Topics in Physical Chemistry and Biophysics

1 Review of probability

1.1 Rules of probability

18.04.23 lec 1

Definition 1.1

Probability. If N is the total number of outcomes, and n_A fall in category A, then

$$p_A = \frac{n_A}{N} = \frac{\text{outcomes cat. } A}{\text{all outcomes}}.$$

Rules of composite events:

1. Mutually exclusive: outcomes $(A_1, A_{2,\dots})$ are *mutually exclusive* if one outcome precludes another outcomes. (Event A_1 prevents even A_2 from happening simultaneously.)

2. Collectively exhaustive: if all known outcomes are also all possible outcomes. $\sum p_i = 1$.

3. Independence: outcomes do not depend on each other.

4. Multiplicity: total number of ways in which outcomes occur.

Rules of calculation:

1. Let there be 3 outcomes A, B, C with probability p_A, p_B, p_C. What is the probability that either one occurs (A or B or C)?

$$p(A \cup B \cup C) = p_A + p_B + p_C.$$

   That's the addition rule.

2. Probability that all outcomes occur? (Assuming independence)

$$p(A \cap B \cap C) = p_A p_B p_C.$$

3. Probability that an event A is not happening? $p = 1 - p_A$.

Example. We roll a die twice. What is the probability of rolling a 1 first **or** a 4 second?

Split the problem to parts. Note that the events are not mutually exclusive. Condition applies if:

- 1 first and not a 4 second: $\frac{1}{6} \cdot \frac{5}{6}$
- not a 1 first and a 4 second: $\frac{5}{6} \cdot \frac{1}{6}$
- 1 first and 4 second: $\frac{1}{6} \cdot \frac{1}{6}$

Now sum up all of the options to get result.

Definition 1.2

Correlated events. $p(B|A)$ is the probability that B occurs given A has occurred.

Joint probability. $p(AB)$ that both A and B occur.

Definition 1.3

General multiplication rule.

$$p(AB) = p(B|A)\, p(A)\ .$$

$P(A)$ is called the a priori probability and $p(B|A)$ is called the a posterior probability

Theorem 1.4

Bayes theorem.

$$p(B|A)\, p(A) = p(A|B)\, p(B)\ .$$

Example. 1% of population has breast cancer. We use mammography to detect cancer.

Event A: breast cancer. $p(A) = 0.01$. $p(\bar{A}) = 1 - p(A) = 0.99$.

Event B: diagnosis. $p(B|A) = 0.8$. $p(B|\bar{A}) = 0.096$. (i.e. false positive0

What is the chance that a doctor has diagnosed someone with cancer? i.e. $p(A|B)$

$$p(A|B) = \frac{p(B|A)\, p(A)}{p(B)}\ .$$

$p(B)$ is the diagnosis of breast cancer irrespective whether it's there or not there.

$$p(B) = p(BA) + p(B\bar{A}) = p(B|A)\, p(A) + p(B|\bar{A})\, p(\bar{A}) = 0.8 \cdot 0.01 + 0.096 \cdot 0.99 = 0.103,$$

$$p(A|B) = \frac{0.8 \cdot 0.01}{0.103} = 0.078 = 7.8\%.$$

The reason that $p(A|B)$ is so small is that the rate of false positive is really low and the rate of having breast cancer is really low.

1.2 Combinatorics and probability distributions

Combinatorics. Concerned with composition of events, and not with their order.

Example. How many combinations there are of N amino acids?

$$W = N! = N\,(N-1)\,(N-2)\cdots$$

Example. Distinguish or not Distinguish: What are the possible number of ways to arrange N amino acids? Divide all permutations (assuming objects are distinguishable) by the number of permutations of objects that are indistinguishable.

$$W = \frac{N!}{N_A}\ .$$

In general, for N objects consisting of t categories in which the objects are indistinguishable:

$$W = \frac{N!}{(n_1!)\,(n_2!)\cdots(n_t!)}\,.$$

So, if $t=2$, (e.g. possible number of ways to arrange three acids A,A,H)

$$W = \frac{N!}{n_1!\cdot n_2!} = \frac{N!}{n_1!\,(N-n_1)!} = \binom{N}{n}\,.$$

Definition 1.5

Distribution functions. Describe collections of probabilities. Relevant for continuous variables.

$$\sum_i p_i \to \int_a^b p(x)\,\mathrm{d}x\,.$$

Popular distributions:

1. *Binomial Distribution.* Relevant when there are only two outcomes.

   Example. What is the probability that a series of N trials has n_H heads and n_T tails in any order?

   p_H, p_T are mutually exclusive, so the probability of one sequence is

   $$p_H^{n_H}\cdot p_T^{n_T} = p_H^{n_H}\,(1-p_H)^{N-n_H}; \quad N = n_H + n_T\,.$$

   and the number of ways to arrange the coins is

   $$W = \frac{N!}{n_H!\,(N-n_H)!}\,.$$

   Therefore, the possibility for the outcome (getting n_H and n_T) in any order is

   $$\boxed{p(n_H, N) = \binom{N}{n_H} p_H^{n_H}\,(1-p_H)^{N-N_H}\,.}$$

   That's the binomial distribution.

   Example. Given the molecule $C_{27}H_{44}O$ such that 1.1% is ^{13}C and the rest are ^{12}C, the fraction of molecules without a single ^{13}C is given by the binomial distribution.

2. *Multinomial distribution.* Basically the extension of the binomial distribution.

   $$\boxed{p(n_1, n_2, \ldots, n_t, N) = \left(\frac{N!}{n_1!\,n_2!\cdots n_t!}\right) p_1^{n_1} p_2^{n_2}\cdots p_t^{n_t}\,.}$$

Definition 1.6

Moments of distributions. Averages and Variances of distribution functions.

Given $p(i)$ s.t. $\sum_i p_{(i)} = 1$, the **Average** is defined as

$$\langle i \rangle = \sum_i i\, p(i) \rightarrow \langle x \rangle = \int x\, p(x)\, \mathrm{d}x.$$

Given $f(x)$,

$$\langle f(x) \rangle = \int f(x)\, p(x)\, \mathrm{d}x.$$

Given $a \in \mathbb{R}$

$$\langle a f(x) \rangle = \int a f(x)\, p(x)\, \mathrm{d}x = a \langle f(x) \rangle .$$

Given 2 functions $f(x), g(x)$,

$$\langle f(x) + g(x) \rangle = \langle f(x) \rangle + \langle g(x) \rangle,$$

$$\langle f(x) \cdot g(x) \rangle \neq \langle f(x) \rangle \langle g(x) \rangle .$$

The 2nd and 3nd **Moments** of the distributions $p(x)$ are

$$\langle x^2 \rangle = \int x^2 p(x)\, \mathrm{d}x,$$

$$\langle x^3 \rangle = \int x^3 p(x)\, \mathrm{d}x.$$

The **Variance** of the distribution, σ^2 is defined as

$$\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2 = \langle (x - \langle x \rangle)^2 \rangle .$$

2 Entropy

2.1 Definition and Stirling's approximation

02.05.23 lec 2 Carved on the tombstone of Ludwig Boltzmann in the central cemetery in Vienna is the definition of *entropy*.

Definition 2.1

Entropy.

$$S = k \ln W .$$

- $k = 1.3806 \times 10^{-23}\,\mathrm{J\,K^{-1}}$ is Boltzmann's constant.
- W is the multiplicity; the microscopic degrees of freedom of a system.

Entropy can help describe the state of a system in equilibrium, as systems tend toward their states of maximum multiplicity W (and minimum energy).

Note. Entropy is an *extensive* (thus additive) quantity. Consider a thermodynamic system having two subsystems, A and B, with multiplicities W_A and W_B, respectively. The multiplicity of the total system is $W_{\text{total}} = W_A W_B$. Following def. 2.1, $S_{\text{total}} = S_A + S_B = k \ln W_A + k \ln W_B$. This is why incorporating the multiplicity in a logarithm makes sense.

Why does def. 2.1 assume this particular mathematical form? The multiplicity could be maximized as $W^2, 15W^3$ etc. We first show that expressing the entropy in terms of a *set of probabilities* p_i,

$$\frac{S}{k} = -\sum_{i=1}^{t} p_i \ln p_i. \tag{2.1}$$

is equivalent to def. 2.1. Roll a t-sided die N times. The multiplicity of outcomes is given by

$$W = \frac{N!}{n_1! n_2! \cdots n_t!},$$

where n_i is the number of times that side i appears face up. Use Stirling's approximation,

Definition 2.2

Stirling's approximation. For $n \gg 1$ (at least 1000),

$$\ln n! \approx n \ln n - n.$$

$$n! \approx \left(\frac{n}{\mathrm{e}}\right)^n.$$

and define the probabilities $p_i = n_i / N$, to get

$$W = \frac{(N/\mathrm{e})^N}{(n_1/\mathrm{e})^{n_1} (n_2/\mathrm{e})^{n_2} \cdots (n_t/\mathrm{e})^{n_t}} = \frac{N^N}{n_1^{n_1} n_2^{n_2} \cdots n_t^{n_t}} = \frac{1}{p_1^{n_1} p_2^{n_2} \cdots p_t^{n_t}}.$$

Take the logarithm of both sides and divide by N to get

$$\ln W = -\sum_{i=1}^{t} n_i \ln p_i \quad \Rightarrow \quad \frac{1}{N} \ln W = -\sum_{i=1}^{t} p_i \ln p_i = \frac{S_N}{Nk} = \frac{S}{k},$$

where S_N is the total entropy for N trials, so the entropy per trial is $S = S_N / N$.

Note. def. 2.1 is reformulated from eq. (2.1) if the microstates whose permutations sum up to multiplicity W are *degenerate*, or equivalently, if $p_i = 1/W$.

Note. Boltzmann's constant k puts entropy into units that inter-convert with energy for thermodynamics. Basically, k is the entropy per particle.

Sometimes, it is more convenient to express the entropy per mole of particles,

$$S = R \ln W.$$

where $R = \mathcal{N}k$ is the *gas constant* and $\mathcal{N}$ is Avogadro's number—the number of molecules per mole.

2.2 Lattice models

Example. Entropy of mixing calculation for *lattice models*.

Two solutions, A and B, are allowed to mix. Calculate the change in entropy of the system due to mixing.

Method. Describe the solution (space) as a lattice, or grid, which has N lattice sites, which are filled by n particles.

W is the number of ways to arrange particles in the available sites.

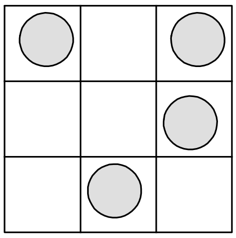


Figure 2.1. Each particle may occupy a lattice site.

Each site can either be occupied or vacant, so W behaves according to the binomial distribution.

$$W_A = \frac{N!}{n!\,(N-n)!}\,.$$

Using def. 2.2,

$$W_A \approx \frac{N^N}{n^n\,(N-n)^{N-n}}\,.$$

Similarly, for the second solution,

$$W_B \approx \frac{M^M}{m^m\,(M-m)^{M-m}}\,.$$

As entropy is extensive,

$$S = S_A + S_B = k \ln\,(W_A\,W_B)\,.$$

After mixing there are $M+N$ lattice sites and $m+n$ particles. Note that $W_{AB} = W_A\,W_B$ is the combined multiplicity without mixing. The multiplicity after mixing, $W^\star_{AB}$ behaves according to the multinomial distribution.

$$W^\star_{AB} = \frac{(N+M)!}{n!\,m!\,(N+M-m-n)!} \approx \frac{(N+M)^{N+M}}{n^n\,m^m\,(N+M-m-n)^{N+m-m-n}}\,.$$

For simplification, assume $N=M$ and $n=m$.

$$W^\star_{AB} = \frac{(2N)^{2N}}{n^{2n}\,(2N-2n)^{2N-2n}}\,.$$

In this case, the combined multiplicity prior to mixing is

$$W_{AB} = \frac{N^{2N}}{n^{2n}\,(N-n)^{2(N-n)}}\,.$$

What is the change in entropy due to mixing?

$$\Delta S = S^\star_{AB} - S_{AB} = k \ln \frac{W^\star_{AB}}{W_{AB}} = \cdots = k \ln\,(2^{2n}) = 2\,n\,k \ln 2\,.$$

Obviously, mixing increased the total entropy of the system.

2.3 Predicting distributions by maximizing entropy

In an isolated environment (no exchange of energy or matter), entropy maximization predicts the distribution of states of the system.

From eq. (2.1) it is clear that S is a *function* of the possible states: $S = f(p_1, \ldots, p_t)$. a function $f(x)$ is maximized when $\mathrm{d}f/\mathrm{d}x = 0$ and $\mathrm{d}^2 f/\mathrm{d}x^2 < 0$.

Additionally, if we know the value of $f(x)$ at some point $x = a$, we can use *Taylor series expansion* to compute $f(x)$ near that point:

$$\Delta f = f(x) - f(a) = \left(\frac{\mathrm{d}f}{\mathrm{d}x}\right)_{x=a} \Delta x + \frac{1}{2}\left(\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}\right)_{x=a} \Delta x^2 + \frac{1}{6}\left(\frac{\mathrm{d}^3 f}{\mathrm{d}x^3}\right)_{x=a} \Delta x^3 + \cdots.$$

For very small changes, $\Delta x = (x - a) \to \mathrm{d}x$, non-linear terms in the series expansions are negligible, and thus $\mathrm{d}f \approx \left(\frac{\mathrm{d}f}{\mathrm{d}x}\right)_{x=a} \mathrm{d}x$.

In the case of a bivariate function, $f(x, y)$,

$$\mathrm{d}f = \left(\frac{\partial f}{\partial x}\right)_y \mathrm{d}x + \left(\frac{\partial f}{\partial y}\right)_x \mathrm{d}y.$$

We can generalize to multivariate functions, and define

$$\mathrm{d}f = \sum_{i=1}^{t} \left(\frac{\partial f}{\partial x_i}\right)_{x_{j \neq i}} \mathrm{d}x_i.$$

The extrema of multivariate functions occur where the partial derivatives are zero. The *global* extremum occurs where *all* partial derivatives are zero: $\left(\frac{\partial f}{\partial x_i}\right)_{x_{j \neq i}} = 0$ for $i = 1, 2, \ldots, t$.

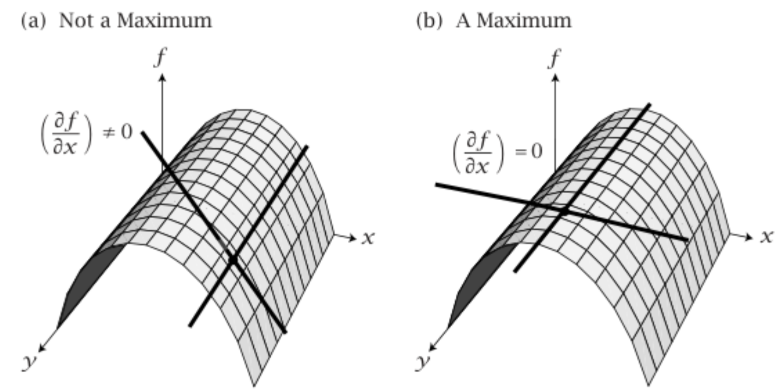


Figure 2.2. To identify the maximum of this function, both $(\partial f/\partial x)$ and $(\partial f/\partial y)$ must equal zero, as they do in (b), but not in (a).

Note that when calculating entropy we are constrained by $\sum p_i = 1$. How do we find the extrema of a function that is subject to a constraint? We must find a set of values that satisfy *both* the extremum equation

$$\mathrm{d}f = \sum_{i=1}^{t} \left(\frac{\partial f}{\partial x_i}\right)_{x_{j \neq i}} \mathrm{d}x_i = 0,$$

and the constraint equation, $\sum p_i = 1$.

Note. A constraint equation has the form $g(x_1, x_2, \ldots, x_t) = \text{constant}$. For example, if we require $x = y$, the constraint function $g(x,y)$ would be:

$$g(x,y) = x - y = 0.$$

Because x and y are related through the equation $g(x,y) = \text{constant}$, they are **not** independent variables. To satisfy both the extremum equation and the constraint equation, put the constraint equation into differential form and combine it with the extremum equation.

$$\mathrm{d}g = \left(\frac{\partial g}{\partial x}\right)_y \mathrm{d}x + \left(\frac{\partial g}{\partial y}\right)_x \mathrm{d}y = 0.$$

In this example,

$$\mathrm{d}g = 1 \cdot \mathrm{d}x - 1 \cdot \mathrm{d}y = 0 \rightarrow \mathrm{d}x = \mathrm{d}y.$$

If the extremum equation is

$$\mathrm{d}f = 0 = \left(\frac{\partial f}{\partial x}\right)_y \mathrm{d}x + \left(\frac{\partial f}{\partial y}\right)_x \mathrm{d}y,$$

then if we replace $\mathrm{d}y$ by $\mathrm{d}x$ in the extremum equation, we get:

$$\mathrm{d}f = 0 = \left[\left(\frac{\partial f}{\partial x}\right)_y + \left(\frac{\partial f}{\partial y}\right)_x\right] \mathrm{d}x = 0,$$

which gives

$$\left(\frac{\partial f}{\partial x}\right)_y = -\left(\frac{\partial f}{\partial y}\right)_x.$$

Solving this equation identifies the point that is both an extremum of f and also satisfies $g(x, y) = \text{constant}$.

2.3.1 Extrema with constraints: Method of Lagrange Multipliers

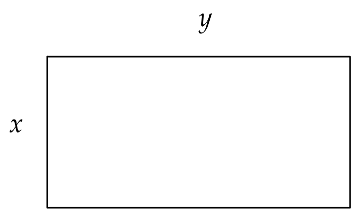
Suppose you want to find the extremum of $f(x,y)$ subject to the constraint $g(x,y) = \text{constant}$.

It can be shown that the derivatives of f and g need only be the same to within an arbitrary constant λ, called the *Lagrange multiplier*:

$$\left(\frac{\partial f}{\partial x}\right)_y = \lambda \left(\frac{\partial g}{\partial x}\right)_y \quad \text{and} \quad \left(\frac{\partial f}{\partial y}\right)_x = \lambda \left(\frac{\partial g}{\partial y}\right)_x. \tag{2.2}$$

The values $x = x^\star$ and $y = y^\star$ that satisfy (2.2) are at the extremum of f and satisfy the constraint.

Example. Suppose you want to find a rectangle of the largest possible area that is 40 cm in circumference.



Mathematically, find the extremum of

$$f(x,y) = xy,$$

with the constraint equation

$$g(x,y) = 2x + 2y = 40.$$

Use Lagrange multipliers to maximize f subject to g:

$$\left(\frac{\partial f}{\partial x}\right) = y, \quad \left(\frac{\partial g}{\partial x}\right) = 2 \quad \Rightarrow \quad y^\star = 2\lambda,$$

$$\left(\frac{\partial f}{\partial y}\right) = x, \quad \left(\frac{\partial g}{\partial y}\right) = 2 \quad \Rightarrow \quad x^\star = 2\lambda.$$

Substituting these into the constraint equation and solving for λ gives $\lambda = 5$ and $x^\star = y^\star = 10$.

Note. For the extremum of $f(x_1, x_2, \ldots, x_t)$ subject to more than one constraint, $g(x_1, \ldots, x_t) = c_1$ and $h(x_1, \ldots, x_t) = c_2$, etc., where the c_i are constants, the Lagrange multiplier method gives the solutions

$$\begin{aligned}
\left(\frac{\partial f}{\partial x_1}\right) - \lambda\left(\frac{\partial g}{\partial x_1}\right) - \beta\left(\frac{\partial h}{\partial x_1}\right) - \cdots &= 0, \\
\left(\frac{\partial f}{\partial x_2}\right) - \lambda\left(\frac{\partial g}{\partial x_2}\right) - \beta\left(\frac{\partial h}{\partial x_2}\right) - \cdots &= 0, \\
&\vdots \\
\left(\frac{\partial f}{\partial x_t}\right) - \lambda\left(\frac{\partial g}{\partial x_t}\right) - \beta\left(\frac{\partial h}{\partial x_t}\right) - \cdots &= 0,
\end{aligned}$$

where $\lambda, \beta, \ldots$ are the Lagrange multipliers for each constraint. Each multiplier is found from its appropriate constraint equation.

An alternative representation:

$$\mathrm{d}(f - \lambda g - \beta h) = \sum_{i=1}^{t} \left[\left(\frac{\partial f}{\partial x_i}\right) - \lambda\left(\frac{\partial g}{\partial x_i}\right) - \beta\left(\frac{\partial h}{\partial x_i}\right)\right] \mathrm{d}x_i = 0\,. \tag{2.3}$$

2.3.2 Maximizing entropy of an isolated system

Back to entropy, we had

$$\frac{S}{k} = -\sum_i p_i \ln p_i\,.$$

This function can only have a maximum.

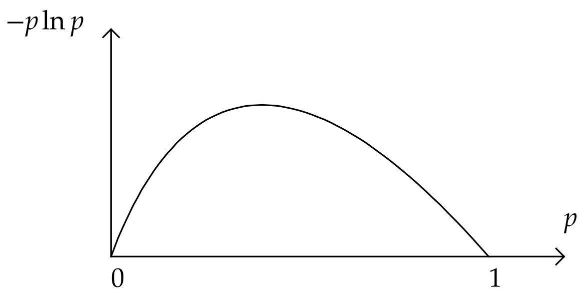


Figure 2.3. $-p\ln p$ versus p has a maximum.

The only (trivial) constraint is normalization: $g = \sum_{i=1}^{t} p_i = 1$. For a simple case of $t = 2$ states,

$$S = -k\,(p_1 \ln p_1 + p_2 \ln p_2)\,.$$

$$g = p_1 + p_2 = 1.$$

Calculate partial derivatives:

$$\frac{1}{k}\left(\frac{\partial S}{\partial p_1}\right)_{p_2} = -\left(\ln p_1 + p_1 \frac{1}{p_1}\right) = -\ln p_1 - 1, \quad \left(\frac{\partial g}{\partial p_1}\right)_{p_2} = 1,$$

$$\frac{1}{k}\left(\frac{\partial S}{\partial p_2}\right) = -\ln p_2 - 1, \quad \left(\frac{\partial g}{\partial p_2}\right)_{p_1} = 1.$$

Via eq. (2.2),

$$\begin{cases} -1 - \ln p_1 - \lambda = 0 \\ -1 - \ln p_2 - \lambda = 0 \end{cases} \quad \Rightarrow \quad p_1^\star = p_2^\star = \mathrm{e}^{-1-\lambda}\,.$$

Plug these to the constraint equation to get

$$g = 2\,\mathrm{e}^{-1-\lambda} = 1 \quad \Rightarrow \quad \mathrm{e}^{-1-\lambda} = \frac{1}{2} \quad \Rightarrow \quad p_1 = p_2 = \frac{1}{2}\,.$$

Conclusion. Maximum entropy predicts a *flat* distribution of states. All states are equally likely.

In non-isolated systems, there may be additional constraints such as conservation of energy, conservation of mass, etc.

2.3.3 Maximizing entropy with an energy constraint

Roll a dice having t sides, with faces numbered $i = 1, 2, 3, \ldots, t$. You don't know the distribution of outcomes of each face, but you know the total score after N rolls. You want to predict the distribution function.

First, let's generalize our dice problem. instead of having the numbers $i = \mathbf{1}, \mathbf{2}, \ldots, \mathbf{6}$ painted on its six sides, the die has a more general set of numbers painted on its t sides. When side i appears face up, the score is ε_i. The total score after N rolls will be $E = \sum_{i=1}^{t} \varepsilon_i n_i$, where n_i is the number of times that you observe face i.

Let $p_i = n_i / N$ represent the fraction of the N rolls on which you observe face i. The average score per roll, $\langle \varepsilon \rangle$ is:

$$\langle \varepsilon \rangle = \frac{E}{N} = \sum_{i=1}^{t} p_i \varepsilon_i\,.$$

What is the distribution of outcomes $(p_1^\star, p_2^\star, \ldots, p_t^\star)$ consistent with the average score $\langle \varepsilon \rangle$? We seek the distribution that maximizes the entropy, subject to two constraints: (1) that all probabilities sum to one, and (2) that the average score agrees with the observed value $\langle \varepsilon \rangle$.

$$g(p_1, p_2, \ldots, p_t) = \sum_{i=1}^{t} p_i = 1 \quad \Rightarrow \quad \sum_{i=1}^{t} \mathrm{d}p_i = 0,$$

$$h(p_1, p_2, \ldots, p_t) = \langle \varepsilon \rangle = \sum_{i=1}^{t} p_i \varepsilon_i \quad \Rightarrow \quad \sum_{i=1}^{t} \varepsilon_i \mathrm{d}p_i = 0.$$

Solve via method of Lagrange multipliers.

$$\left(\frac{\partial S}{\partial p_i}\right) - \alpha\left(\frac{\partial g}{\partial p_i}\right) - \beta\left(\frac{\partial h}{\partial p_i}\right) = 0 \quad \text{for } i = 1, 2, \ldots, t.$$

The partial derivatives are evaluated for each p_i:

$$\left(\frac{\partial S}{\partial p_i}\right) = -1 - \ln p_i, \quad \left(\frac{\partial g}{\partial p_i}\right) = 1, \quad \left(\frac{\partial h}{\partial p_i}\right) = \varepsilon_i\,.$$

Substitute into the above equation to get t equations of the form

$$-1 - \ln p_i^\star - \alpha - \beta\,\varepsilon_i = 0.$$

Solve for each $p_i^\star$ to get

$$p_i^\star = \mathrm{e}^{-1-\alpha-\beta\varepsilon_i}\,.$$

To eliminate α, use the normalization constraint to divide both sides by one. The result is an *exponential distribution law*:

$$p_i^\star = \frac{p_i^\star}{\sum_{i=1}^t p_i^\star} = \frac{\mathrm{e}^{(-1-\alpha)}\,\mathrm{e}^{-\beta\varepsilon_i}}{\sum_{i=1}^t \mathrm{e}^{(-1-\alpha)}\,\mathrm{e}^{-\beta\varepsilon_i}} = \frac{\mathrm{e}^{-\beta\varepsilon_i}}{\sum_{i=1}^t \mathrm{e}^{-\beta\varepsilon_i}}\,. \tag{2.4}$$

In Statistical Mechanics, the average score is translated to *average energy* of the system, and eq. (2.4) is called the *Boltzmann distribution law*. The quantity in the denominator is called the *partition function*, q:

$$q \equiv \sum_{i=1}^t \mathrm{e}^{-\beta\varepsilon_i} \tag{2.5}$$

Using the score constraint and the above, we get

$$\langle\varepsilon\rangle = \sum_{i=1}^t \varepsilon_i p_i^\star = \frac{1}{q}\sum_{i=1}^t \varepsilon_i \mathrm{e}^{-\beta\varepsilon_i}\,. \tag{2.6}$$

16.05.23 lec 3

3 Toward the Fundamental Thermodynamic Equations

3.1 Definitions and conventions

Definition 3.1

Energy.

Energy is system property, and describes the capacity of a system to perform work. Energy is **conserved** and can flow, so that capacity to perform work can be moved from one place to another.

Energy is ubiquitous, and can take any form, e.g. mechanical, potential, or electrical.

Definition 3.2

The First Law of Thermodynamics. (JR Von Mayer, 1842.)

The internal energy ΔU of a system changes when it takes up or gives off heat q or work w:

$$\Delta U = q + w. \tag{3.1}$$

The internal energy is conserved; if ΔU increases in the system, the energy decreases in the surroundings.

General conventions:

- If heat goes **into** the system, $q > 0$, and vice versa.
- If work is done **on** the system, $w > 0$, and vice versa.

3.2 Energy is quantized

Quantum theory debuted in the 20th century showed that energies of atoms and molecules are quantized. Each particle has discrete amounts of energy associated with each of its allowed degrees of freedom, some of which are translations rotations, vibrations, and electronic excitations.

The allowed energies for a given system are indicated in *energy-level diagrams*, and they predict thermodynamic properties.

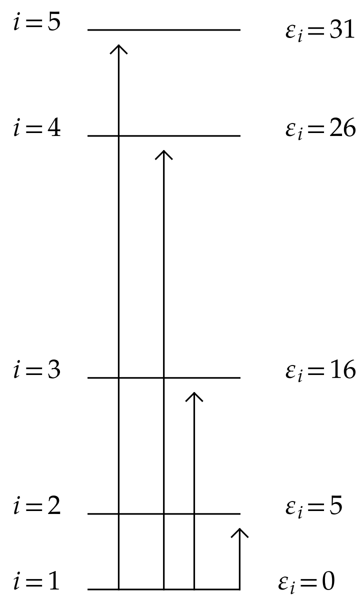


Figure 3.1. An exemplary 5-level energy diagram. Each state has its corresponding energy, and particles may occupy the different states.

For simple systems of independent, non-interacting particles, such as ideal gases, we can express the total *internal energy* of a thermodynamic system as the sum of the particle energies:

$$U = \sum_i N_i \varepsilon_i, \tag{3.2}$$

where ε_i is the energy of any particle at level i and N_i is the number of particles at energy level i. When the total internal energy of a system is increased by heating it, the energy levels **do not** change, but the populations $\{N_i\}$ change.

3.3 Flow of heat

What drives molecules or materials to exchange energy? Why does heat flow? The First Law of Thermodynamics cannot explain this phenomenon. Heat flow is a consequence of the tendency toward maximum multiplicity, which is the *Second Law of Thermodynamics*.

Previously, we've seen that lattice models predict that:

- Gases expand because the multiplicity W increases with volume V.
    - The dependence of W on V defines the force called *pressure*.
- Particles mix because the multiplicity W increases as the particle segregation decreases.
    - The tendency to mix defines the *chemical potential*.

These are both manifestations of the Second Law of Thermodynamics.

As a system absorbs heat, the internal energy increases, and the possible distributions of particles within the energy states (that sum up to the total internal energy) increases.

Why, then, does heat flow from hot objects to cold ones?

Consider two systems, A and B, both having two possible energy states: $\varepsilon_1 = 0, \varepsilon_2 = 1$. Each system has 10 particles. Suppose that system A starts with energy $U_A = 2$ and system B starts with energy $U_B = 4$.

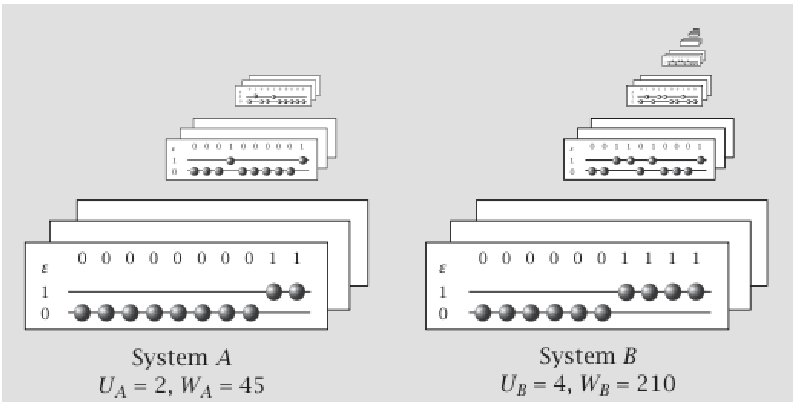


Figure 3.2. Energy-level diagrams for two different systems with 10 particles each. System B has a higher internal energy, and thus a greater multiplicity of states.

What would happen if we let the systems exchange energy? The multiplicities $W(U)$ of the isolated systems are given by binomial distribution.

$$W_A = \frac{10!}{2!\,8!} = 45, \qquad W_B = \frac{10!}{4!\,6!} = 210.$$

The combined multiplicity of the isolated systems is $W_{AB} = W_A W_B = 9450$. When the systems exchange energy, their total internal energy must be conserved—$U_A + U_B = \text{const}$.

One possibility of heat flow is $U_A = 3, U_B = 3$. The total multiplicity would be

$$W_{AB} = \frac{10!}{3!7!}\frac{10!}{3!7!} = 14{,}400.$$

As the multiplicity increased, we can infer that heat flows from B to A, in this case to equalize energies. Consider an alternative outcome of heat flow: $U_A = 1, U_B = 5$. The total multiplicity in this case is

$$W_{AB} = \frac{10!}{1!9!}\frac{10!}{5!5!} = 2520.$$

The principle of maximal multiplicity predicts that heat flow from cold to hot objects is unlikely.

Note. The tendency to maximize multiplicity does not always result in a draining of energy from higher to lower.

Consider two systems ($\varepsilon_1 = 0, \varepsilon_2 = 1$) having the same energies, but different particle numbers. Suppose system A has $N_A = 10, U_A = 2$ and system B has $N_B = 4, U_B = 2$.

The total multiplicity of the isolated systems is

$$W_{AB} = W_A W_B = \frac{10!}{2!8!}\frac{4!}{2!2!} = 270.$$

Now let the systems thermally interact. If the larger system absorbs energy from the smaller one, so that $U_A = 3, U_B = 1$,

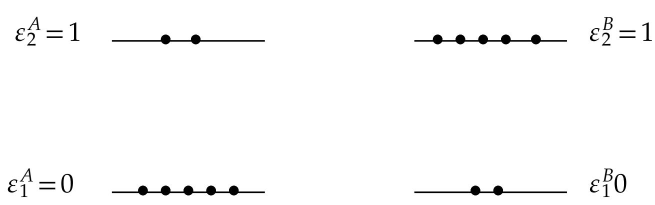
$$W_{AB} = W_A W_B = \frac{10!}{3!7!}\frac{4!}{1!3!} = 480,$$

the multiplicity increased.

We shall later see that the concept of *temperature* describes the driving force for energy exchange. The tendency toward maximum multiplicity is a tendency toward equal temperatures.

Example. *Calculating maximized multiplicity distribution of particles.*

Let there be two non-interacting systems, A and B, each having two possible states:



$N_A = 1000$ $\quad$ $N_B = 1000$

$N_A U_A = 200$ $\quad$ $N_B U_B = 900$

The total energy of the system per particle is $\frac{U}{N} = \sum_i p_i \varepsilon_i$.

What is the distribution of states within each system?

For system A,

$$U_A = \frac{200}{1000} = \varepsilon_1^A p_1^A + \varepsilon_2^A p_2^A = p_2^A\,,$$

$$p_1^A = 1 - p_2^A = 0.8.$$

Similarly, for system B,

$$U_B = \frac{900}{1000} = \varepsilon_1^B p_1^B + \varepsilon_2^B p_2^B = p_2^B\,,$$

$$p_1^B = 1 - p_2^B = 0.1\,.$$

Now we let the systems exchange energy. What would be the new equilibrium distribution of states?

According to the First Law of Thermodynamics, energy is conserved.

$$U_{\text{total}} = U_A + U_B = \frac{N_A U_A + N_B U_B}{N_A + N_B} = 0.55. \tag{3.3}$$

This is a constraint equation!

We wish to maximize the entropy.

$$\frac{S}{k} = -\sum p_i \ln p_i = -p_1^A \ln p_1^A - p_2^A \ln p_2^A - p_1^B \ln p_1^B - p_2^B \ln p_2^B\,. \tag{3.4}$$

The other constraint equations:

$$g = p_1^A + p_2^A + p_1^B + p_2^B = 1 \tag{3.5}$$

$$h = \varepsilon_1^A p_1^A + \varepsilon_2^A p_2^A + \varepsilon_1^B p_1^B + \varepsilon_2^B p_2^B \tag{3.6}$$

According to Boltzmann's law,

$$p_i^\star = \frac{\mathrm{e}^{-\beta\varepsilon_i}}{q}, \qquad q = \mathrm{e}^{-\beta\varepsilon_1^A} + \mathrm{e}^{-\beta\varepsilon_2^A} + \mathrm{e}^{-\beta\varepsilon_1^B} + \mathrm{e}^{-\beta\varepsilon_2^B} = 2\,(1 + \mathrm{e}^{-\beta})\,.$$

Accordingly,

$$p_1^{A,\star} = p_1^{B,\star} = \frac{1}{2\,(1+\mathrm{e}^{-\beta})}, \qquad p_2^{B,\star} = p_2^{A,\star} = \frac{\mathrm{e}^{-\beta}}{2\,(1+\mathrm{e}^{-\beta})}\,. \tag{3.7}$$

What is the value of β? Insert into the constraint equation, eq. (3.4).

$$U_{\text{total}} = 2\left(0 \cdot \frac{1}{2\,(1+\mathrm{e}^{-\beta})} + 1 \cdot \frac{\mathrm{e}^{-\beta}}{2\,(1+\mathrm{e}^{-\beta})}\right) = 0.55,$$

$$\beta = -\ln \frac{U_{\text{total}}}{1 - U_{\text{total}}} = -0.2.$$

Plug β into eq. (3.5) to get (after normalization per system) $p_1^{A,\star} = p_1^{B,\star} = 0.54$ and $p_2^{A\star} = p_2^{B,\star} = 0.46$. We can infer that particles in system A moved to the higher energy state, and particles in system B moved to the lower energy state.

3.4 Thermodynamic systems and the fundamental thermodynamic equations

A thermodynamic system is a collection of matter in any form, delineated from its surroundings by (real or imaginary) boundaries. Defining the boundaries is important, as its specifies thermodynamic properties of the system.

Types of systems:

Open system. An open system can exchange energy, volume, and matter with its surroundings.

Closed system. Energy can cross the boundary of a closed system, but matter cannot.

Isolated system. Energy and matter cannot cross the boundaries of an isolated system. Also, volume does not change. The total internal energy of an isolated system is constant.

Types of boundaries:

Semipermeable membrane. A semipermeable membrane is a boundary that restricts the flow of some kinds of particle, while allowing others to cross.

Adiabatic boundary. Prevents heat from flowing between the system and its surroundings.

We've seen that maximizing entropy with respect to different parameters, such as volume, number of particles, and internal energy, we predict different phenomena, such as expansion of gas, change in composition, and heat flow, respectively.

Many systems, however, allow multiple parameters to change simultaneously. The *fundamental thermodynamic equation* for entropy is multivariate: $S = S(U, V, \mathbf{N})$. In this formulation, the energy U, the volume V, and the number of particles of different categories $N_1, N_2, \ldots, N_M$ are all free to change.

Note that history first conjured the relations in the form of energy: $U = U(S, V, \mathbf{N})$. The fundamental definitions of pressure, chemical potential, and temperature are based on the form of $U = U(S, V, \mathbf{N})$. The microscopic driving forces, though, are better understood in terms of the entropy equation $S = S(U, V, \mathbf{N})$, so we need a way to switch between them. In any case, both formulations completely specify the state of a system.

Note. Thermodynamics does not en-tell the specific mathematical dependence of S on $(U, V, \mathbf{N})$ or U on $(S, V, \mathbf{N})$. *Equations of state*, which come from microscopic models or experiments, specify interrelations among these variables.

In this section, we transition from Statistical Mechanics to Classical Thermodynamics.

$$S = -k \sum p_i \ln p_i \quad \rightarrow \quad S = f(U, V, \mathbf{N})\,.$$

3.5 The fundamental equations define the thermodynamic driving forces

According to the fundamental entropy equation,

$$\mathrm{d}S = \left(\frac{\partial S}{\partial U}\right)_{V,\mathbf{N}} \mathrm{d}U + \left(\frac{\partial S}{\partial V}\right)_{U,\mathbf{N}} \mathrm{d}V + \sum_{j=1}^{M} \left(\frac{\partial S}{\partial N_j}\right)_{U,V,N_{i\neq j}} \mathrm{d}N_j\,. \tag{3.8}$$

Similarly, using the fundamental energy equation,

$$\mathrm{d}U = \left(\frac{\partial U}{\partial S}\right)_{V,\mathbf{N}} \mathrm{d}S + \left(\frac{\partial U}{\partial V}\right)_{S,\mathbf{N}} \mathrm{d}V + \sum_{j=1}^{M} \left(\frac{\partial U}{\partial N_j}\right)_{S,V,N_{i\neq j}} \mathrm{d}N_j \tag{3.9}$$

It turns out that the partial derivatives in (3.9) correspond to measurable physical quantities.

Definition 3.3

Temperature, pressure, and chemical potential.

$$T = \left(\frac{\partial U}{\partial S}\right)_{V,\mathbf{N}}, \qquad p = -\left(\frac{\partial U}{\partial V}\right)_{S,\mathbf{N}}, \qquad \mu_j = \left(\frac{\partial U}{\partial N_j}\right)_{S,V,N_{i\neq j}}. \tag{3.10}$$

Note. T, P and μ are intensive properties, that are conjugate to extensive quantities, U, S, V and N.

Substituting (3.10) into (3.9) gives the *differential form of the fundamental energy equation*:

$$\mathrm{d}U = T\,\mathrm{d}S - p\,\mathrm{d}V + \sum_{j=1}^{M} \mu_j \mathrm{d}N_j. \tag{3.11}$$

Alternatively, substituting into (3.8) gives the *differential form of the fundamental entropy equation*:

$$\mathrm{d}S = \left(\frac{1}{T}\right)\mathrm{d}U + \left(\frac{p}{T}\right)\mathrm{d}V - \sum_{j=1}^{M}\left(\frac{\mu_j}{T}\right)\mathrm{d}N_j. \tag{3.12}$$

From which we identify,

$$\frac{1}{T} = \left(\frac{\partial S}{\partial U}\right)_{V,\mathbf{N}}, \qquad \frac{p}{T} = \left(\frac{\partial S}{\partial V}\right)_{U,\mathbf{N}}, \qquad \frac{\mu_j}{T} = -\left(\frac{\partial S}{\partial N_j}\right)_{U,V,N_{i\neq j}}. \tag{3.13}$$

We shall later see how equations (?) and (?) can be used to identify states of equilibrium.

4 Laboratory conditions and free energies

Up until now we dealt with systems with known energy that exchange it across their boundaries. These systems tend toward states of maximum entropy. That logic helped explain gas expansion, particle mixing, and the interconversion of heat. We now wish to explore a different class of systems, such as test tubes in laboratory heat baths, processes open to the air, and processes in biological systems. In these systems, it is not work or heat flow that is controlled at the boundaries, but temperature and pressure. This change requires new thermodynamic quantities—the *free energy* and the *enthalpy*— and reformulated extremum principles. Systems held at constant temperature do not tend toward their states of maximum entropy. They tend toward their states of *minimum free energy*.

When an intensive variable, such as T, p, or μ is controlled or measured at the boundary, it means that the conjugate variables, U, V, or N, can exchange freely back and forth across the boundary with the *bath*, the external reservoir that is large enough that it can hold V, p, or μ fixed, no matter what happens in the system. Such exchanges are called *fluctuations*.

When T is constant, heat can exchange between the system and the surroundings, so the energy of the system fluctuates. When p is held constant, the volume fluctuates. When μ is constant, a particle bath is in contact with the system—particles leave or enter the system to and from the particle bath. In this case, the number of particles in the system can fluctuate.

Consider a process in a system that we call the *tset tube*, immersed in a *heat bath*. a heat bath refers to any surroundings of a system that hold the temperature of the system constant. If the combined test tube plus heat bath are isolated from the greater surroundings, equilibrium will be the state of maximum entropy for the total system. However, we are only interested in what happens in the test tube itself. We need a new extremum principle that applies to the test tube, where the independent variables are $(T, V, \mathbf{N})$.

Note. If the extremum of a function such as $S(U)$ predicts equilibrium, the variable U is called the *natural variable* of S. T is not a natural variable of S. Now we show that $(T, V, \mathbf{N})$ are natural variables of a function F, the *Helmholtz free energy*.

An extremum in $F(T, V, \mathbf{N})$ predicts equilibria in systems that are constrained to constant temperature at their boundaries.

4.1 Free energy defines another extremum principle

4.1.1 The Helmholtz free energy

Consider a process inside a test tube, sealed so that it has constant volume V and no interchange of its $\mathbf{N}$ particles with the surroundings. A heat bath holds the test tube at constant temperature T.

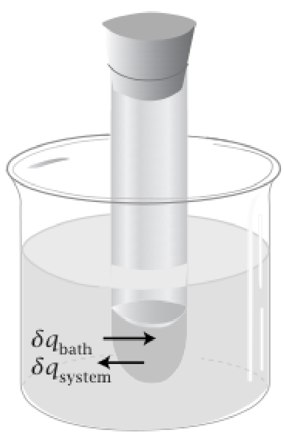


Figure 4.1. A heat bath is a reservoir that holds the system (the test tube in this case) at constant temperature by allowing heat flow in or out, as required. The properties that do not change inside the system are $(T, V, \mathbf{N})$.

The process inside the test tube might be complex. It might vary in rate from a quasi-static process to an explosion. It might or might not involve chemical or phase changes. It might give off or absorb heat. Processes within the test tube will influence the heat bath only through heat exchange, because its volume does not change and no work is done.

If the combined system (sub-system plus heat bath) is isolated, equilibrium will be the state of maximum entropy of the combined system. Any change toward equilibrium must increase the entropy of the combined system, $\mathrm{d}S_{\text{combined}} \geqslant 0$.

Because the entropy is extensive,

$$\mathrm{d}S_{\text{combined}} = \mathrm{d}S_{\text{sys}} + \mathrm{d}S_{\text{bath}} \geqslant 0. \tag{4.1}$$

Since the combined system is isolated,

$$\mathrm{d}U_{\text{bath}} + \mathrm{d}U_{\text{sys}} = 0. \tag{4.2}$$

We wish to relate $\mathrm{d}S_{\text{bath}}$ to some property of the test tube system. Use (?) to get, in our case,

$$\mathrm{d}S_{\text{bath}} = \left(\frac{1}{T}\right)\mathrm{d}U_{\text{bath}}. \qquad (\mathbf{N}, V) = \text{constant} \tag{4.3}$$

Combine with eq. (4.2) to get

$$\mathrm{d}S_{\text{bath}} = -\frac{\mathrm{d}U_{\text{sys}}}{T}. \tag{4.4}$$

Substitute eq. (4.4) into (4.1) to get

$$\mathrm{d}S_{\text{sys}} = -\frac{\mathrm{d}U_{\text{sys}}}{T} \geqslant 0 \qquad \Longrightarrow \qquad \mathrm{d}U_{\text{sys}} - T\,\mathrm{d}S_{\text{sys}} \leqslant 0. \tag{4.5}$$

We got an expression describing the approach to equilibrium in terms of the test tube sub-system alone. Define a quantity F, the *Helmholtz free energy*:

Definition 4.1

Helmholtz free energy.

$$F \equiv U - TS. \tag{4.6}$$

$$\mathrm{d}F = \mathrm{d}U - T\,\mathrm{d}S - S\,\mathrm{d}T. \tag{4.7}$$

Comparison of eq. (4.7) with eq. (4.5) shows that when a system in which $(T, V, \mathbf{N})$ are constant is at equilibrium, the quantity F is at minimum. It also follows from eq. (4.6) that to minimize F, the system in the tube will tend toward *both* high entropy and low energy, depending on the temperature. At high temperatures, the entropy dominates. At low temperatures, the energy dominates.

4.1.2 The fundamental equation for the Helmholtz free energy

Just as the functional form $S(U, V, \mathbf{N})$ implies a fundamental entropy equation for $\mathrm{d}S$, the form $F(T, V, \mathbf{N})$ implies a fundamental equation for $\mathrm{d}F$:

$$\mathrm{d}F = \mathrm{d}(U - TS) = \mathrm{d}U - T\,\mathrm{d}S - S\,\mathrm{d}T.$$

Substitute the fundamental energy equation (?) into (4.7) to get:

$$\begin{aligned} \mathrm{d}F &= \left(T\,\mathrm{d}S - p\,\mathrm{d}V + \sum_{j=1}^{M} \mu_j \mathrm{d}N_j\right) - T\,\mathrm{d}S - S\,\mathrm{d}T \\ &= -S\,\mathrm{d}T - p\,\mathrm{d}V + \sum_{j=1}^{M} \mu_j \mathrm{d}N_j. \end{aligned} \tag{4.8}$$

Because $\mathrm{d}F$ is also defined by its partial derivative expression,

$$\mathrm{d}F = \left(\frac{\partial F}{\partial T}\right)_{V,\mathbf{N}} \mathrm{d}T + \left(\frac{\partial F}{\partial V}\right)_{T,\mathbf{N}} \mathrm{d}V + \sum_{j=1}^{M} \left(\frac{\partial F}{\partial N_j}\right)_{V,T,N_{i\neq j}} \mathrm{d}N_j, \tag{4.9}$$

we get additional thermodynamic relations by comparing eq. (4.9) with (4.8):

$$S = -\left(\frac{\partial F}{\partial T}\right)_{V,\mathbf{N}}, \qquad p = -\left(\frac{\partial F}{\partial V}\right)_{T,\mathbf{N}}, \qquad \mu_j = \left(\frac{\partial F}{\partial N_j}\right)_{V,T,N_{i\neq j}}. \tag{4.10}$$

Note. We derived $F(T,V,\mathbf{N})$ from $S(U,V,\mathbf{N})$ by physical arguments. You can also switch from one set of independent variables to another by purely mathematical arguments, called *Legendre transforms.*

A function $y=f(x)$ can be described as a list of pairs $(x_1,y_1),(x_2,y_2),\ldots$. You can express the same function instead as a list of different pairs: the slopes $c(x)$ and the intercepts $b(x)$: (c_1,b_1), $(c_2,b_2),\ldots$.

For a small change $\mathrm{d}x$, the change $\mathrm{d}y$ in the function can be described by the slope $c(x)$ at that point:

$$\mathrm{d}y=\left(\frac{\partial y}{\partial x}\right)\mathrm{d}x=c(x)\,\mathrm{d}x. \tag{4.11}$$

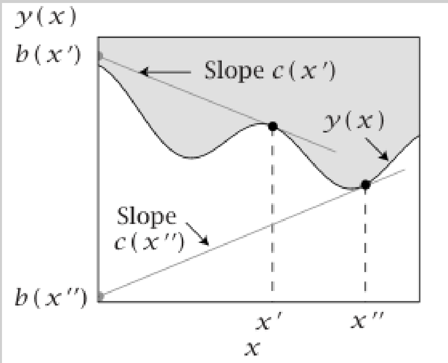


Figure 4.2. To create the Legendre transform, a function $y(x)$ is expressed as a tangent slope function $c(x)$, and a tangent intercept function $b(x)$. The tangent slopes and intercepts of points x' and x'' are shown here.

Actually, the full function $y(x)$ (not just small changes) can be regarded as a set of slopes and intercepts—one slope $c(x)$ and one intercept $b(x)$ for each point x:

$$y(x)=c(x)\,x+b(x) \qquad \Longrightarrow \qquad b(x)=y(x)-c(x)\,x. \tag{4.12}$$

We are interested in the function that expresses the series of intercepts versus slopes, $b(c)$. To see how small changes in the slope c lead to small changes in the intercept b, take the differential of (4.12) and substitute in (4.11), to get

$$\mathrm{d}b=\mathrm{d}y-c\,\mathrm{d}x-x\,\mathrm{d}c=-x\,\mathrm{d}c. \tag{4.13}$$

Now generalize to a multivariate function $y=y(x_1,x_2,x_3)$. The differential element is

$$\mathrm{d}y=c_1\,\mathrm{d}x_1+c_2\,\mathrm{d}x_2+c_3\,\mathrm{d}x_3, \tag{4.14}$$

where

$$c_1=\left(\frac{\partial y}{\partial x_1}\right)_{x_2,x_3}, \qquad c_2=\left(\frac{\partial y}{\partial x_2}\right)_{x_1,x_3}, \qquad c_3=\left(\frac{\partial y}{\partial x_3}\right)_{x_1,x_2}.$$

We want the intercept function b_1 along the x_1 axis:

$$b_1(c_1,x_2,x_3)=y-c_1x_1\,. \tag{4.15}$$

Take the differential of (4.15) and substitute (4.14) to get

$$\mathrm{d}b_1=\mathrm{d}y-c_1\,\mathrm{d}x_1-x_1\,\mathrm{d}c_1=-x_1\,\mathrm{d}c_1+c_2\,\mathrm{d}x_2+c_3\,\mathrm{d}x_3\,. \tag{4.16}$$

From (4.16), you can see that

$$x_1=-\left(\frac{\partial b_1}{\partial c_1}\right)_{x_2,x_3}, \qquad c_2=\left(\frac{\partial b_1}{\partial x_2}\right)_{c_1,x_3}, \qquad c_3=\left(\frac{\partial b_1}{\partial x_3}\right)_{c_1,x_2}. \tag{4.17}$$

4.1.3 The enthalpy

The enthalpy is a function of the natural variables $(S, p, \mathbf{N})$. Enthalpy is seldom used an an extremum principle, because it is not usually convenient to control the entropy. However, it can be obtained from calorimetry experiments, and it gives an experimental route to the Gibbs free energy, which is of central importance in chemistry and biology.

To find the enthalpy, you could reason in the same way as we did for the Helmholtz free energy, but instead let's use a simple math argument. Start with the internal energy $U(S, V, \mathbf{N})$. We seek to replace a $\mathrm{d}V$ term in the energy function with a $\mathrm{d}p$ term to get the enthalpy function $\mathrm{d}H$. Add a pV term to the energy so that when you differentiate it, the $\mathrm{d}V$ term will disappear and a $\mathrm{d}p$ term will appear:

$$H = H(S,p,\mathbf{N}) \equiv U + pV. \tag{4.18}$$

Now differentiate:

$$\mathrm{d}H = \mathrm{d}U + p\,\mathrm{d}V + V\,\mathrm{d}p. \tag{4.19}$$

Substitute eq. (3.11) into (4.19) to get

$$\mathrm{d}H = T\,\mathrm{d}S - p\,\mathrm{d}V + \sum_{j=1}^{M} \mu_j \mathrm{d}N_j + p\,\mathrm{d}V + V\,\mathrm{d}p$$

$$\mathrm{d}H = T\,\mathrm{d}S + V\,\mathrm{d}p + \sum_{j=1}^{M} \mu_j \mathrm{d}N_j\,. \tag{4.20}$$

4.1.4 The Gibbs free energy

The Gibbs free energy G is a function of $(T, p, \mathbf{N})$. Constant temperature and pressure are the easiest constraints to impose in the laboratory, because the atmosphere provides them.

$G = G(T, p, \mathbf{N})$ has a minimum at equilibrium. To find the fundamental equation, start with the enthalpy, $H = H(S, p, \mathbf{N})$. You want to replace the $\mathrm{d}S$ term with a $\mathrm{d}T$ term in eq. (4.20). Define a function G:

$$G \equiv H - TS. \tag{4.21}$$

The total differential $\mathrm{d}G$ is

$$\mathrm{d}G = \mathrm{d}H - T\,\mathrm{d}S - S\,\mathrm{d}T. \tag{4.22}$$

Substitute eq. (4.20) into (4.22) to get

$$\mathrm{d}G = -S\,\mathrm{d}T + V\,\mathrm{d}p + \sum_{j=1}^{M} \mu_j \mathrm{d}N_j\,. \tag{4.23}$$

Similarly to $F(T, V, \mathbf{N})$, if a process occurs in a test tube held at constant pressure and temperature, it will be at equilibrium when the Gibbs free energy is at minimum.

Note. Equilibrium is the state at which the entropy of the combined system *plus* surroundings is at maximum. However, for the test tube system itself, which is at constant $(T,p,\mathbf{N})$, equilibrium occurs when the Gibbs free energy is at minimum.

$\mathrm{d}G$ can be expressed as

$$\mathrm{d}G=\left(\frac{\partial G}{\partial T}\right)_{p,\mathbf{N}}\mathrm{d}T+\left(\frac{\partial G}{\partial p}\right)_{T,\mathbf{N}}\mathrm{d}p+\sum_{j=1}^{M}\left(\frac{\partial G}{\partial N_j}\right)_{p,T,N_{i\neq j}}\mathrm{d}N_j\,. \tag{4.24}$$

So,

$$S=-\left(\frac{\partial G}{\partial T}\right)_{p,\mathbf{N}},\qquad V=\left(\frac{\partial G}{\partial p}\right)_{T,\mathbf{N}},\qquad \mu_j=\left(\frac{\partial G}{\partial N_j}\right)_{p,T,N_{i\neq j}}. \tag{4.25}$$

Note. For equilibrium phase changes, which occur at constant temperature, pressure, and particle number, the Gibbs free energy does not change.

4.1.5 Summary

Function	Fundamental equation	Definition
$U(S,V,\mathbf{N})$	$\mathrm{d}U=T\,\mathrm{d}S-p\,\mathrm{d}V+\sum_j \mu\,\mathrm{d}N_j$	
$S(U,V,\mathbf{N})$	$\mathrm{d}S=\left(\frac{1}{T}\right)\mathrm{d}U+\left(\frac{p}{T}\right)\mathrm{d}V-\sum_j\left(\frac{\mu_j}{T}\right)\mathrm{d}N_j$	
$H(S,p,\mathbf{N})$	$\mathrm{d}H=T\,\mathrm{d}S+V\,\mathrm{d}p+\sum_j \mu_j\mathrm{d}N_j$	$H=U+pV$
$F(T,V,\mathbf{N})$	$\mathrm{d}F=-S\,\mathrm{d}T-p\,\mathrm{d}V+\sum_j \mu_j\mathrm{d}N_j$	$F=U-TS$
$G(T,p,\mathbf{N})$	$\mathrm{d}G=-S\,\mathrm{d}T+V\,\mathrm{d}p+\sum_j \mu_j\mathrm{d}N_j$	$G=H-TS=F+pV$

Table 4.1. Fundamental equations and their natural variables.

4.2 Uses of internal energy, entropy and enthalpy; heat capacity

These non-fundamental component functions are important because they can be measured in calorimeters, and can be combined to give the fundamental functions, such as $F(T,V,\mathbf{N})$ and $G(T,p,\mathbf{N})$. You measure the temperature dependence of a material' heat capacity $C_V(T)$ in a constant-volume calorimeter. Then, using $\Delta U=\int C_V(T)\,\mathrm{d}T$ and $\Delta S=\int\frac{C_V}{T}\mathrm{d}T$ to find ΔF. Alternatively, if ΔG is desired, then you can measure the heat capacity in a constant pressure calorimeter, and use $\mathrm{d}H=\mathrm{d}U+p\,\mathrm{d}V+V\,\mathrm{d}p=\delta q$. The heat capacity C_p is

$$C_p=\left(\frac{\delta q}{\mathrm{d}T}\right)_p=\left(\frac{\partial H}{\partial T}\right)_p=T\left(\frac{\partial S}{\partial T}\right)_p. \tag{4.26}$$

Rearranging, you get

$$\Delta H(T,p)=\int_{T_A}^{T_B}C_p(T)\,\mathrm{d}T. \tag{4.27}$$

$$\Delta S(T,p)=\int_{T_A}^{T_B}\frac{C_p(T)}{T}\mathrm{d}T.$$

And then use these to calculate ΔG.

4.2.1 The third law of thermodynamics

Suppose you want to know the absolute entropy of a material at a temperature T. You can integrate the heat capacity, $S(T) = \int_0^T (C_V/T')\,dT' + S(0)$, where $S(0)$ is the entropy at absolute zero temperature. The third law of thermodynamics states that $S(0) = 0$ at for a pure perfectly crystalline substance at zero Kelvin.

4.2.2 Thermodynamic cycles

The following example shows how to combine a thermodynamic cycle with heat capacities to compute properties for which measurements are difficult or impossible.

Example. Measuring enthalpies under standard conditions and computing them for other conditions. Suppose you want to know the enthalpy of boiling water, $\Delta H_{\mathrm{boil}(0°C)}$, at the freezing point of water 0°C and $p = 1$ atm. boiling water, $\Delta H_{\mathrm{boil}(0°C)}$, at the freezing point of water 0°C and p = 1 atm. Since you cannot boil water at water's freezing point, why would you want to know that quantity? We will see later that $\Delta H_{\mathrm{boil}(0°C)}$ from the heat capacities of water and steam and the enthalpy of vaporization of water, $\Delta H_{\mathrm{boil}(100°C)}$, under more standard boiling conditions, by using a simple thermodynamic cycle.

The standard state enthalpy has been measured to be $\Delta H_{\mathrm{boil}(100°C)} = 540\,\mathrm{cal\,g^{-1}}$. The heat capacity of steam is $C_p = 0.448\,\mathrm{cal\,K^{-1}\,g^{-1}}$ and the heat capacity of liquid water is $C_p = 1.00\,\mathrm{cal\,K^{-1}\,g^{-1}}$. To obtain $\Delta H_{\mathrm{boil}(0°C)}$, construct the following thermodynamic cycle.

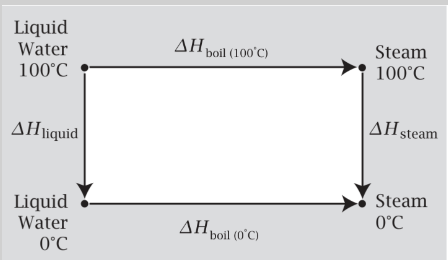


Figure 4.3. A thermodynamic cycle allows you to calculate the enthalpy for boiling water at the freezing temperature of water if you have measured the enthalpy at the boiling temperature.

With the directions of the arrows shown, summing to zero around a cycle means that

$$\Delta H_{\mathrm{boil}(0°C)} = \Delta H_{\mathrm{boil}(100°C)} - \Delta H_{\mathrm{liquid}} + \Delta H_{\mathrm{steam}}\,.$$

Because there is no phase change for the steam or liquid, and because the heat capacities are reasonably independent of temperature, you have

$$\Delta H_{\mathrm{liquid}} = \int_{100}^{0} C_{p,\mathrm{liquid}}\,dT = C_{p,\mathrm{liquid}}\Delta T = \left(1.00\,\frac{\mathrm{cal}}{\mathrm{K\,g}}\right)(-100\,\mathrm{K}) = -100\,\mathrm{cal\,g^{-1}}$$

and

$$\Delta H_{\mathrm{steam}} = \int_{100}^{0} C_{p,\mathrm{steam}}\,dT = C_{p,\mathrm{steam}}\Delta T = \left(0.448\,\frac{\mathrm{cal}}{\mathrm{K\,g}}\right)(-100\,\mathrm{K}) = -44.8\,\mathrm{cal\,g^{-1}}\,.$$

Thus,

$$\Delta H_{0°C} = (540 + 100 - 44.8)\,\mathrm{cal\,g^{-1}} = 585.2\,\mathrm{cal\,g^{-1}}\,.$$

5 The Boltzmann distribution law and partition functions

lec 4. 30.05.23

This chapter deals with modeling of the probability distributions of the energies of atoms and molecules. The motivations is that averages over these distributions correspond to experimental measurements, and may be used to predict equilibria.

5.1 Developing the Boltzmann distribution law

Consider a system having N particles of a single type, and suppose the system has t different energy levels, $E_j, j=1,\dots,t$. We aim to compute the probabilities p_j that the system is in each level j. Suppose (T,V,N) are held constant. Then the condition for equilibrium is $\mathrm{d}F=\mathrm{d}U-T\,\mathrm{d}S=0$. We need to find $\mathrm{d}S$ and $\mathrm{d}U$.

$\mathrm{d}S$ is given by eq. (2.1):

$$\frac{S}{k}=-\sum_{j=1}^{t} p_j \ln p_j\,.$$

Differentiating with respect to p_j gives

$$\mathrm{d}S=-k\sum_{j=1}^{t}\left(1+\ln p_j\right)\mathrm{d}p_j\,. \tag{5.1}$$

$\mathrm{d}U$ is given by eq. (2.6):

$$U=\langle E\rangle=\sum_{j=1}^{t} p_j E_j\,.$$

Take the derivative:

$$\mathrm{d}U=\sum_{j=1}^{t}\left(E_j\mathrm{d}p_j+p_j\mathrm{d}E_j\right)\,. \tag{5.2}$$

Like the macroscopic energy U, the energy levels E_j depend on (V,N), but not on S or T. We take as a fundamental principle of quantum mechanics that only the populations $p_j(T)$, and not the energies E_j, depend on temperature. However, the *average energy* $\langle E\rangle$ does depend on the temperature. $\mathrm{d}E_j=(\partial E_j/\partial V)\,\mathrm{d}V+(\partial E_j/\partial N)\,\mathrm{d}N=0$ because both V and N are held constant. Thus,

$$\mathrm{d}\langle E\rangle=\sum_{j=1}^{t} E_j\mathrm{d}p_j\,. \tag{5.3}$$

The first law of thermodynamics gives $\mathrm{d}U=\delta q+\delta w$, which reduces to $\mathrm{d}\langle E\rangle=\mathrm{d}U=\delta q$ when V,N are constant. Because eq. (5.3) applies when V is constant, it follows that the term $\sum E_j\mathrm{d}p_j$ is the heat and $\sum p_j\mathrm{d}E_j$ is the work.

We want the probability distribution that satisfies the equilibrium condition $\mathrm{d}F=\mathrm{d}\langle E\rangle-T\,\mathrm{d}S=0$ subject to the normalization constraint, which can be expressed in terms of a Lagrange multiplier α:

$$\alpha\sum_{j=1}^{t}\mathrm{d}p_j=0. \tag{5.4}$$

Substitute eq. (5.1) and (5.2)–(5.4) into $\mathrm{d}F=\mathrm{d}U-T\,\mathrm{d}S=0$ to get

$$\mathrm{d}F=\sum_{j=1}^{t}\left[E_j+kT\left(1+\ln p_j^\star\right)+\alpha\right]\mathrm{d}p_j^\star=0. \tag{5.5}$$

According to the Lagrange multiplier equation (2.3), the term in the brackets must equal zero for each j, so we have t equations of the form

$$\ln p_j^\star = -\frac{E_j}{kT} - \frac{\alpha}{kT} - 1. \tag{5.6}$$

Exponentiate eq. (5.6) to find

$$p_j^\star = \mathrm{e}^{-E_j/kT}\,\mathrm{e}^{(-\alpha/kT)-1}\,. \tag{5.7}$$

To eliminate α, write the constraint equation

$$\sum_{j=1}^{t} p_j^\star = 1$$

as

$$1 = \sum_{j=1}^{t} \mathrm{e}^{-E_j/kT}\,\mathrm{e}^{(-\alpha/kT)-1}\,.$$

Divide eq. (5.7) by this form to get the *Boltzmann distribution law*

$$p_j^\star = \frac{\mathrm{e}^{-E_j/kT}}{\sum_{j=1}^{t}\mathrm{e}^{-E_j/kT}} = \frac{\mathrm{e}^{-E_j/kT}}{Q}, \tag{5.8}$$

where Q is the *partition function*,

$$Q \equiv \sum_{j=1}^{t} \mathrm{e}^{-E_j/kT}\,. \tag{5.9}$$

The relative populations of particles in energy levels i and j at equilibrium are given by

$$\frac{p_i^\star}{p_j^\star} = \mathrm{e}^{-(E_i-E_j)/kT}\,. \tag{5.10}$$

Note. Comparison of eq. (5.9), which we obtained by minimizing the free energy, with eq. (2.5), which we obtained by maximizing entropy subject to normalization constraint, shows that the Lagrange multiplier that enforces the constraint of average energy is $\beta = 1/(kT)$.

Note. The Boltzmann distribution says that more particles will have low energies and fewer particles will have high energies, because there are more rearrangements of the system that way. It is extremely unlikely that one particle would have such a high energy that it would leave all the others no energy. There are far more arrangements in which most particles have energies that are relatively low, but nonzero.

5.2 Properties of the partition function

The partition function is the connection between macroscopic thermodynamic properties and microscopic models. It is a sum of *Boltzmann factors* $\mathrm{e}^{-E_j/kT}$ that specify how particles are partitioned throughout the accessible states. eq. (5.9) gives

$$Q = \mathrm{e}^{-E_1/kT} + \mathrm{e}^{-E_2/kT} + \cdots + \mathrm{e}^{-E_t/kT}\,.$$

It is also common to express Q in an alternative form. Experiments usually give information in form of energy differences (rather than absolute energies). So it is often convenient to define the ground-state energy as zero, $E_1 = 0$, and re-write the partition function as follows:

$$Q = 1 + e^{-(E_2-E_1)/kT} + e^{-(E_3-E_1)/kT} + \cdots + e^{-(E_t-E_1)/kT} . \tag{5.11}$$

Another way to think about the partition function is as the number of states that are *effectively* accessible to the system. Look at limits of Q. when the energies are small, or the temperature is high, all the states become equally populated.

$$\begin{cases} E_j \to 0 \\ \text{or} \\ T \to \infty \end{cases} \quad \Longrightarrow \quad \frac{E_j}{kT} \to 0 \quad \Longrightarrow \quad p_j^\star \to \frac{1}{t} \quad \Longrightarrow \quad Q \to t . \tag{5.12}$$

In this case, all t states become accessible. At the other extreme, as the energy intervals become large or as the temperature approaches zero, the particles only occupy the ground state.

$$\begin{cases} E_{j\neq 1} \to \infty \\ \text{or} \\ T \to 0 \end{cases} \quad \Longrightarrow \quad \frac{E_{j\neq 1}}{kT} \to \infty \quad \Longrightarrow \quad \begin{cases} p_1^\star \to 1 \\ \text{and} \\ p_{j\neq 1}^\star \to 0 \end{cases} \quad \Longrightarrow \quad Q \to 1 . \tag{5.13}$$

In other words, only the ground state becomes accessible. The magnitude E_j/kT determines whether state j is “effectively accessible”. Therefore, kT may be used as a reference value. States that have energies $E_j > kT$ are relatively inaccessible at temperature T, while states with $E_j < kT$ are well populated. Increasing kT increases the *threshold* for effective population. Note that the term 'effective' is used, because the number of accessible states is always t, which is determined by the physics of the system.

5.2.1 Densities of states

Sometimes there is a different number of ways that a system can occupy one energy level than another. In such a case, we define $W(E)$ as the *density of states*—the total number of ways a system can occur in energy level E. When $W(E) > 1$, an energy level is called *degenerate*.

When we have a density of states, we can express the partition function as a sum over energy levels $\ell = 1, 2, \ldots, \ell_{\max}$:

$$Q = \sum_{\ell=1}^{\ell_{\max}} W(E_\ell)\, e^{-E_\ell/kT} . \tag{5.14}$$

The probability that a system is in **macrostate** energy level ℓ is

$$p_\ell = Q^{-1} W(E_\ell)\, e^{-E_\ell/kT} . \tag{5.15}$$

5.2.2 Partition functions for independent and distinguishable particles

The Boltzmann distribution law applies to system of any degree of complexity. In the simplest case, the particles do not interact, and therefore are considered independent. In this case, the **system** partition function is the product of **particle** partition functions.

Before we prove the statement above, we need to emphasize the difference between distinguishable particles and indistinguishable ones. For example, the atoms in a crystal are spatially distinguishable because each one has its own private location in the crystal over the timescale of a typical experiment. Its location serves as a marker. In contrast, the particles in a gas are indistinguishable. As they interchange locations, you can't tell which is which.

Consider distinguishable particles in a system with energy levels E_j. Suppose the system has two independent subsystems (e.g., two particles), labeled A and B, with energy levels ε_i^A and ε_m^B, where $i=1,2,\ldots,a$ and $m=1,2,\ldots,b$. The system energy is

$$E_j = \varepsilon_i^A + \varepsilon_m^B .$$

Because the subsystems are independent, we can write partition functions for each subsystem.

$$q_A = \sum_{i=1}^{a} \mathrm{e}^{-\varepsilon_i^A/kT}, \qquad q_B = \sum_{m=1}^{b} \mathrm{e}^{-\varepsilon_m^B/kT} . \tag{5.16}$$

The partition function Q for the entire system is the sum of Boltzmann factors over all $j=a\,b$ energy levels.

$$Q = \sum_{i=1}^{a}\sum_{m=1}^{b} \mathrm{e}^{-(\varepsilon_i^A+\varepsilon_m^B)/kT} = \sum_{i=1}^{a}\sum_{m=1}^{b} \mathrm{e}^{-\varepsilon_i^A/kT}\,\mathrm{e}^{-\varepsilon_m^B/kT} = q_A\,q_B . \tag{5.17}$$

More generally, for a system having N independent and distinguishable particles, each with partition function q, the partition function Q for the whole system is

$$Q = q^N . \tag{5.18}$$

5.2.3 Partition functions for independent and indistinguishable particles

For a system of two indistinguishable particles, the total energy is $E_j = \varepsilon_i + \varepsilon_m$, where $i=1,2,\ldots,t_1$ and $m=1,2,\ldots,t_2$. The system partition function is

$$Q = \sum_{j=1}^{t} \mathrm{e}^{-E_j/kT} = \sum_{i=1}^{t_1}\sum_{m=1}^{t_2} \mathrm{e}^{-(\varepsilon_i+\varepsilon_m)/kT} . \tag{5.19}$$

If one particle occupied energy level 27 and other particle occupied energy level 56, we could not distinguish that from the reverse. Because of this indistinguishability, we would have overcounted by a factor of 2!.

For this system, we have $Q=q^2/2!$, to a good approximation. (We are neglecting the case that both particles occupy the same energy level, which doesn't need a correction factor, but for a large number of available energy levels and limited amount of particles, the chance of the above event occurring is very small.) For N indistinguishable particles, the system partition function is

$$Q \approx \frac{q^N}{N!} . \tag{5.20}$$

5.3 Thermodynamic properties can be predicted from partition functions

5.3.1 Computing the internal energy

Consider a system having fixed (T,V,N). The internal energy for a system with energies E_j, substitute eq. (5.8) into eq. (2.6):

$$U = \sum_{j=1}^{t} p_j^\star E_j = Q^{-1}\sum_{j=1}^{t} E_j \mathrm{e}^{-\beta E_j}, \tag{5.21}$$

where $\beta = 1/kT$. Notice that it follows from eq. (5.9) that we can write

$$\left(\frac{\mathrm{d}Q}{\mathrm{d}\beta}\right) = \frac{\mathrm{d}}{\mathrm{d}\beta}\sum_{j=1}^{t} \mathrm{e}^{-\beta E_j} = -\sum_{j=1}^{t} E_j \mathrm{e}^{-\beta E_j}. \tag{5.22}$$

Substituting eq. (5.22) into (5.21) simplifies it:

$$U = -\frac{1}{Q}\left(\frac{\mathrm{d}Q}{\mathrm{d}\beta}\right) = -\left(\frac{\mathrm{d}\ln Q}{\mathrm{d}\beta}\right). \tag{5.23}$$

Since $\beta = 1/kT$, we have

$$\left(\frac{\mathrm{d}\beta}{\mathrm{d}T}\right) = -\frac{1}{kT^2}. \tag{5.24}$$

So we can multiply the left side of (5.23) by $-1/kT^2$ and the right side by $\mathrm{d}\beta/\mathrm{d}T$ to get

$$\frac{U}{kT^2} = \left(\frac{\mathrm{d}\ln Q}{\mathrm{d}T}\right). \tag{5.25}$$

A useful alternative expression is

$$\frac{U}{kT} = \frac{\mathrm{d}\ln Q}{\mathrm{d}\ln T} = \frac{T}{Q}\left(\frac{\mathrm{d}Q}{\mathrm{d}T}\right). \tag{5.26}$$

5.3.2 Computing the average particle energy

If particles are independent and distinguishable, the average energy $\langle \varepsilon \rangle$ **per particle** is

$$\langle \varepsilon \rangle = \frac{U}{N} = \frac{kT^2}{N}\left(\frac{\partial \ln q^N}{\partial T}\right)_{V,N} = kT^2\left(\frac{\partial \ln q}{\partial T}\right) = -\left(\frac{\partial \ln q}{\partial \beta}\right). \tag{5.27}$$

5.3.3 Computing the entropy

The entropy of a system is given by

$$\frac{S}{k} = -\sum_{j=1}^{t} p_j \ln p_j.$$

Substituting the Boltzmann distribution $p_j^\star = Q^{-1}\mathrm{e}^{-E_j/kT}$ gives

$$\frac{S}{k} = -\sum_{j=1}^{t}\left(\frac{1}{Q}\mathrm{e}^{-E_j/kT}\right)\left[\ln\left(\frac{1}{Q}\right) - \frac{E_j}{kT}\right]. \tag{5.28}$$

Substituting eq. (5.9) and (5.21) into (5.28) gives

$$S = k\ln Q + \frac{U}{T} = k\ln Q + kT\left(\frac{\partial \ln Q}{\partial T}\right). \tag{5.29}$$

For systems of N independent distinguishable particles ($Q = q^N$),

$$S = kN\ln q + \frac{U}{T}. \tag{5.30}$$

Because S increases linearly with N, the system entropy is the num of the entropies of the independent particles.

5.3.4 Computing the free energy and chemical potential

From U and S derived above, thermodynamic relationships can produce the rest—the Helmholtz free energy, chemical potential, and pressure.

$$F = U - TS = -kT \ln Q. \tag{5.31}$$

$$\mu = \left(\frac{\partial F}{\partial N}\right)_{T,V} = -kT \left(\frac{\partial \ln Q}{\partial N}\right)_{T,V}. \tag{5.32}$$

$$p = -\left(\frac{\partial F}{\partial V}\right)_{T,N} = kT \left(\frac{\partial \ln Q}{\partial V}\right)_{T,N}. \tag{5.33}$$

5.4 What is an ensemble?

A term commonly used in statistical mechanics is *ensemble*. The term is usually used in one of two ways. First, it can refer to which set of variables you are controlling: 'the (U,V,N) ensemble' or ' the (T,p,N) ensemble'.

- The (T,V,N) ensemble is called the *canonical ensemble*.
- The (U,V,N) ensemble is called the *microcanonical ensemble*.
- The (T,p,N) ensemble is called the *isobaric-isothermal ensemble*.
- The (T,V,μ) ensemble is called the *grand canonical ensemble*.

The term ensemble also has another meaning. An ensemble is the collection of all the possible microstates, or snapshots, of a system.

5.4.1 The microcanonical ensemble

The microcanonical ensemble is qualitatively different from the canonical and grand canonical ensembles. In the canonical ensemble, the temperature is fixed, which is equivalent to fixing the average energy $U = \langle E \rangle$. The energy can fluctuate. But in the microcanonical ensemble, every microstate has exactly the same fixed energy, so $U = E$, and there are no fluctuations.

For the microcanonical ensemble, it is more useful to focus on the $i = 1,2,...,W(E,V,N)$ microstates of the system, than of t energy levels, since there is only one energy level. In the microcanonical ensemble, each microstate is equivalent. So you can express the probability that the system is in microstate $i = 1,2,\ldots,W$ as

$$p_i^\star = \frac{1}{W}. \tag{5.34}$$

Using the definition of the entropy, we have

$$\frac{S}{k} = -\sum_{i=1}^{W} p_i \ln p_i = -\sum_{i=1}^{W} \left(\frac{1}{W}\right) \ln\left(\frac{1}{W}\right) = \ln W(E,V,N). \tag{5.35}$$