c^{p_n}$, i.e., the product of all cumulants, one each from each group. How many different ways can we group the original m dots, given the set $\{p_n | \sum_n np_n = m\}$? It is $\frac{m!}{\prod_n p_n!(n!)^{p_n}}$.

What we have here is an elementary example of the so-called **cluster expansion**.

Namely, to find how $\langle x^m \rangle$ is expanded in terms of $\langle x^m \rangle_c$, all we have to do is to figure out how many different ways we can group m dots. For each way of grouping, we assign a product of cumulants, one n-th cumulant assigned to each group of size n. Then, the desired expansion is simply the sum of all such products, with the multiplicity ($\frac{m!}{\prod_n p_n!(n!)^{p_n}}$) for each way of grouping taken into account (this is easy to do case by case, for these following first few expansions, rather than by remembering the multiplicity formula and applying it).

$$\langle x \rangle = \langle x \rangle_c \tag{3.9}$$
$$\left\langle x^2 \right\rangle = \langle x \rangle_c^2 + \left\langle x^2 \right\rangle_c \tag{3.10}$$
$$\left\langle x^3 \right\rangle = \langle x \rangle_c^3 + 3\left\langle x^2 \right\rangle_c \langle x \rangle_c + \left\langle x^3 \right\rangle_c \tag{3.11}$$
$$\left\langle x^4 \right\rangle = \langle x \rangle_c^4 + 6\left\langle x^2 \right\rangle_c \langle x \rangle_c^2 + 3\left\langle x^2 \right\rangle_c^2 + 4\left\langle x^3 \right\rangle_c + \left\langle x^4 \right\rangle_c \tag{3.12}$$

Here is a graphical representation of the above results. Note that, for the diagrams on the right hand side, if a dot is not enclosed in a group, then it means it forms a group of size 1, giving rise to a contribution of a factor of $\langle x \rangle_c$.



It is easy to invert these identities from the top to obtain

$$\langle x \rangle_c = \langle x \rangle \qquad \textbf{mean} \qquad (3.13)$$
$$\langle x^2 \rangle_c = \langle x^2 \rangle - \langle x \rangle^2 \qquad \textbf{variance} \qquad (3.14)$$
$$\langle x^3 \rangle_c = \langle x^3 \rangle - 3 \langle x^2 \rangle \langle x \rangle + 2 \langle x \rangle^3 \qquad \textbf{skewness} \qquad (3.15)$$
$$\langle x^4 \rangle_c = \langle x^4 \rangle - 4 \langle x^3 \rangle \langle x \rangle - 3 \langle x^2 \rangle^2 + 12 \langle x^2 \rangle \langle x \rangle^2 - 6 \langle x \rangle^4 \qquad \textbf{curtosis} \qquad (3.16)$$

Another symbol for mean that will be often used in this course is $\bar{x}$. Also, note that the standard deviation (square root of the variance) will be denoted as σ (or σ_x), and it follows that

$$\sigma \equiv \sqrt{\langle x^2 \rangle_c} = \sqrt{\langle (x - \bar{x})^2 \rangle} \qquad (3.17)$$

Note that the notation σ will be used for any probability distribution, not just the Gaussian distribution.

It is left for your exercise to show the following identities.

$$\langle x^2 \rangle_c = \left\langle (x - \bar{x})^2 \right\rangle \equiv \sigma^2 \qquad \textbf{variance} \qquad (3.18)$$
$$\langle x^3 \rangle_c = \left\langle (x - \bar{x})^3 \right\rangle \qquad \textbf{skewness} \qquad (3.19)$$
$$\langle x^4 \rangle_c = \left\langle (x - \bar{x})^4 \right\rangle - 3\sigma^4 \qquad \textbf{curtosis} \qquad (3.20)$$

Notice that these definitions of skewness and curtosis are dimensionful. It is often the case that skewness or curtosis is divided by σ^3 or σ^4 respectively, so that they become dimensionless.

3.1.5 Shifted random variable and cumulants

If a random variable x is shifted, $y = x + b$, (or if the PDF is shifted in the opposite direction, equivalently), then what happens to its cumulants? The answer is that

$$\langle y \rangle_c = \langle x \rangle_c + b \qquad (3.21)$$
$$\langle y^n \rangle_c = \langle x^n \rangle_c \qquad n = 2, 3, ... \qquad (3.22)$$

This can be figured out by noting that $\tilde{p}_y(k) = \langle \exp(-ik(x+b)) \rangle = \exp(-ikb)\tilde{p}_x(k)$. Thus, the 2nd characteristic function is given by $\log \tilde{p}_y(k) = -ikb + \log \tilde{p}_x(k)$. By definition, $\log \tilde{p}_y(k) = \sum_{n=1}^{\infty} \frac{(-ik)^n}{n!} \langle y^c \rangle_c$, and similarly for x. We see that the term $-ikb$ simply adds to the first order term, but leave all other terms of $\log \tilde{p}_x(k)$ intact.

This fact is useful, e.g., in understanding the arguments that go into the **central limit theorem** (cf. next lecture). It also means that cumulants of 2nd order or higher are related to the shape of the PDF only, as opposed to its average position.

3.1.6 Scaled random variable and cumulants

If a random variable x is scaled, $y = ax$, then what happens to its cumulants and its moments?

$$\langle y^n \rangle_c = a^n \langle x^n \rangle_c \qquad n = 0, 1, 2, \ldots \tag{3.23}$$

$$\langle y^n \rangle = a^n \langle x^n \rangle \qquad n = 0, 1, 2, \ldots \tag{3.24}$$

The proof is left for your exercise. These results, combined with the result of the previous section, can be used to express all cumulants for $y = ax + b$ in terms of the cumulants for x.

3.2 Some notable probability distributions

3.2.1 Binomial distribution

Consider an event where two outcomes are possible. A coin toss can be an example. We will assume that the coin is somehow rigged so that the chance of head is h ($0 \le h \le 1$) and the chance of tail is $1 - h$, where h is not necessarily 1/2. Here, we rule out any sideways landing event. For N tosses, we can consider the probability for getting n heads. Let us call this probability $P_N(n)$. Clearly, this formalism is applicable to other types of events as well: e.g., a particle decay or a particle detection – for both of these events, h would be the probability for the event to occur during a small preset time window (cf. Section 3.2.3).

$$P_N(n) = \frac{N!}{n!(N-n)!} h^n (1-h)^{N-n} \qquad n = 0, 1, 2, 3, \ldots, N \tag{3.25}$$

So, in the current case, we have a PDF, $p(x)$, which is non-zero at discrete values of n: so $p(x) = \sum_{n=0}^{N} P_N(x)\, \delta(x - n)$, where $\delta(x)$ is the Dirac delta function. If we simply recognize and accept that integrals over x become summations over n, all our previous results apply.

The coefficients above are the well-known binomial coefficients in the binomial expansion formula.

$$(x + y)^N = \sum_{n=0}^{N} \frac{N!}{n!(N-n)!} x^n y^{N-n} \tag{3.26}$$

By plugging in $x = h$ and $y = 1 - h$, we see that the probability sum rule is satisfied indeed.

$$\sum_n P_N(n) = \sum_n \frac{N!}{n!(N-n)!} h^n (1-h)^{N-n} = 1^N = 1 \tag{3.27}$$

What is the first and second characteristic functions for this distribution?

$$\begin{aligned}
\tilde{P}_N(k) &\equiv \left\langle e^{-ikn} \right\rangle \\
&= \sum_n \frac{N!}{n!(N-n)!} h^n (1-h)^{N-n} e^{-ikn} \\
&= (he^{-ik} + 1 - h)^N \qquad \text{Eq. 3.26 with } x = he^{-ikn} \text{ and } y = 1-h \qquad (3.28) \\
\log \tilde{P}_N(k) &= N \log(he^{-ik} + 1 - h) \qquad (3.29)
\end{aligned}$$

Note right away that $\log \tilde{P}_N(k) = N \log \tilde{P}_1(k)$. Therefore, it follows that the cumulants for N trials are N times the cumulants for a single trial. This is due to the independent nature of each trial. Clearly for a single trial,

$$\begin{aligned}
\hat{P}_1(k) &= he^{-ik} + 1 - h = 1 + h \sum_{l=1}^{\infty} \frac{(-ik)^l}{l!} \qquad (3.30) \\
\left\langle n^l \right\rangle_{N=1} &= h \qquad \text{single trial, } l = 1, 2, 3, \ldots \qquad (3.31)
\end{aligned}$$

Thus, all moments are h for a single trial. This is easy to understand, since $n = 0$ or 1 for a single trial. To get cumulants for N trials, one can evaluate cumulants for $N = 1$, and then multiply by N. In particular, we get these elementary results, quite worth remembering.

$$\begin{aligned}
\langle n \rangle &= Nh \qquad \text{mean} \qquad (3.32) \\
\sigma = \sqrt{\langle n^2 \rangle_c} &= \sqrt{N(h-h^2)} = \sqrt{Nh(1-h)} \qquad \text{standard deviation} \qquad (3.33)
\end{aligned}$$

Most emphatically, we see that

$$\frac{\sigma}{\langle n \rangle} = \frac{1}{\sqrt{N}} \sqrt{\frac{1-h}{h}} \qquad (3.34)$$

As long as h is not zero, we note that this "relative uncertainty" of the outcome vanishes as $O(N^{-1/2})$ as $N \to \infty$. This simple math gives a glimpse of *one way a certainty arises in a randomly fluctuating many body system*, considering that the typical number of particles in a small grain of substance is $O(10^{20})$, whether each particle is decaying or flipping between two possible states reacting to an external field.

Lastly, note that the binomial expansion, Eq. 3.26, is a special case of the so-called multinomial expansion.

$$(x_1 + x_2 + ...x_M)^N = \sum_{N_1,N_2,...,N_M;\sum_{j=1}^{M} N_j = N} \frac{N!}{N_1!...N_M!} x_1^{N_1}...x_M^{N_M} \tag{3.35}$$

And the corresponding multinomial probability distribution can be defined by taking x_j ($j = 1, ..., M$) as the probability for the j-th outcome out of M possible outcomes.

3.2.2 Gaussian distribution

An elementary way to obtain the Gaussian distribution is to take the limit of $N \to \infty$ for the binomial distribution. Then, $(n-Nh)/N$ can be taken as a continuous variable, say x. One can show that, for this continuous variable, the following probability density function is obtained.

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \tag{3.36}$$

This is the standard Gaussian PDF, with zero mean and standard deviation σ. For a non-zero mean, $x \to x - \bar{x}$.

Here, the exponential function $\exp(x)$ means, of course, e^x, where $e = 2.71828...$ can be defined as

$$e = \lim_{\delta \to 0} (1 + \delta)^{1/\delta}. \tag{3.37}$$

In this binomial to Gaussian example, $\sigma = \sqrt{h(1-h)}$ (cf. Sections 3.1.5,3.1.6). However, as we shall see later, the Gaussian distribution can arise in many contexts (**central limit theorem**), and so, in general, we must regard σ as a general parameter.

The normalization of the Gaussian function depends on the following integral, which can be shown to be equal to $\Gamma(0.5) = (-0.5)!$ by a simple substitution of variable ($x^2 \to t$):

$$\int_{-\infty}^{\infty} dx \exp(-x^2) = \sqrt{\pi}. \tag{3.38}$$

This integral can be conveniently calculated using the following trick (let us assume that I is the above integral value)

$$
\begin{aligned}
I^2 &= \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \exp(-x^2 - y^2) \\
&= \int_0^{\infty} dr\, d\theta\, r \exp(-r^2) && \text{change to polar coordinates} \\
&= 2\pi \int_0^{\infty} d(r^2/2) \exp(-r^2) \\
&= \pi \int_0^{\infty} d\xi \exp(-\xi) \;=\; \pi && \xi \equiv r^2
\end{aligned}
$$

The characteristic function for the Gaussian PDF is written as

$$\tilde{p}(k) = \int_{-\infty}^{\infty} dx \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2} - ikx\right).$$

By completing the square in the exponent, and effecting the necessary contour integral, we get

$$\tilde{p}(k) = \exp\left(-\frac{k^2\sigma^2}{2}\right) \tag{3.39}$$

from which we see that $\log \tilde{p}(k) = -\frac{k^2\sigma^2}{2}$ and thus

$$\left\langle x^2 \right\rangle_c = \sigma^2 \tag{3.40}$$

$$\left\langle x^n \right\rangle_c = 0 \qquad n = 3, 4, \ldots \tag{3.41}$$

Note that if x is shifted, e.g., $x \to x - \bar{x}$, then the mean value of x changes from 0 to $\bar{x}$, but no other cumulants change (cf. Section 3.1.5). Further scaling of x will affect the first and second cumulants only. **Thus, all cumulants of order 3 or higher are zero for any Gaussian distribution.** In this regard, one can say that the Gaussian function is, by definition of a statistician, a "perfect shape" or a "standard shape."

Lastly, note that the standard deviation $\sigma = \sqrt{\langle x - \bar{x}^2 \rangle}$ is a useful measure of the width. Another common measure of the width of any peak is the so-called FWHM (full width at half maxima). For the Gaussian distribution function, we get

$$\text{FWHM} = \sigma \times 2\sqrt{2\log 2} \approx 2.35\,\sigma. \tag{3.42}$$

3.2.3 Poisson distribution

The Poisson distribution is very important, not only in the (radioactive) particle decay example given here, but also in other contexts such as particle detections, as in many particle counting experiments. Another important application of the Poisson distribution is the coherent state of photons (as in the description of laser in quantum optics) or the coherent state of phonons.

It applies when a single event is very unlikely to occur, but there are so many trials and so the total average number of events is a finite number. That is, a Poisson distribution is obtained when $h \to 0$, while $N \to \infty$, and $Nh \to$ finite, from the binomial distribution.

For a concrete example, let us consider a particle decay problem. This can be a radioactive particle decay. Or, a charge carrier being knocked off a state in an energy band in a semiconductor.

We suppose that at the beginning there are $\mathcal{N}$ particles. The particle has a decay time constant τ. This means that during a small time interval dt, the particle's chance to decay is given by dt/τ. As we shall see, this defines the so-called "1/e" decay time (cf. Section 3.2.4). Perhaps the more common measure of decay time is the half life, $\tau_{1/2}$, which is related to τ as

$$\tau_{1/2} = \tau \log 2 \approx 0.693\,\tau. \tag{3.43}$$

Let us consider a finite time interval $T \ll \tau$. So, for any one particle the probability that a decay will happen during time T is very small. Then, we can assume that the probability that a particle will decay in any infinitesimal time interval dt is the same for any infinitesimal time interval picked out randomly out of $N = T/dt$ intervals[3]. Furthermore, we can also assume that, by taking dt arbitrarily small, the actual number of decay per dt is very small: $h = \mathcal{N} dt/\tau \to 0$, even if we have a very large number of particles $\mathcal{N} \to \infty$.

Here is a summary of symbols and their conditions.

$$\begin{aligned}
\mathcal{N} &\to \infty && \text{number of particles} && (3.44)\\
T/\tau &\to 0 && T \text{ is a finite time window of observation} && (3.45)\\
h &= \mathcal{N} dt/\tau \to 0 && \text{number of particle decays in one } dt && (3.46)\\
N &= T/dt \to \infty && \text{number of } dt \text{ time intervals} && (3.47)\\
\lambda \equiv hN &= \mathcal{N} T/\tau = \text{finite} && \text{mean number of particle decays within } T && (3.48)
\end{aligned}$$

How does a binomial distribution become a Poisson distribution when $h \to 0$ but

[3] If T is too large, then the particle will most likely have decayed at an early time, so the probability of decay in a dt interval will decrease as time progresses.

$Nh \to$ finite? Let us see.

$$\begin{aligned}
& \text{Binomial distribution} \\
= \; & \frac{N!}{n!(N-n)!}\, h^n (1-h)^{N-n} \\
= \; & \frac{N!}{n!(N-n)!}\, (\lambda/N)^n \frac{(1-\lambda/N)^N}{(1-\lambda/N)^n} && \text{assume } \lambda = hN = \text{finite} \\
= \; & \frac{\lambda^n}{n!} \cdot \frac{N(N-1)...(N-n+1)}{N^n} \cdot \frac{(1-\lambda/N)^N}{(1-\lambda/N)^n} \\
= \; & \frac{\lambda^n}{n!} \cdot (1-\lambda/N)^N \cdot \frac{N(N-1)...(N-n+1)}{N^n} \cdot \frac{1}{(1-\lambda/N)^n} \\
= \; & \frac{\lambda^n}{n!} \cdot (1-\lambda/N)^N \cdot \frac{(1-1/N)(1-2/N)...(1-(n-1)/N)}{(1-\lambda/N)^n} \\
\to \; & \frac{\lambda^n}{n!} \cdot (1-\lambda/N)^N && \text{for any fixed } n, \text{ as } N \to \infty \\
\to \; & \frac{\lambda^n \exp(-\lambda)}{n!} && \text{by Eq. 3.37} \\
= \; & \text{Poisson distribution function}
\end{aligned}$$

Note that we just obtained the Poisson PDF.

$$p(x) = \sum_{n=0}^{\infty} \frac{\lambda^x}{x!} e^{-\lambda} \delta(x-n) \tag{3.49}$$

The characteristic function is obtained either by Fourier transforming this PDF, or, directly from Eq. 3.28:

$$\begin{aligned}
\log \tilde{p}(k) &= N \log(he^{-ik} + t) && (3.50) \\
&= N \log(1 + h(e^{-ik} - 1)) && (3.51) \\
&\approx \lambda(e^{-ik} - 1) \qquad h \to 0, Nh \to \lambda && (3.52) \\
&= \lambda \sum_{l=1}^{\infty} \frac{(-ik)^l}{l!} && (3.53)
\end{aligned}$$

Thus, we get a very simple result for the Poisson distribution. Every cumulant is simply given by λ.

$$\langle x^l \rangle_c = \langle n^l \rangle_c = \lambda \qquad l = 1, 2, 3, ... \tag{3.54}$$

Couple of notes are in order. (1) In the particle detector case, dt corresponds to the gate time of a particle detector. For a particle detector to count events correctly, it is generally the case that only one event, if ever, must be recorded per gate time, since two or three pulses of events merged in a single gate time would be misinterpreted as one event by the detector. So, the probability of particle detection per dt must be very small, satisfying the condition for Poisson statistics. (2) It can be shown that, if the limit of λ becoming very large ($\lambda \to \infty$) is taken, then the Poisson distribution function converges to a Gaussian distribution function with mean λ and $\sigma = \sqrt{\lambda}$.

3.2.4 Lorentzian distribution

This important distribution function is given by

$$p(\omega) = \frac{1}{\pi}\frac{\Gamma}{\omega^2 + \Gamma^2} \qquad \Gamma > 0 \tag{3.55}$$

This form is centered at zero, and has the FWHM 2Γ. If we change ω to $\omega - \omega_1$, then we can describe the Lorentzian distribution, which is centered at the ω_1. By contour integral, one can show that this function does satisfy the probability sum rule $\int_{-\infty}^{\infty} d\omega\, p(\omega) = 1$. However, note that none of higher moments exist. By symmetry, however, one might argue that all odd moments are zero, evaluating the integrals as the Cauchy principal value. In any case, any even moments for order greater than 0 are divergent, and so this quite useful distribution function defies the description in terms of moments or cumulants.

In what context might this distribution arise? One common occurrence is related to the problem considered for the Poisson distribution. Let us consider a particle that is prone to decay. We consider only one particle, not many particles. Let us assume that at $t = 0$ we have prepared the particle and made sure that it exists. Then, we ask, what is the probability that the particle is alive at $t > 0$, $P(t)$? For the particle to be alive at t, it must be alive at $t - dt$, also, and it must survive the time interval $[t - dt, t]$. Thus, we can set up the differential equation: $P(t + dt) = P(t)\left(1 - \frac{dt}{\tau}\right)$. Solving this, we get

$$P(t) = \exp\left(-\frac{t}{\tau}\right) \qquad t \geq 0 \tag{3.56}$$

recognizing $P(t = 0) = 1$ as the initial condition imposed by our experimental constraint. This is the familiar **Rutherford decay law**, generally applicable to any unstable particle.

What quantum mechanical principle is actually responsible for this decay? It is the energy-time uncertainty relation. Indeed, using a time-dependent perturbation theory[4], one can show that the energy of an excited state gets shifted as well as acquiring an imaginary part. The result is that the energy eigenvalue "becomes

[4] See, e.g., Sakurai, Section 5.8.

complex:" $\hbar\omega_1 - i\hbar\Gamma$.

$$\Psi(\vec{x},t) \;=\; \psi(\vec{x})\exp(-i\omega_1 t - \Gamma t) \qquad\qquad t \geq 0. \tag{3.57}$$

where $\hbar\omega_1$ is the perturbed energy level position and $-i\hbar\Gamma$ can be interpreted as a purely imaginary part of energy, arising due to the transition induced by perturbation.

Thus, the survival probability of the excited state is then given by

$$P(t) \;=\; \int d^3\vec{x}|\Psi(\vec{x},t)|^2 = \exp(-2\Gamma t) \qquad\qquad t \geq 0$$

as it must be that $\int d^3\vec{x}|\psi(\vec{x})|^2 = 1$. Indeed, this is exactly in the form of the Rutherford decay law, and thus we identify

$$\frac{1}{\tau} \;=\; 2\Gamma. \tag{3.58}$$

This relation is none other than the energy-time uncertainty relation. By taking the inverse Fourier transformation, one sees that[5]

$$\left|\int_0^\infty dt\,\Psi(\vec{x},t)e^{i\omega t}\right|^2 \propto \frac{1}{\pi}\frac{\Gamma}{(\omega-\omega_1)^2+\Gamma^2} \tag{3.59}$$

This is the Lorentz distribution function of Eq. 3.55, except that for the current problem the mean value is ω_1, not necessarily zero. Therefore, the frequency distribution of the quantum state for our prepared particle is given by the Lorentzian function. Upon spectroscopy of light emitted by the decay of the state, e.g., one would discover that its functional form is precisely this type of Lorentzian function. Indeed, measuring the center position and the FWHM of such a Lorentzian function is the standard way of measuring the energy eigenvalue and the lifetime of a quantum state.

The Lorentzian distribution function has a long power law tail, and thus can not be described in terms of cumulants. It also does not obey the central limit theorem (cf. next lecture). It is *very distinct* from the Gaussian distribution function. In practice, the Gaussian distribution function and the Lorentzian distribution function are two of the most commonly occurring probability distribution functions in physics.

[5] The integral is done only from $t = 0$, since we assume that the particle did not exist for $t < 0$, and came into existence by our creation. This is indeed how actual spectroscopies are performed on quantum particles to measure their energy eigenvalues and lifetime.