Notes for Lecture 3

Probability

In the last class, 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 the result of the maximum entropy principle, although we did not explicitly discuss E in Section 2.3. Let us reflect upon the classical mechanics.

You might ask, how does the maximum entropy principle work when one says, for example, "a pendulum will settle down to the lowest energy state" in classical mechanics. What we really mean is that the total entropy will be maximized if as much mechanical work is converted, and lost, to heat (through friction). This type of situation admits a simple analysis, if one realizes that the ideal pure mechanical situation corresponds to $T = 0$ and $S = 0$ state, as in the pure mechanical modeling of a pendulum, only one particle is envisioned. So, this is the case of the Helmholtz free energy with $T = 0$. However, one might protest that a pendulum is clearly not one particle, and it must have a non-zero temperature, to be realistic. Then, one could put this situation in a more general light, as follows. Energy is minimized if the entropy of the system is held constant and no work is exchanged (so no changes in the displacement thermodynamic variables) between the system and the environment. Then, since $dS_{tot} = dS + dS_{env} = dS_{env} = dE_{env}/T = -dE/T \geq 0$, we see that the maximum entropy leads to the minimum energy. Note that in this situation, the system experiences $đQ \leq TdS$ (as a consequence of the Clausius theorem). That is, the system loses some internal energy to heat. Note that in this case the system must be considered not as the pendulum alone, but as the pendulum plus the Earth.

So, you see, the maximum entropy principle is really the overarching 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.

3.1 Various points

Hoping to help you follow the textbook with more ease, let me just add some comments on various parts.

3.1.1 log or ln?

I think the logarithm is automatically a natural logarithm, unless specified otherwise. So, I will use the symbol log for the natural log. I will not use the symbol ln. If ever we need to use the log with the base of 10, then it can be used as $\log_{10}$.

3.1.2 e = 2.71828...

$$e = \lim_{\delta \to 0} (1+\delta)^{1/\delta} \tag{3.1}$$

This is useful for understanding Eq. T2.25[1], for instance.

3.1.3 Average and standard deviation

I will use the average symbol $\bar{x}$, interchangeably with $\langle x \rangle$. Also, the standard deviation symbol σ (or σ_x) will be used to mean $\sqrt{\langle x^2 \rangle_c}$, for any distribution, not just the Gaussian distribution.

3.1.4 Cluster expansion

Here, we examine the coefficient of Eq. T2.14, which is the basis of the connected cluster analysis used to calculate moments from cumulants. Oddly, it is easier to do this, and then when one wants to calculate cumulants from moments, one can simply

[1] When I refer to equations or items in the textbook, the number will prepended by "T."

reverse equations. Namely, the "correct process" is to derive Eq. T2.12, and then invert this to get Eq. T2.11, as was demonstrated in class.

Now let us look at Eq. T2.14, which is at the heart of the mathematics as to how this cluster expansion work. For Eq. T2.14, it must be clear that, for a given set of non-negative integers $\{p_n\}$, and with $n = 1, 2, 3, ...$ and $\sum_n np_n = m$, the quantity $A_{m,\{p_n\}} = \frac{m!}{\prod_n p_n!(n!)^{p_n}}$, is the total number of ways to group m points into distinct sets, where each set is defined as p_n clusters of n points. Two important points are (1) each set of p_n cluster of n points is distinct from one another since n values are different between them and (2) each (smaller) set of n points in a given cluster is identical with any other set of n points in the same cluster.

That is a bit mouthful, and it may help to view this factor as follows, if necessary.

$$A_{m,\{p_n\}} = \frac{m!}{\prod_n (np_n)!} \prod_n \left(\frac{(np_n)!}{p_n!(n!)^{p_n}} \right) = \frac{m!}{\prod_n (np_n)!} \prod_n \left(\frac{(np_n)!}{(n!)^{p_n}} \frac{1}{p_n!} \right) \tag{3.2}$$

Note that here, the first factor $\frac{m!}{\prod_n (np_n)!}$ is the total number of ways to divide m points into sets of sizes np_n. Then, in the big product term, note that $(np_n)!/(n!)^{p_n}$ is the total number of ways to divide np_n points into sets of n points, assuming that each n point set is distinct. However, since each such set is not distinct from one another (point (2) above), we have to divide by $p_n!$.

3.1.5 Cumulants

It is clear that the second cumulant is the variance.

$$\langle x^2 \rangle_c = \langle x^2 \rangle - \langle x \rangle^2 = \sigma^2 \tag{3.3}$$

From the third and on, the conventional definitions for the skewness and the curtosis are the following.

$$\text{standard skewness} \equiv \frac{\langle (x - \bar{x})^3 \rangle}{\sigma^3} \tag{3.4}$$

$$\text{standard curtosis} \equiv \frac{\langle (x - \bar{x})^4 \rangle}{\sigma^4} - 3 \tag{3.5}$$

Note that these standard skewness and curtosis are dimensionless, while the skewness and curtosis as defined (as the third and the forth cumulants, respectively) in the textbook are dimension-ful. From Eq. T2.11, one can show the following (please verify this yourself).

$$\langle x^3 \rangle_c = \sigma^3 \times (\text{standard skewness}) \tag{3.6}$$

$$\langle x^4 \rangle_c = \sigma^4 \times (\text{standard curtosis}) \tag{3.7}$$

3.1.6 Gaussian distribution

The normalization of the Gaussian function depends on the following integral

$$\int_{-\infty}^{\infty} dx\ \exp(-x^2) \;=\; \sqrt{\pi} \tag{3.8}$$

which can be conveniently derived from the following trick (let us assume that I is the above integral value)

$$\begin{aligned}
I^2 &= \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \exp(-x^2 - y^2) && (3.9)\\
&= \int_{0}^{\infty} dr\, d\theta\, r \exp(-r^2) \qquad \text{change to polar coordinates} && (3.10)\\
&= 2\pi \int_{0}^{\infty} d(r^2/2) \exp(-r^2) && (3.11)\\
&= \pi \int_{0}^{\infty} d\xi \exp(-\xi) \qquad \xi \equiv r^2 && (3.12)\\
&= \pi && (3.13)
\end{aligned}$$

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

$$\text{FWHM} \;=\; \sigma \times 2\sqrt{2\log 2} \approx 2.35\,\sigma \tag{3.14}$$

3.1.7 Poisson distribution

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

The textbook content presented in pages T42-43 need to be qualified a bit more, as follows.

We suppose that one initially starts with $\mathcal{N}$ particles. The particle has a decay time constant τ. This is the so-called "1/e" decay time. Perhaps the more common measure, half life, $\tau_{1/2}$ is given by

$$\tau_{1/2} = \tau \log 2 \approx 0.693\,\tau \tag{3.15}$$

The discussion in pages T42-43 is valid only if $T \ll \tau$. So, for any one particle the probability that the decay will happen during time T is very small. Only then, we can assume that the probably that a particle will decay in any infinitesimal time interval dt is the same for any infinitesimal time interval picked out randomly.

The symbol α used in the textbook is then given by $\alpha = \mathcal{N}/\tau$. To summarize, these are the conditions for this discussion.

$$\begin{aligned}
\alpha &= \mathcal{N}/\tau && (3.16)\\
\mathcal{N} &\rightarrow \infty && (3.17)\\
T/\tau &\rightarrow 0 && (3.18)\\
p &= \mathcal{N}dt/\tau \rightarrow 0 && (3.19)\\
N &= T/dt \rightarrow \infty && (3.20)\\
\lambda \equiv pN &= \alpha T = \mathcal{N}T/\tau = \text{finite} && (3.21)
\end{aligned}$$

It is well-known that the Poisson distribution is the limit of the binomial distribution when $p \rightarrow 0$ and $N \rightarrow \infty$. Here is a direct proof (alternative to the textbook content).

$$\begin{aligned}
& \text{binomial distribution} && (3.22)\\
= & \frac{N!}{k!(N-k)!}\,p^k(1-p)^{N-k} && (3.23)\\
= & \frac{N!}{k!(N-k)!}\,(\lambda/N)^k\frac{(1-\lambda/N)^N}{(1-\lambda/N)^k} \qquad \text{assume } \lambda = pN = \text{finite} && (3.24)\\
= & \frac{\lambda^k}{k!}\cdot\frac{N(N-1)...(N-k+1)}{N^k}\cdot\frac{(1-\lambda/N)^N}{(1-\lambda/N)^k} && (3.25)\\
= & \frac{\lambda^k}{k!}\cdot(1-\lambda/N)^N\cdot\frac{N(N-1)...(N-k+1)}{N^k}\cdot\frac{1}{(1-\lambda/N)^k} && (3.26)\\
= & \frac{\lambda^k}{k!}\cdot(1-\lambda/N)^N\cdot\frac{(1-1/N)(1-2/N)...(1-(k-1)/N)}{(1-\lambda/N)^k} && (3.27)\\
\rightarrow & \frac{\lambda^k}{k!}\cdot(1-\lambda/N)^N \qquad \text{for any fixed } k\text{, as } N \rightarrow \infty && (3.28)\\
\rightarrow & \frac{\lambda^k\,\exp(-\lambda)}{k!} \qquad \text{by Eq. 3.1} && (3.29)\\
= & \text{ Poission distrubtion function (same as Eq. T2.28)} && (3.30)
\end{aligned}$$