# 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.

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

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

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

   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.

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

**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

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

*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, $\sigma^2$ 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* (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 assumes this particular mathematical form? The multiplicity could be maximized as $W^2, 15\, W^3$ 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 $n \gg 1$,

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

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

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

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

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

**Note.** 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.

**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.

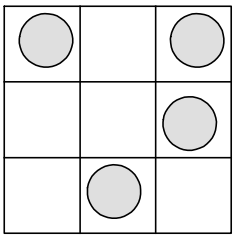

**Figure 2.1.** Each particle may occupy a lattice site.

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,

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

As entropy is extensive,

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

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 $n = m$.

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

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?

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

Obviously, mixing increased the total entropy of the system.

## 2.2 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, $\Delta x = (x - a) \longrightarrow \mathrm{d}x$, non-linear terms in the series expansions are neligible, 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, $f(x, y)$,

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

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 $i = 1, 2, \ldots, t$.

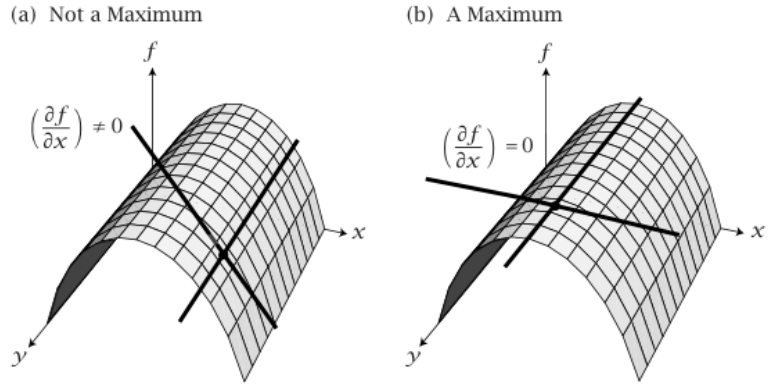


**Figure 2.2.** To identify the maximum of this function, both $(\partial f/\partial x)$ and $(\partial f/\partial y)$ must equal zero, as they do in (b), but not in (a).

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 constaint 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,

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

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 $g(x, y) = \text{constant}$.

### 2.2.1 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 arbitraty constant $\lambda$, 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

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

with the constraint equation

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

Use Lagrange multipliers to maximize $f$ subject to $g$:

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

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

Substituting these into the constraint equation and solving for $\lambda$ 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:

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

### 2.2.2 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.

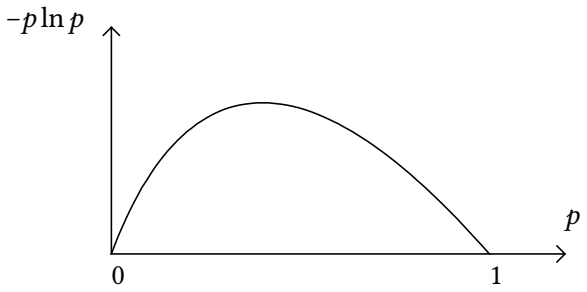


**Figure 2.3.** $-p\ln p$ versus $p$ has a maximum.

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 partical 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),

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

Plug these to the constraint equation to get

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

**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.

### 2.2.3 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 thte 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 $\varepsilon_i$. 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 $\langle\varepsilon\rangle$.

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

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

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 evalueated 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 $\alpha$, 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

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