The goal of these notes is to outline some steps in deriving the partition function in classical statistical mechanics.

We will not go into the physics, which might take a little long to explain, even though it is what really motivates the exercises below. Let E denote the mean energy of the system, and E_i the energy at the ith state. There are states $n_1, n_2, \ldots$ and there are N systems. There are infinitely many states, and all but finitely many of them are nonzero. The N systems can be thought of particles —of which there are a large number— and the n_i states are boxes that some of the N particles can occupy. The number of states adds up to the number of systems. The probability of state i occurring is $p_i = n_i/N$.

- Find the number of ways of distributing the N systems among the $n_1, n_2, \ldots$ states. This is given by the multinomial coefficient:

$$\binom{N}{n_1, n_2, \ldots} = \frac{N!}{n_1! n_2! \cdots}.$$

To show this, try to show first as a warm-up exercise that the number of ways of choosing a k-subset from a collection of N elements is given by the binomial coefficient

$$\binom{N}{k} = N!/((N-k)!k!).$$

Remark: While there are other ways of maximizing the multinomial coefficient, e.g., see `http://www.maths.qmul.ac.uk/~twm/MTH742U/AdvCombEx.pdf`, the notion of entropy —defined below— can be readily extended to continuous probability distributions...

- Show that $N! \sim N^N e^{-N}$, that is, $N!$ is approximately $N^N e^{-N}$ for N large. (An informal argument can be given by approximating $\ln N!$ by the integral $\int_1^N \ln x \, dx$. A more rigorous argument will look like the proof for Stirling's approximation formula.)

- The multinomial coefficient involves products and quotients. Since what we want to find are the $n_1, n_2, \ldots$ making the multinomial coefficient as large as possible, we can alternatively maximize instead the logarithm of the multinomial coefficient. Show that

$$\ln\left(\binom{N}{n_1, n_2, \ldots}\right) \sim -N \sum p_i \ln p_i,$$

where $p_i = n_i/N$ is the probability of the ith state. The above quantity is often denoted by $S(p_1, p_2, \ldots)$, and it is called **entropy**.

Warning: This is the other bit of a leap of faith in this series of exercises —namely, optimizing the approximation (in this case the entropy) will lead to optimizing the original function (the natural logarithm of the multinomial coefficient). There are silly examples in which a function approximates another one but things go wrong in the optimization. Here is one such example: Let $f(x) = x^2 + x - 2$ be a function defined from $(-1, \infty)$ to $\mathbb{R}$. Suppose we want to "approximate" this function by another one. Pick $g(x) = x^2 + \frac{1}{4}\frac{x}{x+1} + 1$, also defined from $(-1, \infty)$ to $\mathbb{R}$. One might say that g is a reasonable "approximation" to f since

$$\lim_{x \to \infty} \frac{f(x)}{g(x)} = 1.$$

(We remark that Stirling's approximation satisfies the above property, that is, when we say that $n! \sim \sqrt{2\pi n}(n/e)^n$, that means that $\lim_{n\to\infty} n!/(\sqrt{2\pi n}(n/e)^n) = 1$.) Single-variable calculus tells us that $x = -1/2$ is a critical point of both f and g. However, the second derivative test shows that $x = -1/2$ is a local maximum of f, whereas for g it is a local minimum. (One issue is that we have not been precise with what we mean by "approximating" a function by another.)

- In order to maximize the entropy, solve the following optimization problem:

$$\begin{aligned} \text{minimize} \quad & \sum p_i \ln p_i \\ \text{subject to} \quad & \sum p_i = 1 \\ & \sum p_i E_i = E \end{aligned}$$

  (Notice that the function we want to minimize is $-S(p_1, p_2, \dots)$.) The minimization problem above can be solved with Lagrange multipliers:

$$\sum p_i \ln p_i + \alpha \left(\sum_i p_i - 1 \right) + \beta \left(\sum p_i E_i - E \right),$$

  where α, β are the Lagrange multipliers. Show that

$$p_j = e^{-(1+\alpha)} e^{-\beta E_j} = \frac{1}{Z} e^{-\beta E_j},$$

  where $Z = e^{1+\alpha}$.

Using the constraint $\sum p_j = 1$, we have

$$\sum \frac{1}{Z} e^{-\beta E_j} = 1 \quad \Longrightarrow \quad Z(\beta) = \sum_j e^{-\beta E_j}.$$

The quantity $Z(\beta)$ is called the **partition function**.